

# Handbook of Philosophical Logic

2nd Edition

Volume 1

edited by Dov M. Gabbay and F. Guentner



## CONTENTS

Editorial Preface <b>Dov M. Gabbay</b>	vii
Elementary Predicate Logic <b>Wilfrid Hodges</b>	1
Systems Between First- and Second-order Logic <b>Stewart Shapiro</b>	131
Higher-Order Logic <b>Johan van Benthem and Kees Doets</b>	189
Algorithms and Decision Problems: A Crash Course in Recursion Theory <b>Dirk van Dalen</b>	245
Mathematics of Logic Programming <b>Hans Dieter Ebbinghaus and Jörg Flum</b>	313
Index	371





## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good.!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London

Logic	IT			
	Natural language processing	Program control specification, verification, concurrency	Artificial intelligence	Logic programming
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs

<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical fragments</b>	Basic back-ground language	Program synthesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely useful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calculus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syntax	Modules. Combining languages	Logics of space and time	Combining features
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game semantics gaining ground		
<b>Object level/metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action-revision models</b>			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model





WILFRID HODGES

## ELEMENTARY PREDICATE LOGIC

### INTRODUCTION

Elementary (first-order) predicate logic is a child of many parents. At least three different groups of thinkers played their part in its conception, with three quite distinct motives. Maybe the mixture gave it hybrid strength. But whatever the reason, first-order logic is both the simplest, the most powerful and the most applicable branch of modern logic.

The first group who can claim paternity are the *Traditional Logicians*. For these scholars the central aim of logic was to schematise valid arguments. For present purposes an argument consists of a string of sentences called *premises*, followed by the word ‘*Therefore*’, followed by a single sentence called the *conclusion*. An argument is called *valid* when its premises *entail* its conclusion, in other words, if the premises can’t be true without the conclusion also being true.

A typical valid argument schema might be:

1.  $a$  is more  $X$  than  $b$ .  $b$  is more  $X$  than  $c$ .  
*Therefore*  $a$  is more  $X$  than  $c$ .

This becomes a valid argument whenever we substitute names for  $a, b, c$  respectively and an adjective for  $X$ ; as for example

2. Oslo is more clean than Ydstebøhavn. Ydstebøhavn is more clean than Trondheim. *Therefore* Oslo is more clean than Trondheim.

Arguments like (2) which result from such substitutions are called *instances* of the schema (1). Traditional logicians collected valid argument schemas such as (1). This activity used to be known as *formal logic* on the grounds that it was concerned with the forms of arguments. (Today we more often speak of formal versus informal logic, just as formal versus informal semantics, meaning mathematically precise versus mathematically imprecise.)

The ancients and the medievals had concerned themselves with small numbers of argument schemas gathered more or less *ad hoc*. Aristotle’s syllogisms give twenty-four schemas, of which Aristotle himself mentions nineteen. The watershed between classical and modern logic lies in 1847, when George Boole (1815–1864) published a calculus which yielded infinitely many valid argument schemas of arbitrarily high complexity (Boole [1847; 1854]). Today we know Boole’s calculus as *propositional logic*. Other early researchers who belong among the Traditionals are Augustus De Morgan (1806–1871) and C. S. Peirce (1839–1914). Their writings are lively with

examples of people  $i$  being enemies to people  $j$  at time  $k$ , and other people overdrawing their bank accounts.

The second group of originators were the *Proof Theorists*. Among these should be included Gottlob Frege (1848–1925), Giuseppe Peano (1858–1932), David Hilbert (1862–1943), Bertrand Russell (1872–1970), Jacques Herbrand (1908–1931) and Gerhard Gentzen (1909–1945). Their aim was to systematise mathematical reasoning so that all assumptions were made explicit and all steps rigorous. For Frege this was a matter of integrity and mental hygiene. For Hilbert the aim was to make mathematical reasoning itself the object of mathematical study, partly in order to justify infinitary mathematics but partly also as a new method of mathematical research. This group devised both the notation and the proof theory of first-order logic. The earliest calculus adequate for first-order logic was the system which Frege published in his *Begriffsschrift* [1879]. This was also the first work to discuss quantifiers.

With a slight anachronism I call the third group the *Model Theorists*. Their aim was to study mathematical structures from the point of view of the laws which these structures obey. The group includes Ernst Schröder (1841–1902), Leopold Löwenheim (1878–1957), Thoralf Skolem (1887–1963), C. H. Langford (1895?–1964), Kurt Gödel (1906–1978) and Alfred Tarski (1901–1983). The notion of a first-order property is already clear in Schröder’s work [1895], though the earliest use I could find of the term ‘first-order’ in the modern sense is in Langford [1927]. (Langford quotes the slightly different use of the term *Principia Mathematica*, Whitehead and Russell [1910].)

Our present understanding of what first-order logic is about was painstakingly built up by this group of workers during the years 1915 to 1935. The progress was conceptual as much as technical; a historian of logic feels his fingers tingle as he watches it. Increasing precision was an important part of it. But it is worth reflecting that by 1935 a logician could safely say ‘The formal sentence  $S$  is true in the structure  $A$ ’ and *mean it*. Frege [1906] had found such language morally reprehensible (cf. Section 12 below). Skolem [1922] talked of formal axioms ‘holding in a domain’, but he felt obliged to add that this was ‘only a manner of speaking, which can lead only to purely formal propositions—perhaps made up of very beautiful *words*. . .’. (On taking truth literally, see above all Kurt Gödel’s letters to Hao Wang, [1974, p. 8 ff] and the analysis by Solomon Feferman [1984]. R. L. Vaught’s historical paper [1974] is also valuable.)

Other groups with other aims have arisen more recently and found first-order logic helpful for their purposes. Let me mention two.

One group (if we can lump together such a vast army of workers) are the computer scientists. There is wide agreement that trainee computer scientists need to study logic, and a range of textbooks have come onto the market aimed specifically at them. (To mention just two, Reeves and

Clarke [1990] is an introductory text and Gallier [1986] is more advanced.) But this is mainly for training; first-order logic itself is not the logic of choice for many computer science applications. The artificial intelligence community consume logics on a grand scale, but they tend to prefer logics which are modal or intensional. By and large, specification languages need to be able to define functions, and this forces them to incorporate some higher-order features. Very often the structures which concern a computer scientist are finite, and (as Yuri Gurevich [1984] argued) first-order logic seems not to be the best logic for classifying finite structures.

Computer science has raised several questions which cast fresh light on first-order logic. For example, how does one search for a proof? The question itself is not new—philosophers from Aristotle to Leibniz considered it. What is completely new is the mathematical analysis of systematic searches through all possible proofs in a formal calculus. Searches of this kind arise naturally in automated theorem proving. Robert Kowalski [1979] proposed that one could read some first-order sentences as instructions to search for a proof; the standard interpretation of the programming language PROLOG rests on his idea. Another question is the cost of a formal proof, in terms of the number of assumptions which are needed and the number of times each assumption is used; this line of enquiry has led to fragments of first-order logic in which one has some control over the cost (see for example Jean-Yves Girard [1987; 1995] on linear logic and Došen and Schroeder-Heister [1993] on substructural logics in general).

Last but in no way least come the linguists. After Chomsky had revolutionised the study of syntax of natural languages in the 1950s and 60s, many linguists shifted the spotlight from grammar to meaning. It was natural to presume that the meaning of a sentence in a natural language is built up from the meanings of its component words in a way which reflects the grammatical structure of the sentence. The problem then is to describe the structure of meanings. One can see the beginnings of this enterprise in Bertrand Russell's theory of propositions and the 'logical forms' beloved of English philosophers earlier in this century; but the aims of these early investigations were not often clearly articulated. Round about 1970 the *generative semanticists* (we may cite G. Lakoff and J. D. McCawley) began to use apparatus from first-order logic in their analyses of natural language sentences; some of their analyses looked very much like the formulas which an up-to-date Traditional Logician might write down in the course of knocking arguments into tractable forms. Then Richard Montague [1974] opened a fruitful line of research by using tools from logic to give extremely precise analyses of both the grammar and semantics of some fragments of English. (Cf. Dowty *et al.* [1981] for an introduction to Montague grammar.) I should add that many researchers on natural language semantics, from Montague onwards, have found that they needed logical devices which go far beyond first-order logic. More recently some of the apparatus of first-

order proof theory has turned up unexpectedly in the analysis of grammar; see for example Morrill [1994] and Kempson [1995].

Logicians like to debate over coffee when ‘real’ first-order logic first appeared in print. The earliest textbook account was in the *Grundzüge der theoretischen Logik* of Hilbert and Ackermann [1928], based on Hilbert’s lectures of 1917–1922. Skolem’s paper [1920] is undeniably about first-order logic. But Whitehead and Russell’s *Principia Mathematica* [1910] belongs to an earlier era. It contains notation, axioms and theorems which we now regard as part of first-order logic, and for this reason it was quoted as a reference by Post, Langford, Herbrand and Gödel up to 1931, when it figured in the title of Gödel’s famous paper on incompleteness, [Gödel, 1931b]. But the first-order part of *Principia* is not distinguished from the rest; and more important, its authors had no notion of a precise syntax or the interpretation of formulas in structures.

## I: Propositional Logic

### 1 TRUTH FUNCTORS

In propositional logic we use six artificial symbols  $\neg, \wedge, \vee, \rightarrow, \leftrightarrow, \perp$ , called *truth-functors*. These symbols all have agreed meanings. They can be used in English, or they can have an artificial language built around them.

Let me explain one of these symbols,  $\wedge$ , quite carefully. The remainder will then be easy.

We use  $\wedge$  between sentences  $\phi, \psi$  to form a new sentence

$$(1) \quad (\phi \wedge \psi).$$

The brackets are an essential part of the notation. Here and below, ‘sentence’ means ‘indicative sentence’. If  $\phi$  and  $\psi$  are sentences, then in any situation,

$$(2) \quad (\phi \wedge \psi) \text{ is true iff } \phi \text{ is true and } \psi \text{ is true; otherwise it is false.}$$

(‘Iff’ means ‘if and only if’.) This defines the meaning of  $\wedge$ .

Several points about this definition call for comment. First, we had to mention the situation, because a sentence can be true in one situation and not true in another. For example, the sentence may contain demonstrative pronouns or other indexicals that need to be given a reference, or words that need to be disambiguated. (The situation is not necessarily the ‘context of utterance’—a sentence can be true in situations where it is never uttered.)

In propositional logic we assume that in every situation, each sentence under discussion is determinately either true or false and not both. This assumption is completely innocent. We can make it correct by adopting

either or both of the following conventions. First, we can agree that although we intend to use the word ‘true’ as it is normally used, we shall take ‘false’ to mean simply ‘not true’. And second, we can take it as understood that the term ‘situation’ covers only situations in which the relevant sentences are either true or false and not both. (We may also wish to put an embargo on nonsensical sentences, but this is not necessary.) There are of course several ways of being not true, but propositional logic doesn’t distinguish between them.

Logicians always make one further assumption here: they assume that truth and falsehood— $T$  and  $F$  for short—are objects. Then they say that the *truth-value* of a sentence is  $T$  if the sentence is true, and  $F$  otherwise. (Frege [1912]: ‘... in logic we have only two objects, in the first place: the two truth-values.’) But I think in fact even the most scrupulous sceptic could follow the literature if he *defined* the truth-value of all true sentences to be his left big toe and that of false sentences to be his right. Many writers take truth to be the number 1, which they identify with the set  $\{0\}$ , and falsehood to be the number 0, which is identified with the empty set. Nobody is obliged to follow these choices, but technically they are very convenient. For example (2) says that if the truth-value of  $\phi$  is  $x$  and the truth-value of  $\psi$  is  $y$ , then that of  $(\phi \wedge \psi)$  is  $xy$ .

With this notation, the definition (2) of the meaning of  $\wedge$  can be written in a self-explanatory chart:

(3)	$\phi$	$\psi$	$(\phi \wedge \psi)$
	$T$	$T$	$T$
	$T$	$F$	$F$
	$F$	$T$	$F$
	$F$	$F$	$F$

The diagram (3) is called the *truth-table* of  $\wedge$ . Truth-tables were first introduced by C. S. Peirce in [1902].

Does (3) really define the meaning of  $\wedge$ ? Couldn’t there be two symbols  $\wedge_1$  and  $\wedge_2$  with different meanings, which both satisfied (3)?

The answer is that there certainly can be. For example, if  $\wedge_1$  is any symbol whose meaning agrees with (3), then we can introduce another such symbol  $\wedge_2$  by declaring that  $(\phi \wedge_2 \psi)$  shall mean the same as the sentence

(4)  $(\phi \wedge_1 \psi)$  and the number  $\pi$  is irrational.

(Wittgenstein [1910] said that  $\wedge_1$  and  $\wedge_2$  then mean the same! *Tractatus* 4.46ff, 4.465 in particular.) But this is the wrong way to read (3). Diagram (3) should be read as stating *what one has to check in order to determine that  $(\phi \wedge \psi)$  is true*. One can verify that  $(\phi \wedge \psi)$  is true without knowing that  $\pi$  is irrational, but not without verifying that  $\phi$  and  $\psi$  are true. (See

Michael Dummett [1958/59; 1975] on the relation between meaning and truth-conditions.)

Some logicians have claimed that the sentence  $(\phi \wedge \psi)$  means the same as the sentence

(5)  $\phi$  and  $\psi$ .

Is this correct? Obviously the meanings are very close. But there are some apparent differences. For example, consider Mr Slippery who said in a court of law:

(6) I heard a shot and I saw the girl fall.

when the facts are that he saw the girl fall and *then* heard the shot. Under these circumstances

(7) (I heard a shot  $\wedge$  I saw the girl fall)

was true, but Mr Slippery could still get himself locked up for perjury. One might maintain that (6) does mean the same as (7) and was equally true, but that the conventions of normal discourse would have led Mr Slippery to choose a different sentence from (6) if he had not wanted to mislead the jury. (See Grice [1975] for these conventions; Cohen [1971] discusses the connection with truth-tables.)

Assuming, then, that the truth-table (3) does adequately define the meaning of  $\wedge$ , we can define the meanings of the remaining truth-functors in the same way. For convenience I repeat the table for  $\wedge$ .

(8)	$\phi$	$\psi$	$\neg\phi$	$\phi \wedge \psi$	$\phi \vee \psi$	$\phi \rightarrow \psi$	$\phi \leftrightarrow \psi$	$\perp$
	$T$	$T$	$F$	$T$	$T$	$T$	$T$	$F$
	$T$	$F$		$F$	$T$	$F$	$F$	
	$F$	$T$	$T$	$F$	$T$	$T$	$F$	
	$F$	$F$		$F$	$F$	$T$	$T$	

$\neg\phi$  is read 'Not  $\phi$ ' and called the *negation* of  $\phi$ .  $(\phi \wedge \psi)$  is read ' $\phi$  and  $\psi$ ' and called the *conjunction* of  $\phi$  and  $\psi$ , with *conjuncts*  $\phi$  and  $\psi$ .  $(\phi \vee \psi)$  is read ' $\phi$  or  $\psi$ ' and called the *disjunction* of  $\phi$  and  $\psi$ , with *disjuncts*  $\phi$  and  $\psi$ .  $(\phi \rightarrow \psi)$  is read 'If  $\phi$  then  $\psi$ ' or ' $\phi$  arrow  $\psi$ '; it is called a *material implication* with *antecedent*  $\phi$  and *consequent*  $\psi$ .  $(\phi \leftrightarrow \psi)$  is read ' $\phi$  if and only if  $\psi$ ', and is called the *biconditional* of  $\phi$  and  $\psi$ . The symbol  $\perp$  is read as 'absurdity', and it forms a sentence by itself; this sentence is false in all situations.

There are some alternative notations in common use; for example

(9)  $-\phi$  or  $\sim\phi$  for  $\neg\phi$ .  
 $(\phi\&\psi)$  for  $(\phi \wedge \psi)$ .  
 $(\phi \supset \psi)$  for  $(\phi \rightarrow \psi)$ .  
 $(\phi \equiv \psi)$  for  $(\phi \leftrightarrow \psi)$ .

Also the truth-functor symbols are often used for other purposes. For example the intuitionists use the symbols  $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$  but not with the meanings given in (8); cf. van Dalen's chapter on Intuitionistic Logic in a later volume. Some writers use the symbol  $\rightarrow$  for other kinds of implication, or even as a shorthand for the English words 'If ... then'.

*A remark on metavariables.* The symbols ' $\phi$ ' and ' $\psi$ ' are not themselves sentences and are not the names of particular sentences. They are used as above, for making statements about any and all sentences. Symbols used in this way are called (*sentence*) *metavariables*. They are part of the *metalanguage*, i.e. the language we use for talking about formulas. I follow the convention that when we talk about a formula, symbols which are not metavariables are used as names for themselves. So for example the expression in line (1) means the same as: the formula consisting of '(' followed by  $\phi$  followed by ' $\wedge$ ' followed by  $\psi$  followed by ')'. I use quotation marks only when clarity or style demand them. These conventions, which are normal in mathematical writing, cut down the clutter but put some obligation on reader and writer to watch for ambiguities and be sensible about them. Sometimes a more rigorous convention is needed. Quine's corners  $\ulcorner \urcorner$  supply one; see Quine [1940, Section 6]. There are some more remarks about notation in Section 4 below.

## 2 PROPOSITIONAL ARGUMENTS

Besides the truth-functors, propositional logic uses a second kind of symbol, namely the *sentence letters*

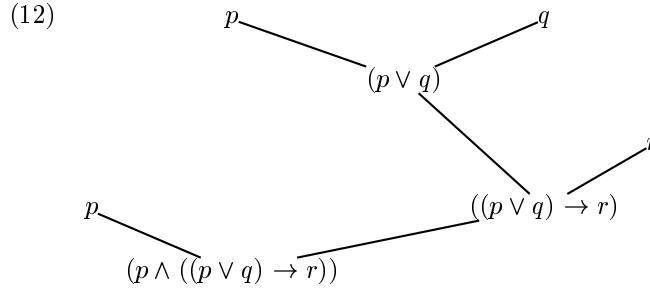
$$(10) \ p, q, r, \dots, p_1, p_2, \dots,$$

These letters have no fixed meaning. They serve to mark spaces where English sentences can be written. We can combine them with the truth-functors to produce expressions called *formulas*, which become sentences when the sentence letters are replaced by sentences.

For example, from the sentence letters  $p, q$  and  $r$  we can build up the formula

$$(11) \ (p \wedge ((p \vee q) \rightarrow r))$$

as follows:



We call (12) the *formation tree* of the formula (11). Sentence letters themselves are reckoned to be *atomic formulas*, while formulas which use truth-functores are called *compound formulas*. In a compound formula there is always a truth-functor which was added last in the formation tree; this occurrence of the truth-functor is called the *main connective* of the formula. In (11) the main connective is the occurrence of  $\wedge$ . The main connective of  $\perp$  is reckoned to be  $\perp$  itself.

Suppose  $\phi$  is a formula. An *instance* of  $\phi$  is a sentence which is got from  $\phi$  by replacing each sentence letter in  $\phi$  by an English sentence, in such a way that no sentence letter gets replaced by different sentences at different occurrences. (Henceforth, the symbols ' $\phi$ ', ' $\psi$ ' are metavariables for formulas as well as sentences. The letters ' $p$ ', ' $q$ ' etc. are not metavariables; they are the actual symbols of propositional logic.)

Now if we know the truth-values of the inserted sentences in an instance of  $\phi$ , then we can work out by table (8) what the truth-value of the whole instance must be. Taking (11) as an example, consider the following table:

(13)

	$p$	$q$	$r$	$(p \wedge ((p \vee q) \rightarrow r))$		
(i)	$T$	$T$	$T$	$TT$	$TTT$	$TT$
(ii)	$T$	$T$	$F$	$TF$	$TTT$	$FF$
(iii)	$T$	$F$	$T$	$TT$	$TTF$	$TT$
(iv)	$T$	$F$	$F$	$TF$	$TTF$	$FF$
(v)	$F$	$T$	$T$	$FF$	$FTT$	$TT$
(vi)	$F$	$T$	$F$	$FF$	$FTT$	$FF$
(vii)	$F$	$F$	$T$	$FF$	$FFF$	$TT$
(viii)	$F$	$F$	$F$	$FF$	$FFF$	$TF$
				1 7	2 5 3	6 4

The rows (i)–(viii) on the left list all the possible ways in which the sentences put for  $p$  and  $q$  can have truth-values. The columns on the right are computed in the order shown by the numbers at the bottom. (The numbers at left and bottom are not normally written—I put them in to help the explanation.) Columns 1, 2, 3, 4 just repeat the columns on the left. Column 5 shows the truth-value of  $(p \vee q)$ , and is calculated from columns 2 and 3 by means of table (8). Then column 6 is worked out from columns 5 and



4, using the truth-table for  $(\phi \rightarrow \psi)$  in (8). Finally, column 7 comes from columns 1 and 6 by the table for  $(\phi \wedge \psi)$ . Column 7 is written under the main connective of (11) and shows the truth-value of the whole instance of (11) under each of the eight possibilities listed on the left.

Table (13) is called the *truth-table* of the formula (11). As we constructed it, we were working out truth-tables for all the formulas shown in the formation tree (12), starting at the top and working downwards.

We are now equipped to use propositional logic to prove the validity of an argument. Consider:

- (14) That was a hornet, and soda only makes hornet and wasp stings worse. So you don't want to use soda.

This contains an argument along the following lines:

- (15) (You were stung by a hornet  $\wedge$  ((you were stung by a hornet  $\vee$  you were stung by a wasp)  $\rightarrow$  soda will make the sting worse)).  
Therefore soda will make the sting worse.

We replace the component sentences by letters according to the scheme:

- (16)  $p$  : You were stung by a hornet.  
 $q$  : You were stung by a wasp.  
 $r$  : Soda will make the sting worse.

The result is:

- (17)  $(p \wedge ((p \vee q) \rightarrow r))$ . Therefore  $r$ .

Then we calculate truth-tables for both premise and conclusion of (17) at the same time. Only the main columns are shown below.

(18)	$p$	$q$	$r$	$(p \wedge ((p \vee q) \rightarrow r))$ .	Therefore $r$
(i)	$T$	$T$	$T$	$T$	$T$
(ii)	$T$	$T$	$F$	$F$	$F$
(iii)	$T$	$F$	$T$	$T$	$T$
(iv)	$T$	$F$	$F$	$F$	$F$
(v)	$F$	$T$	$T$	$F$	$T$
(vi)	$F$	$T$	$F$	$F$	$F$
(vii)	$F$	$F$	$T$	$F$	$T$
(viii)	$F$	$F$	$F$	$F$	$F$

Table (18) shows that if the premise of (15) is true then so is the conclusion. For if the premise is true, then the column under the premise shows that we are in row (i) or row (iii). In both of these rows, the last column in (18) shows that the conclusion is true. There is no row which has a  $T$  below  $(p \wedge ((p \vee q) \rightarrow r))$  and an  $F$  below  $r$ . Hence, (15) is valid.

In the language of the traditional logician, these calculations showed that (17) is a valid argument schema. Every instance of (17) is a valid argument.

Note how the proof of the validity of an argument falls into two parts. The first is to translate the argument into the symbols of propositional logic. This involves no calculation, though a gauche translation can frustrate the second part. I say no more about this first part—the elementary textbooks give hundreds of examples [Kalish and Montague, 1964; Mates, 1965; Thomason, 1970; Hodges, 1977]. The second part of the proof is pure mechanical calculation using the truth-table definitions of the truth-functors. What remains to discuss below is the theory behind this mechanical part.

First and foremost, why does it work?

### 3 WHY TRUTH-TABLES WORK

*If  $\phi$  is any formula of propositional logic, then any assignment of truth-values to the sentence letters which occur in  $\phi$  can be extended, by means of the truth-table definitions of the truth-functors, to give a truth-value to  $\phi$ ; this truth-value assigned to  $\phi$  is uniquely determined and it can be computed mechanically.*

This is the central thesis of propositional logic. In Section 2 I showed how the assignment to  $\phi$  is calculated, with an example. But we shouldn't rest satisfied until we see, first, that this procedure *must always work*, and second, that the outcome is *uniquely determined by the truth-table definitions*. Now there are infinitely many formulas  $\phi$  to be considered. Hence we have no hope of setting out all the possibilities on a page; we need to invoke some abstract principle to see why the thesis is true.

There is no doubt what principle has to be invoked. It is the principle of *induction on the natural numbers*, otherwise called *mathematical induction*. This principle says the following:

- (19) Suppose that the number 0 has a certain property, and suppose also that whenever all numbers from 0 to  $n$  inclusive have the property,  $n + 1$  must also have the property. Then all natural numbers from 0 upwards have the property.

This principle can be put in several forms; the form above is called *course-of-values induction*. (See Appendix B below.) For the moment we shall only be using one or two highly specific instances of it, where the property in question is a mechanically checkable property of arrays of symbols. Several writers have maintained that one knows the truth of any such instance of (19) by a kind of inspection (*Anschauung*). (See for example [Herbrand, 1930, Introduction] and [Hilbert, 1923]. There is a discussion of the point in [Steiner, 1975].)

Essentially what we have to do is to tie a number  $n$  to each formula  $\phi$ , calling  $n$  the *complexity* of  $\phi$ , so that we can then use induction to prove:

- (20) For each number  $n$  from 0 upwards, the thesis stated at the beginning of this section is true for all formulas of complexity  $n$ .

There are several ways of carrying this through, but they all rest on the same idea, namely this: *all formulas are generated from atomic formulas in a finite number of steps and in a unique way; therefore each formula can be assigned a complexity which is greater than the complexities assigned to any formulas that went into the making of it.* It was Emil Post, one of the founders of formal language theory, who first showed the importance of this idea in his paper on truth-tables:

- (21) “It is desirable in what follows to have before us the vision of the totality of these [formulas] streaming out from the unmodified [sentence letters] through forms of ever-growing complexity ...”  
(Post [1921], p. 266 of van Heijenoort [1967]).

For an exact definition of formulas and their complexities, we need to say precisely what sentence letters we are using. But it would be a pity to lumber ourselves with a set of letters that was inconvenient for some future purposes. So we adopt a compromise. Let  $X$  be any set of symbols to be used as sentence letters. Then we shall define the *propositional language of similarity type  $X$* , in symbols  $L(X)$ . The set  $X$  is not fixed in advance; but as soon as it is fixed, the definition of  $L(X)$  becomes completely precise. This is the usual modern practice.

The notions ‘formula of similarity type  $X$ ’ (we say ‘formula’ for short) and ‘complexity of a formula’ are defined as follows.

1. Every symbol in  $X$  is a formula of complexity 0.  $\perp$  is a formula of complexity 1.
2. If  $\phi$  and  $\psi$  are formulas of complexities  $m$  and  $n$  respectively, then  $\neg\phi$  is a formula with complexity  $m + 1$ , and  $(\phi \wedge \psi)$ ,  $(\phi \vee \psi)$ ,  $(\phi \rightarrow \psi)$  and  $(\phi \leftrightarrow \psi)$  are formulas of complexity  $m + n + 1$ .
3. Nothing is a formula except as required by (1) and (2).

For definiteness the *language of similarity type  $X$* ,  $L(X)$ , can be defined as the ordered pair  $\langle X, F \rangle$  where  $F$  is the set of all formulas of similarity type  $X$ . A *propositional language* is a language  $L(X)$  where  $X$  is a set of symbols; the *formulas of  $L(X)$*  are the formulas of similarity type  $X$ .

Frege would have asked: How do we know there is a unique notion ‘formula of similarity type  $X$ ’ with the properties (1)–(3)? A full answer to this question lies in the theory of inductive definitions; cf. Appendix B below. But for the present it will be enough to note that by (1) and (2), every formation tree has a formula as its bottom line, and conversely by (3) every formula is the bottom line of a formation tree. We can prove rigorously by induction that if a formula has complexity  $n$  by definition (1)–(3) then it

can't also have complexity  $m$  where  $m \neq n$ . This is actually not trivial. It depends on showing that the main connective in a compound formula is uniquely determined, and—ignoring  $\neg$  and  $\perp$  for simplicity—we can do that by showing that the main connective is the only truth-functor occurrence which has one more ‘(’ than ‘)’ to the left of it. (Cf. [Kleene, 1952, pp. 21ff].) The proof shows at the same time that every formula has a unique formation tree.

The atomic formulas are those which have complexity 0. A formula is called *basic* if it is either atomic or the negation of an atomic formula.

Now that the language has been adequately formulated, we come back to truth-tables. Let  $L$  be a propositional language with similarity type  $X$ . Then we define an *L-structure* to be a function from the set  $X$  to the set  $\{T, F\}$  of truth-values. (Set-theoretic notions such as ‘function’ are defined in Appendix C below, or in any elementary textbook of set theory.) So an L-structure assigns a truth-value to each sentence letter of  $L$ . For each sentence letter  $\phi$  we write  $I_{\mathfrak{A}}(\phi)$  for the truth-value assigned to  $\phi$  by the L-structure  $\mathfrak{A}$ . In a truth-table where the sentence letters of  $L$  are listed at top left, each row on the left will describe an L-structure, and every L-structure corresponds to just one row of the table.

Now we shall define when a formula  $\phi$  of  $L$  is *true in* an L-structure  $\mathfrak{A}$ , or in symbols

$$(22) \quad \mathfrak{A} \models \phi.$$

The definition of (22) will be by induction of the complexity of  $\phi$ . This means that when  $\phi$  has low complexity, the truth or falsity of (22) will be determined outright; when  $\phi$  has higher complexity the truth of (22) depends in an unambiguous way on the truth of statements ‘ $\mathfrak{A} \models \psi$ ’ for formulas  $\psi$  of lower complexity than  $\phi$ . (Cf. Appendix B.) We can prove by induction on the natural numbers that this definition determines exactly when (22) is true, and in fact that the truth or otherwise of (22) can be calculated mechanically once we know what  $\mathfrak{A}$  and  $\phi$  are. The definition is as follows:

$$(23) \quad \text{For each sentence letter } \phi, \mathfrak{A} \models \phi \text{ iff } I_{\mathfrak{A}}(\phi) = T.$$

It is false that  $\mathfrak{A} \models \perp$ .

For all formulas  $\phi, \psi$  of  $L$ ,

$\mathfrak{A} \models \neg\phi$  if it is not true that  $\mathfrak{A} \models \phi$ ;

$\mathfrak{A} \models (\phi \wedge \psi)$  iff  $\mathfrak{A} \models \phi$  and  $\mathfrak{A} \models \psi$ ;

$\mathfrak{A} \models (\phi \vee \psi)$  iff either  $\mathfrak{A} \models \phi$  or  $\mathfrak{A} \models \psi$  or both;

$\mathfrak{A} \models (\phi \rightarrow \psi)$  iff not:  $\mathfrak{A} \models \phi$  but not  $\mathfrak{A} \models \psi$ .

$\mathfrak{A} \models (\phi \leftrightarrow \psi)$  iff either  $\mathfrak{A} \models \phi$  and  $\mathfrak{A} \models \psi$ , or neither  $\mathfrak{A} \models \phi$  nor  $\mathfrak{A} \models \psi$ .

Definition (23) is known as the *truth definition* for the language  $L$ . The statement ‘ $\mathfrak{A} \models \phi$ ’ is sometimes read as:  $\mathfrak{A}$  is a *model of*  $\phi$ .

The reader can verify that (23) matches the truth-table definitions of the truth-functors, in the following sense. The left-hand part of any row of a truth-table for  $\phi$  describes an L-structure  $\mathfrak{A}$  (for some appropriate language L). The truth-table gives  $\phi$  the value  $T$  in this row if and only if  $\mathfrak{A} \models \phi$ ; moreover the steps by which we calculated this value for  $\phi$  in the table exactly match the steps by which the definition (23) above determines whether  $\mathfrak{A} \models \phi$ . In this sense, and only in this sense, (23) is a correct ‘definition of truth for L’. Nobody claims that (23) explains what is meant by the word ‘true’.

I should mention a useful piece of notation. We can write  $\|\phi\|_{\mathfrak{A}}$  for the truth-value assigned to the formula  $\phi$  by the structure  $\mathfrak{A}$ . Then  $\|\phi\|_{\mathfrak{A}}$  can be defined in terms of  $\models$  by:

$$(24) \quad \|\phi\|_{\mathfrak{A}} = \begin{cases} T & \text{if } \mathfrak{A} \models \phi, \\ F & \text{otherwise.} \end{cases}$$

Some writers prefer to define  $\|\cdot\|_{\mathfrak{A}}$  directly, and then  $\models$  in terms of  $\|\cdot\|_{\mathfrak{A}}$ . If we write 1 for  $T$  and 0 for  $F$ , an inductive definition of  $\|\cdot\|_{\mathfrak{A}}$  will contain clauses such as

$$(25) \quad \|\neg\phi\|_{\mathfrak{A}} = 1 - \|\phi\|_{\mathfrak{A}}; \quad \|(\phi \vee \psi)\|_{\mathfrak{A}} = \max \{\|\phi\|_{\mathfrak{A}}, \|\psi\|_{\mathfrak{A}}\}.$$

#### 4 SOME POINTS OF NOTATION

In Section 3 we put the truth-table method onto a more solid footing. We extended it a little too, because we made no assumption that the language L had just finitely many sentence letters. The original purpose of the exercise was to prove valid argument schemas, and we can now redefine these in sharper terms too.

Let L be a fixed propositional language and  $\phi_1, \dots, \phi_n, \psi$  any formulas of L. Then the statement

$$(26) \quad \phi_1, \dots, \phi_n \models \psi$$

will mean: for every L-structure  $\mathfrak{A}$ , if  $\mathfrak{A} \models \phi_1$  and  $\dots$   $\mathfrak{A} \models \phi_n$ , then  $\mathfrak{A} \models \psi$ . We allow  $n$  to be zero; thus

$$(27) \quad \models \psi$$

means that for every L-structure  $\mathfrak{A}$ ,  $\mathfrak{A} \models \psi$ . To say that (26) is false, we write

$$(28) \quad \phi_1, \dots, \phi_n \not\models \psi.$$

Note that (26)–(28) are statements about formulas of L and not themselves formulas of L.

It is a misfortune that custom requires us to use the same symbol  $\vDash$  both in ' $\mathfrak{A} \vDash \phi$ ' (cf. (22) above) and in ' $\phi_1, \dots, \phi_n \vDash \psi$ '. It means quite different things in the two cases. But one can always see which is meant, because in the first case a structure  $\mathfrak{A}$  is mentioned immediately to the left of  $\vDash$ , and in the second usage  $\vDash$  follows either a formula or an empty space.  $\vDash$  can be pronounced 'double turnstile' or 'semantic turnstile', to contrast it with the symbol  $\vdash$  ('turnstile' or 'syntactic turnstile') which occurs in the study of formal proof calculi (cf. Section 7 below).

The point of definition (26) should be clear. It says in effect that if we make any consistent replacement of the sentence letters by sentences of English, then in any situation where the sentences resulting from  $\phi_1, \dots, \phi_n$  are true, the sentence got from  $\psi$  will be true too. In short (26) says that

$$(29) \quad \phi_1, \dots, \phi_n. \text{ Therefore } \psi.$$

is a valid argument schema. What's more, it says it without mentioning either English sentences or possible situations. Statements of form (26) or (27) are called *sequents* (= 'things that follow' in Latin). When (26) is true,  $\phi_1, \dots, \phi_n$  are said to *logically imply*  $\psi$ . When (27) is true,  $\psi$  is said to be a *tautology*; for a language with a finite number of sentence letters, this means that the truth-table of  $\psi$  has *T* all the way down its main column. Some elementary texts give long lists of tautologies (e.g. Kalish and Montague [1964, pp. 80–84]).

While we are introducing notation, let me mention some useful abbreviations. Too many brackets can make a formula hard to read. So we shall agree that when naming formulas we can leave out some of the brackets. First, we can leave off the brackets at the two ends of an occurrence of  $(\phi \wedge \psi)$  or  $(\phi \vee \psi)$  provided that the only truth-functor which occurs immediately outside them is either  $\rightarrow$  or  $\leftrightarrow$ . For example we can abbreviate

$$(30) \quad (p \leftrightarrow (q \wedge r)) \text{ and } ((p \wedge q) \rightarrow (r \vee s))$$

to

$$(31) \quad (p \leftrightarrow q \wedge r) \text{ and } (p \wedge q \rightarrow r \vee s)$$

respectively; but we can *not* abbreviate

$$(32) \quad (\neg(p \wedge q) \rightarrow r) \text{ and } ((p \leftrightarrow q) \wedge r)$$

to

$$(33) \quad (\neg p \wedge q \rightarrow r) \text{ and } (p \leftrightarrow q \wedge r)$$

respectively.

Second, we can leave off brackets at the ends of a formula. So the formulas in (31) can also be written

$$(34) \quad p \leftrightarrow q \wedge r \text{ and } p \wedge q \rightarrow r \vee s$$

respectively.

Third, if we have a string of  $\wedge$ 's with their associated brackets bunched up towards the left end of the formula, as in

$$(35) \quad (((q \wedge r) \wedge s) \wedge t),$$

then we can omit all but the outermost brackets:

$$(36) \quad (q \wedge r \wedge s \wedge t).$$

Formula (36) is called a *conjunction* whose *conjuncts* are  $q, r, s, t$ . Likewise we can abbreviate  $((q \vee r) \vee s) \vee t$  to the *disjunction*  $(q \vee r \vee s \vee t)$  with *disjuncts*  $q, r, s, t$ . (But the corresponding move with  $\rightarrow$  or  $\leftrightarrow$  is not allowed.)

All these conventions can be applied together, as when we write

$$(37) \quad p \wedge q \wedge r \rightarrow s$$

for

$$(38) \quad (((p \wedge q) \wedge r) \rightarrow s).$$

When only these abbreviations are used, it is always possible to work out exactly which brackets have been omitted, so that there is no loss of information.

Jan Łukasiewicz pointed out that if we always write connectives to the left of the formulas they connect, then there is no need for any brackets at all. In this style the second formula of (30) could be written

$$(39) \quad \rightarrow \wedge pq \vee rs, \text{ or in Łukasiewicz's notation } CKpqArs.$$

Prior [1962] uses Łukasiewicz's notation throughout.

Note that the abbreviations described above only affect the way we talk about formulas of L—the formulas themselves remain untouched. The definition of 'formula of similarity type  $X$ ' given in Section 3 stands without alteration. Some early writers were rather carefree about this point, making it difficult to follow what language L they really had in mind. If anybody wants to do calculations *in* L but still take advantage of our abbreviations, there is an easy way he can do it. He simply writes down abbreviated *names* of formulas instead of the formulas themselves. In other words, he works always in the metalanguage and never in the object language. This cheap trick will allow him the best of both worlds: a rigorously defined language and a relaxed and generous notation. Practising logicians do it all the time.

5 PROPERTIES OF  $\vDash$ 

This section gathers up some properties of  $\vDash$  which can be proved directly from the definitions in Sections 3 and 4 above. They are rather a ragbag, but there are some common themes.

**THEOREM 1.** *If  $\mathfrak{A}$  and  $\mathfrak{B}$  are structures which assign the same truth-values as each other to each sentence letter occurring in  $\phi$ , then  $\mathfrak{A} \vDash \phi$  iff  $\mathfrak{B} \vDash \phi$ .*

This is obvious from (23), but it can also be proved rigorously by induction on the complexity of  $\phi$ . The most important consequence of Theorem 1 is:

**THEOREM 2.** *The truth of the sequent ' $\phi_1, \dots, \phi_n \vDash \psi$ ' doesn't depend on what language  $L$  the formulas  $\phi_1, \dots, \phi_n$  and  $\psi$  come from.*

In other words, although the definition of ' $\phi_1, \dots, \phi_n \vDash \psi$ ' was stated in terms of one language  $L$  containing  $\phi_1, \dots, \phi_n$  and  $\psi$ , any two such languages would give the same outcome. At first sight Theorem 2 seems a reasonable property to expect of any decent notion of entailment. But in other logics, notions of entailment which violate Theorem 2 have sometimes been proposed. (There is an example in Dunn and Belnap [1968], and another in Section 15 below.)

The next result turns all problems about sequents into problems about tautologies.

**THEOREM 3 (Deduction Theorem).**  $\phi_1, \dots, \phi_n \vDash \psi$  if and only if  $\phi_1, \dots, \phi_{n-1} \vDash \phi_n \rightarrow \psi$ .

Theorem 3 moves formulas to the right of  $\vDash$ . It has a twin that does the opposite:

**THEOREM 4.**  $\phi_1, \dots, \phi_n \vDash \psi$  iff  $\phi_1, \dots, \phi_n, \neg\psi \vDash \perp$ .

We say that the formula  $\phi$  is *logically equivalent* to the formula  $\psi$  if  $\phi \vDash \psi$  and  $\psi \vDash \phi$ . This is equivalent to saying that  $\vDash \phi \leftrightarrow \psi$ . Intuitively speaking, logically equivalent formulas are formulas which behave in exactly the same way inside arguments. Theorem 5 makes this more precise:

**THEOREM 5.** *If  $\phi_1, \dots, \phi_n \vDash \psi$ , and we take an occurrence of a formula  $\chi$  inside one of  $\phi_1, \dots, \phi_n, \psi$  and replace it by an occurrence of a formula which is logically equivalent to  $\chi$ , then the resulting sequent holds too.*

For example,  $\neg p \vee q$  is logically equivalent to  $p \rightarrow q$  (as truth-tables will confirm). Also we can easily check that

$$(40) \quad r \rightarrow (\neg p \vee q), p \vDash r \rightarrow q.$$

Then Theorem 5 tells us that the following sequent holds too:

$$(41) \quad r \rightarrow (p \rightarrow q), p \vDash r \rightarrow q.$$



An interesting consequence of Theorem 5 is:

**THEOREM 6.** *Every formula  $\phi$  is logically equivalent to a formula which uses the same sentence letters as  $\phi$ , but no truth-functors except  $\perp, \neg$  and  $\rightarrow$ .*

**Proof.** Truth-tables will quickly show that

$$(42) \quad \begin{aligned} \psi \wedge \chi &\text{ is logically equivalent to } \neg(\psi \rightarrow \neg\chi), \\ \psi \vee \chi &\text{ is logically equivalent to } (\neg\psi \rightarrow \chi), \text{ and} \\ \psi \leftrightarrow \chi &\text{ is logically equivalent to } \neg((\psi \rightarrow \chi) \rightarrow \neg(\chi \rightarrow \psi)). \end{aligned}$$

But then by Theorem 5, if we replace a part of  $\phi$  of form  $(\psi \wedge \chi)$  by  $\neg(\psi \rightarrow \neg\chi)$ , the resulting formula will be logically equivalent to  $\phi$ . By replacements of this kind we can eliminate in turn all the occurrences of  $\wedge, \vee$  and  $\leftrightarrow$  in  $\phi$ , and be left with a formula which is logically equivalent to  $\phi$ . This proves Theorem 6. Noting that

$$(43) \quad \neg\phi \text{ is logically equivalent to } \phi \rightarrow \perp,$$

we can eliminate  $\neg$  too, at the cost of introducing some more occurrences of  $\perp$ . ■

An argument just like the proof of Theorem 6 shows that every formula is logically equivalent to one whose only truth-functors are  $\neg$  and  $\wedge$ , and to one whose only truth-functors are  $\neg$  and  $\vee$ . But there are some limits to this style of reduction: there is no way of eliminating  $\neg$  and  $\perp$  in favour of  $\wedge, \vee, \rightarrow$  and  $\leftrightarrow$ .

The next result is a useful theorem of Post [1921]. In Section 2 we found a truth-table for each formula. Now we go the opposite way and find a formula for each truth-table.

**THEOREM 7.** *Let  $P$  be a truth-table which writes either  $T$  or  $F$  against each possible assignment of truth-values to the sentence letters  $p_1, \dots, p_n$ . Then  $P$  is the truth-table of some formula using no sentence letters apart from  $p_1, \dots, p_n$ .*

**Proof.** I sketch the proof. Consider the  $j$ th row of the table, and write  $\phi_j$  for the formula  $p'_1 \wedge \dots \wedge p'_n$ , where each  $p'_i$  is  $p_i$  if the  $j$ th row makes  $p_i$  true, and  $\neg p_i$  if the  $j$ th row makes  $p_i$  false. Then  $\phi_j$  is a formula which is true at just the  $j$ th row of the table. Suppose the rows to which the table gives the value  $T$  are rows  $j_1, \dots, j_k$ . Then take  $\phi$  to be  $\phi_{j_1} \vee \dots \vee \phi_{j_k}$ . If the table has  $F$  all the way down, take  $\phi$  to be  $\perp$ . Then  $P$  is the truth-table of  $\phi$ . ■

Theorem 7 says in effect that we could never get a more expressive logic by inventing new truth-functors. Anything we could say with the new truth-functors could also be said using the ones we already have.

A formula is said to be in *disjunctive normal form* if it is either  $\perp$  or a disjunction of conjunctions of basic formulas (basic = atomic or negated atomic). The proof of Theorem 7 actually shows that  $P$  is the truth-table of some formula in disjunctive normal form. Suppose now that we take any formula  $\psi$ , work out its truth-table  $P$ , and find a formula  $\phi$  in disjunctive normal form with truth-table  $P$ . Then  $\psi$  and  $\phi$  are logically equivalent, because they have the same truth-table. So we have proved:

**THEOREM 8.** *Every formula is logically equivalent to a formula in disjunctive normal form.*

One can also show that every formula is logically equivalent to one in *conjunctive normal form*, i.e. either  $\neg\perp$  or a conjunction of disjunctions of basic formulas.

**LEMMA 9** (Craig's Interpolation Lemma for propositional logic). *If  $\psi \vDash \chi$  then there exists a formula  $\phi$  such that  $\psi \vDash \phi$  and  $\phi \vDash \chi$ , and every sentence letter which occurs in  $\phi$  occurs both in  $\psi$  and in  $\chi$ .*

**Proof.** Let  $L$  be the language whose sentence letters are those which occur both in  $\psi$  and in  $\chi$ , and  $L^+$  the language whose sentence letters are those in either  $\psi$  or  $\chi$ . Write out a truth-table for the letters in  $L$ , putting  $T$  against a row if and only if the assignment of truth-values in that row can be expanded to form a model of  $\psi$ . By Theorem 7, this table is the truth-table of some formula  $\phi$  of  $L$ . Now we show  $\phi \vDash \chi$ . Let  $\mathfrak{A}$  be any  $L^+$ -structure such that  $\mathfrak{A} \vDash \phi$ . Let  $\mathfrak{C}$  be the  $L$ -structure which agrees with  $\mathfrak{A}$  on all letters in  $L$ . Then  $\mathfrak{C} \vDash \phi$  by Theorem 1. By the definition of  $\phi$  it follows that some model  $\mathfrak{B}$  of  $\psi$  agrees with  $\mathfrak{C}$  on all letters in  $L$ . Now we can put together an  $L^+$ -structure  $\mathfrak{D}$  which agrees with  $\mathfrak{B}$  on all letters occurring in  $\psi$ , and with  $\mathfrak{A}$  on all letters occurring in  $\chi$ . (The overlap was  $L$ , but  $\mathfrak{A}$  and  $\mathfrak{B}$  both agree with  $\mathfrak{C}$  and hence with each other on all letters in  $L$ .) Then  $\mathfrak{D} \vDash \psi$  and hence  $\mathfrak{D} \vDash \chi$  since  $\psi \vDash \chi$ ; but then  $\mathfrak{A} \vDash \chi$  too. The proof that  $\psi \vDash \phi$  is easier and I leave it to the reader. ■

Craig's Lemma is the most recent fundamental discovery in propositional logic. It is easy to state and to prove, but it was first published over a hundred years after propositional logic was invented [Craig, 1957a]. The corresponding lemma holds for full first-order logic too; this is much harder to prove. (See Lemma 32 below.)

Most of the topics in this section are taken further in Hilbert and Bernays [1934], Kleene [1952], Rasiowa and Sikorski [1963] and Bell and Machover [1977].



Now there are just two ways of making  $\neg(p \wedge r)$  true, namely to make  $\neg p$  true and to make  $\neg r$  true. (Of course these ways are not mutually exclusive.) So in our attempt to refute (45) we have two possible options to try, and the diagram accordingly branches in two directions:

$$(48) \quad \begin{array}{c} p \wedge q, \neg(p \wedge r), \neg\neg r \vDash \perp \\ | \\ p \wedge q, \neg(p \wedge r), r \vDash \perp \\ | \\ p, q, \neg(p \wedge r), r \vDash \perp \\ / \quad \backslash \\ p, q, \neg p, r \vDash \perp \quad p, q, \neg r, r \vDash \perp \end{array}$$

But there is no chance of having both  $p$  and  $\neg p$  true in the same structure. So the left-hand fork is a non-starter, and we block it off with a line. Likewise the right-hand fork expects a structure in which  $\neg r$  and  $r$  are both true, so it must be blocked off:

$$(49) \quad \begin{array}{c} p \wedge q, \neg(p \wedge r), \neg\neg r \vDash \perp \\ | \\ p \wedge q, \neg(p \wedge r), r \vDash \perp \\ | \\ p, q, \neg(p \wedge r), r \vDash \perp \\ / \quad \backslash \\ \underline{p, q, \neg p, r \vDash \perp} \quad \underline{p, q, \neg r, r \vDash \perp} \end{array}$$

Since every possibility has been explored and closed off, we conclude that there is no possible way of refuting (45), and so (45) is correct.

What happens if we apply the same technique to an incorrect sequent? Here is an example:

$$(50) \quad p \vee \neg(q \rightarrow r), q \rightarrow r \vDash q.$$

I leave it to the reader to check the reasons for the steps below—he should note that  $q \rightarrow r$  is true if and only if either  $\neg q$  is true or  $r$  is true:

$$(51) \quad \begin{array}{c} p \vee \neg(q \rightarrow r), q \rightarrow r, \neg q \vDash \perp \\ / \quad \backslash \\ p, q \rightarrow r, \neg q \vDash \perp \quad \underline{\neg(q \rightarrow r), q \rightarrow r, \neg q \vDash \perp} \\ / \quad \backslash \\ p, \neg q, \neg q \vDash \perp \quad p, r, \neg q \vDash \perp \end{array}$$

Here two branches remain open, and since all the formulas in them have been decomposed into atomic formulas or negations of atomic formulas,

there is nothing more we can do with them. In every such case it turns out that each open branch describes a structure which refutes the original sequent. For example, take the leftmost branch in (51). The formulas on the left side of the bottom sequent describe a structure  $\mathfrak{A}$  in which  $p$  is true and  $q$  is false. The sequent says nothing about  $r$ , so we can make an arbitrary choice: let  $r$  be false in  $\mathfrak{A}$ . Then  $\mathfrak{A}$  is a structure in which the two formulas on the left in (50) are true but that on the right is false.

This method always leads in a finite time *either* to a tree diagram with all branches closed off, in which case the beginning sequent was correct; *or* to a diagram in which at least one branch remains resolutely open, in which case this branch describes a structure which shows that the sequent was incorrect.

Diagrams constructed along the lines of (49) or (51) above are known as *semantic tableaux*. They were first invented, upside-down and with a different explanation, by Gentzen [1934]. The explanation given above is from Beth [1955] and Hintikka [1955].

We can cut out a lot of unnecessary writing by omitting the ‘ $\models \perp$ ’ at the end of each sequent. Also in all sequents below the top one, we need only write the new formulas. In this abbreviated style the diagrams are called *truth-trees*. Written as truth-trees, (49) looks like this:

$$(52) \quad \begin{array}{c} p \vee q, \neg(p \wedge r), \neg\neg r \\ | \\ r \\ | \\ p \\ | \\ q \\ / \quad \backslash \\ \underline{\neg p} \quad \underline{\neg r} \end{array}$$

and (51) becomes

$$(53) \quad \begin{array}{c} p \vee \neg(q \rightarrow r), q \rightarrow r, \neg q \\ / \quad \backslash \\ p \quad \underline{\neg(q \rightarrow r)} \\ / \quad \backslash \\ \neg q \quad r \end{array}$$

The rules for breaking down formulas in truth-trees can be worked out straight from the truth-table definitions of the truth-functors, but for the reader's convenience I list them:

$$\begin{array}{ccccc}
 (54) & \neg\neg\phi & \phi \wedge \psi & \neg(\phi \wedge \psi) & \phi \vee \psi & \neg(\phi \vee \psi) \\
 & | & | & / \quad \backslash & / \quad \backslash & | \\
 & \phi & \phi & \neg\phi & \phi & \neg\phi \\
 & & \psi & \neg\psi & \psi & \neg\psi \\
 \\
 & \phi \rightarrow \psi & \neg(\phi \rightarrow \psi) & \phi \leftrightarrow \psi & \neg(\phi \leftrightarrow \psi) \\
 & / \quad \backslash & | & / \quad \backslash & / \quad \backslash \\
 & \neg\phi & \phi & \phi & \phi & \neg\phi \\
 & \psi & \neg\psi & \neg\phi & \neg\psi & \psi
 \end{array}$$

One is allowed to close off a branch as soon as either  $\perp$  or any outright contradiction  $\phi, \neg\phi$  appears among the formulas in a branch. (Truth-trees are used in Jeffrey [1967]; see [Smullyan, 1968; Bell and Machover, 1977] for mathematical analyses.) Truth-trees are one dialect of semantic tableaux. Here is another. We shall understand the generalised sequent

$$(55) \quad \phi_1, \dots, \phi_n \vDash \psi_1, \dots, \psi_m$$

to mean that there is no structure which makes  $\phi_1, \dots, \phi_n$  all true and  $\psi_1, \dots, \psi_m$  all false. A structure in which  $\phi_1, \dots, \phi_n$  are true and  $\psi_1, \dots, \psi_m$  are false is called a *counterexample* to (55). When there is only one formula to the right of  $\vDash$ , (55) means just the same as our previous sequents (26).

Generalised sequents have the following two symmetrical properties:

$$(56) \quad \phi_1, \dots, \phi_n, \neg\chi \vDash \psi_1, \dots, \psi_m \quad \text{iff} \quad \phi_1, \dots, \phi_n \vDash \psi_1, \dots, \psi_m, \chi.$$

$$(57) \quad \phi_1, \dots, \phi_n \vDash \psi_1, \dots, \psi_m, \neg\chi \quad \text{iff} \quad \phi_1, \dots, \phi_n, \chi \vDash \psi_1, \dots, \psi_m.$$

Suppose now that we construct semantic tableaux as first described above, but using *generalised* sequents instead of sequents. The effect of (56) and (57) is that we handle  $\neg$  *by itself*; as (54) shows, our previous tableaux could only tackle  $\neg$  two at a time or in combination with another truth-functor.

Using generalised sequents, a proof of (44) goes as follows:

$$\begin{array}{l}
 (58) \qquad p \wedge q, \neg(p \wedge r) \models \neg r \\
 \qquad (i) \qquad \quad | \\
 \qquad \qquad p \wedge q, \neg(p \wedge r), r \models \\
 \qquad (ii) \qquad \quad | \\
 \qquad \qquad p \wedge q, r \models p \wedge r \\
 \qquad (iii) \qquad \quad | \\
 \qquad \qquad p, q, r \models p \wedge r \\
 \qquad (iv) \qquad \quad / \qquad \backslash \\
 \qquad \qquad \underline{p, q, r \models p} \qquad \underline{p, q, r \models r}
 \end{array}$$

Steps (i) and (ii) are by (57) and (56) respectively. Step (iv) is justified as follows. We are trying to build a structure in which  $p, q$  and  $r$  are true but  $p \wedge r$  is false, as a counterexample to the sequent ' $p, q, r \models p \wedge r$ '. By the truth-table for  $\wedge$ , it is necessary and sufficient to build *either* a structure in which  $p, q, r$  are true and  $p$  is false, *or* a structure in which  $p, q, r$  are true and  $r$  is false. We can close off under the bottom left sequent ' $p, q, r \models p$ ' because a formula  $p$  occurs both on the right and on the left of  $\models$ , so that in a counterexample it would have to be both false and true, which is impossible. Likewise at bottom right.

Proofs with generalised sequents are virtually identical with the *cut-free sequent proofs* of [Gentzen, 1934], except that he wrote them upside down. Beth [1955; 1962] used them as a method for testing sequents. He wrote them in a form where, after the first sequent, one only needs to mention the new formulas.

Quine [1950] presents another quite fast decision method which he calls *fell swoop* (to be contrasted with the 'full sweep' of truth-tables).

I turn to the question how fast a decision method of testing sequents can be in the long run, i.e. as the number and lengths of the formulas increase. At the time of writing, this is one of the major unsolved problems of computation theory. A function  $p(n)$  of the number  $n$  is said to be *polynomial* if it is calculated from  $n$  and some other fixed numbers by adding, subtracting and multiplying. (So for example  $n^2 + 3$  and  $2n^3 - n$  are polynomial functions of  $n$  but  $3^n, n!$  and  $1/(n^2 + 1)$  are not.) It is not known whether there exist a decision method  $M$  for sequents of propositional logic, and a polynomial function  $p(n)$ , such that for every sequent  $S$ , if  $n$  is the number of symbols in  $S$  then  $M$  can determine in less than  $p(n)$  steps whether or not  $S$  is correct. If the answer is Yes there are such  $M$  and  $p(n)$ , then we say that the decision problem for propositional logic is *solvable in polynomial time*. Cook [1971] showed that a large number of other interesting computational problems will be solvable in polynomial time if this one is. (See [Garey and Johnson, 1979].) I have the impression that everybody working in the field

expects the answer to be No. This would mean in effect that for longer sequents the problem is too hard to be solved efficiently by a deterministic computer.

## 7 FORMAL PROOF CALCULI

During the first third of this century, a good deal of effort was put into constructing various formal proof calculi for logic. The purpose of this work was to reduce reasoning—or at least a sizeable part of mathematical reasoning—to precise mechanical rules. I should explain at once what a *formal proof calculus* (or *formal system*) is.

A formal proof calculus, call it  $\Sigma$ , is a device for proving sequents in a language  $L$ . First,  $\Sigma$  gives us a set of rules for writing down arrays of symbols on a page. An array which is written down according to the rules is called a *formal proof* in  $\Sigma$ . The rules must be such that one can check by inspection and calculation whether or not an array is a formal proof. Second, the calculus contains a rule to tell us how we can mechanically work out what are the *premises* and the *conclusion* of each formal proof.

We write

$$(59) \quad \phi_1, \dots, \phi_n \vdash_{\Sigma} \psi \text{ or more briefly } \phi_1, \dots, \phi_n \vdash \psi$$

to mean that there is a formal proof in the calculus  $\Sigma$  whose premises all appear in the list  $\phi_1, \dots, \phi_n$ , and whose conclusion is  $\psi$ . Some other ways of expressing (59) are:

‘ $\phi_1, \dots, \phi_n \vdash \psi$ ’ is a *derivable sequent* of  $\Sigma$ ;  
 $\psi$  is *deducible from*  $\phi_1, \dots, \phi_n$  in  $\Sigma$ ;  
 $\phi_1, \dots, \phi_n$  *yield*  $\psi$  in  $\Sigma$ .

We call  $\psi$  a *derivable formula* of  $\Sigma$  if there is a formal proof in  $\Sigma$  with conclusion  $\psi$  and no premises. The symbol  $\vdash$  is called *turnstile* or *syntactic turnstile*.

We say that the calculus  $\Sigma$  is:

*sound* if  $\phi_1, \dots, \phi_n \vdash_{\Sigma} \psi$  implies  $\phi_1, \dots, \phi_n \vDash \psi$   
*strongly complete* if  $\phi_1, \dots, \phi_n \vDash \psi$  implies  $\phi_1, \dots, \phi_n \vdash_{\Sigma} \psi$ ,  
*weakly complete* if  $\vDash \psi$  implies  $\vdash_{\Sigma} \psi$ ,

where  $\phi_1, \dots, \phi_n, \psi$  range over the formulas of  $L$ . These definitions also make sense when  $\vDash$  is defined in terms of other logics, not necessarily first-order. In this chapter ‘complete’ will always mean ‘strongly complete’.

The formal proofs in a calculus  $\Sigma$  are in general meaningless arrays of symbols. They need not be genuine proofs, that is, demonstrations that something is the case. But if we know that  $\Sigma$  is sound, then the fact that a certain sequent is derivable in  $\Sigma$  will prove that the corresponding sequent



with  $\vDash$  is correct. In some proof calculi the formal proofs are made to look as much as possible like intuitively correct reasoning, so that soundness can be checked easily.

We already have the makings of one formal proof calculus in Section 6 above: the *cut-free sequent proofs* using generalised sequents. As proofs, these are usually written the other way up, with  $\vdash$  in place of  $\vDash$ , and with horizontal lines separating the sequents. Also there is no need to put in the lines which mark the branches that are closed, because every branch is closed.

For example, here is a cut-free sequent proof of the sequent ' $p \wedge q, \neg(p \wedge r) \vdash \neg r$ '; compare it with (58):

$$(60) \quad \frac{\frac{\frac{p \vdash p}{p, q, r \vdash p} \quad \frac{r \vdash r}{p, q, r \vdash r}}{p, q, r \vdash p \wedge r}}{p \wedge q, r \vdash p \wedge r}}{p \wedge q, \neg(p \wedge r), r \vdash \quad}{p \wedge q, \neg(p \wedge r) \vdash \neg r}$$

To justify this proof we would show, working upwards from the bottom, that if there is a counterexample to the bottom sequent then at least one of the top sequents has a counterexample, which is impossible. Or equivalently, we could start by noting that the top sequents are correct, and then work *down* the tree, showing that each of the sequents must also be correct. By this kind of argument we can show that the cut-free sequent calculus is sound.

To prove that the calculus is complete, we borrow another argument from Section 6 above. Assuming that a sequent  $S$  is not derivable, we have to prove that it is not correct. To do this, we try to construct a cut-free sequent proof, working upwards from  $S$ . After a finite number of steps we shall have broken down the formulas as much as possible, but the resulting diagram can't be a proof of  $S$  because we assumed there isn't one. So at least one branch must still be 'open' in the sense that it hasn't revealed any immediate contradiction. Let  $B$  be such a branch. Let  $B_L$  be the set of all formulas which occur to the left of  $\vdash$  in some generalised sequent in  $B$ , and let  $B_R$  be the same with 'right' for 'left'. We can define a structure  $\mathfrak{A}$  by

$$(61) \quad I_{\mathfrak{A}}(\phi) = \begin{cases} T & \text{if } \phi \text{ is a sentence letter which is in } B_L, \\ F & \text{if } \phi \text{ is a sentence letter not in } B_L. \end{cases}$$

Then we can prove, by induction on the complexity of the formula  $\psi$ , that if  $\psi$  is any formula in  $B_L$  then  $\mathfrak{A} \vDash \psi$ , and if  $\psi$  is any formula in  $B_R$  then  $\mathfrak{A} \vDash \neg\psi$ . It follows that  $\mathfrak{A}$  is a counterexample to the bottom sequent  $S$ , so that  $S$  is not correct.

The cut-free sequent calculus itself consists of a set of mechanical rules for constructing proofs, and it could be operated by somebody who had not the least idea what  $\vdash$  or any of the other symbols mean. These rules are listed in Sundholm (in Volume 2 of this *Handbook*).

Gentzen [1934] had another formal proof calculus, known simply as the *sequent calculus*. This was the same as the cut-free sequent calculus, except that it allowed a further rule called the *cut rule* (because it cuts out a formula):

$$(62) \frac{\dots \vdash ***, \chi \quad \dots, \chi \vdash ***}{\dots \vdash ***}$$

This rule often permits much shorter proofs. Gentzen justified it by showing that any proof which uses the cut rule can be converted into a cut-free proof of the same sequent. This *cut elimination theorem* is easily the best mathematical theorem about proofs. Gentzen himself adapted it to give a proof of the consistency of first-order Peano arithmetic. By analysing Gentzen's argument we can get sharp information about the degree to which different parts of mathematics rely on infinite sets. (Cf. [Schütte, 1977]. Gentzen's results on cut-elimination were closely related to deep but undigested work on quantifier logic which Jacques Herbrand had done before his death in a mountaineering accident at the age of 23; see [Herbrand, 1930] and the Introduction to [Herbrand, 1971].) Further details of Gentzen's sequent calculi, including the intuitionistic versions, are given in [Kleene, 1952, Ch XV] and Sundholm (in Volume 2 of this *Handbook*).

In the same paper, Gentzen [1934] described yet a third formal proof calculus. This is known as the *natural deduction calculus* because proofs in this calculus start with their premises and finish at their conclusions (unlike sequent calculi and semantic tableaux), and all the steps between are intuitively natural (unlike the Hilbert-style calculi to be described below).

A proof in the natural deduction calculus is a tree of formulas, with a single formula at the bottom. The formulas at the tops of the branches are called the *assumptions* of the proof. Some of the assumptions may be *discharged* or *cancelled* by having square brackets [ ] written around them. The *premises* of the proof are its uncanceled assumptions, and the *conclusion* of the proof is the formula at the bottom.

Sundholm (in his chapter in Volume D2 of this *Handbook*) gives the full rules of the natural deduction calculus. Here are a few illustrations. Leaving aside  $\neg$  and  $\perp$  for the moment, there are two rules for each truth-functor, namely an *introduction* rule and an *elimination* rule. The introduction rule for  $\wedge$  is:

$$(63) \frac{\phi \quad \psi}{\phi \wedge \psi}$$

i.e. from  $\phi$  and  $\psi$  deduce  $\phi \wedge \psi$ . The elimination rule for  $\wedge$  comes in a left-hand version and a right-hand version:

$$(64) \quad \frac{\phi \wedge \psi}{\phi} \quad \frac{\phi \wedge \psi}{\psi}.$$

The introduction rule for  $\rightarrow$  says that if we have a proof of  $\psi$  from certain assumptions, then we can deduce  $\phi \rightarrow \psi$  from those assumptions less  $\phi$ :

$$(65) \quad \frac{\begin{array}{c} [\phi] \\ \vdots \\ \psi \end{array}}{\phi \rightarrow \psi}$$

The elimination rule for  $\rightarrow$  is the *modus ponens* of the medievals:

$$(66) \quad \frac{\phi \quad \phi \rightarrow \psi}{\psi}.$$

For example, to prove

$$(67) \quad q, p \wedge q \rightarrow r \vDash p \rightarrow r$$

in the natural deduction calculus we write:

$$(68) \quad \frac{\frac{\frac{[p] \quad q}{p \wedge q} \quad p \wedge q \rightarrow r}{r}}{p \rightarrow r}}$$

Note that the assumption  $p$  is discharged at the last step when  $p \rightarrow r$  is introduced.

The calculus reads  $\neg\phi$  as a shorthand for  $\phi \rightarrow \perp$ . So for example, from  $\phi$  and  $\neg\phi$  we deduce  $\perp$  by (66). There is an elimination rule for  $\perp$ . It says: given a proof of  $\perp$  from certain assumptions, derive  $\phi$  from the same assumptions less  $\phi \rightarrow \perp$ :

$$(69) \quad \frac{\begin{array}{c} [\phi \rightarrow \perp] \\ \vdots \\ \perp \end{array}}{\phi}$$

This is a form of *reductio ad absurdum*.

The rule about cancelling assumptions in (65) should be understood as follows. When we make the deduction, we are *allowed* to cancel  $\phi$  wherever it occurs as an assumption. But we are not obliged to; we can cancel some

occurrences of  $\phi$  and not others, or we can leave it completely uncanceled. The formula  $\phi$  may not occur as an assumption anyway, in which case we can forget about cancelling it. The same applies to  $\phi \rightarrow \perp$  in (69). So (69) implies the following weaker rule in which we make no cancellations:

$$(70) \frac{\perp}{\phi}$$

(‘Anything follows from a contradiction’.) Intuitionist logic accepts (70) but rejects the stronger rule (69) (cf. van Dalen (Volume 7)).

Belnap [1962] and Prawitz [1965] have explained the idea behind the natural deduction calculus in an interesting way. For each truth-functor the rules are of two sorts, the introduction rules and the elimination rules. In every case the elimination rules *only allow us to infer from a formula what we had to know in order to introduce the formula*. For example we can remove  $\phi \rightarrow \psi$  only by rule (66), i.e. by using it to deduce  $\psi$  from  $\phi$ ; but  $\phi \rightarrow \psi$  can only be introduced either as an explicit assumption or (by (65)) when we already know that  $\psi$  can be deduced from  $\phi$ . (Rule (69) is in a special category. It expresses (1) that everything is deducible from  $\perp$ , and (2) that for each formula  $\phi$ , at least one of  $\phi$  and  $\phi \rightarrow \perp$  is true.)

Popper [1946/47, particularly p. 284] rashly claimed that he could define truth-functors just by writing down natural deduction rules for them. Prior [1960] gave a neat example to show that this led to absurdities. He invented the new truth-functor *tonk*, which is defined by the rules

$$(71) \frac{\phi}{\phi \text{ tonk } \psi} \quad \frac{\phi \text{ tonk } \psi}{\psi}$$

and then proceeded to infer everything from anything. Belnap [1962] points out that Prior’s example works because its introduction and elimination rules fail to match up in the way described above. Popper should at least have imposed a requirement that the rules must match up. (Cf. [Prawitz, 1979], [Tennant, 1978, p. 74ff], and Sundholm (Volume 2).)

Natural deduction calculi, all of them variants of Gentzen’s, are given by Anderson and Johnstone [1962], Fitch [1952], Kalish and Montague [1964], Lemmon [1965], Prawitz [1965], Quine [1950], Suppes [1957], Tennant [1978], Thomason [1970] and van Dalen [1980]. Fitch (followed e.g. by Thomason) makes the trees branch to the right. Some versions (e.g. Quine’s) disguise the pattern by writing the formulas in a vertical column. So they have to supply some other way of marking which formulas depend on which assumptions; different versions do this in different ways.

Just as a semantic tableau with its branches closed is at heart the same thing as a cut-free sequent proof written upside down, Prawitz [1965] has shown that after removing redundant steps, a natural deduction proof is really the same thing as a cut-free sequent proof written sideways. (See

also Zucker [1974].) The relationship becomes clearer if we adapt the natural deduction calculus so as to allow a proof to have several alternative conclusions, just as it has several premises. Details of such calculi have been worked out by Kneale [1956] and more fully by Shoesmith and Smiley [1978].

A proof of  $p \vee \neg p$  in Gentzen's natural deduction calculus takes several lines. This is a pity, because formulas of the form  $\phi \vee \neg \phi$  are useful halfway steps in proofs of other formulas. So some versions of natural deduction allow us to quote a few tautologies such as  $\phi \vee \neg \phi$  whenever we need them in a proof. These tautologies are then called *axioms*. Technically they are formulas deduced from no assumptions, so we draw a line across the top of them, as at top right in (72) below.

If we wanted to undermine the whole idea of natural deduction proofs, we could introduce axioms which replace all the natural deduction rules except modus ponens. For example we can put (63) out of a job by using the axiom  $\phi \rightarrow (\psi \rightarrow \phi \wedge \psi)$ . Whenever Gentzen used (63) in a proof, we can replace it by

$$(72) \quad \frac{\psi \quad \frac{\phi \quad \overline{\phi \rightarrow (\psi \rightarrow \phi \wedge \psi)}}{\psi \rightarrow \phi \wedge \psi}}{\phi \wedge \psi}$$

using (66) twice. Likewise (64) become redundant if we use the axioms  $\phi \wedge \psi \rightarrow \phi$  and  $\phi \wedge \psi \rightarrow \psi$ . Rule (65) is a little harder to dislodge, but it can be done, using the axioms  $\phi \rightarrow (\psi \rightarrow \phi)$  and  $(\phi \rightarrow \psi) \rightarrow ((\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow (\phi \rightarrow \chi))$ .

At the end of these manipulations we have what is called a *Hilbert-style* proof calculus. A Hilbert-style calculus consists of a set of formulas called *axioms*, together with one or two *derivation rules* for getting new formulas out of given ones. To prove  $\phi_1, \dots, \phi_n \vDash \psi$  in such a calculus, we apply the derivation rules as many times as we like to  $\phi_1, \dots, \phi_n$  and the axioms, until they give us  $\psi$ .

One Hilbert-style system is described in Appendix A below. Mates [1965] works out another such system in detail. Hilbert-style calculi for propositional logic were given by Frege [1879; 1893], Peirce [1885], Hilbert [1923] and Łukasiewicz (see [Łukasiewicz and Tarski, 1930]). (Cf. Sundholm (Volume 2 of this *Handbook*).

The typical Hilbert-style calculus is inefficient and barbarously unintuitive. But they do have two merits. The first is that their mechanics are usually very simple to describe—many Hilbert-style calculi for propositional logic have only one derivation rule, namely modus ponens. This makes them suitable for encoding into arithmetic (Section 24 below). The second merit is that we can strengthen or weaken them quite straightforwardly by tampering with the axioms, and this commends them to researchers in non-classical

logics.

Soundness for these calculi is usually easy to prove: one shows (a) that the axioms are true in every structure and (b) that the derivation rules never lead from truth to falsehood. One way of proving completeness is to show that every natural deduction proof can be converted into a Hilbert-style proof of the same sequent, as hinted above. (Kleene [1952] Section 77 shows how to convert sequent proofs into Hilbert-style proofs and *vice versa*; see Sundholm (Volume 2 of this *Handbook*).)

Alternatively we can prove their completeness directly, using maximal consistent sets. Since this is a very un-proof-theoretic approach, and this section is already too long, let me promise to come back to the matter at the end of Section 16 below. (Kalmár [1934/5] and Kleene independently found a neat proof of the weak completeness of Hilbert-style calculi, by converting a truth-table into a formal proof; cf. Kleene [1952, p. 132ff] or Mendelson [1987, p. 34].)

## II: Predicate Logic

### 8 BETWEEN PROPOSITIONAL LOGIC AND PREDICATE LOGIC

If we asked a Proof Theorist to explain what it means to say

(73)  $\phi_1, \dots, \phi_n$  logically imply  $\psi$ ,

where  $\phi_1, \dots, \phi_n$  and  $\psi$  are formulas from propositional logic, he would explain that it means this: there is a proof of  $\psi$  from  $\phi_1, \dots, \phi_n$  in one of the standard proof calculi. A Model Theorist would prefer to use the definition we gave in Section 4 above, and say that (73) means: whenever  $\phi_1, \dots, \phi_n$  are true in a structure, then  $\psi$  is true in that structure too. The Traditional Logician for his part would explain it thus: every argument of the form ' $\phi_1, \dots, \phi_n$ . Therefore  $\psi$ ' is valid. There need be no fight between these three honest scholars, because it is elementary to show that (73) is true under any one of these definitions if and only if it is true under any other.

In the next few sections we shall turn from propositional logic to predicate logic, and the correct interpretation of (73) will become more contentious.

When  $\phi_1, \dots, \phi_n$  and  $\psi$  are sentences from predicate logic, the Proof Theorist has a definition of (73) which is a straightforward extension of his definition for propositional logic, so he at any rate is happy.

But the Traditional Logician will be in difficulties, because the quantifier expressions of predicate logic have a quite different grammar from all locutions of normal English; so he is hard put to say what would count as an argument of the form ' $\phi_1, \dots, \phi_n$ . Therefore  $\psi$ '. He will be tempted to

say that really we should look at sentences whose deep structures (which he may call logical forms) are like the formulas  $\phi_1, \dots, \phi_n, \psi$ . This may satisfy him, but it will hardly impress people who know that in the present state of the linguistic art one can find experts to mount convincing arguments for any one of seventeen deep structures for a single sentence. A more objective but admittedly vague option would be for him to say that (73) means that any argument which can be *paraphrased* into this form, using the apparatus of first-order logic, is valid.

But the man in the worst trouble is the Model Theorist. On the surface all is well—he has a good notion of ‘structure’, which he took over from the algebraists, and he can say just what it means for a formula of predicate logic to be ‘true in’ a structure. So he can say, just as he did for propositional logic, that (73) means that whenever  $\phi_1, \dots, \phi_n$  are true in a structure, then  $\psi$  is true in that structure too. His problems start as soon as he asks himself what a structure really is, and how he knows that they exist.

Structures, as they are presented in any textbook of model theory, are abstract set-theoretic objects. There are uncountably many of them and most of them are infinite. They can’t be inspected on a page (like proofs in a formal calculus) or heard at Hyde Park Corner (like valid arguments). True, several writers have claimed that the only structures which exist are those which somebody constructs. (E.g. Putnam [1980, p. 482]: ‘Models are ... constructions within our theory itself, and they have names from birth.’) Unfortunately this claim is in flat contradiction to about half the major theorems of model theory (such as the Upward Löwenheim–Skolem Theorem, Theorem 14 in Section 17 below).

Anybody who wants to share in present-day model theory has to accept that structures are as disparate and intangible as sets are. One must handle them by set-theoretic principles and not by explicit calculation. Many model theorists have wider horizons even than this. They regard the whole universe  $V$  of sets as a structure, and they claim that first-order formulas in the language of set theory are true or false in this structure by just the same criteria as in smaller structures. The axioms of Zermelo–Fraenkel set theory, they claim, are simply true in  $V$ .

It is actually a theorem of set theory that a notion of truth adequate to cope with the whole universe of sets *cannot be formalised within set theory*. (We prove this in Section 24 below.) So a model theorist with this wider horizon is strictly not entitled to use formal set-theoretic principles either, and he is forced back onto his intuitive understanding of words like ‘true’, ‘and’, ‘there is’ and so forth. In mathematical practice this causes no problems whatever. The problems arise when one tries to justify what the mathematicians are doing.

In any event it is a major exercise to show that these three interpretations of (73) in predicate logic—or four if we allow the Model Theorist his wider and narrower options—agree with each other. But logicians pride

themselves that it can be done. Section 17 will show how.

## 9 QUANTIFIERS

First-order predicate logic comes from propositional logic by adding the words ‘every’ and ‘some’.

Let me open with some remarks about the meaning of the word ‘every’. There is no space here to rebut rival views (Cf. Leblanc (see Volume 2 of this *Handbook*); on substitutional quantification see [Dunn and Belnap, 1968; Kripke, 1976; Stevenson, 1973].) But anybody who puts a significantly different interpretation on ‘every’ from the one presented below will have to see first-order logic in a different light too.

A person who understands the words ‘every’, ‘Pole’, the sentence

(74) Richard is a Catholic.

and the principles of English sentence construction must also understand the sentence

(75) Every Pole is a Catholic.

How?

First, (74) is true if and only if Richard satisfies a certain condition, namely that

(76) He is a Catholic.

I underline the pronoun that stands for whatever does or does not satisfy the condition. Note that the condition expressed by (76) is one which people either satisfy or do not satisfy, regardless of how or whether we can identify them. Understanding the condition is a necessary part of understanding (74). In Michael Dummett’s words [1973, p. 517]:

... given that we understand a sentence from which a predicate has been formed by omission of certain occurrences of a name, we are capable of recognising what concept that predicate stands for in the sense of knowing what it is for it to be true of or false of any arbitrary object, whether or not the language contains a name for that object.

Second, the truth or otherwise of (75) in a situation depends on what class of Poles is on the agenda. Maybe only Poles at this end of town are under discussion, maybe Poles anywhere in the world; maybe only Poles alive now, maybe Poles for the last hundred years or so. Possibly the speaker was a little vague about which Poles he meant to include. I count the specification of the relevant class of Poles as part of the situation in which (75) has a



truth-value. This class of Poles is called the *domain of quantification* for the phrase ‘every Pole’ in (75). The word ‘Pole’ is called the *restriction term*, because it restricts us to Poles; any further restrictions on the domain of quantification are called *contextual restrictions*.

So when (75) is used in a context, the word ‘Pole’ contributes a domain of quantification and the words ‘is a Catholic’ contribute a condition. The contribution of the word ‘Every’ is as follows: *In any situation, (75) is true iff every individual in the domain of quantification satisfies the condition.*

This analysis applies equally well to other simple sentences containing ‘Every’, such as:

(77) She ate every flower in the garden.

For (77), the situation must determine what the garden is, and hence what is the class of flowers that were in the garden. This class is the domain of quantification; ‘flower in the garden’ is the restriction term. The sentence

(78) She ate it.

expresses a condition which things do or do not satisfy, once the situation has determined who ‘she’ refers to. So in this example the condition varies with the situation. The passage from condition and domain of quantification to truth-value is exactly as before.

The analysis of

(79) Some Pole is a Catholic

(80) She ate some flower (that was) in the garden,

is the same as that of (75), (77) respectively, except at the last step. For (79) or (80) to be true we require that *at least one individual in the domain of quantification satisfies the condition.*

In the light of these analyses we can introduce some notation from first-order logic. In place of the underlined pronoun in (76) and (78) we shall use an *individual variable*, i.e. (usually) a lower-case letter from near the end of the alphabet, possibly with a subscript. Thus:

(81)  $x$  is a Catholic.

Generalising (81), we use the phrase *1-place predicate* to mean a string consisting of words and one individual variable (which may be repeated), such that if the variable is understood as a pronoun referring to a certain person or object, then the string becomes a sentence which expresses that the person or object referred to satisfies a certain condition. The condition may depend on the situation into which the sentence is put.

For an example in which a variable occurs twice,

(82)  $x$  handed the melon to Schmidt, who gave it back to  $x$ .

is a 1-place predicate. It expresses the condition which Braun satisfies if and only Braun handed the melon to Schmidt and Schmidt gave it back to Braun.

To return to (75), ‘Every Pole is a Catholic’: we have now analysed this sentence into (a) a *quantifier word* ‘Every’, (b) the restriction term ‘Pole’, and (c) the predicate ‘ $x$  is a Catholic’.

The separating out of the predicate (by [Frege, 1879], see also [Mitchell, 1883] and [Peirce, 1883]) was vital for the development of modern logic. Predicates have the grammatical form of sentences, so that they can be combined by truth-functors. For example

(83) ( $x$  is a Catholic  $\wedge$   $x$  is a philatelist)

is a predicate which is got by conjoining two other predicates with  $\wedge$ . It expresses the condition which a person satisfies if he is both a Catholic and a philatelist. Incidentally I have seen it suggested that the symbol  $\wedge$  must have a different meaning in (83) from its meaning in propositional logic, because in (83) it stands between predicates which do not have truth-values. The answer is that predicates *do* gain truth-values when their variables are either replaced by or interpreted as names. The truth-value gained in this way by the compound predicate (83) is related to the truth-values gained by its two conjuncts in exactly the way the truth-table for  $\wedge$  describes.

(A historical aside: Peirce [1885] points out that by separating off the predicate we can combine quantifiers with propositional logic; he says that all attempts to do this were ‘more or less complete failures until Mr Mitchell showed how it was to be effected’. Mitchell published in a volume of essays by students of Peirce at Johns Hopkins [Members of the Johns Hopkins University, Boston, 1883]. Christine Ladd’s paper in the same volume mentions both Frege’s *Begriffsschrift* [1879] and Schröder’s review of it. It is abundantly clear that nobody in Peirce’s group had read either. The same happens today.)

The account of quantifiers given above agrees with what Frege said in his *Funktion und Begriff* [1891] and *Grundgesetze* [1893], except in one point. Frege required that all conditions on possible values of the variable should be stated in the predicate. In other words, he allowed only one domain of quantification, namely absolutely everything. For example, if someone were to say, à propos of Poles in New York, ‘Every Pole is a Catholic’, Frege would take this to mean that absolutely everything satisfies the condition

(84) If  $x$  is a Pole in New York City then  $x$  is a Catholic.

If a person were to say

(85) Somebody has stolen my lipstick.

Frege's first move would be to interpret this as saying that at least one thing satisfies the condition expressed by

(86)  $x$  is a person and  $x$  has stolen my lipstick.

Thus Frege removed the restriction term, barred all contextual restrictions, and hence trivialised the domain of quantification.

There are two obvious advantages in getting rid of the restriction term: we have fewer separate expressions to deal with, and everything is thrown into the predicate where it can be analysed by way of truth-functors.

However, it is often useful to keep the restriction terms, if only because it makes formulas easier to read. (There are solid technical dividends too, see Feferman [1968b; 1974].) Most logicians who do this follow the advice of Peirce [1885] and use a special style of variable to indicate the restriction. For example set theorists use Greek variables when the restriction is to ordinals. Variables that indicate a special restriction are said to be *sorted* or *sortal*. Two variables marked with the same restriction are said to be *of the same sort*. Logics which use this device are said to be *many-sorted*.

One can also go halfway with Frege and convert the restriction term into another predicate. In this style, 'Every Pole is a Catholic' comes out as a combination of three units: the quantifier word 'Every', the predicate ' $x$  is a Catholic', and a second *relativisation predicate* ' $x$  is a pole'. The mathematical literature is full of *ad hoc* examples of this approach. See for example the bounded quantifiers of number theory in Section 24 below.

When people started to look seriously at other quantifier words besides 'every' and 'some', it became clear that Frege's method of eliminating the restriction term won't always work. For example, the sentence 'Most judges are freemasons' can't be understood as saying that most things satisfy a certain condition. (For a proof of this, and many other examples, see the study of natural language quantifiers by Barwise and Cooper [1981].) For this reason Neil Tennant [Altham and Tennant, 1975] and Barwise [1974] proposed very general formalisms which keep the relativisation predicate separate from the main predicate.

Frege also avoided contextual restrictions. Given his aim, which was to make everything in mathematical reasoning fully explicit, this might seem natural. But it was a bad move. Contextual restrictions do occur, and a logician ought to be prepared to operate with them. In any case various writers have raised philosophical objections to Frege's habit of talking about just everything. Do we really have an undefinable notion of 'object', as Frege supposed? Is it determinate what objects there are? Don't we falsify the meanings of English sentences if we suppose that they state something about everything there is, when on the face of it they are only about Poles?

For a historical study of quantifiers in first-order logic, consult Goldfarb [1979].

## 10 SATISFACTION

As a convenient and well-known shorthand, we shall say that a person or thing *satisfies* the 1-place predicate  $\phi$  if he or it satisfies the condition which the predicate  $\phi$  expresses. (Notice that we are now allowing the metavariables ' $\phi$ ', ' $\psi$ ' etc. to range over predicates as well as sentences and formulas. This shouldn't cause any confusion.)

Many writers put it a little differently. They say that a person or thing satisfies  $\phi$  if the result of putting a name of the person or thing in place of every occurrence of the variable in  $\phi$  is a true sentence. This way of phrasing matters is fine as a first approximation, but it runs into two hazards.

The first hazard is that not everything has a name, even if we allow phrases of the form 'the such-and-such' as names. For example there are uncountably many real numbers and only countably many names.

I can dispose of this objection quickly, as follows. I decree that for purposes of naming arbitrary objects, any ordered pair whose first term is an object and whose second term is the Ayatollah Khalkhali shall be a name of that object. There is a problem about using these names in sentences, but that's just a matter of finding an appropriate convention. So it is clear that if we have an abstract enough notion of what a name is, then every object can have a name.

More conscientious authors have tried to mount reasoned arguments to show that everything is in principle nameable. The results are not always a success. In one paper I recall, the author was apparently under the impression that the nub of the problem was to find a symbol that could be used for naming hitherto nameless objects. After quoting quite a lot of formulas from Quine's *Methods of Logic*, he eventually announced that lower-case italic  $w$  can always be used for the purpose. No doubt it can!

There is a second hazard in the 'inserted name' definition of satisfaction. If we allow phrases of the form 'the such-and-such' to count as names, it can happen that on the natural reading, a name means something different within the context of the sentence from what it means in isolation. For example, if my uncle is the mayor of Pinner, and in 1954 he fainted during the opening ceremony of the Pinner Fair, then the mayor of Pinner satisfies the predicate:

(87) In 1954  $x$  fainted during the opening ceremony of the Pinner Fair.

But on the natural reading the sentence

(88) In 1954 the mayor of Pinner fainted during the opening ceremony of the Pinner Fair.

says something quite different and is probably false. One can avoid this phenomenon by sticking to names like 'the present mayor of Pinner' which automatically extract themselves from the scope of surrounding temporal

operators (cf. [Kamp, 1971]). But other examples are less easily sorted out. If the programme note says simply ‘Peter Warlock wrote this song’, then Philip Heseltine, one of whose pen-names was ‘Peter Warlock’, surely satisfies the predicate

(89) The programme note attributes this song to  $x$ .

But my feeling is that on the natural reading, the sentence

(90) The programme note attributes this song to Philip Heseltine

is false. Examples like these should warn us to be careful in applying first-order formalisms to English discourse. (Cf. Bäuerle and Cresswell’s chapter ‘Propositional Attitudes’ to be found in a later Volume of this *Handbook*.)

I turn to some more technical points. We shall need to handle expressions like

(91)  $x$  was observed handing a marked envelope to  $y$

which expresses a condition on *pairs* of people or things. It is, I think, quite obvious how to generalize the notion of a 1-place predicate to that of an *n-place predicate*, where  $n$  counts the number of distinct individual variables that stand in place of proper names. (Predicates with any positive number of places are also called *open sentences*.) Expression (91) is clearly a 2-place predicate. The only problem is to devise a convention for steering the right objects to the right variables. We do it as follows.

By the *free variables* of a predicate, we mean the individual variables which occur in proper name places in the predicate; so an  $n$ -place predicate has  $n$  free variables. (In Section 11 we shall have to revise this definition and exclude certain variables from being free.) A predicate with no free variables is called a sentence. We define an *assignment*  $g$  to a set of variables (in a situation) to be a function whose domain is that set of variables, with the stipulation that if  $x$  is a sorted variable then (in that situation)  $g(x)$  meets the restriction which goes with the variable. So for example  $g(y_{\text{raccoon}})$  has to be a raccoon.

We say that an assignment  $g$  is *suitable for* a predicate  $\phi$  if every free variable of  $\phi$  is in the domain of  $g$ . Using the inserted name definition of satisfaction as a temporary expedient, we define: if  $\phi$  is a predicate and  $g$  is an assignment which is suitable for  $\phi$ , then  $g$  *satisfies*  $\phi$  (in a given situation) iff a true sentence results (in that situation) when we replace each variable  $x$  in  $\phi$  by a name of the object  $g(x)$ .

We shall write

(92)  $\alpha/x, \beta/y, \gamma/z, \dots$

to name the assignment  $g$  such that  $g(x) = \alpha, g(y) = \beta, g(z) = \gamma$  etc. If  $\mathfrak{A}$  is a situation,  $\phi$  a predicate and  $g$  an assignment suitable for  $\phi$ , then we write

$$(93) \mathfrak{A} \models \phi[g]$$

to mean that  $g$  satisfies  $\phi$  in the situation  $\mathfrak{A}$ . The notation (93) is basic for all that follows, so let me give some examples. For simplicity I take  $\mathfrak{A}$  to be the real world here and now. The following are true:

$$(94) \mathfrak{A} \models \text{In the year } y, x \text{ was appointed Assistant Professor of Mathematics at } w \text{ at the age of 19 years. [Dr Harvey Friedman}/x, 1967/y, \text{Stanford University California}/w].$$

Example (94) asserts that in 1967 Dr Harvey Friedman was appointed Assistant Professor of Mathematics at Stanford University California at the age of 19 years; which must be true because the *Guinness Book of Records* says so.

$$(95) \mathfrak{A} \models v \text{ is the smallest number which can be expressed in two different ways as the sum of two squares. [65}/v].$$

$$(96) \mathfrak{A} \models x \text{ wrote poems about the physical anatomy of } x. [\text{Walt Whitman}/x].$$

This notation connects predicates with *objects*, not with names of objects. In (96) it is Mr Whitman himself who satisfies the predicate shown.

In the literature a slightly different and less formal convention is often used. The first time that a predicate  $\phi$  is mentioned, it is referred to, say, as  $\phi(y, t)$ . This means that  $\phi$  has at most the free variables  $y$  and  $t$ , and that these variables are to be considered *in that order*. To illustrate, let  $\phi(w, x, y)$  be the predicate

$$(97) \text{In the year } y, x \text{ was appointed Assistant Professor of Mathematics at } w \text{ at the age of 19 years.}$$

Then (94) will be written simply as

$$(98) \mathfrak{A} \models \phi [\text{Stanford University California, Dr Harvey Friedman, 1967}].$$

This handy convention can save us having to mention the variables again after the first time that a predicate is introduced.

There is another variant of (93) which is often used in the study of logics. Suppose that in situation  $\mathfrak{A}$ ,  $g$  is an assignment which is suitable for the predicate  $\phi$ , and  $S$  is a sentence which is got from  $\phi$  by replacing each free variable  $x$  in  $\phi$  by a name of  $g(x)$ . Then the truth-value of  $S$  is determined by  $\mathfrak{A}$ ,  $g$  and  $\phi$ , and it can be written

$$(99) g_{\mathfrak{A}}^*(\phi) \text{ or } \|\phi\|_{\mathfrak{A},g}.$$

So we have

$$(100) \mathfrak{A} \models \phi[g] \text{ iff } g_{\mathfrak{A}}^*(\phi) = T.$$

In (99),  $g_{\mathfrak{A}}^*$  can be thought of as a function taking predicates to truth-values. Sometimes it is abbreviated to  $g_{\mathfrak{A}}$  or even  $g$ , where this leads to no ambiguity.

## 11 QUANTIFIER NOTATION

Let us use the symbols  $x_{\text{boy}}, y_{\text{boy}}$  etc. as sorted variables which are restricted to boys. We shall read the two sentences

(101)  $\forall x_{\text{boy}} (x_{\text{boy}} \text{ has remembered to bring his woggle}).$

(102)  $\exists x_{\text{boy}} (x_{\text{boy}} \text{ has remembered to bring his woggle}).$

as meaning exactly the same as (103) and (104) respectively:

(103) Every boy has remembered to bring his woggle.

(104) Some boy has remembered to bring his woggle.

In other words, (101) is true in a situation if and only if in that situation, every member of the domain of quantification of  $\forall x_{\text{boy}}$  satisfies the predicate

(105)  $x_{\text{boy}}$  has remembered to bring his woggle.

Likewise (102) is true if and only if some member of the domain of quantification of  $\exists x_{\text{boy}}$  satisfies (105). The situation has to determine what the domain of quantification is, i.e. what boys are being talked about.

The expression  $\forall x_{\text{boy}}$  is called a *universal quantifier* and the expression  $\exists x_{\text{boy}}$  is called an *existential quantifier*. Because of the restriction ‘boy’ on the variable, they are called *sorted* or *sorted* quantifiers. The symbols  $\forall, \exists$  are called respectively the *universal* and *existential quantifier symbols*;  $\forall$  is read ‘for all’,  $\exists$  is read ‘for some’ or ‘there is’.

For unsorted quantifiers using plain variables  $x, y, z$ , etc., similar definitions apply, but now the domain of quantification for such a quantifier can be any class of things. Most uses of unsorted quantifiers are so remote from anything in ordinary language that we can’t rely on the conventions of speech to locate a domain of quantification for us. So instead we have to assume that *each situation specifies a class which is to serve as the domain of quantification for all unsorted quantifiers*. Then

(106)  $\forall x$  (if  $x$  is a boy then  $x$  has remembered to bring his woggle).

counts as true in a situation if and only if in that situation, every object in the domain of quantification satisfies the predicate

(107) if  $x$  is a boy then  $x$  has remembered to bring his woggle.

There is a corresponding criterion for the truth of a sentence starting with the unsorted existential quantifier  $\exists x$ ; the reader can easily supply it.

The occurrences of the variable  $x_{\text{boy}}$  in (101) and (102), and of  $x$  in (106), are no longer doing duty for pronouns or marking places where names can be inserted. They are simply part of the quantifier notation. We express this by

saying that these occurrences are *bound in* the respective sentences. We also say, for example, that the quantifier at the beginning of (101) *binds* the two occurrences of  $x_{\text{boy}}$  in that sentence. By contrast an occurrence of a variable in a predicate is called *free in* the predicate if it serves the role we discussed in Sections 9 and 10, of referring to whoever or whatever the predicate expresses a condition on. What we called the *free variables* of a predicate in Section 10 are simply those variables which have free occurrences in the predicate. Note that the concepts ‘free’ and ‘bound’ are relative: the occurrence of  $x_{\text{boy}}$  before ‘has’ in (101) is bound in (101) but free in (105). Consider also the predicate

(108)  $x_{\text{boy}}$  forgot his whistle, but  $\forall x_{\text{boy}}$  ( $x_{\text{boy}}$  has remembered to bring his woggle).

Predicate (108) expresses the condition which Billy satisfies if Billy forgot his whistle but every boy has remembered to bring his woggle. So the first occurrence of  $x_{\text{boy}}$  in (108) is free in (108) but the other two occurrences are bound in (108).

I should recall here the well-known fact that in natural languages, a pronoun can be linked to a quantifier phrase that occurs much earlier, even in a different sentence:

(109) HE: This evening I heard a nightingale in the pear tree.  
SHE: It was a thrush—we don’t get nightingales here.

In our notation this can’t happen. *Our quantifiers bind only variables in themselves and the clause immediately following them.* We express this by saying that the *scope* of an occurrence of a quantifier consists of the quantifier itself and the clause immediately following it; a quantifier occurrence  $\forall x$  or  $\exists x$  binds all and only occurrences of the same variable  $x$  which lie within its scope.

It is worth digressing for a moment to ask why (109) makes life hard for logicians. The crucial question is: just when is the woman’s remark ‘It was a thrush’ a true statement? We want to say that it’s true if and only if the object referred to by ‘It’ is a thrush. But what is there for ‘It’ to refer to? Arguably the man hasn’t referred to any nightingale, he has merely said that there was at least one that he heard in the pear tree. Also we want to say that if her remark is true, then it follows that he heard a thrush in the pear tree. But if this follows, why doesn’t it also follow that the nightingale in the pear tree was a thrush? (which is absurd.)

There is a large literature on the problems of cross-reference in natural languages. See for example [Chastain, 1975; Partee, 1978; Evans, 1980]. In the early 1980s Hans Kamp and Irene Heim independently proposed formalisms to handle the matter systematically ([Kamp, 1981; Heim, 1988]; see also [Kamp and Reyle, 1993]). These new formalisms are fundamentally different from first-order logic. Jeroen Groenendijk and Martin Stokhof



[1991] gave an ingenious new semantics for first-order logic which is based on Kamp's ideas and allows a quantifier to pick up a free variable in a later sentence. Their underlying idea is that the meaning of a sentence is the change which it makes to the information provided by earlier sentences in the conversation. This opens up new possibilities, but it heads in a very different direction from the usual first-order logic.

Returning to first-order logic, consider the sentence

(110)  $\exists x_{\text{boy}}(x_{\text{boy}} \text{ kissed Brenda}).$

This sentence can be turned into a predicate by putting a variable in place of 'Brenda'. Naturally the variable we use has to be different from  $x_{\text{boy}}$ , or else it would get bound by the quantifier at the beginning. Apart from that constraint, any variable will do. For instance:

(111)  $\exists x_{\text{boy}}(x_{\text{boy}} \text{ kissed } y_{\text{girlwithpigtails}}).$

We need to describe the conditions in which Brenda satisfies (111). Brenda must of course be a girl with pigtails. She satisfies (111) if and only if there is a boy  $\beta$  such that the assignment

(112)  $\beta/x_{\text{boy}}, \text{ Brenda}/y_{\text{girlwithpigtails}}$

satisfies the predicate ' $x_{\text{boy}} \text{ kissed } y_{\text{girlwithpigtails}}$ '. Formal details will follow in Section 14 below.

## 12 AMBIGUOUS CONSTANTS

In his *Wissenschaftslehre II* [1837, Section 147] Bernard Bolzano noted that we use demonstrative pronouns at different times and places to refer now to this, now to that. He continued:

Since we do this anyhow, it is worth the effort to undertake this procedure with full consciousness and with the intention of gaining more precise knowledge about the nature of such propositions by observing their behaviour with respect to truth. Given a proposition, we could merely inquire whether it is true or false. But some very remarkable properties of propositions can be discovered if, in addition, we consider the truth values of all those propositions which can be generated from it, if we take some of its constituent ideas as variable and replace them by any other ideas whatever.

We can abandon to the nineteenth century the notion of 'variable ideas'. What Bolzano did in fact was to introduce *totally ambiguous symbols*. When a writer uses such a symbol, he has to indicate what it means, just as he has

to make clear what his demonstrative pronouns refer to. In our terminology, the situation must fix the meanings of such symbols. Each totally ambiguous symbol has a certain grammatical type, and the meaning supplied must fit the grammatical type; but that apart, anything goes.

Let us refer to a sentence which contains totally ambiguous symbols as a *sentence schema*. Then an *argument schema* will consist of a string of sentence schemas called *premises*, followed by the word ‘*Therefore*’, followed by a sentence schema called the *conclusion*. A typical argument schema might be:

(113) *a* is more *X* than *b*. *b* is more *X* than *c*. *Therefore a* is more *X* than *c*.

A traditional logician would have said that (113) is a valid argument schema if and only if all its instances are valid arguments (cf. (1) in the Introduction above). Bolzano said something different. Following him, we shall say that (113) is *Bolzano-valid* if for every situation in which *a, b, c* are interpreted as names and *X* is interpreted as an adjective, either one or more of the premises are not true, or the conclusion is true. We say that the premises in (113) *Bolzano-entail* the conclusion if (113) is Bolzano-valid.

Note the differences. For the traditional logician entailment is from sentences to sentences, not from sentence schemas to sentence schemas. Bolzano’s entailment is between schemas, not sentences, and moreover he defines it without mentioning entailment between sentences. The schemas become sentences of a sort when their symbols are interpreted, but Bolzano never asks whether these sentences “can’t be true without certain other sentences being true” (to recall our definition of entailment in the Introduction)—he merely asks when they *are* true.

The crucial relationship between Bolzano’s ideas and the traditional ones is that *every instance of a Bolzano-valid argument schema is a valid argument*. If an argument is an instance of a Bolzano-valid argument schema, then that fact itself is a reason why the premises can’t be true without the conclusion also being true, and so the argument is valid. The traditional logician may want to add a caution here: the argument need not be *logically* valid unless the schema is Bolzano-valid for *logical* reasons—whatever we take ‘logical’ to mean. Tarski [1936] made this point. (Let me take the opportunity to add that recent discussions of the nature of logical consequence have been clouded by some very unhistorical readings of [Tarski, 1936]. Fortunately there is an excellent historical analysis by Gómez-Torrente [1996].)

In first-order logic we follow Bolzano and study entailments between schemas. We use two kinds of totally ambiguous constants. The first kind are the *individual constants*, which are normally chosen from lower-case letters near the beginning of the alphabet: *a, b, c* etc. These behave grammatically as singular proper names, and are taken to stand for objects. The other kind are the *predicate (or relation) constants*. These are usually cho-

sen from the letters  $P, Q, R$  etc. They behave as verbs or predicates, in the following way. To specify a meaning for the predicate constant  $P$ , we could write

(114)  $Pxyz$  means  $x$  aimed at  $y$  and hit  $z$ .

The choice of variables here is quite arbitrary, so (114) says the same as:

(115)  $Pyst$  means  $y$  aimed at  $s$  and hit  $t$ .

We shall say that under the interpretation (114), an ordered 3-tuple  $\langle \alpha, \beta, \gamma \rangle$  of objects *satisfies*  $P$  if and only if the assignment

(116)  $\alpha/x, \beta/y, \gamma/z$

satisfies the predicate ‘ $x$  aimed at  $y$  and hit  $z$ ’. So for example the ordered 3-tuple  $\langle \text{Bert}, \text{Angelo}, \text{Chen} \rangle$  satisfies  $P$  under the interpretation (114) or (115) if and only if Bert aimed at Angelo and hit Chen. (We take  $P$  to be satisfied by ordered 3-tuples rather than by assignments because, unlike a predicate, the symbol  $P$  comes without benefit of variables.) The collection of all ordered 3-tuples which satisfy  $P$  in a situation where  $P$  has the interpretation (114) is called the *extension* of  $P$  in that situation. In general a collection of ordered  $n$ -tuples is called an  *$n$ -place relation*.

Since  $P$  is followed by three variables in (114), we say that  $P$  in (114) is serving as a *3-place predicate constant*. One can have  $n$ -place predicate constants for any positive integer  $n$ ; the extension of such a constant in a situation is always an  $n$ -place relation. In theory a predicate constant could be used both as a 3-place and as a 5-place predicate constant in the same setting without causing mishap, but in practice logicians try to avoid doing this.

Now consider the sentence

(117)  $\forall x$  (if  $Rxc$  then  $x$  is red).

with 2-place predicate constant  $R$  and individual constant  $c$ . What do we need to be told about a situation  $\mathfrak{A}$  in order to determine whether (117) is true or false in  $\mathfrak{A}$ ? The relevant items in  $\mathfrak{A}$  seem to be:

- (a) the domain of quantification for  $\forall x$ .
- (b) the object named by the constant  $c$ . (Note: it is irrelevant what meaning  $c$  has over and above naming this object, because  $R$  will be interpreted by a predicate.) We call this object  $I_{\mathfrak{A}}(c)$ .
- (c) the extension of the constant  $R$ . (Note: it is irrelevant what predicate is used to give  $R$  this extension; the extension contains all relevant information.) We call this extension  $I_{\mathfrak{A}}(R)$ .

(d) the class of red things.

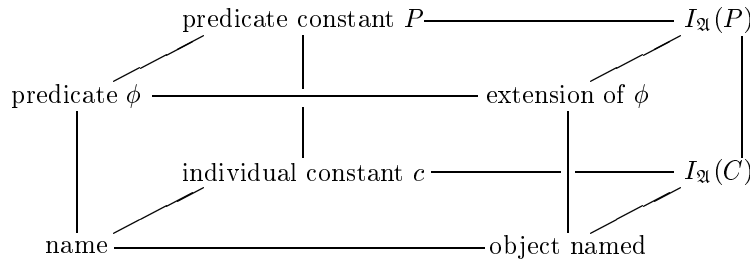
In Section 14 we shall define the important notion of a *structure* by extracting what is essential from (a)–(d). Logicians normally put into the definition of ‘structure’ some requirements that are designed to make them simpler to handle. Before matters get buried under symbolism, let me say what these requirements amount to in terms of  $\mathfrak{A}$ . (See Appendix C below for the set-theoretic notions used.)

1. There is to be a collection of objects called the *domain* of  $\mathfrak{A}$ , in symbols  $|\mathfrak{A}|$ .
2.  $|\mathfrak{A}|$  is the domain of quantification for all unsorted quantifiers. Two sorted quantifiers with variables of the same sort (if there are any) always have the same domain of quantification, which is included in  $|\mathfrak{A}|$ .
3. For every individual constant  $c$ , the interpretation  $I_{\mathfrak{A}}(c)$  is a member of  $|\mathfrak{A}|$ ; for every predicate constant  $R$ , the relation  $I_{\mathfrak{A}}(R)$  is a relation on  $|\mathfrak{A}|$ .
4. Some authors require  $|\mathfrak{A}|$  to be a pure set. Most authors require it to have at least one member. A very few authors (e.g. [Carnap, 1956; Hintikka, 1955]) require it to be at most countable.

Requirements (1)–(3) mean in effect that first-order logicians abandon any pretence of following the way that domains of quantification are fixed in natural languages. Frege’s device of Section 9 (e.g. (84)) shows how we can meet these requirements and still say what we wanted to say, though at greater length. Requirements (4) are an odd bunch; I shall study their reasons and justifications in due course below.

Logicians also allow one important relaxation of (1)–(4). They permit an  $n$ -place predicate symbol to be interpreted by *any*  $n$ -place relation on the domain, not just one that comes from a predicate. Likewise they permit an individual constant to stand for any member of the domain, regardless of whether we can identify that member. The point is that the question whether *we* can describe the extension or the member is totally irrelevant to the question what is true in the structure.

Note here the 3-way analogy



The front face of this cube is essentially due to Frege. Would he have accepted the back?

No, he would not. In 1899 Hilbert published a study of the axioms of geometry. Among other things, he asked questions of the form ‘Do axioms  $A, B, C$  together entail axiom  $D$ ?’ (The famous problem of the independence of Euclid’s parallel postulate is a question of this sort.) Hilbert answered these questions by regarding the axioms as schemas containing ambiguous signs, and then giving number-theoretic interpretations which made the premises  $A, B$  and  $C$  true but the conclusion  $D$  false. Frege read the book [Hilbert, 1899] and reacted angrily. After a brief correspondence with Hilbert (Frege and Hilbert [1899–1900]), he published a detailed critique [1906], declaring [Frege, 1971, p. 66]: “Indeed, if it were a matter of deceiving oneself and others, there would be no better means than ambiguous signs.”

Part of Frege’s complaint was that Hilbert had merely shown that certain argument schemas were not Bolzano-valid; he had not shown that axioms  $A, B$  and  $C$ , taken literally as statements about points, lines etc. in real space, do not entail axiom  $D$  taken literally. This is true and need not detain us—Hilbert had answered the questions he wanted to answer. Much more seriously, Frege asserted that Hilbert’s propositions, being ambiguous, did not express determinate thoughts and hence could not serve as the premises or conclusions of inferences. In short, Frege refused to consider Bolzano-valid argument schemas as any kind of valid argument. So adamant was he about this that he undertook to translate the core of Hilbert’s reasoning into what he considered an acceptable form which never mentioned schematic sentences. This is not difficult to do—it is a matter of replacing statements of the form ‘Axiom  $A$  entails axiom  $B$ ’ by statements of the form ‘For all relations  $P$  and  $R$ , if  $P$  and  $R$  do this then they do that’. But the resulting translation is quite unreadable, so good mathematics is thrown away and all for no purpose.

Frege’s rejection of ambiguous symbols is part and parcel of his refusal to handle indexical expressions; see [Perry, 1977] for some discussion of the issue. It is sad to learn that the grand architect of modern logic fiercely rejected the one last advance which was needed to make his ideas fruitful.

In fact it took some years for logicians to accept the use of ambiguous symbols in the semantics of first-order logic. For example Tarski's paper [1936] on logical deduction made no use of them; Tarski found another device with the same effect (at the cost of adapting the word 'model' to mean 're-interpretation' rather than 'interpretation'). But in his model-theoretic work of the 1950s and later, Tarski used ambiguous constants wholesale in the modern fashion, as a form of indexical. (Cf. [Hodges, 1985/86].)

### 13 FIRST-ORDER SYNTAX FORMALISED

The main purpose of this section and the next is to extract the formal content of Sections 9–12 above. I give the definitions first under the assumption that there are no sorted variables. Also I ignore for the moment the fact that some first-order logicians use = and function symbols. Section 18 below will be more broad-minded.

A *similarity type* is defined to be a set of individual constants together with a set of predicate constants; each predicate constant is assumed to be labelled somehow to indicate that it is an  $n$ -place predicate constant, for some positive integer  $n$ . Some writers include the  $n$  as a superscript:  $R^{133}$  is a 133-place predicate constant.

We shall define the *first-order language*  $L$  of *similarity type*  $X$ . For definiteness,  $L$  shall be an ordered triple  $\langle X, T(X), F(X) \rangle$  where  $X$  is the similarity type, and  $T(X)$  and  $F(X)$  are respectively the set of all terms and formulas of similarity type  $X$  (known more briefly as the terms and formulas of  $L$ ). Grammatically speaking, the terms of  $L$  are its noun phrases and the formulas are its sentences. Metavariables  $\sigma, \tau$  will range over terms, and metavariables  $\phi, \psi, \chi$  will range over formulas.

We start the definition by defining the *variables* to be the countably many symbols

$$(118) \quad x_0, x_1, x_2, \dots$$

Unofficially everybody uses the symbol  $x, y, z$  etc. as variables. But in the spirit of Section 4 above, these can be understood as metavariables ranging over variables. The *terms* of  $L$  are defined to be the variables of  $L$  and the individual constants in  $X$ .

An *atomic formula* of  $L$  is an expression of form  $P(\sigma_1, \dots, \sigma_n)$  where  $P$  is an  $n$ -place predicate constant in  $X$  and  $\sigma_1, \dots, \sigma_n$  are terms of  $L$ . The class of *formulas* of  $L$  is defined inductively, and as the induction proceeds we shall define also the set of subformulas of the formula  $\phi$ , and the set  $FV(\phi)$  of free variables of  $\phi$ :

- (a) Every atomic formula  $\phi$  of  $L$  is a formula of  $L$ ; it is its only subformula, and  $FV(\phi)$  is the set of all variables which occur in  $\phi$ .  $\perp$  is a formula of  $L$ ; it is its only subformula, and  $FV(\perp)$  is empty.

- (b) Suppose  $\phi$  and  $\psi$  are formulas of  $L$  and  $x$  is a variable. Then:  $\neg\phi$  is a formula of  $L$ ; its subformulas are itself and the subformulas of  $\phi$ ;  $FV(\neg\phi)$  is  $FV(\phi)$ . Also  $(\phi \wedge \psi)$ ,  $(\phi \vee \psi)$ ,  $(\phi \rightarrow \psi)$  and  $(\phi \leftrightarrow \psi)$  are formulas of  $L$ ; the subformulas of each of these formulas are itself, the subformulas of  $\phi$  and the subformulas of  $\psi$ ; its free variables are those of  $\phi$  together with those of  $\psi$ . Also  $\forall x\phi$  and  $\exists x\phi$  are formulas of  $L$ ; for each of these, its subformulas are itself and the subformulas of  $\phi$ ; its free variables are those of  $\phi$  excluding  $x$ .
- (c) Nothing is a formula of  $L$  except as required by (a) and (b).

The *complexity* of a formula  $\phi$  is defined to be the number of subformulas of  $\phi$ . This definition disagrees with that in Section 3, but it retains the crucial property that every formula has a higher complexity than any of its proper subformulas. (The *proper subformulas* of  $\phi$  are all the subformulas of  $\phi$  except  $\phi$  itself.) A formula is said to be *closed*, or to be a *sentence*, if it has no free variables. Closed formulas correspond to sentences of English, non-closed formulas to predicates or open sentences of English. Formulas of a formal language are sometimes called *well-formed formulas* or *wffs* for short.

If  $\phi$  is a formula,  $x$  is a variable and  $\tau$  is a term, then there is a formula  $\phi[\tau/x]$  which ‘says the same thing about the object  $\tau$  as  $\phi$  says about the object  $x$ ’. At a first approximation,  $\phi[\tau/x]$  can be described as the formula which results if we put  $\tau$  in place of each free occurrence of  $x$  in  $\phi$ ; when this description works, we say  $\tau$  is *free for  $x$  in  $\phi$*  or *substitutable for  $x$  in  $\phi$* . Here is an example where the approximation doesn’t work:  $\phi$  is  $\exists yR(x, y)$  and  $\tau$  is  $y$ . If we put  $y$  for  $x$  in  $\phi$ , the resulting formula  $\exists yR(y, y)$  says nothing at all about ‘the object  $y$ ’, because the inserted  $y$  becomes bound by the quantifier  $\exists y$ —a phenomenon known as *clash of variables*. In such cases we have to define  $\phi[\tau/x]$  to be  $\exists zR(y, z)$  where  $z$  is some other variable. (There is a good account of this messy matter in Bell and Machover [1977, Chapter 2, Section 3].)

Note the useful shorthand: if  $\phi$  is described at its first occurrence as  $\phi(x)$ , then  $\phi(\tau)$  means  $\phi[\tau/x]$ . Likewise if  $\phi$  is introduced as  $\phi(y_1, \dots, y_n)$  then  $\phi(\tau_1, \dots, \tau_n)$  means the formula which says about the objects  $\tau_1, \dots, \tau_n$  the same thing as  $\phi$  says about the objects  $y_1, \dots, y_n$ .

Not much in the definitions above needs to be changed if you want a system with sorted variables. You must start by deciding what kind of sortal system you want. There will be a set  $S$  of sorts  $s, t$  etc., and for each sort  $s$  there will be sorted variables  $x_0^s, s_1^s, x_2^s$  etc. But then (a) do you want every object to belong to some sort? If so, the similarity type must assign each individual constant to at least one sort. (b) Do you want the sorts to be mutually exclusive? Then the similarity type must assign each individual constant to at most one sort. (c) Do you want to be able to say

‘everything’, rather than just ‘everything of such-and-such a sort’? If not then the unsorted variables (118) should be struck out.

Some formal languages allow restricted quantification. For example in languages designed for talking about numbers, we have formulas  $(\forall x < y)\phi$  and  $(\exists x < y)\phi$ , read respectively as ‘For all numbers  $x$  less than  $y$ ,  $\phi$ ’ and ‘There is a number  $x$  less than  $y$  such that  $\phi$ ’. These expressions can be regarded as metalanguage abbreviations for  $\forall x(x < y \rightarrow \phi)$  and  $\exists x(x < y \wedge \phi)$  respectively (where ‘ $x < y$ ’ in turn is an abbreviation for ‘ $< (x, y)$ ’). Or we can alter the definition of ‘formula of L’ to allow restricted quantifiers in L itself.

One often sees abbreviations such as ‘ $\forall xy\phi$ ’ or ‘ $\exists z\phi$ ’. These are metalanguage abbreviations.  $\forall xy$  is short for  $\forall x\forall y$ .  $z$  means a finite sequence  $z_1, \dots, z_n$ . Furthermore, the abbreviations of Section 4 remain in force.

All the syntactic notions described in this section can be defined using only concrete instances of the induction axiom as in Section 3 above.

#### 14 FIRST-ORDER SEMANTICS FORMALISED

We turn to the definition of structures. (They are also known as *models*—but it is better to reserve this term for the context ‘model of  $\phi$ ’.) Let L be a language with similarity type  $X$ . Then an L-*structure*  $\mathfrak{A}$  is defined to be an ordered pair  $\langle A, I \rangle$  where:

1.  $A$  is a class called the *domain* of  $\mathfrak{A}$ , in symbols  $|\mathfrak{A}|$ . The elements of  $A$  are called the *elements* of  $\mathfrak{A}$ , and the cardinality of  $A$  is called the *cardinality* of  $\mathfrak{A}$ . So for example we call  $\mathfrak{A}$  *finite* or *empty* if  $A$  is finite or empty. Many writers use the convention that  $A, B$  and  $C$  are the domains of  $\mathfrak{A}, \mathfrak{B}$  and  $\mathfrak{C}$  respectively.
2.  $I$  is a function which assigns to each individual constant  $c$  of  $X$  an element  $I(c)$  of  $A$ , and to each  $n$ -place predicate symbol  $R$  of  $X$  an  $n$ -place relation  $I(R)$  on  $A$ .  $I$  is referred to as  $I_{\mathfrak{A}}$ .

*Structure* means: L-structure for some language L.

If  $Z$  is a set of variables, then an *assignment* to  $Z$  in  $\mathfrak{A}$  is defined to be a function from  $Z$  to  $A$ . If  $g$  is an assignment to  $Z$  in  $\mathfrak{A}$ ,  $x$  is a variable not in  $Z$  and  $\alpha$  is an element of  $\mathfrak{A}$ , then we write

$$(119) \quad g, \alpha/x$$

for the assignment  $h$  got from  $g$  by adding  $x$  to  $g$ ’s domain and putting  $h(x) = \alpha$ . (Some writers call assignments *valuations*.)

For each assignment  $g$  in  $\mathfrak{A}$  and each individual constant  $c$  we define  $c[g]$  to be the element  $I_{\mathfrak{A}}(c)$ . For each variable  $x$  and assignment  $g$  whose domain contains  $x$ , we define  $x[g]$  to be the element  $g(x)$ . Then  $\tau[g]$  is ‘the element named by the term  $\tau$  under the assignment  $g$ ’.



For each formula  $\phi$  of  $L$  and each assignment  $g$  to the free variables of  $\phi$  in  $\mathfrak{A}$ , we shall now define the conditions under which  $\mathfrak{A} \models \phi[g]$  (cf. (93) above). The definition is by induction on the complexity of  $\phi$ .

- (a) If  $R$  is an  $n$ -place predicate constant in  $X$  and  $\tau_1, \dots, \tau_n$  are terms, then  $\mathfrak{A} \models R(\tau_1, \dots, \tau_n)$  iff the ordered  $n$ -tuple  $\langle \tau_1[g], \dots, \tau_n[g] \rangle$  is in  $I_{\mathfrak{A}}(R)$ .
- (b) It is never true that  $\mathfrak{A} \models \perp$ .
- (c)  $\mathfrak{A} \models \neg\phi[g]$  iff it is not true that  $\mathfrak{A} \models \phi[g]$ .  
 $\mathfrak{A} \models \phi \wedge \psi[g]$  iff  $\mathfrak{A} \models \phi[g_1]$  and  $\mathfrak{A} \models \psi[g_2]$ , where  $g_1$  and  $g_2$  are the results of restricting  $g$  to the free variables of  $\phi$  and  $\psi$  respectively.  
 Etc. as in (23).
- (d) If  $x$  is a free variable of  $\phi$ , then:  
 $\mathfrak{A} \models \forall x\phi[g]$  iff for every element  $\alpha$  of  $A$ ,  $\mathfrak{A} \models \phi[g, \alpha/x]$ ;  
 $\mathfrak{A} \models \exists x\phi[g]$  iff for at least one element  $\alpha$  of  $A$ ,  $\mathfrak{A} \models \phi[g, \alpha/x]$ .  
 If  $x$  is not a free variable of  $\phi$ , then  $\mathfrak{A} \models \forall x\phi[g]$  iff  $\mathfrak{A} \models \phi[g]$ , and  
 $\mathfrak{A} \models \exists x\phi[g]$  iff  $\mathfrak{A} \models \phi[g]$ .

We say an assignment  $g$  in  $\mathfrak{A}$  is *suitable for* the formula  $\phi$  if every free variable of  $\phi$  is in the domain of  $g$ . If  $g$  is suitable for  $\phi$ , we say that  $\mathfrak{A} \models \phi[g]$  if and only if  $\mathfrak{A} \models \phi[h]$ , where  $h$  comes from  $g$  by throwing out of the domain of  $g$  those variables which are not free variables of  $\phi$ .

If  $\phi$  is a sentence, then  $\phi$  has no free variables and we can write just  $\mathfrak{A} \models \phi$  in place of  $\mathfrak{A} \models \phi[ ]$ . This notation agrees with (22) above. When  $\mathfrak{A} \models \phi$ , we say that  $\mathfrak{A}$  is a *model of*  $\phi$ , or that  $\phi$  is *true in*  $\mathfrak{A}$ . ' $\mathfrak{A} \models \phi[g]$ ' can be pronounced '*g satisfies  $\phi$  in  $\mathfrak{A}$* '.

To anybody who has mastered the symbolism it should be obvious that clauses (a)–(d) really do determine whether or not  $\mathfrak{A} \models \phi$ , for every  $L$ -structure  $\mathfrak{A}$  and every sentence  $\phi$  of  $L$ . If  $\mathfrak{A}$  is a set then we can formalise the definition in the language of set theory and prove that it determines  $\models$  uniquely, using only quite weak set-theoretic axioms (cf. [Barwise, 1975, Chapter 3]). Set structures are adequate for most applications of first-order logic in mathematics, so that many textbooks simply state without apology that a structure has to be a set. We shall return to this point in Section 17 below.

The definition of  $\models$  given above is called the *truth-definition*, because it specifies exactly when a symbolic formula is to count as 'true in' a structure. It solves no substantive problems about what is true—we are just as much in the dark about the Riemann hypothesis or the Reichstag fire after writing it down as we were before. But it has attracted a lot of attention as a possible answer to the question of what is Truth. Many variants of it have appeared in the literature, which can cause anguish to people anxious to get to the

heart of the matter. Let me briefly describe three of these variants; they are all mathematically equivalent to the version given above. (Cf. Leblanc [Volume 2 of this *Handbook*].)

In the first variant, *assignments are sequences*. More precisely an assignment in  $\mathfrak{A}$  is defined to be a function  $g$  from the natural numbers  $N$  to the domain  $A$  of  $\mathfrak{A}$ . Such a function can be thought of as an infinite sequence  $\langle g(0), g(1), g(2), \dots \rangle$ . The element  $g(i)$  is assigned to the  $i$ th variable  $x_i$ , so that  $x_i[g]$  is defined to be  $g(i)$ . In (c) and (d) we have to make some changes for the purely technical reason that  $g$  assigns elements to *every* variable and not just those free in  $\phi$ . In (c) the clause for  $\phi \wedge \psi$  becomes

$$\mathfrak{A} \models \phi \wedge \psi[g] \quad \text{iff} \quad \mathfrak{A} \models \phi[g] \text{ and } \mathfrak{A} \models \psi[g],$$

which is an improvement (and similarly with  $(\phi \vee \psi)$ ,  $(\phi \rightarrow \psi)$  and  $(\phi \leftrightarrow \psi)$ ). But (d) becomes distorted, because  $g$  already makes an assignment to the quantified variable  $x$ ; this assignment is irrelevant to the truth of  $\mathfrak{A} \models \forall x \phi[g]$ , so we have to discard it as follows. For each number  $i$  and element  $\alpha$  of  $\mathfrak{A}$ , let  $g(\alpha/i)$  be the assignment  $h$  which is exactly like  $g$  except that  $h(i) = \alpha$ . Then (d) is replaced by:

$$(d') \text{ For each variable } x_i : \mathfrak{A} \models \forall x_i \phi[g] \text{ iff for every element } \alpha \text{ of } A, \mathfrak{A} \models \phi[g(\alpha/i)].$$

together with a similar clause for  $\exists x_i \phi$ .

In the second variant, we copy (24) and define the *truth-value* of  $\phi$  in  $\mathfrak{A}$ ,  $\|\phi\|_{\mathfrak{A}}$ , to be the set of all assignments  $g$  to the free variables of  $\phi$  such that  $\mathfrak{A} \models \phi[g]$ . When  $\phi$  is a sentence, there is only one assignment to the free variables of  $\phi$ , namely the empty function 0; so  $\|\phi\|_{\mathfrak{A}}$  is  $\{0\}$  if  $\phi$  is true in  $\mathfrak{A}$ , and the empty set (again 0) if  $\phi$  is false in  $\mathfrak{A}$ . This variant is barely more than a change of notation. Instead of ' $\mathfrak{A} \models \phi[g]$ ' we write ' $g \in \|\phi\|_{\mathfrak{A}}$ '. The clauses (a)–(d) can be translated easily into the new notation.

Some writers combine our first and second variants, taking  $\|\phi\|_{\mathfrak{A}}$  to be the set of all sequences  $g$  such that  $\mathfrak{A} \models \phi[g]$ . In this style, the clause for  $\phi \wedge \psi$  in (c) becomes rather elegant:

$$\|\phi \wedge \psi\|_{\mathfrak{A}} = \|\phi\|_{\mathfrak{A}} \cap \|\psi\|_{\mathfrak{A}}.$$

However, when  $\phi$  is a sentence the definition of ' $\phi$  is true in  $\mathfrak{A}$ ' becomes 'every sequence is in  $\|\phi\|_{\mathfrak{A}}$ ', or equivalently 'at least one sequence is in  $\|\phi\|_{\mathfrak{A}}$ '. I have heard students repeat this definition with baffled awe as if they learned it in the Eleusinian Mysteries.

The third variant dispenses with assignments altogether and adds new constant names to the language  $L$ . Write  $L(c)$  for the language got from  $L$  by adding  $c$  as an extra individual constant. If  $\mathfrak{A}$  is an  $L$ -structure and  $\alpha$  is an element of  $\mathfrak{A}$ , write  $(\mathfrak{A}, \alpha)$  for the  $L(c)$ -structure  $\mathfrak{B}$  which is the same as  $\mathfrak{A}$  except that  $I_{\mathfrak{B}}(c) = \alpha$ . If  $\phi$  is a formula of  $L$  with just the free variable  $x$ , one can prove by induction on the complexity of  $\phi$  that

$$(120) \quad (\mathfrak{A}, \alpha) \models \phi[c/x] \quad \text{iff} \quad \mathfrak{A} \models \phi[\alpha/x].$$

(Warning:  $[c/x]$  on the left is a substitution in the formula  $\phi$ ;  $\alpha/x$  on the right is an assignment to the variable  $x$ .) The two sides in (120) are just different ways of expressing that  $\alpha$  satisfies  $\phi$  in  $\mathfrak{A}$ . Hence we have

$$(121) \quad \mathfrak{A} \models \forall x\phi \quad \text{iff} \quad \text{for every element } \alpha \text{ of } \mathfrak{A}, (\mathfrak{A}, \alpha) \models \phi[c/x],$$

and a similar clause for  $\exists x\phi$ . In our third variant, (121) is taken as the *definition* of  $\models$  for sentences of form  $\forall x\phi$ . This trick sidesteps assignments. Its disadvantage is that we have to alter the language and the structure each time we come to apply clause (d). The great merit of assignments is that they enable us to keep the structure fixed while we wiggle around elements in order to handle the quantifiers.

There are L-structures whose elements are all named by individual constants of L. For example, the natural numbers are sometimes understood as a structure in which every number  $n$  is named by a numeral constant  $\bar{n}$  of the language. For such structures, *and only for such structures*, (121) can be replaced by

$$(122) \quad \mathfrak{A} \models \forall x\phi \quad \text{iff} \quad \text{for every individual constant } c \text{ of } L, \mathfrak{A} \models \phi[c/x].$$

Some writers confine themselves to structures for which (122) applies.

Alfred Tarski's famous paper on the concept of truth in formalised languages [1935] was the first paper to present anything like our definition of  $\models$ . Readers should be aware of one vital difference between his notion and ours. His languages have no ambiguous constants. True, Tarski says they have constants. But he explains that by 'constants' he means negation signs, quantifier symbols and suchlike, together with symbols of fixed meaning such as the inclusion sign  $\subseteq$  in set theory. (See Section 20 below on symbols with an 'intended interpretation'.) The only concession that Tarski makes to the notion of an L-structure is that he allows the domain of elements to be any class, not necessarily the class of everything. Even then he says that relativising to a particular class is 'not essential for the understanding of the main theme of this work'! (Cf. pages 199, 212 of the English translation of [Tarski, 1935].) Carnap's truth-definition [1935] is also little sideways from modern versions.

There is no problem about adapting Tarski's definition to our setting. It can be done in several ways. Probably the simplest is to allow some of his constants to turn ambiguous; then his definition becomes our first variant.

Finally I should mention structures for many-sorted languages, if only to say that no new issues of principle arise. If the language L has a set  $S$  of sorts, then for each sort  $s$  in  $S$ , an L-structure  $\mathfrak{A}$  must carry a class  $s(\mathfrak{A})$  of *elements of sort*  $s$ . In accordance with Section 12,  $s(\mathfrak{A})$  must be included in  $|\mathfrak{A}|$ . If the individual constant  $c$  is of sort  $s$ , then  $I_{\mathfrak{A}}(c)$  must be an element of  $s(\mathfrak{A})$ . If we have required that every element should be of at least one sort, then  $|\mathfrak{A}|$  must be the union of the classes  $s(\mathfrak{A})$ .

## 15 FIRST-ORDER IMPLICATIONS

Let me make a leap that will seem absurd to the Traditional Logician, and define sequents with infinitely many premises.

Suppose  $L$  is a first-order language. By a *theory in  $L$*  we shall mean a set of sentences of  $L$ —it can be finite or infinite. The metavariables  $\Delta, \Gamma, \Theta, \Lambda$  will range over theories. If  $\Delta$  is a theory in  $L$  and  $\mathfrak{A}$  is an  $L$ -structure, we say that  $\mathfrak{A}$  is a *model of  $\Delta$*  if  $\mathfrak{A}$  is a model of every sentence in  $\Delta$ .

For any theory  $\Delta$  in  $L$  and sentence  $\phi$  of  $L$ , we define

$$(123) \quad \Delta \vDash \phi \quad (\text{'}\Delta \text{ logically implies } \phi\text{'}, \text{'}\phi \text{ is a logical consequence of } \Delta\text{'})$$

to mean that every  $L$ -structure which is a model of  $\Delta$  is also a model of  $\phi$ . If  $\Delta$  has no models, (123) is reckoned to be true by default. A *counterexample* to (123) is an  $L$ -structure which is a model of  $\Delta$  but not of  $\phi$ . We write

$$(124) \quad \vDash \phi \quad (\text{'}\phi \text{ is logically valid'})$$

to mean that every  $L$ -structure is a model of  $\phi$ ; a *counterexample* to (124) is an  $L$ -structure which is not a model of  $\phi$ . The expressions (123) and (124) are called *sequents*. This definition of logical implication was first set down by Tarski [1936], though it only makes precise what Bolzano [1837, Section 155] and Hilbert [1899] already understood.

Warning: (123) is a definition of logical consequence *for first-order schemas*. It doesn't make sense as a definition of logical consequence between meaningful sentences, even when the sentences are written in first-order notation; logical consequence might hold between the sentences for reasons not expressed in the first-order notation. This is obvious: let ' $p$ ' stand for your favourite logical truth, and consider ' $\vDash p$ '. I mention this because I have seen a small river of philosophical papers which criticise (123) under the impression that it is intended as a definition of logical consequence between sentences of English (they call it the 'model-theoretic definition of logical consequence'). In one case where I collared the author and traced the mistake to source, it turned out to be a straight misreading of that excellent textbook [Enderton, 1972]; though I am not sure the author accepted my correction. One can track down some of these confusions to the terminology of Etchemendy [1990], who uses phrases such as 'the set of logical truths of any given first-order language' [Etchemendy, 1990, p. 148] to mean those sentences of a *fully interpreted* first-order language which are (in Etchemendy's sense) intuitively logically true. In his Chapter 11 especially, Etchemendy's terminology is way out of line with that of the authors he is commenting on.

If the language  $L$  has at least one individual constant  $c$ , then every  $L$ -structure must have an element  $I_{\mathfrak{A}}(c)$ , so the domain of  $\mathfrak{A}$  can't be empty. It follows that in this language the sentence  $\exists x \neg \perp$  must be logically valid, so we can 'prove' that at least one thing exists.

On the other hand if  $L$  has no individual constants, then there is an  $L$ -structure whose domain is empty. This is not just a quirk of our conventions: one can quite easily think of English sentences uttered in contexts where the natural domain of quantification happens to be empty. In such a language  $L$ ,  $\exists x\neg\perp$  is not logically valid.

This odd state of affairs deserves some analysis. Suppose  $L$  does have an individual constant  $c$ . By the Bolzano–Tarski definition (123), when we consider logical implication in  $L$  we are only concerned with structures in which  $c$  names something. In other words, the Bolzano–Tarski definition slips into every argument a tacit premise that *every name does in fact name something*. If we wanted to, we could adapt the Traditional Logician’s notion of a valid argument in just the same way. For a traditional example, consider

(125) Every man runs. *Therefore* Socrates, if he is a man, runs.

On the traditional view, (125) is not a valid argument—it could happen that every man runs and yet there is no such entity as Socrates. On the Bolzano–Tarski view we must consider only situations in which ‘Socrates’ names something or someone, and on that reckoning, (125) is valid. (According to Walter Burleigh in the fourteenth century, (125) is not valid outright, but it is valid at the times when Socrates exists. Cf. Bocheński [1970, p. 193]; I have slightly altered Burleigh’s example. I don’t know how one and the same argument can be valid at 4 p.m. and invalid at 5 p.m.).

Once this much is clear, we can decide whether we want to do anything about it. From the Traditional Logician’s point of view it might seem sensible to amend the Bolzano–Tarski definition. This is the direction which *free logic* has taken. Cf. Bencivenga, (Volume 7 of this *Handbook*).

The mainstream has gone the other way. Non-referring constants are anathema in most mathematics. Besides, Hilbert-style calculi with identity always have  $\exists x(x = x)$  as a provable formula. (See Remark 6 in Appendix A below. On the other hand semantic tableau systems which allow empty structures, such as Hodges [1977], are arguably a little simpler and more natural than versions which exclude them.) If  $\exists x\neg\perp$  is logically valid in some languages and not in others, the easiest remedy is to make it logically valid in all languages, and we can do that by *requiring all structures to have non-empty domains*. Henceforth we shall do so (after pausing to note that Schröder [1895, p. 5] required all structures to have at least two elements).

Let us review some properties of  $\models$ . Analogues of Theorems 1–4 (allowing infinitely many premises!) and Theorem 5 of Section 5 now hold. The relevant notion of logical equivalence is this: the formula  $\phi$  is *logically equivalent* to the formula  $\psi$  if for every structure  $\mathfrak{A}$  and every assignment  $g$  in  $\mathfrak{A}$  which is suitable for both  $\phi$  and  $\psi$ ,  $\mathfrak{A} \models \phi[g]$  if and only if  $\mathfrak{A} \models \psi[g]$ . For example

(126)  $\forall x\phi$  is logically equivalent to  $\neg\exists x\neg\phi$ ,  
 $\exists x\phi$  is logically equivalent to  $\neg\forall x\neg\phi$ .

A formula is said to be *basic* if it is either atomic or the negation of an atomic formula. A formula is in *disjunctive normal form* if it is either  $\perp$  or a disjunction of conjunctions of basic formulas. One can show:

(127) *Every formula of  $L$  is logically equivalent to a formula of  $L$  with the same free variables, in which all quantifiers are at the left-hand end, and the part after the quantifiers is in disjunctive normal form.*

A formula with its quantifiers all at the front is said to be in *prenex form*. (In Section 25 below we meet Skolem normal forms, which are different from (127) but also prenex.)

Proof calculi for propositional logic are generally quite easy to adapt to predicate logic. Sundholm (Volume 2 of this *Handbook*) surveys the possibilities. Usually in predicate logic one allows arbitrary formulas to occur in a proof, not just sentences, and this can make it a little tricky to say exactly what is the informal idea expressed by a proof. (This applies particularly to Hilbert-style calculi; cf. Remarks 4 and 5 in Appendix A below. Some calculi paper over the difficulty by writing the free variables as constants.) When one speaks of a formal calculus for predicate logic as being *sound* or *complete* (cf. Section 7 above), one always ignores formulas which have free variables.

Gentzen's natural deduction calculus can be adapted to predicate logic simply by adding four rules, namely introduction and elimination rules for  $\forall$  and  $\exists$ . The *introduction rule* for  $\exists$  says:

(128) From  $\phi[\tau/x]$  infer  $\exists x\phi$ .

(If the object  $\tau$  satisfies  $\phi$ , then at least one thing satisfies  $\phi$ .) The *elimination rule* for  $\exists$  says:

(129) Given a proof of  $\psi$  from  $\phi[y/x]$  and assumptions  $\chi_1, \dots, \chi_n$ , where  $y$  is not free in any of  $\exists x\phi, \psi, \chi_1, \dots, \chi_n$ , deduce  $\psi$  from  $\exists x\phi$  and  $\chi_1, \dots, \chi_n$ .

The justification of (129) is of some philosophical interest, as the following example will show. We want to deduce an absurdity from the assumption that there is a greatest integer. So we let  $y$  be a greatest integer, we get a contradiction  $y < y + 1 \leq y$ , whence  $\perp$ . Then by (129) we deduce  $\perp$  from  $\exists x$  ( $x$  is a greatest integer). Now the problem is: How can we possibly 'let  $y$  be a greatest integer', since there aren't any? Some logicians exhort us to '*imagine* that  $y$  is a greatest integer', but I always found that this one defeats my powers of imagination.

The Bolzano–Tarski definition of logical implication is a real help here, because it steers us away from matters of 'If it were the case that ...' towards questions about what actually is the case in structures which do

exist. We have to decide how natural deduction proofs are supposed to match the Bolzano–Tarski definition, bearing in mind that formulas with free variables may occur. The following interpretation is the right one: the existence of a natural deduction proof with conclusion  $\psi$  and premises  $\chi_1, \dots, \chi_n$  should tell us that for every structure  $\mathfrak{A}$  and every assignment  $g$  in  $\mathfrak{A}$  which is suitable for all of  $\psi, \chi_1, \dots, \chi_n$ , we have  $\mathfrak{A} \models (\chi_1 \wedge \dots \wedge \chi_n \rightarrow \psi)[g]$ . (This is *not* obvious— for Hilbert-style calculi one has to supply a quite different rationale, cf. Remark 5 on Hilbert-style calculi in Appendix A.)

Now we can justify (129). Let  $\mathfrak{A}$  be a structure and  $g$  an assignment in  $\mathfrak{A}$  which is suitable for  $\exists x\phi, \chi_1, \dots, \chi_n$  and  $\psi$ . We wish to show that:

$$(130) \quad \mathfrak{A} \models (\exists x\phi \wedge \chi_1 \wedge \dots \wedge \chi_n \rightarrow \psi)[g].$$

By the truth-definition in Section 14 we can assume that the domain of  $g$  is just the set of variables free in the formulas listed, so that in particular  $y$  is not in the domain of  $g$ . There are now two cases. The first is that  $\mathfrak{A} \models \neg(\exists x\phi \wedge \chi_1 \wedge \dots \wedge \chi_n)[g]$ . Then truth-tables show that (130) holds. The second case is that  $\mathfrak{A} \models (\exists x\phi \wedge \chi_1 \wedge \dots \wedge \chi_n)[g]$ , so there is an element  $\alpha$  of  $\mathfrak{A}$  such that  $\mathfrak{A} \models (\phi[y/x] \wedge \chi_1 \wedge \dots \wedge \chi_n \rightarrow \psi)[g, \alpha/y]$ , so  $\mathfrak{A} \models \psi[g, \alpha/y]$ . But then since  $y$  is not free in  $\psi$ ,  $\mathfrak{A} \models \psi[g]$ , which again implies (130).

I do not think this solves all the philosophical problems raised by (129). Wiredu [1973] seems relevant.

The references given for the proof calculi discussed in Section 7 remain relevant, except Lukasiewicz and Tarski [1930] which is only about propositional logic. The various theorems of Gentzen [1934], including the cut-elimination theorem, all apply to predicate logic. From the point of view of these calculi, the difference between propositional and predicate logic is relatively slight and has to do with checking that certain symbols don't occur in the wrong places in proofs.

Proof calculi for many-sorted languages are also not hard to come by. See [Schmidt, 1938; Wang, 1952; Feferman, 1968a].

Quantifiers did provoke one quite new proof-theoretic contrivance. In the 1920s a number of logicians (notably Skolem, Hilbert, Herbrand) regarded quantifiers as an intrusion of infinity into the finite-minded world of propositional logic, and they tried various ways of—so to say—deactivating quantifiers. Hilbert proposed the following: replace  $\exists x\phi$  everywhere by the sentence  $\phi[\varepsilon x\phi/x]$ , where ' $\varepsilon x\phi$ ' is interpreted as 'the element I choose among those that satisfy  $\phi$ '. The interpretation is of course outrageous, but Hilbert showed that his  $\varepsilon$ -calculus proved exactly the same sequents as more conventional calculi. See Hilbert and Bernays [1939] and Leisenring [1969].

It can easily be checked that any sequent which can be proved by the natural deduction calculus sketched above (cf. Sundholm's Chapter in a following volume of this *Handbook* for details) is correct. But nobody could claim to see, just by staring at it, that this calculus can prove *every* correct sequent of predicate logic. Nevertheless it can, as the next section will show.

## 16 CREATING MODELS

The natural deduction calculus for first-order logic is *complete* in the sense that if  $\Delta \vDash \psi$  then the calculus gives a proof of  $\psi$  from assumptions in  $\Delta$ . This result, or rather the same result for an equivalent Hilbert-style calculus, was first proved by Kurt Gödel in his doctoral dissertation [1930]. Strictly Thoralf Skolem had already proved it in his brilliant papers [1922; 1928; 1929], but he was blissfully unaware that he had done so. (See [Vaught, 1974; Wang, 1970]; Skolem's finitist philosophical leanings seem to have blinded him to some mathematical implications of his work.)

A theory  $\Delta$  in the language  $L$  is said to be *consistent* for a particular proof calculus if the calculus gives no proof of  $\perp$  from assumptions in  $\Delta$ . (Some writers say instead: 'gives no proof of a contradiction  $\phi \wedge \neg\phi$  from assumptions in  $\Delta$ '. For the calculi we are considering, this amounts to the same thing.) We shall demonstrate that *if  $\Delta$  is consistent for the natural deduction calculus then  $\Delta$  has a model*. This implies that the calculus is complete, as follows. Suppose  $\Delta \vDash \psi$ . Then  $\Delta, \psi \rightarrow \perp \vDash \perp$  (cf. Theorem 4 in Section 5), hence  $\Delta$  together with  $\psi \rightarrow \perp$  has no model. But then the theory consisting of  $\Delta$  together with  $\psi \rightarrow \perp$  is not consistent for the natural deduction calculus, so we have a proof of  $\perp$  from  $\psi \rightarrow \perp$  and sentences in  $\Delta$ . One can then quickly construct a proof of  $\psi$  from sentences in  $\Delta$  by the rule (69) for  $\perp$ .

So the main problem is to show that every consistent theory has a model. This involves constructing a model—but out of what? Spontaneous creation is not allowed in mathematics; the pieces must come from somewhere. Skolem [1922] and Gödel [1930] made their models out of natural numbers, using an informal induction to define the relations. A much more direct source of materials was noticed by Henkin [1949] and independently by Rasiowa and Sikorski [1950]: they constructed the model of  $\Delta$  out of the theory  $\Delta$  itself. (Their proof was closely related to Kronecker's [1882] method of constructing extension fields of a field  $K$  out of polynomials over  $K$ . Both he and they factored out a maximal ideal in a ring.)

Hintikka [1955] and Schütte [1956] extracted the essentials of the Henkin–Rasiowa–Sikorski proof in an elegant form, and what follows is based on their account. For simplicity we assume that the language  $L$  has infinitely many individual constants but its only truth-functors are  $\neg$  and  $\wedge$  and its only quantifier symbol is  $\exists$ . A theory  $\Delta$  in  $L$  is called a *Hintikka set* if it satisfies these seven conditions:

1.  $\perp$  is not in  $\Delta$ .
2. If  $\phi$  is an atomic formula in  $\Delta$  then  $\neg\phi$  is not in  $\Delta$ .
3. If  $\neg\neg\psi$  is in  $\Delta$  then  $\psi$  is in  $\Delta$ .
4. If  $\psi \wedge \chi$  is in  $\Delta$  then  $\psi$  and  $\chi$  are both in  $\Delta$ .



5. If  $\neg(\psi \wedge \chi)$  is in  $\Delta$  then either  $\neg\psi$  is in  $\Delta$  or  $\neg\chi$  is in  $\Delta$ .
6. If  $\exists x\psi$  is in  $\Delta$  then  $\psi[c/x]$  is in  $\Delta$  for some individual constant  $c$ .
7. If  $\neg\exists x\psi$  is in  $\Delta$  then  $\neg\psi[c/x]$  is in  $\Delta$  for each individual constant  $c$ .

We can construct an L-structure  $\mathfrak{A}$  out of a theory  $\Delta$  as follows. The elements of  $\mathfrak{A}$  are the individual constants of L. For each constant  $c$ ,  $I_{\mathfrak{A}}(c)$  is  $c$  itself. For each  $n$ -place predicate constant  $R$  of L the relation  $I_{\mathfrak{A}}(R)$  is defined to be the set of all ordered  $n$ -tuples  $\langle c_1, \dots, c_n \rangle$  such that the sentence  $R(c_1, \dots, c_n)$  is in  $\Delta$ .

Let  $\Delta$  be a Hintikka set. We claim that the structure  $\mathfrak{A}$  built out of  $\Delta$  is a model of  $\Delta$ . It suffices to show the following, by induction on the complexity of  $\phi$ : if  $\phi$  is in  $\Delta$  then  $\phi$  is true in  $\mathfrak{A}$ , and if  $\neg\phi$  is in  $\Delta$  then  $\neg\phi$  is true in  $\mathfrak{A}$ . I consider two sample cases. First let  $\phi$  be atomic. If  $\phi$  is in  $\Delta$  then the construction of  $\mathfrak{A}$  guarantees that  $\mathfrak{A} \models \phi$ . If  $\neg\phi$  is in  $\Delta$ , then by clause (2),  $\phi$  is not in  $\Delta$ ; so by the construction of  $\mathfrak{A}$  again,  $\mathfrak{A}$  is not a model of  $\phi$  and hence  $\mathfrak{A} \models \neg\phi$ . Next suppose  $\phi$  is  $\psi \wedge \chi$ . If  $\phi$  is in  $\Delta$ , then by clause (4), both  $\psi$  and  $\chi$  are in  $\Delta$ ; since they have lower complexities than  $\phi$ , we infer that  $\mathfrak{A} \models \psi$  and  $\mathfrak{A} \models \chi$ ; so again  $\mathfrak{A} \models \phi$ . If  $\neg\phi$  is in  $\Delta$  then by clause (5) either  $\neg\psi$  is in  $\Delta$  or  $\neg\chi$  is in  $\Delta$ ; suppose the former. Since  $\psi$  has lower complexity than  $\phi$ , we have  $\mathfrak{A} \models \neg\psi$ ; it follows again that  $\mathfrak{A} \models \neg\phi$ . The remaining cases are similar. So *every Hintikka set has a model*.

It remains to show that if  $\Delta$  is consistent, then by adding sentences to  $\Delta$  we can get a Hintikka set  $\Delta^+$ ;  $\Delta^+$  will then have a model, which must also be a model of  $\Delta$  because  $\Delta^+$  includes  $\Delta$ . The strategy is as follows.

**Step 1.** Extend the language L of  $T$  to a language  $L^+$  which has infinitely many new individual constants  $c_0, c_1, c_2, \dots$ . These new constants are known as the *witnesses* (because in (6) above they will serve as witnesses to the truth of  $\exists x\psi$ ).

**Step 2.** List all the sentences of  $L^+$  as  $\phi_0, \phi_1, \dots$  in an infinite list so that every sentence occurs infinitely often in the list. This can be done by some kind of zigzagging back and forth.

**Step 3.** At this very last step there is a parting of the ways. Three different arguments will lead us home. Let me describe them and then compare them.

The first argument we may call the *direct* argument: we simply add sentences to  $\Delta$  as required by (3)–(7), making sure as we do so that (1) and (2) are not violated. To spell out the details, we define by induction theories  $\Delta_0, \Delta_1, \dots$  in the language  $L^+$  so that (i) every theory  $\Delta_i$  is consistent; (ii) for all  $i$ ,  $\Delta_{i+1}$  includes  $\Delta_i$ ; (iii) for each  $i$ , only finitely many of the witnesses appear in the sentences in  $\Delta_i$ ; (iv)  $\Delta_0$  is  $\Delta$ ; and (v) for each  $i$ , if  $\phi_i$  is in  $\Delta_i$  then:

- 3' if  $\phi_i$  is of form  $\neg\neg\psi$  then  $\Delta_{i+1}$  is  $\Delta_i$  together with  $\psi$ ;
- 4' if  $\phi_i$  is of form  $\psi \wedge \chi$  then  $\Delta_{i+1}$  is  $\Delta_i$  together with  $\psi$  and  $\chi$ ;
- 5' if  $\phi_i$  is of form  $\neg(\psi \wedge \chi)$  then  $\Delta_{i+1}$  is  $\Delta_i$  together with at least one of  $\neg\psi, \neg\chi$ ;
- 6' if  $\phi_i$  is of form  $\exists x\psi$  then  $\Delta_{i+1}$  is  $\Delta_i$  together with  $\psi[c/x]$  for some witness  $c$  which doesn't occur in  $\Delta_i$ ;
- 7' if  $\phi_i$  is of form  $\neg\exists x\psi$  then  $\Delta_{i+1}$  is  $\Delta_i$  together with  $\neg\psi[c/x]$  for the first witness  $c$  such that  $\neg\psi[c/x]$  is not already in  $\Delta_i$ .

It has to be shown that theories  $\Delta_i$  exist meeting conditions (1)–(5). The proof is by induction. We satisfy (1)–(5) for  $\Delta_0$  by putting  $\Delta_0 = \Delta$  (and this is the point where we use the assumption that  $\Delta$  is consistent for natural deduction). Then we must show that if we have got as far as  $\Delta_i$  safely,  $\Delta_{i+1}$  can be constructed too. Conditions (2) and (3) are actually implied by the others and (4) is guaranteed from the beginning. So we merely need to show that

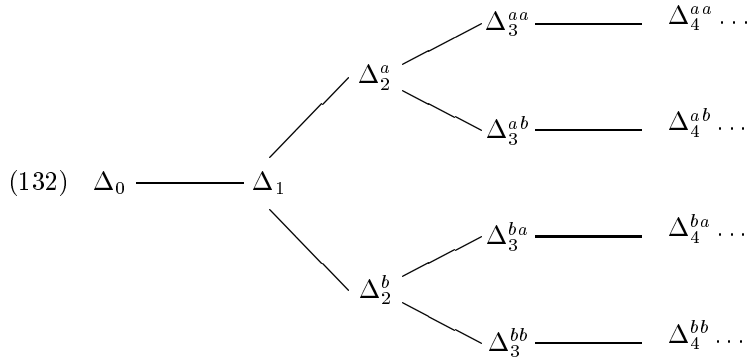
- (131) assuming  $\Delta_i$  is consistent,  $\Delta_{i+1}$  can be chosen so that it is consistent and satisfies the appropriate one of (3')–(7').

There are five cases to consider. Let me take the hardest, which is (6'). It is assumed that  $\phi_i$  is  $\exists x\psi$  and is in  $\Delta_i$ . By (3) so far, some witness has not yet been used; let  $c$  be the first such witness and let  $\Delta_{i+1}$  be  $\Delta_i$  together with  $\psi[c/x]$ . If by misfortune  $\Delta_{i+1}$  was inconsistent, then since  $c$  never occurs in  $\Delta_i$  or  $\phi_i$ , the elimination rule for  $\exists$  (section 15 or Sundholm, Volume 2 of this *Handbook*) shows that we can prove  $\perp$  already from  $\exists x\psi$  and assumptions in  $\Delta_i$ . But  $\exists x\psi$  was in  $\Delta_i$ , so we have a contradiction to our assumption that  $\Delta_i$  was consistent. Hence  $\Delta_{i+1}$  is consistent as required.

When the theories  $\Delta_i$  have been constructed, let  $\Delta^+$  be the set of all sentences which are in at least one theory  $\Delta_i$ . Since each  $\Delta_i$  was consistent,  $\Delta^+$  satisfies conditions (1) and (2) for a Hintikka set. The requirements (3')–(7'), and the fact that in the listing  $\phi_0, \phi_1, \dots$  we keep coming round to each sentence infinitely often, ensure that  $\Delta^+$  satisfies conditions (3)–(7) as well. So  $\Delta^+$  is a Hintikka set and has a model, which completes the construction of a model of  $\Delta$ .

The second argument we may call the *tree* argument. A hint of it is in [Skolem, 1929]. We imagine a man constructing the theories  $\Delta_i$  as in the direct argument above. When he faces clauses (3'), (4'), (6') or (7'), he knows at once how he should construct  $\Delta_{i+1}$  out of  $\Delta_i$ ; the hardest thing he has to do is to work out which is the first witness not yet used in  $\Delta_i$  in the case of clause (6'). But in (5') we can only prove for him that at least one of  $\neg\psi$  and  $\neg\chi$  can consistently be added to  $\Delta_i$ , so he must check for

himself whether  $\Delta_i$  together with  $\neg\psi$  is in fact consistent. Let us imagine that he is allergic to consistency calculations. Then the best he can do is to make *two alternative suggestions* for  $\Delta_{i+1}$ , viz.  $\Delta_i$  with  $\neg\psi$ , and  $\Delta_i$  with  $\neg\chi$ . Thus he will make not a chain of theories  $\Delta_0, \Delta_1, \dots$  but a branching tree of theories:



Now he no longer knows which of these theories are consistent. So he forgets about consistency and looks directly at conditions (1) and (2) in the definition of a Hintikka set. At least he can tell by inspection whether a theory violates these. So he prunes off the tree all theories which fail (1) or (2)—he can do this as he goes along. Some theories in the tree will become dead ends. But the argument we gave for the earlier direct approach shows that at every level in the tree there must be some theory which can be extended to the next level.

Now a combinatorial theorem known as *König's tree lemma* says that if a tree has a positive but finite number of items at the  $n$ th level, for every natural number  $n$ , then the tree has a branch which runs up through all these levels. So we know that (132) has an infinite branch. Let  $\Delta_0, \Delta_1, \Delta_2, \dots$  be such a branch and let  $\Delta^+$  be the set of all sentences which occur in at least one theory  $\Delta_i$  in the branch. The previous argument shows that  $\Delta^+$  satisfies (3)–(7), and we know that  $\Delta^+$  satisfies (1) and (2) because otherwise it would have been pruned off at some finite stage. So again  $\Delta^+$  is a Hintikka set.

The third argument is the *maximising* argument, sometimes known as the *Henkin-style* argument, though Skolem's argument in [1922] seems to be of this type. This argument is an opposite to the second kind of argument: instead of using (1)–(7) in the construction and forgetting consistency, we

exploit consistency and leave (1)–(7) on one side until the very end. We define by induction theories  $\Delta_0, \Delta_1, \dots$  in the language  $L^+$  so that (i) every theory  $\Delta_i$  is consistent; (ii) for all  $i$ ,  $\Delta_{i+1}$  includes  $\Delta_i$ ; (iii) for each  $i$ , only finitely many of the witnesses appear in the sentences in  $\Delta_i$ ; (iv)  $\Delta_0$  is  $\Delta$ ; and (v) for each  $i$ ,

- ( $\alpha$ ) if  $\Delta_i$  together with  $\phi_i$  is consistent then  $\Delta_{i+1}$  contains  $\phi_i$ ;
- ( $\beta$ ) if  $\phi_i$  is in  $\Delta_{i+1}$  and is of form  $\exists x\psi$ , then for some witness  $c$  which doesn't occur in  $\Delta_i$  or in  $\phi_i$ ,  $\psi[c/x]$  is in  $\Delta_{i+1}$ .

The argument to justify this construction is the same as for the direct argument, except that (3'), (4'), (5') and (7') are now irrelevant. As before, let  $\Delta^+$  be the set of sentences which occur in at least one theory  $\Delta_i$ . Clause ( $\alpha$ ) in the construction guarantees that

- (133) for every sentence  $\phi$  of  $L^+$ , if  $\Delta^+$  together with  $\phi$  is consistent, then  $\phi$  is in  $\Delta^+$ .

From (133) and properties of natural deduction we infer

- (134) for every sentence  $\phi$  of  $L^+$ , if  $\phi$  is provable from assumptions in  $\Delta^+$  then  $\phi$  is in  $\Delta^+$ .

Knowing (133) and (134), we can show that  $\Delta^+$  satisfies (3)–(7). For example, take (5) and suppose that  $\neg(\psi \wedge \chi)$  is in  $\Delta^+$  but  $\neg\psi$  is not in  $\Delta^+$ . Then by (133) there is a proof of  $\perp$  from  $\Delta^+$  and  $\neg\psi$ . Using the natural deduction rules we can adapt this proof to get a proof of  $\neg\chi$  from  $\Delta^+$ , and it follows by (134) that  $\neg\chi$  is in  $\Delta^+$ . Since the  $\Delta_i$  are all consistent,  $\Delta^+$  also satisfies (1) and (2). So once again  $\Delta^+$  is a Hintikka set.

Some authors take care of clause ( $\beta$ ) before the main construction. They can do it by adding to  $\Delta$  a collection of sentences of the form  $\exists x\psi \rightarrow \psi[c/x]$ . The argument which justified (6') will justify this too.

The first and third arguments above are very closely related. I gave both of them in the form that would serve for a countable language, but they adapt to first-order languages of any cardinality. The merit of the maximising argument is that the construction is easy to describe. (For example, the listing  $\phi_0, \phi_1, \dots$  need not repeat any formulas.)

The first and second arguments have one advantage over the third. Suppose  $\Delta$  is a finite set of prenex sentences of form  $\exists \vec{x}\forall \vec{y}\psi$ , with no quantifiers in  $\psi$ . Then these two arguments find  $\Delta^+$  after only a finite number of steps in the construction. So  $\Delta^+$  is finite and has a finite model, and it follows that we can compute whether or not a sentence of this form has a model. (This is no longer true if function symbols are added to the language as in Section 18 below.) The decidability of propositional logic is a special case of this. So also are various theorems about finite models for modal logics.

When  $\Delta_0$  is finite, closer inspection of the trees (132) shows that they are just the natural extension to predicate logic of the semantic tableaux of propositional logic. If  $\Delta_0$  has no models then every branch comes to a dead end after a finite number of steps. If  $\Delta_0$  has a model, then the tree has a branch which never closes, and we can read this branch as a description of a model. So the tree argument has given us a complete proof calculus for predicate logic. (Cf. Beth [1955; 1962], Jeffrey [1967], Smullyan [1968], Bell and Machover [1977] for predicate logic semantic tableaux.) Incidentally it is most unpleasant to prove the completeness of semantic tableaux by the direct or maximising arguments. One needs facts of the form: if  $\Delta \vdash \psi$  and  $\Delta, \psi \vdash \chi$  then  $\Delta \vdash \chi$ . To prove these is to prove Gentzen's cut-elimination theorem.

Notice that even when  $\Delta_0$  is finite, semantic tableaux no longer provide a method for deciding whether  $\Delta_0$  has a model. If it does have a model, the tree may simply go on branching forever, and we may never know whether it is going to close off in the next minute or the next century. In Section 24 below we prove a theorem of Church [1936] which says that there is not and cannot be any mechanical method for deciding which sentences of predicate logic have models.

## 17 CONSEQUENCES OF THE CONSTRUCTION OF MODELS

Many of the most important consequences of the construction in the previous section are got by making some changes in the details. For example, instead of using the individual constants of the language as elements, we can number these constants as  $b_0, b_1, \dots$ , and use the number  $n$  in place of the constant  $b_n$ . Since numbers can be thought of as pure sets ([Mendelson, 1987, pp. 187 ff.] or Appendix C below), the structure which emerges at the end will be a pure set structure. Hence, for any standard proof calculus for a language  $L$  of predicate logic:

**THEOREM 10.** *Suppose  $T$  is a theory and  $\psi$  a sentence of  $L$ , such that the calculus doesn't prove  $\psi$  from  $T$ . Then there is a pure set structure which is a model of  $T$  and not of  $\psi$ .*

In terms of the discussion in Section 8 above, this shows that the Proof Theorist's notion of logical implication agrees with the Model Theorist's, whether or not the Model Theorist restricts himself to pure set structures.

We can take matters one step further by encoding all symbols and formulas of  $L$  as natural numbers. So a theory in  $L$  will be a set of numbers. Suppose the theory  $T$  is in fact the set of all numbers which satisfy the first-order formula  $\phi$  in the language of arithmetic; then by analysing the proof of Theorem 10 we can find another first-order formula  $\chi$  in the language of arithmetic, which defines a structure with natural numbers as its elements,

so that:

**THEOREM 11.** *In first-order Peano arithmetic we can prove that if some standard proof calculus doesn't prove  $T$  is inconsistent, then the structure defined by  $\chi$  is a model of  $T$ .*

(Cf. [Kleene, 1952, p. 394] and [Hasenjaeger, 1953] for a sharper result.)

Theorem 11 is philosophically very interesting. Suppose  $T$  is a finite theory, and proof-theoretically  $T$  doesn't imply  $\psi$ . Applying Theorem 11 to the theory  $T \cup \{\neg\psi\}$ , we get a formula  $\chi$  which defines a natural number structure  $\mathfrak{A}$  in which  $T$  is true and  $\psi$  is false. By means of  $\chi$ , the formulas of  $T$  and  $\psi$  can be read as meaningful statements about  $\mathfrak{A}$  and hence about the natural numbers. The statements in  $T$  are true but  $\psi$  is false, so we have found an invalid argument of the form ' $T$ . Therefore  $\psi$ '. It follows that if a first-order sequent is correct by the Traditional Logician's definition, then it is correct by the Proof Theorist's too. Since the converse is straightforward to prove, we have a demonstration that *the Traditional Logician's notion of validity exactly coincides with the Proof Theorist's*. The proof of this result uses nothing stronger than the assumption that the axioms of first-order Peano arithmetic have a model.

The Traditional Logician's notion of logical implication is quite informal—on any version it involves the imprecise notion of a 'valid English argument'. Nevertheless we have now proved that it agrees exactly with the mathematically precise notion of logical implication given by the Proof Theorist. (Cf. [Kreisel, 1967].) People are apt to say that it is impossible to prove that an informal notion and a formal one agree exactly. Since we have just done the impossible, maybe I should add a comment. Although the notion of a valid argument is vague, there is no doubt that (i) if there is a formal proof of a sequent, then any argument with the form of that sequent must be valid, and (ii) if there is an explicitly definable counterexample to the sequent, then there is an invalid argument of that form. We have shown, by strict mathematics, that every finite sequent has either a formal proof or an explicitly definable counterexample. So we have trapped the informal notion between two formal ones. Contrast *Church's thesis*, that the effectively computable functions (informal notion) are exactly the recursive ones (formal). There is no doubt that the existence of a recursive definition for a function makes the function effectively computable. But nobody has yet thought of any kind of mathematical object whose existence undeniably implies that a function is *not* effectively computable. So Church's thesis remains unproved. (Van Dalen's chapter in this Volume discusses Church's thesis.)

I return to the completeness proof. By coding all expressions of  $L$  into numbers or sets, we made it completely irrelevant that the symbols of  $L$  can be written on a page, or even that there are at most countably many of them. *So let us now allow arbitrary sets to serve instead of symbols.* Languages

of this abstract type can be called *set languages*. They are in common use today even among proof theorists. Of course to use these languages we have to rely either on our intuitions about sets or on proofs in axiomatic set theory; there is no question of checking by inspection. Henkin's [1949] completeness proof was given in this setting. In fact he proved:

**THEOREM 12.** *If  $L$  is a first-order set language and  $T$  a theory in  $L$  whose cardinality is at most the infinite cardinal  $\kappa$ , then either a finite part of  $T$  can be proved inconsistent by a proof calculus, or  $T$  has a model with at most  $\kappa$  elements.*

Theorem 12 has several important mathematical consequences. For example, the *Compactness Theorem* says:

**THEOREM 13.** *Let  $T$  be a first-order theory (in a set language). If every finite set of sentences in  $T$  has a model, then  $T$  has a model.*

Theorem 13 for countable languages was proved by Gödel in [1930]. For propositional logic with arbitrarily many symbols it was proved by Gödel [1931a], in answer to a question of Menger. The first proof of Theorem 13 was sketched rather inadequately by Anatolii Mal'tsev in [1936] (see the review of [Mal'tsev, 1941] by Henkin and Mostowski [1959]). But in [1941] Mal'tsev showed that Theorem 13 has interesting and far from trivial consequences in group theory, thus beginning one of the most important lines of application of first-order logic in mathematics.

The last consequence I shall draw from Theorem 12 is not really interesting until identity is added to the language (see the next section); but this is a convenient place to state it. It is the *Upward and Downward Löwenheim-Skolem Theorem*:

**THEOREM 14.** *Let  $T$  be a first-order theory in a language with  $\lambda$  formulas, and  $\kappa$  an infinite cardinal at least as great as  $\lambda$ . If  $T$  has a model with infinitely many elements then  $T$  has one with exactly  $\kappa$  elements.*

Theorem 13 was proved in successively stronger versions by Löwenheim [1915], Skolem [1920; 1922], Tarski in unpublished lectures in 1928, Mal'tsev [1936] and Tarski and Vaught [1956]; see [Vaught, 1974] for a thorough history of this and Theorems 12 and 13. The texts of Bell and Slomson [1969], Chang and Keisler [1973] and Hodges [1993a] develop these theorems, and Sacks [1972] and Cherlin [1976] study some of their applications in algebra. Skolem [1955] expressly dissociated himself from the Upward version of Theorem 14, which he regarded as nonsense.

## 18 IDENTITY

The symbol '=' is reserved for use as a 2-place predicate symbol with the intended meaning

(135)  $a = b$  iff  $a$  and  $b$  are one and the same thing.

When  $\mathfrak{A}$  is a structure for a language containing ‘=’, we say that  $\mathfrak{A}$  has *standard identity* if the relation  $I_{\mathfrak{A}}(=)$  holds between elements  $\alpha$  and  $\beta$  of  $\mathfrak{A}$  precisely when  $\alpha$  and  $\beta$  are the same element.

‘ $x = y$ ’ is read as ‘ $x$  equals  $y$ ’, rather misleadingly—all men may be created equal but they are not created one and the same man. Another reading is ‘ $x$  is identical with  $y$ ’. As far as English usage goes, this is not much improvement on ‘equals’: there are two identical birds feeding outside my window, but they aren’t the same bird (and think of identical twins). Be that as it may, ‘=’ is called the *identity* sign and the relation it expresses in (135) is called *identity*.

Let  $L$  be a language containing the symbol ‘=’. It would be pleasant if we could find a theory  $\Delta$  in  $L$  whose models are exactly the  $L$ -structures with standard identity. Alas, there is no such theory. *For every  $L$ -structure  $\mathfrak{A}$  with standard identity there is an  $L$ -structure  $\mathfrak{B}$  which is a model of the same sentences of  $L$  as  $\mathfrak{A}$  but doesn’t have standard identity.* Let us prove this.

Take an  $L$ -structure  $\mathfrak{A}$  with standard identity and let  $\delta_1, \dots, \delta_{2,000,000}$  be two million objects which are not in the domain of  $\mathfrak{A}$ . Let  $\beta$  be an element of  $\mathfrak{A}$ . We construct the  $L$ -structure  $\mathfrak{B}$  thus. The elements of  $\mathfrak{B}$  are those of  $\mathfrak{A}$  together with  $\delta_1, \dots, \delta_{2,000,000}$ . For each individual constant  $c$  we put  $I_{\mathfrak{B}}(c) = I_{\mathfrak{A}}(c)$ . For each element  $\alpha$  of  $\mathfrak{B}$  we define an element  $\hat{\alpha}$  of  $\mathfrak{A}$  as follows: if  $\alpha$  is in the domain of  $\mathfrak{A}$  then  $\hat{\alpha}$  is  $\alpha$ , and if  $\alpha$  is one of the  $\delta_j$ ’s then  $\hat{\alpha}$  is  $\beta$ . For every  $n$ -place predicate constant  $R$  we choose  $I_{\mathfrak{B}}(R)$  so that if  $\langle \alpha_1, \dots, \alpha_n \rangle$  is any  $n$ -tuple of elements of  $\mathfrak{B}$ , then:

(136)  $\langle \alpha_1, \dots, \alpha_n \rangle$  is in  $I_{\mathfrak{B}}(R)$  iff  $\langle \hat{\alpha}_1, \dots, \hat{\alpha}_n \rangle$  is in  $I_{\mathfrak{A}}(R)$ .

This defines  $\mathfrak{B}$ . By induction on the complexity of  $\phi$  we can prove that for every formula  $\phi(x_1, \dots, x_n)$  of  $L$  and every  $n$ -tuple  $\langle \alpha_1, \dots, \alpha_n \rangle$  of elements of  $\mathfrak{B}$ ,

(137)  $\mathfrak{B} \models \phi[\alpha_1/x_1, \dots, \alpha_n/x_n]$  iff  $\mathfrak{A} \models \phi[\hat{\alpha}_1/x_1, \dots, \hat{\alpha}_n/x_n]$ .

In particular  $\mathfrak{A}$  and  $\mathfrak{B}$  are models of exactly the same sentences of  $L$ . Since  $\mathfrak{A}$  has standard identity,  $\mathfrak{A} \models (x = x)[\beta/x]$ . Then from (136) it follows that the relation  $I_{\mathfrak{B}}(=)$  holds between any two of the elements  $\delta_1, \dots, \delta_{2,000,000}$ , and so  $I_{\mathfrak{B}}(=)$  is vastly different from standard identity.

So we look for a second best. Is there a theory  $\Delta$  which is true in all  $L$ -structures with standard identity, and which logically implies every sentence of  $L$  that is true in all such  $L$ -structures? This time the answer is positive. The following theory will do the job:

(138)  $\forall x x = x$ .



(139) All sentences of the form  $\forall zxy(x = y \rightarrow (\phi \rightarrow \phi[y/x]))$ .

Formula (138) is known as the *law of reflexivity of identity*. (139) is not a single sentence but an infinite family of sentences, namely all those which can be got by putting any formula  $\phi$  of  $L$  into the expression in (139);  $z$  are all the free variables of  $\phi$  except for  $x$  and  $y$ . These sentences (139) are collectively known as *Leibniz' Law*. They are the nearest we can get within  $L$  to saying that if  $a = b$  then anything which is true of  $a$  is true of  $b$  too.

By inspection it is clear that every  $L$ -structure with standard identity is a model of (138) and (139). To show that (138) and (139) logically imply every sentence true in all structures with standard identity, let me prove something stronger, namely: *For every  $L$ -structure  $\mathfrak{B}$  which is a model of (138) and (139) there is an  $L$ -structure  $\mathfrak{A}$  which is a model of exactly the same sentences of  $L$  as  $\mathfrak{B}$  and has standard identity*. Supposing this has been proved, let  $\Delta$  be the theory consisting of (138) and (139), and let  $\psi$  be a sentence of  $L$  which is not logically implied by  $\Delta$ . Then some  $L$ -structure  $\mathfrak{B}$  is a model of  $\Delta$  and  $\neg\psi$ ; so some structure  $\mathfrak{A}$  with standard identity is also a model of  $\neg\psi$ . It follows that  $\psi$  is not true in all structures with standard identity.

To prove what I undertook to prove, let  $\mathfrak{B}$  be a model of  $\Delta$ . Then we can show that the following hold, where we write  $=_{\mathfrak{B}}$  for  $I_{\mathfrak{B}}(=)$ :

(140) the relation  $I_{\mathfrak{B}}(=)$  is an equivalence relation;

(141) for every  $n$ -place predicate constant  $R$  of  $L$ , if  $\alpha_1 =_{\mathfrak{B}} \beta_1, \dots, \alpha_n =_{\mathfrak{B}} \beta_n$  and  $\langle \alpha_1, \dots, \alpha_n \rangle$  is in  $I_{\mathfrak{B}}(R)$  then  $\langle \beta_1, \dots, \beta_n \rangle$  is in  $I_{\mathfrak{B}}(R)$ .

Statement (141) can be proved by applying Leibniz' Law  $n$  times. Then (140) follows from (141) and reflexivity of identity, taking '=' for  $R$ . Statements (140) and (141) together are summarised by saying that the relation  $=_{\mathfrak{B}}$  is a *congruence* for  $L$ . For each element  $\alpha$  of  $\mathfrak{B}$ , we write  $\alpha^=$  for the equivalence class of  $\alpha$  under the relation  $=_{\mathfrak{B}}$ .

Now we define the  $L$ -structure  $\mathfrak{A}$  as follows. The domain of  $\mathfrak{A}$  is the class of all equivalence classes  $\alpha^=$  of elements  $\alpha$  of  $\mathfrak{B}$ . For each individual constant  $c$  we define  $I_{\mathfrak{A}}(c)$  to be  $I_{\mathfrak{B}}(c)^=$ . For each  $n$ -place predicate symbol  $R$  of  $L$  we define  $I_{\mathfrak{A}}(R)$  by:

(142)  $\langle \alpha_1^=, \dots, \alpha_n^= \rangle$  is in  $I_{\mathfrak{A}}(R)$  iff  $\langle \alpha_1, \dots, \alpha_n \rangle$  is in  $I_{\mathfrak{B}}(R)$ .

Definition (142) presupposes that the right-hand side of (142) is true or false depending only on the equivalence classes of  $\alpha_1, \dots, \alpha_n$ ; but (141) assured this.

In particular,  $\alpha^= =_{\mathfrak{A}} \beta^=$  if and only if  $\alpha =_{\mathfrak{B}} \beta$ , in other words, if and only if  $\alpha^=$  equals  $\beta^=$ . Hence,  $\mathfrak{A}$  has standard identity. It remains only to show that for every formula  $\phi(x_1, \dots, x_n)$  of  $L$  and all elements  $\alpha_1, \dots, \alpha_n$  of  $\mathfrak{B}$ ,

$$(143) \quad \mathfrak{A} \models \phi[\alpha_1^-/x_1, \dots, \alpha_n^-/x_n] \text{ iff } \mathfrak{B} \models \phi[\alpha_1/x_1, \dots, \alpha_n/x_n].$$

Statement (143) is proved by induction on the complexity of  $\phi$ .

Most logicians include ‘=’ as part of the vocabulary of every language for predicate logic, and interpret it always to mean standard identity. Since it is in every language, it is usually not mentioned in the similarity type. The proof calculi have to be extended to accommodate ‘=’. One way to extend the natural deduction calculus is to add two new rules:

$$(144) \quad \frac{}{x = x} \quad \frac{x = y \quad \phi}{\phi[y/x]}$$

The first rule deduces  $x = x$  from no premises.

Identity is needed for virtually all mathematical applications of logic. It also makes it possible to express in formulas the meanings of various English phrases such as ‘the’, ‘only’, ‘at least one’, ‘at most eight’, etc. (see e.g. Section 21 below).

Many mathematical applications of logic need symbols of another kind, called *function symbols*. The definitions given above can be stretched to allow function symbols as follows. Symbols  $f, g, h$  etc., with or without subscripts, are called *function constants*. A similarity type may contain function constants, each of which is labelled as an *n-place constant* for some positive integer  $n$ . If the language  $L$  has an  $n$ -place function constant  $f$  and  $\mathfrak{A}$  is an  $L$ -structure, then  $f$  is interpreted by  $\mathfrak{A}$  as an *n-place function*  $I_{\mathfrak{A}}(f)$  which assigns one element of  $\mathfrak{A}$  to each ordered  $n$ -tuple of elements of  $\mathfrak{A}$ . For example the 2-place function constant ‘+’ may be interpreted as a function which assigns 5 to  $\langle 2, 3 \rangle$ , 18 to  $\langle 9, 9 \rangle$  and so forth—though of course it can also be interpreted as some quite different function.

There are various ways of writing functions, such as

$$(145) \quad \sin x, \sqrt{x}, x^2, \hat{x}, y^y, x + y, \langle x, y \rangle.$$

But the general style is ‘ $f(x_1, \dots, x_n)$ ’, and logicians’ notation tends to follow this style. The details of syntax and proof theory with function symbols are rather messy, so I omit them and refer the reader to [Hilbert and Bernays, 1934] for details.

One rarely needs function symbols outside mathematical contexts. In any case, provided we have ‘=’ in our language, everything that can be said with function symbols can also be said without them. Briefly, the idea is to use a predicate constant  $R$  in such a way that ‘ $R(x_1, \dots, x_{n+1})$ ’ means ‘ $f(x_1, \dots, x_n) = x_{n+1}$ ’. When the function symbol  $f$  is in the language, it is true in all structures—and hence logically valid—that for all  $x_1, \dots, x_n$  there is a unique  $x_{n+1}$  such that  $f(x_1, \dots, x_n) = x_{n+1}$ . Translating  $f$  into  $R$ , this becomes

$$(146) \quad \forall x_1 \cdots x_n z t \exists y ((R(x_1, \dots, x_n, z) \wedge R(x_1, \dots, x_n, t) \rightarrow z = t) \wedge R(x_1, \dots, x_n, y)).$$

Since (146) is not logically valid, it may have to be assumed as an extra premise when we translate arguments involving  $f$  into arguments involving  $R$ .

## 19 AXIOMS AS DEFINITIONS

Axioms are, roughly speaking, the statements which one writes down at the beginning of a book in order to define the subject-matter of the book and provide a basis for deductions made in the book. For example any textbook of group theory will start by telling you that a group is a triple  $\langle G, *, e \rangle$  where  $*$  is a binary operation in the set  $G$  and  $e$  is an element of  $G$  such that

(147)  $*$  is associative, i.e. for all  $x, y$  and  $z$ ,  $x * (y * z) = (x * y) * z$ ,

(148)  $e$  is an identity, i.e. for all  $x$ ,  $x * e = e * x = x$ ,

(149) every element  $x$  has an inverse, i.e. an element  $y$  such that  $x * y = y * x = e$ .

Statements (147)–(149) are known as the *axioms for groups*. I could have chosen examples from physics, economics or even ethics.

It is often said that in an ‘axiomatic theory’ such as group theory, the axioms are ‘assumed’ and the remaining results are ‘deduced from the axioms’. This is completely wrong. W. R. Scott’s textbook *Group Theory* [1964] contains 457 pages of facts about groups, and the last fact which can by any stretch of the imagination be described as being ‘deduced from (147)–(149)’ occurs on page 8. We could indeed rewrite Scott’s book as a set of deductions from assumed axioms, but the axioms would be those of set theory, not (147)–(149). These three group axioms would appear, not as assumptions but as *part of the definition of ‘group’*.

The definition of a group can be paraphrased as follows. First we can recast the triple  $\langle G, *, e \rangle$  as an L-structure  $\mathfrak{G} = \langle G, I_{\mathfrak{G}} \rangle$  in a first-order language L with one 2-place function symbol  $*$  and one individual constant  $e$ . Then  $\mathfrak{G}$  is a group if and only if  $\mathfrak{G}$  is a model of the following three sentences:

(150)  $\forall xyz \ x * (y * z) = (x * y) * z$ ,

(151)  $\forall x(x * e = x \wedge e * x = x)$ ,

(152)  $\forall x \exists y(x * y = e \wedge y * x = e)$ .

Generalising this, let  $\Delta$  be any theory in a first-order language L. Let  $\mathbf{K}$  be a class of L-structures. Then  $\Delta$  is said to *axiomatise*  $\mathbf{K}$ , and  $\mathbf{K}$  is

called  $Mod(\Delta)$ , if  $\mathbf{K}$  is the class of all L-structures which are models of  $\Delta$ . The sentences in  $\Delta$  are called *axioms* for  $\mathbf{K}$ . Classes of form  $Mod(\{\phi\})$  for a single first-order sentence  $\phi$  are said to be *first-order definable*. Classes of form  $Mod(\Delta)$  for a first-order theory  $\Delta$  are said to be *generalised first-order definable*. The class of groups is first-order definable—we can use the conjunction of the three sentences (150)–(152).

Many other classes of structure which appear in pure or applied mathematics are (generalised) first-order definable. To give examples I need only list the axioms. First, *equivalence relations*:

$$(153) \quad \forall x R(x, x) \quad \text{'R is reflexive'}$$

$$(154) \quad \forall xy (R(x, y) \rightarrow R(y, x)) \quad \text{'R is symmetric'}$$

$$(155) \quad \forall xyz (R(x, y) \wedge R(y, z) \rightarrow R(x, z)) \quad \text{'R is transitive'}$$

Next, *partial orderings*:

$$(156) \quad \forall x x \leq x \quad \text{' $\leq$  is reflexive'}$$

$$(157) \quad \forall xyz (x \leq y \wedge y \leq z \rightarrow x \leq z) \quad \text{' $\leq$  is transitive'}$$

$$(158) \quad \forall xy (x \leq y \wedge y \leq x \rightarrow x = y) \quad \text{' $\leq$  is antisymmetric'}$$

Then *total* or *linear orderings* are axiomatised by (157) and (158) and

$$(159) \quad \forall xy (x \leq y \vee y \leq x) \quad \text{' $\leq$  is connected'}$$

Total orderings can also be axiomatised as follows, using  $<$  instead of  $\leq$ :

$$(160) \quad \forall xyz (x < y \wedge y < z \rightarrow x < z)$$

$$(161) \quad \forall x \neg x < x$$

$$(162) \quad \forall xy (x < y \vee y < x \vee x = y).$$

A total ordering in the second style can be converted into a total ordering in the first style by reading  $x \leq y$  as meaning  $x < y \vee x = y$ . There is a similar conversion from the first style to the second. We can express various conditions on linear orderings by adding further axioms to (157)–(159):

$$(163) \quad \exists x \forall y y \leq x \quad \text{'there is a last element'}$$

$$(164) \quad \forall x \exists y (\neg x = y \wedge \forall z (x \leq z \leftrightarrow x = z \vee y \leq z)) \quad \text{'every element has an immediate successor'}$$

Algebra is particularly rich in first-order or generalised first-order definable classes, for example rings, fields, lattices, categories, toposes, algebraically closed fields, vector spaces over a given field. *Commutative groups* are axiomatised by adding to (150)–(152) the axiom

(165)  $\forall xy \ x * y = y * x$ .

All the examples mentioned so far are first-order definable except for algebraically closed fields and vector spaces over an infinite field, which need infinitely many sentences to define them.

The notion of first-order definable classes was first made explicit in a paper of Tarski [1954]. If we know that a class of structures is generalised first-order definable then we immediately know various other facts about it, for example that it is closed under taking ultraproducts (cf. [Chang and Keisler, 1973] or [Bell and Slomson, 1969]—they are defined in Appendix C below) and that implicit definitions in the class can all be made explicit ('Beth's theorem'—Theorem 33 in Section 27 below). On the other hand, if one is not interested in model-theoretic facts like these, the informal style of (147)–(149) makes just as good a definition of a class as any set of first-order formulas. (In the philosophy of science, structuralists have given reasons for preferring the informal set-theoretic style; see [Sneed, 1971] and [Stegmüller, 1976].)

It was Hilbert and his school who first exploited axioms, higher-order as well as first-order, as a means of defining classes of structures. Hilbert was horrifically inaccurate in describing what he was doing. When he set up geometric axioms, he said that they defined what was meant by a point. Frege then caustically asked how he could use this definition to determine whether his pocket watch was a point ([Frege and Hilbert, 1899–1900]). Hilbert had simply confused defining a class of structures with defining the component relations and elements of a single structure. (Cf. the comments of [Bernays, 1942].) In this matter Hilbert was a spokesman for a confusion which many people shared. Even today one meets hopeful souls who believe that the axioms of set theory define what is meant by 'set'.

Hilbert added the lunatic remark that 'If . . . arbitrarily posited axioms together with all their consequences do not contradict one another, then they are true and the things defined by these axioms exist' [Frege and Hilbert, 1899–1900]. For example, one infers, if the axioms which say there is a measurable cardinal are consistent, then there is a measurable cardinal. If the axioms which say there is no measurable cardinal are consistent, then there is no measurable cardinal. If both sets of axioms are consistent . . . . In later years he was more cautious. In fairness to Hilbert, one should set his remark against the background beliefs of his time, one of which was the now happily discredited theory of 'implicit definition' (nothing to do with Beth's theorem of that name). See [Coffa, 1991], who puts the Frege-Hilbert debate helpfully into a broad historical context. Be that as it may, readers of Hilbert's philosophical remarks should always bear in mind his slogan '*Wir sind Mathematiker*' [Hilbert, 1926].

## 20 AXIOMS WITH INTENDED MODELS

Axioms are not always intended to define a class of structures as in Section 19 above. Often they are written down *in order to set on record certain facts about a particular structure*. The structure in question is then called the *intended interpretation* or *standard model* of the axioms. The best known example is probably the axioms of Peano arithmetic, which were set down by Dedekind [1888; 1967] as a statement of the ‘fundamental properties’ of the natural number sequence (the first-order formalisation is due to Gödel [1931b], cf. Appendix B below). Euclid’s axioms and postulates of geometry are another example, since he undoubtedly had space in mind as the intended interpretation.

The object in both Dedekind’s case and Euclid’s was to write down some elementary facts about the standard model so that further information could be got by making deductions from these facts. With this aim it becomes very natural to write the axioms in a first-order language, because we understand first-order deducibility well and so we shall know exactly what we are entitled to deduce from the axioms.

However, there is no hope at all of *defining* the natural numbers, even up to isomorphism, by means of any first-order axioms. Let me sketch a proof of this—it will be useful later. Suppose  $L$  is the first-order language of arithmetic, with symbols to represent plus and times, a 2-place predicate constant  $<$  (‘less than’), and a name  $n^*$  for each natural number  $n$ . Let  $L^+$  be  $L$  with a new individual constant  $c$  added. Let  $\Delta$  be the set of all sentences of  $L$  which are true in the standard model. Let  $\Delta^+$  be  $\Delta$  together with the sentences

$$(166) \quad 0^* < c, \quad 1^* < c, \quad 2^* < c, \dots$$

Now if  $\Gamma$  is any finite set of sentences from  $\Delta^+$  then  $\Gamma$  has a model: take the standard model of  $\Delta$  and let  $c$  stand for some natural number which is greater than every number mentioned in  $\Gamma$ . So by the Compactness Theorem (Theorem 13 in Section 17 above),  $\Delta^+$  has a model  $\mathfrak{A}$ . Since  $\Delta^+$  includes  $\Delta$ ,  $\mathfrak{A}$  is a model of  $\Delta$  and hence is a model of exactly the same sentences of  $L$  as the standard model. But  $\mathfrak{A}$  also has an element  $I_{\mathfrak{A}}(c)$  which by (166) is ‘greater than’  $I_{\mathfrak{A}}(0^*)$ ,  $I_{\mathfrak{A}}(1^*)$ ,  $I_{\mathfrak{A}}(2^*)$  and all the ‘natural numbers’ of  $\mathfrak{A}$ . So  $\mathfrak{A}$  is a model of  $\Delta$  with an ‘infinite element’. Such models of  $\Delta$  are called *non-standard models of arithmetic*. They were first constructed by Skolem [1934], and today people hold conferences on them.

But one can reasonably ask whether, say, the first-order Peano axioms (cf. Appendix B) imply all first-order sentences which are true in the standard model. This is equivalent to asking whether the axioms are a *complete* theory in the sense that if  $\phi$  is any sentence of their language, then either  $\phi$  or  $\neg\phi$  is a consequence of the axioms. Gödel’s epoch-making paper [1931b]

showed that the first-order Peano axioms are not complete; in fact no mechanically describable theory in this language is both complete and true in the standard model. In Section 24 below I shall sketch a proof of this.

There is a halfway house between the use of axioms to define a class and their use to say things about a standard model. Often we want to work with a class  $\mathbf{K}$  of L-structures which may not be generalised first-order definable. In such cases we say that a theory  $\Delta$  is a *set of axioms* for  $\mathbf{K}$  if every structure in  $\mathbf{K}$  is a model of  $\Delta$ ; we call it a *complete* set of axioms for  $\mathbf{K}$  if moreover every sentence of L which is true in all structures in  $\mathbf{K}$  is a logical consequence of  $\Delta$ .

Let me give three examples. (i) For the first, paraphrasing Carnap [1956, p. 222 ff] I consider the class of all structures which represent possible worlds, with domain the set of all people, ' $Bx$ ' interpreted as ' $x$  is a bachelor' and ' $Mx$ ' as ' $x$  is married'. Obviously this class is not generalised first-order definable. But the following sentence is a complete set of axioms:

$$(167) \quad \forall x(Bx \rightarrow \neg Mx).$$

In Carnap's terminology, when  $\mathbf{K}$  is the class of all structures in which certain symbols have certain fixed meanings, axioms for  $\mathbf{K}$  are called *meaning postulates*. (Lakoff [1972] discusses some trade-offs between meaning postulates and deep structure analysis in linguistics.)

(ii) For a second sample, consider second-order logic (cf. [Chapter 4, below]). In this logic we are able to say 'for all subsets  $P$  of the domain, ...', using second-order quantifiers ' $\forall P$ '. For reasons explained in Chapter 4 below, there is no hope of constructing a complete proof calculus for second-order logic. But we do have some incomplete calculi which are good for most practical purposes. They prove, among other things, the formula

$$(168) \quad \forall PQ(\forall z(P(z) \leftrightarrow Q(z)) \rightarrow P = Q)$$

which is the second-order logician's version of the axiom of extensionality.

Second-order logic can be translated wholesale into a kind of two-sorted first-order logic by the following device. Let L be any (first-order) language. Form a two-sorted language  $L^\downarrow$  with the same predicate and individual constants as L, together with one new 2-place predicate constant  $\varepsilon$ . For each L-structure  $\mathfrak{A}$ , form the  $L^\downarrow$ -structure  $\mathfrak{A}^\downarrow$  as follows. The domain of  $\mathfrak{A}^\downarrow$  is  $|\mathfrak{A}| \cup \mathcal{P}|\mathfrak{A}|$ ,  $|\mathfrak{A}|$  is the domain for the first sort and  $\mathcal{P}|\mathfrak{A}|$  is the domain for the second. ( $\mathcal{P}X$  = the set of all subsets of  $X$ .) If  $\alpha$  and  $\beta$  are elements of  $\mathfrak{A}^\downarrow$ , then

$$(169) \quad \langle \alpha, \beta \rangle \text{ is in } I_{\mathfrak{A}^\downarrow}(\varepsilon) \quad \text{iff} \quad \begin{array}{l} \alpha \text{ is an element of the first sort,} \\ \beta \text{ of the second sort, and } \alpha \in \beta. \end{array}$$

The constants of L are interpreted in the first sort of  $\mathfrak{A}^\downarrow$  just as they were in  $\mathfrak{A}$ . Now each second-order statement  $\phi$  about L-structures  $\mathfrak{A}$  is equivalent to

a first-order statement  $\phi^\downarrow$  about  $L^\downarrow$ -structures  $\mathfrak{A}^\downarrow$ . For example, if we use number superscripts to distinguish the first and second sorts of variables, the axiom of extensionality (168) translates into

$$(170) \quad \forall x^2 y^2 (\forall z^1 (z^1 \varepsilon x^2 \leftrightarrow z^1 \varepsilon y^2) \rightarrow x^2 = y^2).$$

Axiom (170) is a first-order sentence in  $L^\downarrow$ .

Let  $\mathbf{K}$  be the class of all  $L^\downarrow$ -structures of form  $\mathfrak{A}^\downarrow$  for some  $L$ -structure  $\mathfrak{A}$ . Let  $\mathbf{QC}^2$  be some standard proof calculus for second-order logic, and let  $\Delta$  be the set of all sentences  $\phi^\downarrow$  such that  $\phi$  is provable by  $\mathbf{QC}^2$ . Then  $\Delta$  is a set of axioms of  $\mathbf{K}$ , though not a complete one. The  $L^\downarrow$ -structures in  $\mathbf{K}$  are known as the *standard* models of  $\Delta$ . There will be plenty of non-standard models of  $\Delta$  too, but because of (170) they can all be seen as ‘parts of’ standard models in the following way. For each element  $\beta$  of the second sort in the model  $\mathfrak{B}$  of  $\Delta$ , let  $\beta^+$  be the set of elements  $\alpha$  such that  $\langle \alpha, \beta \rangle \in I_{\mathfrak{B}}(\varepsilon)$ . By (170),  $\beta^+ = \gamma^+$  implies  $\beta = \gamma$ . So in  $\mathfrak{B}$  we can replace each element  $\beta$  of the second sort by  $\beta^+$ . Then the second sort consists of subsets of the domain of the first sort, but not necessarily all the subsets. All the subsets are in the second domain if and only if this doctored version of  $\mathfrak{B}$  is a standard model. (Models of  $\Delta$ , standard or non-standard, are known as *Henkin models of second-order logic*, in view of [Henkin, 1950].)

How can one distinguish between a proof calculus for second-order logic on the one hand, and on the other hand a first-order proof calculus which also proves the sentences in  $\Delta$ ? The answer is easy: one can’t. In our notation above, the proof calculus for second-order logic has ‘ $P(z)$ ’ where the first-order calculus has ‘ $z^1 \varepsilon x^2$ ’, but this is no more than a difference of notation. Take away this difference and the two calculi become exactly the same thing. Don’t be misled by texts like Church [1956] which present ‘calculi of first order’ in one chapter and ‘calculi of second order’ in another. The latter calculi are certainly different from the former, because they incorporate a certain amount of set theory. But what makes them second-order calculi, as opposed to two-sorted first-order calculi with extra non-logical axioms, is *solely their intended interpretation*.

It follows, incidentally, that it is quite meaningless to ask whether the proof theory of actual mathematics is first-order or higher-order. (I recently saw this question asked. The questioner concluded that the problem is ‘not easy’.)

Where then can one meaningfully distinguish second-order from first-order? One place is the *classification of structures*. The class  $\mathbf{K}$  of standard models of  $\Delta$  is not a first-order definable class of  $L^\downarrow$ -structures, but it is second-order definable.

More controversially, we can distinguish between *first-order and second-order statements about a specific structure*, even when there is no question of classification. For example the sentence (168) says about an  $L$ -structure



$\mathfrak{A}$  something which can't be expressed in the first-order language of  $\mathfrak{A}$ . This is not a matter of classification, because (168) is true in *all* L-structures.

(iii) In Section 18 we studied the class of all L-structures with standard identity. Quine [1970, p. 63f] studies them too, and I admire his nerve. He first demonstrates that in any language L with finite similarity type there is a formula  $\phi$  which defines a congruence relation in every L-structure. From Section 18 we know that  $\phi$  cannot always express identity. Never mind, says Quine, let us *redefine* identity by the formula  $\phi$ . This happy redefinition instantly makes identity first-order definable, at least when the similarity type is finite. It also has the consequence, not mentioned by Quine, that for any two different things there is some language in which they are the same thing. (Excuse me for a moment while I redefine exams as things that I don't have to set.)

## 21 NOUN PHRASES

In this section I want to consider whether we can make any headway by adding to first-order logic some symbols for various types of noun phrase. Some types of noun phrase, such as 'most Xs', are not really fit for formalising because their meanings are too vague or too shifting. Of those which can be formalised, some never give us anything new, in the sense that any formula using a symbol for them is logically equivalent to a formula of first-order logic (with =); to express this we say that these formalisations give *conservative extensions* of first-order logic. Conservative extensions are not necessarily a waste of time. Sometimes they enable us to say quickly something that can only be said lengthily in first-order symbols, sometimes they behave more like natural languages than first-order logic does. So they may be useful to linguists or to logicians in a hurry.

Many (perhaps most) English noun phrases have to be symbolised as *quantifiers and not as terms*. For example the English sentence

(171) I have inspected every batch.

finds itself symbolised by something of form

(172) For every batch  $x$ , I have inspected  $x$ .

Let me recall the reason for this. If we copied English and simply put the noun phrase in place of the variable  $x$ , there would be no way of distinguishing between (i) the negation of 'I have inspected every batch' and (ii) the sentence which asserts, of every batch, that I have not inspected it. In style (172) there is no confusion between (i), viz.

(173)  $\neg$  For every batch  $x$ , I have inspected  $x$ .

and (ii), viz.

(174) For every batch  $x$ ,  $\neg$  I have inspected  $x$ .

Confusions like that between (i) and (ii) are so disastrous in logic that it constantly amazes logicians to see that natural languages, using style (171), have not yet collapsed into total anarchy.

In the logician's terminology, the *scope* of the quantifier 'For every batch  $x$ ' in (174) is the whole sentence, while in (173) it is only the part after the negation sign. Unlike its English counterpart, the quantifier doesn't *replace* the free occurrences of  $x$  in the predicate, it *binds* them. (More precisely, an occurrence of a quantifier with variable  $x$  binds all occurrences of  $x$  which are within its scope and not already bound.) This terminology carries over at once to the other kinds of quantifier that we shall consider, for example

(175)  $\neg$  For one in every three men  $x$ ,  $x$  is colour blind.

The quantifier 'For one in every three men  $x$ ' binds both occurrences of the variable, and doesn't include the negation in its scope.

I shall consider three groups of noun phrases. The first yield conservative extensions of first-order logic and are quite unproblematic. The second again give conservative extensions and are awkward. The third don't yield conservative extensions—we shall prove this. In all cases I assume that we start with a first-order language  $L$  with identity.

The *first* group are noun phrases such as 'At least  $n$  things  $x$  such that  $\phi$ '. We do it recursively:

(176)  $\exists_{\geq 0} x\phi$  is  $\neg\perp$ ;  $\exists_{\geq 1} x\phi$  is  $\exists x\phi$ .

(177)  $\exists_{\geq n+1} x\phi$  is  $\exists y(\phi[y/x] \wedge \exists_{\geq n} x(\neg x = y \wedge \phi))$  when  $n \geq 1$ .

To these definitions we add:

(178)  $\exists_{\leq n} x\phi$  is  $\neg\exists_{\geq n+1} x\phi$ .

(179)  $\exists_{=n} x\phi$  is  $\exists_{\geq n} x\phi \wedge \exists_{\leq n} x\phi$ .

$\exists_{=1} x\phi$  is sometimes written  $\exists!x\phi$ .

Definitions (176)–(179) are in the metalanguage; they simply select formulas of  $L$ . But there is no difficulty at all in adding the symbols  $\exists_{\geq n}$ ,  $\exists_{\leq n}$  and  $\exists_{=n}$  for each natural number to the language  $L$ , and supplying the needed extra clauses in the definition of  $\models$ , together with a complete formal calculus.

The *second* group are singular noun phrases of the form 'The such-and-such'. These are known as *definite descriptions*. Verbal variants of definite descriptions, such as 'My father's beard' for 'The beard of my father', are generally allowed to be definite descriptions too.

According to Bertrand Russell [1905], Whitehead and Russell [1910, Introduction, Chapter III], the sentence

(180) The author of ‘Slawkenburgius on Noses’ was a poet.

can be paraphrased as stating three things: (1) at least one person wrote ‘Slawkenburgius on Noses’; (2) at most one person wrote ‘Slawkenburgius on Noses’; (3) some person who did write ‘Slawkenburgius on Noses’ was a poet. I happily leave to Bencivenga [4.5] and Salmon [8.5] the question whether Russell was right about this. But assuming he was, his theory calls for the following symbolisation. We write ‘ $\{ix\psi\}$ ’ to represent ‘the person or thing  $x$  such that  $\psi$ ’, and we define

(181)  $\{ix\psi\}\phi$  to mean  $\exists_{=1}x\psi \wedge \exists x(\psi \wedge \phi)$ .

Expression (181) can be read either as a metalinguistic definition of a formula  $L$ , or as a shorthand explanation of how the expressions  $\{ix\psi\}$  can be added to  $L$ . In the latter case the definition of  $\models$  has to sprout one extra clause:

(182)  $\mathfrak{A} \models \{ix\psi\}\phi[g]$  iff there is a unique element  $\alpha$  of  $\mathfrak{A}$  such that  $\mathfrak{A} \models \psi[g, \alpha/x]$ , and for this  $\alpha$ ,  $\mathfrak{A} \models \phi[g, \alpha/x]$ .

There is something quite strongly counterintuitive about the formulas on either side in (181). It seems in a way obvious that when there is a unique such-and-such, we can refer to it by saying ‘the such-and-such’. But Russell’s paraphrase never allows us to use the expression  $\{ix\psi\}$  this way. For example if we want to say that the such-and-such equals 5, Russell will not allow us to render this as ‘ $\{ix\psi\} = 5$ ’. The expression  $\{ix\psi\}$  has the wrong grammatical type, and the semantical explanation in (182) doesn’t make it work like a name. On the right-hand side in (181) the position is even worse—the definition description has vanished without trace.

Leaving intuition on one side, there are any number of places in the course of formal calculation where one wants to be able to say ‘the such-and-such’, and then operate with this expression *as a term*. For example formal number theorists would be in dire straits if they were forbidden use of the term

(183)  $\mu x\psi$ , i.e. the least number  $x$  such that  $\psi$ .

Likewise formal set theorists need a term

(184)  $\{x|\psi\}$ , i.e. the set of all sets  $x$  such that  $\psi$ .

Less urgently, there are a number of mathematical terms which bind variables, for example the integral  $\int_b^a f(x)dx$  with bound variable  $x$ , which are naturally defined as ‘the number  $\lambda$  such that ... (here follows half a page of calculus)’. If we are concerned to formalise mathematics, the straightforward way to formalise such an integral is by a definite description term.

Necessity breeds invention, and in the event it is quite easy to extend the first-order language  $L$  by adding *terms*  $ix\psi$ . (The definitions of ‘term’ and

‘formula’ in Section 13 above have to be rewritten so that the classes are defined by simultaneous induction, because now we can form terms out of formulas as well as forming formulas out of terms.) There are two ways to proceed. One is to take  $\iota x\psi$  as a name of the unique element satisfying  $\psi$ , if there is such a unique element, and as undefined otherwise; then to reckon an atomic formula false whenever it contains an undefined term. This is equivalent to giving each occurrence of  $\iota x\psi$  the smallest possible scope, so that the notation need not indicate any scope. (Cf. [Kleene, 1952, p. 327]; [Kalish and Montague, 1964, Chapter VII].) The second is to note that questions of scope only arise if there is not a unique such-and-such. So we can choose a constant of the language, say 0, and read  $\iota x\psi$  as

(185) the element which is equal to the unique  $x$  such that  $\psi$  if there is such a unique  $x$ , and is equal to 0 if there is not.

(Cf. [Montague and Vaught, 1959; Suppes, 1972].)

Russell himself claimed to believe that definite descriptions ‘do not name’. So it is curious to note (as Kaplan does in his illuminating paper [1966] on Russell’s theory of descriptions) that Russell himself didn’t use the notation (181) which makes definite descriptions into quantifiers. What he did instead was to invent the notation  $\iota x\psi$  and then use it both as a quantifier and as a term, even though this makes for a contorted syntax. Kaplan detects in this ‘a lingering ambivalence’ in the mind of the noble lord.

The *third* group of noun phrases express things which can’t be said with first-order formulas. Peirce [1885] invented the *two-thirds* quantifier which enables us to say ‘At least  $\frac{2}{3}$  of the company have white neckties’. (His example.) Peirce’s quantifier was unrestricted. It seems more natural, and changes nothing in principle, if we allow a relativisation predicate and write  $\frac{2}{3}x(\psi, \phi)$  to mean ‘At least  $\frac{2}{3}$  of the things  $x$  which satisfy  $\psi$  satisfy  $\phi$ ’.

Can this quantifier be defined away in the spirit of (176)–(179)? Unfortunately not. Let me prove this. By a *functional* I shall mean an expression which is a first-order formula except that formula metavariables may occur in it, and it has no constant symbols except perhaps =. By substituting actual formulas for the metavariables, we get a first-order formula. Two functionals will be reckoned *logically equivalent* if whenever the same formulas are substituted for the metavariables in both functionals, the resulting first-order formulas are logically equivalent. For example the expression  $\exists_{\geq 2}x\phi$ , viz.

(186)  $\exists y(\phi[y/x] \wedge \exists x(\neg x = y \wedge \phi))$ ,

is a functional which is logically equivalent to  $\exists_{\geq 3}x\phi \vee \exists_{=2}x\phi$ . Notice that we allow the functional to change some variables which it binds, so as to avoid clash of variables.

A theorem of Skolem [1919] and Behmann [1922] (cf. [Ackermann, 1962, pp. 41–47]) states that *if a functional binds only one variable in each in-*

serted formula, then it is logically equivalent to a combination by  $\neg, \wedge$  and  $\vee$  of equations  $y = z$  and functionals of the form  $\exists_{=n}x\chi$  where  $\chi$  is a functional without quantifiers. Suppose now that we could define away the quantifier  $\frac{2}{3}x(\cdot)$ . The result would be a functional binding just the variable  $x$  in  $\psi$  and  $\phi$ , so by the Skolem–Behmann theorem we could rewrite it as a propositional compound of a finite number of functionals of the form  $\exists_{=n}x\chi$ , and some equations. (The equations we can forget, because the meaning of  $\frac{2}{3}x(\psi, \phi)$  shows that it has no significant free variables beyond those in  $\psi$  or  $\phi$ .) If  $n$  is the greatest integer for which  $\exists_{=n}x$  occurs in the functional, then the functional is incapable of distinguishing any two numbers greater than  $n$ , so that it can't possibly express that one of them is at least  $\frac{2}{3}$  times the other.

A harder example is

(187) The average Briton speaks at least two-thirds of a foreign language.

I take this to mean that if we add up the number of foreign languages spoken by each Briton, and divide the sum total by the number of Britons, then the answer is at least  $\frac{2}{3}$ . Putting  $\psi(x)$  for ‘ $x$  is a Briton’ and  $\phi(x, y)$  for ‘ $y$  is a foreign language spoken by  $x$ ’, this can be symbolised as  $\{Av\frac{2}{3}xy\}(\psi, \phi)$ . Can the quantifier  $\{Av\frac{2}{3}xy\}$  be defined away in a first-order language? Again the answer is no. This time the Skolem–Behmann result won't apply directly, because  $\{Av\frac{2}{3}xy\}$  binds two variables,  $x$  and  $y$ , in the second formula  $\phi$ . But indirectly the same argument will work.  $\frac{2}{3}x(\psi, \phi)$  expresses just the same thing as  $\forall z(\psi[z/x] \rightarrow \{Av\frac{2}{3}xy\}(\psi, z = x \wedge \phi[y/x] \wedge \psi[y/x]))$ . Hence if  $\{Av\frac{2}{3}xy\}$  could be defined away, then so could  $\frac{2}{3}x$ , and we have seen that this is impossible.

Barwise and Cooper [1981] made a thorough study of the logical properties of natural language noun phrases. See also [Montague, 1970; Montague, 1973], particularly his discussion of ‘the’. Van Benthem and Doets (this Volume) have a fuller discussion of things not expressible in first-order language.

### III: The Expressive Power of First-order Logic

#### 22 AFTER ALL THAT, WHAT IS FIRST-ORDER LOGIC?

It may seem perverse to write twenty-one sections of a chapter about elementary (i.e. first-order) logic without ever saying what elementary logic is. But the easiest definition is ostensive: elementary logic is the logic that we have been doing in Sections 1–18 above. But then, why set *that* logic apart from any other? What particular virtues and vices does it have?

At first sight the Traditional Logician might well prefer a stronger logic. After all, the more valid argument schemas you can find him the happier he is. But in fact Traditional Logicians tend to draw a line between what is ‘genuinely logic’ and what is really mathematics. The ‘genuine logic’ usually turns out to be a version of first-order logic.

One argument often put forward for this choice of ‘genuine logic’ runs along the following lines. In English we can group the parts of speech into two groups. The first group consists of *open classes* such as nouns, verbs, adjectives. These classes expand and contract as people absorb new technology or abandon old-fashioned morality. Every word in these classes carries its own meaning and subject-matter. In the second group are the *closed classes* such as pronouns and conjunctions. Each of these classes contains a fixed, small stock of words; these words have no subject-matter, and their meaning lies in the way they combine with open-class words to form phrases. Quirk and Greenbaum [1973, p.18] list the following examples of closed-class words: the, a, that, this, he, they, anybody, one, which, of, at, in, without, in spite of, and, that, when, although, oh, ah, ugh, phew.

The Traditional Logicians’ claim is essentially this: ‘genuine logic’ is the logic which assembles those valid argument schemas in which open-class words are replaced by schematic letters and closed-class words are not. Quirk and Greenbaum’s list already gives us  $\wedge$  ‘and’,  $\neg$  ‘without’,  $\forall$  ‘anybody’,  $\exists$  ‘a’, and of course the words ‘not’, ‘if’, ‘then’, ‘or’ are also closed-class words. The presence of ‘at’, ‘in spite of’ and ‘phew’ in their list doesn’t imply we ought to have added any such items to our logic, because these words don’t play any distinctive role in arguments. (The presence of ‘when’ is suggestive though.) Arguably it is impossible to express second-order conditions in English without using open-class words such as ‘set’ or ‘concept’.

It’s a pretty theory. Related ideas run through Quine’s [1970]. But for myself I can’t see why features of the surface grammar of a few languages that we know and love should be considered relevant to the question what is ‘genuine logic’.

We turn to the Proof Theorist. His views are not very helpful to us here. As we saw in Section 20 above, there is in principle no difference between a first-order proof calculus and a non-first-order one. Still, he is likely to make the following comment, which is worth passing on. For certain kinds of application of logic in mathematics, a stronger logic may lead to weaker results. To quote one example among thousands: in a famous paper [1965] Ax and Kochen showed that for each positive integer  $d$  there are only finitely many primes which contradict a conjecture of Artin about  $d$ . Their proof used heavy set theory and gave no indication what these primes were. Then Cohen [1969] found a proof of the same result using no set-theoretic assumptions at all. From his proof one can calculate, for each  $d$ , what the bad primes are. By using the heavy guns, Ax and Kochen had

gained intuition but lost information. The moral is that we should think twice before strengthening our logic. The mere fact that a thing is provable in a weaker logic may lead us to further information.

We turn to the Model Theorist. He was probably taught that ‘first-order’ means we only quantify over elements, not over subsets of the domain of a structure. By now he will have learned (Section 21 above) that some kinds of quantification over elements are not first-order either.

What really matters to a Model Theorist in his language is the interplay of strength and weakness. Suppose he finds a language which is so weak that it can’t tell a Montagu from a Capulet. Then at once he will try to use it to prove things about Capulets, as follows. First he shows that something is true for all Montagus, and then he shows that this thing is expressible in his weak language  $L$ . Then this thing must be true for at least one Capulet too, otherwise he could use it to distinguish Montagus from Capulets in  $L$ . If  $L$  is bad enough at telling Montagus and Capulets apart, he may even be able to deduce that *all* Capulets have the feature in question. These methods, which are variously known as *overspill* or *transfer* methods, can be extremely useful if Montagus are easier to study than Capulets.

It happens that first-order languages are excellent for encoding finite combinatorial information (e.g. about finite sequences or syntax), but hopelessly bad at distinguishing one infinite cardinal or infinite ordering from another infinite cardinal or infinite ordering. This particular combination makes first-order model theory very rich in transfer arguments. For example the whole of Abraham Robinson’s non-standard analysis [Robinson, 1967] is one vast transfer argument. The Model Theorist will not lightly give up a language which is as splendidly weak as the Upward and Downward Löwenheim–Skolem Theorem and the Compactness Theorem (Section 17 above) show first-order languages to be.

This is the setting into which Per Lindström’s theorem came (Section 27 below). He showed that any language which has as much coding power as first-order languages, but also the same weaknesses which have just been mentioned, must actually be a first-order language in the sense that each of its sentences has exactly the same models as some first-order sentence.

## 23 SET THEORY

In 1922 Skolem described a set of first-order sentences which have become accepted, with slight variations, as the definitive axiomatisation of set theory and hence in some sense a foundation for mathematics. Skolem’s axioms were in fact a first-order version of the informal axioms which Zermelo [1908] had given, together with one extra axiom (Replacement) which Fraenkel [1922] had also seen was necessary. The axioms are known as ZFC—Zermelo–Fraenkel set theory with Choice. They are listed in Ap-

pendix C below and developed in detail in [Suppes, 1972] and [Levy, 1979].

When these axioms are used as a foundation for set theory or any other part of mathematics, they are read as being about a particular collection  $V$ , the class of all sets. Mathematicians differ about whether we have any access to this collection  $V$  independently of the axioms. Some writers [Gödel, 1947] believe  $V$  is the standard model of the axioms, while others [von Neumann, 1925] regard the symbol ' $V$ ' as having no literal meaning at all. But everybody agrees that the axioms have a standard reading, namely as being about  $V$ . In this the axioms of ZFC differ from, say, the axioms for group theory, which are never read as being about The Group, but simply as being true in any group.

These axioms form a foundation for mathematics in two different ways. First, some parts of mathematics are directly about sets, so that all their theorems can be phrased quite naturally as statements about  $V$ . For example the natural numbers are now often taken to be sets. If they are sets, then the integers, the rationals, the reals, the complex numbers and various vector spaces over the complex numbers are sets too. Thus the whole of real and complex analysis is now recognised as being part of set theory and can be developed from the axioms of ZFC.

Some other parts of mathematics are not about sets, but can be *encoded* in  $V$ . We already have an example in Section 17 above, where we converted languages into sets. There are two parts to an encoding. First the entities under discussion are replaced by sets, and we check that all the relations between the original entities go over into relations in  $V$  that can be defined within the language of first-order set theory. In the case of our encoded languages, it was enough to note that any finite sequence of sets  $a_1, \dots, a_n$  can be coded into an ordered  $n$ -tuple  $\langle a_1, \dots, a_n \rangle$ , and that lengths of sequences, concatenations of sequences and the result of altering one term of a sequence can all be defined. (Cf. [Gandy, 1974].)

The second part of an encoding is to check that all the theorems one wants to prove can be deduced from the axioms of ZFC. Most theorems of elementary syntax can be proved using only the much weaker axioms of Kripke–Platek set theory (cf. [Barwise, 1975]); these axioms plus the axiom of infinity suffice for most elementary model theory too. (Harnik [1985] and [1987] analyses the set-theoretic assumptions needed for various theorems in model theory.) Thus the possibility of encoding pieces of mathematics in set theory rests on two things: first the expressive power of the first-order language for talking about sets, and second the proving power of the set-theoretic axioms. Most of modern mathematics lies within  $V$  or can be encoded within it in the way just described. Not all the encodings can be done in a uniform way; see for example Feferman [1969] for a way of handling tricky items from category theory, and the next section below for a trickier item from set theory itself. I think it is fair to say that all of modern mathematics can be encoded in set theory, but it has to be done locally and



not all at once, and sometimes there is a perceptible loss of meaning in the encoding. (Incidentally the rival system of *Principia Mathematica*, using a higher-order logic, came nowhere near this goal. As Gödel says of *Principia* in his [1951]: ‘it is clear that the theory of real numbers in its present form cannot be obtained’.)

One naturally asks how much of the credit for this universality lies with first-order logic. Might a weaker logic suffice? The question turns out to be not entirely well-posed; if this other logic can in some sense express everything that can be expressed in first-order logic, then in what sense is it ‘weaker’? In case any reader feels disposed to look at the question and clarify it, let me mention some reductions to other logics.

First, workers in logic programming or algebraic specification are constantly reducing first-order statements to universal Horn expressions. One can systematise these reductions; see for example Hodges [1993b, Section 10], or Padawitz [1988, Section 4.8]. Second, using very much subtler methods, Tarski and Givant [1987] showed that one can develop set theory within an equational relational calculus  $\mathcal{L}^\times$ . In their Preface they comment:

...  $\mathcal{L}^\times$  is equipollent (in a natural sense) to a certain fragment ... of first-order logic having one binary predicate and containing just three variables. ... It is therefore quite surprising that  $\mathcal{L}^\times$  proves adequate for the formalization of practically all known systems of set theory and hence for the development of all of classical mathematics.

And third, there may be some mileage in the fact that essentially any piece of mathematics can be encoded in an elementary topos (cf. [Johnstone, 1977]).

Amazingly, Skolem’s purpose in writing down the axioms of ZFC was to debunk the enterprise: ‘But in recent times I have seen to my surprise that so many mathematicians think that these axioms of set theory provide the ideal foundation for mathematics; therefore it seemed to me that the time had come to publish a critique’ [Skolem, 1922].

In fact Skolem showed that, since the axioms form a countable first-order theory, they have a countable model  $\mathfrak{A}$ . In  $\mathfrak{A}$  there are ‘sets’ which satisfy the predicate ‘ $x$  is uncountable’, but since  $\mathfrak{A}$  is countable, these ‘sets’ have only countably many ‘members’. This has become known as Skolem’s Paradox, though in fact there is no paradox. The set-theoretic predicate ‘ $x$  is uncountable’ is written so as to catch the uncountable elements of  $V$ , and there is no reason at all to expect it to distinguish the uncountable elements of other models of set theory. More precisely, this predicate says ‘there is no 1–1 function from  $x$  to the set  $\omega$ ’. In a model  $\mathfrak{A}$  which is different from  $V$ , this only expresses that there is no function which is an element of  $\mathfrak{A}$  and which is 1–1 from  $x$  to  $\omega$ .

According to several writers the real moral of Skolem's Paradox is that there is no standard model of ZFC, since for any model  $\mathfrak{A}$  of ZFC there is another model  $\mathfrak{B}$  which is not isomorphic to  $\mathfrak{A}$  but is indistinguishable from  $\mathfrak{A}$  by first-order sentences. If you have already convinced yourself that the only things we can say about an abstract structure  $\mathfrak{A}$  are of the form 'Such-and-such first-order sentences are true in  $\mathfrak{A}$ ', then you should find this argument persuasive. (See [Klenk, 1976; Putnam, 1980] for further discussion.)

Skolem's own explanation of why his argument debunks axiomatic set-theoretic foundations is very obscure. He says in several places that the conclusion is that the meaning of 'uncountable' is relative to the axioms of set theory. I have no idea what this means. The obvious conclusion, surely, is that the meaning of 'uncountable' is relative to the *model*. But Skolem said that he didn't believe in the existence of uncountable sets anyway, and we learn he found it disagreeable to review the articles of people who did [Skolem, 1955].

Contemporary set theorists make free use of non-standard—especially countable—models of ZFC. One usually requires the models to be well-founded, i.e. to have no elements which descend in an infinite sequence

$$(188) \quad \cdots \in a_2 \in a_1 \in a_0.$$

It is easy to see that this is not a first-order condition on models (for example, Hodges [1972] constructs models of full first-order set theory with arbitrarily long descending sequences of ordinals but no uncountable increasing well-ordered sequences—these models are almost inversely well-founded.) However, if we restrict ourselves to models which are subsets of  $V$ , then the statement that such a model contains no sequence (188) can be written as a first-order formula in the language of  $V$ . The moral is that it is simply meaningless to classify mathematical statements absolutely as 'first-order' or 'not first-order'. One and the same statement can perfectly well express a second-order condition on structure  $\mathfrak{A}$  but a first-order condition on structure  $\mathfrak{B}$ . (Cf. Section 20 above.)

Meanwhile since the 1950s a number of set theorists have been exploring first-order axioms which imply that the universe of sets is *not* well-founded. Axioms of this kind are called *anti-foundation axioms*; they are rivals to the Foundation (or Regularity) axiom ZF3 in Appendix C below. For many years this work went largely unnoticed, probably because nobody saw any foundational use for it (forgive the pun). But in the 1980s Aczel [1988] saw how to use models of anti-foundation axioms in order to build representations of infinite processes. Barwise generalised Aczel's idea and used non-well-founded sets to represent self-referential phenomena in semantics and elsewhere (cf. [Barwise and Moss, 1996]). Of course there is no problem about describing non-well-founded relations in conventional set theory. The advantage of models of anti-foundation axioms is that they take the

membership relation  $\in$  itself to be non-well-founded, and it is claimed that this allows us to fall back on other intuitions that we already have about set membership.

## 24 ENCODING SYNTAX

I begin by showing that the definition of truth in the class  $V$  of all sets is not itself expressible in  $V$  by a first-order formula. This will demonstrate that there is at least one piece of mathematics which can't be encoded in set theory without serious change of meaning.

As we saw in the previous section, there is no problem about encoding the first-order language  $L$  of set theory into  $V$ . Without going into details, let me add that we can go one stage further and add to the language  $L$  a name for each set; the resulting language  $L^+$  can still be encoded in  $V$  as a definable proper class. Let us assume this has been done, so that every formula of  $L^+$  is in fact a set. For each set  $b$ , we write  $\ulcorner b \urcorner$  for the constant of  $L^+$  which names  $b$ . (This is nothing to do with Quine's corners  $\ulcorner \urcorner$ .) When we speak of sentences of  $L^+$  being true in  $V$ , we mean that they are true in the structure whose domain is  $V$  where ' $\in$ ' is interpreted as set membership and each constant  $\ulcorner b \urcorner$  is taken as a name of  $b$ .

A class  $X$  of sets is said to be *definable* by the formula  $\psi$  if for every set  $\alpha$ ,

$$(189) \quad V \models \psi[\alpha/x] \text{ iff } \alpha \in X.$$

Since every set  $\alpha$  has a name  $\ulcorner \alpha \urcorner$ , (189) is equivalent to:

$$(190) \quad V \models \psi(\ulcorner \alpha \urcorner/x) \text{ iff } \alpha \in X$$

where I now write  $\psi(\ulcorner \alpha \urcorner/x)$  for the result of putting  $\ulcorner \alpha \urcorner$  in place of free occurrences of  $x$  in  $\psi$ .

Suppose now that the class of true sentences of  $L^+$  can be defined by a formula *True* of  $L^+$  with the free variable  $x$ . Then for every sentence  $\phi$  of  $L^+$ , according to (190),

$$(191) \quad V \models \text{True}(\ulcorner \phi \urcorner/x) \text{ iff } V \models \phi.$$

But since the syntax of  $L^+$  is definable in  $V$ , there is a formula  $\chi$  of  $L^+$  with just  $x$  free, such that for every formula  $\phi$  of  $L^+$  with just  $x$  free, if  $\ulcorner \phi \urcorner = b$  then

$$(192) \quad V \models \chi(\ulcorner b \urcorner/x) \text{ iff } V \models \neg \text{True}(\ulcorner \phi(\ulcorner b \urcorner/x) \urcorner/x).$$

Now put  $b = \ulcorner \chi \urcorner$ . Then by (191) and (192),

$$(193) \quad V \models \chi(\ulcorner b \urcorner/x) \text{ iff } V \models \text{True}(\ulcorner \chi(\ulcorner b \urcorner/x) \urcorner/x) \text{ iff } V \models \neg \chi(\ulcorner b \urcorner/x).$$

Evidently the two ends of (193) make a contradiction. Hence the class of true sentences of  $L$  can't be defined by any formula of  $L$ . Thus we have shown that

**THEOREM 15.** *The class of pairs  $\langle \phi, g \rangle$  where  $\phi$  is a formula of the language  $L$  of set theory,  $g$  is an assignment in  $V$  and  $V \models \phi[g]$ , is not definable in  $V$  by any formula of the language  $L^+$  of set theory with names for arbitrary sets.*

This is one version of Tarski's [1935] *theorem on the undefinability of truth*. Another version, with essentially the same proof, is:

**THEOREM 16.** *The class of sentences  $\phi$  of  $L$  which are true in  $V$  is not definable in  $V$  by any formula of  $L$ .*

Of course the set  $b$  of all true sentences of  $L$  would be definable in  $V$  if we allowed ourselves a name for  $b$ . Hence the difference between Theorems 15 and 16. These two theorems mean that the matter of truth in  $V$  has to be handled either informally or not at all.

Lévy [1965] gives several refined theorems about definability of truth in  $V$ . He shows that truth for certain limited classes of sentences of  $L^+$  can be defined in  $V$ ; in fact each sentence of  $L^+$  lies in one of his classes. As I remarked earlier, everything can be encoded, but not all at once.

Tarski's argument was based on a famous paper of Gödel [1931b], to which I now turn. When formalising the language of arithmetic it is common to include two restricted quantifiers ( $\forall x < y$ ) and ( $\exists x < y$ ), meaning respectively 'for all  $x$  which are less than  $y$ ' and 'there is an  $x$  which is less than  $y$ , such that'. A formula in which every quantifier is restricted is called a  $\Delta_0$  formula. Formulas of form  $\forall \vec{x}\phi$  and  $\exists \vec{x}\phi$ , where  $\phi$  is a  $\Delta_0$  formula, are said to be  $\Pi_1$  and  $\Sigma_1$  respectively. (See under 'Arithmetical hierarchy' in van Dalen (this Volume).)

$N$  shall be the structure whose elements are the natural numbers; each natural number is named by an individual constant  $\ulcorner n \urcorner$ , and there are relations or functions giving 'plus' and 'times'. A relation on the domain of  $N$  which is defined by a  $\Pi_1$  or  $\Sigma_1$  formula is said to be a  $\Pi_1$  or  $\Sigma_1$  relation respectively. Some relations can be defined in both ways; these are said to be  $\Delta_1$  relations. The interest of these classifications lies in a theorem of Kleene [1943].

**THEOREM 17.** *An  $n$ -place relation  $R$  on the natural numbers is  $\Delta_1$  iff there is a computational test which decides whether any given  $n$ -tuple is in  $R$ ; an  $n$ -tuple relation  $R$  on the natural numbers is  $\Sigma_1$  iff a computer can be programmed to print out all and only the  $n$ -tuples in  $R$ .*

Hilbert in [1926], the paper that started this whole line of enquiry, had laid great stress on the fact that we can test the truth of a  $\Delta_0$  sentence in a finite number of steps, because each time we meet a restricted quantifier we have only to check a finite number of numbers. This is the central idea of

the proofs from left to right in Kleene's equivalences. The other directions are proved by encoding computers into  $N$ ; see Theorems 2.5 and 2.14 in Van Dalen (this Volume).

Now all grammatical properties of a sentence can be checked by mechanical computation. So we can encode the language of first-order Peano arithmetic into  $N$  in such a way that all the grammatical notions are expressed by  $\Delta_1$  relations. (This follows from Theorem 17, but Gödel [1931b] wrote out an encoding explicitly.) We shall suppose that this has been done, so that from now on every formula or symbol of the language of arithmetic is simply a number. Thus every formula  $\phi$  is a number which is named by the individual constant  $\ulcorner \phi \urcorner$ . Here  $\ulcorner \phi \urcorner$  is also a number, but generally a different number from  $\phi$ ;  $\ulcorner \phi \urcorner$  is called the *Gödel number* of  $\phi$ . Note that if  $T$  is any mechanically describable theory in the language of arithmetic, then a suitably programmed computer can spew out all the consequences of  $T$  one by one, so that by Kleene's equivalences (Theorem 17), the set of all sentences  $\phi$  such that  $T \vdash \phi$  is a  $\Sigma_1$  set.

We need one other piece of general theory. Tarski *et al.* [1953] describe a sentence  $Q$  in the language of arithmetic which is true in  $N$  and has the remarkable property that for every  $\Sigma_1$  sentence  $\phi$ ,

$$(194) \quad Q \vdash \phi \text{ iff } N \vDash \phi.$$

We shall use these facts to show that the set of numbers  $n$  which are not sentences deducible from  $Q$  is not a  $\Sigma_1$  set. Suppose it were a  $\Sigma_1$  set, defined by the  $\Sigma_1$  formula  $\psi$ . Then for every number  $n$  we would have

$$(195) \quad N \vDash \psi(\ulcorner n \urcorner/x) \quad \text{iff} \quad \text{not}(Q \vdash n).$$

Now since all syntactic notions are  $\Delta_1$ , with a little care one can find a  $\Sigma_1$  formula  $\chi$  with just  $x$  free, such that for every formula  $\phi$  with just  $x$  free, if  $\ulcorner \phi \urcorner = n$  then

$$(196) \quad N \vDash \chi(\ulcorner n \urcorner/x) \quad \text{iff} \quad N \vDash \psi(\ulcorner \phi(\ulcorner n \urcorner/x) \urcorner/x).$$

Putting  $n = \ulcorner \chi \urcorner$  we get by (194), (195) and (196):

$$(197) \quad \begin{aligned} N \vDash \chi(\ulcorner n \urcorner/x) & \text{ iff } N \vDash \psi(\ulcorner \chi(\ulcorner n \urcorner/x) \urcorner/x) \\ & \text{ iff not}(Q \vdash \chi(\ulcorner n \urcorner/x)) \\ & \text{ iff not}(N \vDash \chi(\ulcorner n \urcorner/x)) \end{aligned}$$

where the last equivalence is because  $\chi(\ulcorner n \urcorner/x)$  is a  $\Sigma_1$  sentence. The two ends of (197) make a contradiction; so we have proved that the set of numbers  $n$  which are not sentences deducible from  $Q$  is not  $\Sigma_1$ . Hence the set of numbers which *are* deducible is not  $\Delta_1$ , and therefore by Theorem 17 there is no mechanical test for what numbers belong to it. We have proved: there is no mechanical test which determines, for any given sentence  $\phi$  of

the language of arithmetic, whether or not  $\vdash (Q \rightarrow \phi)$ . This immediately implies Church's theorem [1936]:

**THEOREM 18.** *There is no mechanical test to determine which sentences of first-order languages are logically valid.*

Now we can very easily prove a weak version of Gödel's [1931b] incompleteness theorem too. Let  $P$  be first-order Peano arithmetic. Then it can be shown that  $P \vdash Q$ . Hence from (194) we can infer that (194) holds with  $P$  in place of  $Q$ . So the same argument as above shows that the set of non-consequences of  $P$  is not  $\Sigma_1$ . If  $P$  had as consequences all the sentences true in  $N$ , then the non-consequences of  $P$  would consist of (i) the sentences  $\phi$  such that  $P \vdash \neg\phi$ , and (ii) the numbers which are not sentences. But these together form a  $\Sigma_1$  set. Hence, as Gödel proved,

**THEOREM 19.** *There are sentences which are true in  $N$  but not deducible from  $P$ .*

Finally Tarski's theorem (Theorems 15, 16) on the undefinability of truth applies to arithmetic just as well as to set theory. A set of numbers which is definable in  $N$  by a first-order formula is said to be *arithmetical*. Tarski's theorem on the undefinability of truth in  $N$  states:

**THEOREM 20.** *The class of first-order sentences which are true in  $N$  is not arithmetical.*

Van Benthem and Doets (this Volume) show why Theorem 19 implies that there can be no complete formal proof calculus for second-order logic.

For work connecting Gödel's argument with modal logic, see Boolos [1979; 1993] and Smoryński (Volume 9 of this *Handbook*).

## 25 SKOLEM FUNCTIONS

When Hilbert interpreted  $\exists x\phi$  as saying in effect 'The element  $x$  which I choose satisfies  $\phi$ ' (cf. Section 15 above), Brouwer accused him of 'causing mathematics to degenerate into a game' [Hilbert, 1928]. Hilbert was delighted with this description, as well he might have been, since games which are closely related to Hilbert's idea have turned out to be an extremely powerful tool for understanding quantifiers.

Before the technicalities, here is an example. Take the sentence

(198) Everybody in Croydon owns a dog.

Imagine a game  $G$ : you make the first move by producing someone who lives in Croydon, and I have to reply by producing a dog. I win if and only if the dog I produced belongs to the person you produced. Assuming that I have free access to other people's dogs, (198) is true if and only if I can always win the game  $G$ . This can be rephrased: (198) is true if and only if

there is a function  $F$  assigning a dog to each person living in Croydon, such that whenever we play  $G$ , whatever person  $x$  you produce, if I retaliate with dog  $F(x)$  then I win. A function  $F$  with this property is called a *winning strategy* for me in the game  $G$ . By translating (198) into a statement about winning strategies, we have turned a statement of form  $\forall x \exists y \phi$  into one of form  $\exists F \forall x \psi$ .

Now come the technicalities. For simplicity, I shall assume that our language  $L$  doesn't contain  $\perp$ ,  $\rightarrow$  or  $\leftrightarrow$ , and that all occurrences of  $\neg$  are immediately in front of atomic formulas. The arguments of Sections 5 and 15 show that every first-order formula is logically equivalent to one in this form, so the theorems proved below hold without this restriction on  $L$ .  $\mathfrak{A}$  shall be a fixed  $L$ -structure. For each formula  $\phi$  of  $L$  and assignment  $g$  in  $\mathfrak{A}$  to the free variables of  $\phi$ , we shall define a game  $G(\mathfrak{A}, \phi; g)$  to be played by two players  $\forall$  and  $\exists$  (male and female). The definition of  $G(\mathfrak{A}, \phi; g)$  is by induction on the complexity of  $\phi$ , and it very closely follows the definition of  $\models$  in Section 14:

1. If  $\phi$  is atomic then neither player makes any move in  $G(\mathfrak{A}, \phi; g)$  or  $G(\mathfrak{A}, \neg\phi; g)$ ; player  $\exists$  wins  $G(\mathfrak{A}, \phi; g)$  if  $\mathfrak{A} \models \phi[g]$ , and she wins  $G(\mathfrak{A}, \neg\phi; g)$  if  $\mathfrak{A} \models \neg\phi[g]$ ; player  $\forall$  wins iff player  $\exists$  doesn't win.
2. Suppose  $\phi$  is  $\psi \wedge \chi$ , and  $g_1$  and  $g_2$  are respectively the restrictions of  $g$  to the free variables of  $\psi$ ,  $\chi$ ; then player  $\forall$  has the first move in  $G(\mathfrak{A}, \phi; g)$ , and the move consists of deciding whether the game shall proceed as  $G(\mathfrak{A}, \psi; g_1)$  or as  $G(\mathfrak{A}, \chi; g_2)$ .
3. Suppose  $\phi$  is  $\psi \vee \chi$ , and  $g_1, g_2$  are as in (2); then player  $\exists$  moves by deciding whether the game shall continue as  $G(\mathfrak{A}, \psi; g_1)$  or  $G(\mathfrak{A}, \chi; g_2)$ .
4. If  $\phi$  is  $\forall x \psi$  then player  $\forall$  chooses an element  $\alpha$  of  $\mathfrak{A}$ , and the game proceeds as  $G(\mathfrak{A}, \psi; g, \alpha/x)$ .
5. If  $\phi$  is  $\exists x \psi$  then player  $\exists$  chooses an element  $\alpha$  of  $\mathfrak{A}$ , and the game proceeds as  $G(\mathfrak{A}, \psi; g, \alpha/x)$ .

If  $g$  is an assignment suitable for  $\phi$ , and  $h$  is the restriction of  $g$  to the free variables of  $\phi$ , then  $G(\mathfrak{A}, \phi; g)$  shall be  $G(\mathfrak{A}, \phi; h)$ . When  $\phi$  is a sentence,  $h$  is empty and we write the game simply as  $G(\mathfrak{A}, \phi)$ .

The quantifier clauses for these games were introduced in [Henkin, 1961]. It is then clear how to handle the other clauses; see [Hintikka, 1973, Chapter V]. Lorenzen [1961; 1962] (cf. also Lorenzen and Schwemmer [1975]) described similar games, but in his versions the winning player had to *prove* a sentence, so that his games turned out to define intuitionistic provability where ours will define truth. (Cf. Felscher (Volume 7 of this *Handbook*.) In Lorenzen [1962] one sees a clear link with cut-free sequent proofs.

A *strategy* for a player in a game is a set of rules that tell him how he should play, in terms of the previous moves of the other player. The strategy is called *winning* if the player wins every time he uses it, regardless of how the other player moves. Leaving aside the game-theoretic setting, the next result probably ought to be credited to Skolem [1920]:

**THEOREM 21.** *Assume the axiom of choice (cf. Appendix C). Then for every L-structure  $\mathfrak{A}$ , every formula  $\phi$  of L and every assignment  $g$  in  $\mathfrak{A}$  which is suitable for  $\phi$ ,  $\mathfrak{A} \models \phi[g]$  iff player  $\exists$  has a winning strategy for the game  $G(\mathfrak{A}, \phi; g)$ .*

Theorem 21 is proved by induction on the complexity of  $\phi$ . I consider only clause (4), which is the one that needs the axiom of choice. The ‘if’ direction is not hard to prove. For the ‘only if’, suppose that  $\mathfrak{A} \models \forall x\psi[g]$ , where  $g$  is an assignment to the free variables of  $\forall x\psi$ . Then  $\mathfrak{A} \models \psi[g, \alpha/x]$  for every element  $\alpha$ ; so by the induction assumption, player  $\exists$  has a winning strategy for each  $G(\mathfrak{A}, \psi; g, \alpha/x)$ . Now *choose* a winning strategy  $S_\alpha$  for player  $\exists$  in each game  $G(\mathfrak{A}, \psi; g, \alpha/x)$ . Player  $\exists$ ’s winning strategy for  $G(\mathfrak{A}, \phi; g)$  shall be as follows: wait to see what element  $\alpha$  player  $\forall$  chooses, and then follow  $S_\alpha$  for the rest of the game.

Theorem 21 has a wide range of consequences. First, it shows that games can be used to give a definition of truth in structures. In fact this was Henkin’s purpose in introducing them. See Chapter III of Hintikka [1973] for some phenomenological reflections on this kind of truth-definition.

For the next applications we should bear in mind that *every first-order formula can be converted into a logically equivalent first-order formula which is prenex, i.e. with all its quantifiers at the left-hand end.* (Cf. (127).) When  $\phi$  is prenex, a strategy for player  $\exists$  takes a particularly simple form. It consists of a set of functions, one for each existential quantifier in  $\phi$ , which tell player  $\exists$  what element to choose, depending on what elements were chosen by player  $\forall$  at earlier universal quantifiers.

For example if  $\phi$  is  $\forall x\exists y\forall z\exists tR(x, y, z, t)$ , then a strategy for player  $\exists$  in  $G(\mathfrak{A}, \phi)$  will consist of two functions, a 1-place function  $F_y$  and a 2-place function  $F_t$ . This strategy will be winning if and only if

$$(199) \text{ for all elements } \alpha \text{ and } \gamma, \mathfrak{A} \models R(x, y, z, t)[\alpha/x, F_y(\alpha)/y, \gamma/z, F_t(\alpha, \gamma)/t].$$

Statement (199) can be paraphrased as follows. Introduce new function symbols  $f_y$  and  $f_t$ . Write  $\phi^\wedge$  for the sentence got from  $\phi$  by removing the existential quantifiers and then putting  $f_y(x), f_t(x, z)$  in place of  $y, t$  respectively. So  $\phi^\wedge$  is  $\forall x\forall zR(x, f_y(x), z, f_t(x, z))$ . We expand  $\mathfrak{A}$  to a structure  $\mathfrak{A}^\wedge$  by adding interpretations  $I_{\mathfrak{A}^\wedge}(f_y)$  and  $I_{\mathfrak{A}^\wedge}(f_t)$  for the new function symbols; let  $F_y$  and  $F_t$  be these interpretations. Then by (199),

$$(200) F_y, F_t \text{ are a winning strategy for player } \exists \text{ in } G(\mathfrak{A}, \phi) \text{ iff } \mathfrak{A}^\wedge \models \phi^\wedge.$$



Functions  $F_y, F_t$  which do satisfy either side of (200) are called *Skolem functions* for  $\phi$ . Putting together (200) and Theorem 21, we get

(201)  $\mathfrak{A} \models \phi$  iff by adding functions to  $\mathfrak{A}$  we can get a structure  $\mathfrak{A}^\wedge$  such that  $\mathfrak{A}^\wedge \models \phi^\wedge$ .

A sentence  $\phi^\wedge$  can be defined in the same way whenever  $\phi$  is any prenex sentence; (201) will still apply. Note that  $\phi^\wedge$  is of the form  $\forall \vec{x}\psi$  where  $\psi$  has no quantifiers; a formula of this form is said to be *universal*.

From (201) we can deduce:

**THEOREM 22.** *Every prenex first-order sentence  $\phi$  is logically equivalent to a second-order sentence  $\exists \vec{f}\phi^\wedge$  in which  $\phi^\wedge$  is universal.*

In other words, we can always push existential quantifiers to the left of universal quantifiers, provided that we convert the existential quantifiers into second-order function quantifiers  $\exists \vec{f}$ . Another consequence of (201) is:

**LEMMA 23.** *For every prenex first-order sentence  $\phi$  we can effectively find a universal sentence  $\phi^\wedge$  which has a model iff  $\phi$  has a model.*

Because of Lemma 23,  $\phi^\wedge$  is known as the *Skolem normal form of  $\phi$  for satisfiability*.

Lemma 23 is handy for simplifying various logical problems. But it would be handier still if no function symbols were involved. At the end of Section 18 we saw that anything that can be said with a function constant can also be said with a relation constant. However, in order to make the implication from right to left in (201) still hold when relations are used instead of functions, we have to require that the relations really do represent functions, in other words some sentences of form (146) must hold. These sentences are  $\forall\exists$  sentences, i.e. they have form  $\forall \vec{x}\exists \vec{y}\psi$  where  $\psi$  has no quantifiers. The upshot is that for every prenex first-order sentence  $\phi$  *without function symbols* we can effectively find an  $\forall\exists$  first-order sentence  $\phi_\wedge$  *without function symbols but with extra relation symbols*, such that  $\phi$  has a model if and only if  $\phi_\wedge$  has a model. The sentence  $\phi_\wedge$  is also known as the *Skolem normal form of  $\phi$  for satisfiability*.

For more on Skolem normal forms see [Kreisel and Krivine, 1967, Chapter 2].

Skolem also applied Theorem 21 to prove his part of the Löwenheim–Skolem Theorem 14. We say that L-structures  $\mathfrak{A}$  and  $\mathfrak{B}$  are *elementarily equivalent* to each other if exactly the same sentences of L are true in  $\mathfrak{A}$  as in  $\mathfrak{B}$ . Skolem showed:

**THEOREM 24.** *If L is a language with at most countably many formulas and  $\mathfrak{A}$  is an infinite L-structure, then by choosing countably many elements of  $\mathfrak{A}$  and throwing out the rest, we can get a countable L-structure  $\mathfrak{B}$  which is elementarily equivalent to  $\mathfrak{A}$ .*

This is proved as follows. There are countably many sentences of  $\mathcal{L}$  which are true in  $\mathfrak{A}$ . For each of these sentences  $\phi$ , player  $\exists$  has a winning strategy  $S_\phi$  for  $G(\mathfrak{A}, \phi)$ . All we need to do is find a countable set  $X$  of elements of  $\mathfrak{A}$  such that if player  $\forall$  chooses his elements from  $X$ , all the strategies  $S_\phi$  tell player  $\exists$  to pick elements which are in  $X$  too. Then  $X$  will serve as the domain of  $\mathfrak{B}$ , and player  $\exists$  will win each  $G(\mathfrak{B}, \phi)$  by playing the same strategy  $S_\phi$  as for  $G(\mathfrak{A}, \phi)$ . Starting from any countable set  $X_0$  of elements of  $\mathfrak{A}$ , let  $X_{n+1}$  be  $X_n$  together with all elements called forth by any of the strategies  $S_\phi$  when player  $\forall$  chooses from  $X_n$ ; then  $X$  can be the set of all elements which occur in  $X_n$  for at least one natural number  $n$ .

In his paper [1920], Skolem noticed that the proof of Theorem 21 gives us information in a rather broader setting too. Let  $\mathcal{L}_{\omega_1\omega}$  be the logic we get if, starting from first-order logic, we allow formulas to contain conjunctions or disjunctions of countably many formulas at a time. For example, in  $\mathcal{L}_{\omega_1\omega}$  there is an infinite sentence

$$(202) \quad \forall x(x = 0 \vee x = 1 \vee x = 2 \vee \dots)$$

which says ‘Every element is a natural number’. If we add (202) to the axioms of first-order Peano arithmetic we get a theory whose only models are the natural number system and other structures which are exact copies of it. This implies that the Compactness Theorem (Theorem 13) and the Upward Löwenheim–Skolem Theorem (Theorem 14) both fail when we replace first-order logic by  $\mathcal{L}_{\omega_1\omega}$ .

Skolem noticed that the proof of Theorem 21 tells us:

**THEOREM 25.** *If  $\phi$  is a sentence of the logic  $\mathcal{L}_{\omega_1\omega}$  and  $\mathfrak{A}$  is a model of  $\phi$ , then by choosing at most countably many elements of  $\mathfrak{A}$  we can get an at most countable structure  $\mathfrak{B}$  which is also a model of  $\phi$ .*

So a form of the Downward Löwenheim–Skolem Theorem (cf. Theorem 14) does hold in  $\mathcal{L}_{\omega_1\omega}$ .

To return for a moment to the games at the beginning of this section: Hintikka [1996] has pointed out that there is an unspoken assumption that each player is allowed to know the previous choices of the other player. (If I don’t know what person in Croydon you have produced, how can I know which dog to choose?) He has proposed that we should recast first-order logic so that this assumption need no longer hold. For example, in his notation, if  $\phi$  is the sentence

$$(203) \quad \forall x(\exists y/\forall x)x = y$$

then in the game  $G(\mathfrak{A}, \phi)$ , player  $\forall$  chooses an element  $a$  of  $\mathfrak{A}$ , then player  $\exists$  chooses an element  $b$  of  $\mathfrak{A}$  *without being told what  $a$  is*. Player  $\exists$  wins if and only if  $a = b$ . (One easily sees that if  $\mathfrak{A}$  has at least two elements, then neither player has a winning strategy for this game.) These added slash quantifiers greatly add to the expressive power of first-order logic. For

example there is now a sentence which is true in a structure  $\mathfrak{A}$  if and only if  $\mathfrak{A}$  has infinitely many elements; there is no such sentence of ordinary first-order logic. As a result, the compactness theorem fails for Hintikka's logic, and hence in turn the logic has no complete proof calculus. One can construct a Tarski-style semantics for the new logic (by a slight adaptation of [Hodges, 1997b]), but it has some very odd features. It no longer makes sense to talk of an element *satisfying* a formula; instead one has to use the notion of a set of elements *uniformly satisfying* the formula, where 'uniform' means essentially that player  $\exists$  doesn't need any forbidden information about which element within the set has been chosen. Hintikka claims, boldly, that the extended logic is in several ways more natural than the usual first-order logic.

## 26 BACK-AND-FORTH EQUIVALENCE

In this section and the next, we shall prove that certain things are definable by first-order formulas. The original versions of the theorems we prove go back to the mid 1950s. But for us their interest lies in the proofs which Per Lindström gave in [1969]. He very cleverly used the facts (1) that first-order logic is good for encoding finite sequences, and (2) that first-order logic is bad for distinguishing infinite cardinals. His proofs showed that anything we can say using a logic which shares features (1) and (2) with first-order logic can also be said with a first-order sentence; so first-order logic is essentially the only logic with these features.

I should say what we mean by a logic. A *logic*  $\mathcal{L}$  is a family of languages, one for each similarity type, together with a definition of what it is for a sentence of a language  $L$  of  $\mathcal{L}$  to be true in an  $L$ -structure. Just as in first-order logic, an  $L$ -structure is a structure which has named relations and elements corresponding to the similarity type of  $L$ . We shall always assume that the analogue of Theorem 1 holds for  $\mathcal{L}$ , i.e., that the truth-value of a sentence  $\phi$  in a structure  $\mathfrak{A}$  doesn't depend on how  $\mathfrak{A}$  interprets constants which don't occur in  $\phi$ .

We shall say that a logic  $\mathcal{L}$  is an *extension of first-order logic* if, roughly speaking, it can do everything that first-order logic can do and maybe a bit more. More precisely, it must satisfy three conditions. (i) Every first-order formula must be a formula of  $\mathcal{L}$ . (ii) If  $\phi$  and  $\psi$  are formulas of  $\mathcal{L}$  then so are  $\neg\phi$ ,  $\phi \wedge \psi$ ,  $\phi \vee \psi$ ,  $\phi \rightarrow \psi$ ,  $\phi \leftrightarrow \psi$ ,  $\forall x\phi$ ,  $\exists x\phi$ ; we assume the symbols  $\neg$  etc. keep their usual meanings. (iii)  $\mathcal{L}$  is *closed under relativisation*. This means that for every sentence  $\phi$  of  $\mathcal{L}$  and every 1-place predicate constant  $P$  not in  $\phi$ , there is a sentence  $\phi^{(P)}$  such that a structure  $\mathfrak{A}$  is a model of  $\phi^{(P)}$  if and only if the part of  $\mathfrak{A}$  with domain  $I_{\mathfrak{A}}(P)$  satisfies  $\phi$ . For example, if  $\mathcal{L}$  can say 'Two-thirds of the elements satisfy  $R(x)$ ', then it must also be able to say 'Two-thirds of the elements which satisfy  $P(x)$  satisfy  $R(x)$ '. First-order

logic itself is closed under relativisation; although I haven't called attention to it earlier, it is a device which is constantly used in applications.

The logic  $\mathcal{L}_{\omega_1\omega}$  mentioned in the previous section is a logic in the sense defined above, and it is an extension of first-order logic. Another logic which extends first-order logic is  $\mathcal{L}_{\infty\omega}$ ; this is like first-order logic except that we are allowed to form conjunctions and disjunctions of arbitrary sets of formulas, never mind how large. Russell's logic, got by adding definite description operators to first-order logic, is another extension of first-order logic though it never enables us to say anything new.

We shall always require logics to obey one more condition, which needs some definitions. L-structures  $\mathfrak{A}$  and  $\mathfrak{B}$  are said to be *isomorphic* to each other if there is a function  $F$  from the domain of  $\mathfrak{A}$  to the domain of  $\mathfrak{B}$  which is bijective, and such that for all elements  $\alpha_0, \alpha_1, \dots$ , of  $\mathfrak{A}$  and every atomic formula  $\phi$  of L,

$$(204) \quad \mathfrak{A} \models \phi[\alpha_0/x_0, \alpha_1/x_1, \dots] \text{ iff } \mathfrak{B} \models \phi[F(\alpha_0)/x_0, F(\alpha_1)/x_1, \dots].$$

It will be helpful in this section and the next if we omit the  $x_i$ 's when writing conditions like (204); so (205) means the same as (204) but is briefer:

$$(205) \quad \mathfrak{A} \models \phi[\alpha_0, \alpha_1, \dots] \text{ iff } \mathfrak{B} \models \phi[F(\alpha_0), F(\alpha_1), \dots].$$

If (204) or equivalently (205) holds, where  $F$  is a bijection from the domain of  $\mathfrak{A}$  to that of  $\mathfrak{B}$ , we say that  $F$  is an *isomorphism* from  $\mathfrak{A}$  to  $\mathfrak{B}$ . Intuitively,  $\mathfrak{A}$  is isomorphic to  $\mathfrak{B}$  when  $\mathfrak{B}$  is a perfect copy of  $\mathfrak{A}$ .

If  $\mathcal{L}$  is a logic, we say that structures  $\mathfrak{A}$  and  $\mathfrak{B}$  are  $\mathcal{L}$ -*equivalent* to each other if every sentence of  $\mathcal{L}$  which is true in one is true in the other. Thus 'elementarily equivalent' means  $\mathcal{L}$ -equivalent where  $\mathcal{L}$  is first-order logic. The further condition we impose on logics is this: structures which are isomorphic to each other must also be  $\mathcal{L}$ -equivalent to each other. Obviously this is a reasonable requirement. Any logic you think of will meet it.

Now we shall introduce another kind of game. This one is used for comparing two structures. Let  $\mathfrak{A}$  and  $\mathfrak{B}$  be L-structures. The game  $\text{EF}_\omega(\mathfrak{A}; \mathfrak{B})$  is played by two players  $\forall$  and  $\exists$  as follows. There are infinitely many moves. At the  $i$ th move, player  $\forall$  chooses one of  $\mathfrak{A}$  and  $\mathfrak{B}$  and then selects an element of the structure he has chosen; then player  $\exists$  must pick an element from the other structure. The elements chosen from  $\mathfrak{A}$  and  $\mathfrak{B}$  at the  $i$ th move are written  $\alpha_i$  and  $\beta_i$  respectively. Player  $\exists$  wins the game if and only if for every atomic formula  $\phi$  of L,

$$(206) \quad \mathfrak{A} \models \phi[\alpha_0, \alpha_1, \dots] \text{ iff } \mathfrak{B} \models \phi[\beta_0, \beta_1, \dots].$$

We say that  $\mathfrak{A}$  and  $\mathfrak{B}$  are *back-and-forth equivalent* to each other if player  $\exists$  has a winning strategy for this game.

The game  $\text{EF}_\omega(\mathfrak{A}; \mathfrak{B})$  is known as the *Ehrenfeucht–Fraïssé* game of length  $\omega$ , for reasons that will appear in the next section. One feels that the more

similar  $\mathfrak{A}$  and  $\mathfrak{B}$  are, the easier it ought to be for player  $\exists$  to win the game. The rest of this section is devoted to turning this feeling into theorems. For an easy start:

**THEOREM 26.** *If  $\mathfrak{A}$  is isomorphic to  $\mathfrak{B}$  then  $\mathfrak{A}$  is back-and-forth equivalent to  $\mathfrak{B}$ .*

Given an isomorphism  $F$  from  $\mathfrak{A}$  to  $\mathfrak{B}$ , player  $\exists$  should always choose so that for each natural number  $i$ ,  $\beta_i = F(\alpha_i)$ . Then she wins. Warning: we are talking set theory now, so  $F$  may not be describable in terms which any human player could use, even if he could last out the game.

As a partial converse to Theorem 26:

**THEOREM 27.** *If  $\mathfrak{A}$  is back-and-forth equivalent to  $\mathfrak{B}$  and both  $\mathfrak{A}$  and  $\mathfrak{B}$  have at most countably many elements, then  $\mathfrak{A}$  is isomorphic to  $\mathfrak{B}$ .*

For this, imagine that player  $\forall$  chooses his moves so that he picks each element of  $\mathfrak{A}$  or  $\mathfrak{B}$  at least once during the game; he can do this if both structures are countable. Let player  $\exists$  use her winning strategy. When all the  $\alpha_i$ 's and  $\beta_i$ 's have been picked, define  $F$  by putting  $F(\alpha_i) = \beta_i$  for each  $i$ . (The definition is possible because (206) holds for each atomic formula ' $x_i = x_j$ '.) Comparing (205) with (206), we see that  $F$  is an isomorphism. The idea of this proof was first stated by Huntington [1904] and Hausdorff [1914, p. 99] in proofs of a theorem of Cantor about dense linear orderings. Fraïssé [1954] noticed that the argument works just as well for structures as for orderings.

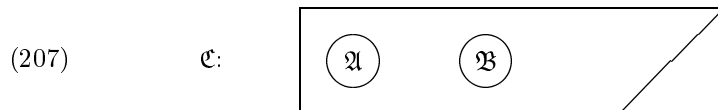
Now we are going to show that whether or not  $\mathfrak{A}$  and  $\mathfrak{B}$  have countably many elements, if  $\mathfrak{A}$  and  $\mathfrak{B}$  are back-and-forth equivalent then they are elementarily equivalent. This was known to Fraïssé [1955], and Karp [1965] gave a direct proof of the stronger result that  $\mathfrak{A}$  is back-and-forth equivalent to  $\mathfrak{B}$  if and only if  $\mathfrak{A}$  is  $\mathcal{L}_{\infty\omega}$ -equivalent to  $\mathfrak{B}$ . The interest of our proof (which was extracted from Lindström [1969] by Barwise [1974]) is that it works for any extension of first-order logic which obeys the Downward Löwenheim–Skolem Theorem. To be precise:

**THEOREM 28.** *Suppose  $\mathcal{L}$  is an extension of first-order logic, and every structure of at most countable similarity type is  $\mathcal{L}$ -equivalent to a structure with at most countably many elements. Suppose also that every sentence of  $\mathcal{L}$  has at most countably many distinct symbols. Then any two structures which are back-and-forth equivalent are  $\mathcal{L}$ -equivalent to each other.*

Theorem 28 can be used to prove Karp's result too, by a piece of set-theoretic strong-arm tactics called 'collapsing cardinals' (as in [Barwise, 1973]). By Skolem's observation (Theorem 25), Theorem 28 applies almost directly to  $\mathcal{L}_{\omega_1\omega}$  (though one still has to use 'countable fragments' of  $\mathcal{L}_{\omega_1\omega}$ —I omit details).

Let me sketch the proof of Theorem 28. Assume all the assumptions of Theorem 28, and let  $\mathfrak{A}$  and  $\mathfrak{B}$  be L-structures which are back-and-forth

equivalent. We have to show that  $\mathfrak{A}$  and  $\mathfrak{B}$  are  $\mathcal{L}$ -equivalent. Replacing  $\mathfrak{B}$  by an isomorphic copy if necessary, we can assume that  $\mathfrak{A}$  and  $\mathfrak{B}$  have no elements in common. Now we construct a jumbo structure:



The language of  $\mathfrak{C}$  shall contain two 1-place predicate constants  $\partial^{\mathfrak{A}}$  and  $\partial^{\mathfrak{B}}$ . Also for each predicate constant  $R$  and individual constant  $c$  of  $\mathcal{L}$  the language of  $\mathfrak{C}$  shall contain two symbols  $R^{\mathfrak{A}}, R^{\mathfrak{B}}$  and  $c^{\mathfrak{A}}, c^{\mathfrak{B}}$ . The elements in  $I_{\mathfrak{C}}(\partial^{\mathfrak{A}})$  are precisely the elements of  $\mathfrak{A}$ , and each  $I_{\mathfrak{C}}(R^{\mathfrak{A}})$  and  $I_{\mathfrak{C}}(c^{\mathfrak{A}})$  is to be identical with  $I_{\mathfrak{A}}(R)$  and  $I_{\mathfrak{A}}(c)$  respectively. Thus  $\mathfrak{C}$  contains an exact copy of  $\mathfrak{A}$ . Likewise with  $\mathfrak{B}$  in place of  $\mathfrak{A}$ . The remaining pieces of  $\mathfrak{C}$  outside  $\mathfrak{A}$  and  $\mathfrak{B}$  consist of enough set-theoretic apparatus to code up all finite sequences of elements of  $\mathfrak{A}$  and  $\mathfrak{B}$ . Finally the language of  $\mathfrak{C}$  shall have a 2-place predicate constant  $S$  which encodes the winning strategy of player  $\exists$  in the game  $\text{EF}_{\omega}(\mathfrak{A}; \mathfrak{B})$  as follows:

(208)  $I_{\mathfrak{C}}(S)$  contains exactly those ordered pairs  $\langle\langle\gamma_0, \dots, \gamma_{n-1}\rangle, \gamma_n\rangle$  such that  $\gamma_n$  is the element which player  $\exists$ 's winning strategy tells her to play if player  $\forall$ 's previous moves were  $\gamma_0, \dots, \gamma_{n-1}$ .

Now we wish to show that any sentence  $\mathcal{L}$  which is true in  $\mathfrak{A}$  is true also in  $\mathfrak{B}$ , and *vice versa*. Since each sentence of  $\mathcal{L}$  contains at most countably many symbols, we can assume without any loss of generality that the similarity type of  $\mathfrak{A}$  and  $\mathfrak{B}$  has just countably many symbols; hence the same is true for  $\mathfrak{C}$ , and thus by the assumption in Theorem 28,  $\mathfrak{C}$  is  $\mathcal{L}$ -equivalent to a structure  $\mathfrak{C}'$  with at most countably many elements. The sets  $I_{\mathfrak{C}'}(\partial^{\mathfrak{A}})$  and  $I_{\mathfrak{C}'}(\partial^{\mathfrak{B}})$  of  $\mathfrak{C}'$  define  $\mathcal{L}$ -structures  $\mathfrak{A}'$  and  $\mathfrak{B}'$  which are  $\mathcal{L}$ -equivalent to  $\mathfrak{A}$  and  $\mathfrak{B}$  respectively, since everything we say in  $\mathcal{L}$  about  $\mathfrak{A}$  can be rewritten as a statement about  $\mathfrak{C}$  using  $\partial^{\mathfrak{A}}$  and the  $R^{\mathfrak{A}}$  and  $c^{\mathfrak{A}}$ . (Here we use the fact that  $\mathcal{L}$  allows relativisation.)

Since  $\mathcal{L}$  contains all first-order logic, everything that we can say in a first-order language about  $\mathfrak{C}$  must also be true in  $\mathfrak{C}'$ . For example we can say in first-order sentences that for every finite sequence  $\gamma_0, \dots, \gamma_{n-1}$  of elements of  $\mathfrak{A}$  or  $\mathfrak{B}$  there is a unique element  $\gamma_n$  such that  $\langle\langle\gamma_0, \dots, \gamma_{n-1}\rangle, \gamma_n\rangle$  is in  $I_{\mathfrak{C}}(S)$ ; also that if player  $\exists$  in  $\text{EF}_{\omega}(\mathfrak{A}; \mathfrak{B})$  reads  $I_{\mathfrak{C}}(S)$  as a strategy for her, then she wins. So all these things must be true also for  $\mathfrak{A}', \mathfrak{B}'$  and  $I_{\mathfrak{C}'}(S)$ . (The reader can profitably check for himself that all this can be coded into first-order sentences, but if he gets stuck he can consult [Barwise, 1974] or [Flum, 1975].)

Therefore  $\mathfrak{A}'$  is back-and-forth equivalent to  $\mathfrak{B}'$ . But both  $\mathfrak{A}'$  and  $\mathfrak{B}'$  are bits of  $\mathfrak{C}'$ , so they have at most countably many elements. Hence by Theorem 27,  $\mathfrak{A}'$  is isomorphic to  $\mathfrak{B}'$  and therefore  $\mathfrak{A}'$  is  $\mathcal{L}$ -equivalent to  $\mathfrak{B}'$ .

But  $\mathfrak{A}'$  was  $\mathcal{L}$ -equivalent to  $\mathfrak{A}$  and  $\mathfrak{B}'$  was  $\mathcal{L}$ -equivalent to  $\mathfrak{B}$ . So finally we deduce that  $\mathfrak{A}$  and  $\mathfrak{B}$  are  $\mathcal{L}$ -equivalent.

In our definition of logics, we allowed the formulas to include some items that go beyond first-order logic, but we made no change in the class of  $L$ -structures. The methods of this section, and many of those of the next section too (in particular Theorem 29), still work if one restricts attention to finite structures. Ebbinghaus and Flum [1995] explore the implications of this fact, with an eye on complexity theory.

## 27 LINDSTRÖM'S THEOREM

Theorem 28 showed that any extension of first-order logic which obeys a form of the Downward Löwenheim–Skolem Theorem is in a sense no stronger than the infinitary logic  $\mathcal{L}_{\infty\omega}$ . This result is relatively shallow and not terribly useful; the logic  $\mathcal{L}_{\infty\omega}$  is quite powerful and not very well understood. (See Van Benthem and Doets [this Volume].) Lindström [1969] found a stronger and more subtle result: he showed that if in addition  $\mathcal{L}$  obeys a form of the Compactness Theorem or the Upward Löwenheim–Skolem Theorem then every sentence of  $\mathcal{L}$  has exactly the same models as some first-order sentence. Since a first-order sentence contains only finitely many symbols, this result evidently needs some finiteness restriction on the sentences of  $\mathcal{L}$ . So from now on we shall assume that *all similarity types are finite and have no function symbols*.

Lindström's argument relies on some detailed information about Ehrenfeucht–Fraïssé games. The Ehrenfeucht–Fraïssé game  $\text{EF}_n(\mathfrak{A}; \mathfrak{B})$  of length  $n$ , where  $n$  is a natural number, is fought and won exactly like  $\text{EF}_\omega(\mathfrak{A}; \mathfrak{B})$  except that the players stop after  $n$  moves. We say that the structures  $\mathfrak{A}$  and  $\mathfrak{B}$  are  *$n$ -equivalent* if player  $\exists$  has a winning strategy for the game  $\text{EF}_n(\mathfrak{A}; \mathfrak{B})$ . If  $\mathfrak{A}$  and  $\mathfrak{B}$  are back-and-forth equivalent then they are  $n$ -equivalent for all  $n$ ; the converse is not true.

Ehrenfeucht–Fraïssé games of finite length were invented by Ehrenfeucht [1960] as a means of showing that two structures are elementarily equivalent. He showed that if two structures  $\mathfrak{A}$  and  $\mathfrak{B}$  are  $n$ -equivalent for all finite  $n$  then  $\mathfrak{A}$  and  $\mathfrak{B}$  are elementarily equivalent (which follows easily from Theorem 28), and that if the similarity type is finite and contains no function symbols, then the converse holds too. Fraïssé's definitions were different, but in his [1955] he proved close analogues of Ehrenfeucht's theorems, including an analogue of the following:

**THEOREM 29.** *Let  $L$  be a first-order language. Then for every natural number  $n$  there is a finite set of sentences  $\sigma_{n,1}, \dots, \sigma_{n,j_n}$  of  $L$  such that:*

1. *every  $L$ -structure  $\mathfrak{A}$  is a model of exactly one of  $\sigma_{n,1}, \dots, \sigma_{n,j_n}$ ; if  $\mathfrak{A} \models \sigma_{n,i}$  we say that  $\mathfrak{A}$  has  $n$ -type  $\sigma_{n,i}$ ;*

2. L-structures  $\mathfrak{A}$  and  $\mathfrak{B}$  are  $n$ -equivalent iff they have the same  $n$ -type.

Theorem 29 is best proved by defining a more complicated game. Suppose  $\gamma_0, \dots, \gamma_{k-1}$  are elements of  $\mathfrak{A}$  and  $\delta_0, \dots, \delta_{k-1}$  are elements of  $\mathfrak{B}$ . Then the game  $\text{EF}_n(\mathfrak{A}, \gamma_0, \dots, \gamma_{k-1}; \mathfrak{B}, \delta_0, \dots, \delta_{k-1})$  shall be played exactly like  $\text{EF}_n(\mathfrak{A}; \mathfrak{B})$ , but at the end when elements  $\alpha_0, \dots, \alpha_{n-1}$  of  $\mathfrak{A}$  and  $\beta_0, \dots, \beta_{n-1}$  of  $\mathfrak{B}$  have been chosen, player  $\exists$  wins if and only if for every atomic formula  $\phi$ ,

$$(209) \quad \begin{aligned} \mathfrak{A} \models \phi[\gamma_0, \dots, \gamma_{k-1}, \alpha_0, \dots, \alpha_{n-1}] \\ \text{iff } \mathfrak{B} \models \phi[\delta_0, \dots, \delta_{k-1}, \beta_0, \dots, \beta_{n-1}]. \end{aligned}$$

So this game is harder for player  $\exists$  to win than  $\text{EF}_n(\mathfrak{A}; \mathfrak{B})$  was. We say that  $\langle \mathfrak{A}, \gamma_0, \dots, \gamma_{k-1} \rangle$  is  $n$ -equivalent to  $\langle \mathfrak{B}, \delta_0, \dots, \delta_{k-1} \rangle$  if player  $\exists$  has a winning strategy for the game  $\text{EF}_n(\mathfrak{A}, \gamma_0, \dots, \gamma_{k-1}; \mathfrak{B}, \delta_0, \dots, \delta_{k-1})$ . We assert that for each finite  $k$  and  $n$  there is a finite set of formulas  $\sigma_{n,1}^k, \sigma_{n,2}^k$  etc. of L such that

1. for every L-structure  $\mathfrak{A}$  and elements  $\gamma_0, \dots, \gamma_{k-1}$  of  $\mathfrak{A}$  there is a unique  $i$  such that  $\mathfrak{A} \models \sigma_{n,i}^k[\gamma_0, \dots, \gamma_{k-1}]$ ; this  $\sigma_{n,i}^k$  is called the  $n$ -type of  $\langle \mathfrak{A}, \gamma_0, \dots, \gamma_{k-1} \rangle$ ;
2.  $\langle \mathfrak{A}, \gamma_0, \dots, \gamma_{k-1} \rangle$  and  $\langle \mathfrak{B}, \delta_0, \dots, \delta_{k-1} \rangle$  are  $n$ -equivalent iff they have the same  $n$ -type.

Theorem 29 will then follow by taking  $k$  to be 0. We prove the assertion above for each  $k$  by induction on  $n$ .

When  $n = 0$ , for each  $k$  there are just finitely many sequences  $\langle \mathfrak{A}, \gamma_0, \dots, \gamma_{k-1} \rangle$  which can be distinguished by atomic formulas. (Here we use the fact that the similarity type is finite and there are no function symbols.) So we can write down finitely many formulas  $\sigma_{0,1}^k, \sigma_{0,2}^k$  etc. which distinguish all the sequences that can be distinguished.

When the formulas have been constructed and (1), (2) proved for the number  $n$ , we construct and prove them for  $n + 1$  as follows. Player  $\exists$  has a winning strategy for  $\text{EF}_{n+1}(\mathfrak{A}, \gamma_0, \dots, \gamma_{k-1}; \mathfrak{B}, \delta_0, \dots, \delta_{k-1})$  if and only if she can make her first move so that she has a winning strategy from that point onwards, i.e. if she can ensure that  $\alpha_0$  and  $\beta_0$  are picked so that

$$\langle \mathfrak{A}, \gamma_0, \dots, \gamma_{k-1}, \alpha_0 \rangle \text{ is } n\text{-equivalent to } \langle \mathfrak{B}, \delta_0, \dots, \delta_{k-1}, \beta_0 \rangle.$$

In other words, using (2) for  $n$  which we assume has already been proved, player  $\exists$  has this winning strategy if and only if for every element  $\alpha$  of  $\mathfrak{A}$  there is an element  $\beta$  of  $\mathfrak{B}$  so that

$$\langle \mathfrak{A}, \gamma_0, \dots, \gamma_{k-1}, \alpha \rangle \text{ has the same } n\text{-type as } \langle \mathfrak{B}, \delta_0, \dots, \delta_{k-1}, \beta \rangle,$$



and *vice versa* with  $\mathfrak{A}$  and  $\mathfrak{B}$  reversed. But this is equivalent to the condition:

$$\text{for every } i, \\ \mathfrak{A} \models \exists x_k \sigma_{n,i}^{k+1}[\gamma_0, \dots, \gamma_{k-1}] \text{ iff } \mathfrak{B} \models \exists x_k \sigma_{n,i}^{k+1}[\delta_0, \dots, \delta_{k-1}].$$

It follows that we can build suitable formulas  $\sigma_{n+1,i}^k$  by taking conjunctions of formulas of form  $\exists x_k \sigma_{n,i}^{k+1}$  or  $\neg \exists x_k \sigma_{n,i}^{k+1}$ , running through all the possibilities.

When the formulas  $\sigma_{n,i}^k$  have all been defined, we take  $\sigma_{n,i}$  to be  $\sigma_{n,i}^0$ . Thus Theorem 29 is proved.

Barwise [1975, Chapter VII.6] describes the formulas  $\sigma_{n,i}^k$  in detail in a rather more general setting. The sentences  $\sigma_{n,i}$  were first described by Hintikka [1953] (cf. also [Hintikka, 1973, Chapter XI]), but their meaning was mysterious until Ehrenfeucht's paper appeared. We shall call the sentences *Hintikka sentences*. Hintikka proved that every first-order sentence is logically equivalent to a (finite) disjunction of Hintikka sentences. We shall prove this too, but by Lindström's proof [1969] which assumes only some general facts about the expressive power of first-order logic; so the proof will show that any sentence in any logic with this expressive power has the same models as some first-order sentence, viz. a disjunction of Hintikka sentences. Lindström proved:

**THEOREM 30.** *Let  $\mathcal{L}$  be any extension of first-order logic with the two properties:*

- (a) *(Downward Löwenheim–Skolem) If a sentence  $\phi$  of  $\mathcal{L}$  has an infinite model then  $\phi$  has a model with at most countably many elements.*
- (b) *Either (Upward Löwenheim–Skolem) if a sentence of  $\mathcal{L}$  has an infinite model then it has one with uncountably many elements; or (Compactness) if  $\Delta$  is a theory in  $\mathcal{L}$  such that every finite set of sentences from  $\Delta$  has a model then  $\Delta$  has a model.*

*Then every sentence of  $\mathcal{L}$  has exactly the same models as some first-order sentence.*

The proof is by the same kind of coding as the proof of Theorem 28. Instead of proving Theorem 30 directly, we shall show:

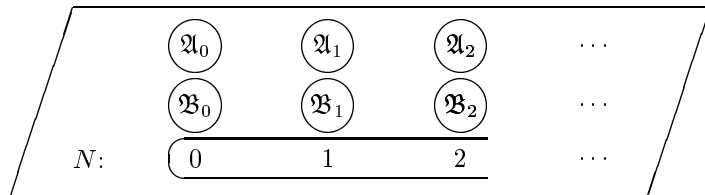
**THEOREM 31.** *Let  $\mathcal{L}$  be any extension of first-order logic obeying (a) and (b) as in Theorem 30, and let  $\phi$  and  $\psi$  be sentences of  $\mathcal{L}$  such that no model of  $\phi$  is also a model of  $\psi$ . Then for some integer  $n$  there is a disjunction  $\sigma$  of Hintikka sentences  $\sigma_{n,i}$  such that  $\phi \models \sigma$  and  $\psi \models \neg \sigma$ .*

To get Theorem 30 from Theorem 31, let  $\psi$  be  $\neg \phi$ .

Suppose then that Theorem 31 is false. This means that there exist sentences  $\phi$  and  $\psi$  of  $\mathcal{L}$  with no models in common, and for every natural

number  $n$  there is no disjunction of Hintikka sentences  $\sigma_{n,i}$  which separates the models of  $\phi$  from the models of  $\psi$ . So by Theorem 29 there are, for each  $n$ ,  $n$ -equivalent structures  $\mathfrak{A}_n$  and  $\mathfrak{B}_n$  such that  $\mathfrak{A}_n$  is a model of  $\phi$  and  $\mathfrak{B}_n$  is a model of  $\psi$ . By (a) we can assume that  $\mathfrak{A}_n$  and  $\mathfrak{B}_n$  have at most countably many elements (since the sentences  $\sigma_{n,i} \wedge \phi$  and  $\sigma_{n,i} \wedge \psi$  are both in  $\mathcal{L}$ ).

So now once again we build a mammoth model  $\mathfrak{C}$ :



The coding is more complicated this time.  $\mathfrak{C}$  contains a copy of the natural numbers  $N$ , picked out by a predicate constant  $\partial^N$ . There are 2-place predicate constants  $\partial^{\mathfrak{A}}, \partial^{\mathfrak{B}}$ .  $I_{\mathfrak{C}}(\partial^{\mathfrak{A}})$  contains just those pairs  $\langle \alpha, n \rangle$  such that  $n$  is a natural number and  $\alpha$  is an element of  $\mathfrak{A}_n$ . Similarly with the  $\mathfrak{B}_n$ . Also  $\mathfrak{C}$  has constants which describe each  $\mathfrak{A}_n$  and  $\mathfrak{B}_n$  completely, and  $\mathfrak{C}$  contains all finite sequences of elements taken from any  $\mathfrak{A}_n$  or  $\mathfrak{B}_n$ , together with enough set theory to describe lengths of sequences etc. There is a relation  $I_{\mathfrak{C}}(S)$  which encodes the winning strategies for player  $\exists$  in all games  $\text{EF}_n(\mathfrak{A}_n, \mathfrak{B}_n)$ . Finally  $\mathfrak{C}$  can be assumed to have just countably many elements, so we can incorporate a relation which sets up a bijection between  $N$  and the whole of the domain of  $\mathfrak{C}$ .

We shall need the fact that everything salient about  $\mathfrak{C}$  can be said in *one single sentence*  $\chi$  of  $\mathcal{L}$ . Since  $N$  is in  $\mathfrak{C}$  and we can build in as much set-theoretic equipment as we please, this is no problem, bearing in mind that  $\mathcal{L}$  is an extension of first-order logic. Barwise [1974] and Flum [1975] give details.

Now by (b), the sentence  $\chi$  has a model  $\mathfrak{C}'$  in which some ‘infinite’ number  $\infty$  comes after all the ‘natural numbers’  $I_{\mathfrak{C}'}(0), I_{\mathfrak{C}'}(1), I_{\mathfrak{C}'}(2), \dots$  in  $I_{\mathfrak{C}'}(\partial^N)$ . If the Upward Löwenheim–Skolem property holds, then this is because the  $N$ -part of any uncountable model of  $\chi$  must have the same cardinality as the whole model, in view of the bijection which we incorporated. If on the other hand the Compactness property holds, we follow the construction of non-standard models in Section 20 above.

By means of  $I_{\mathfrak{C}'}(\partial^{\mathfrak{A}})$  and  $I_{\mathfrak{C}'}(\partial^{\mathfrak{B}})$ , the structure  $\mathfrak{C}'$  encodes structures  $\mathfrak{A}'_{\infty}$  and  $\mathfrak{B}'_{\infty}$ , and  $I_{\mathfrak{C}'}(S)$  encodes a winning strategy for player  $\exists$  in the game  $\text{EF}_{\infty}(\mathfrak{A}'_{\infty}; \mathfrak{B}'_{\infty})$ . All this is implied by a suitable choice of  $\chi$ . The game  $\text{EF}_{\infty}(\mathfrak{A}'_{\infty}; \mathfrak{B}'_{\infty})$  turns out to be bizarre and quite unplayable; but the important point is that if player  $\exists$  has a winning strategy for this game,

then she has one for the shorter and entirely playable game  $\text{EF}_\omega(\mathfrak{A}'_\infty; \mathfrak{B}'_\infty)$ . Hence  $\mathfrak{A}'_\infty$  and  $\mathfrak{B}'_\infty$  are back-and-forth equivalent.

But now  $\chi$  records that all the structures encoded by  $\partial^{\mathfrak{A}}$  are models of  $\phi$ , while those encoded by  $\partial^{\mathfrak{B}}$  are models of  $\psi$ . Hence  $\mathfrak{A}'_\infty \models \phi$  but  $\mathfrak{B}'_\infty \not\models \psi$ . Since  $\phi$  and  $\psi$  have no models in common, it follows that  $\mathfrak{B}'_\infty \models \neg\phi$ . The final step is to use assumption (a), the Downward Löwenheim–Skolem property, to prove a slightly sharpened version of Theorem 28. To be precise, since  $\mathfrak{A}'_\infty$  and  $\mathfrak{B}'_\infty$  are back-and-forth equivalent and  $\mathfrak{A}'_\infty$  is a model of the sentence  $\phi$  of  $\mathfrak{L}$ ,  $\mathfrak{B}'_\infty$  must also be a model of  $\phi$ . (The proof is like that in Section 26, but we use the fact that the similarity type is finite and has no function symbols in order to boil down the essential properties of  $\mathfrak{C}$  into a single sentence.) So we have reached a contradiction, and Theorem 31 is proved.

The proof of Theorem 31, less the last paragraph, adapts to give a proof of *Craig’s Interpolation Lemma* for predicate logic:

**LEMMA 32.** *Let  $\phi$  and  $\psi$  be sentences of first-order predicate logic such that  $\phi \models \neg\psi$ . Then there is a first-order sentence  $\sigma$  such that  $\phi \models \sigma$ ,  $\psi \models \neg\sigma$ , and every constant symbol which occurs in  $\sigma$  occurs both in  $\phi$  and  $\psi$ .*

Let  $\mathfrak{L}$  in the proof of Theorem 31 be first-order logic and let  $L$  be the first-order language whose constants are those which occur both in  $\phi$  and in  $\psi$ . Using Section 18, we can assume that  $L$  has no function symbols. If  $\mathfrak{A}$  is any model of  $\phi$ , then we get an  $L$ -structure  $\mathfrak{A}|L$  by discarding all constant symbols not in  $L$ , without changing the elements or the interpretations of the symbols which are in  $L$ . Likewise for every model  $\mathfrak{B}$  of  $\psi$ . Now suppose that the conclusion of Lemma 32 fails. Then for each natural number  $n$  there is no disjunction  $\sigma$  of Hintikka sentences  $\sigma_{n,i}$  in the language  $L$  such that  $\phi \models \sigma$  and  $\psi \models \neg\sigma$ , and hence there are models  $\mathfrak{A}_n, \mathfrak{B}_n$  of  $\phi, \psi$  respectively, such that  $\mathfrak{A}_n|L$  is  $n$ -equivalent to  $\mathfrak{B}_n|L$ . Proceed now as in the proof of Theorem 31, using the Compactness and Downward Löwenheim–Skolem Theorems to find a countable  $\mathfrak{C}'$  with an infinite natural number  $\infty$ . Excavate models  $\mathfrak{A}'_\infty, \mathfrak{B}'_\infty$  of  $\phi, \psi$  from  $\mathfrak{C}'$  as before, noting this time that  $\mathfrak{A}'_\infty|L$  is back-and-forth equivalent to  $\mathfrak{B}'_\infty|L$ . Then by Theorem 27, since  $\mathfrak{A}'_\infty|L$  and  $\mathfrak{B}'_\infty|L$  are countable and back-and-forth equivalent, they are isomorphic. It follows that we can add to  $\mathfrak{A}'_\infty$  interpretations of those symbols which are in  $\psi$  but not in  $L$ , using  $\mathfrak{B}'_\infty$  as a template. Let  $\mathfrak{D}$  be the resulting structure. Then  $\mathfrak{D} \models \phi$  since  $\mathfrak{A}'_\infty \models \phi$ , and  $\mathfrak{D} \models \psi$  since  $\mathfrak{B}'_\infty \models \psi$ . This contradicts the assumption that  $\phi \models \neg\psi$ . Hence Lemma 32 is proved.

Craig himself [1957b] used his interpolation result to give a proof of *Beth’s Definability Theorem* [Beth, 1953]:

**THEOREM 33.** *Let  $L$  be a first-order language and  $\Delta$  a first-order theory which uses the language  $L$  together with one extra  $n$ -place predicate constant  $R$ . Suppose that for every  $L$ -structure  $\mathfrak{A}$  there is at most one way of adding to  $\mathfrak{A}$  an interpretation of  $R$  so that the resulting structure is a model of  $\Delta$ .*

Then  $\Delta$  has a consequence of form  $\forall x_1, \dots, x_n (R(x_1, \dots, x_n) \leftrightarrow \phi)$ , where  $\phi$  is a formula in the language  $L$ .

Time's wingèd chariot prevents a proper discussion of implicit and explicit definitions here, but Beth's theorem is proved in Section 5.5 of [Hodges, 1997a], and Section 2.2 of Chang and Keisler [1973]. There is some useful background on implicit definitions in [Suppes, 1957, Chapter 8]. Craig's and Beth's results have interested philosophers of science; see e.g. [Sneed, 1971].

## 28 LAWS OF THOUGHT?

This section is new in the second edition. I am not sure that it belongs at Section 28, but this was the simplest place to add it.

Frege fought many battles against the enemies of sound reason. One battle which engaged some of his best energies was that against *psychologism*. Psychologism, put briefly, was the view that the proper definitions of logical notions (such as validity) make essential reference to the contents of minds. Today psychologism in first-order logic is a dead duck; not necessarily because Frege convinced anybody, but simply because there is no room for any mention of minds in the agreed definitions of the subject. The question whether the sequent

$$p \wedge q \vdash p$$

is valid has nothing more to do with minds than it has to do with the virginity of Artemis or the war in Indonesia.

Still, psychology fights back. The next generation has to learn the subject—and so we find ourselves asking: How does one teach logic? How does one learn it? How far do people think logically anyway, without benefit of logic texts? and what are the mental mechanisms involved?

During the 1980s a number of computer programs for teaching elementary logic came onto the market. Generally they would give the student a sequent and allow him or her to build a formal proof on the screen; then they would check it for correctness. Sometimes they would offer hints on possible ways to find a proof. One can still find such programs today, but mostly they are high-tech practical aids for working computer scientists, and they work in higher-order logic as happily as in first-order. (There is a review of teaching packages in [Goldson, Reeves and Bornat, 1993].) To a great extent the introductory teaching packages were driven out by a better program, *Tarski's World*. This was a sophisticated stand-alone Macintosh program put on the market in 1986 by a team of logicians and computer scientists at Stanford University led by Jon Barwise and John Etchemendy [1991].

Tarski's World teaches the notation of first-order logic, by means of the Hintikka games which we studied in Section 25 above. The student sees on the screen a formal sentence, together with a 'world' which consists of a checker board with various objects on it, some labelled with constant symbols. The predicate symbols in the sentence all have fixed meanings such as ' $x$  is a tetrahedron' or ' $x$  is between  $y$  and  $z$ '. The student is invited to guess whether the given world makes the sentence true or false, and to defend the guess by playing a game against the machine. (A little later but independently, a group in Omsk produced a similar package for teaching logic to students in Siberia. The Russian version didn't use the notion of games, and its 'worlds' consisted of graphs.)

As it stands, *Tarski's World* is no use for learning about logical consequence: in the first place it contains no proof theory, and in the second place the geometrical interpretations of the predicate symbols are built into the program, so that there is no possibility of constructing counterexamples in general—even small ones. Barwise and Etchemendy found an innovative way to plug the gap. Their next computer package, *Hyperproof* [Barwise and Etchemendy, 1994], consists of a natural deduction theorem prover for first-order logic, together with a device that allows students to represent facts pictorially rather than by sentences. Thus the picture for ' $a$  is a small tetrahedron' is a small tetrahedron labelled  $a$ . The picture for ' $a$  is small' is subtler: we have to represent  $a$  without showing what shape it is, so the picture is a small paper bag labelled  $a$ . There are devices for reading off sentences from pictures, and for adjusting pictures to fit stated sentences. Proofs are allowed to contain both sentences and pictures.

The language is limited to a small number of predicates with fixed meanings: ' $x$  is between  $y$  and  $z$ ', ' $x$  likes  $y$ ' and a few others. The student is allowed (in fact encouraged) to use geometrical knowledge about the properties of betweenness and the shape of the picture frame. As this suggests, the package aims to teach the students to reason, rather than teaching them logical theory. (On pictorial reasoning in first-order logic, see [Hammer, 1995] and his references.)

There has already been some research on how good *Hyperproof* is at teaching students to reason, compared with more 'syntactic' logic courses.

Stenning, Cox and Oberlander [1995] found that one can divide students into two groups—which they call DetHi and DetLo—in terms of their performance on reasoning tests before they take a logic course. DetHi students benefit from *Hyperproof*, whereas a syntactic logic course tends if anything to make them less able to reason about positions of blocks in space. For this spatial reasoning, DetLo students gain more advantage from a syntactic course than from *Hyperproof*. Different patterns emerge on other measures of reasoning skill. Stenning *et al.* comment:

... the evidence presented here already indicates both that dif-

ferent teaching methods can induce opposite effects in different groups of students, and that the *same* teaching method administered in a strictly controlled computerised environment using the same examples, and the same advice can induce different groups of students to develop quite distinct reasoning styles.

We need replications and extensions of this research, not least because there are several ways in which logic courses can differ. *Hyperproof* is more pictorial than any other logic course that I know. But it also belongs with those courses that give equal weight to deduction and consistency, using both proofs and counterexamples; this is a different dimension, and Stenning *et al.* suggest that it might account for some of their findings. Another feature is that students using computer logic programs get immediate feedback from the computer, unlike students learning in a class from a textbook.

These findings are a good peg to hang several other questions on. First, do classes in first-order logic really help students to do anything except first-order logic? Before the days of the Trade Descriptions Act, one early twentieth-century textbook of syllogisms advertised them as a cure for blushing and stammering. (I quote from memory; the book has long since disappeared from libraries.) Psychological experimenters have usually been much more pessimistic, claiming that there is very little transfer of skills from logic courses to any other kind of reasoning. For example Nisbett, Fong, Lehman and Cheng [1987] found that if you want to improve a student's logical skills (as measured by the Wason selection task mentioned below—admittedly a narrow and untypical test), you should teach her two years of law, medicine or psychology; a standard undergraduate course in logic is completely ineffectual. On the other hand Stenning *et al.* [1995] found that a logic course gave an average overall improvement of about 12% on the Analytical Reasoning score in the US Graduate Record Exam (I thank Keith Stenning for this figure). Their results suggest that the improvement may vary sharply with the kind of logic course, the kind of student and the kind of test.

Second, what is the brute native competence in first-order reasoning of a person with average intelligence and education but no specific training in logic? One of the most thorough-going attempts to answer this question is the work of Lance Rips [1994]. Rips writes a theorem-proving program called PSYCOP, which is designed to have more or less the same proficiency in first-order reasoning as the man on the Clapham omnibus. He defends it with a large amount of empirical evidence. A typical example of a piece of reasoning which is beyond PSYCOP is:

NOT (IF Calvin passes history THEN Calvin will graduate).  
*Therefore* Calvin passes history.

One has to say straight away that the man on the Clapham omnibus has never seen the basic symbols of first-order logic, and there could be a great

deal of slippage in the translation between first-order formalism and the words used in the experiments. In Rips' work there certainly is some slippage. For example he regards  $\forall x\exists y\neg\phi(x,y)$  as the same sentence as  $\neg\exists x\forall y\phi(x,y)$ , which makes it impossible for him to ask whether people are successful in deducing one from the other—even though the two forms suggest quite different sentences of English.

It might seem shocking that there are simple first-order inferences which the average person can't make. One suspects that this must be a misdescription of the facts. Anybody who does suspect as much should look at the astonishing 'selection task' experiment of P. C. Wason [1966], who showed that in broad daylight, with no tricks and no race against a clock, average subjects can reliably and repeatedly be brought to make horrendous mistakes of truth-table reasoning. This experiment has generated a huge amount of work, testing various hypotheses about what causes these mistakes; see [Manktelow and Over, 1990].

Third, what are the mental mechanisms that an untrained person uses in making logical deductions? Credit for raising this as an experimental issue goes to P. N. Johnson-Laird, who with his various collaborators has put together a considerable body of empirical facts (summarised in Johnson-Laird and Byrne [1991], see also the critiques in *Behavioral and Brain Sciences*, **16**, 323–380, 1993). Unfortunately it is hard for an outsider to see what thesis Johnson-Laird is aiming to prove with these facts. He uses some of the jargon of logical theory to set up a dichotomy between rule-based reasoning and model-based reasoning, and he claims that his evidence supports the latter against the former. But for anybody who comes to it from the side of logical theory, Johnson-Laird's dichotomy is a nonsense. If it has any meaning at all, it can only be an operational one in terms of the computer simulation which he offers, and I hope the reader can make more sense of that than I could. Perhaps two things emerge clearly. The first is that what he calls model-based reasoning is meta-level—it is reasoning about reasoning; which leaves us asking what his theory of object-level reasoning can be. The second claim to emerge from the mist is that we regularly use a form of proof-by-cases, and the main cause of making deductions that we shouldn't have done is that we fail to list all the necessary cases. This is an interesting suggestion, but I was unable to see how the theory explains the cases where we fail to make deductions that we should have done.

It would be a pity to end on a negative note. This section has shown, I hope, that at the end of the millenium first-order logic is still full of surprises for the old hands and new opportunities for young researchers.

## ACKNOWLEDGEMENTS

I thank the many people who commented on the first version of this chapter, and especially Franz Guenther, Hans Kamp and Dirk van Dalen for their detailed improvements and corrections. I am also very much in debt to Jan and Emilia Mycielski who allowed me to type the chapter in their idyllic house in the foothills of the Rockies.

Finally I thank Keith Stenning for his help with the new Section 28 in the second edition, with the usual caution that he is not to be held responsible for any of the opinions expressed there (except those quoted from a paper of his).

*Queen Mary and Westfield College, London.*

## IV: Appendices

These three appendices will show in outline how one can construct a formal calculus of set theory, which in some sense formalises the whole of mathematics. I have put this material into appendices, first because it is turgid, and second because I should hate to resuscitate the dreadful notion that the business of logicians is to produce all-embracing formal systems.

## A. A FORMAL PROOF SYSTEM

We shall define a formal proof system for predicate logic with identity. To cover propositional logic too, the language will have some sentence letters. The calculus is a Hilbert-style system.

First we define the *language*  $L$ , by describing its similarity type, its set of terms and its set of formulas (cf. Sections 3 and 13 above).

The *similarity type* of  $L$  is made up of the following sentence letters, individual constants, predicate constants and function constants. The *sentence letters* are the expressions  $p_n$ , where  $n$  is a natural number subscript. The *individual constants* are the expressions  $c_n$ , where  $n$  is a natural number subscript. The *predicate constants* are the expressions  $P_n^m$ , where  $n$  is a natural number subscript and  $m$  is a positive integer superscript. The *function constants* are the expressions  $f_n^m$ , where  $n$  is a natural number subscript and  $m$  is a positive integer superscript. A predicate or function constant is said to be *m-place* if its superscript is  $m$ .

The *terms* of  $L$  are defined inductively as follows: (i) Every variable is a term, where the *variables* are the expressions  $x_n$  with natural number subscript  $n$ . (ii) For each function symbol  $f_n^m$ , if  $\tau_1, \dots, \tau_m$  are terms then the expression  $f_n^m(\tau_1, \dots, \tau_m)$  is a term. (iii) Nothing is a term except as required by (i) and (ii).



The *formulas* of **L** are defined inductively as follows: (i) Every sentence letter is a formula. (ii) The expression  $\perp$  is a formula. (iii) For each predicate constant  $R_n^m$ , if  $\tau_1, \dots, \tau_m$  are terms then the expression  $R_n^m(\tau_1, \dots, \tau_m)$  is a formula. (iv) If  $\sigma$  and  $\tau$  are terms then the expression  $(\sigma = \tau)$  is a formula. (v) If  $\phi$  and  $\psi$  are formulas, then so are the expressions  $\neg\phi$ ,  $(\phi \wedge \psi)$ ,  $(\phi \vee \psi)$ ,  $(\phi \rightarrow \psi)$ ,  $(\phi \leftrightarrow \psi)$ . (vi) For each variable  $x_n$ , if  $\phi$  is a formula then so are the expressions  $\forall x_n \phi$  and  $\exists x_n \phi$ . (vii) Nothing is a formula except as required by (i)–(vi).

A full account would now define two further notions,  $FV(\phi)$  (the set of variables with free occurrences in  $\phi$ ) and  $\phi[\tau_1 \dots \tau_k / x_{i_1} \dots x_{i_k}]$  (the formula which results when we simultaneously replace all free occurrences of  $x_{i_j}$  in  $\phi$  by  $\tau_j$ , for each  $j, 1 \leq j \leq k$ , avoiding clash of variables). Cf. Section 13 above.

Now that **L** has been defined, formulas occurring below should be read as metalinguistic names for formulas of **L**. Hence we can make free use of the metalanguage abbreviations in Sections 4 and 13.

Now we define the proof system—let us call it **H**. We do this by describing the axioms, the derivations, and the way in which a sequent is to be read off from a derivation. (Sundholm (see Volume 2) describes an alternative Hilbert-style system **CQC** which is equivalent to **H**.)

The *axioms* of **H** are all formulas of the following forms:

- H1.**  $\phi \rightarrow (\psi \rightarrow \phi)$
- H2.**  $(\phi \rightarrow \psi) \rightarrow ((\phi \rightarrow (\psi \rightarrow \chi)) \rightarrow (\phi \rightarrow \chi))$
- H3.**  $(\neg\phi \rightarrow \psi) \rightarrow ((\neg\phi \rightarrow \neg\psi) \rightarrow \phi)$
- H4.**  $((\phi \rightarrow \perp) \rightarrow \perp) \rightarrow \phi$
- H5.**  $\phi \rightarrow (\psi \rightarrow \phi \wedge \psi)$
- H6.**  $\phi \wedge \psi \rightarrow \phi, \quad \phi \wedge \psi \rightarrow \psi$
- H7.**  $\phi \rightarrow \phi \vee \psi, \quad \psi \rightarrow \phi \vee \psi$
- H8.**  $(\phi \rightarrow \chi) \rightarrow ((\psi \rightarrow \chi) \rightarrow (\phi \vee \psi \rightarrow \chi))$
- H9.**  $(\phi \rightarrow \psi) \rightarrow ((\psi \rightarrow \phi) \rightarrow (\phi \leftrightarrow \psi))$
- H10.**  $(\phi \leftrightarrow \psi) \rightarrow (\phi \rightarrow \psi), \quad (\phi \leftrightarrow \psi) \rightarrow \psi \rightarrow \phi$
- H11.**  $\phi[\tau/x] \rightarrow \exists x\phi$
- H12.**  $\forall x\phi \rightarrow \phi[\tau/x]$
- H13.**  $x = x$
- H14.**  $x = y \rightarrow (\phi \rightarrow \phi[y/x])$

A *derivation* (or *formal proof*) in  $\mathbf{H}$  is defined to be a finite sequence

$$(A.1) \quad \langle \langle \phi_1, m_1 \rangle, \dots, \langle \phi_n, m_n \rangle \rangle$$

such that  $n \geq 1$ , and for each  $i$  ( $1 \leq i \leq n$ ) one of the five following conditions holds:

1.  $m_i = 1$  and  $\phi_i$  is an axiom;
2.  $m_i = 2$  and  $\phi_i$  is any formula of  $\mathbf{L}$ ;
3.  $m_i = 3$  and there are  $j$  and  $k$  in  $\{1, \dots, i-1\}$  such that  $\phi_k$  is  $\phi_j \rightarrow \phi_i$ ;
4.  $m_i = 4$  and there is  $j$  ( $1 \leq j < i$ ) such that  $\phi_j$  has the form  $\psi \rightarrow \chi$ ,  $x$  is a variable not free in  $\psi$ , and  $\phi_i$  is  $\psi \rightarrow \forall x\chi$ ;
5.  $m_i = 5$  and there is  $j$  ( $1 \leq j < i$ ) such that  $\phi_j$  has the form  $\psi \rightarrow \chi$ ,  $x$  is a variable not free in  $\chi$ , and  $\phi_i$  is  $\exists x\psi \rightarrow \chi$ .

Conditions 3–5 are called the *derivation rules* of the calculus. They tell us how we can add new formulas to the end of a derivation. Thus (3) says that if  $\psi$  and  $\psi \rightarrow \chi$  occur in a derivation, then we can add  $\chi$  at the end; this is the rule of *modus ponens*.

The *premises* of the derivation (A.1) are those formulas  $\phi_i$  such that  $m_i = 2$ . Its *conclusion* is  $\phi_n$ . We say that  $\psi$  is *derivable from*  $\chi_1, \dots, \chi_k$  *in the calculus*  $\mathbf{H}$ , in symbols

$$(A.2) \quad \chi_1, \dots, \chi_n \vdash_{\mathbf{H}} \psi,$$

if there exists a derivation whose premises are all among  $\chi_1, \dots, \chi_n$  and whose conclusion is  $\psi$ .

### Remarks

1. The calculus  $\mathbf{H}$  is sound and strongly complete for propositional and predicate logic with identity. (Cf. Section 7; as in Section 15, this says nothing about provable sequents in which some variables occur free.)
2. In practice most logicians would write the formulas of a derivation as a column or a tree, and they would omit the numbers  $m_i$ .
3. To prove the completeness of  $\mathbf{H}$  by either the first or the third method in Section 16, one needs to know for all sentences  $\chi_1, \dots, \chi_n$  and  $\psi$ ,

$$(A.3) \quad \text{if } \chi_1, \dots, \chi_n \vdash_{\mathbf{H}} \psi \text{ then } \chi_1, \dots, \chi_{n-1} \vdash_{\mathbf{H}} \chi_n \rightarrow \psi.$$

Statement (A.3) is the *Deduction Theorem* for  $\mathbf{H}$ . It remains true if we allow free variables to occur in the formulas, provided that they occur only in certain ways. See [Kleene, 1952, Sections 21–24] for details.

4. Completeness and soundness tell us that if  $\chi_1, \dots, \chi_n$  and  $\psi$  are sentences, then (A.2) holds if and only if  $\chi_1, \dots, \chi_n \vDash \psi$ . This gives an intuitive meaning to such sequents. But when  $\chi_1, \dots, \chi_n$  and  $\psi$  are allowed to be any formulas of  $\mathbf{L}$ , then to the best of my knowledge there are no natural necessary and sufficient conditions for (A.2) to hold. So it seems impossible to explain what if anything (A.2) tells us, except by referring to the fine details of the calculus  $\mathbf{H}$ . This is a general feature of Hilbert-style calculi for predicate logic, and I submit that it makes them thoroughly inappropriate for introducing undergraduates to logic.
5. If we are thinking of varying the rules of the calculus, or even if we just want a picture of what the calculus is about, it is helpful to have at least a *necessary* condition for (A.2) to hold. The following supplies one. The *universal closure* of  $\phi$  is  $\forall y_1, \dots, y_n \phi$ , where  $y_1, \dots, y_n$  are the free variables of  $\phi$ . Let  $\phi_1$  be the universal closure of  $\chi_1 \wedge \dots \wedge \chi_n$  and  $\phi_2$  the universal closure of  $\psi$ . Then one can show that

$$(A.4) \quad \text{if } \chi_1, \dots, \chi_n \vdash_{\mathbf{H}} \psi \text{ then } \phi_1 \vDash \phi_2.$$

The proof of (A.4) is by induction on the lengths of derivations. Statement (A.4) is one way of showing that  $\mathbf{H}$  is sound.

6. The following derivation shows that  $\vdash_{\mathbf{H}} \exists x(x = x)$ :

$$(A.5) \quad \begin{array}{ll} x = x & \text{(axiom H13)} \\ x = x \rightarrow \exists x(x = x) & \text{(axiom H11)} \\ \exists x(x = x) & \text{(from above by modus ponens)} \end{array}$$

Statement (A.4) shows the reason, namely:

$$(A.6) \quad \forall x(x = x \wedge (x = x \rightarrow \exists x(x = x))) \vDash \exists x(x = x).$$

On any reasonable semantic interpretation (cf. Section 14 above), the left-hand side in (A.6) is true in the empty structure but the right-hand side is false. Suppose now that we want to modify the calculus in order to allow empty structures. Then we must alter the derivation rule which took us from left to right in (A.6), and this is the rule of modus ponens. (Cf. Bencivenga (Volume 7 of this *Handbook*.) It is important to note here that even if (A.4) was a tidy two-way implication, the modus ponens rule would *not* express ' $\phi$  and  $\phi \rightarrow \psi$  imply  $\psi$ ', but rather something of the form ' $\forall \vec{x}(\phi \wedge (\phi \rightarrow \psi))$  implies  $\forall \vec{y}\psi$ '. As it is, the meaning of modus ponens in  $\mathbf{H}$  is quite obscure. (Cf. [Kleene, 1952, Section 24].)

## B. ARITHMETIC

I begin with naive arithmetic, not formal Peano arithmetic. One needs to have at least an intuitive grasp of naive arithmetic in order to understand what a formal system is. In any case [Peano, 1889] reached his axioms by throwing naive arithmetic into fancy symbols.

*Naive arithmetic* is adequately summed up by the following five axioms, which come from Dedekind [1888; 1967]. Here and below, ‘number’ means ‘natural number’, and I start with 0 (Dedekind’s first number was 1).

NA1. 0 is a number.

NA2. For every number  $n$  there is a next number after  $n$ ; this next number is called  $Sn$  or the *successor* of  $n$ .

NA3. Two different numbers never have the same successor.

NA4. 0 is not the successor of any number.

NA5. (Induction axiom) Let  $K$  be any set with the properties (i) 0 is in  $K$ , (ii) for every number  $n$  in  $K$ ,  $Sn$  is also in  $K$ . Then every number is in  $K$ .

These axioms miss one vital feature of numbers, viz. their order. So we define  $<$  as follows. First we define an *initial segment* to be a set  $K$  of numbers such that if a number  $Sn$  is in  $K$  then  $n$  is also in  $K$ . We say:

(B.1)  $m < n$  iff there is an initial segment which contains  $m$  but not  $n$ .

The definition (B.1) implies:

(B.2) If  $m < Sn$  then either  $m < n$  or  $m = n$ .

For future reference I give a proof. Suppose  $m < Sn$  but not  $m = n$ . Then there is an initial segment  $K$  such that  $m$  is in  $K$  and  $Sn$  is not in  $K$ . Now there are two cases. *Case 1:*  $n$  is not in  $K$ . Then by (B.1),  $m < n$ . *Case 2:*  $n$  is in  $K$ . Then let  $M$  be  $K$  with  $n$  omitted. Since  $m \neq n$ ,  $M$  contains  $m$  but not  $n$ . Also  $M$  is an initial segment; for if  $Sk$  is in  $M$  but  $k$  is not, then by the definition of  $M$  we must have  $k = n$ , which implies that  $Sn$  is in  $M$  and hence in  $K$ ; contradiction. So we can use  $M$  in (B.1) to show  $m < n$ .

(B.3) For each number  $m$  it is false that  $m < 0$ .

(B.3) is proved ‘by induction on  $m$ ’, using the induction axiom NA5. Proofs of this type are written in a standard style, as follows:

**Case 1.**  $m = 0$ . Then  $m < 0$  would imply by (B.1) that there was a set containing 0 but not 0, which is impossible.

**Case 2.**  $m = Sk$ , assuming it proved when  $m = k$ . Suppose  $Sk < 0$ . Then by (B.1) there is an initial segment containing  $Sk$  and not 0. Since  $K$  is an initial segment containing  $Sk$ ,  $k$  is also in  $K$ . So by (B.1) again,  $K$  shows that  $k < 0$ . But the induction hypothesis states that not  $k < 0$ ; contradiction.

This is all one would normally say in the proof. To connect it with NA5, let  $M$  be the set of all numbers  $m$  such that not  $m < 0$ . The two cases show exactly what has to be shown, according to NA5, in order to prove that every number is in  $M$ .

Here are two more provable facts.

(B.4) The relation  $<$  is a linear ordering of the numbers (in the sense of (157)–(159) in Section 19 above).

(B.5) Every non-empty set of numbers has a first element.

Fact (B.5) states that the numbers are *well-ordered*, and it is proved as follows. Let  $X$  be any set of numbers without a first element. Let  $Y$  be the set of numbers not in  $X$ . Then by induction on  $n$  we show that every number  $n$  is in  $Y$ . So  $X$  is empty.

Fact (B.5) is one way of justifying *course-of-values induction*. This is a style of argument like the proof of (B.3) above, except that in Case 2, instead of proving the result for  $Sk$  assuming it was true for  $k$ , we prove it for  $Sk$  assuming it was true for *all numbers*  $\leq k$ . In many theorems about logic, one shows that every formula has some property  $A$  by showing (i) that every atomic formula has property  $A$  and (ii) that if  $\phi$  is a compound formula whose proper subformulas have  $A$  then  $\phi$  has  $A$ . Arguments of this type are course-of-values inductions on the complexity of formulas.

In naive arithmetic we can justify two important types of definition. The first is sometimes called *recursive definition* and sometimes *definition by induction*. It is used for defining functions whose domain is the set of natural numbers. To define such a function  $F$  recursively, we first say outright what  $F(0)$  is, and then we define  $F(Sn)$  in terms of  $F(n)$ . A typical example is the recursive definition of addition:

$$(B.6) \quad m + 0 = m, \quad m + Sn = S(m + n).$$

Here  $F(n)$  is  $m + n$ ; the definition says first that  $F(0)$  is  $m$  and then that for each number  $n$ ,  $F(Sn)$  is  $SF(n)$ . To justify such a definition, we have to show that there is exactly one function  $F$  which satisfies the stated conditions. To show there is *at most* one such function, we suppose that  $F$  and  $G$  are two functions which meet the conditions, and we prove by induction on  $n$  that for every  $n$ ,  $F(n) = G(n)$ ; this is easy. To show that there is *at least* one is harder. For this we define an *n-approximation* to be a function whose domain is the set of all numbers  $< n$ , and which obeys the conditions

in the recursive definition for all numbers in its domain. Then we show by induction on  $n$  (i) that there is at least one  $n$ -approximation, and (ii) that if  $m < k < n$ ,  $f$  is a  $k$ -approximation and  $g$  is an  $n$ -approximation, then  $f(m) = g(m)$ . Then finally we define  $F$  explicitly by saying that  $F(m)$  is the unique number  $h$  such that  $f(m) = h$  whenever  $f$  is an  $n$ -approximation for some number  $n$  greater than  $m$ .

After defining  $+$  by (B.6), we can go on to define  $\cdot$  by:

$$(B.7) \quad m \cdot 0 = 0, \quad m \cdot Sn = m \cdot n + m.$$

The functions definable by a sequence of recursive definitions in this way, using equations and previously defined functions, are called *primitive recursive* functions. Van Dalen [this Volume] discusses them further.

There is a course-of-values recursive definition too: in this we define  $F(0)$  outright, and then  $F(Sn)$  in terms of values  $F(k)$  for numbers  $k \leq n$ . For example if  $F(n)$  is the set of all formulas of complexity  $n$ , understood as in Section 3 above, then the definition of  $F(n)$  will have to refer to the sets  $F(k)$  for all  $k < n$ . Course-of-values definitions can be justified in the same way as straightforward recursive definitions.

The second important type of definition that can be justified in naive arithmetic is also known as *inductive definition*, though it is quite different from the ‘definition by induction’ above. Let  $H$  be a function and  $X$  a set. We say that  $X$  is *closed under  $H$*  if for every element  $x$  of  $X$ , if  $x$  is in the domain of  $H$  then  $H(x)$  is also in  $X$ . We say that  $X$  is the *closure* of  $Y$  under  $H$  if (i) every element of  $Y$  is in  $X$ , (ii)  $X$  is closed under  $H$ , and (iii) if  $Z$  is any set which includes  $Y$  and is closed under  $H$  then  $Z$  also includes  $X$ . (Briefly, ‘ $X$  is the smallest set which includes  $Y$  and is closed under  $H$ ’.) Similar definitions apply if we have a family of functions  $H_1, \dots, H_k$  instead of the one function  $H$ ; also the functions can be  $n$ -place functions with  $n > 1$ .

A set is said to be *inductively defined* if it is defined as being the closure of some specified set  $Y$  under some specified functions  $H_1, \dots, H_k$ . A typical inductive definition is the definition of the set of terms of a language  $L$ . The usual form for such a definition is:

1. Every variable and every individual constant is a term.
2. For each function constant  $f$ , if  $f$  is  $n$ -place and  $\tau_1, \dots, \tau_n$  are terms, then the expression  $f(\tau_1, \dots, \tau_n)$  is a term.
3. Nothing is a term except as required by (1) and (2).

Here we are defining the set  $X$  of terms. The so-called *basic* clause (1) describes  $Y$  as the set of all variables and all individual constants. The *inductive* clause (2) describes the functions  $H_i$ , one for each function constant.

Finally the *extremal* clause (3) says that  $X$  is the closure of  $Y$  under the  $H_i$ . (Many writers omit the extremal clause, because it is rather predictable.)

Frege [1884] may have been the first to argue that inductive definitions need to be justified. He kept asking: How do we know that there *is* a smallest set which includes  $Y$  and is closed under  $H$ ? One possible justification runs as follows. We recursively define  $F(n)$ , for each positive integer  $n$ , to be the set of all sequences  $\langle b_1, \dots, b_n \rangle$  such that  $b_1$  is in  $Y$  and for every  $i$  ( $1 \leq i < n$ ),  $b_{i+1}$  is  $H(b_i)$ . Then we define  $X$  to be the set of all  $b$  such that for some number  $n$  there is a sequence in  $F(n)$  whose last term is  $b$ . Clearly  $Y$  is included in  $X$ , and we can show that  $X$  is closed under  $H$ . If  $Z$  is any set which is closed under  $H$  and includes  $Y$ , then an induction on the lengths of sequences shows that every element of  $X$  is in  $Z$ .

Naive arithmetic, as described above, is an axiomatic system but not a formal one. Peano [1889] took the first step towards formalising it, by inventing a good symbolism. But the arguments above use quite an amount of set theory, and Peano made no attempt to write down what he was assuming about sets. Skolem [1923] threw out the set theory and made his assumptions precise, but his system was rather weak. First-order Peano arithmetic, a formalisation of the first-order part of Peano's axioms, was introduced in [Gödel, 1931b].

**P**, or *first-order Peano Arithmetic*, is the following formal system. The constants of the language are an individual constant  $\bar{0}$ , a 1-place function symbol  $S$  and 2-place function symbols  $+$  and  $\bullet$ , forming terms of form  $Sx$ ,  $(x+y)$ ,  $(x \bullet y)$ . Write  $\bar{n}$  as an abbreviation for  $S \dots (n \text{ times}) \dots S\bar{0}$ ; the symbols  $\bar{n}$  are called *numerals*. We use a standard proof calculus for first-order logic (e.g. the calculus **H** of Appendix A) together with the following axioms:

$$\text{P1. } \forall xy(Sx = Sy \rightarrow x = y)$$

$$\text{P2. } \forall x \neg(Sx = 0)$$

$$\text{P3. (Axiom schema of induction) All sentences of the form } \forall z(\phi[\bar{0}/x] \wedge \forall x(\phi \rightarrow \phi[Sx/x]) \rightarrow \forall x\phi)$$

$$\text{P4. } \forall x(x + \bar{0} = x)$$

$$\text{P5. } \forall xy(x + Sy = S(x + y))$$

$$\text{P6. } \forall x(x \bullet \bar{0} = \bar{0})$$

$$\text{P7. } \forall xy(x \bullet Sy = (x \bullet y) + x)$$

The axioms are read as being just about numbers, so that  $\forall x$  is read as 'for all numbers  $x$ '. In this way the symbols  $\bar{0}$  and  $S$  in the language take care of axioms NA1 and NA2 without further ado. Axioms NA3 and NA4

appear as **P1** and **P2**. Since we can refer only to numbers and not to sets, axiom NA5 has to be recast as a condition on those sets of numbers which are definable by first-order formulas; this accounts for the axiom schema of induction, **P3**.

**P4–P7** are the recursive definitions of addition and multiplication, cf. (B.6) and (B.7) above. In naive arithmetic there was no need to assume these as axioms, because we could *prove* that there are unique functions meeting these conditions. However, the proof used some set-theoretic notions like ‘function defined on the numbers  $0, \dots, n-1$ ’, which can’t be expressed in a first-order language using just  $\bar{0}$  and  $S$ . So we have to put the symbols  $+$ ,  $\bullet$  into the language—in particular they occur in formulas in the axiom schema of induction—and we have to assume the definitions **P4 – P7** as axioms.

Gödel showed that with the aid of first-order formulas involving only  $\bar{0}, S, +$  and  $\bullet$ , he could explicitly define a number of other notions. For example

$$(B.8) \quad x < y \text{ iff } \exists z(x + Sz = y).$$

Also by using a clever trick with prime numbers he could encode each finite sequence  $\langle m_1, m_2, \dots \rangle$  of numbers as a single number

$$(B.9) \quad 2^{m_1+1}.3^{m_2+1}.5^{m_3+1} \dots$$

and he could express the relation ‘ $x$  is the  $y$ th term of the sequence coded by  $z$ ’ by a first-order formula. But then he could carry out ‘in **P**’ all the parts of naive arithmetic which use only numbers, finite sequences of numbers, finite sequences of finite sequences of numbers, and so on. This includes the argument which justifies primitive recursive definitions. In fact:

1. *For every recursive definition  $\delta$  of a number function, using just first-order formulas, there is a formula  $\phi(x, y)$  such that in **P** we can prove that  $\phi$  defines a function obeying  $\delta$ . (If  $\delta$  is primitive recursive then  $\phi$  can be chosen to be  $\Sigma_1$ , cf. Section 24.)*
2. *For every inductive definition of a set, where a formula  $\psi$  defines the basic set  $Y$  and formulas  $\chi$  define the functions  $H$  in the inductive clause, there is a formula  $\phi(x)$  such that we can prove in **P** that the numbers satisfying  $\phi$  are those which can be reached in a finite number of steps from  $Y$  by  $H$ . (If  $\psi$  and  $\chi$  are  $\Sigma_1$  then  $\phi$  can be chosen to be  $\Sigma_1$ .)*

These two facts state in summary form why the whole of elementary syntax can be formalised within **P**.

There are some things that can be said in the language of **P** but not proved or refuted from the axioms of **P**. For example the statement that **P**



itself is consistent (i.e. doesn't yield  $\perp$ ) can be formalised in the language of  $\mathbf{P}$ . In [1931b] Gödel showed that this formalised statement is not deducible from  $\mathbf{P}$ , although we all hope it is true.

There are some other things that can't even be said in the language of  $\mathbf{P}$ . For example we can't say in this language that the set  $X$  defined by  $\phi$  in (2) above really is the closure of  $Y$  under  $H$ , because that would involve us in saying that 'if  $Z$  is *any set* which includes  $Y$  and is closed under  $H$  then  $Z$  includes  $X$ '. In the first-order language of  $\mathbf{P}$  there is no way of talking about 'all sets of numbers'. For the same reason, many statements about real numbers can't be expressed in the language of  $\mathbf{P}$ —even though some can by clever use of rational approximations.

In second-order arithmetic we can talk about real numbers, because real numbers can be represented as sets of natural numbers. Actually the natural numbers themselves are definable up to isomorphism in second-order logic without special arithmetical axioms. In third-order logic we can talk about sets of real numbers, fourth-order logic can talk about sets of sets of real numbers, and so on. Most of the events that take place in any standard textbook of real analysis can be recorded in, say, fifth-order logic. See Van Benthem and Doets [this Volume] for these higher-order logics.

## C. SET THEORY

The efforts of various nineteenth-century mathematicians reduced all the concepts of real and complex number theory to one basic notion: classes. So when Frege, in his *Grundgesetze der Arithmetik I* [1893], attempted a formal system which was to be adequate for all of arithmetic and analysis, the backbone of his system was a theory of classes. One of his assumptions was that for every condition there is a corresponding class, namely the class of all the objects that satisfy the condition. Unfortunately this assumption leads to contradictions, as Russell and Zermelo showed. Frege's approach has now been abandoned.

Today the most commonly adopted theory of classes is Zermelo–Fraenkel set theory, ZF. This theory was propounded by Zermelo [1908] as an informal axiomatic theory. It reached its present shape through contributions from Mirimanoff, Fraenkel, Skolem and von Neumann. (Cf. Fraenkel's historical introduction to [Bernays and Fraenkel, 1958].)

Officially ZF is a set of axioms in a first-order language whose only constant is the 2-place predicate symbol  $\in$  ('is a member of'). But all set theorists make free use of symbols introduced by definition.

Let me illustrate how a set theorist introduces new symbols. The axiom of Extensionality says that no two different sets have the same members. The Pair-set axiom says that if  $x$  and  $y$  are sets then there is at least one set which has just  $x$  and  $y$  as members. Putting these two axioms together, we

infer that there is exactly one set with just  $x$  and  $y$  as members. Introducing a new symbol, we call this set  $\{x, y\}$ . There are also some definitions which don't depend on the axioms. For example we say  $x$  is *included in*  $y$ , or a *subset of*  $y$ , if every member of  $x$  is a member of  $y$ . This prompts the definition

$$(C1) \quad x \subseteq y \quad \text{iff} \quad \forall t(t \in x \rightarrow t \in y).$$

The language with these extra defined symbols is in a sense impure, but it is much easier to read than the pure set language with only  $\in$ , and one can always paraphrase away the new symbols when necessary. In what follows I shall be relentlessly impure. (On introducing new terms by definition, cf. Section 21 above. Suppes [1972] and Levy [1979] are careful about it.)

The first three axioms of ZF are about what kind of things we choose to count as sets. The axiom of Extensionality says that sets will count as equal when they have the same members:

$$\text{ZF1.} \quad (\text{Extensionality}) \quad \forall xy(x \subseteq y \wedge y \subseteq x \rightarrow x = y)$$

We think of sets as being built up by assembling their members, starting with the empty or null set  $0$  which has no members:

$$\text{ZF2.} \quad (\text{Null-set}) \quad \forall t \quad t \neq 0 \quad (x \notin y \text{ means } \neg(x \in y)).$$

In a formal calculus which proves  $\exists x \quad x = x$ , the Null-set axiom is derivable from the Separation axiom below and can be omitted. The axiom of Regularity (also known as the axiom of Foundation) expresses—as well as one can express it with a first-order statement—that  $X$  will not count as a set unless each of the members of  $x$  could be assembled together at an earlier stage than  $x$  itself. (So for example there is no 'set'  $x$  such that  $x \in x$ .)

$$\text{ZF3.} \quad (\text{Regularity}) \quad \forall x(x = 0 \vee \exists y(y \in x \wedge \forall z(z \in y \rightarrow z \notin x))).$$

The next three axioms state that certain collections can be built up:

$$\text{ZF4.} \quad (\text{Pair-set}) \quad \forall xy(t \in \{x, y\} \leftrightarrow t = x \vee t = y)$$

$$\text{ZF5.} \quad (\text{Union}) \quad \forall xt(t \in \bigcup x \leftrightarrow \exists y(t \in y \wedge y \in x))$$

$$\text{ZF6.} \quad (\text{Power-set}) \quad \forall xt(t \in \mathcal{P}x \leftrightarrow t \subseteq x).$$

Axioms ZF3–ZF6 allow some constructions. We write  $\{x\}$  for  $\{x, x\}$ ,  $x \cup y$  for  $\bigcup\{x, y\}$ ,  $\{x_1, x_2, x_3\}$  for  $\{x_1, x_2\} \cup \{x_3\}$ ,  $\{x_2, \dots, x_4\}$  for  $\{x_1, x_2, x_3\} \cup \{x_4\}$ , and so on. Likewise we can form ordered pairs  $\langle x, y \rangle = \{\{x\}, \{x, y\}\}$ , ordered triplets  $\langle x, y, z \rangle = \langle \langle x, y \rangle, z \rangle$  and so on. Building up from  $0$  we can form  $1 = \{0\}$ ,  $2 = \{0, 1\}$ ,  $3 = \{0, 1, 2\}$  etc.; the axiom of Regularity implies that  $0, 1, 2, \dots$  are all distinct. We can regard  $0, 1, 2, \dots$  as the natural numbers.

We need to be able to express ‘ $x$  is a natural number’ in the language of set theory, without using informal notions like ‘and so on’. It can be done as follows. First, following von Neumann, we define  $\text{Ord}(x)$ , ‘ $x$  is an ordinal’, by:

$$(C.2) \quad \text{Ord}(x) \text{ iff } \bigcup x \subseteq x \wedge \forall yz(y \in x \wedge z \in x \rightarrow y \in z \vee z \in y \vee y = z).$$

This somewhat technical definition implies that the ordinals are linearly ordered by  $\in$ , and that they are well-ordered (i.e. every non-empty set of them has a least element, cf. (B.5) above). We can prove that the first ordinals are  $0, 1, 2, \dots$ . Greek letters  $\alpha, \beta, \gamma$  are used for ordinals. For every ordinal  $\alpha$  there is a first greater ordinal; it is written  $\alpha + 1$  and defined as  $\alpha \cup \{\alpha\}$ . For every set  $X$  of ordinals there is a first ordinal  $\beta$  which is greater than or equal to every ordinal in  $X$ , viz.  $\beta = \bigcup X$ . Each ordinal  $\beta$  has just one of the following three forms: either  $\beta = 0$ , or  $\beta$  is a *successor* (i.e. of form  $\alpha + 1$ ), or  $\beta$  is a *limit* (i.e. of form  $\bigcup X$  for a non-empty set  $X$  of ordinals which has no greatest member). Now the natural numbers can be defined as follows:

$$(C.3) \quad x \text{ is a natural number} \text{ iff } \text{Ord}(x) \wedge \forall y(y \in x + 1 \rightarrow y = 0 \vee y \text{ is a successor}).$$

The remaining four axioms, ZF7–ZF10, are needed for talking about infinite sets. Each of them says that sets exist with certain properties. Nothing in ZF1–ZF6 implies that there are any infinite sets. We fill the gap by decreeing that the set  $\omega$  of all natural numbers exists:

$$\text{ZF7. (Infinity)} \quad \forall t(t \in \omega \leftrightarrow t \text{ is a natural number}).$$

The next axiom says that within any given set  $x$  we can collect together those members  $w$  which satisfy the formula  $\phi(\vec{z}, w)$ . Here  $\phi$  is allowed to be any first-order formula in the language of set theory, and it can mention other sets  $\vec{z}$ . Strictly ZF8 is an axiom schema and not a single axiom.

$$\text{ZF8. (Separation)} \quad \forall \vec{z}x t(t \in \{w \in x | \phi\} \leftrightarrow t \in \{x \wedge \phi[t/w]\}).$$

For example this tells us that for any sets  $x$  and  $y$  there is a set whose members are exactly those members  $w$  of  $x$  which satisfy the formula  $w \in y$ ; in symbols this set is  $\{w \in x | w \in y\}$ . So we can introduce a new symbol for this set, and write  $x \cap y = \{w \in x | w \in y\}$ . Similarly we can define:  $\bigcap x = \{w \in \bigcup x | \forall z(z \in x \rightarrow w \in z)\}$ ,  $x \times y = \{t \in \mathcal{P}\mathcal{P}(x \cup y) | \exists zw(z \in x \wedge w \in y \wedge t = \langle z, w \rangle)\}$ ,  $x^2 = x \times x$  and more generally  $x^{n+1} = x^n \times x$ . An *n-place relation* on the set  $x$  is a subset of  $x^n$ . We can define ‘ $f$  is a function from  $x$  to  $y$ ’, in symbols  $f : x \rightarrow y$ , by:

$$(C.4) \quad f : x \rightarrow y \text{ iff } f \subseteq x \times y \wedge \forall w(w \in x \rightarrow \exists z \forall t(t = z \leftrightarrow \langle w, t \rangle \in f)).$$

We say  $f$  is an  $n$ -place function from  $x$  to  $y$  if  $f : x^n \rightarrow y$ . When  $f : x \rightarrow y$ , we call  $x$  the *domain* of  $f$ , and we can define it in terms of  $f$  by:  $\text{dom} f = \{w \in \bigcup \bigcup f \mid \exists z \langle w, z \rangle \in f\}$ . We define the *value* of  $f$  for *argument*  $w$ , in symbols  $f(w)$ , as  $\{t \in \bigcup \bigcup \bigcup f \mid \exists z (\langle w, z \rangle \in f \wedge t \in z)\}$ . A *bijection* (or *one-one correspondence*) from  $x$  to  $y$  is a function  $f$  such that  $f : x \rightarrow y$  and every element  $z$  of  $y$  is of form  $f(w)$  for exactly one  $w$  in  $x$ . A *sequence of length*  $\alpha$  is defined to be a function with domain  $\alpha$ .

The system of axioms ZF1–ZF8 is sometimes known as *Zermelo set theory*, or  $Z$  for short. It is adequate for formalising all of naive arithmetic, not just the finite parts that can be axiomatised in first-order Peano arithmetic. The Separation axiom is needed. For example in the proof of (B.2) we had to know that there is a set  $M$  whose members are all the members of  $K$  except  $n$ ;  $M$  is  $\{w \in K \mid w \neq n\}$ .

First-order languages can be defined formally within  $Z$ . For example we can define a *similarity type* for predicate logic to be a set whose members each have one of the following forms: (i)  $\langle 1, x \rangle$ , (ii)  $\langle 2, m, x \rangle$  where  $m$  is a positive natural number, (iii)  $\langle 3, m, x \rangle$  where  $m$  is a positive natural number. The elements of form (i) are called *individual constants*, those of form (ii) are the  *$m$ -place predicate constants* and those of form (iii) are the  *$m$ -place function constants*. *Variables* can be defined as ordered pairs of form  $\langle 4, n \rangle$  where  $n$  is a natural number. *Terms* can be defined inductively by: (a) Every variable or individual constant is a term. (b) If  $f$  is an  $m$ -place function constant and  $\tau_1, \dots, \tau_m$  are terms then  $\langle 5, f, \tau_1, \dots, \tau_m \rangle$  is a term. (c) Nothing is a term except as required by (a) and (b). By similar devices we can define the whole language  $L$  of a given similarity type  $X$ .  $L$ -structures can be defined to be ordered pairs  $\langle A, I \rangle$  where  $A$  is a non-empty set and  $I$  is a function with domain  $X$ , such that for each individual constant  $c$  of  $X$ ,  $I(c) \in A$  (and so on as in Section 14). Likewise we can define  $\models$  for  $L$ -structures.

The two remaining axioms of ZF are needed for various arguments in infinite arithmetic.

In Appendix B we saw how one can define functions with domain the natural numbers, by recursion. We want to be able to do the same in set theory, but with any ordinal as the domain. For example if the language  $L$  is not countable, then the proof of completeness in Section 16 above will need to be revised so that we build a chain of theories  $\Delta_i$  for  $i \in \alpha$ , where  $\alpha$  is some ordinal greater than  $\omega$ . One can try to justify recursive definitions on ordinals, just as we justified definitions in Appendix B. It turns out that one piece of information is missing. We need to know that if a formula defines a function  $f$  whose domain is an ordinal, then  $f$  is a set. The following axiom supplies this missing information. It says that if a formula  $\phi$  defines a function with domain a set, then the image of this function is again a set:

ZF9. (Replacement)

$$\forall z x (\forall y w t (y \in x \wedge \phi \wedge \phi[w/t] \rightarrow t = w) \rightarrow \\ \exists u \forall t (t \in u \leftrightarrow \exists y (y \in x \wedge \phi))).$$

Like Separation, the Replacement axiom is really an axiom schema.

The final axiom is the axiom of Choice, which is needed for most kinds of counting argument. This axiom can be given in many forms, all equivalent in the sense that any one can be derived from any other using ZF1–ZF9. The form given below, Zermelo’s Well-ordering principle, means intuitively that the elements of any set can be checked off one by one against the ordinals, and that the results of this checking can be gathered together into a set.

ZF10. (Well-ordering)

$$\forall x \exists f \alpha \text{ (}\alpha \text{ is an ordinal and } f \text{ is a bijection from } \alpha \text{ to } x \text{)}.$$

Axiom ZF10 is unlike axioms ZF4–ZF9 in a curious way. These earlier axioms each said that there is a set with just such-and-such members. But ZF10 says that a certain set exists (the function  $f$ ) without telling us what the members of the set are. So arguments which use the axiom of Choice have to be less explicit than arguments which only use ZF1–ZF9.

Using ZF10, the theory of ‘cardinality proceeds as follows. The *cardinality*  $|x|$  or  $x^\#$  of a set  $x$  is the first ordinal  $\alpha$  such that there is a bijection from  $\alpha$  to  $x$ . Ordinals which are the cardinalities of sets are called *cardinals*. Every cardinal is equal to its own cardinality. Every natural number is a cardinal. A set is said to be *finite* if its cardinality is a natural number. The cardinals which are not natural numbers are said to be *infinite*. The infinite cardinals can be listed in increasing order as  $\omega_0, \omega_1, \omega_2, \dots$ ;  $\omega_0$  is  $\omega$ . For every ordinal  $\alpha$  there is an  $\alpha$ th infinite cardinal  $\omega_\alpha$ , sometimes also written as  $\aleph_\alpha$ . It can be proved that there is no greatest cardinal, using Cantor’s theorem that for every set  $x$ ,  $\mathcal{P}(x)$  has greater cardinality than  $x$ .

Let me give an example of a principle equivalent to ZF10. If  $I$  is a set and for each  $i \in I$  a set  $A_i$  is given, then  $\prod_I A_i$  is defined to be the set of all functions  $f : I \rightarrow \bigcup \{A_i | i \in I\}$  such that for each  $j \in I$ ,  $f(j) \in A_j$ .  $\prod_I A_i$  is called the *product* of the sets  $A_i$ . Then ZF10 is equivalent to the statement: If the sets  $A_i$  in a product are all non-empty then their product is also not empty.

The compactness theorem for propositional logic with any set of sentence letters is not provable from ZF1–ZF9. *A fortiori* neither is the compactness theorem for predicate logic. Logicians have dissected the steps between ZF10 and the compactness theorem, and the following notion is one of the results. (It arose in other parts of mathematics too.)

Let  $I$  be any set. Then an *ultrafilter* on  $I$  is defined to be a subset  $D$  of  $\mathcal{P}(I)$  such that (i) if  $a$  and  $b \in D$  then  $a \cap b \in D$ , (ii) if  $a \in D$  and  $a \subseteq b \subseteq I$  then  $b \in D$ , and (iii) for all subsets  $a$  of  $I$ , exactly one of  $a$  and  $I - a$  is in  $D$  (where  $I - a$  is the set of all elements of  $I$  which are not in  $a$ ). For example if  $i \in I$  and  $D = \{a \in \mathcal{P}(I) | i \in a\}$  then  $D$  is an ultrafilter on  $I$ ; ultrafilters

of this form are called *principal* and they are uninteresting. From ZF1–ZF9 it is not even possible to show that there exist any non-principal ultrafilters at all. But using ZF10 one can prove the following principle:

**THEOREM C.5** *Let  $I$  be any infinite set. Then there exist an ultrafilter  $D$  on  $I$  and for each  $i \in I$  an element  $a_i \in D$ , such that for every  $j \in I$  the set  $\{i \in I \mid j \in a_i\}$  is finite.*

An ultrafilter  $D$  with the property described in Theorem C.5 is said to be *regular*. Regular ultrafilters are always non-principal.

To derive the compactness theorem from Theorem C.5, we need to connect ultrafilters with structures. This is done as follows. For simplicity we can assume that the language  $L$  has just one constant symbol, the 2-place predicate constant  $R$ . Let  $D$  be an ultrafilter on the set  $I$ . For each  $i \in I$ , let  $\mathfrak{A}_i$  be an  $L$ -structure with domain  $A_i$ . Define a relation  $\sim$  on  $\Pi_I A_i$  by:

$$(C.6) \quad f \sim g \text{ iff } \{i \in I \mid f(i) = g(i)\} \in D.$$

Then since  $D$  is an ultrafilter,  $\sim$  is an equivalence relation; write  $f^\sim$  for the equivalence class containing  $f$ . Let  $B$  be  $\{f^\sim \mid f \in \Pi_I A_i\}$ . Define an  $L$ -structure  $\mathfrak{B} = \langle B, I_{\mathfrak{B}} \rangle$  by putting

$$(C.7) \quad \langle f^\sim, g^\sim \rangle \in I_{\mathfrak{B}}(R) \text{ iff } \{i \in I \mid \langle f(i), g(i) \rangle \in I_{\mathfrak{A}_i}(R)\} \in D.$$

(Using the fact that  $D$  is an ultrafilter, this definition makes sense.) Then  $\mathfrak{B}$  is called the *ultraproduct* of the  $\mathfrak{A}_i$  by  $D$ , in symbols  $\Pi_D \mathfrak{A}_i$  or  $D$ -prod  $\mathfrak{A}_i$ . By a theorem of Jerzy Łoś, if  $\phi$  is any sentence of the first-order language  $L$ , then

$$(C.8) \quad \Pi_D \mathfrak{A}_i \models \phi \text{ iff } \{i \in I \mid \mathfrak{A}_i \models \phi\} \in D.$$

Using the facts above, we can give another proof of the compactness theorem for predicate logic. Suppose that  $\Delta$  is a first-order theory and every finite subset of  $\Delta$  has a model. We have to show that  $\Delta$  has a model. If  $\Delta$  itself is finite, there is nothing to prove. So assume now that  $\Delta$  is infinite, and let  $I$  in Theorem C.5 be  $\Delta$ . Let  $D$  and the sets  $a_\phi$  ( $\phi \in \Delta$ ) be as in Theorem C.5. For each  $i \in \Delta$ , the set  $\{\phi \mid i \in a_\phi\}$  is finite, so by assumption it has a model  $\mathfrak{A}_i$ . Let  $\mathfrak{B}$  be  $\Pi_D \mathfrak{A}_i$ . For each sentence  $\phi \in \Delta$ ,  $a_\phi \subseteq \{i \in \Delta \mid \mathfrak{A}_i \models \phi\}$ , so by (ii) in the definition of an ultrafilter,  $\{i \in \Delta \mid \mathfrak{A}_i \models \phi\} \in D$ . It follows by Łoś's theorem (C.8) that  $\mathfrak{B} \models \phi$ . Hence  $\Delta$  has a model, namely  $\mathfrak{B}$ .

There are full accounts of ultraproducts in Bell and Slomson [1969] and Chang and Keisler [1973]. One principle which often turns up when ultraproducts are around is as follows. Let  $X$  be a set of subsets of a set  $I$ . We say that  $X$  has the *finite intersection property* if for every finite subset  $\{a_1, \dots, a_n\}$  of  $X$ , the set  $a_1 \cap \dots \cap a_n$  is not empty. The principle states that *if  $X$  has the finite intersection property then there is an ultrafilter  $D$  on  $I$  such that  $X \subseteq D$* . This can be proved quite quickly from ZF10.

Some writers refer to ZF1–ZF9, without the axiom of Choice, as ZF; they write ZFC when Choice is included. There are a number of variants of ZF. For example the set-class theory of Gödel and Bernays (cf. [Mendelson, 1987]) allows one to talk about ‘the class of all sets which satisfy the formula  $\phi$ ’ provided that  $\phi$  has no quantifiers ranging over classes. This extension of ZF is only a notational convenience. It enables one to replace axiom schemas by single axioms, so as to get a system with just finitely many axioms.

Another variant allows elements which are not sets—these elements are called *individuals*. Thus we can talk about the set {Geoffrey Boycott} without having to believe that Geoffrey Boycott is a set. In informal set theory of course one considers such sets all the time. But there seems to be no mathematical advantage in admitting individuals into formal set theory; rather the contrary, we learn nothing new and the proofs are messier. A set is called a *pure set* if its members, its members’ members, its members’ members’ members etc. are all of them sets. In ZF all sets are pure.

### BIBLIOGRAPHY

The text of Church [1956] is a reliable and thorough source of information on anything that happened in first-order logic before the mid 1950s. The historical survey by Moore [1980] is also valuable.

- [Ackermann, 1962] W. Ackermann. *Solvable Cases of the Decision Problem*. North-Holland, Amsterdam, 1962.
- [Aczel, 1988] P. Aczel. *Non-well-founded Sets*. CSLI, Stanford CA, 1988.
- [Altham and Tennant, 1975] J. E. J. Altham and N. W. Tennant. Sortal quantification. In E. L. Keenan, editor, *Formal Semantics of Natural Language*, pages 46–58. Cambridge University Press, 1975.
- [Anderson and Johnstone Jr., 1962] J. M. Anderson and H. W. Johnstone Jr. *Natural Deduction: The Logical Basis of Axiom Systems*. Wadsworth, Belmont, CA., 1962.
- [Ax and Kochen, 1965] J. Ax and S. Kochen. Diophantine problems over local fields: I. *American Journal of Mathematics*, 87:605–630, 1965.
- [Barwise, 1973] J. Barwise. Abstract logics and  $L_{\infty\omega}$ . *Annals Math Logic*, 4:309–340, 1973.
- [Barwise, 1974] J. Barwise. Axioms for abstract model theory. *Annals Math Logic*, 7:221–265, 1974.
- [Barwise, 1975] J. Barwise. *Admissible Sets and Structures*. Springer, Berlin, 1975.
- [Barwise and Cooper, 1981] J. Barwise and R. Cooper. Generalized quantifiers and natural languages. *Linguistics and Philosophy*, 4:159–219, 1981.
- [Barwise and Etchemendy, 1991] J. Barwise and J. Etchemendy. *Tarski’s World 3.0*. Cambridge University Press, 1991.
- [Barwise and Etchemendy, 1994] J. Barwise and J. Etchemendy. *Hyperproof*. CSLI, Stanford, 1994.
- [Barwise and Moss, 1996] J. Barwise and L. Moss. *Vicious Circles*. CSLI, Stanford CA, 1996.
- [Behmann, 1922] H. Behmann. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Math Annalen*, 86:163–229, 1922.
- [Bell and Machover, 1977] J. L. Bell and M. Machover. *A Course in Mathematical Logic*. North-Holland, Amsterdam, 1977.

- [Bell and Slomson, 1969] J. L. Bell and A. B. Slomson. *Models and Ultraproducts*. North-Holland, Amsterdam, 1969.
- [Belnap, 1962] N. D. Belnap. Tonk, plonk and plink. *Analysis*, 22:130–134, 1962. Reprinted in [Strawson, 1967, pp. 132–137].
- [Benacerraf and Putnam, 1983] P. Benacerraf and H. Putnam, editors. *Philosophy of Mathematics: Selected Readings*. Cambridge University Press, second edition, 1983.
- [Bernays, 1942] P. Bernays. Review of Max Steck, ‘Ein unbekannter Brief von Gottlob Frege über Hilberts erste Vorlesung über die Grundlagen der Geometrie’. *Journal of Symbolic Logic*, 7:92 f., 1942.
- [Bernays and Fraenkel, 1958] P. Bernays and A. A. Fraenkel. *Axiomatic Set Theory*. North-Holland, Amsterdam, 1958.
- [Beth, 1953] E. W. Beth. On Padoa’s method in the theory of definition. *Koninklijke Nederlandse Akad. van Wetensch*, 56 (ser. A, Math Sciences):330–339, 1953.
- [Beth, 1955] E. W. Beth. Semantic entailment and formal derivability. *Mededelingen der Koninklijke Nederlandse Akad. van Wetensch*, afd letterkunde 18, 1955. Reprinted in [Hintikka, 1969, pp. 9–41].
- [Beth, 1962] E. W. Beth. *Formal Methods*. Reidel, Dordrecht, 1962.
- [Bocheński, 1970] I. M. Bocheński. *A History of Formal Logic, translated by I. Thomas*. Chelsea Publishing Co, New York, 1970.
- [Bolzano, 1837] B. Bolzano. *Wissenschaftslehre*. 1837. Edited and translated by R. George as *Theory of Science*, UCLA Press, Berkeley and Los Angeles, 1972.
- [Boole, 1847] G. Boole. *The Mathematical Analysis of Logic*. Macmillan, Barclay and Macmillan, Cambridge, 1847. Also pp. 45–124 of George Boole, *Studies in Logic and Probability*, Open Court, La Salle, IL, 1952.
- [Boole, 1854] G. Boole. *An Investigation of the Laws of Thought*. Walton and Maberley, London, 1854. Republished by Open Court, La Salle, IL, 1952.
- [Boolos and Jeffrey, 1989] G. S. Boolos and R. C. Jeffrey. *Computability and Logic*. Cambridge University Press, Cambridge, 1989.
- [Boolos, 1979] G. Boolos. *The Unprovability of Consistency: An Essay in Modal Logic*. Cambridge University Press, 1979.
- [Boolos, 1993] G. Boolos. *The Logic of Provability*. Cambridge University Press, 1993.
- [Carnap, 1935] R. Carnap. Ein Gültigkeitskriterium für die Sätze der klassischen Mathematik. *Monatshefte Math und Phys*, 42:163–190, 1935.
- [Carnap, 1956] R. Carnap. *Meaning and Necessity*. University of Chicago Press, second edition, 1956.
- [Chang and Keisler, 1973] C. C. Chang and H. J. Keisler. *Model Theory*. North-Holland, Amsterdam, 1973.
- [Chastain, 1975] C. Chastain. Reference and context. In K. Gunderson, editor, *Minnesota Studies in the Philosophy of Science, VII, Language, Mind and Knowledge*, pages 194–269. University of Minnesota Press, MI, 1975.
- [Cherlin, 1976] G. Cherlin. *Model Theoretic Algebra: Selected Topics*, volume 521 of *Lecture Notes in Maths*. Springer, Berlin, 1976.
- [Church, 1936] A. Church. A note on the Entscheidungsproblem. *Journal of Symbolic Logic*, 1:40f, 101f, 1936.
- [Church, 1956] A. Church. *Introduction to Mathematical Logic, I*. Princeton University Press, Princeton, NJ, 1956.
- [Coffa, 1991] J. A. Coffa. *The Semantic Tradition from Kant to Carnap: To the Vienna Station*. Cambridge University Press, Cambridge, 1991.
- [Cohen, 1969] P. J. Cohen. Decision procedures for real and  $p$ -adic fields. *Comm Pure Appl Math*, 22:131–151, 1969.
- [Cohen, 1971] L. J. Cohen. Some remarks on Grice’s views about the logical particles of natural language. In Y. Bar-Hillel, editor, *Pragmatics of Natural Languages*, pages 60–68. Reidel, Dordrecht, 1971.
- [Cook, 1971] S. A. Cook. The complexity of theorem-proving procedures. In *Proceedings of the Third Annual ACM Symposium on Theory of Computing*, pages 151–158. ACM Press, NY, 1971.
- [Craig, 1957a] W. Craig. Linear reasoning. A new form of the Herbrand–Gentzen theorem. *Journal of Symbolic Logic*, 22:250–268, 1957.



- [Craig, 1957b] W. Craig. Three uses of the Herbrand–Gentzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic*, 22:269–285, 1957.
- [Dalen, 1980] D. van Dalen. *Logic and Structure*. Springer, Berlin, 1980.
- [Dedekind, 1888] R. Dedekind. *Was sind und was sollen die Zahlen?* Brunswick, 1888.
- [Dedekind, 1967] R. Dedekind. Letter to Keferstein, 1890. In J. Van Heijenoort, editor, *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*, pages 90–103. Harvard University Press, Cambridge, MA, 1967.
- [Došen and Schroeder-Heister, 1993] K. Došen and P. Schroeder-Heister, editors. *Substructural Logics*. Oxford University Press, Oxford, 1993.
- [Dowty *et al.*, 1981] D. Dowty, R. Wall, and S. Peters. *Introduction to Montague Semantics*. Reidel, Dordrecht, 1981.
- [Dummett, 1958/59] M. A. E. Dummett. Truth. *Proc Aristotelian Soc*, 59:141–162, 1958/59. Reprinted in [Strawson, 1967; pp. 49–68].
- [Dummett, 1973] M. A. E. Dummett. *Frege: Philosophy of Language*. Duckworth, London, 1973.
- [Dummett, 1975] M. A. E. Dummett. What is a theory of meaning? In Samuel Guttenplan, editor, *Mind and Language*. Clarendon Press, Oxford, 1975.
- [Dunn and Belnap, 1968] J. M. Dunn and N. D. Belnap. The substitution interpretation of the quantifiers. *Noûs*, 2:177–185, 1968.
- [Ebbinghaus and Flum, 1995] H.-D. Ebbinghaus and J. Flum. *Finite model theory*. Springer, Berlin, 1995.
- [Ehrenfeucht, 1960] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fundamenta Math*, 49:129–141, 1960.
- [Enderton, 1972] H. B. Enderton. *A Mathematical Introduction to Logic*. Academic Press, New York, 1972.
- [Etchemendy, 1990] J. Etchemendy. *The Concept of Logical Consequence*. Harvard University Press, Cambridge MA, 1990.
- [Evans, 1980] G. Evans. Pronouns. *Linguistic Inquiry*, 11:337–362, 1980.
- [Feferman, 1968a] S. Feferman. Lectures on proof theory. In *Proc Summer School of Logic, Leeds 1967*, Lecture Notes in Mathematics 70, pages 1–109. Springer, Berlin, 1968.
- [Feferman, 1968b] S. Feferman. Persistent and invariant formulas for outer extensions. *Compositio Math*, 20:29–52, 1968.
- [Feferman, 1969] S. Feferman. Set-theoretical foundations of category theory. In *Reports of the Midwest Category Seminar III*, Lecture Notes in Mathematics 106, pages 201–247. Springer, Berlin, 1969.
- [Feferman, 1974] S. Feferman. Applications of many-sorted interpolation theorems. In L. Henkin *et al.*, editor, *Proceedings of the Tarski Symposium, Proc Symposia in Pure Math. XXV*, pages 205–223. American Mathematical Society, Providence, RI, 1974.
- [Feferman, 1984] S. Feferman. Kurt Gödel: conviction and caution. *Philosophia Naturalis*, 21:546–562, 1984.
- [Fitch, 1952] F. B. Fitch. *Symbolic Logic*. Ronald Press, New York, 1952.
- [Flum, 1975] J. Flum. First-order logic and its extensions. In *ISILC Logic Conference*, Lecture Notes in Mathematics 499, pages 248–307. Springer, Berlin, 1975.
- [Fraenkel, 1922] A. Fraenkel. Zu den Grundlagen der Cantor–Zermeloschen Mengenlehre. *Math Annalen*, 86:230–237, 1922.
- [Fraïssé, 1954] R. Fraïssé. Sur l’extension aux relations de quelques propriétés des ordres. *Ann Sci École Norm Sup*, 71:363–388, 1954.
- [Fraïssé, 1955] R. Fraïssé. Sur quelques classifications des relations, basées sur des isomorphismes restreints. *Alger-Mathématiques*, 2:16–60 and 273–295, 1955.
- [Frege, 1879] G. Frege. *Begriffsschrift*. Halle, 1879. Translated in [Heijenoort, 1967, pp. 1–82].
- [Frege, 1884] G. Frege. *Die Grundlagen der Arithmetik*. Breslau, 1884. Translated by J. L. Austin, *The Foundations of Arithmetic*, 2nd edn., Blackwell, Oxford, 1953.
- [Frege, 1891] G. Frege. *Funktion und Begriff*. Jena, 1891. Also in [Frege, 1967, pp. 125–142] and translated in [Frege, 1952].

- [Frege, 1893] G. Frege. *Grundgesetze der Arithmetik I*. Jena, 1893. Partial translation with introduction by M. Furth, *The Basic Laws of Arithmetic*, University California Press, Berkeley, 1964.
- [Frege, 1906] G. Frege. Über die Grundlagen der Geometrie. *Jahresbericht der Deutschen Mathematiker-Vereinigung*, 15:293–309, 377–403 and 423–430, 1906. Translated in [Frege, 1971].
- [Frege, 1912] G. Frege. *Anmerkungen zu: Philip E. B. Jourdain*. The development of the theories of mathematical logic and the principles of mathematics, 1912. In [Frege, 1967, pp. 334–341].
- [Frege, 1952] G. Frege. *Translations from the Philosophical Writings of Gottlob Frege*. Blackwell, Oxford, 1952.
- [Frege, 1967] G. Frege. *Kleine Schriften*. Georg Olms Verlagsbuchhandlung, Hildesheim, 1967.
- [Frege, 1971] G. Frege. *On the Foundations of Geometry and Formal Theories of Arithmetic*. Yale University Press, New Haven, 1971. Translated with introduction by E. W. Kluge.
- [Frege and Hilbert, 1899–1900] G. Frege and D. Hilbert. Correspondence leading to ‘On the foundations of geometry’, 1899–1900. In [Frege, 1967; pp. 407–418], translated in [Frege, 1971; pp. 6–21].
- [Gallier, 1986] J. H. Gallier. *Logic for Computer Science: foundations of Automatic Theorem Proving*. Harper and Row, 1986.
- [Gandy, 1974] R. O. Gandy. Set-theoretic functions for elementary syntax. In T. J. Jech, editor, *Axiomatic Set Theory II*, pages 103–126. American Mathematical Society, Providence, RI, 1974.
- [Garey and Johnson, 1979] M. R. Garey and D. S. Johnson. *Computers and Intractability*. W. H. Freeman, San Francisco, 1979.
- [Gentzen, 1934] G. Gentzen. Untersuchungen über das logische Schliessen. *Math Zeitschrift*, 39:176–210 and 405–431, 1934.
- [Girard, 1987] J.-Y. Girard. Linear logic. *Theoretical Computer Science*, 50:1–102, 1987.
- [Girard, 1995] J.-Y. Girard. Linear logic: its syntax and semantics. In J.-Y. Girard et al., editor, *Advances in Linear Logic*, pages 1–42. Cambridge University Press, 1995.
- [Gödel, 1930] K. Gödel. Die Vollständigkeit der Axiome des logischen Funktionenkalküls. *Monatshefte für Mathematik und Physik*, 37:349–360, 1930. Translated in [Gödel, 1986, pp. 102–123] and [Heijenoort, 1967, pp. 582–591].
- [Gödel, 1931a] K. Gödel. Eine Eigenschaft der Realisierungen des Aussagenkalküls. *Ergebnisse Math Kolloq*, 3:20–21, 1931. Translated in [Gödel, 1986, pp. 238–241].
- [Gödel, 1931b] K. Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. Translated in [Gödel, 1986, pp. 144–195] and [Heijenoort, 1967, pp. 596–616].
- [Gödel, 1947] K. Gödel. What is Cantor’s continuum problem? *American Mathematical Monthly*, 54:515–525, 1947. Revised and expanded version in [Gödel, 1990, pp. 254–270].
- [Gödel, 1951] K. Gödel. Russell’s mathematical logic. In P. A. Schilpp, editor, *The Philosophy of Bertrand Russell*, pages pp. 123–153. Tudor Publ. Co, New York, 1951. Also in [Gödel, 1990, pp. 119–141].
- [Gödel, 1986] K. Gödel. *Collected Works. Volume I*. Oxford University Press, New York, 1986. Edited by S. Feferman et al.
- [Gödel, 1990] K. Gödel. *Collected Works. Volume II*. Oxford University Press, New York, 1990. Edited by S. Feferman et al.
- [Goldblatt, 1982] R. Goldblatt. *Axiomatizing the Logic of Computer Programming*. Lecture Notes in Computer Science, 130, Springer, Berlin, 1982.
- [Goldfarb, 1979] W. D. Goldfarb. Logic in the twenties: the nature of the quantifier. *Journal of Symbolic Logic*, 44:351–368, 1979.
- [Goldson, Reeves and Bornat, 1993] D. Goldson, S. Reeves and R. Bornat. A review of several programs for the teaching of logic, *Computer Journal*, 36:373–386, 1993.
- [Gómez-Torrente, 1996] M. Gómez-Torrente. Tarski on logical consequence, *Notre Dame Journal of Formal Logic*, 37:125–151, 1996.

- [Grice, 1975] H. P. Grice. Logic and conversation. In P. Cole *et al.*, editor, *Syntax and Semantics 3, Speech Acts*, pp. 41–58. Academic Press, New York, 1975. Revised version in P. Grice, *Studies in the Way of Words*, Harvard University Press, Cambridge, MA, 1989, pp. 22–40.
- [Groenendijk and Stokhof, 1991] J. Groenendijk and M. Stokhof. Dynamic predicate logic, *Linguistics and Philosophy*, 14:39–100, 1991.
- [Gurevich, 1984] Y. Gurevich. Toward logic tailored for computational complexity. In M. M. Richter, *et al.*, editors, *Computation and Proof Theory*, Lecture Notes in Mathematics 1104, pp. 175–216, Springer-Verlag, 1984.
- [Hammer, 1995] E. M. Hammer. *Logic and Visual Information*. CSLI and FoLLI, Stanford CA, 1995.
- [Harel, 1979] D. Harel. *First-order Dynamic Logic*. Lecture Notes in Computer Science, 68. Springer, Berlin, 1979.
- [Harnik, 1985] V. Harnik. Stability theory and set existence axioms. *Journal of Symbolic Logic*, 50:123–137, 1985.
- [Harnik, 1987] V. Harnik. Set existence axioms for general (not necessarily countable) stability theory. *Annals of Pure and Applied Logic*, 34:231–243, 1987.
- [Hasenjaeger, 1953] G. Hasenjaeger. Eine Bemerkung zu Henkins Beweis für die Vollständigkeit des Prädikatenkalküls der ersten Stufe. *Journal of Symbolic Logic*, 18:42–48, 1953.
- [Hausdorff, 1914] F. Hausdorff. *Grundzüge der Mengenlehre*. Veit, Leipzig, 1914.
- [Heijenoort, 1967] J. van Heijenoort, editor. *From Frege to Gödel, A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press, Cambridge, MA, 1967.
- [Heim, 1988] I. Heim. *The Semantics of Definite and Indefinite Noun Phrases in English*, Garland, New York, 1988.
- [Henkin, 1949] L. Henkin. The completeness of the first-order functional calculus. *Journal of Symbolic Logic*, 14:159–166, 1949. Reprinted in [Hintikka, 1969].
- [Henkin, 1950] L. Henkin. Completeness in the theory of types. *J. Symbolic Logic*, 15:81–91, 1950. Reprinted in [Hintikka, 1969].
- [Henkin, 1961] L. Henkin. Some remarks on infinitely long formulas. In *Infinitistic Methods: Proc. Symp. on Foundations of Mathematics, Warsaw*, pages 167–183. Pergamon, London, 1961.
- [Henkin and Mostowski, 1959] L. Henkin and A. Mostowski. Review of Mal'tsev [1941]. *Journal of Symbolic Logic*, 24:55–57, 1959.
- [Herbrand, 1930] J. Herbrand. *Recherches sur la théorie de la démonstration*. PhD thesis, University of Paris, 1930. Translated in [Herbrand, 1971, pp. 44–202].
- [Herbrand, 1971] J. Herbrand. *Logical Writings*. Harvard University Press, Cambridge, MA, 1971. Edited by W. D. Goldfarb.
- [Hilbert, 1899] D. Hilbert. *Grundlagen der Geometrie*. Teubner, Leipzig, 1899.
- [Hilbert, 1923] D. Hilbert. Die logischen Grundlagen der Mathematik. *Math Annalen*, 88:151–165, 1923. Also in [Hilbert, 1970, pp. 178–195].
- [Hilbert, 1926] D. Hilbert. Über das Unendliche. *Math Annalen*, 95:161–190, 1926. Translated in [Heijenoort, 1967, pp. 367–392]; partial translation in [Benacerraf and Putnam, 1983, pp. 183–201].
- [Hilbert, 1928] D. Hilbert. Die Grundlagen der Mathematik. *Abhandlungen aus dem Math. Seminar der Hamburgischen Universität*, 6:65–85, 1928. Translated in [Heijenoort, 1967, pp. 464–479].
- [Hilbert, 1970] D. Hilbert. *Gesammelte Abhandlungen III: Analysis, Grundlagen der Mathematik, Physik, Verschiedenes*. Springer, Berlin, 1970.
- [Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*. Springer, Berlin, 1928.
- [Hilbert and Bernays, 1934] D. Hilbert and P. Bernays. *Grundlagen der Mathematik I*. Springer, Berlin, 1934.
- [Hilbert and Bernays, 1939] D. Hilbert and P. Bernays. *Grundlagen der Mathematik II*. Springer, Berlin, 1939.
- [Hintikka, 1953] J. Hintikka. Distributive normal forms in the calculus of predicates. *Acta Philosophica Fennica*, 6, 1953.

- [Hintikka, 1955] J. Hintikka. Form and content in quantification theory. *Acta Philosophica Fennica*, 8:7–55, 1955.
- [Hintikka, 1969] J. Hintikka, editor. *The Philosophy of Mathematics*. Oxford University Press, 1969.
- [Hintikka, 1973] J. Hintikka. *Logic, Language-games and Information*. Oxford University Press, 1973.
- [Hintikka, 1996] J. Hintikka. *The Principles of Mathematics Revisited*, Cambridge University Press, Cambridge, 1996.
- [Hodges, 1972] W. Hodges. On order-types of models. *Journal of Symbolic Logic*, 37:69f, 1972.
- [Hodges, 1977] W. Hodges. *Logic*. Penguin Books, Harmondsworth, Middx, 1977.
- [Hodges, 1985/86] W. Hodges. Truth in a structure, *Proceedings of Aristotelian Society*, 86:135–151, 1985/6.
- [Hodges, 1993a] W. Hodges. *Model Theory*, Cambridge University Press, Cambridge, 1993.
- [Hodges, 1993b] W. Hodges. Logical features of Horn clauses. In *Handbook of Logic in Artificial Intelligence and Logic Programming, Volume 1: Logical Foundations*, D. M. Gabbay, C. J. Hogger and J. A. Robinson, editors. pages 449–503. Clarendon Press, Oxford, 1993.
- [Hodges, 1997a] W. Hodges. *A Shorter Model Theory*, Cambridge University Press, Cambridge, 1997.
- [Hodges, 1997b] W. Hodges. Compositional semantics for a language of imperfect information, *Logic Journal of the IGPL*, 5:539–563, 1997.
- [Huntington, 1904] E. V. Huntington. *The Continuum and Other Types of Serial Order, with an Introduction to Cantor's Transfinite Numbers*. Harvard University Press, Cambridge, MA, 1904.
- [Jeffrey, 1967] R. C. Jeffrey. *Formal Logic: its Scope and Limits*. McGraw-Hill, New York, 1967.
- [Johnson-Laird and Byrne, 1991] P. N. Johnson-Laird and R. M. J. Byrne. *Deduction*. Lawrence Erlbaum Associates, Hove, 1991.
- [Johnstone, 1977] P. T. Johnstone. *Topos Theory*. Academic Press, London, 1977.
- [Kalish and Montague, 1964] D. Kalish and R. Montague, *Logic: Techniques of Formal Reasoning*. Harcourt, Brace and World, New York, 1964.
- [Kalmár, 1934/5] L. Kalmár. Über die Axiomatisierbarkeit des Aussagenkalküls. *Acta Scient. Math. Szeged*, 7:222–243, 1934/5.
- [Kamp, 1971] H. Kamp. Formal properties of ‘Now’. *Theoria*, 37:227–273, 1971.
- [Kamp, 1981] H. Kamp. A theory of truth and semantic representation. In J. A. G. Groenendijk *et al.*, editor, *Formal Methods in the Study of Language*, pages 277–322. Math Centrum, Amsterdam, 1981.
- [Kamp and Reyle, 1993] H. Kamp and U. Reyle. *From Discourse to Logic*, Kluwer, Dordrecht, 1993.
- [Kaplan, 1966] D. Kaplan. What is Russell’s theory of descriptions? In *Proceedings of Internat Colloquium on Logic, Physical Reality and History, Denver, 1966*, pages 227–244. Plenum, New York, 1966. Reprinted in [Pears, 1972, pp. 227–244].
- [Karp, 1965] C. Karp. Finite quantifier equivalence. In J. Addison *et al.*, editor, *The Theory of Models*. North-Holland, Amsterdam, 1965.
- [Kempson, 1995] R. Kempson, editor. *Bulletin of the IGPL*, volume 3 numbers 2, 3: Special Issue on Deduction and Language, 1995.
- [Kleene, 1943] S.C. Kleene. Recursive predicates and quantifiers. *Trans Amer Math Soc*, 53:41–73, 1943.
- [Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics*. North-Holland, Amsterdam, 1952.
- [Klenk, 1976] V. Klenk. Intended models and the Löwenheim–Skolem theorem. *J. Philos. Logic*, 5:475–489, 1976.
- [Kneale, 1956] W. Kneale. The province of logic. In H. D. Lewis, editor, *Contemporary British Philosophy, 3rd Series*, pages 237–261. George Allen and Unwin, London, 1956.

- [Kowalski, 1979] R. Kowalski. *Logic for problem solving*, North-Holland, New York, 1979.
- [Kreisel, 1967] G. Kreisel. Informal rigour and completeness proofs. In Lakatos, editor, *Problems in the Philosophy of Mathematics*, pages 138–157. North-Holland, Amsterdam, 1967. Partially reprinted in [Hintikka, 1969, pp. 78–94].
- [Kreisel and Krivine, 1967] G. Kreisel and J. L. Krivine. *Elements of Mathematical Logic (Model Theory)*. North-Holland, Amsterdam, 1967.
- [Kripke, 1976] S. Kripke. Is there a problem about substitutional quantification? In G. Evans and J. McDowell, editors, *Truth and Meaning: Essays in Semantics*, pages 325–419. Clarendon Press, Oxford, 1976.
- [Kronecker, 1882] L. Kronecker. Grundzüge einer arithmetischen Theorie der algebraischen Grössen. *Crelle's Journal*, 92:1–122, 1882.
- [Lakoff, 1972] G. Lakoff. Linguistics and natural logic. In D. Davidson and G. Harman, editors, *Semantics of Natural Languages*, pages 545–665. Reidel, Dordrecht, 1972.
- [Langford, 1927] C. H. Langford. Some theorems on deducibility. *Annals of Math*, 28:16–40, 1927.
- [Leisenring, 1969] A. C. Leisenring. *Mathematical Logic and Hilbert's  $\epsilon$ -symbol*. Gordon and Breach, New York, 1969.
- [Lemmon, 1965] E. J. Lemmon. *Beginning Logic*. Nelson, London, 1965.
- [Levy, 1965] A. Levy. A hierarchy of formulas in set theory. *Memoirs of the American Mathematical Society*, 57, 1965.
- [Levy, 1979] A. Levy. *Basic Set Theory*. Springer, New York, 1979.
- [Lindström, 1969] P. Lindström. On extensions of elementary logic. *Theoria*, 35:1–11, 1969.
- [Lorenzen, 1961] P. Lorenzen. Ein dialogisches Konstruktivitätskriterium. In *Infinitistic Methods, Proc of a Symp on Foundations of Mathematics, Warsaw*, pages 193–200, Pergamon, London, 1961.
- [Lorenzen, 1962] P. Lorenzen. *Metamathematik*. Bibliographisches Institut, Mannheim, 1962.
- [Lorenzen and Schwemmer, 1975] P. Lorenzen and O. Schwemmer. *Konstruktive Logik, Ethik und Wissenschaftstheorie*. Bibliographisches Institut, Mannheim, 1975.
- [Löwenheim, 1915] L. Löwenheim. Über Möglichkeiten im Relativkalkül. *Math Annalen*, 76:447–470, 1915. Translated in [Heijenoort, 1967, pp. 228–251].
- [Lukasiewicz and Tarski, 1930] J. Lukasiewicz and A. Tarski. Untersuchungen über den Aussagenkalkül. *Comptes Rendus des séances de la Société des Sciences et des Lettres de Varsovie*, 23 cl. iii:30–50, 1930. Translated in [Tarski, 1983, pp. 38–59].
- [Mal'tsev, 1936] A. I. Mal'tsev. Untersuchungen aus dem Gebiete der Mathematischen Logik. *Mat Sbornik*, 1:323–336, 1936. Translated in [Mal'tsev, 1971, pp. 1–14].
- [Mal'tsev, 1941] A. I. Mal'tsev. On a general method for obtaining local theorems in group theory (Russian). *Ivanov Gos. Ped. Inst. Uc. Zap. Fiz.-Mat. Fak.*, 1:3–9, 1941. Translated in [Mal'tsev, 1971, pp. 15–21].
- [Mal'tsev, 1971] A. I. Mal'tsev. *The Metamathematics of Algebraic Systems; Collected Papers 1936–1967*. North-Holland, Amsterdam, 1971. Translated and edited by B. F. Wells III.
- [Manktelow and Over, 1990] K. I. Manktelow and D. E. Over. *Inference and Understanding*, Routledge, London, 1990.
- [Mates, 1965] B. Mates. *Elementary Logic*. Oxford University Press, New York, 1965.
- [Members of the Johns Hopkins University, Boston, 1883] Members of the Johns Hopkins University, Boston. *Studies in Logic*. Little, Brown and Co, 1883.
- [Mendelson, 1987] E. Mendelson. *Introduction to Mathematical Logic*, Third edition. Van Nostrand, Princeton, NJ, 1964.
- [Mitchell, 1883] O. H. Mitchell. On a new algebra of logic. In Members of the Johns Hopkins University, Boston, *Studies in Logic*, pages 72–106. Little, Brown and Co, 1883.
- [Montague, 1970] R. Montague. English as a formal language. In B. Visentini *et al.*, editor, *Linguaggi nella Società e nella Tecnica*. Milan, 1970. Also in [Montague, 1974, pp. 188–221].

- [Montague, 1973] R. Montague. The proper treatment of quantification in ordinary English. In J. Hintikka *et al.*, editor, *Approaches to Natural Language*. Reidel, Dordrecht, 1973. Also in [Montague, 1974, pp. 247–270].
- [Montague, 1974] R. H. Thomason, editor. *Formal Philosophy, Selected Papers of Richard Montague*, Yale University Press, New Haven, 1974.
- [Montague and Vaught, 1959] R. Montague and R. L. Vaught. Natural models of set theory. *Fundamenta Math*, 47:219–242, 1959.
- [Moore, 1980] G. H. Moore. Beyond first-order logic: the historical interplay between mathematical logic and axiomatic set theory. *History and Philosophy of Logic*, 1:95–137, 1980.
- [Morrill, 1994] G. V. Morrill. *Type Logical Grammar: Categorical Logic of Signs*. Kluwer, Dordrecht, 1994.
- [Nisbett *et al.*, 1987] R. E. Nisbett, G. T. Fong, D. R. Lehman and P. W. Cheng. Teaching reasoning. *Science*, 238:625–631, 1987.
- [Padawitz, 1988] P. Padawitz. *Computing in Horn Clause Theories*. Springer, Berlin, 1988.
- [Partee, 1978] B. Partee. Bound variables and other anaphors. In D. Waltz, editor, *Tinlap-2, Theoretical Issues in Natural Language Processing*, pages 248–280. Association for Computing Machinery, New York, 1978.
- [Peano, 1889] G. Peano. *Arithmetices Principia, Nova Methodo Exposita*. Turin, 1889. Translation in [Heijenoort, 1967, pp. 85–97].
- [Pears, 1972] D. F. Pears, editor. *Bertrand Russell. A Collection of Critical Essays*. Anchor Books, Doubleday, New York, 1972.
- [Peirce, 1883] C. S. Peirce. A theory of probable inference. Note B. The logic of relatives. In Boston Members of the Johns Hopkins University, editor, *Studies in Logic*. Little, Brown and Co, 1883. Reprinted in [Peirce, 1933, Vol III, pp. 195–209].
- [Peirce, 1885] C. S. Peirce. On the algebra of logic. *Amer. J. Math.*, 7:180–202, 1885. Reprinted in [Peirce, 1933, Vol. III, pp. 210–238].
- [Peirce, 1902] C. S. Peirce. The simplest mathematics. In C. Hartshorne *et al.*, editor, *Collected Papers of Charles Sanders Peirce*, volume IV, pages 189–262. Harvard University Press, Cambridge, MA, 1902.
- [Peirce, 1933] C. S. Peirce. In C. Hartshorne *et al.*, editor, *Collected Papers of Charles Sanders Peirce*. Harvard University Press, Cambridge, MA, 1933.
- [Perry, 1977] J. Perry. Frege on demonstratives. *Philosophical Review*, 86:474–497, 1977. Reprinted in P. Yourgram, editor, *Demonstratives*, pages 50–70, Oxford University Press, New York, 1990.
- [Popper, 1946/47] K. R. Popper. Logic without assumptions. *Proc. Aristot. Soc*, pages 251–292, 1946/47.
- [Post, 1921] E. Post. Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, 43:163–185, 1921. Reprinted in [Heijenoort, 1967, pp. 264–283].
- [Prawitz, 1965] D. Prawitz. *Natural Deduction: a Proof-theoretical Study*. Almqvist and Wiksell, Stockholm, 1965.
- [Prawitz, 1979] D. Prawitz. Proofs and the meaning and the completeness of the logical constants. In J. Hintikka, I. Niiniluoto, and E. Saarinen, editors, *Essays on Mathematical and Philosophical Logic*, pages 25–40. Reidel, Dordrecht, 1979.
- [Prior, 1960] A. N. Prior. The runabout inference ticket. *Analysis*, 21:38–39, 1960. Reprinted in [Strawson, 1967, pp. 129–131].
- [Prior, 1962] A. N. Prior. *Formal Logic*. Oxford University Press, 1962.
- [Putnam, 1980] H. Putnam. Models and reality. *Journal of Symbolic Logic*, 45:464–482, 1980. Reprinted in [Benacerraf, 1983; pp. 421–444].
- [Quine, 1940] W. V. Quine. *Mathematical Logic*. Harvard University Press, Cambridge, MA, 1940. Revised edition 1951.
- [Quine, 1950] W. V. Quine. *Methods of Logic*. Holt, New York, 1950.
- [Quine, 1970] W. V. Quine. *Philosophy of Logic*. Prentice-Hall, Englewood Cliffs, NJ, 1970.
- [Quirk and Greenbaum, 1973] R. Quirk and S. Greenbaum. *A University Grammar of English*. Longman, London, 1973.

- [Rasiowa and Sikorski, 1950] H. Rasiowa and R. Sikorski. A proof of the completeness theorem of Gödel. *Fundamenta Math*, 37:193–200, 1950.
- [Rasiowa and Sikorski, 1963] H. Rasiowa and R. Sikorski. *The Mathematics of Meta-mathematics*. Monografie Matematyczne, Polska Akad. Nauk, 1963.
- [Reeves and Clarke, 1990] S. Reeves and M. Clarke. *Logic for Computer Science*. Addison-Wesley, 1990.
- [Rips, 1994] L. J. Rips. *The Psychology of Proof*. MIT Press, Cambridge Mass., 1994.
- [Robinson, 1967] A. Robinson. The metaphysics of the calculus. In Lakatos, editor, *Problems in the Philosophy of Mathematics*, pages 28–40. North-Holland, Amsterdam, 1967. Reprinted in [Hintikka, 1969, pp. 153–163], and in Selected papers of Abraham Robinson, Vol. 2, edited by H. J. Keisler *et al.*, pp. 537–555. Yale University Press, New Haven, 1979.
- [Russell, 1905] B. Russell. On denoting. *Mind*, 14:479–493, 1905. Reprinted in [Russell, 1956].
- [Russell, 1956] B. Russell. In R. C. Marsh, editor, *Logic and Knowledge, Essays 1901–1950*. George Allen and Unwin, London, 1956.
- [Sacks, 1972] G. E. Sacks. *Saturated Model Theory*. Benjamin, Reading, MA, 1972.
- [Schmidt, 1938] H. A. Schmidt. Über deduktive Theorien mit mehreren Sorten von Grunddingen. *Math Annalen*, 115:485–506, 1938.
- [Schröder, 1895] E. Schröder. *Vorlesungen über die Algebra der Logik*, volume 3. Leipzig, 1895.
- [Schütte, 1956] K. Schütte. Ein System des verknüpfenden Schliessens. *Arch. Math. Logik Grundlagenforschung*, 2:55–67, 1956.
- [Schütte, 1977] K. Schütte. *Proof Theory*. Springer, Berlin, 1977. Translated by J. N. Crossley.
- [Scott, 1964] W. R. Scott. *Group Theory*. Prentice-Hall, Englewood Cliffs, NJ, 1964.
- [Shoemith and Smiley, 1978] D. J. Shoemith and T. J. Smiley. *Multiple-Conclusion Logic*. Cambridge University Press, 1978.
- [Skolem, 1919] T. Skolem. Untersuchungen über die Axiome des Klassenkalküls und über Produktations- und Summationsprobleme, welche gewisse von Aussagen betreffen. *Videnskapsselskapets Skrifter, I. Matem.-naturv. klasse, no 3*, 1919. Reprinted in [Skolem, 1970, pp. 67–101].
- [Skolem, 1920] T. Skolem. Logisch-kombinatorische Untersuchungen über die Erfüllbarkeit oder Beweisbarkeit mathematischer Sätze nebst einem Theoreme über dichte Mengen. *Videnskapsselskapets Skrifter, I. Matem.-Naturv. Klasse 4*, 1920. Reprinted in [Skolem, 1970, pp. 103–136]; partial translation in [Heijenoort, 1967, pp. 252–263].
- [Skolem, 1922] T. Skolem. Einige Bemerkungen zur axiomatischen Begründung der Mengenlehre. *Matematikerkongressen i Helsingfors den 4–7 Juli 1922*, 1922. Reprinted in [Skolem, 1970, pp. 137–152]; translation in [Heijenoort, 1967, pp. 290–301].
- [Skolem, 1923] T. Skolem. Begründung der elementaren Arithmetik durch die rekurrerende Denkweise ohne Anwendung scheinbarer Veränderlichen mit unendlichem Ausdehnungsbereich. *Videnskapsselskapets Skrifter I, Matem.-naturv. Klasse 6*, 1923. Translation in [Heijenoort, 1967, pp. 303–333].
- [Skolem, 1928] T. Skolem. Über die mathematische Logik. *Norsk. Mat. Tidssk.*, 10:125–142, 1928. Reprinted in [Skolem, 1970, pp. 189–206]; translation in [Heijenoort, 1967, pp. 513–524].
- [Skolem, 1929] T. Skolem. Über einige Grundlagenfragen der Mathematik. *Skr. Norsk. Akad. Oslo I Mat.-Natur Kl 4*, pages 1–49, 1929. Reprinted in [Skolem, 1970, pp. 227–273].
- [Skolem, 1934] T. Skolem. Über die Nichtcharakterisierbarkeit der Zahlenreihe mittels endlich oder abzählbar unendlich vieler Aussagen mit ausschliesslich Zahlenvariablen. *Fundamenta Math*, 23:150–161, 1934. Reprinted in [Skolem, 1970, pp. 355–366].
- [Skolem, 1955] T. Skolem. A critical remark on foundational research. *Kongelige Norsk. Vidensk. Forhand. Trondheim*, 28:100–105, 1955. Reprinted in [Skolem, 1970, pp. 581–586].
- [Skolem, 1970] T. Skolem. In *Selected Works in Logic*. J. E. Fenstad, editor, Universitetsforlaget, Oslo, 1970.

- [Smullyan, 1968] R. Smullyan. *First-Order Logic*. Springer, Berlin, 1968.
- [Sneed, 1971] J. D. Sneed. *The Logical Structure of Mathematical Physics*. Reidel, Dordrecht, 1971.
- [Stegmüller, 1976] W. Stegmüller. *The Structure and Dynamics of Theories*. Springer, New York, 1976.
- [Steiner, 1975] M. Steiner. *Mathematical Knowledge*. Cornell University Press, Ithaca, 1975.
- [Stenning *et al.*, 1995] K. Stenning, R. Cox and J. Oberlander. Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences, *Language and Cognitive Processes*, 10:333–354, 1995.
- [Stevenson, 1973] L. Stevenson. Frege's two definitions of quantification. *Philos. Quarterly*, 23:207–223, 1973.
- [Strawson, 1967] P. F. Strawson, editor. *Philosophical Logic*. Oxford University Press, 1967.
- [Suppes, 1957] P. Suppes. *Introduction to Logic*. Van Nostrand, Princeton, NJ, 1957.
- [Suppes, 1972] P. Suppes. *Axiomatic Set Theory*. Dover, NY, 1972.
- [Tarski, 1935] A. Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen, based on a paper in *Ruch. Filozoficzny xii (1930/1)*, 1935. Translated in [Tarski, 1983, pp. 152–278].
- [Tarski, 1936] A. Tarski. O pojęciu wynikania logicznego. *Przegląd Filozoficzny*, 39:58–68, 1936. Translated as 'On the concept of logical consequence' in [Tarski, 1983, pp. 409–420].
- [Tarski, 1954] A. Tarski. Contributions to the theory of models I, II. *Indag. Math.*, 16:572–588, 1954.
- [Tarski, 1983] A. Tarski. *Logic, Semantics, Metamathematics, Papers from 1923 to 1938*. Hackett, Indianapolis, 1983. Translated by J. H. Woodger with analytical index by J. Corcoran.
- [Tarski and Givant, 1987] A. Tarski and S. Givant. *A Formalization of Set Theory without Variables*. American Mathematical Society, Providence RI, 1987.
- [Tarski and Vaught, 1956] A. Tarski and R. L. Vaught. Arithmetical extensions of relational systems. *Compositio Math*, 13:81–102, 1956.
- [Tarski *et al.*, 1953] A. Tarski, A. Mostowski, and R. M. Robinson. *Undecidable Theories*. North-Holland, Amsterdam, 1953.
- [Tennant, 1978] N. W. Tennant. *Natural Logic*. Edinburgh University Press, 1978.
- [Thomason, 1970] R. H. Thomason. *Symbolic Logic, An Introduction*. Macmillan, London, 1970.
- [Vaught, 1974] R. L. Vaught. Model theory before 1945. In L. Henkin *et al.*, editor, *Proceedings of the Tarski Symposium*, pages 153–172. AMS, Providence, RI, 1974.
- [von Neumann, 1925] J. Von Neumann. Eine Axiomatisierung der Mengenlehre. *J. für die Reine und Angew Math*, 154:219–240, 1925. Translated in [Heijenoort, 1967, pp. 393–413].
- [Wang, 1952] H. Wang. Logic of many-sorted theories. *Journal of Symbolic Logic*, 17:105–116, 1952.
- [Wang, 1970] H. Wang. A survey of Skolem's work in logic, 1970. In [Skolem, 1970, pp. 17–52].
- [Wang, 1974] H. Wang. *From Mathematics to Philosophy*. Routledge and Kegan Paul, NY, 1974.
- [Wason, 1966] P. C. Wason. Reasoning. In *New Horizons in Psychology*, B. Foss, ed., pages 135–151. Penguin, Harmondsworth, 1966.
- [Whitehead and Russell, 1910] A. N. Whitehead and B. Russell. *Principia Mathematica I*. Cambridge University Press, 1910. Up to to 56\*, reprinted 1962.
- [Wiredu, 1973] J. E. Wiredu. Deducibility and inferability. *Mind*, 82:31–55, 1973.
- [Wittgenstein, 1910] L. Wittgenstein. *Tractatus Logico-Philosophicus*. Annalen der Naturphilosophie, 1910. Reprinted with translation by D. F. Pears and B. F. McGuinness, Routledge and Kegan Paul, London, 1961.
- [Zermelo, 1908] E. Zermelo. Untersuchungen über die Grundlagen der Mengenlehre I. *Math Annalen*, 65:261–281, 1908. Translated in [Heijenoort, 1967, pp. 199–215].



[Zucker, 1974] J. Zucker. The correspondence between cut-elimination and normalisation. *Annals of Math Logic*, 7:1–156, 1974.



STEWART SHAPIRO

## SYSTEMS BETWEEN FIRST-ORDER AND SECOND-ORDER LOGICS

### 1 WHY?

The most common logical system taught, used, and studied today is *Elementary predicate logic*, otherwise known as *first-order logic* (see Hodges' chapter in this Volume). First-order logic has a well-studied proof theory and model theory, and it enjoys a number of interesting properties. There is a recursively-enumerable deductive system D1 such that any first-order sentence  $\Phi$  is a consequence of a set  $\Gamma$  of first-order sentences if and only if  $\Phi$  is deducible from  $\Gamma$  in D1. Thus, first-order logic is (strongly) *complete*. It follows that first-order logic is *compact* in the sense that if every finite subset of a set  $\Gamma$  of first-order sentences is satisfiable then  $\Gamma$  itself is satisfiable. The downward Löwenheim–Skolem theorem is that if a set  $\Gamma$  of first-order sentences is satisfiable, then it has a model whose domain is countable (or the cardinality of  $\Gamma$ , whichever is larger). The upward Löwenheim–Skolem theorem is that if a set  $\Gamma$  of first-order sentences has, for each natural number  $n$ , a model whose domain has at least  $n$  elements, then for any infinite cardinal  $\kappa$ ,  $\Gamma$  has a model whose domain is of size at least  $\kappa$  (see Hodges' chapter, and virtually any textbook in mathematical logic, such as Boolos and Jeffrey [1989] or Mendelson [1987]).

Since many arguments in both everyday discourse and mathematics have natural renderings in first-order languages, first-order logic is a good tool to begin the study of validity. First-order languages also capture important features of the semantics of natural language, and so first-order logic is a tool for the study of natural language. However, first-order languages suffer from expressive poverty. It is an easy consequence of compactness that many central concepts—such as finitude, countability, minimal closure, well-foundedness, and well-order—cannot be captured in a first-order language. The Löwenheim–Skolem theorems entail that no infinite structure can be characterized up to isomorphism in a first-order language. Moreover, many important linguistic locutions, distinctions, and constructions fall outside the scope of first-order logic (see van Benthem and Doets' chapter below and Shapiro [1991, Chapter 5]).

The main alternative to first-order logic is *second-order logic* (and *higher-order logic* generally). The aforementioned mathematical notions that lack first-order characterizations all have adequate characterizations in second-order languages. For example, there is a second-order formula  $\mathbf{FIN}(X)$  that is satisfied in a structure if and only if the set assigned to  $X$  is finite. Also,

basic infinite mathematical structures have categorical characterizations in second-order languages. Examples include the natural numbers, the real numbers, Euclidean space, and some initial segments of the set-theoretic hierarchy. Second-order languages, and higher-order languages generally, allow the linguist to model many linguistic constructions that reach beyond first-order.

The expressive richness of second-order languages and logic carries a cost. It follows from the expressive power of second-order logic that it is not compact and the Löwenheim–Skolem theorems fail. Second-order logic is highly complex, and in some ways it is intractable. For example, let  $\text{AR}$  be a categorical characterization of the natural numbers. Then a sentence  $\Phi$  in the (first-order) language of arithmetic is true of the natural numbers if and only if  $\text{AR} \rightarrow \phi$  is a logical truth. Thus, the notion of arithmetic truth is reducible to second-order logical truth. Similarly, the notion of ‘truth of analysis’ and even ‘truth of the first inaccessible rank’, or ‘truth of the rank of the first hyper-Mahlo cardinal’ is reducible to second-order logical truth. It follows that second-order logic is *inherently incomplete* in the sense that there is no sound, recursively enumerable deductive system for it. Indeed, the set of second-order logical truths is not in the analytic hierarchy. A number of central, set-theoretic principles have natural renderings in second-order languages, many of which are independent of Zermelo-Fraenkel set theory. For example, there is a second-order sentence **CH**, which has no non-logical terminology, such that **CH** is a logical truth if and only if the continuum hypothesis fails. There is another sentence which is a logical truth if and only if the generalized continuum hypothesis holds, and there is a sentence which is a logical truth if and only if there are no inaccessible cardinals (again, see [Shapiro, 1991, Chapter 5]).

Of course, whether these features of second-order logic are ‘defects’ depends on what properties a good logical theory should have. This, in turn, depends on what logical theory is supposed to accomplish. On this ancient question, we will rest content with a brief sketch.

The intractability of second-order consequence is a direct and inevitable result of the expressive power of second-order languages. In one sense, this good and bad news is to be expected and welcomed. The informal notion of logical consequence is tied to what sentences (or propositions) mean and what the linguistic items refer to. Thus, one of the purposes of a formal language is to capture the informal semantics of mathematical discourse and, in particular, to replicate the notion of reference and satisfaction. Since informal mathematical discourse appears to have the resources to characterize notions like finitude and structures like the natural numbers and the real numbers (up to isomorphism), our formal language should have this expressive power as well. The richness and intractability of second-order languages is a consequence of the richness and intractability of mathematical discourse generally. From this perspective, one should hold that mathe-

matics and logic are a seamless whole, and it is impossible to draw a sharp boundary between them. In his treatment of second-order logic, Church [1956, p. 332] wrote that ‘logic and mathematics should be characterized, not as different subjects, but as elementary and advanced parts of the same subject’. Barwise [1985, 5] elaborates a similar idea:

... in basic logic courses ... we attempt to draw a line between ‘logical concepts’, as embodied in the so-called ‘logical constants’, and all the rest of the concepts of mathematics. [W]e do not so much question the placement of this line, as question whether there is such a line, or whether all mathematical concepts have their own logic, something that can be investigated by the tools of mathematics ... As logicians, we do our subject a disservice by convincing others that logic is first-order and then convincing them that almost none of the concepts of modern mathematics can really be captured in first-order logic.

Barwise concludes that ‘one thing is certain. There is no going back to the view that logic is first-order logic’. See [Shapiro, 1991] and [Sher, 1991] for articulations of similar theses.

On the other hand, there are reasons to demur from the full expressive power—and intractability—of second-order logic. The mathematical logician desires a system that she can study and shed some light upon, using the ‘tools of mathematics’. Completeness, compactness, and the Löwenheim–Skolem theorems give rise to the main tools developed by the mathematical logician, and these tools only apply to relatively weak formal languages. A logical system that is just as complex as mathematics provides no special handle for the logician. At the extreme of the view articulated in the previous paragraph, logic just is mathematics and so there is nothing for the logician to contribute. The ‘logic’ of arithmetic, say, *is* number theory and so the logician just is a number theorist. The ‘logic’ of Euclidean geometry is Euclidean geometry and so here the logician is just the geometer.

The philosopher also has reasons to keep logic tractable, or at least more tractable than the second-order consequence relation. There is a longstanding view that logic should be free of ontological and metaphysical presuppositions. If that cannot be maintained, then at least these presuppositions should be kept to a minimum. Logical consequence should just turn on the *meanings* of the logical particles. The consequence relation should be transparent and potentially obvious. Something has gone wrong when the continuum hypothesis (or its negation) becomes a logical truth. Quine is a vocal champion of first-order logic, against second-order logic. In [1953, p. 116], he wrote:

The bulk of logical reasoning takes place on a level which does not presuppose abstract entities. Such reasoning proceeds mostly

by quantification theory [i. e., first-order logic], the laws of which can be represented through schemata involving no quantification over class variables. Much of what is commonly formulated in terms of classes, relations, and even number, can easily be reformulated schematically within quantification theory . . .

Quine [1986, p. 68] later argued that second-order logic is not logic, but is ‘set theory in disguise’, a wolf in sheep’s clothing:

Set theory’s staggering existential assumptions are cunningly hidden . . . in the tacit shift from schematic predicate letter to quantifiable variable.

See also [Jané, 1993] and [Wagner, 1987].

Although I am among the advocates of second-order logic [Shapiro, 1991], there is no need to adjudicate this issue here. A safe compromise is that there is motivation to develop logics that are, in a sense, intermediate between first-order and second-order. The philosopher seeks a course between the two extremes delimited above, a logical system that is not as weak as first-order, but has at least some of the traditional *desiderata* of analyticity and transparency. Formally, we desire systems that have greater expressive resources than first-order logic, but are not as intractable as second-order logic. This is the motivation behind the extensive study [Barwise and Feferman, 1985]. Cowles [1979, p. 129] put it well:

It is well-known that first-order logic has a limited ability to express many of the concepts studied by mathematicians . . . However, first-order logic . . . does have an extensively developed and well-understood model theory. On the other hand, full second-order logic has all the expressive power needed to do mathematics, but has an unworkable model theory. Indeed, the search for a logic with a semantics complex enough to say something, yet at the same time simple enough to say something *about*, accounts for the proliferation of logics . . .

There are a growing number of candidates for our mathematical and philosophical logician to consider.

## 2 WHAT?

Just what is a logical system between first-order and second-order? I presume that the reader is familiar with ‘logical system’, ‘first-order’ (Sundholm’s chapter in Volume 2 of this *Handbook*) and ‘second-order’ ([Shapiro, 1991] and van Benthem and Doets’ Chapter below), but I will indulge in a few words on ‘between’.

There is, first, a proof-theoretic sense of ‘between’. The logician begins with an ordinary, second-order language of a particular theory, such as arithmetic or analysis, and studies sub-systems of the full second-order deductive system for that theory. A typical focus is on restricted versions of the comprehension scheme, for example limiting it to  $\Delta_1^0$ -formulas, or to  $\Pi_1^1$ -formulas. Logicians also consider restrictions on the axiom of choice, and restrictions on the schemes used to characterize various structures, such as the induction principle for arithmetic and the completeness principle for analysis. There is an ambitious, fruitful, and growing program developed along these lines. The so-called ‘reverse mathematics’ lies at the heart of this research. Interested readers can begin with [Feferman, 1977] and [Simpson, 1985].

This chapter focuses on a model-theoretic sense of ‘between’. We consider a potpourri of different languages, or to be precise, a potpourri of different logical operators which can be added to a standard, first-order language. Most of the languages have a model-theoretic semantics over the same class of models as first-order and second-order logic, and each of the logics can make more distinctions among models than can be done in first-order logic. That is, each language has more expressive resources than the corresponding first-order language. For example, most of them can characterize the notion of ‘finitude’, and most of the languages allow a categorical characterization of the natural numbers.

Some of the logics have properties enjoyed by first-order logic, such as compactness, completeness, and the Löwenheim–Skolem theorems, and some have weaker versions of these properties. On the other hand, the logics considered here cannot make all of the distinctions that can be accomplished with full second-order languages with standard semantics. Thus, the logics are ‘between’ first-order and second-order. Some of the systems are strictly weaker than second-order, in a sense to be made precise, while others (like the infinitary languages) are not comparable. In light of the theme of this *Handbook*, I will stick (for the most part) to systems that have, or might have, some philosophical interest or application. There is no attempt to be exhaustive.

Logicians have discovered limits to the ability to optimize between expressive power and tractability. Certain of the limitative properties *characterize* first-order logic, in a sense to be made precise, and so we cannot have the bulk of our cake and eat the bulk of it too. If we are to have the main tractable features of first-order logic, we are stuck with its expressive poverty. Conversely, some central non-first-order concepts and structures can be characterized, up to isomorphism, as soon as some of the limitative properties are given up.

Let  $K$  be a set of non-logical terminology. It is convenient to assume that  $K$  contains infinitely many constants and relation symbols of each degree. Sometimes  $K$  is called a ‘vocabulary’ or a ‘signature’. We consider various

languages built upon  $K$ . Let  $\mathcal{L}1[K] =$  be the first-order language, with identity, whose non-logical terminology comes from  $K$ , and let  $\mathcal{L}2[K]$  be the corresponding second-order language.

Suppose that  $\mathcal{L}[K]$  is a language that contains  $\mathcal{L}1[K] =$ . Assume that if  $\Phi$  and  $\Psi$  are formulas in  $\mathcal{L}[K]$ , then so are  $\neg\Phi$ ,  $\Phi \rightarrow \Psi$ , and  $\exists x\Phi$ , for each first-order variable  $x$ . That is, we assume that  $\mathcal{L}[K]$  is closed under the usual first-order connectives and quantifiers. Assume also that  $\mathcal{L}[K]$  has a semantics with the same class of models as that of  $\mathcal{L}1[K] =$  and that the aforementioned connectives and quantifiers have the same role in the satisfaction of formulas as they have in  $\mathcal{L}1[K] =$ . Thus, in particular, the semantics of  $\mathcal{L}[K]$  agrees with that of  $\mathcal{L}1[K] =$  on the satisfaction of first-order formulas. We assume finally that if  $M1$  and  $M2$  are isomorphic models and  $\Phi$  is any formula of  $\mathcal{L}[K]$ , then  $M1 \models \Phi$  if and only if  $M2 \models \Phi$ . This *isomorphism property* seems essential to any model-theoretic semantics worthy of the name. If a language/logic could distinguish between isomorphic structures, then its consequence relation is not formal.<sup>1</sup> Of course,  $\mathcal{L}1[K] =$  and  $\mathcal{L}2[K]$  have the isomorphism property, as do all of the logics considered below.

Many common semantical notions can be formulated in this general setting. The logic  $\mathcal{L}[K]$  is *compact* if for every set  $\Gamma$  of formulas of  $\mathcal{L}[K]$ , if each finite subset of  $\Gamma$  is satisfiable, then  $\Gamma$  itself is satisfiable; and  $\mathcal{L}[K]$  is *countably compact* if for every countable set  $\Gamma$  of formulas of  $\mathcal{L}[K]$ , if each finite subset of  $\Gamma$  is satisfiable, then  $\Gamma$  itself is satisfiable. The logic  $\mathcal{L}[K]$  is *weakly complete* if the collection of logically true sentences of  $\mathcal{L}[K]$  is a recursively enumerable set of strings. If  $\mathcal{L}[K]$  is weakly complete, then there is an effective deductive system whose theorems are the logical truths of  $\mathcal{L}[K]$ . That is, if  $\mathcal{L}[K]$  is weakly complete, then there is an effective, sound, and complete deductive system for it. The logic  $\mathcal{L}[K]$  has the *downward Löwenheim–Skolem property* if each satisfiable, countable set of sentences has a model whose domain is at most countable; and  $\mathcal{L}[K]$  has the *upward Löwenheim–Skolem property* if, for each set  $\Gamma$  of sentences, if  $\Gamma$  has a model whose domain is infinite, then for each infinite cardinal  $\kappa$ ,  $\Gamma$  has a model whose domain has cardinality at least  $\kappa$ . All of these properties are enjoyed by  $\mathcal{L}1[K] =$  (provided that  $K$  is recursive), but decidedly not by  $\mathcal{L}2[K]$ .

We say that  $\mathcal{L}[K]$  is *first-order equivalent* if for each sentence  $\Phi$  of  $\mathcal{L}[K]$ , there is a sentence  $\Phi'$  of  $\mathcal{L}1[K] =$  such that  $\Phi \equiv \Phi'$  is a logical truth, or in other words,  $\Phi$  and  $\Phi'$  are satisfied by the same models. Thus, if  $\mathcal{L}[K]$  is first-order equivalent, then it is not capable of making any distinctions among models that cannot be made by the first-order  $\mathcal{L}1[K] =$ . Clearly, the second-order  $\mathcal{L}2[K]$  is not first-order equivalent. Any categorical sentence with an infinite model is not equivalent to any first-order sentence. There are a number of results that characterize logics that are first-order equivalent,

<sup>1</sup>See [Tarski, 1986] and [Sher, 1991] for an elaboration of this point.



several of which are reported here. A few more definitions are needed.

The logic  $\mathcal{L}[K]$  has the *relativization property* if for each formula  $\Phi$  in  $\mathcal{L}[K]$  and each  $\Psi(x)$  with  $x$  free, there is a formula  $\Phi/\{x|\Psi(x)\}$  asserting that  $\Phi$  holds when the domain is  $\{x|\Psi(x)\}$ .  $\mathcal{L}[K]$  has the *substitution property* if, for each formula  $\Phi$  containing an  $n$ -place relation symbol  $R$ , and each formula  $\Psi(x_1, \dots, x_n)$  (containing no free variables that occur in  $\Phi$ , except possibly  $x_1, \dots, x_n$ ), there is a formula  $\Phi(R|\Psi)$  that is equivalent to the result of substituting  $\Psi(t_1, \dots, t_n)$  for each occurrence of  $Rt_1, \dots, t_n$  in  $\Phi$ . Both  $\mathcal{L}1[K]$  and  $\mathcal{L}2[K]$  have these properties, as do most of the logics considered below.<sup>2</sup> The logic  $\mathcal{L}[K]$  is *effectively regular* if the collection of formulas of  $\mathcal{L}[K]$  is a recursive set of strings, and if the aforementioned relativization and substitution functions are recursive. In the case of first-order and second-order languages, the indicated functions are straightforward.  $\mathcal{L}1[K]$  and  $\mathcal{L}2[K]$  are effectively regular if the set  $K$  is recursive.

Finally,  $\mathcal{L}[K]$  has the *finite occurrence property* if for each formula  $\Phi$  of  $\mathcal{L}[K]$ , there is a finite subset  $K'$  of  $K$  such that  $\Phi$  is in  $\mathcal{L}[K']$ . The idea is that if  $\mathcal{L}[K]$  has the finite occurrence property, then each formula of  $\mathcal{L}[K]$  involves only finitely many non-logical items. For most of the logics considered below, the finite occurrence property holds automatically, since their formulas are finite strings of characters. Only the infinitary logics lack this property.

The most well-known characterizations of first-order equivalence are due to Lindström:

**THEOREM 1** ([Lindström, 1969]). *If  $\mathcal{L}[K]$  has the finite occurrence property, is countably compact, and has the downward Löwenheim–Skolem property, then  $\mathcal{L}[K]$  is first-order equivalent.*

**THEOREM 2** ([Lindström, 1969]). *Let  $\mathcal{L}[K]$  be an effectively regular logic. Then if  $\mathcal{L}[K]$  has the downward Löwenheim–Skolem property and the upward Löwenheim–Skolem property, then  $\mathcal{L}[K]$  is first-order equivalent.*

**THEOREM 3** ([Lindström, 1969]). *Let  $\mathcal{L}[K]$  be an effectively regular logic. If  $\mathcal{L}[K]$  has the downward Löwenheim–Skolem property and is weakly complete then  $\mathcal{L}[K]$  is first-order equivalent, and, moreover, there is a recursive function  $f$  such that for every sentence  $\Phi$  of  $\mathcal{L}[K]$ ,  $f(\Phi)$  is a sentence of  $\mathcal{L}1[K]$  that has exactly the same models as  $\Phi$ .*

See Flum [1985, Section 1] for proofs of these theorems, and further refinements of them.

So we see some limitations to our optimization project. We cannot have both compactness and the downward Löwenheim–Skolem property and get beyond the expressive poverty of first-order logic. If we manage to keep compactness and get beyond first-order, we forgo Löwenheim–Skolem. If

<sup>2</sup>See [Ebbinghaus, 1985, Section 1.2] for more precise definitions of relativization and substitution.

we keep Löwenheim–Skolem and get beyond first-order, we forgo weak completeness.

The proofs of Lindström’s results given in [Flum, 1985, Section 1] reveal that if  $\mathcal{L}[K]$  has the finite occurrence property and the downward Löwenheim–Skolem property, and yet  $\mathcal{L}[K]$  is not first-order equivalent, then it is possible to characterize the notion of *finitude* in  $\mathcal{L}[K]$ . In particular, under these circumstances, there is a sentence  $\Phi$  of  $\mathcal{L}[K]$  containing a monadic predicate letter  $U$  such that (1) in every model of  $\Phi$ , the extension of  $U$  is finite; and (2) for each natural number  $n \geq 1$ , there is a model of  $\Phi$  in which the extension of  $U$  has cardinality  $n$ . There can be no such sentence in any countably compact extension of a first-order language. To see this consider the following countable set of sentences:

$$\Gamma = \{\Phi, \exists x Ux, \exists x \exists y (x \neq y \& Ux \& Uy), \\ \exists x \exists y \exists z (x \neq y \& x \neq z \& y \neq z \& Ux \& Uy \& Uz), \dots\}$$

By hypothesis, every finite subset of  $\Gamma$  is satisfiable and so by countable compactness,  $\Gamma$  itself is satisfiable. But a model of  $\Gamma$  is a model of  $\Phi$  in which the extension of  $U$  is infinite.

Let  $M1$  and  $M2$  be two models of the logic  $\mathcal{L}[K]$  (and of  $\mathcal{L}1[K] =$ ). A *partial isomorphism* between  $M1$  and  $M2$  is defined to be a one-to-one function  $f$  from a subset of the domain of  $M1$  onto a subset of the domain of  $M2$  that preserves the relevant structure. Thus, for example, if  $R$  is a binary relation letter and  $m$  and  $n$  are both in the domain of  $f$ , then  $\langle m, n \rangle$  is in the extension of  $R$  in  $M1$  if and only if  $\langle fm, fn \rangle$  is in the extension of  $R$  in  $M2$ . Now, the structures  $M1$  and  $M2$  are *partially isomorphic* if there is a set  $P$  of partial isomorphisms between  $M1$  and  $M2$  with the *back-and-forth property*: for each  $f \in P$  and each  $m$  in the domain of  $M1$  and each  $m'$  in the domain of  $M2$ , there is an  $f' \in P$  such that  $f \subseteq f'$  and  $m$  is in the domain of  $f'$  and  $m'$  is in the range of  $f'$ .

A well known technique, due to Cantor, establishes that if  $M1$  and  $M2$  are partially isomorphic and both domains are countable, then  $M1$  and  $M2$  are isomorphic. This does not hold for domains with higher cardinalities, since, for example, any two dense linear orderings with neither a first nor a last element are partially isomorphic.

A logic  $\mathcal{L}[K]$  has the *Karp property* if partially isomorphic structures are equivalent. That is,  $\mathcal{L}[K]$  has the Karp property iff for any models  $M1$  and  $M2$ , and any sentence  $\Phi$  of  $\mathcal{L}[K]$ , if  $M1$  and  $M2$  are partially isomorphic, then  $M1 \models \Phi$  iff  $M2 \models \Phi$ . The Karp property gives rise to many of the techniques for the study of first-order model theory. It is part of another characterization of first-order logic:

**THEOREM 4.** *Let  $\mathcal{L}[K]$  be a logic with the relativization, substitution, and finite occurrence properties. If  $\mathcal{L}[K]$  has the Karp property and is countably compact, then  $\mathcal{L}[K]$  is first-order equivalent.*

The proof of this in [Flum, 1985, Section 2] establishes that if a logic  $\mathcal{L}[K]$  (with the relativization, substitution, and finite occurrence properties) has the Karp property and is not first-order equivalent, then the natural numbers under the ‘less than’ relation can be characterized up to isomorphism in  $\mathcal{L}[K]$ . See van Benthem and Doets’ Chapter below for an interesting relationship between partial isomorphism and first-order quantifiers.

One more example: The use of ultraproducts is an extremely fruitful technique in the model theory of first-order logic. In effect, this gives another characterization of first-order equivalence (for the relevant definitions, see [Bell and Slomson, 1971] or [Chang and Keisler, 1973]). If  $\{M_i \mid i \in A\}$  is a family of models of  $\mathcal{L}_1[K]$ , and  $U$  an ultrafilter on  $A$ , then let  $\Pi_U\{M_i\}$  be the resulting ultraproduct. Say that a logic  $\mathcal{L}[K]$  *preserves ultraproducts* if for each sentence  $\Phi$  of  $\mathcal{L}[K]$  and each ultraproduct  $\Pi_U\{M_i\}$ , if  $M_i \models \Phi$  for each  $i \in A$ , then  $\Pi_U\{M_i\} \models \Phi$ .

**THEOREM 5** ([1973, Chapter 6]).  *$\mathcal{L}[K]$  is first-order equivalent if and only if  $\mathcal{L}[K]$  preserves ultraproducts.*

The ‘only if’ part of this equivalence underwrites the ultraproduct construction in first-order logic; the ‘if’ part indicates that only first-order logic can be illuminated this way.

It is surely significant that such a wide variety of properties all converge on first-order semantics. In philosophical jargon, one might call first-order logic a ‘natural kind’. But we should not forget the expressive poverty of first-order languages. First-order logic is important, but it does not have a monopoly on the attention of mathematical and philosophical logicians.

### 3 JUST SHORT OF SECOND-ORDER LOGIC

Here we consider two seemingly minor restrictions to full second-order logic. One is to allow only second-order variables that range over *monadic* predicates or properties (or sets). The other is to allow the full range of second-order variables, but insist that the variables do not occur bound. This is equivalent to using a language with nothing more complex than  $\Pi_1^1$ -formulas. It is interesting how much tractability these restrictions bring, with a minimal loss in expressive power.

#### 3.1 *Monadic second-order logic*

Define a set  $K$  of non-logical terminology to be *monadic* if it does not contain any function symbols or any  $n$ -place relation symbols, for  $n > 1$ . It is well-known that if  $K$  is monadic and recursive, then the set of logical truths of the first-order  $\mathcal{L}_1[K]$  is recursive. Moreover, if the set of non-logical terminology is monadic, the Löwenheim [1915] classic contains a decision

procedure for the logical truths of a language that contains bound first-variables and bound second-order variables ranging over 1-place properties (see [Gandy, 1988, p. 61] and [Dreben and Goldfarb, 1979, Section 8.3]). This sounds like wonderful news, but the languages are too weak to express substantial mathematics. The notion of function is central to modern mathematics, and it is hard to do much without it. However, we may get by without *variables* ranging over functions.

*Monadic second-order languages* contain bound variables ranging over 1-place relations, but there are no variables ranging over functions or  $n$ -place relations, for any  $n > 1$ . That is, all second-order variables are monadic. No restrictions are placed on the non-logical terminology, so that monadic second-order languages lie between first-order and second-order languages. Gurevich [1985] is an extensive treatment of such languages, arguing that they are ‘a good source of theories that are both expressive and manageable’.

There is an important restriction on this statement. A *pair function* on a given domain  $d$  is a one-to-one function from  $d \times d$  into  $d$ . A theory *admits pairing* if there is a definable pair function on it. That is, there is a formula  $\Phi(x, y, z)$ , with only the free variables shown, such that in every model  $M$  of the theory, there is a pair function  $f$  on the domain of  $M$  such that for any  $a, b, c$  in the domain,  $M$  satisfies  $\Phi(a, b, c)$  if and only if  $f(a, b) = c$ . Then if a theory cast in a monadic second-order language admits pairing, it is equivalent to the same theory formulated in an unrestricted second-order language. There is no loss of expressive power and no gain in manageability.<sup>3</sup> The reason, of course, is that a relation can be thought of as a property of pairs. Let  $f$  be a pair function. Then a given binary relation  $R$  is equivalent to the property that holds of an element  $x$  iff there is a  $y$  and  $z$  such that  $f(y, z) = x$  and  $R$  holds of the pair  $\langle y, z \rangle$ .

In arithmetic, the function  $g(x, y) = 2^x 3^y$  is a pair function, and in set theory  $h(x, y) = \{\{x\}, \{x, y\}\}$  is the standard pair function. For this reason, monadic second-order arithmetic and monadic second-order set theory are equivalent to their full second-order versions. However, on the positive side of the ledger, Gurevich [1985] points out that there are theories that do not admit pairing, whose monadic second-order theories are interesting. One is arithmetic, formulated with the successor function alone. Although the monadic second-order theory is categorical, and the natural order can easily be defined in it, the theory is decidable. Addition and multiplication can be defined in the full second-order theory of arithmetic (see [Shapiro, 1991, Chapter 5]), but not in the monadic theory. A second example, also decidable, is the monadic theory of the binary tree—the structure of the set of strings on a two letter alphabet. Rabin [1969] showed how to interpret the theory of strings on a countable alphabet in the monadic second-order

---

<sup>3</sup>Shapiro [1991, Chapter 6, Section 2] contains a theorem that what may be called monadic  $n$ th-order logic (for sufficiently large  $n$ ) admits pairing. Thus, the manageability of monadic second-order logic does not apply to monadic higher-order logic in general.

theory of the binary tree, so the theory does have interesting and useful applications. A third example is the monadic second-order theory of countable ordinals.

Some reducibility results indicate that certain monadic theories are rich and intractable. Shelah [1975] showed that first-order arithmetic can be reduced to the monadic second-order theory of the real numbers under the order relation. It follows that the latter is a rich, undecidable theory—just as rich and unmanageable as first-order arithmetic. More generally, Gurevich and Shelah [1983] established that full second-order logic itself can be reduced to what is called the monadic second-order theory of order, cast in a language with a single binary, non-logical relation symbol  $<$ . In particular, they show that there is a recursive function  $F$  such that for each sentence  $\Phi$  of the second-order language  $\mathcal{L}_2$  (with no non-logical terminology),  $F(\Phi)$  is a sentence in the monadic second-order language of order, and  $\Phi$  is a logical truth iff  $F(\Phi)$  is satisfied by every linear order. It follows that the monadic second-order theory of order is just as rich and unmanageable as second-order logic.

George Boolos [1984; 1985] proposed an alternate way to understand monadic second-order languages—with or without pairing—which promises to overcome at least some of the objections to second-order logic (see also [Boolos, 1985a; Lewis, 1991]). Recall that according to standard semantics for second-order languages, a monadic second-order existential quantifier can be read ‘there is a class’ or ‘there is a property’, in which case, of course, the locution invokes classes or properties. This is the source of Quine’s argument that in order to understand second-order quantifiers, we need to invoke a special subject—the mathematical theory of sets or, even worse, the metaphysical theory of properties. Quine concludes that second-order logic is not logic. Against this, Boolos suggests that the monadic second-order universal quantifier be understood as a *plural* quantifier, like the locution ‘there are (objects)’ in natural language.

Consider the following, sometimes called the ‘Geach–Kaplan sentence’:

Some critics admire only one another.

Taking the class of critics to be the domain of discourse, and symbolizing ‘ $x$  admires  $y$ ’ as  $Axy$ , the Geach–Kaplan sentence has a (more or less) straightforward second-order rendering:

$$(*) \quad \exists X(\exists xXx \& \forall x \forall y((Xx \& Axy) \rightarrow (x \neq y \& Xy))).$$

Kaplan observed that if  $Axy$  is interpreted as  $x = 0 \vee x = y + 1$  in the language of arithmetic, then  $(*)$  is satisfied by all *non-standard* models of first-order arithmetic, but not by the natural number structure  $\mathbb{N}$ . However, a compactness argument establishes the existence of a non-standard model  $M$  such that for any sentence  $\Phi$  of first-order arithmetic,  $M \models \Phi$  if and only if  $\mathbb{N} \models \Phi$ . Thus there is no first-order sentence that is equivalent to  $(*)$ .

The issue concerns how the sentence (\*) is to be understood. According to standard semantics, it would correspond to ‘there is a non-empty *class*  $X$  of critics such that for any  $x$  in  $X$  and any critic  $y$ , if  $x$  admires  $y$ , then  $x \neq y$  and  $y$  is in  $X$ ’. This gloss implies the existence of a class, while the original ‘some critics admire only one another’ does not, at least *prima facie*.

Natural languages, like English, allow the plural construction and, in particular, English contains the plural quantifier. Boolos argues that the informal meta-language—the one we use in developing formal semantics—also contains this construction, and the construction can be employed to interpret monadic second-order existential quantifiers. The relevant locution is ‘there are objects  $X$ , such that . . .’. As in the first-order case, the variable serves as a place-holder, for purposes of cross reference.

In set theory, for example, the ‘Russell sentence’,

$$\exists x \forall x (Xx \equiv x \notin x),$$

is a consequence of the comprehension scheme. According to standard semantics, it corresponds to a statement that there is a *class* (or property) that is not coextensive with any *set*. Admittedly, this takes some getting used to. On Boolos’ interpretation, the Russell sentence has an innocent reading: ‘there are some sets such that any set is one of them just in case it is not a member of itself’. Similarly, the second-order principle of foundation,

$$\forall X (\exists x Xx \rightarrow \exists x (Xx \& \forall y (y \in x \rightarrow \neg Xy))),$$

comes to ‘it is not the case that there are some sets such that every one of them has a member that is also one of them’. Again, neither properties nor proper classes are invoked.

There is a complication here due to the fact that an English sentence in the form ‘there are some objects with a certain property’ implies that there is at least one object with this property, while a sentence that begins with a standard second-order existential quantifier does not have a similar implication. In particular, in standard semantics, a sentence in the form  $\exists X \Phi(X)$  is satisfied by a model even if  $\Phi$  holds only of the empty class in that model.<sup>4</sup> To accommodate this, Boolos takes the comprehension scheme  $\exists X \forall x (Xx \equiv \Phi(x))$ , for example, to correspond to ‘either  $\neg \exists x \Phi(x)$  or else there are some objects such that any object is one of them just in case  $\Phi$  holds of it’.

Boolos [1985] develops a rigorous, model-theoretic semantics for monadic second-order languages. As indicated, the plural quantifier is *used* in the meta-language to interpret the monadic quantifier. If this semantics can

<sup>4</sup>Actually, it seems to me that the locution ‘there are objects with a certain property’ implies that there are at least *two* objects with the property. This detail can be handled in a straightforward manner, if desired.

be sustained, then one can accept monadic second-order languages, without thereby being committed to the existence of classes. Boolos' main claim is that plural quantifiers do not involve any ontology other than the range of the first-order variables. Monadic second-order formulas do not invoke classes at all, unless the corresponding first-order formulas do.

According to the Boolos proposal, then, second-order arithmetic presupposes natural numbers, but not sets of numbers, and second-order geometry presupposes points, but not sets of points. This may be an important distinction for tracking the separate presuppositions of different fields, but ultimately it is not crucial for these fields. Boolos is certainly not out to reject sets altogether, being an advocate set theory. Moreover, if certain reflection principles hold, the second-order consequence relation is the same on both standard semantics and his interpretation. The difference between the interpretations comes to the fore in set theory itself. Boolos does not accept the existence of proper classes (and thus does not regard 'V' as a proper noun). In [1985], he wrote that 'the difficulty of interpreting second-order quantifiers is most acute when the underlying language is the language of set theory ...'. And in [1984]:

... we [do not] want to take the second-order variables as ranging over some set-like objects, sometimes called 'classes', which have members, but are not themselves members of other sets, supposedly because they are 'too big' to be sets. Set theory is supposed to be a theory about *all* set-like objects. [Boolos, 1984, p, 442]

The Boolos program, then, accomplishes a reduction of ontology by employing plural quantifiers, which are found in ordinary language. It is thus a tradeoff between ontology and ideology, and, as such, it is not clear how the case is to be adjudicated. The prevailing criterion is the Quinean assertion that the ontology of a theory is the range of its bound variables. Quine insists that the theory in question be first regimented in a *first-order* language, but the criterion is readily extended to *standard* higher-order languages, since in such systems, higher-order variables have (more or less) straightforward ranges, namely, classes, relations, or functions. In this respect, second-order variables are on a par with first-order variables. Boolos, however, proposes a certain asymmetry between first-order and monadic second-order variables. The latter do not have 'ranges' in the same sense that the former do.

Resnik [1988] argues against the Boolos program, suggesting that plural quantifiers of natural language be understood (after all) in terms of classes. Both Resnik and Boolos [1985] acknowledge that this sort of dispute leads to a standoff, or a regress. Anything either side says can be reinterpreted by the other. The issue concerns whether we have a serviceable grasp of plural quantifiers, sufficient for use in the meta-languages of model-theoretic

semantics. Resnik seems to claim that we do not. What understanding we do have of plural quantifiers is mediated by our understanding of sets. Boolos claims that we do have a reasonable grasp on plural quantifiers, citing the prevalence of plurals in ordinary language. It might be noted, however, that plurals *in general* seem to be rather complex, and there is no consensus among linguists concerning how they are to be understood (see, for example, [Landman, 1989]). But Boolos does not invoke the full range of plural nouns, only plural *quantifiers*. It must be admitted that these seem to be understood reasonably well, about as well as (monadic) second-order quantifiers. Resnik would retort that even this is mediated by set theory, *first-order* set theory. Thus, the regress.

### 3.2 *Free-variable second-order logic*

Our second ‘slight’ restriction on second-order logic consists of restricting the language to free second-order variables. The resulting logic has much of the expressive power of full second-order logic, but is not quite as intractable. Some of the usual arguments against second-order logic do not apply to free-variable second-order logic. Free-variable second-order languages are similar (if not identical) to the ‘schematic’ languages studied in [Lavine, 1994], and they are in the same spirit as the ‘slightly augmented first-order languages’ presented in [Corcoran, 1980]. The latter has only a single, monadic predicate variable, which occurs free.

The language  $\mathcal{L}2[K]-$  is obtained from the first-order  $\mathcal{L}1[K]-$  by adding a stock of relation variables, with the usual formation rules for second-order languages.<sup>5</sup> The point, of course, is that  $\mathcal{L}2[K]-$  has no quantifiers to bind the second-order variables. We follow the usual convention of interpreting the free variables as if they are bound by universal quantifiers whose range is the whole formula. Thus, the formulas envisaged here are equivalent to  $\Pi_1^1$  formulas of a second-order language. We formulate the semantics in terms of the usual model theory for second-order languages.

Let  $M$  be a structure appropriate for  $K$  and let  $d$  be the domain of  $M$ . Let  $s$  be an assignment of a member of  $d$  to each first-order variable and an assignment of an appropriate relation on  $d$  to each second-order variable. Let  $\Phi$  be a formula of  $\mathcal{L}2[K]-$ . In the usual treatments of second-order logic, one defines the notion that  $M$  *satisfies*  $\Phi$  under the assignment  $s$  (see van Benthem and Doets’ Chapter below or [Shapiro, 1991, Chapter 3]). This is not quite what we want here, since in the usual framework, a free variable  $X$  is taken as ‘denoting’ the particular relation  $s(X)$ , whereas here we want the variable to serve generality—we interpret the variable as if it

---

<sup>5</sup>The free-variable system in [Shapiro, 1991] includes variables ranging over functions. This does not affect the expressive power of the language, since a function can be thought of as a relation. The required modifications are straightforward, but they are tedious and a distraction from the present focus.



were bound by a universal quantifier. So we say that  $M$  *quasi-satisfies*  $\Phi$  under the assignment  $s$ , written  $M, s \models \Phi$ , if and only if  $M$  satisfies  $\Phi$  under every assignment  $s'$  that agrees with  $s$  on the first-order variables. Notice that  $M$  quasi-satisfies  $\Phi$  under  $s'$  if and only if  $M$  satisfies  $\forall X\Phi$  under  $s$ . The values assigned to the higher-order variables play no role. As usual, we suppress the assignment if there are no free variables in  $\Phi$ .

Since an  $\mathcal{L}2[K]$ -formula in the form  $\Phi(X)$  amounts to  $\forall X\Phi(X)$ , a formula  $\neg\Phi(X)$  amounts to  $\forall X\neg\Phi(X)$ . Thus,  $\neg\Phi(X)$  is *not* the ‘contradictory opposite’ of  $\Phi(X)$ . There are formulas  $\Phi(X)$  with  $X$  free, such that there is no formula of  $\mathcal{L}2[K]$ -equivalent to its contradictory opposite,  $\neg\forall X\Phi(X)$ . Thus, even though  $\mathcal{L}2[K]$ - has a negation sign, the language is not closed under contradictory opposition.

In standard deductive systems for higher-order languages, the main item is the comprehension scheme:

$$\exists X(\forall x(Xx \equiv \Phi(x))),$$

one instance for each formula  $\Phi$  (not containing  $X$  free). Since this is not a formula of  $\mathcal{L}2[K]$ -, the deductive system for free-variable second-order languages is a bit more complicated.

Let  $\Phi$  and  $\Psi(x_1, \dots, x_n)$  be formulas of  $\mathcal{L}2[K]$ -, the latter possibly containing the indicated free (first-order) variables. Let  $R$  be an  $n$ -place relation variable. Define  $\Phi[R/\Psi(x_1, \dots, x_n)]$  to be the formula obtained from  $\Phi$  by replacing each occurrence of  $Rt_1, \dots, t_n$  (where each  $t_i$  is a term) with  $\Psi(t_1, \dots, t_n)$ , making sure that no free variables in any  $t_i$  become bound in  $\Psi(t_1, \dots, t_n)$  (relettering bound variables if necessary). For example, if  $\Phi$  is  $Rf(w) \vee \forall y(Ry \rightarrow Qy)$  and  $\Psi(x)$  is  $\forall zXxz$ , then  $\Phi[R/\Psi(x)]$  is  $\forall zXf(w)z \vee \forall y(\forall zXyz \rightarrow Qy)$ .

The deductive system for  $\mathcal{L}2[K]$ - consists of the schemes and rules of the corresponding first-order system, together with the following *substitution rule*:

From  $\Phi$  infer  $\Phi[R/\Psi(x_1, \dots, x_n)]$ , where  $\Psi$  does not contain any free variables that are bound in  $\Phi[R/\Psi(x_1, \dots, x_n)]$ .

The substitution rule has the effect of treating any formula with relation variables as a scheme, whose ‘place holders’ are the relation variables, and whose substitution instances are the appropriate formulas of  $\mathcal{L}2[K]$ -.

In the usual deductive system for full second-order logic, one can derive  $\Phi[R/\Psi(x_1, \dots, x_n)]$  from  $\forall R\Phi$ , using an instance of the comprehension scheme, provided that  $\Psi$  does not contain any free variables that become bound in  $\Psi[R/\Psi(x_1, \dots, x_n)]$ . A variant of the substitution rule is thus a derived rule in full second-order logic. Henkin [1953] is an insightful account of the relationship between substitution rules and principles of comprehension.

Call the deductive system for free-variable second-order logic  $D2-$ . Notice that  $D2-$  does *not* have an unrestricted deduction theorem. If it did, then since  $\neg Xx \vdash_{D2-} \neg(x = x)$ , we would have  $\vdash_{D2-} \neg Xx \rightarrow \neg(x = x)$  and so  $\vdash_{D2-} Xx$ . But, from  $Xx$ , any formula can be deduced. Thus, if the deduction theorem held,  $D2-$  would be inconsistent. The following, however, is straightforward:

**THEOREM 6** (Restricted deduction theorem). *If there is a deduction in  $D2-$  of  $\Psi$  from  $\Gamma \cup \{\Phi\}$  in which the substitution rule is not applied to a relation variable that occurs in  $\Phi$ , then  $\Gamma \vdash_{D2-} \Phi \rightarrow \Psi$ .*

This difference between  $D2-$  and a deductive system for full second-order logic is due to a common ambiguity in the interpretation of free variables. Sometimes they are taken as surrogate names for (unspecified) individuals. On this reading, a formula  $\Phi(x)$  asserts that the object  $x$  has the property represented by  $\Phi$ . The phrase *free constant* might be better than ‘free variable’ in such cases. In other contexts, free variables are taken as if they are bound by prenex universal quantifiers. Accordingly,  $\Phi(x)$  says that *everything* has the property represented by  $\Phi$ , in which case, the variable may be called *implicitly bound*. Some authors employ different notation for free constants and (implicitly or explicitly) bound variables. Here, the semantics and the substitution rule presuppose that all second-order variables of  $\mathcal{L}2[K]-$  are *implicitly bound*. Assume that  $\Phi$  and  $\Psi$  are in  $\mathcal{L}2[K]-$ ,  $\Phi$  has only  $X$  free, and  $\Psi$  has no free variables. Suppose also that  $\Phi \vdash_{D2-} \Psi$ . Then in full second-order logic, we would have  $\forall X\Phi \vdash \Psi$ . So, from the deduction theorem,  $(\forall X\Phi) \rightarrow \Psi$  can be deduced from no premises. This is not a formula of  $\mathcal{L}2[K]-$ . In  $D2-$ , the conclusion of a deduction theorem would be  $\vdash \Phi(X) \rightarrow \Psi$ , which amounts to  $\vdash \forall X[\Phi(X) \rightarrow \Psi]$ .

So much for deduction. What of expressive resources? The usual categorical axiomatizations of arithmetic, analysis, complex analysis, Euclidean geometry, etc. each contain a finite number of first-order sentences and a single  $\Pi_1^1$  sentence. Thus, the axiomatization can be written in a free-variable second-order language. A categorical axiomatization of arithmetic consists of the conjunction of the usual first-order Peano axioms and the *induction* principle:

$$(X0 \& \forall x(Xx \rightarrow Xsx)) \rightarrow \forall xXx.$$

A categorical axiomatization of real analysis in a free-variable second-order language consists of the conjunction of the axioms of an ordered field, all of which are first-order, and the principle of *completeness* asserting that every bounded set has a least upper bound:

$$\exists x \forall y (Xy \rightarrow y \leq x) \rightarrow \exists x [\forall y (Xy \rightarrow y \leq x) \& \forall z (\forall y (Xy \rightarrow y \leq z) \rightarrow x \leq z)].$$

In the second-order axiomatization of Zermelo–Fraenkel set theory, every

axiom is first-order except the replacement principle, and that can be rendered in  $\mathcal{L}2[\{\in\}]$ –:

$$\forall x\forall y\forall z(Rxy\&Rxz \rightarrow y = z) \rightarrow \forall x\exists y\forall z(z \in y \equiv \exists w(w \in x\&Rwz)).$$

Thus, when it comes to the ability to characterize central structures, free-variable second-order languages have much of the strength of full second-order languages. A structure quasi-satisfies the axiomatization of arithmetic if and only if it is isomorphic to the natural numbers, and so all models of this axiomatization are countably infinite. A structure quasi-satisfies the axiomatization of analysis if and only if it is isomorphic to the real numbers, and so all such structures have the cardinality of the continuum. A structure quasi-satisfies the axiomatization of set theory if and only if it is isomorphic to an inaccessible rank (or  $V$  itself). Thus, both of the Löwenheim–Skolem theorems fail. There are no countable models of analysis and no uncountable models of arithmetic.

Compactness also fails. To see this add a constant  $c$  to the language of arithmetic and consider a set  $\Gamma$  consisting of the single free-variable second-order axiom of arithmetic and the sentences  $c \neq 0, sc \neq 0, ssc \neq 0, \dots$ . For any finite  $\Gamma' \subset \Gamma$ , one can interpret the constant  $c$  so that  $\Gamma'$  quasi-satisfies the natural numbers. However, there is no structure that quasi-satisfies  $\Gamma$  itself. Similarly, free-variable second-order logic is inherently incomplete, for much that same reason that full second-order logic is. The set of logical consequences of the axiomatization of arithmetic is not recursively enumerable. These results fail because of the expressive strength of the language.

There is some good news, however. Let  $\Phi$  be a formula of  $\mathcal{L}2[K]$ – and let  $\Phi'$  be the result of uniformly replacing each second-order predicate variable with a different non-logical relation letter of the same degree, not in  $\Phi$  already (expanding the set  $K$  if necessary). Then  $\Phi'$  is first-order. Notice that  $\Phi$  is a logical truth (i.e.,  $\Phi$  is quasi-satisfied by every structure) if and only if  $\Phi'$  is a logical truth. It follows that free-variable second-order logic is weakly complete. A formula  $\Phi$  is a logical truth if and only if  $\Phi$  can be deduced in  $D2$ – (without using the substitution rule!). So the relevant notion of logical truth is no more intractable than its first-order counterpart.

This small gain in manageability comes with the cost that free-variable second-order languages are not as expressive as full second-order languages. Let  $A$  be a monadic predicate letter and  $R$  a binary relation (both non-logical). In any interpretation of the language, the *minimal closure* of  $A$  under  $R$  is the smallest set that contains (the extension of)  $A$  and is closed under  $R$ . This is an important construction in mathematics and logic. In general, there is no first-order formula equivalent to ‘ $x$  is in the minimal closure of  $A$  under  $R$ ’ (see [Shapiro, 1991, Chapter 5, Section 1]). There is such a formula in  $\mathcal{L}2[K]$ –. One simply renders the informal definition:

$$\mathbf{MC}(x) : \forall y(Ay \rightarrow Xy)\&\forall y\forall z((Xy\&Ryz) \rightarrow Xz) \rightarrow Xx.$$

However, one cannot state in  $\mathcal{L}2[K]$ — that *there is* a minimal closure of  $A$  under  $R$ —which, in the full second-order system is an instance of the comprehension scheme. More importantly, the use of the minimal closure construction is hampered by the inability to directly state in  $\mathcal{L}2[K]$ — a conditional whose antecedent is ‘ $x$  is in the minimal closure of  $A$  under  $R$ ’. In such a conditional, the variable  $X$  would be implicitly bound by a universal quantifier whose scope is the *entire formula*. In the full second-order system,  $\forall X(\mathbf{MC}(x) \rightarrow \Phi)$  is not equivalent to  $\forall X(\mathbf{MC}(x) \rightarrow \Phi)$ , but only the latter is directly equivalent to a formula in  $\mathcal{L}2[K]$ — . This problem can sometimes be circumvented. Introduce a new non-logical predicate letter  $B$ , with the axiom:

$$\forall y(Ay \rightarrow By) \& \forall y \forall z ((By \& Ryz) \rightarrow Bz) \& \forall x (Bx \rightarrow \mathbf{MC}(x)).$$

This entails that the extension of  $B$  is coextensive with the indicated minimal closure. Then, to make the assertion that ‘if  $x$  is in the minimal closure of  $A$  under  $R$ , then  $\Phi$ ’ one would write

$$\forall x (Bx \rightarrow \Phi).$$

The extension of a predicate  $A$  is finite if there is no one-to-one function from this extension into a proper subset of itself. As noted above, one can just state this in  $\mathcal{L}2[\{A\}]$ — (using a relation variable instead of a function variable):

$$\mathbf{FIN}(A) : \neg[\forall x \forall y \forall z (Rxz \& Ryz \rightarrow x = y) \& \forall x (Ax \rightarrow \exists y (Rxy \& Ay)) \& \exists x (Ax \& \forall y (\neg Ryx))].$$

That is, a structure quasi-satisfies  $\mathbf{FIN}(A)$  if and only if the extension of  $A$  is finite. However, there is no general expression of the complement— infinitude—in this framework. The usual formula expressing infinitude has an *existential* quantifier ranging over relations:

$$\mathbf{INF}(A) : \exists R [\forall x \forall y \forall z (Rxz \& Ryz \rightarrow x = y) \& \forall x (Ax \rightarrow \exists y (Rxy \& Ay)) \& \exists x (Ax \& \forall y (\neg Ryx))].$$

See Shapiro [1991, Chapter 5, Section 1].

Another group of examples concerns the comparison of cardinalities. The formulation of ‘the cardinality of the extension of  $A$  is less than the cardinality of the extension of  $B$ ’,

$$\exists R [\forall x \forall y \forall z (Rxz \& Ryz \rightarrow x = y) \& \forall x (Ax \rightarrow \exists y (Rxy \& By)) \& \exists x (Bx \& \forall y (Ay \rightarrow \neg Ryx))],$$

has an initial existential quantifier, and so the relation cannot be characterized directly in  $\mathcal{L}2[K]$ — . But its complement ‘the cardinality of  $B$  is greater

than or equal to the cardinality of  $A$ ' can be:

$$\neg[\forall x\forall y\forall z(Rxz\&Ryz \rightarrow x = y)\&\forall x(Ax \rightarrow \exists y(Rxy\&By))\& \exists x(Bx\&\forall y(Ay \rightarrow \neg Ryx))],$$

but this last cannot be the antecedent of a conditional (with its intended meaning). Again, the notion 'A and B have the same cardinality' cannot be directly characterized, but its complement 'A and B have different cardinality' can be.

Similarly, one can assert in  $\mathcal{L}2[K]$ — that a given relation is well-founded, or is a well-ordering of its field, but the well-ordering *principle*, that every set has a well-ordering, cannot be stated. The latter requires an existential quantifier ranging over relations.

Many of these features are a consequence of the fact that  $\mathcal{L}2[K]$ — is not closed under contradictory opposition. It is clearly inconvenient to be unable to express the complements of otherwise definable properties and relations, and to be unable to use definable notions in the antecedents of conditionals.

As with monadic second-order logic, some of the motivation for free-variable second-order logic is philosophical. Recall that many thinkers balk at the automatic assumption of the existence of relations, no matter how they are construed. According to Quine, for example, if relations are intensional then they are too obscure for serious scientific work, and if they are extensional then we deal with sets, and have crossed the border out of logic and into mathematics. There is a second tradition, also due to Quine, that regards the ontology of a theory to consist of the range of its bound variables. The point is a simple one. The existential quantifier is a gloss on the ordinary word for existence. Thus, an interpreted theory in a formal language entails the existence of whatever falls under the range of an existential quantifier. In free-variable second-order languages, however, one cannot say that relations *exist*, since relation variables are not bound by quantifiers.

In another context, Hilbert (e.g., [1925]) made a similar distinction. Accordingly, a formula with a free variable expresses a certain generality, in that such a formula can be used to assert *each* of its instances. On the other hand, a formula with a bound variable—called an 'apparent variable'—represents or entails a genuine claim of existence. In articulating his finitism, Hilbert proposed that we develop theories that avoid reference to completed infinite sets. If such finitary theories are to capture any mathematics at all, the formulas need to express generality. His finitary formulas contain free variables, but he banned bound ('apparent') variables. Skolem [1923] expressed a similar idea:

[Arithmetic] can be founded in a rigorous way without the use of Russell and Whitehead's notions 'always' and 'sometimes'.

This can also be expressed as follows: a logical foundation can be provided for arithmetic without the use of apparent logical variables.

Again, the system that Skolem proposed allows free variables, but not bound variables.

Free-variable second-order languages exploit the Hilbert/Skolem distinction in the context of sets and relations—the items of second-order logic. A common complaint against second-order logic emerges from the belief that for a given infinite domain  $d$ , there is no clear understanding of the totality of the subsets of  $d$  (i.e., the powerset of  $d$ ). The skeptic points out that even the powerful axioms of Zermelo–Fraenkel set theory do not suffice to fix the powerset of the set of natural numbers, the simplest infinite powerset. But this powerset is the range of the predicate variables in second-order axiomatizations of arithmetic. The argument concludes that even if the purported range of the second-order variables is unambiguous, the range is too problematic to serve logic and foundational studies. Can one claim to have an intuitive grasp of statements in the form  $\forall X\exists Y\forall Z\Phi$ , even in a simple context like arithmetic?

This is not the place to respond to these skeptical arguments (see [Shapiro, 1991]). The point here is that much of the force of the arguments is deflected from free-variable, second-order logic. An advocate of a second-order free-variable system does not presuppose a far-reaching grasp of the range of the second-order variables. In fact, the advocate need not even presuppose that there is a fixed range of the relation variables. In typical cases, it is enough to recognize *unproblematic* definitions of relations on the domain, as they arise in practice. Recall that the only higher-order rule of inference allowed in the deductive system D2– is the ‘substitution rule’ allowing one to systematically replace a subformula  $Xt$ , for example, with  $\psi(t)$ . Since the formula  $\psi$  determines a subclass  $\{x \mid \psi(x)\}$  of the domain, the substitution rule is a version of universal instantiation. To put it loosely, the rule is that from  $\Phi(X)$ , one can infer  $\Phi(S)$ , where  $S$  is any set. If, in a given case, there is no unclarity about the set  $S$ , then there is no unclarity about the inference. In short, in  $\mathcal{L}2[K]$ –, a formula in the form  $\Phi(X)$  can be interpreted as ‘once a set  $S$  is determined,  $\Phi$  holds of it’.

Consider the axiom of induction in arithmetic, as formulated in a free-variable, second-order language:

$$(X0 \& \forall x(Xx \rightarrow Xsx)) \rightarrow \forall xXx,$$

As interpreted here, the principle asserts that any *given* set of natural numbers that contains 0 and is closed under the successor function contains all of the natural numbers. This axiom, so construed, is enough to establish the categoricity of arithmetic (together with the other axioms, of course). In the proof of categoricity (see [Shapiro, 1991, Chapter 4, Section 2] and,

of course, [Dedekind, 1988]), we consider two models  $M, M'$  of the theory. Using only weak and uncontroversial principles of set theory, one defines a subset  $c$  of the domain of  $M$  in terms of  $M'$  and a subset  $c'$  of the domain of  $M'$  in terms of  $M$ . Then  $c$  and  $c'$  are taken as instances of the induction axiom. That is, we have an application of universal instantiation. Notice that in order to apply the completeness axiom to  $c$ , one need only recognize that  $c$  is a subset of the domain of  $M$ . This, I suggest, is patently obvious (even though the definition of  $c$  goes beyond the resources of the corresponding first-order language). The conclusion is that one can work with theories formulated in free-variable second-order languages, and one can coherently maintain the categoricity of arithmetic (and analysis, Euclidean geometry, etc.) without claiming some sort of absolute grasp on the range of the relation variables—or even claiming that there is a fixed range. One only needs the ability to recognize subsets as they are defined; and in the context of the interpreted formal languages in question, this is not problematic. In sum, the free-variable second-order versions of the various theories involve only a rather weak hold on the range of the second-order variables.

Consider *first-order* logical consequence and logical truth. The standard definition is that a sentence  $\Phi$  is a logical truth if  $\Phi$  is satisfied by every model *under every interpretation* of its non-logical terminology. This is virtually the same as treating the non-logical terminology as implicitly bound free variables, and some of these variables are higher-order. Tarski [1935] explicitly uses second-order variables in his celebrated treatment of logical consequence. The reader is told to replace each non-logical term with an appropriate variable and consider the universal generalization of the formula that results. The free-variable, second-order language would have done just as well, and Tarski's procedure is recognized as the same as the contemporary one (modulo a few possible differences of no concern here). Thus, some grasp of second-order variables is even presupposed in standard treatments of first-order logic, so a skeptic about free-variable second-order languages should also be skeptical of common logical concepts.

The 'dual' to  $\mathcal{L}2[K]-$  would be a language that allowed only  $\Sigma_1^1$ -formulas—formulas with existential second-order quantifiers whose range is the entire formula. Call this a  $\Sigma$ -*language*. Like  $\mathcal{L}2[K]-$ , a  $\Sigma$ -language is not closed under contradictory opposition. The negation of a  $\Sigma_1^1$ -formula is a  $\Pi_1^1$ -formula. Thus, a given notion can be characterized in a  $\Sigma$ -language if and only if its complement can be characterized in a free-variable second-order language. So, for example, *infinitude* can be characterized, but not finitude.

Notice that the satisfiability of a first-order formula is equivalent to the satisfiability of the  $\Sigma_1^1$ -formula obtained by replacing the non-logical terminology with appropriate variables and binding the formula with existential quantifiers over the new variables. Thus, the downward and upward Löwenheim–Skolem theorems hold for  $\Sigma$ -languages. However, it follows from Church's theorem that the set of satisfiable first-order formulas is not

recursively enumerable. Thus, the set of logically true sentences in a  $\Sigma$ -language is also not recursively enumerable and so the logic of  $\Sigma$ -languages is not weakly complete. Flum [1985] establishes an interesting interpolation theorem for  $\Sigma$ -languages: let  $\Phi$  and  $\Psi$  be two sentences in a  $\Sigma$ -language such that  $\Phi \& \Psi$  has no models. Then there is a first-order sentence  $\chi$  such that  $\Phi \rightarrow \chi$  is a logical truth and  $\chi \& \Psi$  has no models. That is, any pair of incompatible sentences in a  $\Sigma$ -language can be ‘separated’ by a first-order sentence.

Consider a language containing both  $\Pi_1^1$ -formulas and  $\Sigma_1^1$ -formulas, but nothing more complex than that. This combines the expressive advantages of free-variable second-order languages and  $\Sigma$ -languages, but it also combines the disadvantages of both languages. The logic is not weakly complete and the Löwenheim–Skolem theorems fail. The language is closed under contradictory opposition, but not under conjunction or disjunction. We briefly return to a language consisting of Boolean combinations of  $\Pi_1^1$ -formulas in Section 6 below.

Note that someone might try to obtain the advantages of both monadic second-order logic and free-variable second-order logic by proposing a language with only free, monadic second-order variables. However, the main philosophical advantage to the monadic system was that the Boolos semantics could be used instead of standard semantics, thus easing the ontological burden of second-order logic. However, the Boolos construction invokes the plural quantifier from ordinary language, which is an *existential* quantifier: ‘there are objects ...’. The universal quantifier is obtained by way of negation:  $\forall X \Phi$  is just  $\neg \exists X \neg \Phi$ . Thus, the Boolos semantics is not available for the free-variable language.

#### 4 FINITUDE PRESUPPOSED

The main strength of full second-order languages is their ability to characterize important mathematical structures and concepts. The simplest infinite mathematical structure is surely that of the natural numbers, and one of the most basic mathematical concepts is finitude. The four logics presented in this section presuppose the notion of finitude/natural number, each in a different way. We can use the logics to see what can be captured in terms of, or relative to, finitude or the natural numbers. After characterizing each of the logics, we show how there is a sense in which they are equivalent to each other. Then their expressive resources are assessed, and they are compared to second-order logic. The succeeding sections take up extensions of these logics, which cover most of the (finitary) intermediate logics under study today.

*Weak second-order logic* employs the same languages as second-order logic, namely  $\mathcal{L}2[K]$ , except that there are no function variables (and we



maintain a symbol for identity). The difference with second-order logic lies in the model-theoretic semantics. In weak second-order logic, the second-order quantifiers range over *finite* relations. Let  $M$  be a model whose domain is  $d$ . Define  $s$  to be a *finite assignment* on  $M$  if  $s$  assigns a member of  $d$  to each first-order variable and a finite  $n$ -place relation on  $d$  to each  $n$ -place relation variable. For example, if  $X$  is an  $n$ -place relation variable, then  $s(X)$  is a finite subset of  $d^n$ . The semantics of weak second-order logic is restricted to finite assignments. The notion of a model  $M$  satisfying a formula under the finite assignment  $s$ , written  $M, s \models \Phi$ , is defined in the straightforward manner. The crucial clause is:

$$M, s \models \forall X \Phi \text{ if and only if } M, s' \models \Phi \text{ for every finite assignment } s' \text{ that agrees with } s \text{ except possibly at } X.$$

If the context does not make it clear which logic is under discussion, we employ the symbol  $\mathbf{w}$  for the satisfaction and consequence relation of weak second-order logic.

Some instances of the comprehension scheme are not logically true in weak second-order logic. In fact, a sentence in the form  $\exists X \forall x (Xx \equiv \Phi(x))$  is satisfied by a structure  $M$  if and only if the extension of  $\Phi$  in  $M$  is finite. Thus, the notion of *finitude* can be expressed in weak second-order logic, but if anything has the advantages of theft over toil, this does. The notion of finitude is built into the semantics from the outset. It does follow from the theft that weak second-order logic is more expressive than first-order logic, since the latter cannot express finitude.

The next logic  $\mathcal{L}(Q_0)$  (or  $\mathcal{L}(Q_0)[K]$ ) employs the language of first-order logic with identity  $\mathcal{L}1[K]=$ , augmented with another quantifier  $Q$ , called a *cardinality quantifier*. Let  $M$  be a model of  $\mathcal{L}1[K]=$  and  $s$  an assignment. The new clause in the semantics is:

$$M, s \models Qx\Phi \text{ iff there are infinitely many distinct assignments } s' \text{ such that } s \text{ agrees with } s' \text{ on every variable except possibly } x, \text{ and } M, s' \models \Phi.$$

The formula  $Qx\Phi$  may be read ‘for infinitely many  $x$ ,  $\Phi$ ’ or ‘ $\Phi$  holds of infinitely many  $x$ ’. The sentence  $Qx(x = x)$  asserts that the domain is infinite;  $Qx\Phi$  asserts that the extension of  $\Phi$  is infinite; and  $\neg Qx\Phi$  asserts that the extension of  $\Phi$  is finite. So we can express finitude in  $\mathcal{L}(Q_0)$ , again assuming the advantages of theft over toil. As above, it follows that  $\mathcal{L}(Q_0)$  is more expressive than first-order logic.

Our third logic is even more explicit about the theft. Assume that the set  $K$  of non-logical terms contains a binary relation symbol  $<$ . Let  $M$  be a model whose domain is  $d$ . Define the *field* of  $<$  in  $M$  to be the set

$$\{a \in d \mid M \models \exists x(a < x \vee x < a)\}.$$

That is, the field of  $<$  consists of the elements of the domain that are either ‘less than’ or ‘greater than’ something. Define  $M$  to be an  $\omega$ -model if the field of  $<$  in  $M$  is isomorphic to the natural numbers under the usual ‘less than’ relation. The idea here is to focus attention on  $\omega$ -models. We say that a set  $\Gamma$  of formulas of  $\mathcal{L}1[K]$  is  $\omega$ -satisfiable if there is an  $\omega$ -model  $M$  and an assignment  $s$  on  $M$  such that  $M, s \models \Phi$ , for every  $\Phi$  in  $\Gamma$ . A single formula  $\Phi$  is  $\omega$ -satisfiable if the singleton  $\{\Phi\}$  is  $\omega$ -satisfiable. And we say that a set of formulas  $\omega$ -implies  $\Phi$ , or  $\langle \Gamma, \Phi \rangle$  is  $\omega$ -valid, written  $\Gamma \vDash_{\omega} \Phi$ , if for every  $\omega$ -model  $M$  and assignment  $s$ , if  $M, s$  satisfies every member of  $\Gamma$ , then  $M, s$  satisfies  $\Phi$ . A formula  $\Phi$  is an  $\omega$ -logical truth if the empty set  $\omega$ -implies  $\Phi$ . For example,  $\exists x \forall y (\neg y < x)$  is an  $\omega$ -logical truth. The resulting system is called  $\omega$ -logic (see [Ebbinghaus, 1985]).

Let  $Sxy$  be an abbreviation of

$$x < y \& \neg \exists z (x < z \& z < y).$$

That is,  $Sxy$  asserts that  $y$  is the successor of  $x$  in the relation  $<$ . Notice that  $\forall x (\exists y (x < y) \rightarrow \exists ! y Sxy)$  is an  $\omega$ -logical truth.

To motivate the fourth logic considered in this section, recall that Frege [1979] defined the *ancestral*  $R^*$  of a given relation  $R$ , and he made brilliant use of this construction. To reiterate,  $R^*xy$  holds if there is a finite sequence  $a_0, \dots, a_n$  such that  $a_0 = x, a_n = y$ , and, for each  $i, 0 \leq i < n, Ra_i a_{i+1}$  holds. Equivalently,  $R^*xy$  if  $y$  is in the minimal closure of  $\{x\}$  under  $R$ . Consider the first-order language  $\mathcal{L}1[K]$ , augmented with an *ancestral operator*  $\mathbf{A}$ . If  $\Phi$  is a formula in which  $x$  and  $y$  occur free, and if  $t_1, t_2$  are terms, then  $\mathbf{A}xy(\Phi)t_1 t_2$  is a well-formed formula in which the variables  $x, y$  are bound. If  $M$  is a model and  $s$  an assignment to the variables, then  $M, s \models \mathbf{A}xy(\Phi)t_1 t_2$  if the denotation of  $t_2$  is an ancestor of the denotation of  $t_1$  under the relation (in  $M$ ) expressed by  $\Phi(x, y)$ . Call the resulting system *ancestral logic*. Immerman [1987] is an interesting treatment of (what amounts to) ancestral logic, restricted to finite models.

This completes the list of logics for this section. They are weak second-order logic,  $\mathcal{L}(Q_0)$ ,  $\omega$ -logic, and ancestral logic. The next item on the agenda is to assess and compare their expressive power.

It should not be surprising that in each of the four languages, there is a single sentence that characterizes the natural numbers, up to isomorphism, in the respective model theory:

**THEOREM 7.** *Assume that the set  $K$  includes  $\{0, s, +, \cdot, <\}$ , the non-logical terminology of arithmetic plus the  $<$  symbol. Each of the languages described in this section contains a sentence  $\Phi$  such that for each model  $M, M \models \Phi$  iff  $M$  is isomorphic to the natural numbers (with the usual operations and relations). In the terminology of [Barwise and Feferman, 1985], the collection of structures isomorphic to the natural numbers is an elementary class (EC) of weak second-order logic,  $\mathcal{L}(Q_0)$ ,  $\omega$ -logic, and ancestral*

logic.

**Proof.** Let  $\psi$  be the conjunction of the following (first-order) sentences.

$$\begin{aligned} & \forall x(sx \neq 0) \& \forall x \forall y (sx = sy \rightarrow x = y) \& \forall x (x \neq 0 \rightarrow \exists y (sy = x)) \\ & \hspace{15em} \text{(successor axiom)} \\ & \forall x (x + 0 = x) \& \forall x \forall y (x + sy = s(x + y)) \hspace{2em} \text{(addition axiom)} \\ & \forall x (x \cdot 0 = 0) \& \forall x \forall y (x \cdot sy = x \cdot y + y) \hspace{2em} \text{(multiplication axiom)} \\ & \forall x \forall y (x < y \equiv \exists z (x + sz = y)) \hspace{10em} \text{(order axiom)} \end{aligned}$$

In any model of the order, successor, and addition axioms, the field of  $<$  is the entire domain. Thus, it is straightforward that if  $M$  is an  $\omega$ -model of  $\psi$  then  $M$  is isomorphic to the natural numbers, and so  $\psi$  itself characterizes the natural numbers, up to isomorphism, in  $\omega$ -logic. In the other cases,  $\psi$  must be augmented with a statement that entails that  $0, s0, ss0, \dots$  (i.e. the minimal closure of  $0$  under  $s$ ) is the whole domain. In ancestral logic, there is a formula that just says this. Let  $\Phi_A$  be the following ancestral sentence:

$$\forall z (\mathbf{A}xy (y = sx) 0z),$$

In effect,  $\Phi_A$  asserts that everything is a successor-ancestor of  $0$ . So,  $\psi \& \Phi_A$  characterizes the natural numbers up to isomorphism in ancestral logic. Notice that it would also suffice to conjoin  $\psi$  with an assertion that for every object  $x$  there are only finitely many elements smaller than  $x$ . This can be said in  $\mathcal{L}(Q_0)$ . Let  $\Phi_Q$  be the following sentence:

$$\forall y \neg Qx (x < y).$$

Then  $\psi \& \Phi_Q$  is a categorical characterization of the natural numbers. Finally, for weak second-order logic, we add a statement asserting that for each  $x$  there is a *finite* set  $X$  that contains all of the elements smaller than  $x$ . Let  $\Phi_w$  be:

$$\forall x \exists X \forall y (y < x \rightarrow Xy).$$

Once again,  $\psi \& \Phi_w$  is a categorical characterization of the natural numbers. ■

The refutations of compactness, completeness and the upward Löwenheim–Skolem theorems for second-order logic only depend on the existence of a categorical characterization of the natural numbers (see [Shapiro, 1991, Chapter 4, Section 2]). Thus, these theorems fail for the logics under consideration here:

**COROLLARY 8.** *Let  $\mathcal{L}$  be weak second-order logic,  $\mathcal{L}(Q_0)$ ,  $\omega$ -logic, or ancestral logic. Then the upward Löwenheim–Skolem theorem fails for  $\mathcal{L}$ , and  $\mathcal{L}$  is not compact. Moreover, let  $D$  be any effective deductive system that is*

sound for  $\mathcal{L}$ . Then  $D$  is not (weakly) complete: there is a logical truth of  $\mathcal{L}$  that is not a theorem of  $D$ . In short,  $\mathcal{L}$  is inherently incomplete.

This summarizes the aforementioned theft.

Now for some toil. Let  $\mathcal{L}[K]$  and  $\mathcal{L}'[K]$  be languages based on the set  $K$  of non-logical terminology, and let each be equipped with a model-theoretic semantics involving the same class of models as the first-order  $\mathcal{L}1[K]=$ . Then  $\mathcal{L}'[K]$  is said to *include*  $\mathcal{L}[K]$ , written  $\mathcal{L}[K] \leq \mathcal{L}'[K]$ , if for each sentence  $\Phi$  of  $\mathcal{L}[K]$  there is a sentence  $\Phi'$  of  $\mathcal{L}'[K]$  such that for every model  $M$ ,  $M \models \Phi$  in  $\mathcal{L}[K]$  iff  $M \models \Phi'$  in  $\mathcal{L}'[K]$ . The idea is that  $\mathcal{L}'[K]$  is capable of expressing any distinctions among models that is expressible in  $\mathcal{L}[K]$ . In the terminology of [Barwise and Feferman, 1985],  $\mathcal{L}'[K]$  includes  $\mathcal{L}[K]$  if every elementary class of  $\mathcal{L}[K]$  is an elementary class of  $\mathcal{L}'[K]$ , in which case they say that  $\mathcal{L}'[K]$  is ‘as strong as’  $\mathcal{L}[K]$ . Under these circumstances, Cowles [1979] says that  $\mathcal{L}[K]$  is an ‘extension’ of  $\mathcal{L}[K]$ . If both  $\mathcal{L}[K] \leq \mathcal{L}'[K]$  and  $\mathcal{L}'[K] \leq \mathcal{L}[K]$ , the languages are said to be *equivalent*.<sup>6</sup>

We must extend this notion a bit to accommodate  $\omega$ -logic, since it does not have the same class of models as the first-order  $\mathcal{L}1[K]=$ . Assume that the set  $K$  contains the binary relation symbol  $<$ . Then we say that  $\mathcal{L}'[K]$  *includes  $\omega$ -logic* if, for each sentence  $\Phi$  of the first-order  $\mathcal{L}1[K]=$ , there is a sentence  $\Phi'$  of  $\mathcal{L}'[K]$  such that for each model  $M$ ,  $M \models \Phi'$  in  $\mathcal{L}'[K]$  if and only if  $M$  is an  $\omega$ -model and  $M \models \Phi$ . Note that we do not define the notion of  $\omega$ -logic including  $\mathcal{L}[K]$  (but see below for a variation on this theme). The following is immediate:

**LEMMA 9.** *Suppose that  $\mathcal{L}'[K]$  contains the connectives and quantifiers of the first-order  $\mathcal{L}1[K]=$ . Then  $\mathcal{L}'[K]$  includes  $\omega$ -logic if and only if there is a sentence  $\psi$  of  $\mathcal{L}'[K]$  whose only non-logical term is  $<$ , such that for each model  $M$ ,  $M \models \psi$  in  $\mathcal{L}'[K]$  if and only if the field of  $<$  in  $M$  is isomorphic to the natural numbers (i.e.,  $M$  is an  $\omega$ -model).*

This makes it straightforward to deal with  $\omega$ -logic.

**THEOREM 10.** *Weak second-order logic,  $\mathcal{L}(Q_0)$ , and ancestral logic all include  $\omega$ -logic.*

**Proof.** According to the above lemma, for each case, we need a sentence  $\psi$  whose only non-logical term is  $<$ , and which is satisfied by all and only  $\omega$ -models. Let  $\psi'$  be a (first-order) sentence asserting that the field of  $<$  is a non-reflexive, linear order of its field, and that every element in the field of  $<$  has a unique successor. For ancestral logic, we conjoin  $\psi'$  with an assertion that there is an element  $x$  such that every element in the field of  $<$  is an ancestor of  $x$  under the successor relation:

$$\exists x \forall y (\exists z (y < z \vee z < y) \rightarrow (\mathbf{A}pq(Spq)xy)).$$

<sup>6</sup>In these terms, the results reported in Theorems 1–5 above provide conditions under which a logic is equivalent to the first-order  $\mathcal{L}1[K]=$ .

For  $\mathcal{L}(Q_0)$ , we conjoin  $\psi'$  with an assertion that for each  $y$  there are only finitely many  $x$  such that  $x < y$ :

$$\forall y \neg Qx(x < y).$$

And for weak second-order logic, we conjoin  $\psi'$  with an assertion that for each  $y$ , there is a finite set  $X$  containing every element that 'precedes'  $y$  under  $<$ :

$$\forall y \exists X \forall x(x < y \rightarrow Xx).$$

■

This lends some precision to the remark that all of the logics considered here presuppose the natural numbers.

Suppose that  $\Phi$  is a formula of  $\mathcal{L}(Q_0)$  and  $\Phi'$  an equivalent formula in weak second-order logic. Then  $Qx\Phi$  is equivalent to  $\neg \exists X(\forall x(Xx \equiv \Phi'))$  (where  $X$  does not occur in  $\Phi'$ ). Thus a simple induction shows the following:

**THEOREM 11.** *Weak second-order logic includes  $\mathcal{L}(Q_0)$ .*

Now let  $\Phi$  be a formula of an ancestral language and  $\Phi'$  an equivalent formula in the corresponding language of weak second-order logic. Then there is a formula equivalent to  $\mathbf{A}xy(\Phi)t_1t_2$  in weak second-order logic. It is tedious to write out the ancestral formula, but it goes like this:

Either  $t_1 = t_2$  or else there is a finite binary relation  $Y$  such that (1) the extension of  $Y$  is a sub-relation of the extension of  $\Phi'$  ( $\forall x \forall y(Yxy \rightarrow \Phi'((x, y)))$ ), (2)  $Y$  is the graph of a one-to-one function  $f$  whose domain is a subset of the domain of discourse, (3)  $t_2$  is in the range of  $f$  ( $\exists x Yxt_2$ ), (4)  $t_1$  is in the domain of  $f$  ( $\exists x Yt_1x$ ), (5)  $t_1$  is not in the range of  $f$  ( $\neg \exists x Yxt_1$ ), and (6) if  $y$  is in the range of  $f$  and  $y \neq t_2$  then  $y$  is in the domain of  $f$  ( $\forall x \forall y(Yxy \& y \neq t_2 \rightarrow \exists z Yyz)$ ).

So an induction establishes

**THEOREM 12.** *Weak second-order logic includes ancestral logic.*

This completes the list of inclusion relations among the logics of this section. The other combinations fail.

**THEOREM 13.**  *$\mathcal{L}(Q_0)$  does not include weak second-order logic [Cowles, 1979] and  $\mathcal{L}(Q_0)$  does not include ancestral logic.*

**Proof.** Let  $\Phi$  be the conjunction of the (first-order) axioms for an ordered field and let  $\Phi_w$  be

$$\forall x \exists X(X1 \& \forall y((y < x \& Xy) \rightarrow X(y+1))).$$

In weak second-order logic, this sentence asserts that for each  $x$  there is a finite set that contains all of the positive integers less than  $x$ . That is,  $\Phi_w$

says that for each number  $x$ , there are only finitely many positive integers less than  $x$ . In effect,  $\Phi_w$  entails that the structure is *Archimedean*. Thus,  $M \models \Phi \& \Phi_w$  iff  $M$  is an Archimedean field. Similarly, let  $\Phi_A$  be

$$\forall x(1 < x \rightarrow \exists y[\mathbf{A}pq(q = p + 1)1y \& x < y]).$$

The sentence  $\Phi_A$  asserts that for every  $x$ , there is a positive integer  $y$  that is larger than  $x$ . Thus,  $M \models \Phi \& \Phi_A$  iff  $M$  is an Archimedean field. Thus, the class of Archimedean fields is an elementary class of both weak second-order logic and ancestral logic. On the other hand, Cowles [1979] shows that Tarski's theorem concerning the completeness of first-order analysis can be extended to  $\mathcal{L}(Q_0)$ . In particular, for each formula  $\chi$  of  $\mathcal{L}(Q_0)$  whose non-logical terminology is in  $\{0, 1, +, \cdot, <\}$ , there is a formula  $\chi'$ , with the same free variables as  $\chi$ , such that  $\chi'$  has no quantifiers and  $\chi \equiv \chi'$  holds in all models of the theory of real closed fields—first-order analysis. Now, if  $\mathcal{L}(Q_0)$  included weak second-order logic or ancestral logic, it would contain a sentence  $\Phi'$  that is a correlate of  $\Phi \& \Phi_w$  or  $\Phi \& \Phi_A$  above. That is,  $\Phi'$  would be satisfied by all and only Archimedean fields. Let  $\Phi''$  be its quantifier-free equivalent. But a compactness argument establishes that there is no first-order sentence that is satisfied by all and only Archimedean fields (see [Shapiro, 1991, Chapter 5]). ■

There is a sense in which the notion of an Archimedean field can be characterized in  $\omega$ -logic. Given the way we have set up  $\omega$ -logic, we require separate symbols for the 'less than' relation on the domain and the 'less than' relation on the 'natural numbers'. Let us use ' $\prec$ ' for the latter. Let  $\Phi_\omega$  be a (first-order) sentence asserting that (1) 0 is the smallest element in the field of  $\prec$  ( $\forall x(\exists y(x \prec y \vee y \prec x) \rightarrow (x = 0 \vee 0 \prec x))$ ) and (2) for each  $x$  in the field of  $\prec$ , the successor of  $x$  is  $x + 1$  ( $\forall x(\exists y(x \prec y \vee y \prec x) \rightarrow (x \prec x + 1 \& \neg \exists z(x \prec z \& z \prec x + 1)))$ ). Then any  $\omega$ -model of  $\Phi \& \Phi_\omega$  is an Archimedean field (in which the field of  $\prec$  is the 'natural numbers' of the model). It follows from Theorem 10 that there is a sentence  $\Phi^*$  of  $\mathcal{L}(Q_0)$  that is equivalent to  $\Phi \& \Phi_\omega$ . Every model of  $\Phi^*$  is an Archimedean ordered field and any Archimedean ordered field can be made into a model of  $\Phi^*$ . However, this does not contradict the result from [Cowles, 1979] cited in Theorem 13, since  $\Phi^*$  contains another non-logical constant, namely ' $\prec$ '.

**THEOREM 14.** *Ancestral logic does not include  $\mathcal{L}(Q_0)$  or weak second-order logic.*

**Proof.** Let the set  $K$  contain only monadic predicate letters, and let  $\Phi$  be a formula of the first-order  $\mathcal{L}1[K]$ . Then it can be shown that there is a natural number  $n$  such that for any model  $M$  and assignment  $s$  on  $M$ ,  $M, s \models \mathbf{A}xy(\Phi)pq$  iff

$$M, s \models \exists x_1 \dots \exists x_n (x_1 = p \& x_n = q \& (\Phi(x_1, x_2) \vee x_1 = x_2) \& \dots \& (\Phi(x_{n-1}, x_n) \vee x_{n-1} = x_n))$$

(where  $x_1 \dots x_n$  do not occur in  $\mathbf{A}xy(\Phi)pq$ , relettering if necessary). The implication from right to left is immediate. The converse is a consequence of the proof of the decidability of the monadic predicate calculus (see [Dreben and Goldfarb, 1979, Section 8.3]). Thus, for this set  $K$  of non-logical terminology, ancestral logic is equivalent to the first-order  $\mathcal{L}1[K]=$ . Let  $D$  be a monadic predicate letter in  $K$ . It follows from the above, and another result reported in [Dreben and Goldfarb, 1979], that there is no sentence of ancestral logic equivalent to either the sentence  $QxDx$  of  $\mathcal{L}(Q_0)$  or the sentence  $\neg\exists X\forall x(Xx \equiv Dx)$  of weak second-order logic, each of which asserts that the extension of  $D$  is infinite. ■

I suggest that the ‘non-inclusions’ are artifacts of a restriction on the non-logical terminology. We have already seen that  $\mathcal{L}(Q_0)$  can express the notion of an Archimedean ordered field if the non-logical terminology is expanded to include a symbol ‘ $\prec$ ’ for the ‘less than’ relation on the ‘natural numbers’ of the structure. In the terminology of [Barwise and Feferman, 1985], the class of Archimedean fields is a *projective class* (PC) of  $\mathcal{L}(Q_0)$ . Similarly, the class of structures in which the extension of a predicate letter  $D$  is infinite is a projective class of first-order logic, and thus, of ancestral logic.

The relevant insight here is that the four logics under study are equivalent in the sense that any class of (infinite) structures characterized by one of them can be characterized by any of the others, if one can add non-logical terminology. This can be made precise, using the resources of model-theoretic logic.

Let  $K$  and  $K'$  be sets of non-logical terminology such that  $K \subseteq K'$ . Let  $M$  be a model of the first-order language  $\mathcal{L}1[K]=$ , and let  $M'$  be a model of  $\mathcal{L}1[K']=$ . We say that  $M'$  is an *expansion* of  $M$  if they have the same domain and agree on the interpretation of the items in  $K$ .

Let  $\mathcal{L}[K]$  and  $\mathcal{L}'[K]$  be languages built on a set  $K$  of non-logical terminology, and assume that each is equipped with a semantics involving the usual class of models. We say that  $\mathcal{L}'[K]$  *quasi-projects*  $\mathcal{L}[K]$  if for each sentence  $\Phi$  of  $\mathcal{L}[K]$ , if  $\Phi$  has only infinite models, then there is a set  $K' \supseteq K$  and a sentence  $\Phi'$  of  $\mathcal{L}'[K']$ , such that for each model  $M$ ,  $M \models \Phi$  in  $\mathcal{L}[K]$  iff there is an expansion  $M'$  of  $M$  such that  $M' \models \Phi'$  in  $\mathcal{L}'[K']$ . In the terminology of [Barwise and Feferman, 1985],  $\mathcal{L}'[K]$  quasi-projects  $\mathcal{L}[K]$  iff every elementary class of  $\mathcal{L}[K]$  that contains only infinite structures is a projective class of  $\mathcal{L}'[K]$ . We say that  $\mathcal{L}[K]$  and  $\mathcal{L}'[K]$  are *quasi-projectively-equivalent* if  $\mathcal{L}[K]$  quasi-projects  $\mathcal{L}'[K]$  and  $\mathcal{L}'[K]$  quasi-projects  $\mathcal{L}[K]$ . The restriction to sentences without finite models is admittedly inelegant, but convenient here.<sup>7</sup>

<sup>7</sup>The standard model-theoretic notion is that  $\mathcal{L}'[K]$  *projects*  $\mathcal{L}[K]$  if for each sentence  $\Phi$  of  $\mathcal{L}[K]$ , there is a set  $K' \supseteq K$  and a sentence  $\Phi'$  of  $\mathcal{L}'[K']$  such that for each structure  $M$ ,  $M \models \Phi$  in  $\mathcal{L}[K]$  iff there is an expansion  $M'$  of  $M$  such that  $M' \models \Phi'$  in  $\mathcal{L}'[K']$ . The

The special symbol ‘<’ of  $\omega$ -logic complicates the definition. We say that  $\mathcal{L}'[K]$  *quasi-projects  $\omega$ -logic* if, for each sentence of the first-order  $\mathcal{L}1[K]=$ , in which  $K$  includes the symbol <, there is a set  $K' \supseteq K$  and a sentence  $\Phi'$  of  $\mathcal{L}'[K']$ , such that (1) for each structure  $M$  for the language  $\mathcal{L}'[K']$ , if  $M \models \Phi'$  in  $\mathcal{L}'[K']$ , then  $M$  is an  $\omega$ -model and  $M \models \Phi$ ; and (2) for each  $\omega$ -model  $M$  of  $\mathcal{L}1[K]=$ , if  $M \models \Phi$  then there is an expansion  $M'$  of  $M$  such that  $M' \models \Phi'$  in  $\mathcal{L}'[K']$ .

Conversely, we say that  $\omega$ -logic *quasi-projects  $\mathcal{L}[K]$*  if, for each sentence  $\Phi$  of  $\mathcal{L}[K]$ , in which  $K$  does *not* include the symbol <, if  $\Phi$  has only infinite models, then there is a set  $K' \supseteq K$  such that the symbol < is in  $K'$ , and there is a sentence  $\Phi'$  of the first-order  $\mathcal{L}1[K']=$ , such that for each structure  $M$ ,  $M \models \Phi$  in  $\mathcal{L}[K]$  iff there is an expansion  $M'$  of  $M$  such that  $M'$  is an  $\omega$ -model and  $M' \models \Phi'$ . And we say that  $\mathcal{L}[K]$  is *quasi-projectively equivalent to  $\omega$ -logic* if  $\omega$ -logic projects  $\mathcal{L}[K]$  and  $\mathcal{L}[K]$  projects  $\omega$ -logic.

We come, finally, to the equivalence of the logics of this section:

**THEOREM 15.** *Weak second-order logic,  $\mathcal{L}(Q_0)$ , ancestral logic, and  $\omega$ -logic are quasi-projectively equivalent to each other.*

**Proof.** Notice, first, that if  $\mathcal{L}'[K]$  includes  $\mathcal{L}[K]$ , then  $\mathcal{L}'[K]$  quasi-projects  $\mathcal{L}[K]$ . It follows from this, Theorems 10, 11 and 12, and the various definitions, that it suffices to show that  $\omega$ -logic quasi-projects weak second-order logic. This is accomplished by adding terminology for coding ‘finite sets’. The idea is to use a binary relation to represent some subsets of a domain (see [Shapiro, 1991, Chapter 5]). Let  $R$  be a binary relation, and define  $R_x$  to be the set  $\{y \mid Rxy\}$ . We say that  $R_x$  is the set *coded by  $x$  in  $R$* , and the relation  $R$  *represents* the collection of all the sets  $R_x$ , where  $x$  ranges over the domain of discourse. Of course, no relation can represent every subset of the domain (Cantor’s theorem), but if a domain is infinite, then there is a relation that represents the collection of its *finite* subsets. The plan here is to show that such a relation can be characterized in  $\omega$ -logic. Let  $E$  be a binary non-logical relation symbol (not in the given set  $K$  of non-logical terminology) and let  $\psi_1$  be the following (first-order) sentence:

$$\exists x \forall y (\neg Exy) \& \forall x \forall y \exists z \forall w (Ezw \equiv (Exw \vee w = y)).$$

---

logics  $\mathcal{L}[K]$  and  $\mathcal{L}'[K]$  are *PC-equivalent* if each one projects the other. Ancestral logic and  $\mathcal{L}(Q_0)$  are PC-equivalent (see [Shapiro, 1991, Chapter 9, Section 9.1.2]). The question of whether  $\mathcal{L}(Q_0)$  and ancestral logic project weak second-order logic is equivalent to the proposition that for every sentence  $\Phi$  of the second-order  $\mathcal{L}2[K]$ , there is a  $\Sigma_1^1$  sentence  $\Phi^*$  (also of  $\mathcal{L}2[K]$ ) such that for each finite structure  $M$ ,  $M \models \Phi$  iff  $M \models \Phi^*$  (see [Shapiro, 1991, Chapter 9, Section 9.1.2]). This, in turn, is equivalent to the longstanding open problem in complexity theory concerning whether the properties of finite structures recognized by NP algorithms include the full polynomial-time hierarchy. For the relevant complexity results see [Fagin, 1974; Immerman, 1987; Gurevich, 1988; Leivant, 1989], as well as the wealth of papers cited there.



The first conjunct of  $\psi_1$  asserts that the empty set is coded by something in  $E$ , and the second conjunct asserts that if a set  $X$  is coded in  $E$  then, for any element  $y$ ,  $X \cup \{y\}$  is coded in  $E$ . Thus  $\psi_1$  entails that every finite subset of the domain is coded in  $E$ . It remains to assert that *only* finite sets are coded in  $E$ . For this, the resources of  $\omega$ -logic are employed. We introduce a non-logical binary relation  $N$  (not in  $K$ ) such that  $Nxy$  entails  $x$  is in the field of  $<$ , and the cardinality of  $E_y$  is the natural number corresponding to  $x$ . In particular, let  $\psi_2$  be the conjunction of (1) the assertion that if  $x$  is the initial element of  $<$ , then  $Nxy$  holds iff  $\forall z(\neg Eyz)$  and (2) if  $x'$  is the successor of  $x$  in  $<$ , then  $Nx'y$  holds if there is a  $w$  and a  $z$  such that  $Nxw, z$  is not in  $E_w$ , and  $E_y$  is  $E_w \cup \{z\}$ . Finally, let  $\psi_3$  be  $\forall y \exists x (\exists z (x < z) \& Nxy)$ . That is  $\psi$  asserts that for every  $y$  there is an  $x$  in the field of  $<$  that represents the cardinality of  $E_y$ . So let  $\psi$  be  $\psi_1 \& \psi_2 \& \psi_3$ . In any  $\omega$ -model of  $\psi$ ,  $E$  represents the set of all finite subsets of the domain. Let  $\Phi$  be a sentence of weak second-order logic that has no finite models. Then terminology for a pairing function can be introduced, and there is a sentence  $\Phi'$  containing only monadic second-order variables that has the same models as  $\Phi$ . Assume that the relation letters  $<, E, N$  do not occur in  $\Phi'$  (relettering if necessary). To each second-order variable  $X$  that occurs in  $\Phi'$ , associate a unique first-order variable  $x_X$  that does not occur in  $\Phi'$ . Let  $\Phi''$  be the result of replacing each subformula  $Xt$  of  $\Phi'$  with  $Ex_X t$  (i.e. the formula asserting that  $t$  is in the set represented by  $x_X$  in  $E$ ) and replacing each quantifier  $\forall X$  by  $\forall x_X$ . The result is a first-order sentence. Finally, let  $\chi$  be  $\psi \& \Phi''$ . It is routine to establish that for each model  $M$ ,  $M \models \Phi$  iff there is an expansion of  $M$  that satisfies  $\chi$ . ■

Enough comparison. It should be clear that the languages of this section do not have all of the shortcomings of first-order languages, even if some of this comes by way of theft. The natural numbers can be characterized up to isomorphism, and minimal closures of definable sets and relations can be characterized (e.g., in terms of the ancestral). Slightly less trivially, the rational numbers can be characterized up to isomorphism as an infinite field whose domain is the minimal closure of  $\{1\}$  under the field operations and their inverses. As noted above, the logics are not compact and the upward Löwenheim–Skolem theorem fails.

When it comes to expressive resources, the logics presented here fall well short of second-order languages. Recall that the stronger version of the *downward* Löwenheim–Skolem theorem is that for every structure  $M$  with an infinite domain there is an elementarily equivalent *submodel*  $M'$  whose domain is countable. A routine check of the usual proof will verify that if the original structure  $M$  is an  $\omega$ -model, then the countable submodel  $M'$  is also an  $\omega$ -model. Thus, the downward Löwenheim–Skolem theorem holds for  $\omega$ -logic. It follows from the proofs of the above comparison results that the downward theorem holds for weak second-order logic, ancestral logic,

and  $\mathcal{L}(Q_0)$ . In technical terms, the Löwenheim number of each logic is  $\aleph_0$ .

It follows, of course, that the real numbers cannot be characterized up to isomorphism in any of these languages. Nevertheless, the versions of real analysis in these languages are improvements over the first-order version of the theory. For example, to repeat a result rehearsed above, one can guarantee that every model of real analysis is Archimedean (employing extra terminology if needed), and thus one can guarantee that every model is isomorphic to a *subset* of the real numbers. To speak loosely, with the present languages, we cannot establish the existence of every real number, but at least extraneous ‘numbers’—infinitesimals for example—can be excluded. Moreover, the ‘natural numbers’ and the ‘rational numbers’ of each model can be characterized up to isomorphism.

Set theorists define an  $\omega$ -*model* to be a model of the axioms of set theory in which the extension of ‘finite ordinal’ is isomorphic to the natural numbers. This usage of the term is in line with the present one:

**THEOREM 16.** *Let  $Z\omega$  be Zermelo–Fraenkel set theory (ZFC) where the replacement scheme is expanded to include formulas with a new binary relation symbol  $<$ . Let  $M$  be a model of ZFC. Then the extension of ‘finite ordinal’ in  $M$  is isomorphic to the natural numbers if and only if there is a way to interpret the symbol  $<$  in  $M$  so that the field of  $<$  is isomorphic to the natural numbers and the expanded model satisfies  $Z\omega$ . In other words,  $M$  is an  $\omega$ -model of ZFC in the set theorists’ sense if and only if  $M$  is a  $\omega$ -model of  $Z\omega$  in the present sense.*

**Proof.** [Sketch] If the extension of ‘finite ordinal’ is isomorphic to the natural numbers, then make the field of  $<$  the finite ordinals of the model, with membership as the order relation. The result is an  $\omega$ -model (in the present sense) of  $Z\omega$ . For the converse, if  $M$  is an  $\omega$ -model (in the present sense) of  $Z\omega$  then we can define a one-to-one function from the field of  $<$  onto the finite ordinals of  $M$ . It follows from replacement that the field of  $<$  is a set and that it is isomorphic to the finite ordinals of  $M$ . *A fortiori*,  $M$  is an  $\omega$ -model in the set-theorists’ sense ■

The above comparison results yield the following:

**COROLLARY 17.** *A structure  $M$  is an  $\omega$ -model of ZFC if and only if  $M$  is a model of the version of ZFC formulated in weak second-order logic,  $\mathcal{L}(Q_0)$ , and ancestral logic—in each case the replacement scheme is expanded to include formulas with the new vocabulary.*

Thus, the move to one of the languages under study in the present section is an improvement over first-order ZFC. There are, however, countable models of the indicated set theories. Moreover, there are models in which the membership relation is not well-founded and so there are models in which some members in the extension of ‘ordinal’ are not well-ordered under membership.

The general notion of well-ordering cannot be characterized in any of the logics under study here. In particular, let  $\omega_1^{\text{CK}}$  (the ‘Church-Kleene  $\omega_1$ ’) be the least upper bound of all ordinals  $\alpha$  such that there is a recursive well-ordering of the natural numbers whose order-type is  $\alpha$ . If  $\Phi$  is a sentence of one of our languages all of whose models are well-orderings, then there is no model of  $\Phi$  whose order type is  $\omega_1^{\text{CK}}$  or any ordinal greater than  $\omega_1^{\text{CK}}$ . In the terminology of Barwise and Feferman [1985],  $\omega_1^{\text{CK}}$  is not ‘pinned-down’ by weak second-order logic, ancestral logic,  $\omega$ -logic, or  $\mathcal{L}(Q_0)$  and, in fact,  $\omega_1^{\text{CK}}$  is the ‘bound’ of these languages (see [Ebbinghaus, 1985, Section 5.2]).

Barwise [1985] indicates that the Beth definability property fails for the logics of this section, for much the same reason that the general property fails for second-order logic. The interpolation property also fails. Details of the results reported here, and a host of other information about weak second-order logic,  $\omega$ -logic, and  $\mathcal{L}(Q_0)$  can be found in the papers published in [Barwise and Feferman, 1985], especially [Barwise, 1985; Ebbinghaus, 1985; Väanänen, 1985].

## 5 MORE THEFT, MORE TOIL

Since finitude is probably the simplest notion that goes beyond the resources of first-order logic, the logics of the previous section are the minimal intermediate systems. Many of the other intermediate logics under study today are obtained by presupposing other, richer mathematical structures and notions. Recall that the logics of the previous section are all equivalent to each other, in one sense or another. One surprising result is some of the corresponding equivalences fail in the extended cases. Moreover, some of the extended logics are *more* tractable than those of the previous section. Completeness is regained in one case. This section contains a brief account of some of the systems, but it seems to me that as the presupposition—the theft—increases, the philosophical interest and application decreases. I have no desire to legislate or predict what will or will not attract the attention of philosophers.

The system of  $\omega$ -logic is an example of what Ebbinghaus [1985] calls a ‘logic with a *standard part*’. Each  $\omega$ -model includes a copy of the ‘standard’ natural numbers, and the language has the resources to refer to this standard part. In effect, the natural numbers are included in  $\omega$ -logic by fiat. To extend the idea, let  $L$  be any set of non-logical terminology not containing a monadic predicate  $U$ , and let  $\mathfrak{R}$  be any class of structures on the first-order language  $\mathcal{L}[L]$ . We assume that  $\mathfrak{R}$  is closed under isomorphism. Let  $K$  be a set of non-logical terminology such that  $L \subseteq K$  and  $U \in K$ . A structure  $M$  of  $\mathcal{L}[K]$  is an  $\mathfrak{R}$ -model if the restriction of  $M$  to the extension of  $U$  (and the terminology in  $L$ ) is a structure in  $\mathfrak{R}$ . In other words, an  $\mathfrak{R}$ -model has a definable substructure that is a member of  $\mathfrak{R}$ .

The resulting system may be called  $\mathfrak{R}$ -logic, written  $\mathcal{L}(\mathfrak{R})[K]$ . A structure can be characterized, up to isomorphism, in  $\mathfrak{R}$ -logic if and only if it can be characterized in terms of  $\mathfrak{R}$ .

For example,  $\mathbb{R}$ -logic would be the restriction of first-order logic to models that contain an isomorphic copy of the real numbers. ZFC-logic would be the restriction of first-order logic to models that contain an isomorphic copy of an inaccessible rank. ZFC-logic would be the proper framework for those philosophers and logicians who advocate first-order set theory, interpreted standardly, as the foundation of mathematics. That is, ZFC-logic might be a good framework for anyone who wants to develop this foundation more fully. The proof of the Löwenheim–Skolem property can be adapted to establish a weaker version for  $\mathbb{R}$ -logic: for any  $\mathbb{R}$ -model  $M$ , there is an elementarily equivalent  $M'$  whose domain is the cardinality of the continuum (or size of the non-logical terminology, whichever is larger). Ebbinghaus reports that if  $\mathfrak{R}$  is the set of linear orders in which every initial segment is countable—the  $\aleph_1$ -like orders—then  $\mathfrak{R}$ -logic is  $\aleph_0$ -compact: a countable set of sentences has an  $\mathfrak{R}$ -model if each finite subset does.

Moving on, define *quasi-weak second-order logic* to be like weak second-order logic, but with variables ranging over *countable* relations. That is, quasi-weak second-order logic has the same formulas as full second-order languages, but in the semantics each variable assignment consists of a function from the first-order variables to the domain (as usual) and a function from the relation variables to countable relations. So  $\forall X \Phi$  can be read, ‘for all countable  $X$ ,  $\Phi$ ’. Quasi-weak second-order logic is equivalent to augmenting  $\omega$ -logic with bound variables ranging over functions whose domain is the field of  $<$  (i.e. the collection of ‘natural numbers’).

The following is a variation of the above second-order formula that asserts that the extension of  $X$  is finite (see Section 3.2):

$$\forall R \neg [\forall x \forall y \forall z (Rxz \& Ryz \rightarrow x = y) \& \forall x \forall y (Rxy \rightarrow (Xx \& Xy \& \exists z Ryz)) \& \exists x \exists y (Rxy \& \forall z (\neg Rzx))].$$

As interpreted in quasi-weak second-order logic, this formula asserts that there is no *countable* one-to-one relation whose domain and range are contained in  $X$  and whose range is a proper subset of its domain. This is clearly a necessary and sufficient condition for the extension of  $X$  to be finite. It follows that quasi-weak second-order logic includes weak second-order logic.

The converse fails, however: weak second-order logic does not include quasi-weak second-order logic. To see this, consider the completeness principle for real analysis:

$$\forall X (\exists x \forall y (Xy \rightarrow y \leq x) \rightarrow \exists x [\forall y (Xy \rightarrow y \leq x) \& \forall z (\forall y (Xy \rightarrow y \leq z) \rightarrow x \leq z)]).$$

As interpreted in quasi-weak second-order logic, this sentence asserts that every bounded, *countable* set has a least upper bound. This, with the other

axioms for real analysis, is sufficient to establish the categoricity of the theory—all of its models are isomorphic to the real numbers. Since there is no categorical axiomatization of real analysis in weak second-order logic, we conclude that weak second-order logic does not include quasi-weak second-order logic.

Let  $\mathbf{Z2q}$  be the axioms of ZFC formulated in a quasi-weak second-order language. This theory employs a replacement scheme, one instance for each formula of the quasi-weak second-order language. Let  $M$  be a model of  $\mathbf{Z2q}$  and let  $c$  be a countable subset of the domain of  $M$ . Then there is a member of the domain of  $M$  whose ‘elements’ in  $M$  are the members of  $c$ . In other words, every countable *class* is a *set*. It follows from this and the axiom of foundation (and choice) that the membership relation of  $M$  is well-founded. Thus,  $M$  is isomorphic to a transitive set  $m$  under membership. Also,  $m$  contains all of its countable subsets. This is a major improvement over the versions of set theory formulated in weak second-order logic,  $\omega$ -logic, ancestral logic, and  $\mathcal{L}(Q_0)$ .

In general, the notion of well-foundedness can be formulated in quasi-weak second-order logic. Let  $\mathbf{WO}(R)$  be the assertion that  $R$  is a linear order and that every countable set of the domain has a ‘least element’ under  $R$ . This is a straightforward sentence in a quasi-weak second-order language. Assuming the axiom of choice in the meta-theory, it follows that  $\mathbf{WO}(R)$  is satisfied by a structure if and only if  $R$  is a well-ordering of the domain.

We are near the limit of the expressive resources of quasi-weak second-order languages. The categoricity of real analysis entails that the full downward Löwenheim–Skolem theorem fails for quasi-weak second-order logic. There is, however, an attenuated version of the theorem, similar to the one for  $\mathbb{R}$ -logic: for every structure  $M$ , there is a substructure  $M'$  such that the cardinality of the domain of  $M'$  is at most that of the continuum (or the size of the non-logical terminology, whichever is larger) and  $M$  and  $M'$  satisfy the same formulas of the quasi-weak second-order language.

Of course, we need not stop here. One can construct languages with variables ranging over relations of cardinality  $\aleph_1$  or the ever present  $\aleph_{17}$ , or the first measurable cardinal, etc. If the cardinality in question is definable in a second-order language, then the system is intermediate between first-order and second-order (see [Shapiro, 1991, Chapter 5, Section 5.1]).

The extensions of  $\mathcal{L}(Q_0)$  have attracted more attention from logicians than the other logics in this section—even more than  $\mathcal{L}(Q_0)$  itself. For each ordinal  $\alpha$ , there is a logic  $\mathcal{L}(Q_\alpha)$ , with the same language as  $\mathcal{L}(Q_0)$ . That is,  $\mathcal{L}(Q_\alpha)[K]$  is obtained from the first-order  $\mathcal{L1}[K]$  by adding a (monadic) quantifier  $Q$ . Let  $M$  be a structure and  $s$  an assignment to the variables of the language. The model-theoretic semantics of  $\mathcal{L}(Q_\alpha)$  includes the following clause:

$M, s \models Qx\Phi$  if there are at least  $\aleph_\alpha$  distinct assignments  $s'$  such

that  $s'$  agrees with  $s$  on every variable except possibly  $x$ , and  
 $M, s' \models \Phi$ .

In other words, in  $\mathcal{L}(Q_\alpha)$   $Qx\Phi$  amounts to ‘there are at least  $\aleph_\alpha$ -many  $x$  such that  $\Phi$ ’. In  $\mathcal{L}(Q_1)$ ,  $Qx\Phi$  comes to ‘there are uncountably many  $x$  such that  $\Phi$ ’.

It is surprising that  $\mathcal{L}(Q_1)$  has many of the model-theoretic properties of first-order logic. The logic is  $\aleph_0$ -compact, in the sense that if every finite subset of a countable set  $S$  of sentences in the language is satisfiable, then  $S$  itself is satisfiable. Moreover, the logic is weakly complete, and the set of consequences of a recursively enumerable set of sentences is itself recursively enumerable. To obtain a sound and complete deductive system for  $\mathcal{L}(Q_1)$ , one adds the following axioms to a complete axiomatization for first-order logic:

Bound variables can be renamed:  $Qx\Phi(x) \rightarrow Qy\Phi(y)$ , provided  $y$  is not free in  $\Phi(x)$ .

A set of two elements is countable:  $\forall y\forall z\neg Qx(x = y \vee x = z)$ .

The new quantifier is ‘monotone’:  $\forall x(\Phi \rightarrow \Psi) \rightarrow (Qx\Phi \rightarrow Qx\Psi)$ .

Countable unions of countable sets are countable:  $(\forall x\neg Qy\Phi \& \neg Qx\exists y\Phi) \rightarrow \neg Qy\exists x\Phi$ .

The last scheme is a version of the axiom of choice, which is assumed in the meta-theory.

**THEOREM 18.**  $\mathcal{L}(Q_1)$  does not project (or quasi-project)  $\mathcal{L}(Q_0)$ .

**Proof.** This is a straightforward consequence of compactness (or completeness). Let  $\Phi$  be any sentence of  $\mathcal{L}(Q_1)$  that is satisfied by the natural numbers and let  $c$  be an individual constant that does not occur in  $\Phi$ . Consider the set

$$S = \{\Phi, c \neq 0, c \neq s0, c \neq ss0\}.$$

Every finite subset of  $S$  is satisfiable and so, by  $\aleph_0$ -compactness,  $S$  is satisfiable. However, a model of  $S$  is a model of  $\Phi$  which is not isomorphic to the natural numbers. Thus, there is no sentence in any  $\mathcal{L}(Q_1)[K]$  that is equivalent to the  $\mathcal{L}(Q_0)$  characterization of the natural numbers. ■

**THEOREM 19.** Quasi-weak second-order logic includes  $\mathcal{L}(Q_1)$ , but is not (quasi-) projectively equivalent to it.

**Proof.** Suppose that a formula  $\Phi$  of  $\mathcal{L}(Q_1)$  is equivalent to  $\Phi'$  in quasi-weak second-order logic, and let  $X$  be a monadic predicate variable that does not occur in  $\Phi'$ . Then  $Qx\Phi$  is equivalent to  $\neg\exists X(\forall x(Xx \equiv \Phi'))$  in quasi-weak second-order logic. A straightforward induction establishes that quasi-weak second-order logic includes  $\mathcal{L}(Q_1)$ . The second clause of the theorem is a corollary of the previous theorem and the fact that quasi-weak second-order logic includes weak second-order logic and  $\mathcal{L}(Q_0)$ . ■

It follows that  $\mathcal{L}(Q_1)$  does not enjoy the expressive resources of its cousin  $\mathcal{L}(Q_0)$ , but  $\mathcal{L}(Q_1)$  has a more attractive model theory. It is hard to assess the philosophical significance of  $\mathcal{L}(Q_1)$ . The language can express the notion of ‘uncountable’, of course, and thus it can express the disjunctive property ‘either finite or countably infinite’, but it cannot express ‘finite’ nor can it express ‘countably infinite’.

One important class of mathematical objects that can be characterized in  $\mathcal{L}(Q_1)$  is the aforementioned  $\aleph_1$ -like orderings—uncountable linear orderings in which every initial segment is countable. Simply conjoin the (first-order) axioms for a linear order with the following:

$$Qx(x = x) \& \forall y \neg Qx(x < y).$$

Of course there is no first-order sentence equivalent to this, since any first-order sentence with an infinite model has a countable model. Strictly speaking, the logic  $\mathcal{L}(Q_1)$  is not fully compact. To see this let  $\{c_\alpha \mid \alpha < \aleph_1\}$  be an uncountable set of constants, and consider the set

$$S = \{\neg Qx(x = x)\} \cup \{c_\alpha \neq c_\beta \mid \alpha < \beta\}.$$

Every finite subset of  $S$  is satisfiable, but  $S$  itself is not. Clearly, this generalizes to any  $\mathcal{L}(Q_\alpha)$ . However, the non-compactness of  $\mathcal{L}(Q_1)$  invokes countable models, which seem out of place in this context. If we eliminate finite and countable models from the model theory, then  $\mathcal{L}(Q_1)$  is compact. In other words, if for every finite subset  $S'$  of a set  $S$  of sentences, there is a model with an uncountable domain that satisfies  $S'$ , then the set  $S$  itself is satisfiable (in a model with an uncountable domain).

There is a downward Löwenheim–Skolem theorem of sorts for each logic  $\mathcal{L}(Q_\alpha)$ . Let  $M$  be a structure. Then there is a substructure  $M'$  of  $M$  whose domain has at most  $\aleph_\alpha$  elements (or the cardinality of the set of non-logical terminology, whichever is greater) such that  $M$  and  $M'$  satisfy the same sentences of  $\mathcal{L}(Q_\alpha)$ .

In studying  $\mathcal{L}(Q_\alpha)$  with  $\alpha > 1$ , we enter the realm of matters that are (or may be) independent of Zermelo–Fraenkel set theory. Chang [1965] showed that if the generalized continuum hypothesis holds, and if  $\aleph_\alpha$  is a regular cardinal, then  $\mathcal{L}(Q_{\alpha+1})$  is  $\aleph_\alpha$ -compact, and weakly complete. In fact, the same axioms work for any  $\mathcal{L}(Q_{\alpha+1})$  where  $\aleph_\alpha$  is regular. That is, if the generalized continuum hypothesis is true, the sentence  $\Phi$  is a logical truth of  $\mathcal{L}(Q_1)$  if and only if  $\Phi$  is a logical truth of  $\mathcal{L}(Q_{\alpha+1})$ . Jensen [1972] showed that if  $V = L$ , then for any ordinal  $\alpha$ ,  $\mathcal{L}(Q_{\alpha+1})$  is  $\aleph_\alpha$ -compact and weakly complete.

I close this section with a few variations on the theme of  $\mathcal{L}(Q_\alpha)$ . A logic with a *Chang quantifier* employs the same language as  $\mathcal{L}(Q_0)$  with the following clause for the new quantifier:

$M, s \models Qx\Phi$  if and only if the set  $\{s' \mid M, s' \models \Phi \text{ and } s' \text{ agrees with } s \text{ except possibly at } x\}$  has the same cardinality as the domain of  $M$ .

So  $Qx\Phi$  asserts that the extension of  $\Phi(x)$  is as large as the universe. Call the resulting logic *Chang logic*. Schmerl [1985] reports that if the generalized continuum hypothesis holds and we omit finite models from the model theory, then the Chang logic is compact and is weakly complete. In second-order set theory, the Chang quantifier might be used to indicate that the extension of  $\Phi$  is a proper class. Von Neumann once proposed an axiom that if a class is not the size of the universe, then it is a set. In this context, the scheme would be

$$\neg Qx\Phi \equiv \exists y \forall x (x \in y \equiv \Phi).$$

Consider augmenting a first-order language with a two-place *Ramsey* quantifier  $Q^2$  with the following clause in the model theory:

$m, s \models Q^2xy\Phi$  if and only if there is an uncountable subset  $d$  of the domain of  $M$  such that  $M, s' \models \Phi$  for every assignment  $s'$  which assigns members of  $d$  to  $x$  and  $y$ , and agrees with  $s$  at the other variables.

The logic is called  $\mathcal{L}(Q_1^2)$ . It turns out that if  $V = L$ , then  $\mathcal{L}(Q_1^2)$  is  $\aleph_0$ -compact, but it is consistent with Zermelo–Fraenkel set theory that  $\mathcal{L}(Q_1^2)$  is not  $\aleph_0$ -compact. In other words, it is independent of set theory whether this logic enjoys the compactness property. Extensions of these logics have been extensively studied.

The *Rescher quantifier*  $Q^R$  and the *Härtig quantifier*  $Q^I$  each binds two variables and has two formulas in its scope. In words,  $Q^Rxy[\Phi(x), \Psi(y)]$  if and only if the extension of  $\Phi(x)$  is not larger than the extension of  $\Psi(y)$ , and  $Q^Ixy[\Phi(x), \Psi(y)]$  if and only if the extension of  $\Phi(x)$  is the same size as the extension of  $\Psi(y)$ . Rescher logic includes Härtig logic, but not conversely. The natural numbers, under ‘less than’ can be characterized in Härtig logic (and thus in Rescher logic) with a sentence consisting of the axioms for a linear order with a first but no last element and the following:

$$\forall x \forall y (x = y \equiv Q^I uv [u < x, v < y]).$$

Thus, neither of these logics are compact or complete. Härtig logic includes  $\mathcal{L}(Q_0)$  but not conversely.

For details on the logics invoked in this section, see [Ebbinghaus, 1985]. For a more extensive treatment of  $\mathcal{L}(Q_1)$  see [Kaufmann, 1985], and for  $\mathcal{L}(Q_\alpha)$  see [Schmerl, 1985; Mundici, 1985]. There are extensive references in these sources. Cowles [1979] surveys the relations between some of the logics—and a number of others that I neglected to mention.



6 BRANCHING, OR NON-LINEAR QUANTIFIERS: THEFT OR TOIL?

Let  $\Phi(x_1, y_1, x_2, y_2)$  be a formula with only the indicated free variables, and consider the following two sentences:

$$\begin{aligned} &\forall x_1 \forall x_2 \exists y_1 \exists y_2 \Phi(x_1, y_1, x_2, y_2) \\ &\forall x_1 \exists y_1 \forall x_2 \exists y_2 \Phi(x_1, y_1, x_2, y_2). \end{aligned}$$

In words—and very roughly—the first of these says that if we are given an  $x_1$  and  $x_2$  then we can pick a  $y_1$  and a  $y_2$  such that  $\Phi$  holds. The ‘choice’ of the  $y$ ’s is made after we are given both of the  $x$ ’s. The second formula says that if we are given an  $x_1$  then we can pick a  $y_1$  and if we are then given an  $x_2$  we can pick a  $y_2$  such that  $\Phi$  holds. Here also the ‘choice’ of  $y_2$  is made after both  $x$ ’s are ‘given’, and so the ‘choice’ of  $y_2$  ‘depends’ on both  $x_1$  and  $x_2$ .

In general, each existentially quantified variable depends on all of the universally quantified variables that come before it. Some logicians and philosophers suggest that there is a need to introduce *independence* between some of the bound variables in a string of quantifiers. They have developed what are called ‘partially ordered quantifier prefixes’. For example, the two-dimensional formula,

$$\begin{aligned} &\forall x_1 \exists y_1 \\ &\quad \Phi(x_1, y_1, x_2, y_2) \\ &\forall x_2 \exists y_2 \end{aligned}$$

asserts that for every  $x_1$  there is a  $y_1$ , and for every  $x_2$  there is a  $y_2$  *chosen independently of*  $x_1$ , such that  $\Phi$  holds.

This four-place non-linear prefix,

$$\left. \begin{aligned} &\forall x_1 \exists y_1 \\ &\forall x_2 \exists y_2 \end{aligned} \right\}$$

is called the *Henkin quantifier*, and for the sake of typography, we will sometimes abbreviate it  $Hx_1y_1x_2y_2$ . The language  $\mathcal{L}(H)[K]$  is obtained from first-order  $\mathcal{L}1[K]$  = by adding the Henkin quantifier. The relevant formation rule is that if  $\Phi$  is a formula and  $x_1, y_1, x_2, y_2$  are four distinct variables, then  $Hx_1y_1x_2y_2\Phi$  is a formula.<sup>8</sup>

The literature contains several (more or less) equivalent ways to generalize this notion. I will give one, in terms of what are called ‘dependency relations’. A *dependency prefix* is a triple  $Q = (A_Q, E_Q, D_Q)$ , structured as follows:  $A_Q$  is the set of *universal variables* of  $Q$ ;  $E_Q$  is the set of *existential*

<sup>8</sup>Strictly speaking, we should distinguish quantifiers from quantifier prefixes. For convenience, however, I do not enforce the distinction here, relying on context when necessary.

variables of  $Q$ ; and  $D_Q$  is a *dependency relation* between  $A_Q$  and  $E_Q$ . If  $(x, y)$  is in  $D_Q$ , then we say that the existential variable  $y$  depends on the universal variable  $x$  in  $Q$ . See [Krynicky and Mostowski, 1995, Section 1.5].

In these terms, the aforementioned Henkin quantifier is the triple  $(A_H, E_H, D_H)$  where  $A_H$  is  $\{x_1, x_2\}$ ,  $E_H$  is  $\{y_1, y_2\}$ , and  $D_H$  contains the two pairs  $(x_1, y_1)$  and  $(x_2, y_2)$ . Ordinary, linear quantifier prefixes can also be cast in this form. Both of the formulas set off at the top of this section have the same sets of universal and existential variables as  $H$ . The dependency relation of the prefix,  $\forall x_1 \forall x_2 \exists y_1 \exists y_2$  of the first formula is all of  $A_H \times E_H : \{(x_1, y_1), (x_1, y_2), (x_2, y_1), (x_2, y_2)\}$ . The dependency relation of the prefix,  $\forall x_1 \exists y_1 \forall x_2 \exists y_2$  of the second formula is  $\{(x_1, y_1), (x_1, y_2), (x_2, y_2)\}$ .

A dependency prefix  $Q$  is called *linear* if there is a linear ordering  $R$  on the variables of  $Q$  such that for each  $x$  in  $A_Q$  and each  $y$  in  $E_Q$ ,  $(x, y)$  is in  $D_Q$  if and only if  $Rxy$ . Linear prefixes are equivalent to first-order prefixes.

Let  $\mathfrak{S}$  be a set of dependency prefixes. The language  $\mathcal{L}(\mathfrak{S})[K]$  is obtained from the first-order  $\mathcal{L}1[K]$  by adding formulas with dependency prefixes in  $\mathfrak{S}$ . The relevant formation rule is that if  $\Phi$  is a formula and  $Q$  is in  $\mathfrak{S}$ , then  $Q\Phi$  is a formula. The language  $\mathcal{L}^*[K]$  is the language  $\mathcal{L}(\mathfrak{S})[K]$  in which  $\mathfrak{S}$  is the set of all dependency prefixes.

So much for the grammar. Now, what is the model theory? In other words, what do these formulas  $Q\Phi$  mean? We use functions in the meta-language to express the relevant dependency and independency relations among the variables, along the lines of Skolem functions for first-order languages. Here, the relevant functions are denoted by new non-logical terminology in the object language. Suppose that  $Q$  is a dependency prefix and that  $\Phi$  is a formula. Define the *Skolemization* of  $Q\Phi$ , written  $\mathbf{sk}Q\Phi$ , as follows: let  $y$  be an existential variable in  $E_Q$  and let  $x_1, \dots, x_i$  be the universal variables on which  $y$  depends. Pick a unique  $i$ -place non-logical function letter  $f_y$ , which does not occur in  $\Phi$ , and replace each occurrence of  $y$  with  $f_y x_1 \dots x_i$ . Bind the result with universal quantifiers over the variables in  $A_Q$ . To take an example, if  $H$  is the Henkin quantifier, then  $\mathbf{sk}(Hxyzw\Phi(x, y, z, w))$  is:

$$\forall x \forall z \Phi(x, f_y x, z, f_w z).$$

The functions express the requisite dependence of the existential variables. Notice that if the prefix  $Q$  is linear, then  $\mathbf{sk}(Q\Phi)$  is the usual result of invoking Skolem functions to interpret existential variables.

The relevant clause in the semantics is:

Let  $M$  be a structure and  $s$  an assignment to the variables. Then  $M, s \models Q\Phi$  if there are assignments to the new function letters (as appropriate functions on the domain of  $M$ ) such that the resulting structure satisfies  $\mathbf{sk}(Q\Phi)$  under  $s$ .

Suppose that  $f_1$  and  $f_2$  are the only new function letters in  $\mathbf{sk}(Q\Phi)$ . Then, if we can invoke second-order quantifiers,  $M, s \models Q\Phi$  if and only if  $M, s \models \exists f_1 \exists f_2 \mathbf{sk}(Q\Phi)$ .

There is a potential complication for readers with constructivist tendencies. Suppose that  $A_Q$  is  $\{x\}$ ,  $E_Q$  is  $\{y\}$ , and  $D_Q$  is  $\{(x, y)\}$ . Then  $Q$  is a linear quantifier prefix and one would expect that  $Q\Phi(x, y)$  to be equivalent to  $\forall x \exists y \Phi(x, y)$ . However, according to the semantics  $Q\Phi$  is equivalent to  $\exists f \forall x \Phi(x, fx)$ . The inference from  $\forall x \exists y \Phi(x, y)$  to  $\exists f \forall x \Phi(x, fx)$  is a version of the axiom of choice (see [Shapiro, 1991, Chapter 4]). Thus, the plausibility of the given model theory for  $\mathcal{L}(\mathfrak{S})$  presupposes choice. Krynicki and Mostowski [1995, Section 2] provide a straightforward, but tedious, way to avoid this presupposition. An  $n + 1$ -place relation  $R$  is defined to be a *dependency* relation if for each  $x_1, \dots, x_n$  in the domain, there is at least one  $y$  such that  $Rx_1 \dots x_n y$  holds. In what follows, however, we follow Krynicki and Mostowski's practice of assuming the axiom of choice, and using functions instead of dependency relations.

It is straightforward to verify that if a dependency prefix has fewer than 4 variables then it is linear and equivalent to a string of first-order quantifiers. Moreover, if a dependency prefix has exactly four variables then either it is linear or it is a Henkin quantifier. Thus, the simplest non-linear quantifier is the Henkin quantifier. Krynicki and Mostowski [1995, Section 3.3] show that if a set  $\mathfrak{S}$  has at least one non-linear dependency prefix, then  $\mathcal{L}(\mathfrak{S})[K]$  includes  $\mathcal{L}(H)[K]$  in the sense of Section 4 above: every formula in  $\mathcal{L}(H)[K]$  is equivalent to one in  $\mathcal{L}(\mathfrak{S})[K]$ .

Krynicki and Mostowski [1995, Section 3.9] also show that if  $Q$  is any dependency prefix, then  $Q$  can be defined in terms of a prefix  $Q'$  in the following form:

$$\left. \begin{array}{l} \forall x_1 \dots \forall x_n \exists y \\ \forall z_1 \dots \forall z_n \exists w \end{array} \right\}$$

There are  $2n$  variables in  $A_{Q'}$  and 2 variables in  $E_{Q'}$ . The variable  $y$  depends on the  $x$ 's and the variable  $w$  depends on the  $z$ 's. In structures with a pair function, the latter quantifier can be reduced further to the Henkin quantifier  $H$  (see Section 3.1 above). In other words, in a structure with a pair function, any formula using any dependency prefix is equivalent to a formula that just uses the Henkin quantifier  $H$ .

A pair function can be added to any structure whose domain is infinite. This allows a significant reduction of dependency prefixes. Let  $\Phi$  be any formula using quantifier prefixes, such that  $\Phi$  has only infinite models. Then there is a set  $K'$  of non-logical terminology—including a pair function—and a sentence  $\Phi'$  in  $\mathcal{L}(H)[K']$ , such that a structure  $M$  satisfies  $\Phi$  under a given assignment if and only if there is an expansion of  $M$  (to the set  $K'$ ) which satisfies  $\Phi'$  under the same assignment. Recall that  $\mathcal{L}^*$  is the language containing every dependency prefix. We see that  $\mathcal{L}(H)[K]$  quasi-projects

the full  $\mathcal{L}^*[K]$  in the sense of Section 4 above.

Enough of these definitions and internal comparisons. What can we do with these new quantifiers, and how tractable is the semantics? It turns out that  $\mathcal{L}(H)$  and thus  $\mathcal{L}^*$  represents a significant foray into the expressive resources of second-order logic.

Consider the following sentence in  $\mathcal{L}(H)$  (which has no non-logical terminology):

$$\exists t \begin{array}{l} \forall x \exists y \\ \forall x' \exists y' \end{array} ((x = x' \equiv y = y') \& y \neq t),$$

or in one line

$$\exists t H(xy x' y') ((x = x' \equiv y = y') \& y \neq t).$$

According to the given semantics, this holds in a given domain if and only if there is an element  $t$  in the domain and two functions  $f$  and  $f'$  such that for all  $x$  and  $x'$ ,  $x = x'$  if and only if  $fx = f'x'$  and  $fx \neq t$ . This entails that  $f = f'$ , that  $f$  is one-to-one, and that there is an element  $t$  that is not in the range of  $f$ . Thus, the given formula holds in a given structure if and only if its domain is infinite. Thus,  $\mathcal{L}(H)[K]$  does not include the first-order  $\mathcal{L}1[K]$ . It follows that no logic that includes a non-linear quantifier prefix is compact.

Let  $\Phi(x)$  be any formula with  $x$  free. Then the formula

$$\exists t (\Phi(t) \& H(xy x' y') ((x = x' \equiv y = y') \& (\Phi(x) \rightarrow \Phi(y)) \& y \neq t))$$

is satisfied in a structure if and only if the extension of  $\Phi$  is infinite. That is, the above formula is equivalent to  $Qx\Phi(x)$  in the logic  $\mathcal{L}(Q_0)$ . It follows that  $\mathcal{L}(H)[K]$  includes  $\mathcal{L}(Q_0)[K]$ . Thus, there is a categorical characterization of the natural numbers in  $\mathcal{L}(H)$ , and so  $\mathcal{L}(H)$  is not weakly complete.

Recall that the Rescher quantifier  $Q^R$  binds one variable in each of two formulas:  $Q^R xy[\Phi(x), \Psi(y)]$  'says' that the extension of  $\Phi(x)$  is not larger than the extension of  $\Psi(y)$ . This holds if there is a one-to-one function from the extension of  $\Phi$  to the extension of  $\Psi$ . Thus, the Rescher quantifier can be captured with a sentence using the Henkin quantifier:

$$H(xy x' y') ((x = x' \equiv y = y') \& (\Phi(x) \rightarrow \Psi(y))).$$

It follows that the Hartig quantifier and the Chang quantifier can also be characterized in terms of the Henkin quantifier.

The expressive power of the languages  $\mathcal{L}(H)[K]$  is richer than most of the languages considered above. Krynicki and Mostowski [1995, Section 8.4] point out that the notion of well-ordering can be characterized in  $\mathcal{L}(H)$ , using only the non-logical symbol  $<$ . The notion of dense continuous order can be characterized, as can the ordinal structure of  $\aleph_n$  for each natural number  $n$ .

It follows, of course, that there is no simple downward Löwenheim–Skolem theorem for  $\mathcal{L}(H)$  or for the full  $\mathcal{L}^*$ . However, if a sentence  $\Phi$  in  $\mathcal{L}^*[K]$  has no non-logical terminology, then if  $\Phi$  has an infinite model then it has a countable model. It follows that neither  $\mathcal{L}(H)[K]$  nor  $\mathcal{L}^*[K]$  includes any  $\mathcal{L}(Q_\alpha)$  for any  $\alpha > 0$ .

What are the exact bounds to the expressive resources of  $\mathcal{L}(H)$  and  $\mathcal{L}^*$ ? Let  $\Phi$  be first-order and let  $Hxyx'y'\Phi$  be the result of prefixing  $\Phi$  with a Henkin quantifier. We saw above that  $Hxyx'y'\Phi$  is equivalent to a formula in the form  $\exists f\exists f'\forall x\forall x'\Phi'$ , where  $\Phi'$  is first-order. That is,  $Hxyx'y'\Phi$  is equivalent to a  $\Sigma_1^1$ -formula. Krynicki and Mostowski [1995, Section 4] report a converse of sorts:

**THEOREM 20** (Enderton and Walkoe). *There is an effective procedure for assigning to each  $\Sigma_1^1$  formula  $\Phi$  a dependency prefix  $Q$  and a quantifier free formula  $\Psi$  such that  $\Phi$  is equivalent to  $Q\Psi$ .*

It follows from Theorem 20 that every Boolean combination of  $\Sigma_1^1$  formulas is equivalent to a formula in  $\mathcal{L}^*[K]$ . In particular, since any  $\Pi_1^1$  formula is equivalent to the negation of a  $\Sigma_1^1$  formula, it follows that every  $\Pi_1^1$  formula is equivalent to a formula in  $\mathcal{L}^*[K]$ . Thus,  $\mathcal{L}^*[K]$  (and  $\mathcal{L}(H)[K]$  on infinite domains) has all the expressive power of free-variable second-order logic, and then some. Moreover,  $\mathcal{L}^*[K]$  does not have the major shortcoming of free-variable second-order languages, since  $\mathcal{L}^*[K]$  is closed under contradictory opposition: the negation of a  $\mathcal{L}^*[K]$  formula is an  $\mathcal{L}^*[K]$  formula.

Krynicki and Mostowski report that the expressive resources of formulas with quantifier dependencies do not go much further than what is expressed in Theorem 20:

**THEOREM 21.** *For any formula  $\Phi$  in any  $\mathcal{L}^*[K]$ , we can effectively find a  $\Sigma_2^1$ -formula and a  $\Pi_2^1$  formula both equivalent to  $\Phi$ . Thus, any formula in  $\mathcal{L}^*[K]$  is equivalent to a  $\Delta_2^1$  formula.*

They also point out that there are  $\Delta_2^1$  formulas which are not equivalent to any formula in any  $\mathcal{L}^*[K]$ . For example, there is a  $\Delta_2^1$ -sentence  $T$  that gives a ‘truth definition’ for arithmetic in the  $\mathcal{L}^*$  language of arithmetic. That is,  $T$  characterizes structures  $\langle M, c \rangle$  such that  $M$  is a standard model of arithmetic and  $c$  is the code of an  $\mathcal{L}^*$  sentence true in  $M$ . It follows from Tarski’s theorem that  $T$  is not equivalent to any sentence in any  $\mathcal{L}^*[K]$ .

It follows that Theorem 21 does not give the ‘best possible’ characterization of the expressive power of  $\mathcal{L}^*$ . According to Krynicki and Mostowski [1995], it is an open question whether every formula of  $\mathcal{L}^*[K]$  is equivalent to a Boolean combination of  $\Sigma_1^1$  formulas.

It is hard to assess the philosophical significance of languages with dependency prefixes. As we saw, even  $\mathcal{L}(H)[K]$  overcomes the bulk of the shortcomings with first-order logic, such as those elaborated in [Shapiro, 1991, Chapters 4–5]. Yet  $\mathcal{L}(H)[K]$  and even  $\mathcal{L}^*[K]$  only invoke *first-order*

variables, and the ordinary existential and universal quantifiers.

This may be too good to be significant. Recall that the official model-theoretic semantics for these languages invokes functions—or relations if choice is to be avoided. The satisfiability of a formula that starts with a Henkin quantifier is understood in terms of the *existence* of certain functions (or relations). Functions and relations, of course, are higher-type items. Thus, it is no surprise that the expressive resources of the languages hovers somewhere around that of  $\Sigma_1^1$ - and  $\Pi_1^1$ -formulas. A critic of  $\mathcal{L}^*$  might claim that an ‘ontological commitment’ to functions (or relations) is hidden in the model-theoretic semantics. He might argue that there is no way to understand the requisite dependencies except via functions or relations. If the critic is successful, then we would see that the very notion of ‘dependency’ invokes higher-order items, in which case there is no special significance to the expressive resources of  $\mathcal{L}^*$ .

To counter this argument, an advocate of dependency prefixes might try to give a semantics for the languages that does not explicitly invoke functions or relations. One straightforward—and potentially question-begging—way to do so would be to simply *use* quantifier dependencies in the meta-language. One clause might be the following:

$$M, s \models H(xyx'y')\Phi \text{ if and only if in the domain of } M, H(mnm'n') \text{ such that } M, s' \models \Phi \text{ for every assignment } s \text{ that agrees with } s \text{ except possibly at } x, y, x', \text{ and } y' \text{ and } s(x) = m, s(y) = n, s(x') = m', \text{ and } s(y') = n'.$$

This would make the clause for the Henkin quantifier exactly analogous to the clauses for the first-order connectives and quantifiers. We *use* the terminology in the meta-language in giving the model-theoretic semantics.

Is this a vicious circle? The potentially question-begging move is plausible if, but only if, the advocate for dependency prefixes can successfully argue that we already understand these prefixes. Then the situation with dependency prefixes would be no different than the situation with the other logical terminology.

The dialectic here is reminiscent of the clash between Resnik and Boolos over plural quantification (see Section 3.1 above). Boolos claims that we have a decent pre-theoretic grasp of plural quantifiers and uses this construction to interpret monadic existential second-order variables. Resnik claims that whatever understanding we have of the plural construction is mediated by set theory, and thus the plural construction hides the ‘ontological commitment’ to sets. In reply Boolos can cite the prevalence of the plural construction in natural language, pointing out that common folk who are ignorant of set theory are clearly competent in the use of plurals. What of the present case, concerning non-linear dependency prefixes? Are there any natural language constructions which are best interpreted using,

say, Henkin quantifiers? Hintikka [1976] argues that there are, and gives examples like the following:

Some relative of each villager and some relative of each towns-  
person hate each other.

Every writer likes a book of his almost as much as every critic  
dislikes some book he has reviewed.

Readers interested in this issue can also consult [Gabbay and Moravcsik, 1974; Barwise, 1979]. For more on the technical side of quantifier prefixes, the aforementioned [Krynicky and Mostowski, 1995] is a comprehensive and readable treatment. See also [Mundici, 1985, Section 1].

## 7 EXTRA LONG FORMULAS

Let us put philosophical worries aside, and assume that mathematicians are able to refer to and discuss some infinite mathematical sets and structures. Then they can also refer to and discuss infinitely long sentences and infinitely long deductions, themselves construed as abstract objects. In short, infinitary languages are respectable objects of mathematical study. Our question here is whether they are relevant to philosophical logic. Some philosophers reject infinitely long formulas, out of hand, as serious candidates for foundational research. For good reason. One cannot do much communicating if it takes an infinite amount of time and space to write, or speak, or comprehend, a single sentence. Surely, natural languages are not infinitary and so we should not need infinitary languages to model them.

This eminently reasonable observation may not disqualify infinitary languages from every role in foundational studies. Perhaps one can argue that infinitary languages capture *something* important about the logical structure of natural languages. One suggestion is to regard the natural language of mathematics as an informal meta-language for an infinitary object language, whose models are the various structures under study. It may not be too much of a distortion to view the proposal in [Zermelo, 1931] that way.

Less exotically, someone might propose that infinitary formulas come close to the logical forms of propositions, or one might suggest that infinitary languages capture important relations and features underlying mathematics as practiced. For example, first-order arithmetic consists of a finite number of axioms together with each instance of the induction scheme. It is reasonable to interpret such theories as the infinitary *conjunction* of their axioms, or to put it differently, there is not much difference between considering an infinite set of axioms and considering an infinitary conjunction of them. Infinitary *disjunctions* are, of course, another story. They enter via omitting types.

Infinitary languages have been invoked by philosophers for various purposes, often to reduce ontological or other commitments. It is common, for example, for deflationists about truth to regard an assertion like ‘Everything my mother says is true’ as an infinite conjunction of sentences of the form: if my mother says that  $\Phi$  then  $\Phi$ .

Infinitary logic has probably received more attention from mathematical logicians than any of the intermediate systems presented above. Such systems seem to do well in the tradeoff between expressive ability and tractable model theory—a major focus of this chapter. Without further ado, we take a passing glance at infinitary languages.

If  $K$  is a set of non-logical terminology, and  $\kappa \geq \lambda$  are two cardinal numbers, then  $\mathcal{L}_{\kappa\lambda}[K]$  is an infinitary language based on  $K$ . For convenience, we will omit the ‘ $K$ ’ in most contexts. The formation rules of  $\mathcal{L}_{\kappa\lambda}$  are those of the first-order  $\mathcal{L}1[K]=$ , augmented with the following clauses:

If  $\Gamma$  is a set of well-formed formulas whose cardinality is less than  $\kappa$ , then  $\bigwedge \Gamma$  is a well-formed formula.

If  $A$  is a set of variables whose cardinality is less than  $\lambda$ , and  $\Phi$  is a well-formed formula, then  $\forall A\Phi$  is a well-formed formula. In  $\forall A\Phi$ , every variable in  $A$  is bound.

Two technical caveats: Notice that if  $\kappa$  is not regular, then there are, in effect, conjunctions of size  $\kappa$  in  $\mathcal{L}_{\kappa\lambda}$ . Similarly, if  $\lambda$  is not regular, there are formulas with  $\lambda$ -many bound variables. For this reason, some authors require  $\kappa$  and  $\lambda$  to be regular cardinals. Also, for convenience, we stipulate that the formulas in the set  $\Gamma$  of the first clause contain fewer than  $\lambda$  free variables total. Otherwise, there will be formulas of  $\mathcal{L}_{\kappa\lambda}$  that cannot be turned into sentences by binding all of their free variables.

Infinitary disjunctions can be defined in a straightforward manner: if  $\Gamma$  is a set of formulas, let  $\neg\Gamma$  be  $\{\neg\Phi \mid \Phi \in \Gamma\}$ . Then define  $\bigvee \Gamma$  to be  $\neg\bigwedge \neg\Gamma$ . Infinitary existential quantification is similar: if  $A$  is a set of variables, then define  $\exists A\Phi$  to be  $\neg\forall A\neg\Phi$ .

If the cardinality of the set  $K$  of non-logical terminology is not larger than  $\kappa$ , then there are (only)  $2^\kappa$  well-formed formulas in  $\mathcal{L}_{\kappa\lambda}$ . For readers who do not think that this is enough formulas, there are some *really* big languages. If the restriction on the size of the set  $\Gamma$  in the above clauses is dropped, the language is called  $\mathcal{L}_{\infty\lambda}$ . That is to say, if  $\Gamma$  is *any* set of formulas in  $\mathcal{L}_{\infty\lambda}$ , then  $\bigwedge \Gamma$  is a formula. Similarly, if the restriction on the cardinality of the set  $A$  of variables is also dropped, the language is called  $\mathcal{L}_{\infty\infty}$ . Notice that  $\mathcal{L}_{\infty\lambda}$  and  $\mathcal{L}_{\infty\infty}$  each have a proper class of formulas. The latter has a proper class of variables!

At the other end of the scale, notice that  $\mathcal{L}_{\omega\omega}$  is just the first-order  $\mathcal{L}1[K]=$ . The ‘smallest’ infinitary language is  $\mathcal{L}_{\omega_1\omega}$ , which allows countable conjunctions but only finitary quantifiers.



The semantics for all of these infinitary languages is a straightforward extension of the semantics of first-order languages. The new clauses are:

$$\begin{aligned} M, s \models \bigwedge \Gamma & \text{ if } M, s \models \Phi \text{ for every } \Phi \in \Gamma. \\ M, s \models \bigvee A \Phi & \text{ if } M, s' \models \Phi \text{ for every assignment } s' \text{ that agrees with} \\ & s \text{ on the variables not in } A. \end{aligned}$$

Suppose that  $K$  contains at least one binary relation letter. A straightforward transfinite induction establishes that if  $\alpha$  is any ordinal whose cardinality is less than  $\kappa$ , then there is a sentence  $\Phi_\alpha$  of  $\mathcal{L}\kappa\omega[K]$ , such that a structure  $M$  satisfies  $\Phi_\alpha$  iff  $M$  is isomorphic to  $\alpha$ . Thus, there are uncountably many different structures that can be characterized up to isomorphism in  $\mathcal{L}\omega_1\omega$ . On the other hand, if  $K$  is countable, then any finitary language based in  $K$  has only countably many sentences, and so only countably many structures can be characterized up to isomorphism (with a single sentence). Thus, second-order logic does not include  $\mathcal{L}\omega_1\omega$ . Strictly speaking, infinitary logics are not ‘intermediate’ between first-order and second-order.

It might be added that no infinitary language  $\mathcal{L}\kappa\lambda$  includes second-order logic. For example, the notions of compact space and complete linear order can be characterized in a second-order language, but not in any  $\mathcal{L}\kappa\lambda$  (see [Dickmann, 1985, p. 323]). The reason is that there is no bound on the cardinality of the relations in the range of second-order variables.

The expressive power of infinitary languages is often a matter of ‘brute force’. One constructs a formula that simply ‘says’ what is required to characterize a given notion or structure. For example, the extension of a formula is finite if and only if the disjunction of the following formulas holds:

$$\begin{aligned} \exists x \forall y (\Phi(y) \rightarrow x = y), \exists x_1 \exists x_2 \forall y (\Phi(y) \rightarrow (x_1 = y \vee x_2 = y)), \\ \exists x_1 \exists x_2 \exists x_3 \forall y (\Phi(y) \rightarrow (x_1 = y \vee x_2 = y \vee x_3 = y)), \dots \end{aligned}$$

Similarly, let  $\Psi(x)$  be the infinitary disjunction of  $x = 0, x = s0, x = ss0, \dots$ . Any model of the axiom for the successor function and  $\forall x \Psi(x)$  is isomorphic to the natural numbers. Thus, the natural numbers can be characterized, up to isomorphism, in  $\mathcal{L}\omega_1\omega$ . The infinitary  $\forall x \Psi(x)$  guarantees that the numerals exhaust the domain, and so there are no ‘non-standard’ numbers. To take one more example, let  $\chi(x)$  be the disjunction of  $x < 1, x < 1 + 1, x < 1 + 1 + 1, \dots$ . Then  $\forall x \chi(x)$  is satisfied by an ordered field  $F$  if and only if  $F$  is Archimedean.

Let  $\Phi(x, y)$  be any formula with  $x$  and  $y$  free. Then ‘ $w$  is an ancestor of  $x$  under  $\Phi$ ’ is characterized as the disjunction of

$$\begin{aligned} w = x, \Phi(z, w), \exists x (\Phi(z, x) \& \Phi(x, w)), \\ \exists x_1 \exists x_2 (\Phi(z, x_1) \& \Phi(x_1, x_2) \& \Phi(x_2, w)), \dots \end{aligned}$$

This, and similar reasoning, shows that the smallest infinitary language  $\mathcal{L}\omega_1\omega$  includes the logics of Section 4 above—the ones that presuppose the

notion of finitude. That is, if  $\Phi$  is any sentence of weak second-order logic,  $\mathcal{L}(Q_0)$ , ancestral logic, or  $\omega$ -logic, then there is a sentence  $\Phi'$  of  $\mathcal{L}\omega_1\omega$  such that for any model  $M$ ,  $M \models \Phi$  iff  $M \models \Phi'$ . See [Cowles, 1979] for more details on these results.

There is an analogue of the downward Löwenheim–Skolem theorem: if  $\kappa$  is uncountable and  $\Phi$  is any sentence of  $\mathcal{L}\kappa\omega$ , then if  $\Phi$  has a model at all, it has a model whose cardinality is less than  $\kappa$ . It follows that the Löwenheim number of  $\mathcal{L}\kappa\omega$  is at most  $\kappa$ . The ordinary Löwenheim–Skolem theorem holds in  $\mathcal{L}\omega_1\omega$ . If a sentence has a model at all, then it has a countable model.

One consequence of this Löwenheim–Skolem result is that there is no characterization of the real numbers in any  $\mathcal{L}\kappa\omega$  unless  $\kappa$  is larger than the continuum. However, there is a characterization of the real numbers, up to isomorphism, in  $\mathcal{L}\omega_1\omega_1$ , as follows: let  $A$  be the countably infinite set of distinct variables,  $x_1, x_2, \dots$ . If  $v$  is any variable, then let  $A < v$  be the conjunction of the set  $\{x_i < v \mid x_i \in A\}$ . Let  $AR^\infty$  be the conjunction of the axioms for an ordered field and the following version of the completeness principle:

$$\forall A(\exists y A < y \rightarrow \exists z(A < z \& \forall y(A < y \rightarrow z \leq y))).$$

This formula asserts, via brute force, that for any countable (non-empty) set of elements, if that set is bounded, then it has a least upper bound. Thus, the  $\mathcal{L}\omega_1\omega_1$ -sentence  $AR^\infty$  is a categorical characterization of the real numbers.

Let  $A$  be a countable set of variables, as above, and let  $\Phi$  be the conjunction of  $x_2 < x_1, x_3 < x_2, x_4 < x_3, \dots$ . Then, assuming the axiom of choice, the relation  $<$  is well-founded if  $\forall A \neg \Phi$ . This last is a sentence of  $\mathcal{L}\omega_1\omega_1$ . Thus, if we assume the axiom of choice in the meta-theory, then the notion well-ordering can be characterized by a sentence of  $\mathcal{L}\omega_1\omega_1$ . Nadel [1985] reports that there is no sentence of  $\mathcal{L}\infty\omega$  that characterizes the class of well-orderings. However, in  $\mathcal{L}\kappa\omega$ , one can characterize the notion of ‘well-order of size smaller than  $\kappa$ ’. To move up one level,  $\mathcal{L}\omega_2\omega_1$  includes the system called quasi-weak second-order logic in Section 5 above.

Compactness fails, even in  $\mathcal{L}\omega_1\omega$ . Let  $\Gamma$  be an infinite set of (independent) atomic sentences. For example,  $\Gamma$  might consist of  $c \neq 0, c \neq s0, c \neq ss0, c \neq sss0, \dots$ . Then the set  $\Gamma \cup \neg \bigwedge \Gamma$  is clearly unsatisfiable, and yet every finite subset of  $\Gamma \cup \neg \bigwedge \Gamma$  is satisfiable. In fact, every *proper* subset of  $\Gamma \cup \neg \bigwedge \Gamma$  is satisfiable.

For another example, for each  $\alpha < \omega_1$ , let  $c_\alpha$  be an individual constant, and let  $f$  be a unary function symbol. Let  $\Theta$  be the set  $\{c_\alpha \neq c_\beta \mid \alpha < \beta < \omega_1\}$ . Let  $\Psi$  be the disjunction of  $\{fx = c_\alpha \mid \alpha < \omega\}$  and let  $\Phi$  be

$$\forall x \forall y (fx = fy \rightarrow x = y) \& \forall x \Psi.$$

That is,  $\Phi$  is a statement that  $f$  is one-to-one and the range of  $f$  is  $\{c_\alpha \mid \alpha < \omega\}$ . Then  $\Theta$  entails that the domain is uncountable while  $\Phi$  entails that the domain is countable. Thus  $\Theta \cup \{\Phi\}$  has no models. Yet every finite subset of  $\Theta \cup \{\Phi\}$  has a model. Indeed, every *countable* subset of  $\Theta \cup \{\Phi\}$  has a model.

I hope it will not further offend the gentle reader's sensibilities to speak of infinitely long deductions. Hilbert [1925] wrote:

... the literature of mathematics is glutted with ... absurdities which have had their source in the infinite. For example, we find writers insisting, as though it were a restrictive condition, that in rigorous mathematics only a finite number of deductions are admissible in a proof—as if someone had succeeded in making an infinite number of them.

Nevertheless, some of the above motivation for infinitary logic might support a theory of infinitary deduction. Moreover, some of the semantic properties of infinitary languages are revealed via infinitary deduction.

There is a pretty straightforward infinitary deductive system for  $\mathcal{L}\kappa\omega$ . Augment a standard deductive system for  $\mathcal{L}1[K]=$  with the following rules:

Infer  $\wedge \Gamma \rightarrow \Psi$ , if  $\Psi \in \Gamma$ .  
From  $\Phi \rightarrow \psi$ , for all  $\psi$  in  $\Gamma$ , infer  $\Phi \rightarrow \wedge \Gamma$ .

We require the 'length' of a deduction in  $\mathcal{L}\kappa\omega$  to be 'shorter' than  $\kappa$ . If we can be permitted to speak of 'natural deduction' for infinitary languages, the first rule of inference can be replaced by a rule of  $\wedge$ -elimination: if  $\Psi \in \Gamma$ , then infer  $\Psi$  from  $\wedge \Gamma$ , resting on whatever assumptions  $\wedge \Gamma$  rests upon. The second rule can be replaced with a rule of  $\wedge$ -introduction: from  $\Psi$ , for all  $\Psi$  in  $\Gamma$ , infer  $\wedge \Gamma$ , resting on all assumptions that the members of  $\Gamma$  rest upon.

The smallest infinitary logic  $\mathcal{L}\omega_1\omega$  enjoys a certain completeness property: if  $\Phi$  is a logical truth in  $\mathcal{L}\omega_1\omega$ , then  $\Phi$  can be 'deduced' in the above system. This is a 'weak completeness' of sorts. We get a bit more as a corollary: if  $\Gamma$  is a *countable* set of formulas and  $\Phi$  a single formula, then  $\Gamma \vDash \Phi$  in  $\mathcal{L}\omega_1\omega$  if  $\Phi$  can be 'deduced' from  $\Gamma$  in the expanded deductive system.

However, there is no full completeness. Recall the set  $\Theta \cup \{\Phi\}$ , defined just above, which has no models. Thus,  $\Theta \cup \{\Phi\} \vDash c_0 \neq c_0$ . But a 'deduction' from  $\Theta \cup \{\Phi\}$  can involve only countably many members of  $\Theta \cup \{\Phi\}$ , and any such collection is satisfiable and thus consistent. So  $c_0 \neq c_0$  cannot be deduced from  $\Theta \cup \{\Phi\}$ .

The above completeness result indicates that 'logical truth' in  $\mathcal{L}\omega_1\omega$  is 'absolute' in the background meta-theory. That is, if a formula is a logical truth in any transitive model of ZFC, then it is a logical truth in any other

transitive model of ZFC. However, when we consider larger languages we go beyond what can be discerned in the background meta-theory. There are sentences in  $\mathcal{L}_{\infty\omega}$  that are logical truths in some models of the background meta-theory, but are not logical truths in others. In this respect,  $\mathcal{L}_{\infty\omega}$  is like second-order logic. It follows that there is no ‘absolute’ notion of ‘provability’ that will yield a version of weak completeness for even  $\mathcal{L}_{\omega_2\omega}$ .

Logicians have studied infinitary languages even more exotic than  $\mathcal{L}_{\infty\omega}$ . Some have infinite alternations of quantifiers, e.g.  $\forall x_1 \exists y_1 \forall x_2 \exists y_2 \dots \Phi$ . From the opposite perspective, the objections to infinitary languages might be attenuated if we focus attention on a subclass of  $\mathcal{L}_{\omega_1\omega}$ . Logicians have studied certain countable fragments of  $\mathcal{L}_{\omega_1\omega}$ . The idea of an infinitary conjunction of a recursive (or otherwise definable) set of sentences might be less offensive to a sensitive philosophical temperament. Assume that we have cast the syntax for  $\mathcal{L}_{\infty\omega}$  in set theory, so that the formulas are defined to be sets. A transitive set  $B$  of sets is called *admissible* if it satisfies a certain theory, called ‘Kripke–Platek’ set theory, which is weaker than full Zermelo–Fraenkel set theory. A fragment  $\mathcal{L}$  of  $\mathcal{L}_{\infty\omega}$  is *admissible* if there is an admissible set  $B$  such that  $\mathcal{L}$  is  $\mathcal{L}_{\infty\omega} \cap B$ . There is an extensive literature on admissible fragments of  $\mathcal{L}_{\infty\omega}$  (see [Nadel, 1985, Section 5]).

The reader interested in infinitary languages will do well to consult the essays in [Barwise and Feferman, 1985], especially [Dickmann, 1985; Kolaitis, 1985; Nadel, 1985] and the wealth of references provided there.

## 8 SOMETHING COMPLETELY DIFFERENT: SUBSTITUTIONAL QUANTIFICATION

Some philosophers, unhappy with ‘satisfaction’ as the central component of model-theoretic semantics, propose to replace the ‘satisfaction’ of formulas with the ‘truth’ of sentences. The crucial clause in *substitutional semantics* is:

Let  $\Phi(x)$  be a formula whose only free variable is  $x$ . Then  $\forall x\Phi(x)$  is *true substitutionally* in an interpretation if for every term  $t$  of the language,  $\Phi(t)$  is true substitutionally in that interpretation;  $\exists x\Phi(x)$  is true substitutionally in an interpretation if there is a term  $t$  of the language such that  $\Phi(t)$  is true substitutionally in that interpretation.

Sometimes different quantifiers are used, ‘ $\Pi x$ ’ instead of ‘ $\forall x$ ’ and ‘ $\Sigma x$ ’ instead of ‘ $\exists x$ ’, especially if an author wants to have substitutional quantifiers alongside ordinary quantifiers. I do not follow this practice here.

For philosophers, one main purpose of substitutional semantics is to have variables and quantifiers in an interpreted formal language without thereby taking on ‘ontological commitment’. Presumably, variables and quantifiers,

as understood substitutionally, do not have ‘ranges’ (see, for example, [Gottlieb, 1980] and [Leblanc, 1976]). A nice deal for the anti-realist—perhaps.

Our purposes here are different. We are examining languages and semantics capable of expressing substantial mathematical concepts and describing mathematical structures, like the natural and real numbers. Since this presupposes that there is something to describe, we are not out to reduce ‘ontological commitment’. When adapted to present purposes, however, substitutional semantics has some interesting advantages. It happens that the semantics is not compact, and no effective deductive system is both sound and complete for it. Ironically, a system that is supposedly ‘ontologically’ weaker than first-order (whatever that might mean) is semantically stronger than first-order and is, in a sense, intermediate between first-order and second-order.

It is straightforward to adapt model theory to substitutional semantics. Let  $M$  be a model of a first-order language  $\mathcal{L}1[K]=$  and let  $d$  be the domain of  $M$ . Define  $M$  to be a *substitution model* if for every  $b \in d$ , there is a term  $t$  of  $\mathcal{L}1[K]=$  such that  $t$  denotes  $b$  in  $M$ . In other words,  $M$  is a substitution model if every element of its domain is denoted by a term of the language. Substitution models are good candidates for what may be called ‘substitutional interpretations’ of a formal language like  $\mathcal{L}1[K]=$ .

The usual semantic notions are readily defined. A set  $\Gamma$  of sentences is *substitutionally satisfiable* in  $\mathcal{L}1[K]=$  if there is a substitution model  $M$  such that for every  $\Phi \in \Gamma, M \models \Phi$ ; and a sentence  $\Phi$  is *substitutionally satisfiable* in  $\mathcal{L}1[K]=$  if the singleton  $\{\Phi\}$  is substitutionally satisfiable in  $\mathcal{L}1[K]=$ . An argument  $\langle \Gamma, \Phi \rangle$  is *substitutionally valid* in  $\mathcal{L}1[K]=$ , or  $\Phi$  is a *substitutional consequence* of  $\Gamma$  in  $\mathcal{L}1[K]=$ , if for every substitution model  $M$ , if  $M \models \Psi$  for every  $\Psi \in \Gamma$ , then  $M \models \Phi$ . Finally, a sentence  $\Phi$  is a *substitutional logical truth* in  $\mathcal{L}1[K]=$  if  $\Phi$  is a substitutional consequence of the empty set or, in other words, if  $\Phi$  holds in every substitution model.

In the usual semantics for first-order languages, the properties of a formula, a set of formulas, or an argument, depend only on the non-logical items it contains. For example, if a formula  $\Phi$  is in both  $\mathcal{L}1[K]=$  and  $\mathcal{L}1[K']=$ , then  $\Phi$  is a logical truth in  $\mathcal{L}1[K]=$  if and only if  $\Phi$  is a logical truth in  $\mathcal{L}1[K']=$ . The same goes for higher-order languages, and every other logic presented in this chapter, but not for substitutional semantics. The reason is that the extension of ‘substitution model’ depends on the terminology of the language. For example, if  $K$  consists only of the individual constants  $p$  and  $q$ , then

$$\forall x(x = p \vee x = q)$$

is a substitutional logical truth, as is its consequence  $\exists y \exists z \forall x(x = y \vee x = z)$ . Neither of these sentences is a substitutional logical truth if there is a third constant (or a function letter) in  $K$ .

There is thus a close tie between the non-logical terminology available and the semantic properties of a first-order language construed with substitutional semantics. The link is attenuated somewhat with the customary stipulation that the set  $K$  of non-logical terminology contain infinitely many individual constants. We adopt that convention here, unless explicitly noted otherwise.

With this convention in place, we report some comparisons and some meta-theory:

**THEOREM 22.** *A sentence  $\Phi$  of  $\mathcal{L}1[K]$  is substitutionally satisfiable if and only if  $\Phi$  is satisfiable in the usual first-order semantics. A fortiori,  $\Phi$  is a substitutional logical truth if and only if  $\Phi$  is a logical truth.*

**Proof.** Every substitution model is a model. So if  $\Phi$  is substitutionally satisfiable then  $\Phi$  is satisfiable. For the converse, let  $M$  be a model that satisfies  $\Phi$ . Applying the downward Löwenheim–Skolem theorem, let  $M_1$  be a model whose domain is (at most) countable such that  $M_1 \models \Phi$ . Then let  $M_2$  be a substitution model with the same domain as  $M_1$  such that  $M_2$  agrees with  $M_1$  on every non-logical item that occurs in  $\Phi$ . The model  $M_2$  is obtained by reassigning the non-logical individual constants that do not occur in  $\Phi$ , so that every element of the domain is assigned to at least one constant. It is straightforward to verify that  $M_2 \models \Phi$  (citing the aforementioned fact about first-order model theory). ■

The following is then immediate:

**COROLLARY 23.** *Substitutional semantics is weakly complete. A sentence  $\Phi$  is a substitutional logical truth if and only if  $\Phi$  is deducible in a standard deductive system for first-order logic.*

On the other hand, there is no effective deductive system that is complete for substitutional validity or logical consequence. Consider a language with the non-logical terminology of arithmetic  $\{0, s, +, \cdot\}$  together with the infinite list of individual constants  $\{p_0, p_1, \dots\}$ . Let  $\Gamma$  consist of the successor, addition, and multiplication axioms (see Section 4 above) and the sentences  $p_0 = 0, p_1 = s0, p_2 = ss0, \dots$ . Then a substitution model  $M$  satisfies every member of  $\Gamma$  if and only if  $M$  is isomorphic to the natural numbers, with  $p_0, p_1, \dots$  as the numerals. In other words, in substitutional semantics,  $\Gamma$  is a categorical characterization of the natural numbers. It follows that for every sentence  $\Phi$ ,  $\Phi$  is a substitutional consequence of  $\Gamma$  if and only if  $\Phi$  is true of the natural numbers. As above, it is a corollary of the incompleteness of arithmetic that substitutional semantics is inherently incomplete.

Notice that no induction principle is explicitly included in  $\Gamma$ , and yet each instance of the induction scheme is a substitutional consequence of  $\Gamma$ . In any substitution model of  $\Gamma$ , the denotations of the constants  $p_0, p_1$ , etc. exhaust the domain, and so there is no need for an additional axiom to state this.

This looks like another instance of theft over toil. Recall that one major problem in characterizing the natural numbers up to isomorphism is to state, somehow, that  $0, s0, ss0, \dots$  are all the numbers there are. This can be done with a higher-order language, and with most of the languages developed in this chapter, and of course it cannot be done with any first-order language. Indeed, if a first-order theory of arithmetic has an infinite model at all then it has models that contain elements different from the denotations of  $0, s0, ss0, \dots$ . With substitutional semantics, categoricity is achieved by simply excluding those non-standard models from the semantics, by fiat.

Incidentally, in the example at hand, we added the constants  $p_0, p_1, \dots$  in order to satisfy the convention that there be infinitely many individual constants. If that convention is waived, then the characterization of the natural numbers can be accomplished by a single sentence. Let the set of non-logical terminology be  $\{0, s, +, \cdot\}$  and let  $\Phi$  be the conjunction of the successor, addition, and multiplication axioms. Then for every substitution model  $M$ ,  $M \models \Phi$  if and only if  $M$  is isomorphic to the natural numbers. It follows that if we waive the convention and allow finite sets of non-logical terms, then substitutional semantics is not even weakly complete.

Recall that the usual proof of the upward Löwenheim–Skolem theorem involves adding individual constants to the language. This manoeuvre is not kosher here, since with substitutional semantics the properties of a sentence or a set of sentences are dependent on the non-logical terminology available in the language. Adding new constants would change the extension of ‘substitution model’. In any case, the upward Löwenheim–Skolem theorem fails, trivially. If there are only countably many terms of the language, then there are no uncountable substitution models. There is a more substantial result:

**THEOREM 24.** *There is a set  $\Gamma$  of sentences such that for every natural number  $n > 0$ ,  $\Gamma$  has a substitution model whose domain has cardinality  $n$ , but  $\Gamma$  has no substitution model whose domain is infinite.*

**Proof.** Let  $K$  consist of the unary function letter  $f$  and the individual constants  $t_0, t_1, \dots$ . Let  $\Gamma$  consist of the sentences  $ft_0 = t_1, ft_1 = t_2, ft_2 = t_3, \dots$  and  $\exists x(fx = t_0)$ . For each  $n > 0$ , let the domain of  $M_n$  consist of the natural numbers  $\{0, 1, \dots, n-1\}$ . The structure  $M_n$  assigns each constant  $t_i$  to the remainder when  $i$  is divided by  $n$ , and  $M_n$  assigns  $f$  to the function whose value at  $j$  is the remainder when  $j+1$  is divided by  $n$ . Then  $M_n$  is a substitution model that satisfies every member of  $\Gamma$ . Now, let  $M$  be any substitution model of this language that satisfies every member of  $\Gamma$ . If the domain of  $M$  were infinite, then the denotations in  $M$  of the terms  $t_0, ft_0, fft_0$ , etc. must all be distinct and must exhaust the domain. Thus  $M \models \neg \exists x(fx = t_0)$ . A contradiction. Thus, the domain of  $M$  is finite. ■

Despite this result, there is no characterization of *finitude* in substitution semantics. In particular, for every set  $\Gamma$  of formulas, if every finite substitution model satisfies every member of  $\Gamma$ , then there is an infinite substitution model that also satisfies every member of  $\Gamma$ . On the other hand, if we waive the convention that there be infinitely many individual constants, then we can characterize the notion of finitude with a single sentence. Let the non-logical terminology consist of only the individual constant 0 and the unary function letter  $f$ . Then, for any substitution model  $M$  for this language,

$$M \models \exists x(fx = 0) \vee \exists x\exists y(x \neq y \& fx = fy)$$

if and only if the domain of  $M$  is finite.

**THEOREM 25.** *Substitutional semantics is not compact.*

**Proof.** This is a corollary of Theorem 24, and it can be established in the usual way from the categoricity of the natural numbers. There is, however, a direct way to establish this theorem. Let the non-logical terms consist of the constants  $t_0, t_1, \dots$ , and the monadic predicate letter  $D$ , and let  $\Gamma$  consist of  $Dt_0, Dt_1, \dots$ , together with  $\exists x\neg Dx$ . Then every proper subset of  $\Gamma$  is substitutionally satisfiable and so every finite subset is satisfiable. But  $\Gamma$  itself is not substitutionally satisfiable. ■

To belabour the obvious, no structure whose domain is uncountable can be characterized in substitutional semantics, unless uncountably many non-logical terms are employed. On the other hand, every structure whose domain is countable can be characterized up to isomorphism with substitutional semantics. In general, any structure can be characterized in a language that has as many individual constants as the domain has members. Indeed, let  $M$  be any model of a language  $\mathcal{L}1[K] =$ . Assume that no element of the domain  $d$  of  $M$  is a non-logical term of the associated language (relettering the items in  $K$  if necessary). Now expand the language so that every element of  $d$  is a non-logical constant. That is, consider the language  $\mathcal{L}1[K']$ , where  $K'$  is  $K \cup d$ . Expand the model  $M$  to the new ‘language’, so that each  $b \in d$  denotes itself. Call the result  $M'$ . Clearly,  $M'$  is a substitution model for the expanded language. Let  $\Gamma$  be the set of sentences  $\{\phi \mid M' \models \Phi\}$ . Then any substitution model in the expanded language is isomorphic to  $M$  iff it satisfies every member of  $\Gamma$ .

The idea here is to expand the ‘language’ so that the elements of the domain of the model act as singular terms. The procedure can be reversed. If a set  $\Gamma$  has a substitution model at all, then one can construct such a model from equivalence classes of the terms of the language. In short, a theory that is substitutionally satisfiable carries a model in its syntax. This is probably part of the reason that anti-realists find substitutional semantics attractive. We must remain aware of the complexity and depth of this semantics. See [Dunn and Belnap, 1968].



## ACKNOWLEDGEMENTS

Some of the material here is adapted from [Shapiro, 1991, Chapter 9]. Thanks to Timothy Carlson and Crispin Wright for useful conversations.

*The Ohio State University at Newark and The University of St. Andrews*

## BIBLIOGRAPHY

- [Barwise, 1979] J. Barwise. On branching quantifiers in English, *Journal of Philosophical Logic*, **8**, 47–80, 1979.
- [Barwise, 1985] J. Barwise. Model-theoretic logics: background and aims. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 3–23. Springer Verlag, New York, 1985.
- [Barwise and Feferman, 1985] J. Barwise and S. Feferman, eds. *Model-Theoretic Logics*, Springer-Verlag, New York, 1985.
- [Bell and Slomson, 1971] J. Bell and A. Slomson. *Models and Ultraproducts: An Introduction*. North Holland Publishing Company Amsterdam, 1971.
- [Boolos, 1984] G. Boolos. To be is to be a value of a variable (or to be some values of some variables). *Journal of Philosophy*, **81**, 430–449, 1984.
- [Boolos, 1985] G. Boolos. Nominalist platonism. *The Philosophical Review*, **94**, 327–344, 1985.
- [Boolos, 1985a] G. Boolos. Reading the Begriffsschrift. *Mind*, **94**, 331–344, 1985.
- [Boolos and Jeffrey, 1989] G. Boolos and R. Jeffrey. *Computability and Logic*, third edition. Cambridge University Press, Cambridge, 1989.
- [Chang, 1965] C. Chang. A note on the two cardinal problem. *Proceedings of the American Mathematical Society*, **16**, 1148–1155, 1965.
- [Chang and Keisler, 1973] C. Chang and H. J. Keisler. *Model Theory*. North Holland Publishing Company, Amsterdam, 1973.
- [Church, 1956] A. Church. *Introduction to Mathematical Logic*. Princeton University Press, Princeton, 1973.
- [Corcoran, 1980] J. Corcoran. Categoricity. *History and Philosophy of Logic*, **1**, 187–207, 1980.
- [Cowles, 1979] J. Cowles. The relative expressive power of some logics extending first-order logic. *Journal of Symbolic Logic*, **44**, 129–146, 1979.
- [Dedekind, 1988] R. Dedekind. *Was sind und was sollen die Zahlen?*, Vieweg, Brunswick, 1888; tr. as *The nature and meaning of numbers*. In *Essays on the Theory of Numbers*, W. W. Beman, ed. pp. 31–115, Dover Press, New York, 1963.
- [Dickmann, 1985] M. A. Dickmann. Larger infinitary languages. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 317–363. Springer Verlag, New York, 1985.
- [Dreben and Goldfarb, 1979] B. Dreben and W. Goldfarb. *The Decision Problem: Solvable Classes of Quantificational Formulas*. Addison-Wesley Publishing Company, Inc., London, 1979.
- [Dunn and Belnap, 1968] J. M. Dunn and N. Belnap. The substitution interpretation of the quantifier. *Nous*, **2**, 177–185, 1968.
- [Ebbinghaus, 1985] H. D. Ebbinghaus. Extended logics: The general framework. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 25–76. Springer Verlag, New York, 1985.
- [Fagin, 1974] R. Fagin. Generalized first-order spectra and polynomial-time recognizable sets. *SIAM-AMS Proceedings*, **7**, 43–73, 1974.
- [Feferman, 1977] S. Feferman. Theories of finite type related to mathematical practice. In *Handbook of Mathematical Logic*, J. Barwise, ed. pp. 913–971. North Holland, Amsterdam, 1977.
- [Field, 1994] H. Field. Deflationist views of meaning and content. *Mind*, **103**, 249–285, 1994.

- [Flum, 1985] J. Flum. Characterizing logics. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 77–120. Springer Verlag, New York, 1985.
- [Frege, 1979] G. Frege. *Begriffsschrift, eine der arithmetischen nachgebildete Formelsprache des reinen Denkens*, Louis Nebert, Halle, 1879. In *From Frege to Gödel*, J. van Heijenoort, ed. pp. 1–82. Harvard University Press, Cambridge, Massachusetts, 1967.
- [Gabbay and Moravcsik, 1974] D. Gabbay and J. Moravcsik. Branching quantifiers, English, and Montague grammar. *Theoretical Linguistics*, **1**, 141–157, 1974.
- [Gandy, 1988] R. Gandy. The confluence of ideas in 1936. In *The Universal Turing Machine*, R. Herken ed. pp. 55–111. Oxford University Press, New York, 1988.
- [Gottlieb, 1980] D. Gottlieb. *Ontological Economy: Substitutional Quantification and Mathematics*. Oxford University Press Oxford, 1980.
- [Gurevich, 1985] Y. Gurevich. Monadic second-order theories. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 479–506. Springer Verlag, New York, 1985.
- [Gurevich, 1988] Y. Gurevich. Logic and the challenge of computer science. In *Trends in Theoretical Computer Science*, Egon Börger, ed. pp. 1–57, Computer Science Press, Maryland, 1988.
- [Gurevich and Shelah, 1983] Y. Gurevich and S. Shelah. Interpreting second-order logic in the monadic theory of order. *Journal of Symbolic Logic*, **48**, pp. 816–828, 1983.
- [Henkin, 1953] L. Henkin. Banishing the rule of substitution for functional variables. *Journal of Symbolic Logic*, **18**, 201–208, 1953.
- [Hilbert, 1925] D. Hilbert. Über das Unendliche. *Mathematische Annalen*, **95**, 161–190, 1925. tr. as “On the infinite”, in *From Frege to Gödel*, J. van Heijenoort, ed. pp. 369–392. Harvard University Press, Cambridge, Massachusetts, 1967.
- [Hintikka, 1976] J. Hintikka. Partially ordered quantifiers vs. partially ordered ideas. *Dialectica*. **30**, 89–99, 1976.
- [Immerman, 1987] N. Immerman. Languages that capture complexity classes. *SIAM Journal of Computing*, **16**, 760–778, 1987.
- [Jané, 1993] I. Jané. A critical appraisal of second-order logic. *History and Philosophy of Logic*, **14**, 67–86, 1993.
- [Jensen, 1972] R. B. Jensen. The fine structure of the constructible hierarchy. *Annals of Mathematical Logic*, **4**, 229–308, 1972.
- [Kaufmann, 1985] M. Kaufmann. The quantifier ‘there exist uncountably many’ and some of its relatives. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 123–176. Springer Verlag, New York, 1985.
- [Kolaitis, 1985] P. Kolaitis. Game quantification. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 365–421. Springer Verlag, New York, 1985.
- [Krynicky and Mostowski, 1995] M. Krynicky and M. Mostowski. Henkin quantifiers. In *Quantifiers: Logics, Models and Computation 1*, M. Krynicky, M. Mostowski and L. Szczerba, eds. Kluwer Academic Publishers, Dordrecht, Holland, 1995.
- [Landman, 1989] F. Landman. Groups. *Linguistics and Philosophy*, **12**, 559–605, 723–744, 1989.
- [Lavine, 1994] S. Lavine. *Understanding the Infinite*. Harvard University Press, Cambridge, Massachusetts, 1994.
- [Leblanc, 1976] H. Leblanc. *Truth-value Semantics*, North Holland Publishing Company, Amsterdam, 1976.
- [Leivant, 1989] D. Leivant. Descriptive characterizations of computational complexity. *Journal of Computer and System Sciences*, **39**, 51–83, 1989.
- [Lewis, 1991] D. Lewis. *Parts of Classes*. Blackwell, Oxford, 1991.
- [Lindström, 1969] P. Lindström. On extensions of elementary logic. *Theoria*, **35**, 1–11, 1969.
- [Löwenheim, 1915] L. Löwenheim. Über Möglichkeiten im Relativkalkül. *Mathematische Annalen*, **76**, 447–479, 1915. tr. in *From Frege to Gödel*, J. van Heijenoort, ed. pp. 228–251. Harvard University Press, Cambridge, Massachusetts, 1967.
- [Mendelson, 1987] E. Mendelson. *Introduction to Mathematical Logic*, third edition. van Nostrand, Princeton, 1987.
- [Mundici, 1985] D. Mundici. Other quantifiers: an overview. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 211–233. Springer Verlag, New York, 1985.

- [Nadel, 1985] M. Nadel.  $\mathcal{L}_{\omega_1\omega}$  and admissible fragments. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 271–316. Springer Verlag, New York, 1985.
- [Quine, 1953] W. V. O. Quine. *From a Logical Point of View*. Harper and Row, New York, 1953.
- [Quine, 1986] W. V. O. Quine. *Philosophy of Logic*, second edition. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.
- [Rabin, 1969] M. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, **141**, 1–35, 1969.
- [Resnik, 1988] M. Resnik. Second-order logic still wild. *Journal of Philosophy*, **85**, 75–87, 1988.
- [Schmerl, 1985] J. H. Schmerl. Transfer theorems and their applications to logics. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 177–209. Springer Verlag, New York, 1985.
- [Shapiro, 1991] S. Shapiro. *Foundations Without Foundationalism: A Case for Second-order Logic*. Oxford University Press, Oxford, 1991.
- [Shelah, 1975] S. Shelah. The monadic theory of order. *Annals of Mathematics*, **102**, 379–419, 1975.
- [Sher, 1991] G. Sher. *The Bounds of Logic*. The MIT Press, Cambridge, Massachusetts, 1991.
- [Simpson, 1985] S. Simpson. Friedman’s research on subsystems of second order arithmetic. In *Harvey Friedman’s Research on the Foundations of Mathematics*, L. A. Harrington *et al.* (eds.). North Holland Publishing Company, Amsterdam, 1985.
- [Skolem, 1923] T. Skolem. Begründung der elementaren Arithmetik durch die rekurrierende Denkweise. *Videnskapsselskapets skrifter I. Matematisk-naturvidenskabelig klasse, no. 6*. tr. as ‘The foundations of arithmetic established by the recursive mode of thought’ in *From Frege to Gödel*, J. van Heijenoort, ed. pp. 303–333. Harvard University Press, Cambridge, Massachusetts, 1967.
- [Tarski, 1935] A. Tarski. On the concept of logical consequence. In *Logic, Semantics and Metamathematics*, A. Tarski, pp. 417–429. Clarendon Press, Oxford, 1956.
- [Tarski, 1986] A. Tarski. What are logical notions? (ed by John Corcoran). *History and Philosophy of Logic*, **7**, 143–154, 1986.
- [Väänänen, 1985] J. Väänänen. Set-theoretic definability of logics. In *Model-Theoretic Logics*, J. Barwise and S. Feferman, eds. pp. 599–643. Springer Verlag, New York, 1985.
- [Heijenoort, 1967] J. Van Heijenoort, ed. *From Frege to Gödel*. Harvard University Press, Cambridge, Massachusetts, 1967.
- [Wagner, 1987] S. Wagner. The rationalist conception of logic. *Notre Dame Journal of Formal Logic*, **28**, 3–35, 1987.
- [Zermelo, 1931] E. Zermelo. Über stufen der Quantifikation und die Logik des Unendlichen. *Jahresbericht Deutsche Mathematische Verein*, **31**, 85–88, 1931.



JOHAN VAN BENTHEM AND KEES DOETS

## HIGHER-ORDER LOGIC

### INTRODUCTION

What is nowadays the central part of any introduction to logic, and indeed to some the logical theory par excellence, used to be a modest fragment of the more ambitious language employed in the logicist program of Frege and Russell. ‘Elementary’ or ‘first-order’, or ‘predicate logic’ only became a recognized stable base for logical theory by 1930, when its interesting and fruitful meta-properties had become clear, such as completeness, compactness and Löwenheim-Skolem. Richer higher-order and type theories receded into the background, to such an extent that the (re-) discovery of useful and interesting extensions and variations upon first-order logic came as a surprise to many logicians in the sixties.

In this chapter, we shall first take a general look at first-order logic, its properties, limitations, and possible extensions, in the perspective of so-called ‘abstract model theory’. Some characterizations of this basic system are found in the process, due to Lindström, Keisler-Shelah and Fraïssé. Then, we go on to consider the original mother theory, of which first-order logic was the elementary part, starting from second-order logic and arriving at Russell’s theory of finite types. As will be observed repeatedly, a border has been crossed here with the domain of set theory; and we proceed, as Quine has warned us again and again, at our own peril. Nevertheless, first-order logic has a vengeance. In the end, it turns out that higher-order logic can be viewed from an elementary perspective again, and we shall derive various insights from the resulting semantics.

Before pushing off, however, we have a final remark about possible pretensions of what is to follow. Unlike first-order logic and some of its less baroque extensions, second and higher-order logic have no coherent well-established theory; the existent material consisting merely of scattered remarks quite diverse with respect to character and origin. As the time available for the present enterprise was rather limited (to say the least) the authors do not therefore make any claims as to complete coverage of the relevant literature.

### 1 FIRST-ORDER LOGIC AND ITS EXTENSIONS

The starting point of the present story lies somewhere within Hodges’ (this volume). We will review some of the peculiarities of first-order logic, in order to set the stage for higher-order logics.

### 1.1 Limits of Expressive Power

In addition to its primitives *all* and *some*, a first-order predicate language with identity can also express such quantifiers as *precisely one*, *all but two*, *at most three*, etcetera, referring to specific finite quantities. What is lacking, however, is the general mathematical concept of *finiteness*.

EXAMPLE. The notion ‘finiteness of the domain’ is not definable by means of any first-order sentence, or set of such sentences.

It will be recalled that the relevant refutation turned on the *compactness theorem* for first-order logic, which implies that sentences with arbitrarily large finite models will also have infinite ones.

Another striking omission, this time from the perspective of natural language, is that of common quantifiers, such as *most*, *least*, not to speak of *many* or *few*.

EXAMPLE. The notion ‘most  $A$  are  $B$ ’ is not definable in a first-order logic with identity having, at least, unary predicate constants  $A, B$ . This time, a refutation involves both compactness and the (downward) *Löwenheim–Skolem theorem*: Consider any proposed definition  $\mu(A, B)$  together with the infinite set of assertions ‘at least  $n$   $A$  are  $B$ ’, ‘at least  $n$   $A$  are not  $B$ ’ ( $n = 1, 2, 3, \dots$ ). Any finite subset of this collection is satisfiable in some finite domain with  $A - B$  large enough and  $A \cap B$  a little larger. By compactness then, the whole collection has a model with infinite  $A \cap B$ ,  $A - B$ . But now, the Löwenheim–Skolem theorem gives a countably infinite such model, which makes the latter two sets equinumerous — and ‘most’  $A$  are no longer  $B$ : in spite of  $\mu(A, B)$ .

One peculiarity of this argument is its lifting the meaning of colloquial ‘most’ to the infinite case. The use of infinite models is indeed vital in the coming sections. Only in Section 1.4.3 shall we consider the purely *finite* case: little regarded in mathematically-oriented model theory, but rather interesting for the semantics of natural language.

In a sense, these expressive limits of first-order logic show up more dramatically in a slightly different perspective. A given *theory* in a first-order language may possess various ‘non-standard models’, not originally intended. For instance, by compactness, Peano Arithmetic has non-Archimedean models featuring infinite natural numbers. And by Löwenheim–Skolem, Zermelo–Fraenkel set theory has countable models (if consistent), a phenomenon known as ‘Skolem’s Paradox’. Conversely, a given model may not be defined categorically by its complete first-order theory, as is in fact known for al (infinite) mathematical standard structures such as integers, rationals or reals. (These two observations are sides of the same coin, of course.) Weakness or strength carry no moral connotations in logic, however, as one may turn into the other. Non-standard models for analysis

have turned out quite useful for their own sake, and countable models of set theory are at the base of the independence proofs: first-order logic's loss thus can often be the mathematician's or philosopher's gain.

### 1.2 Extensions

When some reasonable notion falls outside the scope of first-order logic, one rather natural strategy is to add it to the latter base and consider the resulting stronger logic instead. Thus, for instance, the above two examples inspire what is called 'weak second-order logic', adding the quantifier 'there exist finitely many', as well as first-order logic with the added 'generalized quantifier' *most*. But, there is a price to be paid here. Inevitably, these logics lose some of the meta-properties of first-order logic employed in the earlier refutations of definability. Here is a telling little table:

	Compactness	Löwenheim–Sk.
First-order logic	yes	yes
Plus 'there exists finitely many'	no	yes
Plus 'there exist uncountably many'	yes	no
Plus 'most'	no	no

For the second and third rows, cf. [Monk, 1976, Chapter 30]. For the fourth row, here is an argument.

EXAMPLE. Let the most-sentence  $\varphi(R)$  express that  $R$  is a discrete linear order with end points, possessing a greatest point with more successors than non-successors (i.e. most points in the order are its successors). Such orders can only be finite, though of arbitrarily large size: which contradicts compactness. Next, consider the statement that  $R$  is a dense linear order without end points, possessing a point with more successors than predecessors. There are uncountable models of this kind, but no countable ones: and hence Löwenheim–Skolem fails.

As it happens, no proposed proper extension of first-order logic ever managed to retain both the compactness and Löwenheim–Skolem properties. And indeed, in 1969 Lindström proved his famous theorem [Lindström, 1969] that, given some suitable explication of a 'logic', first-order logic is indeed characterizable as the strongest logic to possess these two meta-properties.

### 1.3 Abstract Model Theory

Over the past two decades, many types of extension of first-order logic have been considered. Again, the earlier two examples illustrate general patterns. First, there are so-called *finitary* extensions, retaining the (effective) finite

syntax of first-order logic. The *most* example inspires two general directions of this kind.

First, one may add *generalized quantifiers*  $Q$ , allowing patterns

$$Qx \cdot \varphi(x) \text{ or } Qxy \cdot \varphi(x), \psi(y).$$

E.g. ‘the  $\varphi$ s fill the universe’ (*all*), ‘the  $\varphi$ s form the majority in the universe’ (*most*), ‘the  $\varphi$ s form the majority of the  $\psi$ s’ (most  $\psi$  are  $\varphi$ ). But also, one may stick with the old types of quantifier, while employing them with new ranges. For instance, ‘most  $A$  are  $B$ ’ may be read as an ordinary quantification over functions: ‘there exists a 1–1 correspondence between  $A-B$  and some subset of  $A \cap B$ , but not vice versa’. Thus, one enters the domain of *higher-order logic*, to be discussed in later sections.

The earlier example of ‘finiteness’ may lead to finitary extensions of the above two kinds, but also to an *infinitary* one, where the syntax now allows infinite conjunctions and disjunctions, or even quantifications. For instance, finiteness may be expressed as ‘either one, or two, or three, or . . .’ in  $L_{\omega_1\omega}$ : a first-order logic allowing countable conjunctions and disjunctions of formulas (provided that they have only finitely many free variables together) and finite quantifier sequences. Alternatively, it may be expressed as ‘there are no  $x_1, x_2, \dots$ : all distinct’, which would belong to  $L_{\omega_1\omega_1}$ , having a countably infinite quantifier string. In general, logicians have studied a whole family of languages  $L_{\alpha\beta}$ ; but  $L_{\omega_1\omega}$  remains the favourite (cf. [Keisler, 1971]).

Following Lindström’s result, a research area of ‘abstract model theory’ has arisen where these various logics are developed and compared. Here is one example of a basic theme. Every logic  $L$  ‘casts its net’ over the sea of all structures, so to speak, identifying models verifying the same  $L$ -sentences ( $L$ -equivalence). On the other hand, there is the finest sieve of *isomorphism* between models. One of Lindström’s basic requirements on a logic was that the latter imply the former. One measure of strength of the logic is now to which extent the converse obtains. For instance, when  $L$  is first-order logic, we know that elementary equivalence implies isomorphism for *finite* models, but not for countable ones. (Cf. the earlier phenomenon of non-categorical definability of the integers.) A famous result concerning  $L_{\omega_1\omega}$  is Scott’s theorem to the effect that, for *countable* models,  $L_{\omega_1\omega}$ -equivalence and isomorphism coincide. (Cf. [Keisler, 1971, Chapter 2] or [Barwise, 1975, Chapter VII.6].) That such matches cannot last in the long run follows from a simple set-theoretic consideration, however, first made by Hanf. As long as the  $L$ -sentences form a set, they can distinguish at best  $2^{\|L\|}$  models, up to  $L$ -equivalence — whereas the number of models, even up to isomorphism, is unbounded.

A more abstract line of research is concerned with the earlier meta-properties. In addition to compactness and Löwenheim–Skolem, one also considers such properties as recursive axiomatizability of universally valid



sentences (*'completeness'*) or *interpolation* (cf. Hodges' chapter in this Volume). Such notions may lead to new characterization results. For instance, Lindström himself proved that elementary logic is also the strongest logic with an effective finitary syntax to possess the Löwenheim–Skolem property and be complete. (The infinitary language  $L_{\omega_1\omega}$  has both, without collapsing into elementary logic, however; its countable *admissible fragments* even possess compactness in the sense of [Barwise, 1975].) Similar characterizations for stronger logics have proven rather elusive up till now.

But then, there are many further possible themes in this area which are of a general interest. For instance, instead of haphazardly selecting some particular feature of first-order, or any other suggestive logic, one might proceed to a systematic description of meta-properties.

EXAMPLE. A folklore prejudice has it that interpolation was the 'final elementary property of first-order logic to be discovered'. Recall the statement of this meta-property: if one formula implies another, then (modulo some trivial cases) there exists an *interpolant* in their common vocabulary, implied by the first, itself implying the second. Now, this assertion may be viewed as a (first-order) fact about the two-sorted 'meta-structure' consisting of all first-order formulas, their vocabulary types (i.e. all finite sets of non-logical constants), the relations of implication and type-inclusion, as well as the type-assigning relation. Now, the complete first-order theories of the separate components are easily determined. The pre-order (formulas, implication) carries a definable Boolean structure, as one may define the connectives ( $\wedge$  as greatest lower bound,  $\neg$  as some suitable complement). Moreover, this Boolean algebra is countable, and atomless (the latter by the assumption of an infinite vocabulary). Thus, the given principles are complete, thanks to the well-known categoricity and, hence, completeness of the latter theory. The complete logic of the partial order (finite types, inclusion) may be determined in a slightly more complex way. The vindication of the above conviction concerning the above meta-structure would then consist in showing that interpolation provides the essential link between these two separate theories, in order to obtain a complete axiomatization for the whole.

But as it happens, [Mason, 1985] (in response to the original version of this chapter) has shown that the complete first-order theory of this meta-model is effectively equivalent to True Arithmetic, and hence non-axiomatizable.

Even more revolutionary about abstract model theory is the gradual reversal in methodological perspective. Instead of starting from a given logic and proving some meta-properties, one also considers these properties as such, establishes connections between them, and asks for (the ranges of) logics exemplifying certain desirable combinations of features.

Finally, a warning. The above study by no means exhausts the range of logical questions that can be asked about extensions of first-order logic. Indeed, the perspective of meta-properties is very global and abstract. One more concrete new development is the interest in, e.g. generalized quantifiers from the perspective of linguistic semantics (cf. [Barwise and Cooper, 1981; van Benthem, 1984]), which leads to proposals for reasonable constraints on new quantifiers, and to a semantically-motivated classification of reasonable additions to elementary logic.

#### 1.4 Characterization Results

A good understanding of first-order logic is essential to any study of its extensions. To this end, various characterizations of first-order definability will be reviewed here in a little more detail than in Hodges' chapter.

*1.4.1 Lindström's Theorem.* Lindström's result itself gives a definition of first-order logic, in terms of its global properties. Nevertheless, in practice, it is of little help in establishing or refuting first-order definability. To see if some property  $\Phi$  of models is elementary, one would have to consider the first-order language with  $\Phi$  added (say, as a propositional constant), close under the operations that Lindström requires of a 'logic' (notably, the Boolean operations and relativization to unary predicates), and then find out if the resulting logic possesses the compactness and Löwenheim–Skolem properties. Moreover, the predicate logic is to have an *infinite* vocabulary (cf. the proof to be sketched below): otherwise, we are in for surprises.

EXAMPLE. Lindström's theorem fails for the pure identity language. First, it is a routine observation that sentences in this language can only express (negations of) disjunctions 'there are precisely  $n_1$  or ... or precisely  $n_k$  objects in the universe'. Now, add a propositional constant  $C$  expressing *countable infinity* of the universe.

This logic retains compactness. For, consider any finitely satisfiable set  $\Sigma$  of its sentences. It is not difficult to see that either  $\Sigma \cup \{C\}$  or  $\Sigma \cup \{\neg C\}$  must also be finitely satisfiable. In the first case, replace occurrences of  $C$  in  $\Sigma$  by some tautology: a set of first-order sentences remains, each of whose finite subsets has a (countably) infinite model. Therefore, it has an infinite model itself and, hence, a countably infinite one (satisfying  $C$ ) — by ordinary compactness and Löwenheim–Skolem. This model satisfies the original  $\Sigma$  as well. In the second case, replace  $C$  in  $\Sigma$  by some contradiction. The resulting set either has a finite model, or an infinite one, and hence an uncountably infinite one: either way,  $\neg C$  is satisfied — and again, the original  $\Sigma$  is too.

The logic also retains Löwenheim–Skolem. Suppose that  $\varphi$  has no countably infinite models. Then  $\varphi \wedge \neg C$  has a model, if  $\varphi$  has one. Again, replace occurrences of  $C$  inside  $\varphi$  by some contradiction: a pure identity sentence

remains. But such sentences can always be verified on some finite universe (witness the above description) where  $\neg C$  is satisfied too.

*1.4.2 Keisler's Theorem.* A more local description of first-order definability was given by Keisler, in terms of preservation under certain basic operations on models.

**THEOREM.** *A property  $\Phi$  of models is definable by means of some first-order sentence iff both  $\Phi$  and its complement are closed under the formation of isomorphisms and ultraproducts.*

The second operation has not been introduced yet. As it will occur at several other places in this Handbook, a short introduction is given at this point. For convenience, henceforth, our standard example will be that of binary relational models  $F = \langle A, R \rangle$  (or  $F_i = \langle A_i, R_i \rangle$ ).

A *logical fable*. A family of models  $\{F_i \mid i \in I\}$  once got together and decided to join into a common state. As everyone wanted to be fully represented, it was decided to create new composite individuals as functions  $f$  with domain  $I$ , picking at each  $i \in I$  some individual  $f(i) \in A_i$ . But now, how were relations to be established between these new individuals? Many models were in favour of consensus democracy:

$$Rfg \text{ iff } R_i f(i)g(i) \text{ for all } i \in I.$$

But, this led to indeterminacies as soon as models started voting about whether or *not*  $Rfg$ . More often than not, no decision was reached. Therefore, it was decided to ask the gods for an 'election manual'  $U$ , saying which sets of votes were to be 'decisive' for a given atomic statement. Thus, votes now were to go as follows:

$$Rfg \text{ iff } \{i \in I \mid R_i f(i)g(i)\} \in U. \quad (*)$$

Moreover, although one should not presume in these matters, the gods were asked to incorporate certain requirements of consistency

$$\text{if } X \in U, \text{ then } I - X \notin U$$

as well as democracy

$$\text{if } X \in U \text{ and } Y \supseteq X, \text{ then } Y \in U.$$

Finally, there was also the matter of expediency: the voting procedure for atomic statements should extend to complex decisions:

$$\varphi(f_1, \dots, f_n) \text{ iff } \{i \in I \mid F_i \models \varphi[f_1(i), \dots, f_n(i)]\} \in U$$

for all predicate-logical issues  $\varphi$ .

After having pondered these wishes, the gods sent them an *ultrafilter*  $U$  over  $I$ , proclaiming the Łoś Equivalence:

**THEOREM.** *For any ultrafilter  $U$  over  $I$ , the stipulation  $(*)$  creates a structure  $F = \langle \prod_{i \in I} A_i, R \rangle$  such that*

$$F \models \varphi[f_1, \dots, f_n] \text{ iff } \{i \in I \mid F_i \models \varphi[f_1(i), \dots, f_n(i)]\} \in U.$$

**Proof.** The basic case is just  $(*)$ . The negation and conjunction cases correspond to precisely the defining conditions on ultrafilters, viz. (i)  $X \notin U$  iff  $I - X \in U$ ; (ii)  $X, Y \in U$  iff  $X \cap Y \in U$  (or, alternatively, besides consistency and democracy above: if  $X, Y \in U$  then also  $X \cap Y \in U$ ; and: if  $I - X \notin U$  then  $X \in U$ ). And finally, the gods gave them the existential quantifier step for free:

- if  $\exists x \varphi(x, f_1, \dots, f_n)$  holds then so does  $\varphi(f, f_1, \dots, f_n)$  for some function  $f$ . Hence, by the inductive hypothesis for  $\varphi$ , we have that  $\{i \in I \mid F_i \models \varphi[f(i), f_1(i), \dots, f_n(i)]\} \in U$ , which set is contained in  $\{i \in I \mid F_i \models \exists x \varphi[f_1(i), \dots, f_n(i)]\} \in U$ .
- if  $\{i \in I \mid F_i \models \exists x \varphi[f_1(i), \dots, f_n(i)]\} \in U$ , then choose  $f(i) \in A_i$  verifying  $\varphi$  for each of these  $i$  (and arbitrary elsewhere): this  $f$  verifies  $\varphi(x, f_1, \dots, f_n)$  in the whole product, whence  $\exists x \varphi(f_1, \dots, f_n)$  holds. ■

After a while, an unexpected difficulty occurred. Two functions  $f, g$  who did not agree among themselves asked for a public vote, and the outcome was ...

$$\{i \in I \mid f(i) = g(i)\} \in U.$$

Thus it came to light how the gift of the gods had introduced an invisible equality  $\sim$ . By its definition and the Łoś Equivalence, it even turned out to partition the individuals into equivalence classes, whose members were indistinguishable as to  $R$  behaviour:

$$Rfg, f \sim f', g \sim g' \text{ imply } Rf'g'.$$

But then, such classes themselves could be regarded as the building bricks of society, and in the end there were:

**DEFINITION.** For any family of models  $\{F_i \mid i \in I\}$  with an ultrafilter  $U$  on  $I$ , the *ultraproduct*  $\prod_U F_i$  is the model  $\langle A, R \rangle$  with

1.  $A$  is the set of classes  $f_\sim$  for all functions  $f \in \prod_{i \in I} A_i$ , where  $f_\sim$  is the equivalence class of  $f$  in the above relation,
2.  $R$  is the set of couples  $\langle f_\sim, g_\sim \rangle$  for which  $\{i \in I \mid R_i f(i)g(i)\} \in U$ .

By the above observations, the latter clause is well-defined — and indeed the whole Łoś Equivalence remained valid.

Whatever their merits as regards democracy, ultraproducts play an important role in the following fundamental question of model theory:

What structural behaviour makes a class of models *elementary*, i.e. definable by means of some first-order sentence?

First, the Łoś Equivalence implies that first-order sentences  $\varphi$  are preserved under ultraproducts in the following sense:

$$\text{if } F_i \models \varphi \text{ (all } i \in I), \text{ then } \prod_U F_i \models \varphi.$$

(The reason is that  $I$  itself must belong to  $U$ .) But conversely, Keisler's theorem told us that this is also enough. *End of fable.*

The proof of Keisler's theorem (subsequently improved by Shelah) is rather formidable: cf. [Chang and Keisler, 1973, Chapter 6]. A more accessible variant will be proved below, however. First, one relaxes the notion of isomorphism to the following partial variant.

**DEFINITION.** A *partial isomorphism* between  $\langle A, R \rangle$  and  $\langle B, S \rangle$  is a set  $I$  of coupled finite sequences  $(s, t)$  from  $A$  resp.  $B$ , of equal length, satisfying

$$\begin{aligned} (s)_i = (s)_j & \text{ iff } (t)_i = (t)_j \\ (s)_i R (s)_j & \text{ iff } (t)_i S (t)_j \end{aligned}$$

which possesses the *back-and-forth property*, i.e. for every  $(s, t) \in I$  and every  $a \in A$  there exists some  $b \in B$  with  $(s \hat{\ } a, t \hat{\ } b) \in I$ ; and vice versa.

Cantor's zig-zag argument shows that partial isomorphism coincides with total isomorphism on the countable models. Higher up, matters change; e.g.  $\langle \mathbb{Q}, < \rangle$  and  $\langle \mathbb{R}, < \rangle$  are partially isomorphic by the obvious  $I$  without being isomorphic.

First-order formulas  $\varphi$  are preserved under partial isomorphism in the following sense:

$$\text{if } (s, t) \in I, \text{ then } \langle A, R \rangle \models \varphi[s] \text{ iff } \langle B, S \rangle \models \varphi[t].$$

Indeed, this equivalence extends to formulas from arbitrary infinitary languages  $L_{\alpha\omega}$ : cf. [Barwise, 1977, Chapter A.2.9] for further explanation.

**THEOREM.** *A property  $\Phi$  of models is first-order definable iff both  $\Phi$  and its complement are closed under the formation of partial isomorphisms and countable ultraproducts.*

*1.4.3 Fraïssé's Theorem.* Even the Keisler characterization may be difficult to apply in practice, as ultraproducts are such abstract entities. In many cases, a more combinatorial method may be preferable; in some, it's even necessary.

EXAMPLE. As was remarked earlier, colloquial ‘most’ only seems to have natural meaning on the *finite* models. But, as to first-order definability on this restricted class, both previous methods fail us completely, all relevant notions being tied up with infinite models. Nevertheless, *most A are B* is not definable on the finite models in the first-order language with  $A, B$  and identity. But this time, we need a closer combinatorial look at definability.

First, a natural measure of the ‘pattern complexity’ of a first-order formula  $\varphi$  is its *quantifier depth*  $d(\varphi)$ , which is the maximum length of quantifier nestings inside  $\varphi$ . (Inductively,  $d(\varphi) = 0$  for atomic  $\varphi$ ,  $d(\neg\varphi) = d(\varphi)$ ,  $d(\varphi \wedge \psi) = \max(d(\varphi), d(\psi))$ , etcetera,  $d(\exists x\varphi) = d(\forall x\varphi) = d(\varphi) + 1$ .) Intuitively, structural complexity beyond this level will escape  $\varphi$ ’s notice. We make this precise.

Call two sets  $X, Y$  *n-equivalent* if either  $|X| = |Y| < n$  or  $|X|, |Y| \geq n$ . By extension, call two models  $\langle D, A, B \rangle, \langle D', A', B' \rangle$  *n-equivalent* if all four ‘state descriptions’  $A \cap B, A - B, B - A, D - (A \cup B)$  are *n-equivalent* to their primed counterparts.

LEMMA. *If  $\langle D, A, B \rangle, \langle D', A', B' \rangle$  are n-equivalent then all sequences  $d, d'$  with corresponding points in corresponding states verify the same first-order formulas with quantifier depth not exceeding  $n$ .*

COROLLARY. *‘Most A are B’ is not first-order definable on the finite models.*

**Proof.** For no finite number  $n$ , ‘most A are B’ exhibits the required *n*-insensitivity. ■

This idea of insensitivity to structural complexity beyond a certain level forms the core of our third and final characterization, due to Fraïssé. Again, only the case of a binary relation  $R$  will be considered, for ease of demonstration.

First, on the linguistic side, two models are *n-elementarily equivalent* if they verify the same first-order sentences of quantifier depth not exceeding  $n$ . Next, on the structural side, a matching notion of *n-partial isomorphism* may be defined, by postulating the existence of a chain  $I_n, \dots, I_0$  of sets of matching couples  $(s, t)$ , as in the earlier definition of partial isomorphism. This time, the back-and-forth condition is index-relative, however:

if  $(s, t) \in I_{i+1}$  and  $a \in A$ , then for some  $b \in B$ ,  $(s \frown a, t \frown b) \in I_i$ ,  
and vice versa.

PROPOSITION. *Two models are n-elementarily equivalent iff they are n-partially isomorphic.*

The straightforward proof uses the following auxiliary result, for first-order languages with a finite non-logical vocabulary of relations and individual constants.

LEMMA. *For each depth  $n$  and for each fixed number of free variables  $x_1, \dots, x_m$ , there exist only finitely many formulas  $\varphi(x_1, \dots, x_m)$ , up to logical equivalence.*

This lemma allows us to describe all possible  $n$ -patterns in a single first-order formula, a purpose for which one sometimes uses explicit ‘Hintikka normal forms’.

THEOREM. *A property  $\Phi$  of models is first-order definable iff it is preserved under  $n$ -partial isomorphism for some natural number  $n$ .*

**Proof.** The invariance condition is obvious for first-order definable properties. Conversely, for  $n$ -invariant properties, the disjunction of all complete  $n$ -structure descriptions for models satisfying  $\Phi$  defines the latter property. ■

*Applications.* Now, from the Fraïssé theorem, both the weak Keisler and the Lindström characterization may be derived in a perspicuous way. Here is an indication of the proofs.

EXAMPLE. (*Weak Keisler from Fraïssé*) First-order definable properties are obviously preserved under partial isomorphism and (countable) ultraproducts. As for the converse, suppose that  $\Phi$  is not thus definable. By Fraïssé, this implies the existence of a sequence of  $n$ -partially isomorphic model pairs  $\mathfrak{A}_n, \mathfrak{B}_n$  of which only the first verify  $\Phi$ .

The key observation is now simply this. Any *free* ultrafilter  $U$  on  $\mathbb{N}$  (containing all tails of the form  $[n, \infty)$ ) will make the countable ultraproducts  $\Pi_U \mathfrak{A}_n, \Pi_U \mathfrak{B}_n$  partially isomorphic. The trick here is to find a suitable set  $I$  of partial isomorphisms, and this is accomplished by setting, for sequences of functions  $s, t$  of length  $m$

$$((s)_U, (t)_U) \in I \text{ iff } \{n \geq m \mid (s(n), t(n)) \in I_{n-m}^n\} \in U$$

where ‘ $I_n^m, \dots, I_0^0$ ’ is the sequence establishing the  $n$ -partial isomorphism of  $\mathfrak{A}_n, \mathfrak{B}_n$ .

So, by the assumed preservation properties,  $\Phi$  would hold for  $\Pi_U \mathfrak{A}_n$  and hence for  $\Pi_U \mathfrak{B}_n$ . But, so would not- $\Phi$ : a contradiction.

EXAMPLE. (*Lindström from Fraïssé*) Let  $L$  be a logic whose non-logical vocabulary consists of infinitely many predicate constants of all arities.  $L$  is completely specified by its *sentences*  $S$ , each provided with a finite ‘type’ (i.e. set of predicate constants), its *models*  $M$  (this time: ordinary first-order models) and its *truth relation*  $T$  between sentences and models. We assume four basic conditions on  $L$ : the truth relation is invariant for *isomorphisms*, the sentence set  $S$  is closed under *negations* and *conjunctions* (in the obvious

semantic sense), and each sentence  $\varphi$  can be *relativized* by arbitrary unary predicates  $A$ , such that a model verifies  $\varphi^A$  iff its  $A$ -submodel verifies  $\varphi$ . Finally, we say that  $L$  ‘contains elementary logic’ if each first-order sentence is represented by some sentence in  $S$  having the same models. ‘Compactness’ and ‘Löwenheim–Skolem’ are already definable in this austere framework. (By the latter we’ll merely mean: ‘sentences with any model at all have countable models’.)

**THEOREM.** *Any logic containing elementary logic has compactness and Löwenheim–Skolem iff it coincides with elementary logic.*

The non-evident half of this assertion again starts from Fraïssé’s result. Suppose that  $\Phi \in S$  is not first-order. Again, there is a sequence  $\mathfrak{A}_n, \mathfrak{B}_n$  as above. For a natural number  $n$ , consider the complex model (an expanded “model pair”)  $\mathfrak{M}_n = (\mathfrak{A}_n, \mathfrak{B}_n, R_0, \dots, R_n)$ , where the  $2i$ -ary relations  $R_i \subseteq A_n^i \times B_n^i$  ( $i = 0, \dots, n$ ) are defined by

$$R_i(a_1, \dots, a_i, b_1, \dots, b_i) := (\mathfrak{A}_n, a_1, \dots, a_i) \equiv^{n-i} (\mathfrak{B}_n, b_1, \dots, b_i)$$

( $\equiv^{n-i}$  denoting  $(n-i)$ -equivalence here). The model  $\mathfrak{M}_n$  satisfies sentences expressing that

1.  $\Phi$  is true in its first component  $\mathfrak{A}_n$  but false in its second one  $\mathfrak{B}_n$ , (note that we use relativizations here),
2. if  $i \leq n$  and  $R_i(a_1, \dots, a_i, b_1, \dots, b_i)$  holds, then the relation  $\{(a_1, b_1), \dots, (a_i, b_i)\}$  between the component-models  $\mathfrak{A}_n$  and  $\mathfrak{B}_n$  has the properties of a partial isomorphism (preservation of equality and relations) introduced earlier,
3. (a)  $R_0$  (which has 0 arguments) is true (of the empty sequence),  
 (b) if  $i < n$  and  $R_i(a_1, \dots, a_i, b_1, \dots, b_i)$  holds, then for all  $a \in A_n$  there exists  $b \in B_n$  such that  $R_{i+1}(a_1, \dots, a_i, a, b_1, \dots, b_i, b)$ , and vice versa.

By the Downward Löwenheim–Skolem and Compactness property, there is a *countable* complex  $(\mathfrak{A}, \mathfrak{B}, R_0, R_1, R_2, \dots)$  with an *infinite* sequence  $R_0, R_1, R_2, \dots$  that satisfies these requirements for *every*  $i$ . By requirements 2 and 3 and Cantor’s zig-zag argument it then follows that  $\mathfrak{A} \cong \mathfrak{B}$ . However, this outcome contradicts requirement 1. ■

## 2 SECOND-ORDER LOGIC

### 2.1 Language

Quantification over properties and predicates, rather than just objects, has a philosophical pedigree. For instance, Leibniz’s celebrated principle of



Identity of Indiscernibles has the natural form

$$\forall xy(\forall X(X(x) \leftrightarrow X(y)) \rightarrow x = y).$$

There also seems to be good evidence for this phenomenon from natural language, witness Russell's example 'Napoleon had all the properties of a great general'

$$\forall X(\forall y(GG(y) \rightarrow X(y)) \rightarrow X(n)).$$

Moreover, of course, mathematics abounds with this type of discourse, with its explicit quantification over relations and functions. And indeed, logic itself seems to call for this move. For, there is a curious asymmetry in ordinary predicate logic between individuals: occurring both in constant and variable contexts, and predicates: where we are denied the power of quantification. This distinction seems arbitrary: significantly, Frege's *Be-griffsschrift* still lacks it. We now pass on to an account of second-order logic, with its virtues and vices.

The language of second-order logic distinguishes itself from that of first-order logic by the addition of variables for subsets, relations and functions of the universe and the possibility of quantification over these. The result is extremely strong in expressive power; we list a couple of examples in Section 2.2. As a consequence, important theorems valid for first-order languages fail here; we mention the compactness theorem, the Löwenheim–Skolem theorems (Section 2.2) and the completeness theorem (Section 2.3). With second-order logic, one really enters the realm of set theory. This state of affairs will be illustrated in Section 2.4 with a few examples. What little viable logic can be snatched in the teeth of these limitations usually concerns special fragments of the language, of which some are considered in Section 2.5.

## 2.2 Expressive Power

*2.2.1.* An obvious example of a second-order statement is Peano's induction axiom according to which every set of natural numbers containing 0 and closed under immediate successors contains all natural numbers. Using  $S$  for successor, this might be written down as

$$\forall Y[Y(0) \wedge \forall x(Y(x) \rightarrow Y(S(x))) \rightarrow \forall xY(x)] \quad (1)$$

(The intention here is that  $x$  stands for numbers,  $Y$  for sets of numbers, and  $Y(x)$  says, as usual, that  $x$  is an element of  $Y$ .)

Dedekind already observed that the axiom system consisting of the induction axiom and the two first-order sentences

$$\forall x\forall y(S(x) = S(y) \rightarrow x = y) \quad (2)$$

and

$$\forall x(S(x) \neq 0) \quad (3)$$

is categorical. Indeed suppose that  $\langle A, f, a \rangle$  models (1)–(3). Let  $A' = \{a, f(a), f(f(a)), \dots\}$ . Axioms (2) and (3) alone imply that the submodel  $\langle A', f \upharpoonright A', a \rangle$  is isomorphic with  $\langle \mathbb{N}, S, 0 \rangle$  (the isomorphism is clear). But, (1) implies that  $A' = A$  (just let  $X$  be  $A'$ ).

This result should be contrasted with the first-order case. *No* set of first-order sentences true of  $\langle \mathbb{N}, S, 0 \rangle$  is categorical. This can be proved using either the upward Löwenheim–Skolem theorem or the compactness theorem. As a result, neither of these two extend to second-order logic. The nearest one can come to (1) in first-order terms is the ‘schema’

$$\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(S(x))) \rightarrow \forall x\varphi(x) \quad (4)$$

where  $\varphi$  is any first-order formula in the vocabulary under consideration. It follows that in models  $\langle A, f, a \rangle$  of (4) the set  $A'$  above cannot be defined in first-order terms: otherwise one could apply (4) showing  $A' = A$  just as we applied (1) to show this before. (This weakness of first-order logic becomes its strength in so-called ‘overspill arguments’, also mentioned in Hodges’ chapter (this Volume).) We will use the categoricity of (1)–(3) again in Section 2.3 to show non-axiomatizability of second-order logic.

*2.2.2.* The next prominent example of a second-order statement is the one expressing ‘Dedekind completeness’ of the order of the reals: every set of reals with an upper bound has a least upper bound. Formally

$$\begin{aligned} \forall X[\exists x\forall y(X(y) \rightarrow y \leq x) \rightarrow \\ \rightarrow \exists x(\forall y(X(y) \rightarrow y \leq x) \wedge \forall x'[\forall y(X(y) \rightarrow y \leq x') \rightarrow x \leq x'])] \quad (5) \end{aligned}$$

It is an old theorem of Cantor’s that (5) together with the first-order statements expressing that  $\leq$  is a dense linear order without endpoints plus the statement ‘there is a countable dense subset’, is categorical. The latter statement of so-called ‘separability’ is also second-order definable: cf. Section 2.2.5. Without it a system is obtained whose models all *embed*  $\langle \mathbb{R}, \leq \rangle$ . (For, these models must embed  $\langle \mathbb{Q}, \leq \rangle$  for first-order reasons; and such an embedding induces one for  $\langle \mathbb{R}, \leq \rangle$  by (5).) Thus, the downward Löwenheim–Skolem theorem fails for second-order logic.

*2.2.3.* A relation  $R \subseteq A^2$  is *well-founded* if every non-empty subset of  $A$  has an  $R$ -minimal element. In second-order terms w.r.t. models  $\langle A, R, \dots \rangle$

$$\forall X[\exists xX(x) \rightarrow \exists x(X(x) \wedge \forall y[X(y) \rightarrow \neg R(y, x)]] \quad (6)$$

This cannot be expressed in first-order terms. For instance, every first-order theory about  $R$  which admits models with  $R$ -chains of arbitrary large but

finite length must, by compactness, admit models with infinite  $R$ -chains which decrease, and such a chain has no minimal element.

*2.2.4.* Every first-order theory admitting arbitrarily large, finite models has infinite models as well: this is one of the standard applications of compactness. On the other hand, higher-order terms enable one to define finiteness of the universe. Probably the most natural way to do this uses *third-order* means: a set is finite iff it is in every *collection of* sets containing the empty set and closed under the addition of one element. Nevertheless, we can define finiteness in second-order terms as well:  $A$  is finite iff every relation  $R \subseteq A^2$  is well-founded; hence, a second-order definition results from (6) by putting a universal quantifier over  $R$  in front. Yet another second-order definition of finiteness uses Dedekind's criterion: every injective function on  $A$  is surjective. Evidently, such a quantification over functions on  $A$  may be simulated using suitable predicates. By the way, to see that these second-order sentences do indeed define finiteness one needs the axiom of choice.

*2.2.5 Generalized Quantifiers.* Using Section 2.2.4, it is easy to define the quantifier  $\exists_{<\aleph_0}$  (where  $\exists_{<\aleph_0} x\varphi(x)$  means: there are only finitely many  $x$  s.t.  $\varphi(x)$ ) in second-order terms;  $\exists_{\geq\aleph_0}$  simply is its negation. (In earlier terminology, weak second-order logic is part of second-order logic.) What about higher cardinalities? Well, e.g.  $|X| \geq \aleph_1$  iff  $X$  has an infinite subset  $Y$  which cannot be mapped one-one onto  $X$ . This can obviously be expressed using function quantifiers. And then of course one can go on to  $\aleph_2, \aleph_3, \dots$

Other generalized quantifiers are definable by second-order means as well. For instance, the standard example of Section 1 has the following form. *Most  $A$  are  $B$*  becomes 'there is no injective function from  $A \cap B$  into  $A - B$ '.

A highly successful generalized quantifier occurs in *stationary logic*, cf. [Barwise *et al.*, 1978]. Its language is second-order in that it contains monadic second-order variables; but the only quantification over these *almost all* quantifier  $aa$ . A sentence  $aaX\varphi(X)$  is read as: there is a collection  $C$  of countable sets  $X$  for which  $\varphi(X)$ , which is closed under the formation of countable unions and has the property that every countable subset of the universe is subset of a member of  $C$ . (We'll not take the trouble explaining what 'stationary' means here.) The obvious definition of  $aa$  in higher-order terms employs third-order means. Stationary logic can define the quantifier  $\exists_{\geq\aleph_1}$ . It has a complete axiomatization and, as a consequence, obeys compactness and downward Löwenheim-Skolem (in the form: if a sentence has an uncountable model, it has one of power  $\aleph_1$ ).

Other compact logics defining  $\exists_{\geq\aleph_1}$  have been studied by Magidor and Malitz [1977].

*2.2.6.* The immense strength of second-order logic shows quite clearly when set theory itself is considered.

Zermelo's *separation axiom* says that the elements of a given set sharing a given property form again a set. Knowing of problematic properties occurring in the paradoxes, he required 'definiteness' of properties to be used. In later times, Skolem replaced this by 'first-order definability', and the axiom became a first-order schema. Nevertheless, the intended axiom quite simply is the second-order statement

$$\forall X \forall x \exists y \forall z (z \in y \leftrightarrow z \in x \wedge X(z)) \quad (8)$$

Later on, Fraenkel and Skolem completed Zermelo's set theory with the substitution axiom: the complete image of a set under an operation is again a set, resulting from the first by 'substituting' for its elements the corresponding images. Again, this became a first-order schema, but the original intention was the second-order principle

$$\forall F \forall a \exists b \forall y (y \in b \leftrightarrow \exists x [x \in a \wedge y = F(x)]) \quad (9)$$

Here  $F$  is used as a variable for arbitrary operations from the universe to itself;  $F(x)$  denotes application. The resources of set theory allow an equivalent formulation of (9) with a set (i.e. class) variable, of course. Together with the usual axioms, (9) implies (8).

It must be considered quite a remarkable fact that the first-order versions of (8) and (9) have turned out to be sufficient for every mathematical purpose. (By the way, in ordinary mathematical practice, (9) is seldom used; the proof that Borel-games are determined is a notable exception. Cf. also Section 2.4.)

The Zermelo–Fraenkel axioms intend to describe the cumulative hierarchy with its membership structure  $\langle V, \in \rangle$ , where  $V = \cup_{\alpha} V_{\alpha}$  ( $\alpha$  ranging over all ordinals) and  $V_{\alpha} = \cup_{\beta < \alpha} \mathcal{P}V_{\beta}$ . For the reasons mentioned in Section 1, the first-order version  $\text{ZF}^1$  of these axioms does not come close to this goal, as it has many non-standard models as well. The second-order counterpart  $\text{ZF}^2$  using (9) has a much better score in this respect:

**THEOREM.**  $\langle A, E \rangle$  satisfies  $\text{ZF}^2$  iff for some strongly inaccessible cardinal  $\kappa : \langle A, E \rangle \cong \langle V_{\kappa}, \in \rangle$ .

It is generally agreed that the models  $\langle V_{\kappa}, \in \rangle$  are 'standard' to a high degree.

If we add an axiom to  $\text{ZF}^2$  saying there are no inaccessibles, the system even becomes categorical, defining  $\langle V_{\kappa}, \in \rangle$  for the first inaccessible  $\kappa$ .

### 2.3 Non-axiomatizability

First-order logic has an effective notion of proof which is complete w.r.t. the intended interpretation. This is the content of Gödel's completeness theorem. As a result, the set of (Gödel numbers of) universally valid first-order

formulas is recursively enumerable. Using second-order Example 2.2.1, it is not hard to show that the set of second-order validities is not arithmetically definable, let alone recursively enumerable and hence that an effective and complete axiomatization of second-order validity is impossible.

Let  $P^2$  be Peano arithmetic in its second-order form, i.e. the theory in the language of  $\mathfrak{N} = \langle \mathbb{N}, S, 0, +, \times \rangle$  consisting of (1)–(3) above plus the (first-order) recursion equations for  $+$  and  $\times$ .  $P^2$  is a categorical description of  $\mathfrak{N}$ , just as (1)–(3) categorically describe  $\langle \mathbb{N}, S, 0 \rangle$ . Now, let  $\varphi$  be any first-order sentence in the language of  $\mathfrak{N}$ . Then clearly

$$\mathfrak{N} \models \varphi \text{ iff } P^2 \rightarrow \varphi \text{ is valid.}$$

(Notice that  $P^2$  may be regarded as a single second-order sentence.)

Now the left-hand side of this equivalence expresses a condition on (the Gödel number of)  $\varphi$  which is not arithmetically definable by Tarski's theorem on non-definability of truth (cf. Section 3.2 or, for a slightly different setting, see Section 20 of Hodges' chapter in this Volume). Thus, second-order validity cannot be arithmetical either. ■

Actually, this is still a very weak result. We may take  $\varphi$  second-order and show that second-order truth doesn't fit in the analytic hierarchy (again, see Section 3.4). Finally, using Section 2.2.6, we can replace in the above argument  $\mathfrak{N}$  by  $\langle V_\kappa, \in \rangle$ , where  $\kappa$  is the smallest inaccessible, and  $P^2$  by  $\text{ZF}^2 +$  'there are no inaccessibles', and find that second-order truth cannot be (first-order) defined in  $\langle V_\kappa, \in \rangle$ , etc. This clearly shows how frightfully complex this notion is.

Not to end on too pessimistic a note, let it be remarked that the logic may improve considerably for certain fragments of the second-order language, possibly with restricted classes of models. An early example is the decidability of second-order monadic predicate logic (cf. [Ackermann, 1968]). A more recent example is Rabin's theorem (cf. [Rabin, 1969]) stating that the monadic second-order theory (employing only second-order quantification over subsets) of the structure  $\langle 2^\omega, P_0, P_1 \rangle$  is still decidable. Here,  $2^\omega$  is the set of all finite sequences of zeros and ones, and  $P_i$  is the unary operation 'post-fix  $i$ ' ( $i = 0, 1$ ).

Many decidability results for monadic second-order theories have been derived from this one by showing their models to be definable parts of the Rabin structure. For instance, the monadic second-order theory of the natural numbers  $\langle \mathbb{N}, < \rangle$  is decidable by this method.

The limits of Rabin's theorem show up again as follows. The *dyadic* second-order theory of  $\langle \mathbb{N}, < \rangle$  is already non-arithmetical, by the previous type of consideration. (Briefly,  $\mathfrak{N} \models \varphi$  iff  $\langle \mathbb{N}, < \rangle$  verifies  $P \rightarrow \varphi$  for all those choices of  $0, S, +, \times$  whose defined relation 'smaller than' coincides with the actual  $<$ . Here,  $P$  is first-order Peano Arithmetic minus induction. In this

formulation, ternary predicates are employed (for  $+$ ,  $\times$ ), but this can be coded down to the binary case.)

#### 2.4 *Set-Theoretic Aspects*

Even the simplest questions about the model theory of second-order logic turn out to raise problems of set theory, rather than logic. Our first example of this phenomenon was a basic theme in Section 1.3.

If two models are first-order (elementarily) equivalent and one of them is finite, they must be isomorphic. What, if we use second-order equivalence and relax finiteness to, say, countability? Ajtai [1979] contains a proof that this question is undecidable in ZF (of course, the *first-order* system is intended here). One of his simplest examples shows it is consistent for there to be two countable well-orderings, second-order (or indeed higher-order) equivalent but not isomorphic.

The germ of the proof is in the following observation. If the Continuum Hypothesis holds, there must be second-order (or indeed higher-order) equivalent well-orderings of power  $\aleph_1$ : for, up to isomorphism, there are  $\aleph_2$  such well orderings (by the standard representation in terms of ordinals), whereas there are only  $2^{\aleph_0} = \aleph_1$  second-order theories. The consistency-proof itself turns on a refined form of this cardinality-argument, using ‘cardinal collapsing’. On the other hand, Ajtai mentions the ‘folklore’ fact that countable second-order equivalent models *are* isomorphic when the axiom of constructibility holds. In fact, this may be derived from the existence of a second-order definable well-ordering of the reals (which follows from this axiom).

Another example belongs to the field of second-order cardinal characterization (cf. [Garland, 1974]). Whether a sentence without non-logical symbols holds in a model or not depends only on the cardinality of the model. If a sentence has models of one cardinality only, it is said to *characterize* that cardinal. As we have seen in Section 1, first-order sentences can only characterize single finite cardinals. In the meantime, we have seen how to characterize, e.g.  $\aleph_0$  in a second-order way: let  $\varphi$  be the conjunct of (1)–(3) of Section 2.2.1 and consider  $\exists S \exists 0 \varphi$  — where  $S$  and  $0$  are now being considered as variables. Now, various questions about the simplest second-order definition of a given cardinal, apparently admitting of ‘absolute’ answers, turn out to be undecidable set theoretic problems; cf. [Kunen, 1971].

As a third example, we finally mention the question of cardinals characterizing, conversely, a logic  $L$ . The oldest one is the notion of *Hanf number* of a logic, alluded to in Section 1.3. This is the least cardinal  $\gamma$  such that, if an  $L$ -sentence has a model of power  $\geq \gamma$ , it has models of arbitrarily large powers. The *Löwenheim number*  $\lambda$  of a language  $L$  compares to the *downward* Löwenheim–Skolem property just as the Hanf number does to

the *upward* notion: it is the least cardinal with the property that every satisfiable  $L$ -sentence has a model of power  $\leq \lambda$ . It exists by a reasoning similar to Hanf's: for satisfiable  $\varphi$ , let  $|\varphi|$  be the least cardinal which is the power of some model of  $\varphi$ . Then  $\lambda$  clearly is the sup of these cardinals. (By the way, existence proofs such as these may rely heavily on ZF's substitution-axiom. Cf. [Barwise, 1972].)

How large are these numbers pertaining to second-order logic? From Section 2.2.6 it follows, that the first inaccessible (if it exists) can be second-order characterized; thus the Löwenheim and Hanf numbers are at least bigger still. By similar reasoning, they are not smaller than the second, third, ... inaccessible. And we can go on to larger cardinals; for instance, they must be larger than the first *measurable*. The reason is mainly that, like inaccessibility, defining measurability of  $\kappa$  only needs reference to sets of rank not much higher than  $\kappa$ . (In fact, inaccessibility of  $\kappa$  is a first-order property of  $\langle V_{\kappa+1}, \in \rangle$ ; measurability one of  $\langle V_{\kappa+2}, \in \rangle$ .) Only when large cardinal properties refer in an essential way to the *whole* set theoretic universe (the simplest example being that of *strong compactness*) can matters possibly change. Thus, [Magidor, 1971] proves that the Löwenheim number of universal second-order sentences (and hence, by 4.3, of higher-order logic in general) is less than the first *supercompact* cardinal.

As these matters do bring us a little far afield (after all, this is a handbook of *philosophical* logic) we stop here.

In this light, the recommendation in the last problem of the famous list 'Open problems in classical model theory' in Chang and Keisler [1973] remains as problematic as ever: 'Develop the model theory of second and higher-order logic'.

Additional evidence for the view that second-order logic (and, a fortiori, higher-order logic in general) is not so much logic as set theory, is provided by looking directly at existing set-theoretic problems in second-order terms.

Let  $\kappa$  be the first inaccessible cardinal. In Section 2.2.6 we have seen that every  $ZF^2$  model contains (embeds)  $\langle V_\kappa, \in \rangle$ . As this portion is certainly (first-order) definable in all  $ZF^2$  models in a uniform way,  $ZF^2$  decides every set theoretic problem that mentions sets in  $V_\kappa$  only. This observation has led Kreisel to recommend this theory to our lively attention, so let us continue.

Indeed, already far below  $\kappa$ , interesting questions live. Foremost is the *continuum problem*, which asks whether there are sets of reals in cardinality strictly between  $\mathbb{N}$  and  $\mathbb{R}$ . (Cantor's famous *continuum hypothesis* (CH) says there are not.) Thus,  $ZF^2$  decides CH: either it or its negation follows logically from  $ZF^2$ . Since  $ZF^2$  is correct, in the former case CH is true, while it is false in the latter. But of course, this reduction of the continuum problem to second-order truth really begs the question and is of no help whatsoever.

It does refute an analogy, however, which is often drawn between the continuum hypothesis and the Euclidean postulate of parallels in geometry.

For, the latter axiom is not decided by second-order geometry. Its independence is of a different nature; there are different ‘correct’ geometries, but only one correct set theory (modulo the addition of large cardinal axioms):  $ZF^2$ . (In view of Section 2.2.6, a better formal analogy would be that between the parallel postulate and the existence of inaccessibles — though it has shortcomings as well.)

Another example of a set-theoretic question deep down in the universe is whether there are non-constructible reals. This question occurs at a level so low that, using a certain amount of coding, it can be formulated already in the language of  $P^2$ .

$ZF^2$  knows the answers — unfortunately, we’re not able to figure out exactly what it knows.

So, what is the practical use of second-order set theory? To be true, there are *some* things we do know  $ZF^2$  proves while  $ZF^1$  does not; for instance, the fact that  $ZF^1$  is consistent. Such metamathematical gains are hardly encouraging, however, and indeed we can reasonably argue that there is no way of knowing something to follow from  $ZF^2$  unless it is provable in the two-sorted set/class theory of Morse-Mostowski, a theory that doesn’t have many advantages over its subtheory  $ZF = ZF^1$ . (In terms of Section 4.2 below, Morse-Mostowski can be described as  $ZF^2$  under the general-models interpretation with full comprehension-axioms added.)

We finally mention that sometimes, higher-order notions find application in the theory of sets. In Myhill and Scott [1971] it is shown that the class of hereditarily ordinal-definable sets can be obtained by iterating second-order (or general higher-order) definability through the ordinals. (The constructible sets are obtained by iterating first-order definability; they satisfy the  $ZF$ -axioms only by virtue of their first-order character.) Also, interesting classes of large cardinals can be obtained by their reflecting higher-order properties; cf. for instance [Drake, 1974, Chapter 9].

### 2.5 Special Classes: $\Sigma_1^1$ and $\Pi_1^1$

In the light of the above considerations, the scarcity of results forming a subject of ‘second-order logic’ becomes understandable. (A little) more can be said, however, for certain fragments of the second-order language. Thus, in Section 2.3, the monadic quantificational part was considered, to which belong, e.g. second-order Peano arithmetic  $P^2$  and Zermelo-Fraenkel set theory  $ZF^2$ . The more fruitful restriction for general model-theoretic purposes employs quantificational pattern complexity, however. We will consider the two simplest cases here, viz. prenex forms with only existential second-order quantifiers ( $\Sigma_1^1$  formulas) or only universal quantifiers ( $\Pi_1^1$  formulas). For the full prenex hierarchy, cf. Section 3.2; note however that we restrict the discussion here to formulas all of whose free variables are first-order. One useful shift in perspective, made possible by the present restricted language,



is the following.

If  $\exists X_1, \dots, \exists X_k \varphi$  is a  $\Sigma_1^1$  formula in a vocabulary  $L$ , we sometimes consider  $\varphi$  as a first-order formula in the vocabulary  $L \cup \{X_1, \dots, X_k\}$  — now suddenly looking upon the  $X_1, \dots, X_k$  not as second-order *variables* but as non-logical *constants* of the extended language. Conversely, if  $\varphi$  is a first-order  $L$  formula containing a relational symbol  $R$ , we may consider  $\exists R \varphi$  as a  $\Sigma_1^1$  formula of  $L - \{R\}$  — viewing  $R$  now as a second-order *variable*. As a matter of fact, this way of putting things has been used already (in Section 2.4).

*2.5.1 Showing Things to be  $\Sigma_1^1$  or  $\Pi_1^1$ .* Most examples of second-order formulas given in Section 2.2 were either  $\Sigma_1^1$  or  $\Pi_1^1$ ; in most cases, it was not too hard to translate the given notion into second-order terms.

A simple result is given in Section 3.2 which may be used in showing things to be  $\Sigma_1^1$  or  $\Pi_1^1$ -expressible: any formula obtained from a  $\Sigma_1^1$  ( $\Pi_1^1$ ) formula by prefixing a series of first-order quantifications still has a  $\Sigma_1^1$  ( $\Pi_1^1$ ) equivalent.

For more intricate results, we refer to Kleene [1952] and Barwise [1975]. The first shows that if  $\phi$  is a recursive set of first-order formulas, the infinitary conjunct  $\bigwedge \phi$  has a  $\Sigma_1^1$  equivalent (on infinite models). Thus,  $\exists X_1, \dots, \exists X_k \bigwedge \phi$  is also  $\Sigma_1^1$ . This fact has some relevance to resplendency, cf. Section 2.5.4 below. Kleene's method of proof uses absoluteness of definitions of recursive sets, coding of satisfaction and the integer structure on arbitrary infinite models. (It is implicit in much of Barwise [1975, Chapter IV 2/3], which shows that we are allowed to refer to integers in certain ways when defining  $\Sigma_1^1$  and  $\Pi_1^1$  notions.)

We now consider these concepts one by one.

*2.5.2  $\Sigma_1^1$ -sentences.* The key quantifier combination in Frege's predicate logic expresses dependencies beyond the resources of traditional logic:  $\forall \exists$ . This dependency may be made explicit using a  $\Sigma_1^1$  formula:

$$\forall x \exists y \varphi(x, y) \leftrightarrow \exists f \forall x \varphi(x, f(x)).$$

This introduction of so-called *Skolem functions* is one prime source of  $\Sigma_1^1$  statements. The quantification over functions here may be reduced to our predicate format as follows:

$$\exists X (\forall x y z (X(x, y) \wedge X(x, z) \rightarrow y = z) \wedge \forall x \exists y (X(x, y) \wedge \varphi(x, y))).$$

Even at this innocent level, the connection with set theory shows up (Section 2.4): the above equivalence itself amounts to the assumption of the Axiom of Choice (Bernays).

Through the above equivalence, all first-order sentences may be brought into 'Skolem normal form'. E.g.,  $\forall x \exists y \forall z \exists u A(x, y, z, u)$  goes to  $\exists f \forall x \forall z \exists u A(x, f(x), z, u)$ , and thence to  $\exists f \exists g \forall x \forall z A(x, f(x), z, g(x, z))$ . For

another type of Skolem normal form (using relations instead), cf. [Barwise, 1975, Chapter V 8.6].

Conversely,  $\Sigma_1^1$  sentences allow for many other patterns of dependency. For instance, the variant  $\exists f \exists g \forall x \forall z A(x, f(x), z, g(z))$ , with  $g$  only dependent on  $z$ , is not equivalent to any first-order formula, but rather to a so-called ‘branching’ pattern (first studied in [Henkin, 1961])

$$\left( \begin{array}{l} \forall x \exists y \\ \forall z \exists u \end{array} \right) A(x, y, z, u).$$

For a discussion of the linguistic significance of these ‘branching quantifiers’, cf. [Barwise, 1979]. One sentence which has been claimed to exhibit the above pattern is ‘some relative of each villager and some relative of each townsman hate each other’ (Hintikka). The most convincing examples of first-order branching to date, however, rather concern quantifiers such as (precisely) *one* or *no*. Thus, ‘one relative of each villager and one relative of each townsman hate each other’ seems to lack any linear reading. (The reason is that any linear sequence of *precisely one*’s creates undesired dependencies. In this connection, recall that ‘one sailor has discovered one sea’ is not equivalent to ‘one sea has been discovered by one sailor’.) An even simpler example might be ‘no one loves no one’, which has a linear reading  $\neg \exists x \neg \exists y L(x, y)$  (i.e. everyone loves someone), but also a branching reading amounting to  $\neg \exists x \exists y L(x, y)$ . (Curiously, it seems to lack the inverse scope reading  $\neg \exists y \neg \exists x L(x, y)$  predicted by Montague Grammar.) Actually, this last example also shows that the phenomenon of branching does not lead inevitably to second-order readings.

The preceding digression has illustrated the delicacy of the issue whether second-order quantification actually occurs in natural language. In any case, if branching quantifiers occur, then the logic of natural language would be extremely complex, because of the following two facts. As Enderton [1970] observes, universal validity of  $\Sigma_1^1$  statements may be effectively reduced to that of branching statements. Thus, the complexity of the latter notion is at least that of the former. And, by inspection of the argument in Section 2.3 above, we see that

**THEOREM.** *Universal validity of  $\Sigma_1^1$ -sentences is non-arithmetical, etc.*

**Proof.** The reduction formula was of the form  $P^2 \rightarrow \varphi$ , where  $P^2$  is  $\Pi_1^1$  and  $\varphi$  is first-order. By the usual prenex operation, the universal second-order quantifier in the antecedent becomes an existential one in front. ■

Indeed, as will be shown in Section 4.3, the complexity of  $\Sigma_1^1$ -universal validity is essentially that of universal validity for the whole second-order (or higher-order) language. Nevertheless, one observation is in order here.

These results require the availability of non-logical constants and, e.g. universal validity of  $\exists X \varphi(X, R)$  really amounts to universal validity of the

$\Pi_2^1$ -statement  $\forall Y \exists X \varphi(X, Y)$ . When attention is restricted to ‘pure’ cases, it may be shown that universal validity of  $\Sigma_1^1$  statements is much less complex, amounting to truth in all finite models (cf. [van Benthem, 1977]). Thus, in the arithmetical hierarchy (cf. Section 3.2.) its complexity is only  $\Pi_1^0$ .

When is a  $\Sigma_1^1$  sentence, say of the form  $\exists X_1, \dots, \exists X_k \varphi(X_1, \dots, X_k, R)$ , equivalent to a first-order statement about its parameter  $R$ ? An answer follows from Keisler’s theorem (Section 1.4.2), by the following observation.

**THEOREM.** *Truth of  $\Sigma_1^1$  sentences is preserved under the formation of ultra-products.*

(This is a trivial corollary of the preservation of first-order sentences, cf. Section 1.4.2.)

**COROLLARY.** *A  $\Sigma_1^1$  sentence is first-order definable iff its negation is preserved under ultraproducts.*

(That  $\Sigma_1^1$  sentences, and indeed all higher-order sentences are preserved under isomorphism should be clear.)

Moreover, there is a consequence analogous to Post’s theorem in recursion theory:

**COROLLARY.** *Properties of models which are both  $\Sigma_1^1$  and  $\Pi_1^1$  are already elementary.*

(Of course, this is also immediate from the interpolation theorem which, in this terminology, says that disjoint  $\Sigma_1^1$  classes can be separated by an elementary class.)

Next, we consider a finer subdivision of  $\Sigma_1^1$  sentences, according to their first-order matrix. The simplest forms are the following ( $\varphi$  quantifier-free):

1.  $(\exists \exists) \exists X_1 \dots \exists X_k \exists y_1 \dots y_m \varphi(X_1, \dots, X_k, y_1, \dots, y_m, R)$
2.  $(\exists \forall) \exists X_1 \dots \exists X_k \forall y_1 \dots y_m \varphi(X_1, \dots, X_k, y_1, \dots, y_m, R)$
3.  $(\exists \forall \exists) \exists X_1 \dots \exists X_k \forall y_1 \dots y_m \exists z_1 \dots z_n \varphi(X_1, \dots, y_1, \dots, z_1, \dots, R)$ .

We quote a few observations from [van Benthem, 1983]:

- all forms (1) have a first-order equivalent,
- all forms (2) are preserved under elementary (first-order) equivalence, and hence are equivalent to some (infinite) disjunction of (infinite) conjunctions of first-order sentences,
- the forms (3) harbour the full complexity of  $\Sigma_1^1$ .

The first assertion follows from its counterpart for  $\Pi_1^1$  sentences, to be stated below. A proof sketch of the second assertion is as follows. If (2) holds in a model  $\mathfrak{A}$ , then so does its first-order matrix  $(2)^*$  in some expansion  $\mathfrak{A}^+$  of  $\mathfrak{A}$ . Now suppose that  $\mathfrak{B}$  is elementarily equivalent to  $\mathfrak{A}$ . By a standard compactness argument,  $(2)^*$  is satisfiable together with the elementary diagram of  $\mathfrak{B}$ , i.e. in some elementary extension of  $\mathfrak{B}$ . But, restricting  $X_1, \dots, X_k$  to  $B$ , a substructure arises giving the same truth values to formulas of the specific form  $(2)^*$ ; and hence we have an expansion of  $\mathfrak{B}$  to a model for  $(2)^*$  — i.e.  $\mathfrak{B}$  satisfies (2).

Finally, the third assertion follows from the earlier Skolem reduction: with proper care, the Skolem normal form of the first-order matrix will add some predicates to  $X_1, \dots, X_k$ , while leaving a first-order prefix of the form  $\forall\exists$ . ■

Lastly, we mention the Svenonius characterization of  $\Sigma_1^1$ -sentences in terms of quantifiers of infinite length. In chapter I.1 an interpretation is mentioned of finite formulas in terms of games. This is a particularly good way of explaining infinite sequences of quantifiers like

$$\forall x_1 \exists y_1 \forall x_2 \exists y_2 \forall x_3 \exists y_3 \dots \varphi(x_1, y_1, x_2, y_2, \dots). \quad (1)$$

Imagine players  $\forall$  and  $\exists$  alternatively picking  $x_1, x_2, \dots$  resp.  $y_1, y_2, \dots$ :  $\exists$  wins iff  $\varphi(x_1, y_1, x_2, \dots)$ . (1) is counted as *true* iff  $\exists$  has a *winning strategy*, i.e. a function telling him how to play, given  $\forall$ 's previous moves, in order to win. Of course, a winning strategy is nothing more than a bunch of Skolem functions.

Now, Svenonius' theorem says that, on countable models, every  $\Sigma_1^1$  sentence is equivalent to one of the form (1) where  $\varphi$  is the conjunction of an (infinite) recursive set of first-order formulas. The theorem is in Svenonius [1965]; for a more accessible exposition, cf. [Barwise, 1975, Chapter VI.6].

*2.5.3  $\Pi_1^1$ -sentences.* Most examples of second-order sentences in Section 2.2 were  $\Pi_1^1$ : full induction, Dedekind completeness, full substitution. Also, our recurrent example *most* belonged to this category — and so do, e.g. the modal formulas of intensional logic (compare van Benthem's chapter on Correspondence theory in Volume 3 of this *Handbook*).

Results about  $\Pi_1^1$  sentences closely parallel those for  $\Sigma_1^1$ . (One notable exception is universal validity, however: that notion is recursively axiomatizable here, for the simple reason that  $\models_2 \forall X \varphi(X, Y)$  iff  $\models_1 \varphi(X, Y)$ .) For instance, we have

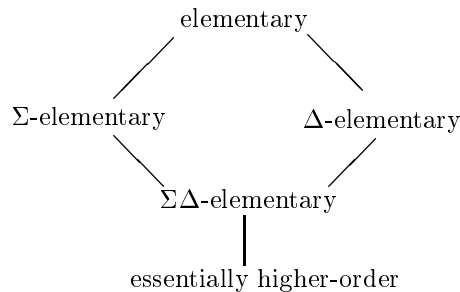
**THEOREM.** *A  $\Pi_1^1$ -sentence has a first-order equivalent iff it is preserved under ultraproducts.*

This time, we shall be little more explicit about various possibilities here. The above theorem refers to *elementary* definitions of  $\Pi_1^1$ -sentences, i.e. in

terms of *single* first-order sentences. The next two more liberal possibilities are  $\Delta$ -*elementary* definitions (allowing an infinite conjunction of first-order sentences) and  $\Sigma$ -*elementary* ones (allowing an infinite disjunction). As was noted in Section 1.2, the non-first-order  $\Pi_1^1$  notion of finiteness is also  $\Sigma$ -elementary: ‘precisely one or precisely two or ...’. The other possibility does not occur, however: all  $\Delta$ -elementary  $\Pi_1^1$  sentences are already elementary. (If the conjunction  $\bigwedge S$  defines  $\forall X_1, \dots, X_n \varphi$ , then the following first-order implication holds:  $S \vDash \varphi(X_1, \dots, X_n)$ . Hence, by compactness  $S_0 \vDash \varphi$  for some finite  $S_0 \subseteq S$  — and  $\bigwedge S_0$  defines  $\forall X_1, \dots, X_n \varphi$  as well.) The next levels in this more liberal hierarchy of first-order definability are  $\Sigma\Delta$  and  $\Delta\Sigma$ . (Unions of intersections and intersections of unions, respectively.) These two, and in fact all putative ‘higher’ ones collapse, by the following observation from Bell and Slomson [1969].

**PROPOSITION.** *A property of models is preserved under elementary equivalence iff it is  $\Sigma\Delta$ -elementary.*

Thus, essentially, there remains a hierarchy of the following form:



Now, by a reasoning similar to the above, we see that  $\Sigma\Delta$ -elementary  $\Pi_1^1$ -sentences are  $\Sigma$ -elementary already. Thus, in the  $\Pi_1^1$ -case, the hierarchy collapses to ‘elementary,  $\Sigma$ -elementary, essentially second-order’. This observation may be connected up with the earlier syntactic classification of  $\Sigma_1^1$ . Using negations, we get for  $\Pi_1^1$ -sentences the types  $\forall\forall(1)$ ,  $\forall\exists(2)$  and  $\forall\exists\forall(3)$ . And these provide precisely instances for each of the above remaining three stages. For instance, that all types (1) are elementary follows from the above characterization theorem, in combination with the observation that type (1)  $\Pi_1^1$ -sentences are preserved under *ultraproducts* (cf. [van Benthem, 1983]).

We may derive another interesting example of failure of first-order model theory here. One of the classical mother results is the Łoś-Tarski theorem: preservation under submodels is syntactically characterized by definability in universal prenex form. But now, consider well-foundedness (Section

2.2.3). This property of models is preserved in passing to submodels, but it cannot even be defined in the universal form (1), lacking first-order definability.

Our final result shows that even this modest, and basic topic of connections between  $\Pi_1^1$  sentences and first-order ones is already fraught with complexity.

**THEOREM.** *The notion of first-order definability for  $\Pi_1^1$ -sentences is not arithmetical.*

**Proof.** Suppose, for the sake of reduction, that it were. We will then derive the arithmetical definability of arithmetical truth — again contradicting Tarski’s theorem. Actually, it is slightly more informative to consider a set-theoretic reformulation (involving only one, binary relation constant): truth in  $\langle V_\omega, \in \rangle$  cannot be arithmetical for first-order sentences  $\psi$ .

Now, consider any categorical  $\Pi_1^1$ -definition  $\Phi$  for  $\langle V_\omega, \in \rangle$ . Truth in  $\langle V_\omega, \in \rangle$  of  $\psi$  then amounts to the implication  $\Phi \vDash_2 \psi$ . It now suffices to show that this statement is effectively equivalent to the following one: ‘ $\Phi \vee \psi$  is first-order definable’. Here, the  $\Pi_1^1$  statement  $\Phi \vee \psi$  is obtained by pulling  $\psi$  into the first-order matrix of  $\Phi$ .

‘ $\implies$ ’: If  $\Phi \vDash_2 \psi$ , then  $\Phi \vee \psi$  is defined by  $\psi$ .

‘ $\impliedby$ ’: Assume that some first-order sentence  $\alpha$  defines  $\Phi \vee \psi$ .

Consider  $\langle V_\omega, \in \rangle$ :  $\Phi$  holds here, and hence so does  $\alpha$ . Now let  $\mathfrak{A}$  be any proper elementary extension of  $\langle V_\omega, \in \rangle$ :  $\Phi$  fails there, while  $\alpha$  still holds. Hence  $(\Phi \vee \psi$  and so)  $\psi$  holds in  $\mathfrak{A}$ . But then,  $\psi$  holds in the elementary submodel  $\langle V_\omega, \in \rangle$ , i.e.  $\Phi \vDash_2 \psi$ . ■

**2.5.4 Resplendent Models.** One tiny corner of ‘higher-order model theory’ deserves some special attention. Models on which  $\Sigma_1^1$  formulas are equivalent with their set of first-order consequences have acquired special interest in model theory. Formally,  $\mathfrak{A}$  is called *resplendent* if for every first-order formula  $\varphi = \varphi(x_1, \dots, x_n)$  in the language of  $\mathfrak{A}$  supplemented with some relation symbol  $R$ :

$$\mathfrak{A} \vDash_2 \forall x_1 \dots x_n (\bigwedge \psi \rightarrow \exists R\varphi),$$

where  $\psi$  is the set of all first-order  $\psi = \psi(x_1, \dots, x_n)$  in the language of  $\mathfrak{A}$  logically implied by  $\varphi$ . Thus,  $\mathfrak{A}$  can be expanded to a model of  $\varphi$  as soon as it satisfies all first-order consequences of  $\varphi$  in its own language.

Resplendency was introduced, in the setting of infinitary admissible languages by Ressayre [1977] under the name *relation-universality*. A discussion of its importance for the first-order case can be found in Barwise and Schlipf [1976]. The notion is closely related to *recursive saturation* (i.e. saturation w.r.t. recursive types of formulas): every resplendent model is

recursively saturated, and, for countable models, the converse obtains as well. In fact, Ressayre was led to (the infinitary version of) this type of saturation by looking at what it takes to prove resplendency.

The importance of resplendent models is derived from the fact that they exist in abundance in all cardinals and can be used to trivialize results in first-order model theory formerly proved by means of saturated and special models of awkward cardinalities. Besides, Ressayre took the applicability of the infinitary notions to great depth, deriving results in descriptive set theory as well. We only mention two easy examples.

**Proof.** (*Craig interpolation theorem*) Suppose that  $\models \varphi(R) \rightarrow \psi(S)$ ; let  $\Phi$  be the set of  $R$ -less consequences of  $\varphi$ . When  $\Phi \models \psi$ , we are finished by one application of compactness. Thus, let  $(\mathfrak{A}, S)$  be a *resplendent* model of  $\Phi$ . By definition, we can expand  $(\mathfrak{A}, S)$  to a model  $(\mathfrak{A}, R, S)$  of  $\varphi$ . Hence,  $(\mathfrak{A}, S) \models \psi$ . But then,  $\Phi \models \psi$ , as *every* model has a resplendent equivalent. ■

As is the case of saturated and special models, many global definability theorems have local companions for resplendent ones. We illustrate this fact again with the interpolation theorem, which in its local version takes the following form: if the resplendent model  $\mathfrak{A}$  satisfies  $\forall x(\exists R\varphi(x) \rightarrow \forall S\psi(x))$ , then there exists a first-order formula  $\eta(x)$  in the  $\mathfrak{A}$ -language such that  $\mathfrak{A}$  satisfies both  $\forall x(\exists R\varphi \rightarrow \eta)$  and  $\forall x(\eta \rightarrow \forall S\psi)$ . To make the proof slightly more perspicuous, we make the statement more symmetrical. Let  $\varphi' = \neg\psi$ . The first sentence then is equivalent with  $\forall x(\neg\exists R\varphi \vee \neg\exists S\varphi')$  (\*), while the last amounts then to  $\forall x(\exists S\varphi' \rightarrow \neg\eta)$ . Hence, interpolation takes the (local) ‘Robinson-consistency’ form: disjoint  $\Sigma_1^1$ -definable sets on  $\mathfrak{A}$  can be separated by a first-order definable one.

Now for the proof, which is a nice co-operation of both resplendency and recursive saturation. Suppose that our resplendent model  $\mathfrak{A} = \langle A, \dots \rangle$  satisfies (\*). By resplendency, the set of logical consequences of either  $\varphi$  or  $\varphi'$  in the language of  $\mathfrak{A}$  is not satisfiable in  $\mathfrak{A}$ .

Applying recursive saturation (the set concerned is only recursively enumerable according to first-order completeness — but we can use Craig’s ‘pleonasm’ trick to get a recursive equivalent), some finite subset  $\Phi \cup \Phi'$  is non-satisfiable already, where we’ve put the  $\varphi$  consequences in  $\Phi$  and the  $\varphi'$  consequences in  $\Phi'$ . We now have  $\models \varphi \rightarrow \bigwedge \Phi$ ,  $\models \varphi' \rightarrow \bigwedge \Phi'$ , and, by choice of  $\Phi \cup \Phi'$ ,  $\mathfrak{A} \models \forall x \neg \bigwedge (\Phi \cup \Phi')$ , which amounts to  $\mathfrak{A} \models \forall x (\neg \bigwedge \Phi \vee \neg \bigwedge \Phi')$ ; hence we may take either  $\bigwedge \Phi$  or  $\bigwedge \Phi'$  as the ‘separating’ formula. The local Beth theorem is an immediate consequence: if the disjoint  $\Sigma_1^1$ -definable sets are each other’s complement, they obviously coincide with the first-order definable separating set and its complement, respectively. In other words, sets which are both  $\Sigma_1^1$  and  $\Pi_1^1$ -definable on  $\mathfrak{A}$  are in fact first-order definable. ■

This situation sharply contrasts with the case for (say)  $\mathfrak{N}$  discussed in Section 3.2, where we mention that arithmetic truth is both  $\Sigma_1^1$  and  $\Pi_1^1$ -definable, but not arithmetical.

### 3 HIGHER-ORDER LOGIC

Once upon the road of second-order quantification, higher predicates come into view. In mathematics, one wants to quantify over functions, but also over functions defined on functions, etcetera. Accordingly, the type theories of the logicist foundational program allowed quantification over objects of all finite orders, as in *Principia Mathematica*. But also natural language offers such an ascending hierarchy, at least in the types of its lexical items. For instance, nouns (such as ‘woman’) denote properties, but then already adjectives become higher-order phrases (‘blond woman’), taking such properties to other properties. In fact, the latter type of motivation has given type theories a new linguistic lease of life, in so-called ‘Montague Grammar’, at a time when their mathematical functions had largely been taken over by ordinary set theory (cf. [Montague, 1974]).

In this section, we will consider a stream-lined relational version of higher-order logic, which leads to the basic logical results with the least amount of effort. Unfortunately for the contemporary semanticist, it does not look very much like the functional Montagovian type theory. In fact, we will not even encounter such modern highlights as lambda-abstraction, because our language can do all this by purely traditional means. Moreover, in Section 4, we shall be able to derive partial completeness results from the standard first-order ones for many-sorted logic in an extremely simple fashion. (In particular, the complicated machinery of [Henkin, 1950] seems unnecessary.) It’s all very elegant, simple, and exasperating. A comparison with the more semantic, categorial grammar-oriented type theories will be given at the end.

#### 3.1 *Syntax and Semantics*

As with first-order languages, higher-order formulas are generated from a given set  $L$  of non-logical constants, among which we can distinguish individual constants, function symbols and relation signs. (Often, we will just think of the latter.) Formulas will be interpreted in the same type of models  $\mathfrak{A} = \langle A, * \rangle$  as used in the first-order case, i.e.  $A \neq \emptyset$ , and  $*$  assigns something appropriate to every  $L$  symbol: (‘distinguished’) elements of  $A$  to individual constants, functions over  $A$  to function symbols (with the proper number of arguments) and relations over  $A$  to relation signs (again, of the proper arity).



Thus, fix any such set  $L$ . Patterns of complexity are now recorded in *types*, defined inductively by

1. 0 is a type
2. a finite sequence of types is again a type.

Here, 0 will be the type of *individuals*,  $(\tau_1, \dots, \tau_n)$  that of *relations* between objects of types  $\tau_1, \dots, \tau_n$ . Notice that, if we read clause (2) as also producing the *empty* sequence, we obtain a type of relation without arguments; i.e. of propositional constants, or *truth values*. Higher up then, we will have propositional functions, etcetera. This possibility will not be employed in the future, as our metatheory would lose some of its elegance. (But see Section 3.3 for a reasonable substitute.)

The *order* of a type is a natural number defined as follows: the order of 0 is 1 (individuals are ‘first-order’ objects), while the order of  $(\tau_1, \dots, \tau_n)$  equals  $1 + \max \text{order}(\tau_i)$  ( $1 \leq i \leq n$ ). Thus, the terminology of ‘first-order’, ‘second-order’, etcetera, now becomes perfectly general.

For each type  $\tau$ , the language has a countably infinite set of variables. The order of a variable is the order of its type. Thus, there is only one kind of first-order variable, because the only order 1-type is 0. The second-order variables all have types  $(0, \dots, 0)$ . Next, the *terms* of type 0 are generated from the type 0 variables and the individual constants by applying function symbols in the proper fashion. A term of type  $\neq 0$  is just a variable of that type. Thus, for convenience, non-logical constants of higher-orders have been omitted: we are really thinking of our former first-order language provided with a quantificational higher-order apparatus. Finally, one might naturally consider a relation symbol with  $n$  places as a term of type  $(0, \dots, 0)$  ( $n$  times); but the resulting language has no additional expressive power, while it becomes a little more complicated. Hence, we refrain from utilising this possibility.

*Atomic formulas* arise as follows:

1.  $R(t_1, \dots, t_n)$  where  $R$  is an  $n$ -place relation symbol and  $t_1, \dots, t_n$  terms of type 0,
2.  $X(t_1, \dots, t_n)$ , where  $X$  is a variable of type  $(\tau_1, \dots, \tau_n)$  and  $t_i$  a type  $\tau_i$ -term ( $1 \leq i \leq n$ ).

We could have added identities  $X = Y$  here for all higher types; but these may be thought of as defined through the scheme  $\forall X_1 \dots \forall X_n (X(X_1, \dots, X_n) \leftrightarrow Y(X_1, \dots, X_n))$ , with appropriate types.

*Formulas* are defined inductively from the atomic ones using propositional connectives and quantification with respect to variables of all types. The resulting set, based on the vocabulary  $L$  is called  $L_\omega$ .  $L_n$  is the set of formulas all of whose variables have order  $\leq n$  ( $n = 1, 2, \dots$ ). Thus,

we can identify  $L_1$  with the first-order formulas over  $L$ , and  $L_2$  with the second-order ones. (A more sophisticated classification of orders is developed in Section 3.2 below, however.) The reader is requested to formulate the examples of Section 2.2 in this language; especially the  $L_3$ -definition of finiteness.

Again, let us notice that we have opted for a rather austere medium: no higher-order constants or identities, no conveniences such as *function quantifiers*, etcetera. One final omission is that of relational *abstracts* taking formulas  $\varphi(X_1, \dots, X_n)$  to terms  $\lambda X_1 \dots X_n \cdot \varphi(X_1, \dots, X_n)$  denoting the corresponding relation. In practice, these commodities do make life a lot easier; but they are usually dispensable in theory. For instance, the statement  $\varphi(\lambda X \cdot \psi(X))$  is equally well expressed by means of  $\exists Y (Y = \lambda X \cdot \psi(X) \wedge \varphi(Y))$ , and this again by  $\exists Y (\forall Z (Y(Z) \leftrightarrow \psi(Z)) \wedge \varphi(Y))$ .

From the syntax of our higher-order language, we now pass on to its semantics. Let  $\mathfrak{A}$  be an ordinary  $L$ -model  $\langle A, * \rangle$  as described above. To interpret the  $L_\omega$ -formulas in  $\mathfrak{A}$  we need the *universes of type  $\tau$  over  $A$*  for all types  $\tau$ :

1.  $D_0(A) = A$
2.  $D_{(\tau_1, \dots, \tau_n)}(A) = \mathcal{P}(D_{\tau_1}(A) \times \dots \times D_{\tau_n}(A))$ .

An  $A$ -assignment is a function  $\alpha$  defined on all variables such that, if  $X$  has type  $\tau$ ,  $\alpha(X) \in D_\tau(A)$ .

We now lift the ordinary satisfaction relation to  $L_\omega$ -formulas  $\varphi$  in the obvious way. For instance, for an  $L$ -model  $\mathfrak{A}$  and an  $A$ -assignment  $\alpha$ ,

$$\mathfrak{A} \models_\omega X(t_1, \dots, t_n)[\alpha] \text{ iff } \alpha(X)(t_1^\mathfrak{A}[\alpha], \dots, t_n^\mathfrak{A}[\alpha]);$$

where  $t^\mathfrak{A}[\alpha]$  is the *value* of the term  $t$  under  $\alpha$  in  $\mathfrak{A}$  defined as usual. Also, e.g.  $\mathfrak{A} \models_\omega \forall X \varphi[\alpha]$  iff for all assignments  $\alpha'$  differing from  $\alpha$  at most in the value given to  $X$ :  $\mathfrak{A} \models_\omega \varphi[\alpha']$ .

The other semantical notions are derived from satisfaction in the usual fashion.

### 3.2 The Prenex Hierarchy of Complexity

The logic and model theory of  $L_\omega$  exhibit the same phenomena as those of  $L_2$ : a fluid border line with set theory, and a few systematic results. Indeed, in a sense, higher-order *logic* does not offer anything new. It will be shown in Section 4.2 that there exists an effective reduction from universal validity for  $L_\omega$  formulas to that for second-order ones, indeed to monadic  $\Sigma_1^1$  formulas [Hintikka, 1955].

As for the connections between  $L_\omega$  and set theory, notice that the present logic is essentially that of arbitrary models  $A$  provided with a natural set-theoretic superstructure  $\bigcup_n V^n(A)$ ; where  $V^0(A) = A$ , and  $V^{n+1}(A) =$

$V^n(A) \cup \mathcal{P}(V^n(A))$ . As a ‘working logic’, this is a sufficient setting for many mathematical purposes. (But cf. [Barwise, 1975] for a smaller, *constructible* hierarchy over models, with a far more elegant metatheory.)

We will not go into the exact relations between the logic of  $L_\omega$  and ordinary set theory, but for the following remark.

Given a structure  $\mathfrak{A} = \langle A, * \rangle$ , the structure  $\mathfrak{A}^+ = \langle \bigcup_n V^n(A), \in, * \rangle$  is a model for a set theory with atoms. There is an obvious translation from the  $L_\omega$ -theory of  $\mathfrak{A}$  into a fragment of the ordinary first-order theory of  $\mathfrak{A}^+$ . The reader may care to speculate about a converse (cf. [Kemeny, 1950]).

What will be considered instead in this section, is one new topic which is typical for a hierarchical language such as the present one. We develop a prenex classification of formulas, according to their patterns of complexity; first in general, then on a specific model, viz. the natural numbers. This is one of the few areas where a coherent body of higher-order theory has so far been developed.

There exists a standard classification of first-order formulas in prenex form.  $\Sigma_0 = \Pi_0$  is the class of quantifier-free formulas;  $\Sigma_{m+1}$  is the class of formulas  $\exists x_1 \dots \exists x_k \varphi$  where  $\varphi \in \Pi_m$ ; and dually,  $\Pi_{m+1}$  is the class of formulas  $\forall x_1 \dots \forall x_k \varphi$  with  $\varphi \in \Sigma_m$ . The well-known Prenex Normal Form Theorem now says that every first-order formula is logically equivalent to one in  $\bigcup_m (\Sigma_m \cup \Pi_m)$ ; i.e. to one in prenex form.

The above may be generalized to arbitrary higher-order formulas as follows. We classify quantificational complexity with respect to the  $n + 1$ st order.  $\Sigma_0^n = \Pi_0^n$  is the class of  $L_\omega$ -formulas all of whose quantified variables have order  $\leq n$ . Thus,  $\Sigma_0^0$  is the class of quantifier-free  $L_\omega$ -formulas. (Notice that the above  $\Sigma_0$  is a proper subclass of  $\Sigma_0^0$ , as we allow free variables of higher type in  $\Sigma_0^0$  formulas. Also, it is not true that  $\Sigma_0^1 \subseteq L_1$ , or even  $\Sigma_0^1 \subseteq L_n$  for some  $n > 1$ .)

Next,  $\Sigma_{m+1}^n$  is the class of formulas  $\exists X_1 \dots \exists X_k \varphi$ , where  $\varphi \in \Pi_m^n$  and  $X_1, \dots, X_k$  have order  $n + 1$ ; and dually,  $\Pi_{m+1}^n$  consists of the formulas  $\forall X_1 \dots \forall X_k \varphi$  with  $\varphi \in \Sigma_m^n$  and  $X_1, \dots, X_k$  ( $n + 1$ )st order. (Notice the peculiar, but well-established use of the upper index  $n$ : a  $\Sigma_{\frac{1}{2}}^1$  formula thus has quantified *second-order* variables.)

The reader may wonder why we did not just take  $\Sigma_0^n$  to be  $L_n$ . The reason is that we do not consider the mere occurrence of, say, second-order variables in a formula a reason to call it (at least) second-order. (Likewise, we do not call first-order formulas ‘second-order’ ones, because of the occurrence of second-order relational constants.) It is *quantification* that counts: we take a formula to be of order  $n$  when its interpretation in a model  $\langle A, \dots \rangle$  presupposes complete knowledge about some  $n$ th order universe  $D_\tau(A)$  over  $A$ . And it is the quantifier over some order  $n$  variable which presupposes such knowledge, not the mere presence of free variables of that order. (After all, we want to call, e.g. a property of type  $((0))$  ‘first-order’ definable, even if its first-order definition contains a second-order free variable — and it

must.) There is an interesting historical analogy here. One way to think of the prenex hierarchy is as one of *definitional complexity*, superimposed upon one of *argument type complexity* (given by the free variable pattern of a formula). This move is reminiscent of Russell's passage from *ordinary* to *ramified* type theory.

**THEOREM.** *Every  $L_{n+1}$ -formula has an equivalent in  $\bigcup_m (\Sigma_m^n \cup \Pi_m^n)$ .*

**Proof.** Let  $\varphi \in L_{n+1}$  be given. First, manipulate it into prenex form, where the order of the quantifiers is immaterial — just as in the first-order case. Now, if we can manage to get quantifiers over  $n+1$ st order variables to the front, we are done. But, this follows by repeated use of the valid equivalence below and its dual.

Let  $x$  have type  $\tau_0$  and order less than  $n+1$ : the order of the type  $(\tau_1, \dots, \tau_k)$  of the variable  $X$ . Let  $Y$  be some type  $(\tau_0, \dots, \tau_k)$  variable; its order is then  $n+1$  too, and we have the equivalence

$$\forall x \exists X \psi \leftrightarrow \exists Y \forall x \psi'.$$

Here  $\psi'$  is obtained from  $\psi$  by replacing subformulas  $X(t_1, \dots, t_k)$  by  $Y(x, t_1, \dots, t_k)$ ; where  $Y$  does not occur in  $\psi$ . Thanks to the restriction to  $L_{n+1}$ , the *only* atomic subformulas of  $\psi$  containing  $X$  are of the above form and, hence,  $\psi'$  does not contain  $X$  any longer. (If  $X$  could occur in argument positions, it would have to be defined away using suitable  $Y$  *abstracts*. But, this addition to the language would bring about a revised account of complexity in any case.)

To show intuitively that the above equivalence is valid, assume that  $\forall x \exists X \psi(x, X)$ . For every  $x$ , choose  $X_x$  such that  $\psi(x, X_x)$ . Define  $Y$  by setting  $Y(x, y_1, \dots, y_n) := X_x(y_1, \dots, y_n)$ . Then clearly  $\forall x \psi(x, \{y_1, \dots, y_n\} \mid Y(x, y_1, \dots, y_n))$  and, hence,  $\exists Y \forall x \psi'$ . The converse is immediate. ■

We will now pass on to more concrete hierarchies of higher-order definable relations on specific models.

Let  $\mathfrak{A} = \langle A, \dots \rangle$  be some model,  $R \in D_\tau(A)$ ,  $\tau = (\tau_1, \dots, \tau_n)$ , and let  $\varphi \in L_\omega$  have free variables  $X_1, \dots, X_n$  of types (respectively)  $\tau_1, \dots, \tau_n$ .  $\varphi$  is said to *define*  $R$  on  $\mathfrak{A}$  if, whenever  $S_1 \in D_{\tau_1}(A), \dots, S_n \in D_{\tau_n}(A)$ ,

$$R(S_1, \dots, S_n) \text{ iff } \mathfrak{A} \models_\omega \varphi[S_1, \dots, S_n].$$

$R$  is called  $\Sigma_m^n(\Pi_m^n)$  on  $\mathfrak{A}$  if it has a defining formula of this kind. It is  $\Delta_m^n$  if it is both  $\Sigma_m^n$  and  $\Pi_m^n$ . We denote these classes of definable relations on  $\mathfrak{A}$  by  $\Sigma_m^n(\mathfrak{A})$ , etcetera.

Now, let us restrict attention to  $\mathfrak{A} =$  the natural numbers  $\mathfrak{N} : \langle \mathbb{N}, +, \times, 0 \rangle$ . (In this particular case it is customary to let  $\Sigma_0^0(\mathfrak{N}) = \Pi_0^0(\mathfrak{N})$  be the wider class of relations definable using formulas in which *restricted* quantification

over first-class variables is allowed.) For any type  $\tau$ ,  $\Delta_1^0(\mathfrak{N}) \cap D_\tau(\mathbb{N})$  is the class of *recursive* relations of type  $\tau$ ; the ones in  $\Sigma_1^0(\mathfrak{N}) \cap D_\tau(\mathbb{N})$  are called *recursively enumerable*. These are the simplest cases of the *arithmetical hierarchy*, consisting of all  $\Sigma_n^0$  and  $\Pi_n^0$ -definable relations on  $\mathfrak{N}$ . Evidently, these are precisely the first-order-definable ones, in any type  $\tau$ .

At the next level, the *analytic hierarchy* consists of the  $\Sigma_n^1$  and  $\Pi_n^1$ -definable relations on  $\mathfrak{N}$ . Those in  $\Delta_1^1(\mathfrak{N})$  are called *hyperarithmetical*, and have a (transfinite) hierarchy of their own. One reason for the special interest in this class is the fact that *arithmetic truth* for first-order sentences is hyper-arithmetical (though not arithmetical, by Tarski's Theorem).

These hierarchies developed after the notion of recursiveness had been identified by Gödel, Turing and Church, and were studied in the fifties by Kleene, Mostowski and others.

Just to give an impression of the more concrete type of investigation in this area, we mention a few results. Methods of proof are rather uniform: positive results (e.g. ' $\varphi \in \Sigma_n^1$ ') by actual inspection of possible definitions, negative results (' $\varphi \notin \Sigma_n^1$ ') by diagonal arguments reminiscent of the mother example in Russell's Paradox.

1. The satisfaction predicate 'the sequence (coded by)  $s$  satisfies the first-order formula (coded by)  $\varphi$  in  $\mathfrak{N}$ ' is in  $\Delta_1^1(\mathfrak{N}) \cap D_{(0,0)}(\mathbb{N})$ .
2. This predicate is not in  $\Sigma_0^1(\mathfrak{N})$ .
3. The *Analytic Hierarchy Theorem* for  $D_{(0)}(\mathbb{N})$  relations. All inclusions in the following scheme are proper (for all  $m$ ):

$$\begin{array}{ccccc}
 & & \Sigma_m^1(\mathfrak{N}) & & \\
 & \subseteq & & \subseteq & \\
 \Delta_m^1(\mathfrak{N}) & & & & \Delta_{m+1}^1(\mathfrak{N}) \\
 & \subseteq & & \subseteq & \\
 & & \Pi_m^1(\mathfrak{N}) & & 
 \end{array}$$

These results may be generalized to higher orders.

4. Satisfaction for  $\Sigma_0^n$ -formulas (with first-order free variables only) on  $\mathfrak{N}$  is in  $\Delta_1^n(\mathfrak{N}) - \Sigma_0^n(\mathfrak{N})$ .
5. The Hierarchy Theorem holds in fact for any upper index  $\geq 1$ .

By allowing second-order parameters in the defining formulas, the analytic hierarchy is transformed into the classical hierarchy of *projective* relations. Stifled in set-theoretic difficulties around the twenties, interest in this theory was revived by the set-theoretic revolution of the sixties. The reader is referred to the modern exposition [Moschovakis, 1980].

### 3.3 *Two Faces of Type Theory*

As was observed earlier, the above language  $L_\omega$  is one elegant medium of description for one natural type superstructure on models with relations. Nevertheless, there is another perspective, leading to a more function-oriented type theory closer to the categorial system of natural language. In a sense, the two are equivalent through codings of functions as special relations, or of relations through characteristic functions. It is this kind of *sous entendu* which would allow an ordinary logic text book to suppress all reference to functional type theories in the spirit of [Church, 1940; Henkin, 1950] or [Montague, 1974]. (It is this juggling with codings and equivalences also, which makes advanced logic texts so impenetrable to the outsider lacking that frame of mind.)

For this reason, we give the outline of a functional type theory, comparing it with the above. As was observed earlier on, in a first approximation, the existential part of natural language can be described on the model of a *categorial grammar*, with basic *entity expressions* (e.g. proper names; type  $e$ ) and *truth value expressions* (sentences; type  $t$ ), allowing arbitrary binary couplings  $(a, b)$ : the type of functional expressions taking an  $a$ -type expression to a  $b$ -type one. Thus, for instance, the intransitive verb ‘walk’ has type  $(e, t)$ , the transitive verb ‘buy’ type  $(e, (e, t))$ , the sentence negation ‘not’ has  $(t, t)$  while sentence conjunction has  $(t, (t, t))$ . More complicated examples are quantifier phrases, such as ‘no man’, with type  $((e, t), t)$ , or determiners, such as ‘no’, with type  $((e, t), ((e, t), t))$ . Again, to a first approximation, there arises the picture of natural language as a huge jigsaw puzzle, in which the interpretable sentences are those for which the types of their component words can be fitted together, step by step, in such a way that the end result for the whole is type  $t$ .

Now, the natural matching type theory has the above types, with a generous supply of variables and constants for each of these. Its basic operations will be, at least, *identity* (between expressions of the same type), yielding truth value expressions, and *functional application* combining  $B$  with type  $(a, b)$  and  $A$  with type  $a$  to form the expression  $B(A)$  of type  $b$ . What about the logical constants? In the present light, these are merely constants of specific categories. Thus, binary connectives (‘and’, ‘or’) are in  $(t, (t, t))$ , quantifiers (‘all’, ‘some’) in the above determiner type  $((e, t), ((e, t), t))$ . (Actually, this makes them into binary relations between properties: a point of view often urged in the logical folklore.) Nevertheless, one can single them out for special treatment, as was Montague’s own strategy. On the other hand, a truly natural feature of natural language seems to be the phenomenon of *abstraction*: from any expression of type  $b$ , we can make a functional one of type  $(a, b)$  by varying some occurrence(s) of component  $a$  expressions. Formally then, our type theory will have so-called ‘lambda abstraction’: if  $B$  is an expression of type  $b$ , and  $x$  a variable of type  $a$ , then

$\lambda x \cdot B$  is an expression of type  $(a, b)$ .

Semantic structures for this language form a function hierarchy as follows:

1.  $D_e$  is some set (of ‘entities’ or ‘individuals’),
2.  $D_t$  is the set of truth values  $\{0, 1\}$  (or some generalization thereof),
3.  $D_{(a,b)} = D_b^{D_a}$

Given a suitable interpretation for constants and assignments for variables, values may be computed for terms of type  $a$  in the proper domain  $D_a$  through the usual compositional procedure. Thus, in particular, suppressing indices,

$$\begin{aligned} \text{val}(B(A)) &= \text{val}(B)(\text{val}(A)) \\ \text{val}(\lambda x \cdot B) &= \lambda a \in D_a \cdot \text{val}(B)_{x \rightarrow a}. \end{aligned}$$

(Just this once, we have refrained from the usual pedantic formulation.)

In Montague’s so-called ‘intensional type theory’, this picture is considerably complicated by the addition of a realm of possible world-times, accompanied by an auxiliary type  $s$  with restricted occurrences. This is a classical example of an unfortunate formalization. Actually, the above set-up remains exactly the same with one additional basic type  $s$  (or two, or ten) with corresponding semantic domains  $D_s$  (all world-times, in Montague’s case). In the terms of [Gallin, 1975]: once we move up from *Ty* to *Ty2*, simplicity is restored.

We return to the simplest case, as all relevant points can be made here. What is the connection with the earlier logic  $L_\omega$ ? Here is the obvious translation, simple in content, a little arduous in combinatorial detail.

First, let us embed the Montague hierarchy of domains  $D_a$  over a given universe  $A$  into our previous hierarchy  $D_\tau(A)$ . In fact, we shall *identify* the  $D_a$  with certain subsets of the  $D_\tau(A)$ . There seems to be one major problem here, viz. what is to correspond to  $D_t = \{0, 1\}$ . (Recall that we opted for an  $L_\omega$ -hierarchy without truth-value types.) We choose to define  $D_t \subseteq D_{(0)}(A) : 0$  becoming  $\emptyset$ , and 1 becoming the whole  $A$ . Next, of course  $D_e = D_0(A)$ . The rule  $D_{(a,b)} = D_b^{D_a}$  then generates the other domains. Thus, every Montague universe  $D_a$  has been identified with a subset of a certain  $D_{\underline{a}}(A)$ ; where  $\underline{a}$  is obviously determined by the rules  $\underline{e} := 0$ ,  $\underline{t} := (0)$  and  $\underline{(a,b)} := (\underline{a}, \underline{b})$ . (Thus, functions have become identified with their graphs; which are binary relations in this case.)

Next, for each Montague type  $a$ , one can write down an  $L_\omega$ -formula  $T_a(x)$  (with  $x$  of type  $\underline{a}$ ) which *defines*  $D_a$  in  $D_{\underline{a}}(A)$ , i.e. for  $b \in D_{\underline{a}}(A)$ ,  $\mathfrak{A} \models_\omega T_a[b]$  iff  $b \in D_a$ .

When  $E = E(x_1, \dots, x_n)$  is any type  $a_0$  expression in the Montague system, with the free variables  $x_1, \dots, x_n$  (with types  $a_1, \dots, a_n$ , respectively) and  $b_i \in D_{a_i}$  ( $1 \leq i \leq n$ ), an object  $E^{\mathfrak{A}}[b_1, \dots, b_n] \in D_{a_0}$  has been defined which is the *value* of  $E$  under  $b_1, \dots, b_n$  in  $\mathfrak{A}$ . We shall indicate now how

to write down an  $L_\omega$ -formula  $V(x_0, E)$  with free variables  $x_0, \dots, x_n$  (where *now*  $x_i$  has type  $\underline{a}_i$  ( $1 < i < n$ )), which says that  $x_0$  is the value of  $E$  under  $x_1, \dots, x_n$ . To be completely precise, we will have

$$\mathfrak{A} \models_\omega V(x_0, E)[b_0, \dots, b_n] \text{ iff } b_0 = E^{\mathfrak{A}}[b_1, \dots, b_n]$$

for objects  $b_0, \dots, b_n$  of the appropriate types.

As a consequence of this, we obtain

$$\mathfrak{A} \models_\omega \exists x(V(x, E_1) \wedge V(x, E_2)) \text{ iff } E_1^{\mathfrak{A}} = E_2^{\mathfrak{A}}$$

for closed expressions  $E_1, E_2$ . Thus, the characteristic assertions of Montagovian type theory have been translated into our higher-order logic.

It remains to be indicated how to construct the desired  $V$ . For perspicuity, three shorthands will be used in  $L_\omega$ . First,  $x(y)$  stands for the unique  $z$  such that  $x(y, z)$ , if it exists. (Elimination is always possible in the standard fashion.) Furthermore, we will always have  $\forall x_1 \dots x_n \exists! x_0 V(x_0, E)$  valid when relativized to the proper types. Therefore, instead of  $V(x_0, E)$ , one may write  $x_0 = V(E)$ . Third, quantifier relativization to  $T_a$  will be expressed by  $\forall x \in T_a (\exists x \in T_a)$  (where  $x$  has type  $\underline{a}$ ). Finally, in agreement with the above definition of the truth values, we abbreviate  $\forall y \in T_e x(y)$  and  $\forall y \in T_e \neg x(y)$  by  $x = \top$ ,  $x = \perp$ , respectively (where  $x$  has type (0)).

Here are the essential cases:

1.  $E$  is a two-place relation symbol of the base vocabulary  $L$ .

$$V(x, E) := x \in T_{(e, (e, t))} \wedge \forall yz \in T_e ((x(y))(z) = \top \leftrightarrow E(y, z)).$$

2.  $E = E_1(E_2)$ .

$$V(x, E) := x = V(E_1)(V(E_2)).$$

3.  $E = \lambda y \cdot F$  ( $y$  of type  $a$ ,  $F$  of type  $b$ ).

$$V(x, E) := x \in T_{(a, b)} \wedge \forall y \in T_a (x(y) = V(F)).$$

4.  $E = (E_1 = E_2)$ .

$$V(x, E) := x \in T_t \wedge (x = \top \leftrightarrow V(E_1) = V(E_2)).$$

That these clauses do their job has to be demonstrated by induction, of course; but this is really obvious.

It should be noted that the procedure as it stands does not handle higher-order constants: but, a generalization is straightforward.

For further details, cf. [Gallin, 1975, Chapter 13]. Gallin also has a converse translation from  $L_\omega$  into functional type theory, not considered here.



The reduction to  $L_\omega$  makes some prominent features of functional type theory disappear. Notably, lambda abstraction is simulated by means of ordinary quantification. It should be mentioned, however, that this also deprives us of some natural and important questions of functional type theory, such as the search for unique *normal forms*. The latter topic will be reviewed briefly at the end of the following Section.

#### 4 REDUCTION TO FIRST-ORDER LOGIC

One weak spot in popular justifications for employing higher-order logic lies precisely in the phrase ‘all predicates’. When we say that Napoleon has all properties of the great generals, we surely mean to refer to some sort of relevant human properties, probably even definable ones. In other words, the lexical item ‘property’ refers to some sort of ‘things’, just like other common nouns. Another, more philosophical illustration of this point is Leibniz’ Principle, quoted earlier, of the identity of indiscernibles. Of course, when  $x, y$  share *all* properties, they will share that of being identical to  $x$  and, hence, they coincide. But this triviality is not what the great German had in mind — witness the charming anecdote about the ladies at court, whom Leibniz made to search for autumn leaves, promising them noticeable differences in colour or shape for any two merely distinct ones.

Thus, there arises the logical idea of re-interpreting second-order, or even higher-order logic as some kind of *many-sorted* first-order logic, with various distinct kinds of objects: a useful, though inessential variation upon first-order logic itself. To be true, properties and predicates are rather abstract kinds of ‘things’; but then, so are many other kinds of ‘individual’ that no one would object to. The semantic net effect of this change in perspective is to allow a greater variety of models for  $L_\omega$ , with essentially smaller ranges of predicates than the original ‘full ones’. Thus, more potential counter-examples become available to universal truths, and the earlier set of  $L_\omega$ -validities decreases; so much so, that we end up with a recursively axiomatizable set. This is the basic content of the celebrated introduction of ‘general models’ in [Henkin, 1950]: the remainder is frills and laces.

##### 4.1 General Models

The type structure  $\langle D_\tau(A) \mid \tau \in \mathcal{T} \rangle$  ( $\mathcal{T}$  the set of types) over a given non-empty set  $A$  as defined in Section 3.1 is called the principal or *full* type structure over  $A$ ; the interpretation of  $L_\omega$  by means of  $\vDash_\omega$  given there the *standard* interpretation. We can generalize these definitions as follows.

$E = \langle E_\tau \mid \tau \in \mathcal{T} \rangle$  is called a *type structure* over  $A$  when

1.  $E_0 = A$  (as before)

$$2. E_{(\tau_1, \dots, \tau_n)} \subseteq \mathcal{P}(E_{\tau_1} \times \dots \times E_{\tau_n}).$$

Thus, not every relation on  $E_{\tau_1} \times \dots \times E_{\tau_n}$  need be in  $E_{(\tau_1, \dots, \tau_n)}$  any more. Restricting assignments to take values in such more general type structures, satisfaction can be defined as before, leading to a notion of truth with respect to arbitrary type structures. This so-called *general models interpretation* of  $L_\omega$  admits of a complete axiomatisation, as we shall see in due course.

First, we need a certain transformation of higher-order logic into first-order terms. Let  $L$  be a given vocabulary.  $L^+$  is the *first-order* language based on the vocabulary

$$L \cup \{\varepsilon_\tau \mid 0 \neq \tau \in \mathcal{T}\} \cup \{T_\tau \mid \tau \in \mathcal{T}\};$$

where  $\varepsilon_\tau$  is an  $n + 1$ ary relation symbol when  $\tau = (\tau_1, \dots, \tau_n)$ , and the  $T_\tau$  are unary relation symbols. Now, define the translation  $^+ : L_\omega \rightarrow L^+$  as follows. Let  $\varphi \in L_\omega$ . First, replace every atom  $X(t_1, \dots, t_n)$  in it by  $\varepsilon_\tau(X, t_1, \dots, t_n)$  when  $X$  has type  $\tau$ . Second, relativize quantification with respect to type  $\tau$  variables to  $T_\tau$ . Third, consider all variables to be (type 0) variables of  $L^+$ . This defines  $\varphi^+$ . (For those familiar with many-sorted thinking (cf. Hodges' chapter, this Volume), the unary predicates  $T_\tau$  may even be omitted, and  $\varphi^+$  just becomes  $\varphi$ , in a many-sorted reading.)

On the model-theoretic level, suppose that  $(\mathfrak{A}, E)$  is a general model for  $L$ ; i.e.  $\mathfrak{A}$  is an  $L$ -model with universe  $A$  and  $E$  is a type structure over  $A$ . We indicate how  $(\mathfrak{A}, E)$  can be transformed into an ordinary (first-order) model  $(\mathfrak{A}, E)^+$  for  $L^+$ :

1. the universe of  $(\mathfrak{A}, E)^+$  is  $\bigcup_{\tau \in \mathcal{T}} E_\tau$
2. the interpretation of  $L$ -symbols is the same as in  $\mathfrak{A}$
3.  $\varepsilon_\tau$  is interpreted by  $(\tau = (\tau_1, \dots, \tau_n))$ :  $\varepsilon_\tau^*(R, S_1, \dots, S_n)$  iff  $R \in E_\tau$ ,  $S_i \in E_{\tau_i}$  ( $1 \leq i \leq n$ ) and  $R(S_1, \dots, S_n)$
4.  $T_\tau$  is interpreted by  $E_\tau$ .

There is a slight problem here. When  $L$  contains function symbols, the corresponding functions in  $\mathfrak{A}$  should be extended on  $\bigcup_{\tau \in \mathcal{T}} E_\tau$ . It is irrelevant how this is done, as arguments outside of  $E_0$  will not be used.

The connection between these transformations is the following

**LEMMA.** *Let  $\alpha$  be an  $E$  assignment, and let  $\varphi \in L_\omega$ . Then  $(\mathfrak{A}, E) \models_\omega \varphi[\alpha]$  iff  $(\mathfrak{A}, E)^+ \models \varphi^+[\alpha]$ .*

The proof is a straightforward induction on  $\varphi$ .

There is semantic drama behind the simple change in clause (2) for  $E_{(\tau_1, \dots, \tau_n)}$  from identity to inclusion. Full type structures are immense; witness their cardinality, which increases exponentially at each level. In

stark contrast, a general model may well have an empty type structure, not ascending beyond the original universe. Evidently, the interesting general models lie somewhere in-between these two extremes.

At least two points of view suggest themselves for picking out special candidates, starting from either boundary.

‘From above to below’, the idea is to preserve as much as possible of the global type structure; i.e. to impose various principles valid in the full model, such as Comprehension or Choice (cf. the end of Section 4.2). In the limit, one might consider general models which are  $L_\omega$ -elementarily equivalent to the full type model. Notice that, by general logic, only  $\Pi_1^1$  truths are automatically preserved in passing from the full model to its general submodels. Such preservation phenomena were already noticed in [Orey, 1959], which contains the conjecture that a higher-order sentence is first-order definable if and only if it has the above persistence property, as well as its converse. (A proof of this assertion is in van Benthem [1977].)

Persistence is of some interest for the semantics of natural language, in that some of its ‘extensional’ fragments translate into persistent fragments of higher-order logic (cf. [Gallin, 1975, Chapter 1.4]). Although the main observation (due to Kamp and Montague) is a little beyond the resources of our austere  $L_\omega$ , it may be stated quite simply. Existential statements  $\exists X A(X)$  may be lost in passing from full standard models to their general variants (cf. the example given below). But, *restricted* existential statements  $\exists X (P(X, Y) \wedge A(X))$  with all their parameters (i.e.  $P(!)$ ,  $Y$ ) in the relevant general model, are thus preserved — and the above-mentioned extensional fragments of natural language translate into these restricted forms, which are insensitive, in a sense, to the difference between a general model and its full parent. Therefore, the completeness of  $L_\omega$  with respect to the general models interpretation (Section 4.2) extends to these fragments of natural language, despite their *prima facie* higher-order nature.

Conversely, one may also look ‘from below to above’, considering reasonable constructions for filling the type universes without the above explosive features. For instance, already in the particular case of  $L_2$ , a natural idea is to consider predicate ranges consisting of all predicates *first-order definable* in the base vocabulary (possibly with individual parameters). Notice that this choice is stable, in the sense that iteration of the construction (plugging in newly defined predicates into first-order definitions) does not yield anything new. (By the way, the simplest proof that, e.g. von Neumann-Bernays-Gödel set theory is conservative over ZF uses exactly this construction.)

EXAMPLE. The first-order definable sets on the base model  $\langle \mathbb{N}, < \rangle$  are precisely all finite and co-finite ones; and a similar characterization may be given for arbitrary predicates. This general model for  $L_2$  is not elementarily equivalent to the standard model, however, as it fails to validate

$$\exists X \forall y ((\exists z (X(z) \wedge y < z) \wedge \exists z (\neg X(z) \wedge y < z)).$$

Second-order general models obtained in this way only satisfy the so-called ‘predicative’ comprehension axioms. (Referring to the end of Section 4.2, these are the sentences (1) where  $\varphi$  does not *quantify* over second-order variables, but may contain them freely.) We can, however, obtain general models of full (‘impredicative’) comprehension if we iterate the procedure as follows. For any second-order general model  $\langle \mathfrak{A}, E \rangle$ , let  $E^+$  consist of all relations on  $\mathfrak{A}$  parametrically second-order definable in  $\langle \mathfrak{A}, E \rangle$ . Thus, the above ‘predicative’ extension is just  $\langle \mathfrak{A}, \emptyset^+ \rangle$ . This time, define  $E_\alpha$  for ordinals  $\alpha$  by  $E_\alpha = \bigcup_{\beta < \alpha} E_\beta^+$ . By cardinality considerations, the hierarchy must stop at some  $\gamma$  (by first-order Löwenheim–Skolem, it can in fact be proved that  $\gamma$  has the same cardinal as  $\mathfrak{A}$ ), which obviously means that  $\langle \mathfrak{A}, E_\gamma \rangle$  satisfies full comprehension.

For  $\mathfrak{A} = \mathfrak{N}$ , the above transfinite hierarchy is called *ramified analysis*,  $\gamma$  is Church-Kleene  $\omega_1$  and there is an extensive literature on the subject. Barwise [1975] studies related things in a more set-theory oriented setting for arbitrary models.

#### 4.2 General Completeness

As a necessary preliminary to a completeness theorem for  $L_\omega$  with its new semantics, we may ask which  $L^+$ -sentences hold in every model of the form  $\langle \mathfrak{A}, E \rangle^+$ , where  $\mathfrak{A}$  is an  $L$ -model and  $E$  a type structure over its universe  $A$ . As it happens, these are of six kinds.

1.  $\exists x T_0 x$ . This is because  $T_0$  is interpreted by  $E_0 = A$ , which is not empty. The other type levels of  $E$  might indeed be empty, if  $E$  is not full.
2. The next sentences express the fact that the  $L$ -symbols stand for distinguished elements, functions and relations over the set denoted by  $T_0$ :
  - (a)  $T_0(c)$ , for each individual constant of  $L$ .
  - (b)  $\forall x_1 \dots \forall x_n (T_0(x_1) \wedge \dots \wedge T_0(x_n) \rightarrow T_0(F(x_1, \dots, x_n)))$ , for all  $n$ -place function symbols  $F$  of  $L$ .
  - (c)  $\forall x_1 \dots \forall x_n (R(x_1, \dots, x_n) \rightarrow T_0(x_1) \wedge \dots \wedge T_0(x_n))$ , for all  $n$ -place relation symbols  $R$  of  $L$ .

Finally, there are sentences about the type levels.

3.  $\forall x (T_\tau(x) \rightarrow \neg T_{\tau'}(x))$ , whenever  $\tau \neq \tau'$ .

As a matter of fact, there is a small problem here. If  $A$  has elements which are sets, then we might have simultaneously  $a \in A$  and  $a \subseteq A$ . It could happen then that  $a \in E_{(0)}$  also, and hence  $E_0 \cap E_{(0)} \neq \emptyset$ . To avoid inessential sophistries, we shall resolutely ignore these eventualities.

4.  $\forall x \bigvee_{\tau \in \mathcal{T}} T_\tau(x)$ .

The content of this statement is clear; but unfortunately, it is not a first-order sentence of  $L^+$ , having an infinite disjunction. We shall circumvent this problem eventually.

5.  $\forall x \forall y_1 \dots \forall y_n (\varepsilon_\tau(x, y_1, \dots, y_n) \rightarrow T_\tau(x) \wedge T_{\tau_1}(y_1) \wedge \dots \wedge T_{\tau_n}(y_n))$ , whenever  $\tau = (\tau_1, \dots, \tau_n)$ . (Compare the earlier definition of  $(\mathfrak{A}, E)^+$ : especially the role of  $\varepsilon_\tau^*$ .)

The sentences (1)–(5) are all rather trivial constraints on the type framework. The following *extensionality axioms* may be more interesting:

6.  $\forall x \forall y (T_\tau(x) \wedge T_\tau(y) \wedge \forall z_1 \dots \forall z_n (T_{\tau_1}(z_1) \wedge \dots \wedge T_{\tau_n}(z_n) \rightarrow (\varepsilon_\tau(x, z_1, \dots, z_n) \leftrightarrow \varepsilon_\tau(y, z_1, \dots, z_n))) \rightarrow x = y)$ ; whenever  $\tau = (\tau_1, \dots, \tau_n)$ .

That this holds in  $(\mathfrak{A}, E)^+$  when  $E$  is full, is due to the extensionality axiom of set theory. But it is also easily checked for general type structures.

This exhausts the obvious validities. Now, we can ask whether, conversely, every  $L^+$  model of (1)–(6) is of the form  $(\mathfrak{A}, E)^+$ , at least, up to isomorphism. (Otherwise, trivial counter-examples could be given.) The answer is positive, by an elementary argument. For any  $L^+$ -model  $\mathfrak{B}$  of our six principles, we may construct a general model  $(\mathfrak{A}, E)$  and an isomorphism  $h : \mathfrak{B} \rightarrow (\mathfrak{A}, E)^+$  as follows.

Writing  $h_\tau := h \upharpoonright T_\tau^{\mathfrak{B}}$ , we shall construct  $h_\tau$  and  $E_\tau$  simultaneously by induction on the order of  $\tau$ , relying heavily on (6). (This construction is really a particular case of the Mostowski collapsing lemma in set theory.)

First, let  $A = E_0 := T_0^{\mathfrak{B}}$ , while  $h_0$  is the identity of  $T_0^{\mathfrak{B}}$ . (1) says that  $A \neq \emptyset$ , and (2) adds that we can define  $\mathfrak{A}$  by taking over the interpretations that  $\mathfrak{B}$  gave to the  $L$ -symbols. Trivially then,  $h_0$  preserves  $L$ -structure. Next, suppose  $\tau = (\tau_1, \dots, \tau_n)$ , where  $E_{\tau_i}, h_{\tau_i}$  ( $1 \leq i \leq n$ ) have been constructed already. Define  $h_\tau$  on  $T_\tau^{\mathfrak{B}}$  by setting

$$h_\tau(b) := \{(h_{\tau_1}(a_1), \dots, h_{\tau_n}(a_n)) \mid \varepsilon_\tau^{\mathfrak{B}}(b, a_1, \dots, a_n)\}$$

(by (5), this stipulation makes sense); putting  $E_\tau := h_\tau[T_\tau^{\mathfrak{B}}]$ . Clearly,  $E_\tau \subseteq \mathcal{P}(E_{\tau_1} \times \dots \times E_{\tau_n})$ . We are finished if it can be shown that  $h_\tau$  is one-one, while  $\varepsilon_\tau^{\mathfrak{B}}(b, a_1, \dots, a_n)$  iff  $h_\tau(b)(h_{\tau_1}(a_1), \dots, h_{\tau_n}(a_n))$ . But, the first assertion is immediate from (6), and it implies the second. Finally, put  $h := \bigcup_{\tau \in \mathcal{T}} h_\tau$ . (3) is our licence to do this. That  $h$  is defined on all of  $\mathfrak{B}$  is implied by (4).

The previous observations yield a conclusion:

LEMMA. *An  $L_\omega$ -sentence  $\varphi$  is true in all general models if its translation  $\varphi^+$  logically follows from (1)–(6) above.*

**Proof.** The direction from right to left is immediate from the definition of the translation  $^+$ , and its semantic behaviour. From left to right, we use the above representation. ■

The value of the Lemma is diminished by the fact that (4) has an infinite disjunction, outside of  $L^+$ . But we can do better.

**THEOREM.**  $\varphi \in L_\omega$  is true in all general models iff  $\varphi^+$  follows from (1), (2), (3), (5) and (6).

**Proof.** The first half is as before. Next, assume that  $\varphi$  is true in all general models, and consider any  $L^+$ -model  $\mathfrak{B}$  satisfying the above five principles. Now, its submodel  $\mathfrak{B}^*$  with universe  $\bigcup_{\tau \in \mathcal{T}} T_\tau^{\mathfrak{B}}$  satisfies these principles as well, but in addition, it also verifies (4). Thus, as before,  $\mathfrak{B}^* \models \varphi^+$ . But then, as all quantifiers in  $\varphi^+$  occur restricted to the levels  $T_\tau$ ,  $\mathfrak{B} \models \varphi^+$ , and we are done after all. ■

This theorem effectively reduces  $L_\omega$ -truth under the general model interpretation to first-order consequence from a recursive set of axioms: which shows it to be recursively enumerable and, hence, recursively axiomatisable (by Craig's Theorem). This strongly contrasts with the negative result in Section 2.3. We conclude with a few comments on the situation.

Henkin's original general models (defined, by the way, with respect to a richer language) form a proper subclass of ours. This is because one may strengthen the theorem a little (or much — depending on one's philosophy) by adding to (1)–(6) translations of  $L_\omega$ -sentences obviously true in the *standard model* interpretation, thereby narrowing the class of admissible general models. Of course, Section 2.3 prevents an effective narrowing down to *exactly* the standard models!

Here are two examples of such additional axioms, bringing the general models interpretation closer to the standard one.

1. Comprehension Axioms for type  $\tau = (\tau_1, \dots, \tau_n)$ :

$$\forall X_1 \dots \forall X_m \exists Y \forall Z_1 \dots \forall Z_n (Y(Z_1, \dots, Z_n) \leftrightarrow \varphi),$$

where  $Y$  has type  $\tau$ ,  $Z_i$  type  $\tau_i$  ( $1 \leq i \leq n$ ) and the free variables of  $\varphi$  are among  $X_1, \dots, X_m, Z_1, \dots, Z_n$ . Thus, all definable predicates are to be actually present in the model.

2. Axioms of Choice for type  $\tau = (\tau_1, \dots, \tau_n, \tau_{n+1})$ :

$$\forall Z_1 \exists Z_2 \forall X_1 \dots \forall X_n (\exists Y Z_1(X_1, \dots, X_n, Y) \rightarrow \rightarrow \exists ! Y Z_2(X_1, \dots, X_n, Y));$$

where  $Z_1, Z_2$  have type  $\tau$ ,  $X_i$  has  $\tau_i$  ( $1 \leq i \leq n$ ) and  $Y$  has type  $\tau_{n+1}$ . Thus, every relation contains a function: cf. Bernays' Axiom of Choice mentioned in Section 2.5.1.

There is also a more ‘deductive’ motivation for these axioms. When one ponders which principles of deduction should enter into any reasonable higher-order logic, one immediate candidate is the ordinary complete first-order axiom set, with quantifiers now also of higher orders (cf. [Enderton, 1972], last chapter, for this line). All usual principles are valid in general models without further ado, except for Universal Instantiation, or equivalently, Existential Generalization:

$$\forall X\varphi(X) \rightarrow \varphi(T) \text{ or } \varphi(T) \rightarrow \exists X\varphi(X).$$

These two axioms are valid in all general models when  $T$  is any variable or constant of the type of  $X$ . But, in actual practice, one wants to substitute further instances in higher-order reasoning. For example, from  $\forall X\varphi(X, R)$ , with  $X$  of type  $(0)$ , one wants to conclude  $\varphi(\psi)$  for any *first-order* definable property  $\psi$  in  $R$ , = (cf. van Benthem’s chapter on Correspondence Theory in Volume 3 of this *Handbook*). In terms of Comprehension, this amounts to closure of predicate ranges under first-order definability, mentioned in Section 4.1. A further possibility is to allow *predicative* substitutions, where  $\psi$  may be higher-order, but with its quantifiers all ranging over orders lower than that of  $X$ . Finally, no holds barred, there is the use of *arbitrary substitutions*, whether predicative or not; as in the above Comprehension Schema.

One consequence of Comprehension is the following Axiom of Descriptiveness:

$$\forall x\exists!y\varphi(x, y) \rightarrow \exists f\forall x\varphi(x, f(x)).$$

If we want to strengthen this to the useful existence of Skolem functions (cf. Section 2.5.2), we have to postulate

$$\forall x\exists y\varphi(x, y) \rightarrow \exists f\forall x\varphi(x, f(x));$$

and this motivates the above Axioms of Choice.

No further obvious logical desiderata seem to have been discovered in the literature.

By the way, our above formulation of the Axiom of Choice cannot be strengthened when all types are present, assuming the comprehension axioms. If this is not the case, it can be. For instance, in the second-order language, the strongest possible formulation is just the implication  $\forall x\exists X\psi \rightarrow \exists Y\forall x\psi'$  (where  $\psi'$  is obtained from  $\psi$  by substituting  $Y(x, t_1, \dots, t_n)$  for  $X(t_1, \dots, t_n)$ ) used to prove the prenex theorem in Section 3.2.

In a sense, this form gives more than just choice; conceived of set-theoretically, it has the flavour of a ‘collection’ principle. It plays a crucial role in proving reflectivity of second-order theories containing it, similar to the role the substitution (or collection) axiom has in proving reflection principles in set theory.

The general picture emerging here is that of an ascending range of recursively axiomatized higher-order logics, formalizing most useful fragments of  $L_\omega$ -validity that one encounters in practice.

### 4.3 Second-Order Reduction

The general completeness theorem, or rather, the family of theorems in Section 4.2, by no means exhausts the uses of the general model idea of Section 4.1. For instance, once upon this track, we may develop a ‘general model theory’ which is much closer to the first-order subject of that description. A case in point are the ‘general ultraproducts’ of [van Benthem, 1983], which allow for an extension of the fundamental characterization theorems of Section 1.4 to higher-order logic. This area remains largely unexplored.

Here we present a rather more unexpected application, announced in Section 3.2:  $L_\omega$ -standard validity is effectively reducible to standard validity in monadic  $L_2$ , in fact in the monadic  $\Sigma_1^1$ -fragment.

Consider the *first-order* language  $L^+$  (relative to a given base language  $L$ ) introduced in Section 4.1. Extend it to a *second-order* language  $L_2^+$  by adding second-order variables of all types  $(0, \dots, 0)$ , with which we can form atoms  $X(t_1, \dots, t_n)$ . Consider the following  $L_2^+$ -principles ( $\tau = (\tau_1, \dots, \tau_n)$ ):

#### Plenitude( $\tau$ )

$$\forall X \exists x \forall y_1 \dots \forall y_n (T_\tau(x) \wedge (T_{\tau_1}(y_1) \wedge \dots \wedge T_{\tau_n}(y_n)) \rightarrow \rightarrow (\varepsilon_\tau(x, y_1, \dots, y_n) \leftrightarrow X(y_1, \dots, y_n))).$$

Evidently, Plenitude holds in all  $^+$ -transforms of all standard models of  $L_\omega$ . Conversely, if the  $L^+$ -model  $\mathfrak{B}$  satisfies Plenitude( $\tau$ ) for all types  $\tau$ , then its submodel  $\mathfrak{B}^*$  (cf. the proof of the main theorem in Section 4.2) is isomorphic to a model of the form  $(\mathfrak{A}, E)^+$  with a full type structure  $E$ .

**THEOREM.**  $\varphi \in L_\omega$  is true in all standard models iff  $\varphi^+$  follows from (1), (2), (3), (5), (6), and the Plenitude axioms.

As  $\varphi$  can only mention a finite number of types and non-logical constants, the relevant axioms of the above-mentioned kinds can be reduced to a finite number and hence to a single sentence  $\psi$ .

**THEOREM.** With every  $\varphi \in L_\omega$ , a  $\Pi_1^1$ -sentence  $\psi$  of  $L_2^+$  can be associated effectively, and uniformly, such that

$$\models_\omega \psi \text{ iff } \models_2 \psi \rightarrow \varphi^+.$$

As  $\psi \in \Pi_1^1$  and  $\varphi^+$  is first-order, this implication is equivalent to a  $\Sigma_1^1$ -sentence; and the promised reduction is there.



But Plenitude has been formulated using second-order variables of an arbitrary type. We finally indicate how this may be improved to the case of only monadic ones. Consider the variant

**Plenitude\***( $\tau$ )

$$\forall X \exists x \forall y_1 \dots \forall y_n (T_\tau(x) \wedge (T_{\tau_1}(y_1) \wedge \dots \wedge T_{\tau_n}(y_n)) \rightarrow \\ \rightarrow (\varepsilon_\tau(x, y_1, \dots, y_n) \leftrightarrow \exists y (T_\tau(y) \wedge X(y) \wedge \varepsilon_\tau(y, y_1, \dots, y_n))))).$$

When  $E$  is full, this will obviously hold in  $(\mathfrak{A}, E)^+$ . To make this monadic variant do its job, it has to be helped by the following first-order principle stating the existence of singleton sets of ordered sequences:

**Singletons**( $\tau$ )

$$\forall z_1 \dots \forall z_n \exists x \forall y_1 \dots \forall y_n (T_{\tau_1}(z_1) \wedge \dots \wedge T_{\tau_n}(z_n) \rightarrow (T_\tau(x) \wedge \\ \wedge (T_{\tau_1}(y_1) \wedge \dots \wedge T_{\tau_n}(y_n)) \rightarrow (\varepsilon_\tau(x, y_1, \dots, y_n) \leftrightarrow \\ \leftrightarrow y_1 = z_1 \wedge \dots \wedge y_n = z_n))))).$$

Suppose now that  $\mathfrak{B}$  satisfies all these axioms and  $(\mathfrak{A}, E)^+ \cong \mathfrak{B}^*$ . Let  $S \subseteq E_{\tau_1} \times \dots \times E_{\tau_n}$  be arbitrary: we must show that  $S \in E_\tau$ . Notice that **Singletons**( $\tau$ ) implies that, if  $s \in E_{\tau_1} \times \dots \times E_{\tau_n}$  (in particular, if  $s \in S$ ), then  $\{s\} \in E_\tau$ . Now let  $S' := \{\{s\} \mid s \in S\}$ . Clearly,  $S = \bigcup S'$  and  $S' \subseteq E_\tau$ . That  $S \in E_\tau$  follows from one application of **Plenitude\***( $\tau$ ), taking  $S'$  as value for  $X$ . ■

#### 4.4 Type Theory and Lambda Calculus

Readers of Section 4.2 may have been a little disappointed at finding no preferred *explicit* axiomatized ‘first-order’ version of  $L_\omega$ -logic. And indeed, an extreme latitude of choices was of the essence of the situation. Indeed, there exist various additional points of view leading to, at least, interesting logics. One of these is provided by the earlier functional type theory of Section 3.3. We will chart the natural road from the perspective of its basic primitives.

*Identity* and *application* inspire the usual identity axioms, *Lambda abstraction* really including replacement of identicals. *Lambda abstraction* really contributes only one further principle, viz. the famous ‘lambda conversion’

$$\lambda x \cdot B(A) = [A/x]B;$$

for  $x, B, A$  of suitable types, and modulo obvious conditions of freedom and bondage. Thus, there arises a simple kind of *lambda calculus*. (Actually, a rule of ‘alphabetic bound variants’ will have to be added in any case, for domestic purposes.)

Lambda conversion is really a kind of simplification rule, often encountered in the semantics of natural, or programming languages. One immediate question then is if this process of simplification ever stops.

**THEOREM.** *Every lambda reduction sequence stops in a finite number of steps.*

**Proof.** Introduce a suitable measure of type complexity on terms, so that each reduction lowers complexity. ■

This theorem does not hold for the more general *type free* lambda calculi of [Barendregt, 1980]; where, e.g.  $\lambda x \cdot x(x)(\lambda x \cdot x(x))$  runs into an infinite regress.

Another immediate follow-up question concerns the *unicity* (in addition to the above *existence*) of such irreducible ‘normal forms’. This follows in fact from the ‘diamond property’:

**THEOREM.** (Church-Rosser) *Every two lambda reduction sequences starting from the same terms can be continued to meet in a common term (up to alphabetic variance).*

Stronger lambda calculi arise upon the addition of further principles, such as *extensionality*:

$$\lambda x \cdot A(x) = \lambda x \cdot B(x) \text{ implies } A = B \text{ (for } x \text{ not free in } A, B).$$

This is the lambda analog of the earlier principle (6) in Section 4.2.

Still further additions might be made reflecting the constancy of the truth value domain  $D_t$ . Up till now, all principles considered would also be valid for arbitrary truth value structures. (In some cases, this will be a virtue, of course.)

Let us now turn to traditional logic. Henkin has observed how all familiar logical constants may be *defined* (under the standard interpretation) in terms of the previous notions. Here is the relevant list [Henkin, 1963]:

$$\begin{aligned} \top \text{ (a tautology)} &:= \lambda x \cdot x = \lambda x \cdot x \\ \perp \text{ (a contradiction)} &:= \lambda x_t \cdot x_t = \lambda x_t \cdot \top \\ \neg \text{ (negation)} &:= \lambda x_t \cdot x_t = \perp \end{aligned}$$

The most tricky case is that of conjunction:

$$\wedge := \lambda x_t \cdot \lambda y_t (\lambda f_{(t,t)} \cdot (f_{(t,t)}(x_t) = y_t) = \lambda f_{(t,t)} \cdot f_{(t,t)} \top)$$

One may then define  $\vee, \rightarrow$  in various ways. Finally, as for the quantifiers,

$$\forall x A := \lambda x \cdot A = \lambda x \cdot \top.$$

The induced logic has not been determined yet, as far as we know.

With the addition of the axiom of *bivalence*, we are on the road to classical logic:

$$\forall x_t \cdot f_{(t,t)} \cdot x_t = f_{(t,t)} \top \wedge f_{(t,t)} \perp.$$

For a fuller account, cf. [Gallin, 1975, Chapter 1.2].

One may prove a general completeness theorem for the above identity, application, abstraction theory in a not inelegant direct manner, along the lines of Henkin's original completeness proof. (Notably, the familiar 'witnesses' would now be needed in order to provide instances  $f(c) \neq g(c)$  when  $f \neq g$ .) But, the additional technicalities, especially in setting up the correct account of general models for functional-type theory, have motivated exclusion here.

Even so, the differences between the more 'logical' climate of functional-type theory and the more 'set-theoretic' atmosphere of the higher-order  $L_\omega$  will have become clear.

## 5 REFLECTIONS

Why should a Handbook of (after all) Philosophical Logic contain a chapter on extensions of first-order logic; in particular, on higher-order logic? There are some very general, but also some more specific answers to this (by now) rather rhetorical question.

One general reason is that the advent of competitors for first-order logic may relativize the intense preoccupation with the latter theory in philosophical circles. No specific theory is sacrosanct in contemporary logic. It is rather a certain logical perspective in setting up theories, weaker or stronger as the needs of some specific application require, that should be cultivated. Of course, this point is equally valid for *alternatives* to, rather than *extensions* of classical first-order logic (such as intuitionistic logic).

More specifically, two themes in Section 1 seem of a wider philosophical interest: the role of limitative results such as the Löwenheim–Skolem, or the Compactness theorem for scientific theory construction; but also the new systematic perspective upon the nature of logical constants (witness the remarks made about generalized quantifiers). Some authors have even claimed that proper applications of logic, e.g. in the philosophy of science or of language, can only get off the ground now that we have this amazing diversity of logics, allowing for conceptual 'fine tuning' in our formal analyses.

As for the specific case study of higher-order logic, there was at least a convincing *prima facie* case for this theory, both from the (logician) foundations of mathematics and the formal semantics of natural language. Especially in the latter area, there have been recurrent disputes about clues from natural language urging higher-order descriptions. (The discussion of branching quantifiers in Section 2.5.1 has been an example; but many others could be cited.) This subject is rather delicate, however, having to do with philosophy as much as with linguistics. (Cf. [van Benthem, 1984] for a discussion of some issues.) For instance, the choice between a standard model or a general model approach to higher-order quantification is semantically

highly significant and will hopefully undercut at present rather dogmatic discussions of the issue. For instance, even on a Montagovian type theoretic semantics, we are not committed to a non-axiomatizable logic, or models of wild cardinalities: contrary to what is usually claimed. (General models on a countable universe may well remain countable throughout, no matter how far the full type structure explodes.)

One might even hazard the conjecture that natural language is partial to restricted predicate ranges which are *constructive* in some sense. For instance, [Hintikka, 1973] contains the suggestion to read branching quantifier statements on countable domains in terms of the existence of Skolem functions which are recursive in the base predicates. If so, our story might end quite differently: for, the higher-order logic of constructive general models might well lapse into non-axiomatizability again. Thus, our chapter is an open-ended one, as far as the philosophy and semantics of language are concerned. It suggests possibilities for semantic description; but on the other hand, this new area of application may well inspire new directions in logical research.

#### ADDENDA

This chapter was written in the summer of 1982, in response to a last-minute request of the editors, to fill a gap in the existing literature. No standard text on higher-order logic existed then, and no such text has emerged in the meantime, as far as our information goes. We have decided to keep the text of this chapter unchanged, as its topics still seem to the point. Nevertheless, there have been quite a few developments concerning different aspects of our exposition. We provide a very brief indication — without any attempt at broad coverage.<sup>1</sup>

##### *Ehrenfeucht-Fraïssé Games*

Game methods have become a common tool in logic for replacing compactness arguments to extend standard meta-properties beyond first-order model theory. Cf. [Hodges, 1993], [Doets, 1996]. They extend to many variations and extensions of first-order logic (cf. [Barwise and van Benthem, 1996]).

##### *Finite Model Theory*

Model theory over finite models has become a topic in its own right. Cf. [Ebbinghaus and Flum, 1995]. For connections with data base theory, cf.

<sup>1</sup>The following people were helpful in providing references: Henk Barendregt, Philip Kremer, Godehard Link, Maria Manzano, Marcin Mostowski, Reinhard Muskens, Mikhail Zakhariashev.

[Kanellakis, 1990]. In particular, over finite models, logical definability links up with computational complexity: cf. [Immerman, 1996].

### *General Models*

[Henkin, 1996] is an exposition by the author of the original discovery. [Manzano, 1996] develops a broad spectrum of applied higher-order logics over general models with partial truth values. [van Benthem, 1996] gives a principled defense of general models in logical semantics, as a ‘geometric’ strategy of replacing predicates by objects.

### *Order-Independent Properties of Logics*

The distinction ‘first-order’/‘higher-order’ is sometimes irrelevant. Many logical properties hold independently of the division into logical ‘orders’. Examples are monotonicity (upward preservation of positive statements) or relativization (quantifier restriction to definable subdomains), whose model-theoretic statements have nothing to do with orders. There is an emerging linguistic interest in such ‘transcendental’ properties: cf. [van Benthem, 1986b], [Sanchez Valencia, 1991].

### *Generalized Quantifier Theory*

The theory of generalized quantifiers has had a stormy development in the 80s and 90s, both on the linguistic and the mathematical side. Cf. [van Benthem, 1986a], [Westerståhl, 1989]. In particular, the latter has systematic game-based (un-) definability results for hierarchies of generalized quantifiers. [van Benthem and Westerståhl, 1995] is a survey of the current state of the field, [Keenan and Westerståhl, 1996] survey the latest linguistic applications, many of which involve the polyadic quantifiers first introduced by [Lindström, 1966].

### *Higher-Order Logic in Computer Science*

Higher-order logics have been proposed for various applications in computer science. Cf. [Leivant, 1994].

### *Higher-Order Logic in Natural Language*

Much discussion has centered around the article [Boolos, 1984], claiming that plurals in natural language form a plausible second-order logic. Strong relational higher-order logics have been proposed by [Muskens, 1995]. The actual extent of higher-order phenomena is a matter of debate: cf. [Lönning,

1996], [Link, 1997, Chapter 14]. In particular, there is a continuing interest in better-behaved ‘bounded fragments’ that arise in natural language semantics.

### *Higher-Order Logic in the Philosophy of Science*

Higher-order logic has been used essentially in the philosophy of time (cf. various temporal postulates and open questions in [van Benthem, 1992]), the foundations of physics and measurement (cf. the higher-order physical theories of [Field, 1980]) and mathematics (cf. [Shapiro, 1991]).

### *Infinitary Logic*

Infinitary logics have become common in computer science: cf. [Harel, 1984], [Goldblatt, 1982]. In particular, fixed-point logics are now a standard tool in the theory of data bases and query languages: cf. [Kanellakis, 1990]. Recently, [Barwise and van Benthem, 1996] have raised the issue just what are the correct formulations of the first-order meta-properties that should hold here. (For instance, the standard interpolation theorem fails for  $L_{\infty\omega}$ , but more sophisticated variants go through.) Similar reformulation strategies might lead to interesting new meta-properties for second-order logic.

### *Lambda Calculus and Type Theories*

There is an exploding literature on (typed) lambda calculus and type theories, mostly in computer science. Cf. [Hindley and Seldin, 1986], [Barendregt, 1980; Barendregt, 1992], [Mitchell, 1996; Gunter and Mitchell, 1994]. In natural language, higher-order logics and type theories have continued their influence. Cf. [Muskens, 1995] for a novel use of relational type theories, and [Lapierre, 1992; Lepage, 1992] for an alternative in partial functional ones. [van Benthem, 1991] develops the mathematical theory of ‘categorical grammars’, involving linear fragments of a typed lambda calculus with added Booleans.

### *Modal Definability Theory*

First-order reductions of modal axioms viewed as  $\Pi_1^1$ -sentences have been considerably extended in [Venema, 1991], [de Rijke, 1993]. In the literature on theorem proving, these translations have been extended to second-order logic itself: cf. [Ohlbach, 1991], [Doherty *et al.*, 1994]. [Zakhariashev, 1992; Zakhariashev, 1996] provides a three-step classification of all second-order forms occurring in modal logic.

### *Propositional Quantification in Intensional Logic*

*Modal Logic.* [Kremer, 1996] considers the obvious interpretation of propositional quantification in the topological semantics for S4, and defines a system  $S4\pi\tau$ , related to the system  $S4\pi^+$  of [Fine, 1970]. He shows that second-order arithmetic can be recursively embedded in  $S4\pi\tau$ , and asks whether second order logic can.

[Fine, 1970] is the most comprehensive early piece on the topic of propositional quantifiers in modal logic. (Contrary to what is stated therein, decidability of  $S4.3\pi^+$  is open.)

*Intuitionistic Logic.* References here are [Löb, 1976], [Gabbay, 1981], [Kreisel, 1981] and [Pitts, 1992].

*Relevance Logic.* Cf. [Kremer, 1994].

### *Higher-Order Proof Theory*

Cf. [Troelstra and Schwichtenberg, 1996, Chapter 11], for a modern exposition of relevant results.

*University of Amsterdam, The Netherlands.*

## BIBLIOGRAPHY

- [Ackermann, 1968] W. Ackermann. *Solvable Cases of the Decision Problem*. North-Holland, Amsterdam, 1968.
- [Ajtai, 1979] M. Ajtai. Isomorphism and higher-order equivalence. *Annals of Math. Logic*, 16:181–203, 1979.
- [Baldwin, 1985] J. Baldwin. Definable second-order quantifiers. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*, pages 445–477. Springer, Berlin, 1985.
- [Barendregt, 1980] H. Barendregt. *The Lambda Calculus*. North-Holland, Amsterdam, 1980.
- [Barendregt, 1992] H. Barendregt. Lambda calculi with types. In S. Abramsky, D. Gabbay, and T. Maibaum, editors, *Handbook of logic in computer science Vol. 2*. Oxford University Press, 1992.
- [Barwise and Cooper, 1981] J. Barwise and R. Cooper. Generalized quantifiers and natural language. *Linguistics and Philosophy*, 4:159–219, 1981.
- [Barwise and Schlipf, 1976] J. Barwise and J. Schlipf. An introduction to recursively saturated and resplendent models. *J. Symbolic Logic*, 41:531–536, 1976.
- [Barwise and van Benthem, 1996] J. Barwise and J. van Benthem. Interpolation, preservation, and pebble games. Technical Report ML-96-12, ILLC, 1996. To appear in *Journal of Symbolic Logic*.
- [Barwise et al., 1978] J. Barwise, M. Kaufman, and M. Makkai. Stationary logic. *Annals of Math Logic*, 13:171–224, 1978. A correction appeared in *Annals of Math. Logic* 16:231–232.
- [Barwise, 1972] J. Barwise. The Hanf-number of second-order logic. *J. Symbolic Logic*, 37:588–594, 1972.
- [Barwise, 1975] J. Barwise. *Admissible Sets and Structures*. Springer, Berlin, 1975.
- [Barwise, 1977] J. Barwise, editor. *Handbook of Mathematical Logic*. North-Holland, Amsterdam, 1977.

- [Barwise, 1979] J. Barwise. On branching quantifiers in English. *J. Philos. Logic*, 8:47–80, 1979.
- [Bell and Slomson, 1969] J.L. Bell and A.B. Slomson. *Models and Ultraproducts*. North-Holland, Amsterdam, 1969.
- [Boolos, 1975] G. Boolos. On second-order logic. *J. of Symbolic Logic*, 72:509–527, 1975.
- [Boolos, 1984] G. Boolos. To be is to be a value of a variable (or to be some values of some variables). *J. of Philosophy*, 81:430–449, 1984.
- [Chang and Keisler, 1973] C.C. Chang and H.J. Keisler. *Model theory*. North-Holland, Amsterdam, 1973. Revised, 3rd edition 1990.
- [Church, 1940] A. Church. A formulation of the simple theory of types. *J. Symbolic Logic*, 5:56–68, 1940.
- [Copi, 1971] I.M. Copi. *The Logical Theory of Types*. Routledge and Kegan Paul, London, 1971.
- [de Rijke, 1993] M. de Rijke. *Extending Modal Logic*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1993.
- [Doets, 1996] K. Doets. *Basic Model Theory*. CSLI, 1996.
- [Doherty *et al.*, 1994] P. Doherty, W. Lukasiewicz, and A. Szalas. Computing circumscription revisited: A reduction algorithm. Technical Report LiTH-IDA-R-94-42, Institutionen för Datavetenskap, University of Linköping, 1994.
- [Drake, 1974] F.R. Drake. *Set Theory. An Introduction to Large Cardinals*. North-Holland, Amsterdam, 1974.
- [Ebbinghaus and Flum, 1995] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Springer, Berlin, 1995.
- [Enderton, 1970] H.B. Enderton. Finite partially-ordered quantifiers. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 16:393–397, 1970.
- [Enderton, 1972] H.B. Enderton. *A Mathematical Introduction to Logic*. Academic Press, New York, 1972.
- [Field, 1980] H. Field. *Science Without Numbers*. Princeton University Press, Princeton, 1980.
- [Fine, 1970] K. Fine. Propositional quantifiers in modal logic. *Theoria*, 36:336–346, 1970.
- [Gabbay, 1981] D. Gabbay. *Semantical investigations in Heyting's intuitionistic logic*. Reidel, Dordrecht, 1981.
- [Gallin, 1975] D. Gallin. *Intensional and Higher-Order Modal Logic*. North-Holland, Amsterdam, 1975.
- [Garland, 1974] S.J. Garland. Second-order cardinal characterisability. In *Proceedings of Symposia in Pure Mathematics*, pages 127–146. AMS, vol. 13, part II, 1974.
- [Goldblatt, 1982] R. Goldblatt. *Axiomatizing the Logic of Computer Programming*. Springer, Berlin, 1982.
- [Gunter and Mitchell, 1994] C.A. Gunter and J.C. Mitchell, editors. *Theoretical Aspects of Object-Oriented Programming: Types, Semantics, and Language Design*. The MIT Press, 1994.
- [Gurevich, 1985] Y. Gurevich. Monadic second-order theories. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*, pages 479–506. Springer, Berlin, 1985.
- [Gurevich, 1987] Y. Gurevich. Logic and the challenge of computer science. In E. Börger, editor, *Current Trends in Theoretical Computer Science*. Computer Science Press, 1987.
- [Harel, 1984] D. Harel. Dynamic logic. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic, II*, pages 497–604. Reidel, Dordrecht, 1984.
- [Henkin, 1950] L.A. Henkin. Completeness in the theory of types. *J. Symbolic Logic*, 15:81–91, 1950.
- [Henkin, 1961] L.A. Henkin. Some remarks on infinitely long formulas. In *Infinitistic Methods. Proceedings of a Symposium on the Foundations of Mathematics*, pages 167–183. Pergamon Press, London, 1961.
- [Henkin, 1963] L.A. Henkin. A theory of propositional types. *Fundamenta Mathematica*, 52:323–344, 1963.
- [Henkin, 1996] L.A. Henkin. The discovery of my completeness proofs. *Bulletin of Symbolic Logic*, 2(2):127–158, 1996.



- [Hindley and Seldin, 1986] J. Hindley and J. Seldin. *Introduction to Combinators and Lambda Calculus*. Cambridge University Press, Cambridge, 1986.
- [Hintikka, 1955] K.J.J. Hintikka. Reductions in the theory of types. *Acta Philosophica Fennica*, 8:61–115, 1955.
- [Hintikka, 1973] K.J.J. Hintikka. Quantifiers versus quantification theory. *Dialectica*, 27:329–358, 1973.
- [Hodges, 1983] W. Hodges. Elementary predicate logic. In *Handbook of Philosophical Logic: Second Edition*, Vol. I, pages 1–120. Kluwer, 2000.
- [Hodges, 1993] W. Hodges. *Model Theory*. Cambridge University Press, Cambridge UK, 1993.
- [Immerman, 1995] N. Immerman. Descriptive complexity: A logician's approach to computation. *Notices of the American Mathematical Society*, 42(10):1127–1133, 1995.
- [Immerman, 1996] N. Immerman. *Descriptive Complexity*. Springer Verlag, Berlin, 1996. To appear.
- [Kanellakis, 1990] P. Kanellakis. Elements of relational database theory. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, pages 1073–1156. Elsevier Science Publishers, Amsterdam, 1990.
- [Keenan and Westerståhl, 1996] E. Keenan and D. Westerståhl. Generalized quantifiers in linguistics and logic. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*. Elsevier Science Publishers, Amsterdam, 1996.
- [Keisler, 1971] H.J. Keisler. *Model Theory for Infinitary Logic*. North-Holland, Amsterdam, 1971.
- [Kemeny, 1950] J. Kemeny. Type theory vs. set theory. *J. Symbolic Logic*, 15:78, 1950.
- [Kleene, 1952] S.C. Kleene. Finite axiomatizability of theories in the predicate calculus using additional predicate symbols. In *Two Papers on the Predicate Calculus, Memoirs of the Amer. Math. Soc. Vol. 10*, pages 27–68. American Mathematical Society, 1952.
- [Kreisel, 1981] G. Kreisel. Monadic operators defined by means of propositional quantification in intuitionistic logic. *Reports on mathematical logic*, 12:9–15, 1981.
- [Kremer, 1994] P. Kremer. Quantifying over propositions in relevance logic: non-axiomatizability of  $\forall p$  and  $\exists p$ . *J. of Symbolic Logic*, 58:334–349, 1994.
- [Kremer, 1996] P. Kremer. Propositional quantification in the topological semantics for S4. Unpublished., 1996.
- [Krynicky and Mostowski, 1985a] M. Krynicky and M. Mostowski. Henkin quantifiers. In M. Krynicky, M. Mostowski, and L. Szczerba, editors, *Quantifiers: logics, models and computation, Vol. I*, pages 193–262. Kluwer, Dordrecht, 1985.
- [Krynicky and Mostowski, 1985b] M. Krynicky and M. Mostowski, editors. *Quantifiers: logics, models and computation, Vols. I and II*. Kluwer, Dordrecht, 1985.
- [Krynicky and Mostowski, 1985c] M. Krynicky and M. Mostowski. Quantifiers, some problems and ideas. In M. Krynicky, M. Mostowski, and L. Szczerba, editors, *Quantifiers: logics, models and computation, Vol. I*, pages 1–20. Kluwer, Dordrecht, 1985.
- [Kunen, 1971] K. Kunen. Indescribability and the continuum. In *Proceedings of Symposium in Pure Mathematics*, pages 199–204. AMS, vol. 13, part I, 1971.
- [Lapierre, 1992] S. Lapierre. A functional partial semantics for intensional logic. *Notre Dame J. of Formal Logic*, 33:517–541, 1992.
- [Leivant, 1994] D. Leivant. Higher order logic. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming Vol. II*, pages 229–321. Oxford University Press, 1994.
- [Lepage, 1992] F. Lepage. Partial functions in type theory. *Notre Dame J. of Formal Logic*, 33:493–516, 1992.
- [Lindström, 1966] P. Lindström. First-order predicate logic with generalized quantifiers. *Theoria*, 32:186–195, 1966.
- [Lindström, 1969] P. Lindström. On extensions of elementary logic. *Theoria*, 35:1–11, 1969.
- [Link, 1997] G. Link. *Algebraic Semantics in Language and Philosophy*. CSLI Publications, Stanford, 1997.
- [Löb, 1976] M.H. Löb. Embedding first order predicate logic in fragments of intuitionistic logic. *J. of Symbolic Logic*, 41:705–718, 1976.

- [Lönnig, 1996] U. Lönnig. Plurals and collectivity. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*. Elsevier Science Publishers, Amsterdam, 1996.
- [Magidor and Malitz, 1977] M. Magidor and J. Malitz. Compact extensions of  $L_Q$ . *Annals of Math. Logic*, 11:217–261, 1977.
- [Magidor, 1971] M. Magidor. On the role of supercompact and extendible cardinals in logic. *Israel J. Math.*, 10:147–157, 1971.
- [Manzano, 1996] M. Manzano. *Extensions of First Order Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, 1996.
- [Mason, 1985] I. Mason. The metatheory of the classical propositional calculus is not axiomatizable. *J. Symbolic Logic*, 50:451–457, 1985.
- [Mitchell, 1996] J.C. Mitchell, editor. *Foundations for Programming Languages*. The MIT Press, 1996. 846 pages.
- [Monk, 1976] J.D. Monk. *Mathematical Logic*. Springer, Berlin, 1976.
- [Montague, 1974] R. Montague. In R.H. Thomason, editor, *Formal Philosophy: Selected Papers of Richard Montague*. Yale University Press, New Haven, 1974.
- [Moschovakis, 1980] Y.N. Moschovakis. *Descriptive Set Theory*. North-Holland, Amsterdam, 1980.
- [Mostowski, 1985] M. Mostowski. Quantifiers definable by second order means. In M. Krynicki, M. Mostowski, and L. Szczerba, editors, *Quantifiers: logics, models and computation, Vol. II*, pages 181–214. Kluwer, Dordrecht, 1985.
- [Muskens, 1989] R. Muskens. A relational reformulation of the theory of types. *Linguistics and Philosophy*, 12:325–346, 1989.
- [Muskens, 1995] R. Muskens. *Meaning and Partiality*. Studies in Logic, Language and Information. CSLI Publications, Stanford, 1995.
- [Myhill and Scott, 1971] J. Myhill and D.S. Scott. Ordinal definability. In *Proceedings of Symposia in Pure Mathematics*, pages 271–278. AMS, vol. 13, part I, 1971.
- [Ohlbach, 1991] H.-J. Ohlbach. Semantic-based translation methods for modal logics. *Journal of Logic and Computation*, 1(5):691–746, 1991.
- [Orey, 1959] S. Orey. Model theory for the higher-order predicate calculus. *Transactions of the AMS*, 92:72–84, 1959.
- [Pitts, 1992] A.M. Pitts. On an interpretation of second order quantification in first order intuitionistic propositional logic. *J. of Symbolic Logic*, 57:33–52, 1992.
- [Rabin, 1969] M.O. Rabin. Decidability of second-order theories and automata on infinite trees. *Transactions of the AMS*, 141:1–35, 1969.
- [Ressayre, 1977] J.P. Ressayre. Models with compactness properties relative to an admissible language. *Annals of Math. Logic*, 11:31–55, 1977.
- [Sanchez Valencia, 1991] V. Sanchez Valencia. *Studies on Natural Logic and Categorical Grammar*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1991.
- [Shapiro, 1991] S. Shapiro. *Foundations without Foundationalism, a case study for second-order logic*. Oxford Logic Guides 17. Oxford University Press, Oxford, 1991.
- [Svenonius, 1965] L. Svenonius. On the denumerable models of theories with extra predicates. In *The Theory of Models*, pages 376–389. North-Holland, Amsterdam, 1965.
- [Troelstra and Schwichtenberg, 1996] A.S. Troelstra and H. Schwichtenberg. *Basic Proof Theory*. Cambridge University Press, 1996.
- [Turner, 1996] R. Turner. Types. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*. Elsevier Science Publishers, Amsterdam, 1996.
- [Väänänen, 1982] J. Väänänen. Abstract logic and set theory: II large cardinals. *J. Symbolic Logic*, 47:335–346, 1982.
- [van Benthem and Westerståhl, 1995] J. van Benthem and D. Westerståhl. Directions in generalized quantifier theory. *Studia Logica*, 55(3):389–419, 1995.
- [van Benthem, 1977] J.F.A.K. van Benthem. Modal logic as second-order logic. Technical Report 77-04, Mathematisch Instituut, University of Amsterdam, 1977.
- [van Benthem, 1983] J.F.A.K. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Naples, 1983.
- [van Benthem, 1984] J.F.A.K. van Benthem. Questions about quantifiers. *Journal of Symbolic Logic*, 49:443–466, 1984.

- [van Benthem, 1986a] J.F.A.K. van Benthem. *Essays in Logical Semantics*. Reidel, Dordrecht, 1986.
- [van Benthem, 1986b] J.F.A.K. van Benthem. The ubiquity of logic in natural language. In W. Leinfellner and F. Wuketits, editors, *The Tasks of Contemporary Philosophy*, Schriftenreihe der Wittgenstein Gesellschaft, pages 177–186. Verlag Hölder-Pichler-Tempsky, Wien, 1986.
- [van Benthem, 1989] J.F.A.K. van Benthem. Correspondence theory. In *Handbook of Philosophical Logic, Second Edition*, Volume 3, Kluwer, 2001. First published in *Handbook of Philosophical Logic*, Volume 2, 1989.
- [van Benthem, 1991] J.F.A.K. van Benthem. *Language in Action. Categories, Lambdas and Dynamic Logic*. North-Holland, Amsterdam, 1991.
- [van Benthem, 1992] J.F.A.K. van Benthem. *The Logic of Time*. Reidel, Dordrecht, 1992. second edition.
- [van Benthem, 1996] J.F.A.K. van Benthem. Content versus wrapping: An essay in semantic complexity. In M. Marx, M. Masuch, and L. Pólos, editors, *Logic at Work*, Studies in Logic, Language and Information. CSLI Publications, 1996.
- [Venema and Marx, 1996] Y. Venema and M. Marx. *Multi-Dimensional Modal Logic*. Kluwer, Dordrecht, 1996.
- [Venema, 1991] Y. Venema. *Many-Dimensional Modal Logics*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1991.
- [Westerståhl, 1989] D. Westerståhl. Quantifiers in formal and natural languages. In D. Gabbay and F. Guentner, editors, *Handbook of Philosophical Logic IV*, pages 1–131. Reidel, Dordrecht, 1989.
- [Zakhariashev, 1992] M. Zakhariashev. Canonical formulas for K4. part I: Basic results. *J. Symb. Logic*, 57:1377–1402, 1992.
- [Zakhariashev, 1996] M. Zakhariashev. Canonical formulas for K4. part II: Cofinal sub-frame logics. *J. Symb. Logic*, 61:421–449, 1996.



## ALGORITHMS AND DECISION PROBLEMS: A CRASH COURSE IN RECURSION THEORY

At first sight it might seem strange to devote in a handbook of philosophical logic a chapter to algorithms. For, algorithms are traditionally the concern of mathematicians and computer scientists. There is a good reason, however, to treat the material here, because the study of logic presupposes the study of languages, and languages are by nature discrete inductively defined structures of words over an alphabet. Moreover, the derivability relation has strong algorithmic features. In almost any (finitary) logical system, the consequences of a statement can be produced by an algorithm. Hence questions about derivability, and therefore also underderivability, ask for an analysis of possible algorithms. In particular, questions about decidability (is there an algorithm that automatically decides if  $\psi$  is derivable from  $\varphi$ ?) boil down to questions about *all* algorithms. This explains the interest of the study of algorithms for logicians.

There is also a philosophical aspect involved: granting the mathematical universe, and by association the logicians universe, an independent status, as providing the basic building blocs for abstract science, it is of supreme importance to discover which basic objects and structures are given to us in a precise and manageable manner. The natural numbers have long remained the almost unique paradigm of a foundationally justified notion, with a degree of universal acceptance. The class of algorithms as given by any of the current systems (Turing machines, Post systems, Markov systems, lambda calculable functions, Herbrand-Gödel computable functions, register machines, etc.), have in this century become the second such class. As Gödel put it, “It seems to me that this importance [i.e. of the notion of recursive function] is largely due to the fact that with this concept one has for the first time succeeded in giving an absolute definition of an interesting epistemological notion, i.e. one not depending on the formalism chosen.” [Gödel, 1965], [Wang, 1974, p. 81]

The reader may feel encouraged to go on and get acquainted with the fascinating insights that are hidden behind a certain amount of technicality.

An acquaintance with such topics as *diagonalization*, *arithmetization*, *self-reference*, *decidability*, *recursive enumerability* is indispensable for any student of logic. The mere knowledge of syntax (and semantics) is not sufficient to elevate him to the desired height.

The present chapter contains the bare necessities of recursion theory, supplemented by some heuristics and some applications to logic. The hard core of the chapter is formed by Sections 1 and 2 on primitive recursive functions and partial recursive functions—a reader who just wants the basic theory of recursivity can stick to those two sections. However, Section 0 provides

a motivation for much that happens in Sections 1 and 2. In particular, it helps the reader to view recursive functions with a machine-oriented picture in mind. Section 3 contains a number of familiar applications, mainly to arithmetical theories.

The author does not claim any originality. There is a large number of texts on recursion theory (or computability) and the reader is urged to consult the literature for a more detailed treatment, or for alternative approaches. Our approach is aimed at a relatively complete treatment of some of the fundamental theorems, accompanied by a running commentary.

Drafts of this chapter have been read by a number of colleagues and students and I have received most helpful comments. I wish to thank all those who have kindly provided comments or criticism, but I would like to mention in particular the editors of the *Handbook* and Henk Barendregt, who tried out the first draft in a course, Karst Koymans and Erik Krabbe for their error detecting and Albert Visser for many helpful discussions.

## 0 INTRODUCTION

Algorithms have a long and respectable history. There are, e.g. Euclid's algorithm for determining the greatest common divisor of two numbers, Sturm's algorithm to find the number of zeros of a polynomial between given bounds.

Let us consider the example of Euclid's algorithm applied to 3900 and 5544.

After division of	5544	by	3900	the remainder is	1644
"	3900	"	1644	"	612
"	1644	"	612	"	420
"	612	"	420	"	192
"	420	"	192	"	36
"	192	"	36	"	12
"	36	"	12	"	0

Hence, the g.c.d. of 3900 and 5544 is 12.

There are three features in the above example:

1. There is a proof that the algorithm does what it is asked to do. In this case, that 12 is actually the g.c.d., but in general that the outcome for any pair  $n, m$  is the g.c.d. (the reader will see the proof after a moment's reflection).
2. The procedure is algorithmic, i.e. at each step it is clear what we have to do, and it can be done 'mechanically' by finite manipulations. This part is clear, assuming we know how to carry out the arithmetical operations on numbers given in decimal representation.

3. The procedure stops after a finite number of steps. In a way (1) presupposes (3), but (1) might give the following result: if the procedure stops, then the answer is correct. So we are still left with the burden of showing the halting of the procedure. In our example we observe that all entries in the last column are positive and that each is smaller than the preceding one. So a (very) rough estimate tells us that we need at most 1644 steps.

Another example:

A palindrome is a word that reads the same forward or backwards, e.g. bob. Is there a decision method to test if a string of symbols is a palindrome? For short strings the answer seems obvious: you can see it at a glance. However, a decision method must be universally applicable, e.g. also to strings of 2000 symbols. Here is a good method: compare the first and last symbol and if they are equal, erase them. If not then the string is not a palindrome. Next repeat the procedure. After finitely many steps we have checked if the string is a palindrome. Here too, we can easily show that the procedure always terminates, and that the answer is correct.

The best-known example from logic is the decidability of classical propositional logic. The algorithm requires us to write down the truth table for a given proposition  $\varphi$  and check the entries in the last column if all of them are 1 (or  $T$ ). If so, then  $\vdash \varphi$ .

If  $\varphi$  has  $n$  atoms and  $m$  subformulas, then a truth table with  $m \cdot 2^n$  entries will do the job, so the process terminates. The truth tables for the basic connectives tell us that the process is effective and give us the completeness theorem.

The need for a notion of effectiveness entered logic in considerations on symbolic languages. Roughly speaking, syntax was assumed (or required) to be decidable, i.e. one either explicitly formulated the syntax in such a way that an algorithm for testing strings of symbols on syntactic correctness was seen to exist, or one postulated such an algorithm to exist, cf. [Carnap, 1937] or [Fraenkel *et al.*, 1973, p. 280 ff]. Since then it is a generally recognized practice to work with a decidable syntax. This practice has vigorously been adopted in the area of computer languages.

The quest for algorithms has been stimulated by the formalist view of logic and mathematics, as being fields described by mechanical (effective) rules. Historically best-known is Hilbert's demand for a decision method for logic and arithmetic. In a few instances there are some a priori philosophical arguments for decidability. For example, the notion ' $p$  is a proof of  $\varphi$ ' should be decidable, i.e. we should be able to recognize effectively whether or not a given proof  $p$  proves a statement  $\varphi$ . Furthermore, it is a basic assumption for the usefulness of language that well-formedness should be effectively testable.

In the thirties, a number of proposals for the codification on the notion

of ‘algorithm’ were presented. A very attractive and suggestive view was presented by Alan Turing, who defined effective procedures, or algorithms, as abstract machines of a certain kind (cf. [Turing, 1936; Kleene, 1952; Davis, 1958; Odifreddi, 1989]).

Without aiming for utmost precision, we will consider these so-called Turing machines a bit closer. This will give the reader a better understanding of algorithms and, given a certain amount of practical experience, he will come to appreciate the ultimate claim that any algorithm can be carried out (or simulated) on a Turing machine. The reason for choosing this particular kind of machine and not, e.g. Markov algorithms or Register machines, is that there is a strong conceptual appeal to Turing machines. Turing has given a very attractive argument supporting the claim that all algorithms can be simulated by Turing machines—known as *Turing’s Thesis*. We will return to the matter later.

A Turing machine can be thought of as an abstract machine (a black box) with a finite number of internal states, say  $q_1, \dots, q_n$ , a reading and a printing device, and a (potentially infinite) tape. The tape is divided into squares and the machine can move one square at a time to the left or right (it may be more realistic to make the tape move, but realism is not in our object). We suppose that a Turing machine can read and print a finite number of symbols  $S_1, \dots, S_n$ . The actions of the machine are strictly local, there are a finite number of instructions of the form: *When reading  $S_j$  and being in state  $q_i$  print  $S_k$ , go into state  $q_l$  and move to the left (or right)*.

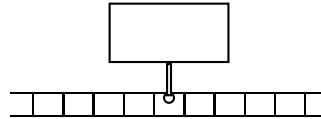


Figure 1.

We collect this instruction into a convenient string  $q_i S_j S_k q_l X$ , where  $X$  is  $L$  or  $R$ . The machine is thus supposed to read a symbol, erase it, print a new one, and move left or right. It would not hurt to allow the machine to remain stationary, but it does not add to the algorithmic power of the class of Turing machines.

Of course we need some conventions or else a machine would just go on operating and we would never be able to speak of computations in a systematic way. Here are our main conventions: (1) we will always present the machine at the beginning with a tape whose squares, except for a finite number, are blank; (2) at the beginning of the process the machine scans the leftmost non-blank square; (3) the machine stops when it is in a state and reads a symbol such that no instruction applies; (4) for any state and symbol read by the machine there is at most one instruction which applies



(i.e. the machine is *deterministic*).

Another convention, which can be avoided at the cost of some complication, is that we always have a symbol  $B$  for ‘blank’ available. This helps us to locate the end of a given string, although even here there are some snags (e.g. suppose you move right until you get to a blank, how do you know that there may not be a non-blank square way out to the right?).

Now it is time for a few examples.

### 0.1 The Palindrome Tester

We use the ideas presented above. The machine runs back and forth checking the end symbols of the string, when it is a matching pair it erases them and proceeds to the next symbol. Let us stipulate that the machine erases all symbols when it finds a palindrome, and leaves at least one non-blank square if the result is negative. Thus, we can see at a glance the *yes* or *no* answer. We introduce the symbols  $a, b, B$ . During the process we will find out how many states we need. Consider the following example: the tape is of the form  $\dots B B a a b a b a a B B \dots$  and the machine reads the first  $a$  while being in the initial state  $q_0$ , we represent this by

$$\dots B a a b a b a a B \dots$$

$q_0$

We now want to move right while remembering that we scanned a first symbol  $a$ . We do that by changing to a new state  $q_a$ . We find out that we have run through the word when we meet our first  $B$ , so then we move back, read the symbol and check if it is an  $a$ , that is when we use our memory—i.e. the  $q_a$ .

Here are the necessary instructions:

- $q_0 a a q_a R$  – in state  $q_0$ , read  $a$ , go to state  $q_a$ , move right,
- $q_a a a q_a R$  – in state  $q_a$ , read  $a$ , do nothing, move right,
- $q_a b b q_a R$  – in state  $q_a$ , read  $b$ , do nothing, move right,
- $q_a B B q_1 L$  – in state  $q_a$ , read  $B$ , go to state  $q_1$ , move left,
- $q_1 a B q_2 L$  – in state  $q_1$ , read  $a$ , erase  $a$ , go to state  $q_2$ , move left,  
and now return to the front of the word.

We indicate the moves of the machine below:

$$\begin{array}{ccccccc} B a a b a a B & \rightarrow & B a a b a a B & \rightarrow & \dots & \rightarrow & B a a b a a B & \rightarrow & B a a b a a B & \rightarrow \\ q_0 & & q_a & & & & q_a & & q_1 & \\ B a a b a a B B & \rightarrow & \dots & \rightarrow & B a a b a a B & & & & & \\ q_2 & & & & q_2 & & & & & \end{array}$$

We now move right, erase the first symbol, look for the next one and repeat the procedure.

More instructions:

$$\begin{array}{l}
 \text{move to the front} \\
 \text{erase the first symbol}
 \end{array}
 \left\{ \begin{array}{l} q_2aaq_2L \\ q_2bbq_2L \\ q_2BBq_3R \\ \\ q_3aBq_0R \\ q_3bBq_0R \end{array} \right.
 \begin{array}{l}
 \text{move right,} \\
 \text{when you see a} \\
 \text{b and check the} \\
 \text{last symbol}
 \end{array}
 \left\{ \begin{array}{l} q_0bbq_bR \\ q_bbbq_bR \\ q_baaq_bR \\ q_bBBq_4L \\ q_4bBq_2L \end{array} \right.$$

We indicate a few more steps in the computation:

$$\begin{array}{ccccccc}
 BaababaB & \rightarrow & BaababaB & \rightarrow & BBababaB & \rightarrow \dots & \rightarrow BbabB \rightarrow BbabB \rightarrow \dots \rightarrow \\
 q_2 & & q_3 & & q_0 & & q_0 & & q_b
 \end{array}$$

$$\begin{array}{ccccccc}
 BbabB & \rightarrow & BbaBB & \rightarrow \dots & \rightarrow BaB & \rightarrow BaB & \rightarrow BaB & \rightarrow BBB & \rightarrow BBB \\
 q_4 & & q_2 & & q_0 & & q_a & & q_1 & & q_2 & & q_3
 \end{array}$$

Here the machine stops, there is no instruction beginning with  $q_3B$ . The tape is blank, so the given word was a palindrome. If the word is not a palindrome, the machine stops at the end of the printed tape in state  $q_1$  or  $q_4$  and a non-blank tape is left.

One can also present the machine in the form of a graph (a kind of flow diagram). Circles represent states and arrows the action of the machine, e.g.

$$\bigcirc(q_i) \xrightarrow{S_j S_k X} \bigcirc(q_l)$$

stands for the instruction  $q_i S_j S_k q_l X$ .

The graph for the above machine is given in Figure 2.

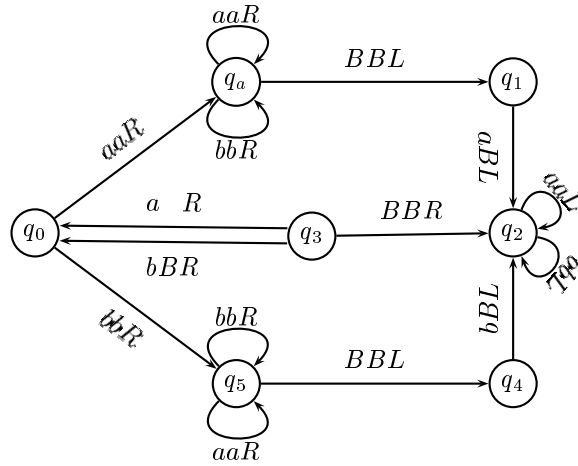


Figure 2.

The expressions consisting of a finite part of the tape containing the non-blank part plus the state symbol indicating which symbol is being read, are called *state descriptions*. For better printing we place the state symbol in

front of the scanned symbol instead of below it. A sequence of successive state descriptions is called a *computation*. Note that a computation may be infinite. In that case there is no output.

**Exercises.** Design Turing machines for the following tasks:

1. check if a word (over the alphabet  $\{a, b\}$ ) contains an  $a$ ,
2. check if a word (over the alphabet  $\{a, b\}$ ) contains two  $a$ 's,
3. check if a word (over the alphabet  $\{a, b\}$ ) has even length,
4. interchange all  $as$  and  $bs$  in a word,
5. produce the mirror image of a word.

### 0.2 Some Arithmetical Operations

We represent natural numbers  $n$  by  $n + 1$  strokes (so that 0 is taken along in accordance with modern usage). A pair of numbers is represented by two strings of strokes separated by a blank. We will denote the sequence of  $n + 1$  strokes by  $\bar{n}$ .

The convention for reading the output is: count all the strokes that are on the tape when the machine stops. It is a simple exercise to convert the tape contents into the above unary representation, but it is not always required to make the conversion.

*The identity function:  $f(x) = x$*

We just have to erase one stroke .

Instructions:

$$q_0 \mid Bq_0L$$

Here is a computation

$$Bq_0 \mid \mid \mid \cdots \mid B \rightarrow q_0BB \mid \mid \cdots \mid B.$$

*The successor function:  $f(x) = x + 1$*

This is a very simple task: do nothing. So the machine has some dummy instruction, e.g.  $q_0BBq_0R$ . This machine stops right when it starts.

*Addition:  $f(x, y) = x + y$*

Here we have to erase two strokes. It is tempting to erase both from the first string; however, the first string may contain only one  $\mid$ , so we have to be somewhat careful.

Here is the informal description: erase the first  $|$  and go into state  $q_1$ , move right until you meet the first  $|$ , erase it and stop. Instructions:

$$\begin{aligned} q_0 &| Bq_1R \\ q_1 &| Bq_2R \\ q_1 &BBq_1R. \end{aligned}$$

Example:

$$q_0 | B | \rightarrow Bq_1B | \rightarrow BBq_1 | \rightarrow BBBq_2 |$$

and

$$q_0 || B | \rightarrow Bq_1 | B | \rightarrow BBq_2B |$$

*Subtraction:*  $f(x, y) = x - y$

Observe that this is a *partial* function, for  $x - y$  is defined only for  $x \geq y$ . The obvious procedure seems to erase alternately a stroke of  $y$  and one of  $x$ . If  $y$  is exhausted before  $x$  is, we stop.

Instructions: array here

For convenience we will write  $\overset{*}{\rightarrow}$  to indicate a finite number of steps  $\rightarrow$ .

EXAMPLE.

$$\begin{aligned} Bq_0 || | B | | B \overset{*}{\rightarrow} | | | Bq_1 | | B \overset{*}{\rightarrow} B | | | B | | q_1B \rightarrow \\ B | | | B | q_2 | B \rightarrow B | | | Bq_3 | a \rightarrow B | | | q_3B | a \rightarrow \\ B | | q_4 | B | a \rightarrow B | | Bq_5B | a \overset{*}{\rightarrow} B | | BBq_2 | a \overset{*}{\rightarrow} \\ B | BBBq_5aa \rightarrow B | BBq_2Baa . \end{aligned}$$

If  $x$  is exhausted before  $y$  is, then by ( $\dagger$ ) the machine keeps moving left, i.e. it never stops. Hence for  $x < y$  there is no output.

*The projection functions:*  $U_i^n(x_0, \dots, x_n) = x_i (0 \leq i \leq n)$

The machine has to erase all the  $x_j$ 's for  $j \neq i$  and also to erase one  $|$  from  $x_i$ .

Instructions:

$$\begin{aligned} q_0 &| Bq_0R \\ q_0 &BBq_1R \\ q_1 &| Bq_1R \\ q_1 &BBq_2R \\ &\vdots \\ q_i &| Bq'_iR \\ q'_i &|| q'_iR \\ q'_i &BBq_{i+1}R \\ &\vdots \\ q_n &| Bq_nR \end{aligned}$$

By now the reader will have reached the point where he realizes that he is simply writing programs in a rather uncomfortable language for an imaginary machine. The awkwardness of the programming language is not accidental, we wanted to perform really atomic acts so that the evidence for the algorithmic character of Turing machines can immediately be read off from those acts. Of course, a high-level programming language is more convenient to handle, but it also stresses some features and neglects some other features, e.g. it might be perfect for numerical calculation and poor for string manipulations.

It is also about time to give a definition of the Turing machine, after all we have treated it so far as a *Gedankenexperiment*. Well, a *Turing machine is precisely a finite set of instructions!* For, given those instructions, we can perform all the computations we wish to perform. So, strictly speaking, adding or changing an instruction gives us a *new* machine. One can, in general, perform operations on Turing machines, e.g. for the purpose of presenting the output in a convenient way, or for creating a kind of memory for the purpose of recording the computation.

**EXAMPLE.** *Carrying out a computation between end markers.*

Let a machine  $M$  (i.e. a set of instructions) be given. We want to add two end markers, so that any computation of  $M$  has descriptions of the form  $\$1-\$2$ , where the tape contains only blanks to the left of  $\$1$  and to the right of  $\$2$ . We add two new symbols  $\$1$  and  $\$2$  to those of  $M$  and a number of instructions that take care of keeping the descriptions between  $\$1$  and  $\$2$ . For, in the course of a computation, one may need more space, so we have to build in a  $\$$ -moving feature. The following instructions will do the job:

$$\begin{array}{l} q_i \$1 B q'_i L \\ q'_i B \$1 q_i R \end{array} \left\{ \begin{array}{l} \text{if } M \text{ reads } \$1 \text{ print a blank, move one step left,} \\ \text{print } \$1, \text{ move back and to into the original state} \end{array} \right.$$

$$\begin{array}{l} q_i \$2 B q''_i R \\ q''_i B \$2 q_i L \end{array} \left\{ \begin{array}{l} \text{same action on the right hand side.} \end{array} \right.$$

Here  $q'_i$  and  $q''_i$  are new states not occurring in  $M$ .

The reader may try his hand at the following operations.

1. Suppose that a computation has been carried out between end markers. Add instructions so that the output is presented in the form  $\$1 \bar{n} \$2$  (sweeping up the strokes).
2. Let  $M$  be given, add a terminal state to it, i.e. a new state  $q_t$  such that the new machine  $M'$  acts exactly like  $M$ , but when  $M$  stops  $M'$  makes one more step so that it stops at the same description with the new  $q_t$  as state.

3. Suppose a tape containing a word between end markers is given. Add instructions to a machine  $M$  such that during a computation  $M$  preserves the word intact, i.e. any time  $M$  reads, e.g. the left end marker, it moves the whole word one square to the right, and resumes its normal activity to the left of this marker.

The last exercise may serve to store, e.g. the input in the tape as memory, so that we can use it later.

We will now consider some operations on Turing machines, required for certain arithmetical procedures. The precise details of those operations can be found in the literature, e.g. [Börger, 1989; Davis, 1958; Minsky, 1967], we will present a rough sketch here.

### *Substitution*

Suppose that machines  $M_1$  and  $M_2$  carry out the computations for the functions  $f$  and  $g$ . How can we compute  $h(x) = f(g(x))$  by means of a Turing machine? To begin with, we make the sets of states of  $M_1$  and  $M_2$  disjoint. The idea is to carry out the computation of  $M_2$  on input  $x$ , we add extra instructions so that  $M_2$  moves into a terminal state  $q_t$  when it stops. Then we add some instructions that collect at the strokes into one string and make the machine scan the leftmost  $|$  in the initial state of  $M_1$ .

As simple as this sounds, it takes a certain amount of precaution to carry out the above plan, e.g. in order to sweep all the strokes together one has to provide end markers so that one knows when all strokes have been counted, cf. the example above.

Schematically, we perform the following operations on the machines  $M_1, M_2$ : (1) change  $M_2$  into a machine  $M'_2$  which carries out the same computations, but between end markers, (2) change  $M'_2$  into  $M''_2$  which goes on to sweep all  $|$ 's together and stops in a terminal state  $q_t$  scanning the leftmost  $|$ , (3) renumber the states  $q_0, \dots, q_m$ , of  $M_1$  into  $q_t, \dots, q_{t+m}$ , the resulting machine is  $M'_1$ . Then the instructions of  $M''_2$  and  $M'_1$ , joined together, define the required machine for  $h$ . Substitution with more variables is merely a more complicated variation of the above.

### *Primitive recursion*

One of the standard techniques for defining new algorithms is that of recursion. We consider the simple parameterless case. If  $g$  is a given algorithm (and a total function) then so is  $f$ , with

$$\begin{cases} f(0) = n \\ f(x+1) = g(f(x), x). \end{cases}$$

We give a rough outline of the specification of the required Turing machine. To begin, we store the input  $x$  together with  $n$  on the tape in the

form  $\$1\bar{x}\$2\bar{n}\$3$ . we check if  $x = 0$ , i.e. we erase one  $|$  and see if no  $|$  is left. If  $x = 0$ , then we erase one  $|$  from  $\bar{n}$ , and terminate the computation. If  $x \neq 0$ , we let the machine  $N$  for  $g$  act on  $\$2\bar{n}\$3|\$4$ , sweep up the strokes between  $\$2$  and  $\$4$ , add one stroke between  $\$3$  and  $\$4$  and rewrite the tape content as  $\$1\bar{x}-1\$2f(1)\$3|\$4$ . Now we test if  $x - 1 = 0$ . If 'yes', we erase one stroke from  $f(1)$ , and all strokes between  $\$3$  and  $\$4$  and terminate. If 'no', let  $N$  operate on  $\$2f(1)\$3|\$4$ , replace the tape content by  $\$1\bar{x}-1\$2f(2)\$3||\$4$ . In  $x$  steps this procedure terminates and yields  $f(x)$ .

The resulting machine eventually stops after  $x$  steps with  $f(x)$  strokes on the tape. The addition of extra parameters is merely a matter of storing the parameters conveniently on the tape.

### *Unbounded search or minimalization*

Suppose we have a Turing machine  $M$  which computes a total function  $g(x, y)$ . Can we find a Turing machine  $M_1$  that for a given  $y$  looks for the first  $x$  such that  $g(x, y) = 0$ ?

Essentially, we will successively compute  $g(0, y), g(1, y), g(2, y), \dots$  and stop as soon as an output 0 has been produced. This is what we will do: (1) start with a tape of the form  $\dots B\$1 | B\bar{y}\$2\$3B\dots$  and read the first  $|$ , (2) copy the string between  $\$1$  and  $\$2$  between  $\$2$  and  $\$3$ , (3) let  $M$  act on the string between  $\$2$  and  $\$3$ , (4) add instructions that test if there is a  $|$  left between  $\$2$  and  $\$3$ , if not erase  $\bar{y}$  and also one stroke to the right of  $\$1$  then stop, otherwise erase everything between  $\$2$  and  $\$3$  while shifting  $\$3$  to the left, then move left and add one  $|$  following  $\$1$ , (5) repeat (2).

Clearly, if the new machine stops, then the tape content yields the desired output. The machine may, however, go on computing indefinitely. Contrary to the cases of substitution and recursion, the minimalization operation leads outside the domain of totally-defined algorithms!

The most striking feature of the family of Turing machines is that it contains a 'master' machine, that can mimic all Turing machines. This was established in Turing's very first paper on the subject. We will first give a loose and imperfect statement of this fact:

There is a Turing machine, such that if it is presented with a tape containing all the instructions of a Turing machine  $M$  plus an input, it will mimic the computations of  $M$  on this input, and yield the same output.

We will indicate the idea of the simulation process by means of a rough sketch of a simple case. Consider the addition-machine (0.2.3). On the tape we print the instructions plus the input separated by suitable symbols.

$$\$1q_0 | Bq_1R * q_1 | Bq_2R * q_1BBq_1R\$2q_0 | B | | \$3.$$

Note that the states of the addition-machine and its symbolism and the  $R$  and  $L$  have become symbols for the new machine Now we start the machine

reading the symbol to the right of  $\$2$ , it moves one square to the right, stores  $q_0$  | in its memory (i.e. it goes into a state that carries this information) and moves left looking for a pair  $q_0$  | left of  $\$2$ . When it finds such a pair, it looks at the three right-hand neighbours, stores them into its memory (again by means of an internal state), and moves right in order to replace the  $q_0$  | following  $\$2$  by  $Bq_1$ . Then the machine repeats the procedure all over again. In this way the machine mimics the original computation.

$$\begin{aligned} \$1 - \$2\bar{q}_0q_0 | B | | \$3 \xrightarrow{*} \$1 - \$2B\bar{q}_kq_kB | | \$3 \xrightarrow{*} \dots \xrightarrow{*} \\ \xrightarrow{*} \$1 - \$2BB\bar{q}_i q_1 | | \$3 \xrightarrow{*} \$1 - \$2BBB\bar{q}_j q_2 | \$3 \xrightarrow{*} \\ \xrightarrow{*} \$2BBB\bar{q}_j q_2 | \$3. \end{aligned}$$

The states of the new machine have been indicated by barred  $q$ 's. The final steps are to erase everything left of  $\$2$ .

Of course, we have in a most irresponsible way suppressed all technical details, e.g. the search procedure, the 'memory' trick. But the worst sin is our oversimplification of the representation of the instructions. In fact we are dealing with an infinite collection of Turing machines and, hence, we have to take care of infinitely many states  $q_i$  and symbols  $S_j$ . We solve this problem by a unary coding of the  $q_i$ 's and  $S_j$ 's, e.g. represent  $q_i$  by  $qq \dots q$  ( $i$  times) and  $S_j$  by  $SS \dots S$  ( $j + 1$  times). This of course complicates the above schema, but not in an insurmountable way.

A more precise formulation of the theorem concerning the so-called *Universal Turing machine* is:

*There is a Turing machine  $U$  such that for each Turing machine  $M$  it can simulate the computation of  $M$  with input  $x$ , when presented with an input consisting of a coded sequence of instructions of  $M$  and  $x$ . The output of  $U$  is identical with that of  $M$  (possibly up to some auxiliary symbols).*

One can find proofs of this theorem in a number of places, e.g. [Davis, 1958; Minsky, 1967; Turing, 1936].

If the reader is willing to accept the above facts for the moment, he can draw some immediate consequences. We will give a few informal sketches.

Let us call the coded sequence  $e$  of instructions of a machine  $M$  its *index*, and let us denote the output of  $M$  with input  $x$  by  $\varphi_e(x)$ . Obviously the universal Turing machine has itself an index; up to some coding  $U$  can act on Turing machines (i.e. their indices), in particular, on itself. In a way we can view this as a kind of self-reference or self-application.

Since Turing machines are algorithmic, i.e. given an input they effectively go through a sequence of well-determined steps and hence, produce in an effective way an output when they stop, they can be used for decision procedures. Decision problems ask for effective yes-no answers, and Turing machines provide a particular framework for dealing with them. We can



design a Turing machine that decides if a number is even, i.e. it produces a 1 if the input  $n$  is even and a 0 if  $n$  is odd.

If there is a Turing machine that produces in such a way 0–1 answers for a problem, we say that the problem is decidable. Question: are there undecidable problems? In a trivial way, yes. A problem can be thought of as a subset  $X$  of  $\mathbb{N}$ , and the question to be answered is: ‘Is  $n$  an element of  $X$ ?’. (In a way this exhausts all reasonably well-posed decision problems.) Since there are uncountably many subsets of  $\mathbb{N}$  and only countably many Turing machines, the negative answer is obvious. Let us therefore reformulate the question: are there interesting undecidable problems? Again the answer is yes, but the solution is not trivial; it makes use of Cantor’s diagonal procedure.

It would be interesting to have a decision method for the question: does a Turing machine (with index  $e$ ) eventually stop (and thus produce an output) on an input  $x$ ? This is Turing’s famous *Halting Problem*. We can make this precise in the following way: is there a Turing machine such that with input  $(e, x)$  it produces an output 1 if the machine with index  $e$  and input  $x$  eventually stops, and an output 0 otherwise. We will show (informally) that there is no such machine.

Suppose there is a machine  $M_0$  with index  $e_0$  such that

$$\varphi_{e_0}(e, x) = \begin{cases} 1 & \text{if } \varphi_e(x) \text{ exists,} \\ 0 & \text{if there is no such output for the machine with} \\ & \text{index } e \text{ on input } x. \end{cases}$$

We can change this machine  $M_0$  slightly such that we get a new machine  $M_1$  with index  $e_1$  such that

$$\varphi_{e_1}(x) = 1 \text{ if } \varphi_{e_0}(x, x) = 0$$

and there is no output if  $\varphi_{e_0}(x, x) = 1$ . One can simply take the machine  $M_0$  and change the output 0 into a 1, and send it indefinitely moving to the left if the output of  $M_0$  was 1.

Now,

$$\varphi_{e_0}(e_1, e_1) = 0 \Leftrightarrow \varphi_{e_1}(e_1) = 1 \Leftrightarrow \varphi_{e_0}(e_1, e_1) = 1.$$

*Contradiction.* So the machine  $M_0$  does not exist: the halting problem is undecidable.

Turing himself has put forward certain arguments to support the thesis that all algorithms (including the partial ones) can be carried out by means of Turing machines. Algorithms are here supposed to be of a ‘mechanical’ nature, i.e. they operate stepwise, each step is completely determined by the instructions and the given configurations (e.g. number-symbols on paper, pebbles, or the memory content of a computer), and everything involved is strictly finite. Since computations have to be performed on (or in) some

device (paper, strings of beads, magnetic tape, etc.) it will not essentially restrict the discussion if we consider computations on paper. In order to carry out the algorithm one (or a machine) has to act on the information provided by the configuration of symbols on the paper. The effectiveness of an algorithm requires that one uses an immediately recognizable portion of this information, so one can use only *local* information (we cannot even copy a number of 20 figures as a whole!), such as three numerals in a row or the letters attached to the vertices of a small-sized triangle. A Turing machine can only read one symbol at a time, but it can, e.g. scan three squares successively and use the information by using internal states for memory purposes. So the limitations of the reading ability to one symbol at a time is not essential.

The finiteness condition on Turing machines, i.e. both on the alphabet and on the number of states, is also a consequence of the effectiveness of algorithms. An infinite number of symbols that can be printed on a square would violate the principle of immediate recognizability, for a number of symbols would become so similar that they would drop below the recognizability threshold.

Taking into account the ability of Turing machines to simulate more complex processes by breaking them into small atomic acts, one realizes that any execution of an algorithm can be mimicked by a Turing machine. We will return to this matter when we discuss Church's Thesis.

There are many alternative but equivalent characterizations of algorithms: recursive functions,  $\lambda$ -calculable functions, Markov Algorithms, Register Machines, etc.—all of which have the discrete character in common. Each can be given by a finite description of some sort. Given this feature, it is a fundamental trick to code these machines, or functions, or whatever they may be, into natural numbers. The basic idea, introduced by Gödel, is simple: a description is given in a particular (finite) alphabet, code each of the symbols by fixed numbers and code the strings, e.g. by the prime-power-method.

EXAMPLE. Code  $a$  and  $b$  as 2 and 3. Then the strings  $aba\ aaba\ \dots$  are coded as  $2^2 \cdot 3^3 \cdot 5^2, 2^2 \cdot 3^2 \cdot 5^3 \cdot 7^2 \cdot 11^3 \cdot 13^3, \dots$ . Note that the coding is fully effective: we can find for each word its numerical code, and conversely, given a natural number, we simply factorize it and, by looking at the exponents, can check if it is a code of a word, and if so, of which word.

Our example is, of course, shockingly simple, but the reader can invent (or look up) more complicated and versatile codings, cf. [Smorynski, 1991].

The coding reduces the study of algorithms and decision methods to that of effective operations on natural numbers.

EXAMPLE. (1) We consider strings of  $a$ 's and  $b$ 's, and we want to test if such a string contains 15 consecutive  $b$ 's.

First we code  $a \rightarrow 1, b \rightarrow 2$  and next each string  $x_1, x_2 \dots x_n$  is coded as  $p_1^{\bar{x}_1} \cdot p_2^{\bar{x}_2} \dots p_n^{\bar{x}_n}$ , where  $p_i$  is the  $i$ th prime and  $\bar{x}_i$  the code of  $x_i (x_i \in \{a, b\})$ , e.g.  $a \rightarrow 2^1 \cdot 3^2 \cdot 5^2 \cdot 7^1 = 3150, bbb \rightarrow 44100$ .

Under this coding, the test for containing 15 consecutive  $b$ 's is taken to be a test for a number to be divisible by 15 squares of consecutive primes, which is a purely number-theoretic test.

(2) We want an algorithm for the same set of strings that counts the number of  $a$ 's. We use the same coding, then the algorithm is translated into a numerical algorithm: compute the prime factorization of  $n$  and count the number of primes with exponent 1.

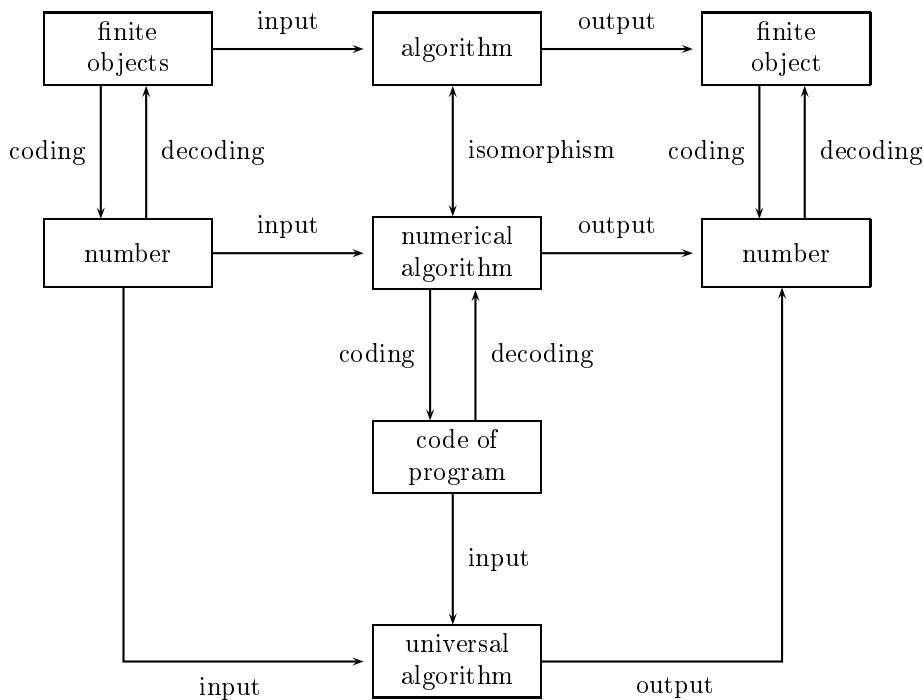


Figure 3.

Figure 3 illustrates the use of codings. the lower half contains the so-called Universal Algorithm. Our working hypothesis is that there is a standard codification of algorithms that is specified in a certain language. By coding the linguistic expression for the algorithm in standard codification into a number, we obtain two inputs for a ‘super’-algorithm that looks at the

number that codes the algorithm and then proceeds to simulate the whole computation. We will meet this so-called universal algorithm in Section 2 under the disguise of clause R7.

We also can see now why the general form of a decision problem can be taken to be of the form ‘ $n \in X?$ ’, for a set of natural numbers  $X$ .

Say we want to decide if an object  $a$  has the property  $A$ ; we consider a coding  $\#$  of the class of objects under consideration into the natural numbers. Then  $A$  is coded as a predicate  $A^\#(x)$  of natural numbers, which in turn determines the set  $A^\# = \{x \in \mathbb{N} \mid A^\#(x)\}$ . So, we have reduced the question ‘Does  $a$  have the property  $A?$ ’ to ‘ $\#(a) \in A^\#?$ ’

For theoretical purposes we can therefore restrict ourselves to the study of algorithms on natural numbers and to decision problems for sets of natural numbers.

We say that a set  $X$  is *decidable* if there is an algorithm  $F$  such that

$$F(n) = \begin{cases} 1 & \text{if } n \in X \\ 0 & \text{if } n \notin X. \end{cases}$$

We say that  $F$  tests for membership of  $X$ . In other words:  $X$  is *decidable* if its characteristic function is given by an algorithm.

## 1 PRIMITIVE RECURSIVE FUNCTIONS

Given the fact that numerical algorithms can simulate arbitrary algorithms, it stands to reason that a considerable amount of time and ingenuity has been invested in that specific area. A historically and methodologically important class of numerical algorithms is that of the *primitive recursive functions*. One obtains the primitive recursive functions by starting with a stock of acknowledged algorithms and constructing new algorithms by means of substitution and recursion. We have presented evidence that, indeed, recursion and substitution transform Turing machines into Turing machines, but the reader can easily provide intuitive arguments for the algorithmic character of functions defined by recursion from algorithmic functions. The primitive recursive functions are so absolutely basic and foundationally unproblematic (or rather, just as problematic as the natural number sequence), that they are generally accepted as a starting point for metamathematical research. Primitive recursive functions provide us with a surprisingly large stock of algorithms, including codings of finite sequences of natural numbers as mentioned above, and one has to do some highly non-trivial tricks to get algorithms which are not primitive recursive.

The basic algorithms one departs from are extremely simple indeed: the successor function, the constant functions and the projection functions  $(x_1, \dots, x_n) \mapsto x_i (i \leq n)$ . The use of recursion was already known to Dedekind, and Landau spelled out the technique in his ‘*Foundations of*

*Analysis*'. The study of primitive recursive functions was initiated in logic by Skolem, Herbrand, Gödel and others.

We will now proceed with a precise definition, which will be given in the form of an inductive definition. First we present a list of initial functions of an unmistakably algorithmic nature, and then we specify how to get new algorithms from old ones. The so-called *initial functions* are the *constant functions*  $C_m^k$  with  $C_m^k(n_1, \dots, n_k) = m$ , the *successor function*  $S$  with  $S(n) = n + 1$ , and the *projection function*  $P_i^k$  with  $P_i^k(n_1, \dots, n_k) = n_i (i \leq k)$ .

The recognized procedures are: *substitution* or *composition*, i.e. when  $f(n_1, \dots, n_k) = g(h_1(n_1, \dots, n_k), \dots, h_p(n_1, \dots, n_k))$  then we say that  $f$  is obtained by substitution from  $g$  and  $h_1, \dots, h_p$ , and *primitive recursion*, i.e. we say that  $f$  is obtained by primitive recursion from  $g$  and  $h$  if

$$\begin{cases} f(0, n_1, \dots, n_k) = g(n_1, \dots, n_k) \\ f(m + 1, n_1, \dots, n_k) = h(f(m, n_1, \dots, n_k), n_1, \dots, n_k, m). \end{cases}$$

A class of functions is *closed under substitution or primitive recursion* if  $f$  belongs to it whenever it is obtained by substitution or primitive recursion from functions that already belong to that class.

DEFINITION 1. The class of *primitive recursive functions* is the smallest class containing the initial functions that is closed under substitution and primitive recursion.

*Notation.* For convenience we abbreviate sequences  $n_1, \dots, n_k$  as  $\vec{n}$ , whenever no confusion arises.

EXAMPLES 2. *The following functions are primitive recursive.*

1.  $x + y$

$$\begin{cases} x + 0 = x \\ x + (y + 1) = (x + y) + 1 \end{cases}$$

This definition can be put in the form that shows immediately that  $+$  is primitive recursive.

$$\begin{aligned} +(0, x) &= P_1^1(x) \\ +(y + 1, x) &= S(P_1^3(+(y, x), x, y)). \end{aligned}$$

In accordance with tradition we write  $x + y$  for  $+(y, x)$ . Note that we have given an  $h$  in the second line, that actually contains all the variables that the schema of recursion prescribes. This is not really necessary since the projection functions allow us to add dummy variables.

EXAMPLE. Let  $g$  contain only the variables  $x$  and  $y$ , then we can add the dummy variable  $z$  as follows  $f(x, y, z) = g(P_1^3(x, y, z), P_2^3(x, y, z))$ .

We will leave such refinements to the reader and proceed along traditional lines.

2.  $x \cdot y$ 

$$\begin{cases} x \cdot 0 = 0 \\ x \cdot (y + 1) = x \cdot y + x \quad (\text{we use (1)}) \end{cases}$$

3.  $x^y$ 

$$\begin{cases} x^0 = 1 \\ x^{y+1} = x^y \cdot x \end{cases}$$

4. the predecessor function,  $p(x) = \begin{cases} x - 1 & \text{if } x > 0 \\ 0 & \text{if } x = 0 \end{cases}$ 

$$\begin{cases} p(0) = 0 \\ p(x + 1) = x \end{cases}$$

5. the cut-off subtraction (monus),  $x \dot{-} y$ , where  $x \dot{-} y = x - y$  if  $x \geq y$  and 0 else.

$$\begin{cases} x \dot{-} 0 = x \\ x \dot{-} (y + 1) = p((x \dot{-} y)) \end{cases}$$

6. the factorial function,  $n! = 1 \cdot 2 \cdot 3 \cdots (n - 1) \cdot n$ .7. the signum function,  $\text{sg}(x) = 0$  if  $x = 0$ , 1 otherwise.8.  $\overline{\text{sg}}$ , with  $\overline{\text{sg}}(x) = 1$  if  $x = 0$ , 0 otherwise.Observe that  $\overline{\text{sg}}(x) = 1 \dot{-} \text{sg}(x)$ .9.  $|x - y|$ , observe that  $|x - y| = (x \dot{-} y) + (y \dot{-} x)$ .10.  $f(\vec{x}, y) = \Sigma_{i=0}^y g(\vec{x}, i)$ , where  $g$  is primitive recursive.11.  $f(\vec{x}, y) = \Pi_{i=0}^y g(\vec{x}, i)$ , idem.12. If  $f$  is primitive recursive and  $\pi$  is a permutation of the set  $\{1, \dots, n\}$ , then  $g$  with  $g(x_1, \dots, x_n) = f(x_{\pi 1}, \dots, x_{\pi n})$  is also primitive recursive.**Proof.**  $g(\vec{x}) = f(P_{\pi 1}^n(\vec{x}), \dots, P_{\pi n}^n(\vec{x}))$ . ■

The reader may find it an amusing exercise to enlarge the stock for primitive recursive functions 'by hand'. We will, however, look for a more systematic way to obtain new primitive recursive functions.

**DEFINITION 3.** A relation  $R$  is primitive recursive if its characteristic function is so.

Note that this corresponds to the idea of testing  $R$  for membership: let  $K_R$  be the characteristic function of  $R$  then we know that

$$\vec{n} \in R \Leftrightarrow K_R(n_1, \dots, n_k) = 1.$$

**EXAMPLES 4.** *The following sets (relations) are primitive recursive*

1.  $\emptyset, K_\emptyset(x) = 0$
2. The set of even numbers,  $E$ .

$$\begin{cases} K_E(0) = 1 \\ K_E(x+1) = \overline{\text{sg}}(K_E(x)) \end{cases}$$

3. The equality relation  $K_=(x, y) = \overline{\text{sg}}(|x - y|)$
4. The order relation:  $K_<(x, y) = \text{sg}(x - y)$ .

Note that relations are subsets of  $\mathbb{N}^k$  for a suitable  $k$ ; when dealing with operations or relations, we assume that we have the correct number of arguments, e.g. when we write  $A \cap B$  we suppose that  $A, B \subseteq \mathbb{N}^k$ .

LEMMA 5. *The primitive recursive relations are closed under  $\cup, \cap, ^c$  and bounded quantification.*

**Proof.** Let  $C = A \cap B$ , then  $x \in C \leftrightarrow x \in A \wedge x \in B$ , so  $K_C(x) = 1 \leftrightarrow K_A(x) = 1 \wedge K_B(x) = 1$ . Therefore we put  $K_C(x) = K_A(x) \cdot K_B(x)$ . For union take  $K_{A \cup B}(x) = \text{sg}(K_A(x) + K_B(x))$ , and for the complement  $K_{A^c}(x) = \overline{\text{sg}}(K_A(x))$ .

We say that  $R$  is obtained by bounded quantification from  $S$  if  $R(n_1, \dots, n_k, m) := Qx \leq mS(n_1, \dots, n_k, x)$ , where  $Q$  is one of the quantifiers  $\forall, \exists$ .

Consider the bounded existential quantification:  $R(\vec{x}, n) := \exists y \leq nS(\vec{x}, y)$ , then  $K_R(\vec{x}, n) = \text{sg}\Sigma_{y \leq n} K_S(\vec{x}, y)$ , therefore  $R$  is primitive recursive if  $S$  is so.

The  $\forall$  case is similar, and is left to the reader. ■

LEMMA 6. *The primitive recursive relations are closed under primitive recursive substitutions, i.e. if  $f_1, \dots, f_n$  and  $R$  are primitive recursive, then so is*

$$S(x_1, \dots, x_k) := R(f_1(\vec{x}), \dots, f_n(\vec{x})).$$

**Proof.**  $K_S(\vec{x}) = K_R(f_1(\vec{x}), \dots, f_n(\vec{x}))$ . ■

LEMMA 7 (definition by cases). *Let  $R_1, \dots, R_p$  be mutually exclusive primitive recursive predicates, such that  $\forall \vec{x}(R_1(\vec{x}) \vee R_2(\vec{x}) \vee \dots \vee R_p(\vec{x}))$  and  $g_1, \dots, g_p$  primitive recursive functions, then  $f$  with*

$$f(\vec{x}) = \begin{cases} g_1(\vec{x}) & \text{if } R_1(\vec{x}) \\ g_2(\vec{x}) & \text{if } R_2(\vec{x}) \\ \vdots & \\ g_p(\vec{x}) & \text{if } R_p(\vec{x}) \end{cases}$$

*is primitive recursive.*

**Proof.** If  $K_{R_i}(\vec{x}) = 1$ , then all the other characteristic functions yield 0, so we put  $f(\vec{x}) = g_1(\vec{x}) \cdot K_{R_1}(\vec{x}) + \dots + g_p(\vec{x}) \cdot K_{R_p}(\vec{x})$ . ■

The natural numbers have the fundamental and convenient property that each non-empty set has a least element ( $\mathbb{N}$  is well-ordered). A natural question to pose is: can we effectively find this least element? In general the answer is negative, but if the set under consideration is non-empty and primitive recursive, then we can simply take the element that ensured its non-emptiness and test the smaller numbers one by one for membership.

Some notation:  $(\mu y)R(\vec{x}, y)$  stands for the least number  $y$  such that  $R(\vec{x}, y)$  if it exists.  $(\mu y < m)R(\vec{x}, y)$  stands for the least number  $y < m$  such that  $R(\vec{x}, y)$  if such a number exists; if not, we simply take it to be  $m$ .

LEMMA 8. *If  $R$  is primitive recursive, then so is  $(\mu y < m)R(\vec{x}, y)$ .*

**Proof.** Consider the following table

$R$	$R(\vec{x}, 0)$	$R(\vec{x}, 1)$	$\dots$	$R(\vec{x}, i)$	$R(\vec{x}, i+1)$	$\dots$	$R(\vec{x}, m)$
$K_R$	0	0	$\dots$	1	0	$\dots$	1
$g$	0	0	$\dots$	1	1	$\dots$	1
$h$	1	1	$\dots$	0	0	$\dots$	0
$f$	1	2	$\dots$	$i$	$i$	$\dots$	$i$

In the first line we write the values of  $K_R(\vec{x}, i)$  for  $0 \leq i \leq m$ , in the second line we make the sequence monotone, e.g. take  $g(\vec{x}, i) = \text{sg} \sum_{j=0}^i K_R(\vec{x}, j)$ . Next we switch 0 and 1:  $h(\vec{x}, i) = \overline{\text{sg}} g(\vec{x}, i)$  and finally we sum the  $h$ :  $f(\vec{x}, i) = \sum_{j=0}^i h(\vec{x}, j)$ . If  $R(\vec{x}, j)$  holds for the first time in  $i$ , then  $f(\vec{x}, m) = i$ , and if  $R(\vec{x}, j)$  does not hold for any  $j < m$ , then  $f(\vec{x}, m-1) = m$ . So  $(\mu y < m)R(\vec{x}, y) = f(\vec{x}, m)$ , and this bounded minimalization yields a primitive recursive function. ■

We put  $(\mu y \leq m)R(\vec{x}, y) := (\mu y < m+1)R(\vec{x}, y)$ .

We now have sufficient equipment to establish the primitive recursiveness of a considerable number of functions and relations.

EXAMPLES 9. *The following are primitive recursive.*

1. The set of primes:  $x$  is a prime  $\leftrightarrow \forall yz \leq x (x = yz \rightarrow y = 1 \vee z = 1) \wedge x \neq 1$ .
2. The divisibility relation:  $x \mid y \leftrightarrow \exists z \leq y (x \cdot z = y)$
3. The exponent of the prime  $p$  in the factorization of  $x$ :

$$f(x) = (\mu y \leq x)(p^y \mid x \wedge \neg p^{y+1} \mid x)$$



4. The ‘ $n$ th prime’ function:

$$\begin{cases} p_1 = 2 \\ p_{n+1} = (\mu x \leq p_n^n)[x \text{ is prime} \wedge x > p_n]. \end{cases}$$

We can use the stock of primitive recursive functions that we built up so far to get a coding of finite sequences of natural numbers into natural numbers:

$$(n_1, \dots, n_k) \mapsto 2^{n_1+1} \cdot 3^{n_2+1} \cdot \dots \cdot p_i^{n_i+1} \cdot \dots \cdot p_k^{n_k+1}.$$

Note that not all numbers figure as codes, e.g. 14 does not.

For convenience we add a code for the so-called ‘empty sequence’.

Recall that, in the framework of set theory a sequence of length  $n$  is a mapping from  $\{1, \dots, n\}$  to  $\mathbb{N}$ , so we define the empty sequence as the unique sequence of length 0, i.e. the unique map from  $\emptyset$  to  $\mathbb{N}$ , which is the empty function (set). The choice of the code is a matter of convenience, we put it 1. Following tradition, we write  $1 = \langle \rangle$ .

The predicate  $\text{Seq}(n)$ , ‘ $n$  is a sequence number’, is clearly primitive recursive, for it boils down to ‘if a prime divides  $n$ , then each smaller prime divides it’:  $\forall p, q \leq n$  (‘ $p$  is a prime’  $\wedge$  ‘ $q$  is a prime’  $\wedge q < p \wedge p \mid n \rightarrow q \mid n$ )  $\wedge n \neq 0$ . If  $n$  is a sequence number, say of  $\langle a_1, \dots, a_k \rangle$  we can find its ‘length’, i.e.  $k$ :

$$\text{lth}(n) := (\mu x \leq n + 1)[\neg p_x \mid n] - 1.$$

Observe that  $\text{lth}(2) = 0$ . We can ‘decode’  $n$ :  $(n)_i =$  (the exponent of the  $i$ th prime in the factorization of  $n$ )  $- 1$  (cf. Example 3 above). Note that  $\text{lth}(n)$  and  $(n)_i$  are primitive recursive. For a fixed  $k$

$$(a_1, \dots, a_k) \mapsto \prod_{i=1}^k p_i^{a_i+1}$$

is primitive recursive. Notation:  $\langle a_1, \dots, a_k \rangle := \prod_{i=1}^k p_i^{a_i+1}$ .

We will use abbreviations for the iterated decoding functions:  $(n)_{i,j} = ((n)_i)_j$ , etc.

We can also code the ‘concatenation’ of two sequence numbers:  $n * m$  is the code of  $\langle a_1, \dots, a_k, b_1, \dots, b_p \rangle$  where  $n$  and  $m$  are the codes of  $\langle a_1, \dots, a_k \rangle$  and  $\langle b_1, \dots, b_p \rangle$ . the definition of  $*$  is as follows (but may be skipped):

$$n * m = n \cdot \prod_{i=1}^{\text{lth}(m)} p_{\text{lth}(n)+i}^{(m)_i+1}.$$

There is one more form of recursion that will come in handy—the one where a value may depend on all preceding values. In order to make this precise we define for a function  $f(y, \vec{x})$  its ‘course of value’ function  $\bar{f}(y, \vec{x})$ :

$$\begin{cases} \bar{f}(0, \vec{x}) = 1 \\ \bar{f}(y + 1, \vec{x}) = \bar{f}(y, \vec{x}) \cdot p_{y+1}^{f(y, \vec{x})+1}, \end{cases}$$

e.g. if  $f(0) = 1, f(1) = 0, f(2) = 7$ , then

$$\bar{f}(0) = 1, \bar{f}(1) = 2^{1+1}, \bar{f}(2) = 2^{1+1} \cdot 3^1, \bar{f}(3) = 2^2 \cdot 3 \cdot 5^8.$$

Clearly, if  $f$  is primitive recursive, then so is  $\bar{f}$ . Since  $\bar{f}(n+1)$  ‘codes’ so to speak all information on  $f$  up to the  $n$ th value, we can use  $\bar{f}$  to formulate course-of-value recursion.

**THEOREM 10.** *If  $g$  is primitive recursive and  $f(y, \vec{x}) = g(\bar{f}(y, \vec{x}), y, \vec{x})$ , then  $f$  is primitive recursive.*

**Proof.** We first define  $\bar{f}$ .

$$\begin{aligned} \bar{f}(0, \vec{x}) &= 1 \\ \bar{f}(y+1, \vec{x}) &= \bar{f}(y, \vec{x}) * \langle g(\bar{f}(y, \vec{x}), y, \vec{x}) \rangle. \end{aligned}$$

By primitive recursion,  $\bar{f}$  is primitive recursive. Now  $f(y, \vec{x}) = (\bar{f}(y+1, \vec{x}))_y$ , and so  $f$  is primitive recursive. ■

By now we have collected enough facts about the primitive recursive functions. We might ask if there are more algorithms than just the primitive recursive functions. The answer turns out to be yes. Consider the following construction: each primitive recursive function  $f$  is determined by its definition, which consists of a string of functions  $f_0, f_1, \dots, f_n = f$  such that each function is either an initial function, or obtained from earlier ones by substitution or primitive recursion.

It is a matter of dull routine to code the whole definition into a natural number such that all information can be effectively extracted from the code (see [Grzegorzcyk, 1961, p. 41]). The construction shows that we may define a function  $F$  such that  $F(x, y) = f_x(y)$ , where  $f_x$  is the primitive recursive function with code  $x$ . Now consider  $D(x) = F(x, x) + 1$ . Suppose that  $D$  is primitive recursive, i.e.  $D = f_n$  for a certain  $n$ , but then  $f_n(n) = D(n) = F(n, n) + 1 = f_n(n) + 1$ . Contradiction.

*Conclusion.* We have ‘diagonalized out’ of the class of primitive recursive functions and yet preserved the algorithmic character. Hence, we have to consider a wider class of algorithms.

In case the reader should have qualms in accepting the above outlined argument, he may set his mind at ease. There are straightforward examples of algorithms that are not primitive recursive, e.g. Ackermann’s function (cf. Section 2.4).

Since our class of primitive recursive functions evidently does not contain all algorithms, we will have to look for ways of creating new algorithms not covered by substitution or primitive recursion. There are various solutions to this problem. The most radical being a switch to a conceptually different framework, e.g. that of Turing machines. We want to stay, however, as close as possible to our mode of generating the primitive recursive functions.

One way out is to generalize the minimalization, e.g. if  $g(\vec{x}, y)$  is an algorithm such that  $\forall \vec{x} \exists y (g(\vec{x}, y) = 0)$  then  $f(\vec{x}) = (\mu y)[g(\vec{x}, y) = 0]$  is an algorithm. This leads to the so-called  $\mu$ -recursive functions.

Although we will ultimately adopt another approach that will quickly yield all the fundamental theorems of the field, we will dwell for a moment on the  $\mu$ -recursive functions.

The operation of *minimalization* associates with each total function  $g(\vec{x}, y)$  a partial function  $f(\vec{x}) = \mu y [g(\vec{x}, y) = 0]$ .

**DEFINITION 11.** The class of  $\mu$ -recursive partial functions is the least set containing the initial functions  $P_i^k$  (projection),  $+$ ,  $\cdot$ ,  $K_<$  (the characteristic function of 'less than') which is closed under substitution and minimalization.

Although the successor and the constant functions are obviously  $\mu$ -recursive, we apparently have lost as much as we have won, for now we no longer have closure under recursion. One can, fortunately, show that the class of  $\mu$ -recursive (partial) functions is closed under recursion. The proof rests on the presence of a coding of finite sequences of numbers, for a computation associated with a function defined by recursion proceeds by computing successively  $f(0), f(1), \dots, f(x)$ . Although we cannot in any obvious way use the coding via the prime factorization—since we cannot make use of the exponential function—we can get an alternative coding. The main tool here is *Gödel's  $\beta$ -function*:

**THEOREM 12.** *there is a  $\mu$ -recursive function  $\beta$  such that  $\beta(n, i) \leq n - 1$  and for any sequence  $a_0, a_1, \dots, a_{n-1}$  there is an  $a$  with  $\beta(a, i) = a_i$  for  $i < n$ .*

For a proof, cf. [Shoenfield, 1967, p. 115].

One then defines the coding of  $a_0, \dots, a_{n-1}$  as  $\mu a [\forall i < n (\beta(a, i) = a_i)]$ . Here we have skipped the traditional lemma's on  $\mu$ -recursive functions and relations (in particular the closure properties), cf. [Shoenfield, 1967] or [Davis, 1958].

If we denote this particular coding temporarily by  $[a_0, \dots, a_{n-1}]$ , then we can get closure under recursion as follows:

Let

$$\begin{cases} f(0, \vec{x}) = g(\vec{x}) \\ f(y + 1, \vec{x}) = h(f(y, \vec{x}), \vec{x}, y) \end{cases}$$

put

$$f'(y, \vec{x}) = [f(0, \vec{x}, \dots, f(y, \vec{x}))]$$

then

$$f'(y, \vec{x}) = \mu z [\text{Seq}(z) \wedge \forall i < y ([z]_0 = g(\vec{x}) \wedge [z]_{i+1} = h([z]_i, \vec{x}, i))].$$

Here Seq is the obvious predicate which states that  $z$  is a coded sequence and  $[ ]_i$  is the decoding function belonging to  $[ ]$ . Taking the closure properties

for granted we see that  $f'(y, \vec{x})$  is  $\mu$ -recursive. But then so is  $f$ , since  $f(y, \vec{x}) = [f'(y, \vec{x})]_{\text{lth}}(y)$ , where  $\text{lth}$  is the proper length function.

The definition of recursiveness via minimalization has the advantage that it does not ask for fancy apparatus, just two innocent closure operations. One has, however, to work harder to obtain the fundamental theorems that concern the properties of algorithms as finite, discrete, structured objects.

The sketch of Turing machine computability that we have presented should, however, make it clear that all (partial)  $\mu$ -recursive functions can be simulated by Turing machines. The converse is also correct: every function that can be computed by a Turing machine is  $\mu$ -recursive (cf. [Davis, 1958]).

The approach to the partial recursive functions that we will use is that of Kleene using indices of recursive functions in the definition. The most striking aspect of that approach is that we postulate right away the existence of a universal function for each class of (partial) recursive functions of  $n$  arguments. The system has, so to speak, its diagonalization built in. Because of this we cannot have total functions only, for suppose that we have a universal recursive function  $g(x, y)$  for the class of all unary recursive functions, i.e. for each  $f$  in the class there is a  $y$  such that  $f(x) = g(x, y)$ . Taking for granted that the recursive functions are closed under identification of variables, we get a unary recursive function  $g(x, x)$ . Evidently  $g(x, x) + 1$  is also recursive, so  $g(x, x) + 1 = g(x, y)$  for some  $y$ . For this particular  $y$ , we get  $g(y, y) + 1 = g(y, y)$ . *Contradiction.* Since  $g(x, y)$  was taken to be recursive, we cannot conclude to have diagonalized out of the class of recursive functions. Instead, we conclude that  $g(y, y)$  is undefined, so not all recursive functions are total.

Surprising as it may seem, we thus escape a diagonalization paradox for recursion theory.

Before we start our definition of the recursive functions in earnest, it may be helpful to the reader to stress an analogy with the theory of Turing machines.

We have seen that there is a universal Turing machine that operates on suitably coded strings of instructions. Calling such a coded string the *index* of the machine that is being simulated by the universal Turing machine, we introduced the notation  $\varphi_e(x) = y$  for ‘the machine with index  $e$  yields output  $y$  on input  $x$ ’. We can now refer to the Turing machines by their indices, e.g. the existence of the universal Turing machine comes to : there is an index  $u$  such that for all indices  $e$   $\varphi_u(e, x) \simeq \varphi_e(x)$ . The last expression has to be read as ‘both sides are undefined, or they are defined and identical’.

Whereas in the case of Turing machines there is quite a lot of work to be done before one gets the universal machine, we will take the easy road and give the ‘universal’ recursive functions by one of the closure properties (clause R7 in Definition 2.1).

One final remark: matters of terminology in recursion theory are somewhat loosely observed. One should always speak of partial recursive func-

tions, and add the predicate *total* when such a function is defined for all arguments. However, the total ‘partial recursive functions are called just ‘recursive’. Moreover, some authors simply drop the adjective ‘partial’ and always speak of ‘recursive functions’. We will steer a middle course and add whatever adjectives that may be helpful. Nonetheless, the reader should be aware!

## 2 PARTIAL RECURSIVE FUNCTIONS

We will now extend the class of algorithms as indicated above. This extension will yield new algorithms *and* it will automatically widen the class to partial functions.

In our context functions have natural domains, i.e. sets of the form  $\mathbb{N}^n$  ( $= \{(m_1, \dots, m_n) \mid m_i \in \mathbb{N}\}$ , so called Cartesian products), a partial function has a domain that is a subset of  $\mathbb{N}^n$ . If the domain is all of  $\mathbb{N}^n$ , then we call the function *total*.

EXAMPLE.  $f(x) = x^2$  is total,  $g(x) = \mu y[y^2 = x]$  is partial and not total, ( $g(x)$  is the square root of  $x$  if it is an integer).

The algorithms that we are going to introduce are called *partial recursive functions*; maybe recursive partial functions would have been a better name, anyway, the name has come to be generally accepted. The particular technique for defining partial recursive functions that we employ here goes back to Kleene. As before, we use an inductive definition; apart from clause R7 below, we could have used a formulation almost identical to that of the definition of the primitive recursive functions. Since we want a built-in universal function, we have to employ a more refined technique that allows explicit reference to the various algorithms. The trick is not esoteric at all, we simply give each algorithm a code number, what we call its *index*. We fix these indices in advance so that we can speak of the ‘algorithm with index  $e$  yields output  $y$  on input  $(x_1, \dots, x_n)$ ’, symbolically represented as  $\{e\}(x_1, \dots, x_n) \simeq y$ .

Note that we do not know in advance that the result is a partial function, i.e. that for each input there is at most one output. However plausible that is, it has to be shown. Kleene has introduced the symbol  $\simeq$  for equality in the context of undefined terms. A proper treatment would be by means of the existence predicate and  $\simeq$  would be the  $\equiv$  of Van Dalen [see the chapter on Intuitionistic Logic in Volume 7 of this *Handbook*]. The convention ruling  $\simeq$  is: if  $g \simeq s$  then  $t$  and  $s$  are simultaneously defined and identical, or they are simultaneously undefined, [Kleene, 1952, p. 327].

DEFINITION 13. The relation  $\{e\}(\vec{x}) \simeq y$  is inductively defined by

- R1  $\{\langle 0, n, q \rangle\}(m_1, \dots, m_n) \simeq q$
- R2  $\{\langle 1, n, i \rangle\}(m_1, \dots, m_n) \simeq m_i$  for  $1 \leq i \leq n$
- R3  $\{\langle 2, n, i \rangle\}(m_1, \dots, m_n) \simeq m_i + 1$  for  $1 \leq i \leq n$
- R4  $\{\langle 3, n + 4 \rangle\}(p, q, r, s, m_1, \dots, m_n) \simeq p$  if  $r = s$   
 $\{\langle 3, n + 4 \rangle\}(p, q, r, s, m_1, \dots, m_n) \simeq q$  if  $r \neq s$
- R5  $\{\langle 4, n, b, c_1, \dots, c_k \rangle\}(m_1, \dots, m_n) \simeq p$  if there are  $q_1, \dots, q_k$   
such that  $\{c_i\}(m_1, \dots, m_n) \simeq q_i (1 \leq i \leq k)$  and  $\{b\}(q_1, \dots, q_k) \simeq p$
- R6  $\{\langle 5, n + 2 \rangle\}(p, q, m_1, \dots, m_n) \simeq S_n^1(p, q)$
- R7  $\{\langle 6, n + 1 \rangle\}(b, m_1, \dots, m_n) \simeq p$  if  $\{b\}(m_1, \dots, m_n) \simeq p$ .

The function  $S_n^1$  from R6 will be specified in the  $S_n^m$  theorem. It is a pure technicality, slipped in to simplify the proof of the normal form theorem. We will comment on it below.

Keeping the above reading of  $\{e\}(\vec{x})$  in mind, we can paraphrase the schema's as follows:

- R1 the machine with index  $\langle 0, n, q \rangle$  yields for input  $(m_1, \dots, m_n)$  output  $q$  (the *constant function*),
- R2 the machine with index  $\langle 1, n, i \rangle$  yields for input  $\vec{m}$  output  $m_i$  (the *projection function*  $p_i^n$ ),
- R3 the machine with index  $\langle 2, n, i \rangle$  yields for input  $\vec{m}$  output  $m_i + 1$  (the *successor function* on the  $i$ th argument),
- R4 the machine with index  $\langle 4, n + 4 \rangle$  tests the equality of the third and fourth argument of the input and puts out the first or second argument accordingly (the *discriminator function*),
- R5 the machine with index  $\langle 4, n, b, c_1, \dots, c_k \rangle$  first simulates the machines with index  $c_1, \dots, c_k$  with input  $\vec{m}$ , then uses the output sequence  $(q_1, \dots, q_k)$  as input and simulates the machine with index  $b$  (*substitution*),
- R7 the machine with index  $\langle 6, n + 1 \rangle$  simulates for a given input  $b, m_1, \dots, m_n$ , the machine with index  $b$  and input  $m_1, \dots, m_n$  (*reflection*).

The machine with index  $\langle 6, n + 1 \rangle$  acts as a *universal machine* for all machines with  $n$ -argument inputs.

Remarks. (1) The index of a machine contains all relevant information, the first co-ordinate tells us which clause to use, the second co-ordinate always gives the number of arguments. The remaining co-ordinates contain the specific information.

(2) R7 is very powerful, it yields an enumeration of all machines with a fixed number of arguments. Exactly the kind of machine we needed above for the diagonalization. Intuitively, the existence of such a machine seems quite reasonable. If one can effectively recognize the indices of machines,

then a machine should be able to do so, and thus to simulate each single machine.

The scrupulous might call R7 a case of cheating, since it does away with all the hard work one has to do in order to obtain a universal machine, e.g. in the case of Turing machines.

The relation  $\{e\}(\vec{x}) \simeq y$  is functional, i.e. we can show

*Fact.*  $\{e\}(\vec{x}) \simeq y, \{e\}(\vec{x}) \simeq z \Rightarrow y = z$ .

**Proof.** Use induction on the definition of  $\{e\}$ . ■

The above definition tells us implicitly what we have to consider a computation: to compute  $\{e\}(\vec{x})$  we look at  $e$ , is the first ‘entry’ of  $e$  if 0, 1, 2, then we compute the output via the corresponding initial function. If the first ‘entry’ is 3, then we determine the output ‘by cases’. First ‘entry’ 5 is handled as indicated in the  $S_n^m$  theorem. If the first entry is 4, then we first carry out the subcomputations with indices  $c_1, \dots, c_k$ , followed by the subcomputation with index  $b$ , and find the output according to R5. At first ‘entry’ 6, we jump to the subcomputation with index  $b$  (cf. R7).

In the presence of R7 we are no longer guaranteed that the process will stop; indeed, we may run into a loop, as the following simple example shows.

By R7 there exists an  $e$  such that  $\{e\}(x) = \{x\}(x)$ .

To compute  $\{e\}$  for the argument  $e$  we pass, according to R7, onto the right-hand side, i.e. we must compute  $\{e\}(e)$ , since  $e$  was introduced by R7, we must repeat the transitions to the right hand side, etc. Evidently our procedure does not get us anywhere!

Loops and non-terminating computations account for algorithms being undefined at some inputs.

There could also be a trivial reason for not producing outputs, e.g.  $\{0\}(\vec{x}) \simeq y$  holds for no  $y$ , since 0 is to an index at all, so  $\{0\}$  stands for the empty function.

Some terminology:

1. If for a partial function  $\varphi \exists y(\varphi(\vec{x}) \simeq y)$ , then we say that  $\varphi$  *converges at  $\vec{x}$* , otherwise  $\varphi$  *diverges at  $\vec{x}$* .
2. If a partial function converges for all inputs, it is called *total*.
3. A total partial recursive function (sic!) will be called a *recursive function*.
4. a set (relation) is called *recursive* if its characteristic function is recursive.

The definition of  $\{e\}(\vec{x}) \simeq y$  has the consequence that a partial recursive function diverges if one of its arguments diverges. This is an important

feature, not shared for example by  $\lambda$ -calculus or combinatory logic. It tells us that we *have* to carry out all subcomputations. We could, for instance, not assert that

$\{e\}x - \{e\}x = 0$  for all  $e$  and  $x$ , we first must show that  $\{e\}x$  converges.

This feature is sometimes inconvenient and slightly paradoxical, e.g. in direct applications of the discriminator scheme  $R4, \{\langle 3, 4 \rangle\}(\varphi(x), \psi(x), 0, 0)$  is undefined when the (seemingly irrelevant) function  $\psi(x)$  is undefined. With a bit of extra work, we can get an index for a partial recursive function that does *definition by cases* on partial recursive functions:

$$\{e\}(\vec{x}) \simeq \begin{cases} \{e_1\}(\vec{x}) & \text{if } g_1(\vec{x}) = g_2(\vec{x}) \\ \{e_2\}(\vec{x}) & \text{if } g_1(\vec{x}) \neq g_2(\vec{x}) \end{cases}$$

for recursive  $g_1, g_2$ .

Define

$$\varphi(\vec{x}) \simeq \begin{cases} e_1 & \text{if } g_1(\vec{x}) = g_2(\vec{x}) \\ e_2 & \text{if } g_1(\vec{x}) \neq g_2(\vec{x}) \end{cases}$$

by  $\varphi(\vec{x}) \simeq \{\langle 3, 4 \rangle\}(e_1, e_2, g_1(\vec{x}), g_2(\vec{x}))$ , use R5. Then an application of R7 and R5 to  $\{\varphi(\vec{x})\}(\vec{x})$  yields the desired  $\{e\}(\vec{x})$ .

We will adopt the following notational convention after Rogers' [1967] book: partial recursive functions will be denoted by  $\varphi, \psi, \dots$ , and the total ones by  $f, g, h, \dots$ . From now on we will indiscriminately use '=' for ' $\simeq$ ', and for the ordinary equality.

After some preliminary work, we will show that all primitive recursive functions are recursive. We could forget about the primitive recursive functions and just discuss partial recursive ones. However, the primitive recursive functions form a very natural class, and they play an important role in metamathematics.

The following important theorem has a neat machine motivation. Consider a machine with index  $e$  operating on two arguments  $x$  and  $y$ . Keeping  $x$  fixed, we have a machine operating on  $y$ . So we get a sequence of machines, one for each  $x$ . Does the index of each such machine depend in a decent way on  $x$ ? The plausible answer seems 'yes'. The following theorem confirms this.

**THEOREM 14** (The  $S_n^m$  Theorem). *For every  $m, n$  such that  $0 < m < n$  there exists a primitive recursive function  $S_n^m$  such that*  
 $\{S_n^m(e, x_1, \dots, x_m)\}(x_{m+1}, \dots, x_n) = \{e\}(\vec{x})$ .

**Proof.** The first function  $S_n^1$  is given by R6, we write down its explicit definition:

$$S_n^1(e, y) = \langle 4, (e)_2 \dot{-} 1, e, \langle 0, (e)_2 \dot{-} 1, 1 \rangle, \dots, \langle 1, (e)_2 \dot{-} 1, n \dot{-} \rangle \rangle.$$



Then  $\{S_n^1(e, y)\}(\vec{x}) = z \Leftrightarrow \exists q_1 \cdots q_n [\{ \langle 0, (e)_2 - 1, y \rangle \}(\vec{x}) = q_1 \wedge \{ \langle 1, (e)_2 - 1, 1 \rangle \}(\vec{x}) = q_2 \wedge \cdots \wedge \{ \langle 1, (e)_2 - 1, n - 1 \rangle \}(\vec{x}) = q_n \wedge \{e\}(q_1, \dots, q_n) = z]$ .

By the clauses R1 and R2 we get  $q_1 = y$  and  $q_{i+1} = x_i$ , so  $\{S_n^1(e, y)\}(\vec{x}) = \{e\}(y, \vec{x})$ . Clearly  $S_n^1$  is primitive recursive.

$S_n^m$  is obtained by applying  $S_n^1$   $m$  times. Note that  $S_n^m$  is primitive recursive. ■

The  $S_n^m$  theorem expresses a *uniformity* property of the partial recursive functions. It is obvious indeed that, say for a partial recursive function  $\varphi(x, y)$ , each individual  $\varphi(n, y)$  is partial recursive (substitute the constant  $n$  function for  $x$ ), but this does not yet show that the index of  $\lambda y \cdot \varphi(x, y)$  is in a systematic, uniform way computable from the index of  $\varphi$  and  $x$ , this is taken care of by the  $S_n^m$ -theorem.

There are numerous applications, we will just give one: define  $\varphi(x) = \{e\}(x) + \{f\}(x)$ , then by 20  $\varphi$  is partial recursive and we would like to express the index of  $\varphi$  as a function of  $e$  and  $f$ . Consider  $\psi(e, f, x) = \{e\}(x) + \{f\}(x)$ .  $\psi$  is partial recursive, so it has an index  $n$ , i.e.  $\{n\}(e, f, x) = \{e\}(x) + \{f\}(x)$ . By the  $S_n^m$  theorem there is a primitive recursive function  $h$  such that  $\{n\}(e, f, x) = \{h(n, e, f)\}(x)$ . Therefore,  $g(e, f) = h(n, e, f)$  is the required function.

Next we will prove a fundamental theorem about partial recursive functions that allows us to introduce partial recursive functions by inductive definitions, or by implicit definition. We have seen that we can define a primitive recursive function by using all (or some) of the preceding values to get a value in  $n$ . We might, however, just as well make the value depend on future values, only then we can no longer guarantee that the resulting function is total (let alone primitive recursive!).

EXAMPLE.

$$\varphi(n) = \begin{cases} 0 & \text{if } n \text{ is a prime, or } 0, \text{ or } 1 \\ \varphi(2n + 1) + 1 & \text{otherwise.} \end{cases}$$

Then  $\varphi(0) = \varphi(1) = \varphi(2) = \varphi(3) = 0$ ,  $\varphi(4) = \varphi(9) + 1 = \varphi(19) + 2 = 2$ ,  $\varphi(5) = 0$ , and, e.g.  $\varphi(85) = 6$ . Prima facie, we cannot say much about such a sequence. The following theorem of Kleene shows that we can always find a partial recursive solution to such an equation for  $\varphi$ .

**THEOREM 15 (The Recursion Theorem).** *There exists a primitive recursive function  $rc$  such that  $\{rc(e)\}(\vec{x}) = \{e\}(rc(e), \vec{x})$ .*

Before we prove the theorem let us convince ourselves that it solves our problem. We want a partial recursive  $\varphi$  such that  $\varphi(\vec{x}) = \{e\}(\dots \varphi \dots \vec{x})$  (where the notation is meant to indicate that  $\varphi$  occurs on the right-hand side). To ask for a partial recursive function is to ask for an index for it, so replace  $\varphi$  by  $\{z\}$ , where  $z$  is the unknown index:

$$\{z\}(\vec{x}) = \{e\}(\dots\{z\}\dots, \vec{x}) = \{e'\}(z, \vec{x}).$$

Now it is clear that  $\text{rc}(e')$  gives us the required index for  $\varphi$ .

**Proof.** Let  $\varphi(m, e, \vec{x}) = \{e\}(S_{n+2}^2(m, m, e), \vec{x})$  and let  $p$  be an index of  $\varphi$ . Put  $\text{rc}(e) = S_{n+2}^2(p, p, e)$ , then

$$\begin{aligned} \{\text{rc}(e)\}(\vec{x}) &= \{S_{n+2}^2(p, p, e)\}(\vec{x}) = \{p\}(p, e, \vec{x}) = \varphi(p, e, \vec{x}) \\ &= \{e\}(S_{n+2}^2(p, p, e), \vec{x}) = \{e\}(\text{rc}(e), \vec{x}). \end{aligned}$$

■

As a special case we get the

**COROLLARY.** *For each  $e$  there exists an  $n$  such that  $\{n\}(\vec{x}) = \{e\}(n, \vec{x})$ .*

**REMARK.** Although we have not yet shown that the class of partial recursive functions contains all primitive recursive functions, we know what primitive recursive functions are and what their closure properties are. In particular, if  $\{e\}$  should happen to be primitive recursive, then by  $\{\text{rc}(e)\}(\vec{x}) = \{e\}(\text{rc}(e), \vec{x})$ ,  $\{\text{rc}(e)\}$  is also primitive recursive.

**EXAMPLES 16.**

1. *There is a partial recursive function  $\varphi$  such that  $\varphi(n) = (\varphi(n+1)+1)^2$ : Consider  $\{z\}(n) = \{e\}(z, n) = (\{z\}(n+1) + 1)^2$ . By the recursion theorem there is a solution  $\text{rc}(e)$ , hence  $\varphi$  exists. A simple argument shows that  $\varphi$  cannot be defined for any  $n$ , so the solution is the empty function (the machine that never gives an output).*
2. *The Ackermann function, see [Smorynski, 1991], p. 70. Consider the following sequence of functions.*

$$\begin{aligned} \varphi_0(m, n) &= n + m \\ \varphi_1(m, n) &= n \cdot m \\ \varphi_2(m, n) &= n^m \\ &\vdots \\ \begin{cases} \varphi_{k+1}(0, n) = n \\ \varphi_{k+1}(m+1, n) = \varphi_k(\varphi_{k+1}(m, n), n) \quad (k \geq 2) \end{cases} \end{aligned}$$

This sequence consists of faster and faster growing functions. We can lump all those functions together in one function

$$\varphi(k, k, n) = \varphi_k(m, n).$$

The above equations can be summarized as

$$\begin{cases} \varphi(0, m, n) = n + m \\ \varphi(k + 1, 0, n) = \begin{cases} 0 & \text{if } k = 0 \\ 1 & \text{if } k = 1 \\ n & \text{else} \end{cases} \\ \varphi(k + 1, m + 1, n) = \varphi(k, \varphi(k + 1, m, n), n). \end{cases}$$

Note that the second equation has to distinguish cases according to the  $\varphi_{k+1}$  being the multiplication, exponentiation, or the general case ( $k \geq 2$ ).

Using the fact that all primitive recursive functions are recursive (Corollary 20) we rewrite the three cases into one equation of the form  $\{e\}(k, m, n) = f(e, k, m, n)$  for a suitable recursive  $f$ . Hence, by the recursion theorem there exists a recursive function with index  $e$  that satisfies the equations above. Ackermann has shown that the function  $\varphi(n, n, n)$  grows eventually faster than any primitive recursive function.

The recursion theorem can also be used for inductive definitions of sets or relations, this is seen by changing over to characteristic functions, e.g. suppose we want a relation  $R(x, y)$  such that

$$R(x, y) \leftrightarrow (x = 0 \wedge y \neq 0) \vee (x \neq 0 \wedge y \neq 0 \wedge R(x - 1, y - 1)).$$

Then we write

$$K_R(x, y) = \text{sg}(\overline{\text{sg}}(x) \cdot \text{sg}(y) + \text{sg}(x) \cdot \text{sg}(y) \cdot K_R(x - 1, y - 1)),$$

so there is an  $e$  such that

$$K_R(x, y) = \{e\}(K_R(-1, y - 1), x, y).$$

Now suppose  $K_R$  has index  $z$  then we have

$$\{z\}(x, y) = \{e'\}(z, x, y).$$

The solution  $\{n\}$  provided by the recursion theorem is the required characteristic function. One immediately sees that  $R$  is the relation ‘less than’, so  $\{n\}$  is total recursive and hence so is  $R$  (cf. 4), note that by the remark following the recursion theorem we even get the primitive recursiveness of  $R$ . The partial recursive functions are a rather wild lot, they have an enormous variety of definitions (in terms of R1–R7). We can, however, obtain them in a uniform way by one minimalization from a fixed predicate.

**THEOREM 17 (Normal Form Theorem).** *There is a primitive recursive predicate  $T$  such that  $\{e\}(\vec{x}) = ((\mu z)T(e, \langle \vec{x} \rangle, z))_1$ .*

**Proof.** See the Appendix. ■

The predicate  $T$  formalizes the statement ‘ $z$  is the computation of the partial recursive function (machine with index  $e$  operating on input  $\langle \vec{x} \rangle$ ’, where ‘computation’ has been defined such that the first projection is the output.

For applications the precise structure of  $T$  is not important. One can obtain the well-known undecidability results from the  $S_n^m$  theorem, the recursion theorem and the normal form theorem.

The partial recursive functions are closed under a general form of minimalization, sometimes called *unbounded search*, which for a given recursive function  $f(y, \vec{x})$  and arguments  $\vec{x}$  runs through the values of  $y$  and looks for the first one that makes  $f(y, \vec{x})$  equal to zero.

**THEOREM 18.** *Let  $f$  be a recursive function, then  $\varphi(\vec{x}) = \mu y[f(y, \vec{x}) = 0]$  is partial recursive.*

**Proof.** Our strategy consists of testing successively all values of  $y$  until we find the first  $y$  such that  $f(y, \vec{x}) = 0$ . We want a function  $\psi$  such that  $\psi(y, \vec{x})$  produces a 0 if  $f(y, \vec{x}) = 0$  and moves on to the next  $y$  while counting the steps if  $f(y, \vec{x}) \neq 0$ . Let this function  $\psi$  have index  $e$ . We introduce auxiliary functions  $\psi_1, \psi_2$  with indices  $b$  and  $c$  such that  $\psi_1(e, y, \vec{x}) = 0$  and  $\psi_2(e, y, \vec{x}) = \psi(y + 1, \vec{x}) + 1 = \{e\}(y + 1, \vec{x}) + 1$ . If  $f(y, \vec{x}) = 0$  then we consider  $\psi_1$ , if not,  $\psi_2$ . So we introduce, by clause R4, a new function  $\chi_0$ :

$$\chi_0(e, y, \vec{x}) = \begin{cases} b & \text{if } f(y, \vec{x}) = 0 \\ c & \text{else} \end{cases}$$

and we put  $\chi(e, y, \vec{x}) = \{\chi_0(e, y, \vec{x})\}(e, y, \vec{x})$ .

The recursion theorem provides us with an index  $e_0$  such that  $\chi(e_0, y, \vec{x}) = \{e_0\}(y, \vec{x})$ .

We claim that  $\{e_0\}(0, \vec{x})$  yields the desired value, if it exists at all, i.e.  $e_0$  is the index of the  $\psi$  we were looking for.

For, if  $f(y, \vec{x}) \neq 0$  then  $\chi(e_0, y, \vec{x}) = \{c\}(e_0, y, \vec{x}) = \psi_2(e_0, y, \vec{x}) = \psi(y + 1, \vec{x}) + 1$ , and if  $f(y, \vec{x}) = 0$  then  $\chi(e_0, y, \vec{x}) = \{b\}(e_0, y, \vec{x}) = 0$ .

So suppose that  $y_0$  is the first value  $y$  such that  $f(y, \vec{x}) = 0$ , then

$$\psi(0, \vec{x}) = \psi(1, \vec{x}) + 1 = \psi(2, \vec{x}) + 2 = \dots = \Psi(y_0, \vec{x}) + y_0 = y_0.$$

■

Note that the given function need not be recursive, and that the above argument also works for partial recursive  $f$ . We then have to reformulate  $\mu y[f(x, \vec{y}) = 0]$  as the  $y$  such that  $f(y, \vec{x}) = 0$  and for all  $z < y$   $f(z, \vec{x})$  is defined and positive.

We need minimalization in our approach to obtain closure under primitive recursion. We could just as well have thrown in an extra clause for primitive recursion, (and deleted R4 and R6), but that would have obscured the power

of the reflection clause R7. Observe that in order to get closure under primitive recursion, we need a simple consequence of it, namely R6.

It is easy to see that the predecessor function,  $x \dot{-} 1$ , can be obtained:

$$\text{define } x \dot{-} 1 = \begin{cases} 0 & \text{if } x = 0 \\ \mu y[y + 1 = x] & \text{else} \end{cases}$$

where  $\mu y[y + 1 = x] = \mu y[f(y, x) = 0]$  with

$$f(y, x) = \begin{cases} 0 & \text{if } y + 1 = x \\ 1 & \text{else} \end{cases}$$

**THEOREM 19.** *The recursive functions are closed under primitive recursion.*

**Proof.** We want to show that if  $g$  and  $h$  are recursive, then so is  $f$ , defined by

$$\begin{cases} f(0, \vec{x}) = g(\vec{x}) \\ f(y + 1, \vec{x}) = h(f(y, \vec{x}), \vec{x}, y). \end{cases}$$

We rewrite the schema as

$$f(y, \vec{x}) = \begin{cases} g(\vec{x}) & \text{if } y = 0 \\ h(f(y \dot{-} 1, \vec{x}), \vec{x}, y \dot{-} 1) & \text{otherwise.} \end{cases}$$

Since the predecessor is recursive, an application of definition by cases yields the following equation for an index of the function  $f : \{e\}(y, \vec{x}) = \{a\}(y, \vec{x}, e)$  (where  $a$  can be computed from the indices of  $g, h$  and the predecessor). By the recursion theorem the equation has a solution  $e_0$ . One shows by induction on  $y$  that  $\{e_0\}$  is total, so  $f$  is a recursive function. ■

We now get the obligatory

**COROLLARY 20.** *All primitive recursive functions are recursive.*

**DEFINITION 21.**

1. A set (relation) is (recursively) *decidable* if it is recursive.
2. A set is *recursively enumerable* (RE) if it is the domain of a partial recursive function.
3.  $W_e^k = \{\vec{x} \in \mathbb{N}^k \mid \exists y(\{e\}(\vec{x}) = y)\}$ , i.e. the domain of the partial recursive function  $\{e\}$ . We call  $e$  the RE index of  $W_e^k$ . If no confusion arises we will delete the superscript.

We write  $\varphi(\vec{x}) \downarrow$  (resp.  $\varphi(\vec{x}) \uparrow$ ) for  $\varphi(\vec{x})$  converges (resp.  $\varphi$  diverges).

One can think of a recursively enumerable set as a set that is accepted by an abstract machine; one successively offers the natural numbers, 0, 1, 2, ..., and when the machine produces an output the input is 'accepted'.

The next theorem states that we could also have defined RE sets as those produced by a machine.

It is good heuristics to think of RE sets as being accepted by machines, e.g. if  $A_i$  is accepted by machine  $M_i$  ( $i = 0, 1$ ), then we make a new machine that simulates  $M_0$  and  $M_1$  running parallel, and so  $n$  is accepted by  $M$  if it is accepted by  $M_0$  or  $M_1$ . Hence the union of two RE sets is also RE.

EXAMPLES 22 (of RE sets).

1.  $\mathbb{N} =$  the domain of the constant function.
2.  $\emptyset =$  the domain of the empty function. This function is partial recursive, as we have already seen.
3. Every recursive set is RE. Let  $A$  be recursive, put

$$\psi(\vec{x}) = \mu y [K_A(\vec{x}) = y \wedge y \neq 0]$$

Then  $Dom(\psi) = A$ .

The recursively enumerable sets derive their importance from the fact that they are effectively given, in the sense that they are produced by partial recursive functions, i.e. they are presented by an algorithm. Furthermore it is the case that the majority of important relations (sets) in logic are RE. For example the set of provable sentences of arithmetic or predicate logic is RE. The RE sets represent the first step beyond the decidable sets, as we will show below.

THEOREM 23. The following statements are equivalent, ( $A \subseteq \mathbb{N}$ ):

1.  $A = Dom(\varphi)$  for some partial recursive  $\varphi$ ,
2.  $A = Ran(\varphi)$  for some partial recursive  $\varphi$ ,
3.  $A = \{x \mid \exists y R(x, y)\}$  for some recursive  $R$ .

**Proof.** (1)  $\Rightarrow$  (2). Define  $\psi(x) = x \cdot \text{sg}(\varphi(x) + 1)$ . If  $x \in Dom(\varphi)$ , then  $\psi(x) = x$ , so  $x \in Ran(\psi)$ , and if  $x \in Ran(\psi)$ , then  $\varphi(x) \downarrow$ , so  $x \in Dom(\varphi)$ .

(2)  $\Rightarrow$  (3) Let  $A = Ran(\{g\})$  then

$$x \in A \leftrightarrow \exists w [T(g, (w)_1, (w)_2) \wedge x = (w)_{2,1}].$$

The relation in the scope of the quantifier is recursive.

Note that  $w$  acts as a pair: first co-ordinate—input, second co-ordinate—computation.

(3)  $\Rightarrow$  (1) Define  $\varphi(x) = \mu y R(x, y)$ .  $\varphi$  is partial recursive and  $Dom(\varphi) = A$ .

Observe that (1)  $\Rightarrow$  (3) also holds for  $A \subseteq \mathbb{N}^k$ . ■

Since we have defined recursive sets by means of characteristic functions, and since we have established closure under primitive recursion, we can copy all the closure properties of primitive recursive sets (and relations) for the recursive sets (and relations).

Next we list a number of closure properties of RE-sets.

**THEOREM 24.**

1. If  $A$  and  $B$  are RE, then so are  $A \cup B$  and  $A \cap B$
2. If  $R(x, \vec{y})$  is RE, then so is  $\exists xR(x, \vec{y})$
3. If  $R(x, \vec{y})$  is RE and  $\varphi$  partial recursive, then  $R(\varphi(\vec{y}, \vec{z}), \vec{y})$  is RE
4. If  $R(x, \vec{y})$  is RE, then so are  $\forall x < zR(x, \vec{y})$  and  $\exists x < zR(x, \vec{y})$ .

**Proof.**

1. There are recursive  $R$  and  $S$  such that

$$A\vec{y} \leftrightarrow \exists xR(x, \vec{y}), B\vec{y} \leftrightarrow \exists xS(x, \vec{y}).$$

Then

$$\begin{aligned} A\vec{y} \wedge B\vec{y} &\leftrightarrow \exists x_1x_2(R(x_1, \vec{y}) \wedge S(x_2, \vec{y})) \\ &\leftrightarrow \exists z(R((z)_1, \vec{y}) \wedge S((z)_2, \vec{y})). \end{aligned}$$

The relation in the scope of the quantifier is recursive, so  $A \cap B$  is RE. A similar argument establishes the recursive enumerability of  $A \cup B$ . The trick of replacing  $x_1$  and  $x_2$  by  $(z)_1$  and  $(z)_2$  and  $\exists x_1x_2$  by  $\exists z$  is called *contraction of quantifiers*.

2. Let  $R(x, \vec{y}) \leftrightarrow \exists zS(z, x, \vec{y})$  for a recursive  $S$ , then  $\exists xR(x, \vec{y}) \leftrightarrow \exists x\exists zS(z, x, \vec{y}) \leftrightarrow \exists uS((u)_1, (u)_2, \vec{y})$ . So the *projection*  $\exists xR(x, \vec{y})$  of  $R$  is RE.

$\exists xR(x, \vec{y})$  is indeed a projection. Consider the two-dimensional case (Figure 4).

The vertical projection  $S$  of  $R$  is given by  $Sx \leftrightarrow \exists yR(x, y)$ .

3. Let  $R$  be the domain of a partial recursive  $\psi$ , then  $R(\varphi(\vec{y}, \vec{z}), \vec{y})$  is the domain of  $\psi(\varphi(\vec{y}, \vec{z}), \vec{y})$ .
4. Left to the reader. ■

**THEOREM 25.** *The graph of a partial function is RE iff the function is partial recursive.*

**Proof.**  $G = \{(\vec{x}, y) \mid y = \{e\}(\vec{x})\}$  is the graph of  $\{e\}$ . Now  $(\vec{x}, y) \in G \Leftrightarrow \exists z(T(e, \langle \vec{x}, z \rangle) \wedge y = (z)_1)$ , so  $G$  is RE. Conversely, if  $G$  is RE, then  $G(\vec{x}, y) \Leftrightarrow \exists zR(\vec{x}, y, z)$  for some recursive  $R$ . Hence  $\varphi(\vec{x}) = (\mu wR(\vec{x}, (w)_1, (w)_2))_1$ , so  $\varphi$  is partial recursive. ■

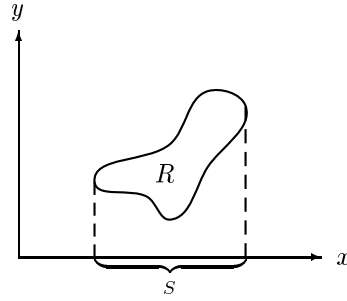


Figure 4.

We can also characterize sets in terms of RE-sets. Suppose both  $A$  and its complement  $A^c$  are RE, then (heuristically) we have two machines enumerating  $A$  and  $A^c$ . Now the test for membership of  $A$  is simple: turn both machines on and wait for  $n$  to turn up as output of the first or second machine. This must necessarily occur in finitely many steps since  $n \in A$  or  $n \in A^c$  (principle of the excluded third). Hence, we have an effective test. We formalize the above:

**THEOREM 26.**  $A$  is recursive  $\Leftrightarrow A$  and  $A^c$  are RE.

**Proof.**  $\Rightarrow$  is trivial,  $A(\vec{x}) \leftrightarrow \exists y A(\vec{x})$ , where  $y$  is a dummy variable. Similarly for  $A^c$ .

$\Leftarrow$  Let  $A(\vec{x}) \leftrightarrow \exists y R(\vec{x}, y)$ ,  $\neg A(\vec{x}) \leftrightarrow \exists z (S(\vec{x}, z))$ . Since  $\forall \vec{x} (A(\vec{x}) \vee \neg A(\vec{x}))$ , we have  $\forall \vec{x} \exists y (R(\vec{x}, y) \vee S(\vec{x}, y))$ , so  $f(\vec{x}) = \mu y [R(\vec{x}, y) \vee S(\vec{x}, y)]$  is recursive and if we plug the  $y$  that we found in  $R(\vec{x}, y)$ , then we know that if  $R(\vec{x}, f(\vec{x}))$  is true, the  $\vec{x}$  belongs to  $A$ . So  $A(\vec{x}) \leftrightarrow R(\vec{x}, f(\vec{x}))$ , i.e.  $A$  is recursive. ■

For partial recursive functions we have a strong form of definition by cases:

**THEOREM 27.** Let  $\psi_1, \dots, \psi_k$  be partial recursive,  $R_1, \dots, R_k$  mutually disjoint RE-relations, then

$$\varphi(\vec{x}) = \begin{cases} \psi_1(\vec{x}) & \text{if } R_1(\vec{x}) \\ \psi_2(\vec{x}) & \text{if } R_2(\vec{x}) \\ \vdots & \\ \psi_k(\vec{x}) & \text{if } R_k(\vec{x}) \\ \uparrow & \text{else} \end{cases}$$

is partial recursive.

**Proof.** We consider the graph of the function  $\varphi$ .

$$G(\vec{x}, y) \leftrightarrow (R_1(\vec{x}) \wedge y = \psi_1(\vec{x})) \vee \dots \vee (R_k(\vec{x}) \wedge y = \psi_k(\vec{x})).$$



By the properties of RE-sets,  $G(\vec{x}, y)$  is RE and, hence,  $\varphi(\vec{x})$  is partial recursive. ■

Note that the last case in the definition is just a bit of decoration.

Now we can show the existence of undecidable RE sets.

**PROBLEM 28** (The Halting Problem (A. Turing)). (1) Consider  $K = \{x \mid \exists zT(x, x, z)\}$ .  $K$  is the projection of a recursive relation, so it is RE. Suppose that  $K^c$  is also RE, then  $x \in K^c \leftrightarrow \exists zT(e, x, z)$  for some index  $e$ . Now  $e \in K \leftrightarrow \exists zT(e, e, z) \leftrightarrow e \in K^c$ . Contradiction. Hence  $K$  is not recursive by the above theorem.

The decision problem for  $K$  is called the *halting problem*, because it can be paraphrased as ‘decide if the machine with index  $x$  performs a computation that halts after a finite number of steps when presented with  $x$  as input. Note that it is *ipso facto* undecidable if ‘the machine with index  $x$  eventually halts on input  $y$ ’.

We will exhibit a few more examples of undecidable problems.

(2) It is not decidable if  $\{x\}$  is a total function.

Suppose it were decidable, then we would have a recursive function  $f$  such that  $f(x) = 0 \leftrightarrow \{x\}$  is total.

Now consider

$$\varphi(x, y) := \begin{cases} 0 & \text{if } x \in K \\ \uparrow & \text{else} \end{cases}$$

By the  $S_n^m$  theorem there is a recursive  $h$  such that  $\{h(x)\}(y) = \varphi(x, y)$ . Now  $\{h(x)\}$  is total  $\leftrightarrow x \in K$ , so for  $f(h(x)) = 0 \leftrightarrow x \in K$ , i.e. we have a recursive characteristic function  $\text{sg}(f(h(x)))$  for  $K$ . Contradiction. Hence such an  $f$  does not exist, that is  $\{x \mid \{x\} \text{ is total}\}$  is not recursive.

(3) The problem ‘ $W_e$  is finite’ is not recursively solvable. Suppose that there was a recursive function  $f$  such that  $f(e) = 0 \leftrightarrow W_e$  is finite.

Consider the  $h(x)$  defined in example (2). Clearly  $W_{h(x)} = \text{Dom}\{h(x)\} = \emptyset \leftrightarrow x \notin K$ , and  $W_{h(x)}$  is infinite for  $x \in K$ .  $f(h(x)) = 0 \leftrightarrow x \notin K$ , and hence  $\text{sg}(f(h(x)))$  is a recursive characteristic function for  $K$ . Contradiction.

Note that  $x \in K \leftrightarrow \{x\}x \downarrow$ , so we can reformulate the above solutions as follows: in (2) take  $\varphi(x, y) = 0$ .  $\{x\}(x)$  and in (3)  $\varphi(x, y) = \{x\}(x)$ .

(4) The equality of RE sets is undecidable, i.e.  $\{(x, y) \mid W_x = W_y\}$  is not recursive. We reduce the problem to the solution of (3) by choosing  $W_y = \emptyset$ .

(5) It is not decidable if  $W_e$  is recursive. Put  $\varphi(x, y) = \{x\}(x) \cdot \{y\}(y)$ , then  $\varphi(x, y) = \{h(x)\}(y)$  for a certain recursive  $h$ , and

$$\text{Dom}\{h(x)\} = \begin{cases} K & \text{if } x \in K \\ \emptyset & \text{otherwise.} \end{cases}$$

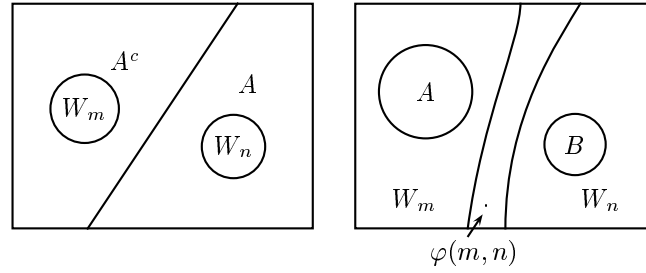


Figure 5.

Suppose there were a recursive function  $f$  such that  $f(x) = 0 \leftrightarrow W_x$  is recursive, then  $f(h(x)) = 0 \leftrightarrow x \notin K$  and, hence,  $K$  would be recursive. Contradiction.

There are several more techniques for establishing undecidability. We will consider the method of inseparability.

**DEFINITION 29.** Two disjoint RE-sets  $W_m$  and  $W_n$  are *recursively separable* (Figure 5) if there is a recursive set  $A$  such that  $W_n \subseteq A$  and  $W_m \subseteq A^c$ . Disjoint sets  $A$  and  $B$  are *effectively inseparable* if there is a partial recursive  $\varphi$  such that for every  $m, n$  with  $A \subseteq W_m, B \subseteq W_n, W_m \cap W_n = \emptyset$  we have  $\varphi(m, n) \downarrow$  and  $\varphi(m, n) \notin W_m \cup W_n$ .

We immediately see that effectively inseparable RE sets are recursively inseparable, i.e. not recursively separable.

**THEOREM 30.** *There exist effectively inseparable RE sets.*

**Proof.** Define  $A = \{x \mid \{x\}(x) = 0\}, B = \{x \mid \{x\}(x) = 1\}$ . Clearly  $A \cap B = \emptyset$  and both are RE.

Let  $W_m \cap W_n = \emptyset$  and  $A \subseteq W_m, B \subseteq W_n$ . To define  $\varphi$  we start testing  $x \in W_m$  or  $x \in W_n$ , if we first find  $x \in W_m$ , then we put an auxiliary function  $\sigma(x)$  equal to 1, if  $x$  turns up first in  $W_n$  then we put  $\sigma(x) = 0$ .

Formally

$$\sigma(m, n, x) = \begin{cases} 1 & \text{if } \exists z(T(m, x, z) \text{ and } \forall y < z \neg T(n, x, y)) \\ 0 & \text{if } \exists z(T(n, x, z) \text{ and } \forall y \leq z \neg T(m, x, y)) \\ \uparrow & \text{else.} \end{cases}$$

By the  $S_n^m$  theorem  $\{h(m, n)\}(x) = \sigma(m, n, x)$  for some recursive  $h$ .

Now

$$\begin{aligned} h(m, n) \in W_m &\Rightarrow h(m, n) \notin W_n. \text{ So } \exists z(T(m, h(m, n), z) \text{ and } \\ &\quad \forall y < z \neg T(n, h(m, n), y)) \\ &\Rightarrow \sigma(m, n, h(m, n)) = 1 \Rightarrow \{h(m, n)\}(h(m, n)) = 1 \\ &\Rightarrow h(m, n) \in B \Rightarrow h(m, n) \in W_n. \end{aligned}$$

Contradiction. Hence  $h(m, n) \notin W_m$ . Similarly  $h(m, n) \notin W_n$ . Thus  $h$  is the required  $\varphi$ . ■

As a corollary we find that  $A^c$  is *productive*, i.e. there is a partial recursive  $\psi$  such that for each  $W_k \subseteq A^c$  we have  $\psi(k) \in A^c - W_k$ . Simply take in the above proof  $W_{m_0} = A$  and  $W_n = B \cup W_k$ .

Using the simple fact that there is a recursive  $f$  such that  $W_x \cup W_y = W_{f(x,y)}$ , we find a recursive  $g$  such that  $B \cup W_k = W_{g(k)}$ . Putting  $\psi(k) = \varphi(m_0, g(k))$  ( $\varphi$  as defined in 30), we find the desired production function:  $\psi(k) \in A^c - W_k$ .

Such a productive set is in a strong sense not RE: if one tries to fit in an RE set then one can uniformly and effectively indicate a point that eludes this RE set.

## 2.1 Relative Recursiveness

Following Turing we can widen the scope of computability (recursiveness) a bit, by allowing one (or finitely many) functions to be added to the initial functions, e.g. let  $f$  be such an extra initial function, then definition 13 yields a wider class of partial functions. We call these functions ‘*recursive in  $f$* ’, and we may think of them as being computable when  $f$  is given beforehand as an *oracle*. The theory of this section can be carried through for the new concept, just replace ‘recursive’ or ‘RE’ by ‘recursive in  $f$ ’ or ‘RE in  $f$ ’.

The notion of ‘recursive in’ is particularly interesting when applied to (characteristic functions of) sets. We say that  $A$  is Turing reducible to  $B$  (notation  $A \leq_T B$ ) if  $K_A$  is recursive in  $K_B$ .  $K_A$  stands for a membership test for  $A$ , so  $A \leq_T B$  means that there is an algorithm such that we can test  $n \in A$  by applying the algorithm to the (given) membership test for  $B$  (i.e.  $K_B$ ).

By assigning an index to  $K_B$ , we can write this as  $K_A(x) = \{e\}^B(x)$ , where  $e$  is computed as before. The superscript  $B$  (or in general  $f$ ) is added to indicate the dependence on  $B$  (or  $f$ ).

It is not terribly difficult to show that in computing  $\{e\}^B(x)$  we can only use a finite part of the function  $K_B$ . Heuristically this means that in order to test  $n \in A$  we carry out an algorithm while during the computation we may ask finitely many questions to the oracle  $K_B$  (or  $B$ ).

It is easily seen that  $\leq_T$  is transitive, but not a partial order. Since  $A \leq_T B$  means roughly that  $A$  is, from a recursive viewpoint, less complicated than  $B$ ,  $A \leq_T B \wedge B \leq_T A$  means that  $A$  and  $B$  are equally complicated. Thus we introduce the relation  $=_T$ :  $A =_T B := A \leq_T B \wedge B \leq_T A$ . It can be shown to be an equivalence relation, the equivalence classes are called *degrees of unsolvability* or *Turing degrees*, cf. [Shoenfield, 1971].

## 2.2 Church's Thesis

Are there more algorithms than just the recursive ones? This question has never been settled, partly due to the nature of the problem. The same question for the primitive recursive functions has been answered positively. We have been able to 'diagonalize out of the class of primitive recursive functions' in an effective way. The same procedure does not work for the recursive functions, since there is no effective (i.e. recursive) way to enumerate them.

If one accepts the fact that the initial functions are algorithmic, and that the closure under substitution and reflection leads from algorithms to algorithms, then there is an inductive proof that all recursive functions are algorithms. Or, if one takes partial functions into consideration, that all partial recursive functions are algorithmic.

The converse poses the real hard question: are all algorithms recursive, or in a negative form: are there any non-recursive algorithms?

The exhibition of a non-recursive algorithm would settle the problem in the negative. A positive solution would require an exact characterization of the class of algorithms, something that is lacking. Actually the partial recursive functions have been introduced precisely for this purpose. To put it succinctly: an algorithm is a function that we recognize as effectively computable. So there is on the one hand the mathematically precise notion of a *partial recursive function* and on the other hand the anthropological, subjective notion of an algorithm.

In 1936 Alonzo Church proposed to identify the two notions, a proposal that since has become known as *Church's Thesis*: A (number theoretic) function is algorithmic if and only if it is recursive. A similar proposal was made by Turing, hence one sometimes speaks of the *Church–Turing Thesis*.

There are a number of arguments that support Church's Thesis.

(1) *A pragmatic argument: all known algorithms are recursive.* As a matter of fact, the search for non-recursive algorithms has not yielded any result. The long experience in the subject has led to acceptance for all practical purposes of the thesis by all who have practised the art of recursion theory. This has led to a tradition of 'proof by Church's Thesis', cf. [Rogers, 1967], which takes the following form: one convinces oneself by any means whatsoever that a certain function is computable and then jumps to the conclusion that it is (partial) recursive. Similarly, for 'effectively enumerable' and 'RE'.

We will demonstrate a 'proof by Church's Thesis' in the following

EXAMPLE. Each infinite RE set contains an infinite recursive set.

**Proof.** Let  $A$  be infinite RE. We list the elements of  $A$  effectively,  $n_0, n_1, n_2, n_3, \dots$

From this list we extract an increasing sublist: put  $m_0 = n_0$ , after finitely many steps we find an  $n_k$  such that  $n_k > n_0$ , put  $m_1 = n_k$ . We repeat this procedure to find  $m_2 > m_1$ , etc. this yields an effective listing of the subset  $B = \{m_0, m_1, m_2, \dots\}$  of  $A$ , with the property  $m_i < m_{i+1}$ .

*Claim.*  $B$  is decidable. For, in order to test  $k \in B$  we must check if  $k = m_i$  for some  $i$ . Since the sequence of  $m_i$ 's is increasing we have to produce at most  $k + 1$  elements of the list and compare them with  $k$ . If none of them is equal to  $k$ , then  $k \notin B$ . Since this test is effective,  $B$  is decidable and, by Church's Thesis, recursive. ■

This practice is not quite above board, but it is very convenient, and most experienced recursion theorists adhere to it.

(2) A conceptual analysis of the *notion of computability*. An impressive specimen is to be found in Alan Turing's fundamental paper [1936], also cf. [Kleene, 1952]. Turing has broken down the human computational procedures in elementary steps formulated in terms of abstract computers, the so-called *Turing machines*.

Robin Gandy has pursued the line of Turing's analysis in his paper 'Church's thesis and principles for mechanisms' [Gandy, 1980], which contains a list of four principles that underlie, so to speak, the conceptual justification of Church's Thesis.

(3) A stability argument: all the codifications of the notion of computability that have been put forward (by, e.g. Gödel–Herbrand, Church, Curry, Turing, Markov, Post, Minsky, Shepherdson–Sturgis) have been shown to be equivalent. Although, as Kreisel pointed out, this does not rule out a systematic mistake, it does carry some weight as a heuristic argument: the existence of a large number of independent but equivalent formulations of the same notion tends to underline the naturalness of the notion.

The algorithms referred to in Church's Thesis must be 'mechanical' in nature, i.e. they should not require any creativity or inventiveness on the part of the human performer. The points to be kept in mind; in the chapter on intuitionistic logic [Volume 7 of this *Handbook*] we will return to it.

One particular consequence of Church's thesis has come to light in the recent literature. In order to appreciate the phenomenon, one has to take into account the constructive meaning of the 'there exists'. That is to say, one has to adopt a constructive logic in order to obtain a formal version of Church's thesis.

For intuitionists the proof interpretation explains  $\forall x \exists y \varphi(x, y)$  as 'there exists an algorithm  $f$  such that  $\forall x \varphi(x, f(x))$ '. There are a few sophisticated conditions that must be observed, but for natural numbers there is no problem:

$$\forall x \in \mathbb{N} \exists y \in \mathbb{N} \varphi(x, y) \rightarrow \exists f \in \mathbb{N}^{\mathbb{N}} \forall x \in \mathbb{N} \varphi(x, f(x))$$

Since  $f$  has to be lawlike, it is an algorithm in the broadest sense, and on the basis of Church's thesis  $f$  must be recursive. This gives us a means to formulate Church's thesis in arithmetic (in intuitionistic arithmetic, **HA**, to be precise):

$$CT_0 \quad \forall x \exists y \varphi(x, y) \rightarrow \exists e \forall x \varphi(x, \{e\}(x))$$

The totality of  $\{e\}$  is implicit in this formulation.  $CT_0$  tells us in particular that all number theoretic functions are recursive.

Kleene, by means of his realizability interpretation, has shown that **HA** +  $CT_0$  is consistent, so it is allowed to assume Church's thesis in the context of intuitionistic arithmetic. In the eighties, the position of CT was further clarified, when it was shown independently by David McCarty and M. Hyland that there are models for higher-order intuitionistic logic (including arithmetic) in which Church's thesis holds, hence the above result was not a mere freak of first-order logic. McCarty employed an amalgamation of Kleene's realizability and von Neumann's cumulative hierarchy for set theory. Hyland constructed a particular category which acts as a higher-order intuitionistic universe in which Church's thesis holds, the so-called *effective topos*, cf. [McCarty, 1986; Hyland, 1982].

McCarty has explored the consequences of Church's thesis in a series of papers. We will mention just two facts here:

(a). Intuitionistic arithmetic has no non-standard models.

The proof runs roughly as follows: Suppose that  $\mathcal{M}$  is a non-standard model of **HA**, then the standard numbers form, exactly as in classical arithmetic, an initial segment of  $\mathcal{M}$ . Let  $a$  be a non-standard element of  $\mathcal{M}$ . Consider the two recursively inseparable RE sets  $A$  and  $B$  of theorem 18. The  $\Sigma_1^0$  formulas  $\varphi(x)$  and  $\psi(x)$  represent  $A$  and  $B$ . It is routine exercise to show that **HA**  $\vdash \forall x \neg \neg \forall y < x (\varphi(y) \vee \neg \varphi(y))$ , and hence  $\mathcal{M} \models \neg \neg \forall y < a (\varphi(y) \vee \neg \varphi(y))$ .

Assume for the sake of argument that  $\mathcal{M} \models \forall y < a (\varphi(y) \vee \neg \varphi(y))$ . Since  $a$  is preceded by all standard numbers,  $\mathcal{M} \models \varphi(n)$  or  $\mathcal{M} \models \neg \varphi(n)$  for all standard  $n$ . Define a 0 – 1 function  $f$  so that  $f(n) = 0 \Leftrightarrow \mathcal{M} \models \varphi(n)$ . By  $CT_0$   $f$  is recursive, moreover it is the characteristic function of a recursive set which separates the standard extensions of  $\varphi(x)$  and  $\psi(x)$ , i.e.  $A$  and  $B$ , contradiction. This shows that  $\mathcal{M}$  cannot be a non-standard model. The technique of the proof goes back to Tenenbaum

(b). Validity for **IQC** (intuitionistic predicate logic) is non-arithmetic, [McCarty, 1986]. The fact goes back to Kreisel (cf. [van Dalen, 1973]; McCarty's proof is an improvement both in elegance and length.

One should also keep in mind that the notion of computability that is under discussion here is an abstract one. Matters of feasibility are not relevant to Church's Thesis, but they *are* of basic interest to theoretical computer

scientists. In particular, the time (or tape) complexity has become a subject of considerable importance. Computations in ‘polynomial time’ are still acceptable from a practical point of view. Unfortunately, many important decision methods (algorithms) require exponential time (or worse), cf. [Börger, 1989; Papadimitriou, 1994].

There is a constructive and a non-constructive approach to the notion of recursiveness. There seems little doubt that the proper framework for a theory of (abstract) computability is the constructive one. Let us illustrate an anomaly of the non-constructive approach: there is a partial recursive function with at most one output, that is the Gödel number of the name of the President of American in office on the first of January of the year 2050, if there is such a president, and which diverges otherwise. This may seem surprising; is the future fully determined? A moments reflection shows that the above statement is a cheap, magician’s trick: consider the empty function and all constant functions (we can even bound the number by putting a bound on the possible length of the name of the future president). Exactly one of those partial (recursive) functions is the required one, we don’t know *which one*, but a repeated application of the principle of the excluded third proves the statement. Here is another one: consider some unsolved problem  $P$  (e.g. the Riemann hypothesis)—there is a recursive function  $f$  such that  $f$  has (constant) output 1 if  $P$  holds and 0 if  $P$  is false. Solution: consider the constant 0 and 1 functions  $f_0$  and  $f_1$ . Since  $P \vee \neg P$  holds (classically) either  $f_1$  or  $f_0$  is the required recursive function.

Constructively viewed, the above examples are defective, since the principle of the excluded third is constructively false (cf. the chapter on intuitionistic logic [see Volume 7 of this *Handbook*]). The constructive reading of ‘there exists a partial recursive function  $\varphi$ ’ is: we can effectively compute an index  $e$ . Rózsa Péter has used the constructive reading of recursion theory as an argument for the circularity of the notion of recursiveness, when based on Church’s Thesis, [Péter, 1959]. The circularity is, however, specious. Recursive functions are not used for computing single numbers, but to yield outputs for given inputs. The computation of isolated discrete objects precedes the manipulations of recursive functions, it is one of the basic activities of constructivism. So the computation of an index of a partial recursive function does itself not need recursive functions.

The notion of a recursive function has received much attention. Historically speaking, its emergence is an event of the first order. It is another example where an existing notion was successfully captured by a precise mathematical notion.

### 3 APPLICATIONS

It is no exaggeration to say that recursion theory was conceived for the sake of the study of arithmetic. Gödel used the machinery of recursion theory to show that theories containing a sufficient portion of arithmetic are incomplete. Subsequent research showed that arithmetic is undecidable, and many more theories to boot. The book [Smorynski, 1991] is an excellent source on arithmetic.

We will briefly sketch some of the methods and results.

#### 3.1 Formal Arithmetic

The first-order theory of arithmetic, **PA** (Peano's arithmetic), has a language with  $S, +, \cdot$  and  $0$ .

Its axioms are

$$\begin{array}{ll} Sx \neq 0 & x + Sy = S(x + y) \\ Sx = Sy \rightarrow x = y & x \cdot 0 = 0 \\ x + 0 = x & x \cdot Sy = x \cdot y + x \\ & \varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(Sx)) \rightarrow \forall x\varphi(x). \\ & \text{(the induction schema).} \end{array}$$

In **PA** we can define the order relation:  $x < y := \exists z(x + Sz = y)$  and we can prove its properties:  $\neg(x < 0), x < Sy \leftrightarrow x < y \vee x = y, x < y \vee x = y \vee y < x, x < y \wedge y < z \rightarrow x < z$ , by induction. The individual natural number symbols are defined by

$$1 = S0, \quad 2 = SS0, \quad 3 = SSS0, \dots$$

R. Robinson introduced a finitely axiomatized subsystem **Q** of **PA** with the schema of induction replaced by one axiom:

$$x \neq 0 \rightarrow \exists y(x = Sy).$$

Another finitely axiomatized subsystem, **N**, of **PA** was introduced by Shoenfield. This system has  $<$  as a primitive symbol, and the schema of induction is replaced by the axioms  $\neg(x < 0), x < Sy \rightarrow x < y \vee x = y, x < y \vee x = y \vee y < x$ .

#### 3.2 Arithmetization

One can code the expressions of arithmetic as natural numbers in such a way that the relevant syntactical properties become primitive recursive predicates of the codes.

There are many ways to carry out the actual coding. Unfortunately this part of the theory is strongly 'coordinate dependent', i.e. it depends on the



underlying coding of finite sequences of natural numbers. Canonical codings have been proposed at various points, cf. [Jeroslow, 1972], but there has always remained a residue of arbitrariness. We will sketch a coding based on the coding of Examples 9. Following the tradition, we will call the codes *Gödel numbers*.

1. We assign Gödel numbers to the symbols of the alphabet.

$$\begin{aligned} x_i &\mapsto 2i, 0 \mapsto 1, \vee \mapsto 3, \neg \mapsto 5, \exists \mapsto 7, S \mapsto 9, + \mapsto 11, \cdot \mapsto 13, \\ &= \mapsto 15, (< \mapsto 17, \text{ if considering } \mathbf{N}) \end{aligned}$$

2. The Gödel numbers of terms are defined by

$$\begin{aligned} \ulcorner x_i \urcorner &= \langle 2i \rangle, \ulcorner 0 \urcorner = \langle 1 \rangle, \ulcorner St \urcorner = \langle 0, \ulcorner t \urcorner \rangle, \ulcorner (t + s) \urcorner = \langle 11, \ulcorner t \urcorner, \ulcorner s \urcorner \rangle, \\ \ulcorner t \cdot s \urcorner &= \langle 13, \ulcorner t \urcorner, \ulcorner s \urcorner \rangle \end{aligned}$$

3. The Gödel numbers of formulas are defined by

$$\begin{aligned} \ulcorner (t = s) \urcorner &= \langle 15, \ulcorner t \urcorner, \ulcorner s \urcorner \rangle, \ulcorner (\varphi \vee \psi) \urcorner = \langle 3, \ulcorner \varphi \urcorner, \ulcorner \psi \urcorner \rangle, \ulcorner \neg \varphi \urcorner = \langle 5, \ulcorner \varphi \urcorner \rangle, \\ \ulcorner (\exists x_i \varphi) \urcorner &= \langle 7, \ulcorner x_i \urcorner, \ulcorner \varphi \urcorner \rangle. \end{aligned}$$

The above functions that assign Gödel numbers to expressions are defined by recursion, and one would like to know if, e.g. the set of Gödel numbers of formulas is decidable. The following lemmas provide answers.

1. **Lemma** *The following predicates are primitive recursive;*

- (a)  *$n$  is the Gödel number of a variable*
- (b)  *$n$  is the Gödel number of a term*
- (c)  *$n$  is the Gödel number of a formula.*

**Proof.** We will write down the predicate in a suitable form such that by means of the usual closure properties the reader can immediately conclude the primitive recursiveness.

- (a)  $Vble(n) \leftrightarrow \exists x \leq n (n = \langle 2x \rangle)$
- (b)  $Term(n) \leftrightarrow [Vble(n) \vee n = \langle 1 \rangle \vee$   
 $\exists x < n (n = \langle 0, x \rangle) \wedge Term(x)] \vee$   
 $\exists x, y < n (n = \langle 11, x, y \rangle \wedge Term(x) \wedge Term(y)) \vee \exists xy < n$   
 $(n = \langle 13, x, y \rangle \wedge Term(x) \wedge Term(y))$
- (c)  $Form(n) \leftrightarrow$   
 $\exists xy < n (n = \langle 15, x, y \rangle \wedge Term(x) \wedge Term(y)) \vee$   
 $\exists xy < n (n = \langle 3, x, y \rangle \wedge Form(x) \wedge Form(y)) \vee$   
 $\exists x < n (n = \langle 5, x \rangle \wedge Form(x)) \vee$   
 $\exists xy < n (n = \langle 7, x, y \rangle \wedge Vble(x) \wedge Form(y)).$

Note that the lemma is established by an appeal to the primitive recursive version of the recursion theorem (cf. Theorem 15, remark); by switching to characteristic functions, one can make use of course of value recursion. Another standard procedure, e.g. for the coding of terms, is to arithmetize the condition ‘there is a finite sequence of which each member is either a variable or 0, or obtained from earlier ones by applying  $S$ ,  $+$  or  $\cdot$ ’. This immediately establishes the recursiveness, for primitive recursiveness one only has to indicate a bound on the code of this sequence.

2. At certain places it is important to keep track of the free variables in terms and formulas, e.g. in  $\forall x\varphi(x) \rightarrow \varphi(t)$ . The following predicate expresses that ‘ $x$  is free in  $A$ ’ (where  $A$  is a term or a formula—we handle them simultaneously): ‘ $x$  is  $A$  itself or  $A$  is built from two parts not by means of  $\exists$  and  $x$  is free in at least one of those, or built from two parts by means of  $\exists$  and  $x$  is not the first part and free in the second part, etc.’.

So put

$$\begin{aligned} Fr(m, n) \leftrightarrow & (m = n \wedge Vble(m)) \vee \exists xy < n (n = \langle 3, x, y \rangle \wedge \\ & (Fr(m, x) \vee Fr(m, y))) \vee (\exists xy < n (n = \langle 11, \dots \rangle \dots) \vee \\ & (\dots n = \langle 13, \dots \rangle \dots) \vee (\dots n = \langle 15, \dots \rangle \dots) \vee \dots \\ & \dots \exists xy < n (n = \langle 7, x, y \rangle \wedge m \neq x \wedge Fr(m, y)) \end{aligned}$$

Again  $Fr$  is primitive recursive.

3. We need a number-theoretic function that mimics the substitution operation, i.e. that from the Gödel numbers of  $\varphi$ ,  $x$  and  $t$  computes the Gödel number of  $\varphi[t/x]$ .

The function  $Sub$  must satisfy  $Sub(\ulcorner A \urcorner, \ulcorner x \urcorner, \ulcorner t \urcorner) = \ulcorner A[t/x] \urcorner$ , where  $A$  is a term or a formula.

So put

$$Sub(m, n, k) = \begin{cases} k & \text{if } Vble(m) \wedge m = n \\ \langle 0, Sub(p, n, k) \rangle & \text{if } m = \langle 0, p \rangle \\ \langle 5, Sub(p, n, k) \rangle & \text{if } m = \langle 5, p \rangle \\ \langle a, Sub(p, n, k), Sub(q, n, k) \rangle & \text{if } m = \langle a, p, q \rangle \\ & \text{for } a = 3, 11, 13, 15 \\ \langle 7, p, Sub(q, n, k) \rangle & \text{if } m = \langle 7, p, q \rangle \wedge p \neq n, \\ m & \text{otherwise.} \end{cases}$$

Clearly,  $Sub$  is primitive recursive.

4. Depending on the logical basis of **PA** (e.g. a Hilbert type system, or sequent calculus, etc.) one can find a primitive recursive predicate

$\text{Prov}(m, n)$  which expresses that  $n$  is the Gödel number of a proof of the formula with Gödel number  $m$ . Every textbook will give the details, but by now the reader will be able to concoct such a predicate by himself.

In order to avoid cumbersome manipulations, one usually assumes the set of (Gödel numbers of) axioms to be recursive. For **PA**, **Q** and **N** this evidently is correct.

5. The predicate  $\text{Thm}(m) := \exists x \text{Prov}(m, x)$  expresses that  $m$  is the Gödel number of a theorem of **PA**. The existential quantifier makes  $\text{Thm}$  recursively enumerable, and we will show that it is *not* recursive.
6. Superficially speaking, natural numbers lead a double life; they occur in the theory of arithmetic and in the real world. In the first case they occur as symbols, the so-called *numerals*. The numeral for the number  $n$  is denoted by  $\bar{n}$ . To be specific  $\bar{0} = 0$  (recall that **PA** had 0 as a constant symbol, we could have used a bold face zero, but the reader will not be confused),  $\overline{n+1} = S(\bar{n})$ . Now numerals are symbols, so they have Gödel numbers. We define the function  $\text{Num}$  which associates with each number  $n$  the Gödel number of  $\bar{n}$ :

$$\begin{aligned} \text{Num}(0) &= \langle 1 \rangle, \\ \text{Num}(n+1) &= \langle 9, \text{Num}(n) \rangle. \end{aligned}$$

So  $\text{Num}(n) = \ulcorner \bar{n} \urcorner$ .

7. In order to avoid needless restrictions, we quote the following theorem of Craig:

**THEOREM.** *Every axiomatizable theory can be axiomatized by a recursive set of axioms.*

Here a theory is called axiomatizable if the set of its axioms is RE. The following informal argument may suffice.

Let  $\varphi_0, \varphi_1, \varphi_2, \dots$  be an effective enumeration of the axioms of  $T$  (it is no restriction to assume an infinite list). then  $\varphi_0, \varphi_0 \wedge \varphi_1, \varphi_0 \wedge \varphi_1 \wedge \varphi_2, \dots$  also axiomatizes  $T$ , and we can effectively test whether a given sentence  $\sigma$  belongs to this set. For the length of the axioms is strictly increasing so after a finite number of those axioms have been listed we know that if  $\sigma$  has not yet occurred, it will not occur at all. In the literature axiomatizable arithmetical theories are also called RE theories. Non-axiomatizable theories are *ipso facto* undecidable, for their class of theorems is even not RE.

### 3.3 Representability of the Recursive Functions and Predicates

In the preceding section we have reduced, so to speak, logic and arithmetic to recursion theory. Now we will reduce recursion theory to the formal

theory of arithmetic. We will show that inside **PA** (or **Q**, or **N**, for that matter) we can speak about recursive functions and relations, be it in a rather complicated way.

For convenience we will treat the three theories of arithmetic above on an equal footing by considering a theory  $T$  which is an axiomatizable extension of **Q** or **N**. If we need special features of  $T$  (e.g.  $T = \mathbf{PA}$ ), then we will explicitly list them.

**DEFINITION 31.** An  $n$ -ary function  $f$  is *represented* by a formula  $\varphi$  with  $n + 1$  variables  $x_1, \dots, x_n, y$  if for all  $k_1, \dots, k_n, l$ .

$$f(k_1, \dots, k_n) = l \Leftrightarrow T \vdash \varphi(\bar{k}_1, \dots, \bar{k}_n, y) \leftrightarrow \bar{l} = y.$$

An  $n$ -ary relation  $R$  is *represented* by a formula  $\varphi$  with  $n$  free variables if  $F(k_1, \dots, k_n) \Rightarrow T \vdash \varphi(\bar{k}_1, \dots, \bar{k}_n)$  and  $\text{not-}R(k_1, \dots, k_n) \Rightarrow T \vdash \neg\varphi(\bar{k}_1, \dots, \bar{k}_n)$ .

The basic theorem states that

**THEOREM.** *All recursive functions and predicates are representable in any of the systems **PA**, **Q**, **N** (and hence in any  $T$ ).*

It is a simple exercise to show that a relation is representable iff its characteristic function is so. Therefore it suffices to prove the theorem for recursive functions. For this purpose it is most convenient to use the characterization of recursive functions by means of  $\mu$ -recursion.

The proof of the theorem is not very difficult but rather clerical, the reader may look it up in, e.g. [Davis, 1958; Kleene, 1952; Shoenfield, 1967; Smorynski, 1991].

As a consequence we now have available formulas in  $T$  for the predicates that we introduced in Section 3.2. In particular, there is a formula that represents *Prov*. We will, for convenience, use the same symbol for the representing formulas. It will always appear from the context which reading must be used.

The soundness theorem for predicate logic tells us that the theorems of a theory  $T$  are true in all models of  $T$ . In particular, all provable sentences of  $T$  are true in the standard model, following the tradition we call sentences true in  $\mathbb{N}$  simply *true*.

Until the late twenties it was hoped and expected that, conversely, all true sentences would also be provable in **PA**. Gödel destroyed that hope in 1931. Nonetheless, one might hope to establish this converse for a significant class of sentences.

The following theorem provides such a class.

*Convention.* We call a formula  $\Sigma_1^0$  ( $\Pi_1^0$ ) if it is (equivalent to) a prenex formula with a prefix of existential (universal) quantifiers followed by a formula containing only bounded quantifiers.

THEOREM 32 ( $\Sigma_1^0$ -completeness).

$$\varphi \Rightarrow \mathbf{PA} \vdash \varphi \text{ for } \Sigma_1^0 \text{ sentences } \varphi.$$

The proof proceeds by induction on the structure of  $\varphi$ , cf. [Shoenfield, 1967,

p. 211]. Note that the premise is ‘ $\varphi$  is true’, or to spell it out ‘ $\mathbb{N} \models \varphi$ ’.

For  $\Pi_1^0$  sentences truth does not imply provability as we will show below.

Before we proceed to the incompleteness theorem, let us have another look at our formal system. The language of arithmetic contains function symbols for  $+$ ,  $\cdot$ ,  $S$ , should we leave it at that? After all, exponentiation, the square, the factorial, etc. belong to the daily routine of arithmetic, so why should we not introduce them into the language? Well, there is no harm in doing so since primitive recursive functions are given by defining equations which can be formulated in arithmetical language. To be specific, for each primitive recursive function  $f$  we can find the representing formula  $\varphi(\vec{x}, y)$  such that not only  $f(\vec{m}) = n \Leftrightarrow \mathbf{PA} \vdash \varphi(\vec{m}, y) \leftrightarrow y = n$ , but also  $\mathbf{PA} \vdash \forall \vec{x} \exists! y \varphi(\vec{x}, y)$ . Note that here one essentially needs induction.

Now it is a well-known fact of elementary logic that one can add a function symbol  $F$  to the language, and an axiom  $\forall \vec{x} \varphi(\vec{x}, F(\vec{x}))$ , without essentially strengthening the theory. To be precise: the extended theory  $T'$  is *conservative* over the original theory  $T$ : for formulas  $\psi$  not containing  $F$  we have  $T \vdash \psi \Leftrightarrow T' \vdash \psi$ .

In this case we even have a translation  $^\circ$  that eliminates the symbol  $F$  so that  $T' \vdash \psi \Leftrightarrow \psi^\circ$  and  $T \vdash \psi \Leftrightarrow T \vdash \psi^\circ$  (cf. [Shoenfield, 1967, p. 55 ff.], [van Dalen, 1997, p. 144 ff.]).

Summing up, we can conservatively add function symbols and defining axioms for all primitive recursive functions to  $\mathbf{PA}$ .

The proof of this fact runs parallel to the proof of the representability of the primitive recursive functions (and makes use of Gödel’s  $\beta$ -function), cf. [Shoenfield, 1967; Smorynski, 1991].

Observe that if  $F$  is related to the primitive recursive  $f$  in the above manner, then  $F$  *represents*  $f$ :

$$f(\vec{m}) = n \Leftrightarrow \mathbf{PA} \vdash F(\vec{m}) = \bar{n}.$$

Note that we can also add function symbols for recursive functions since a recursive  $f$  is represented by a formula  $\varphi$ . For put  $\psi(\vec{x}, y) = [\exists z \varphi(\vec{x}, z) \rightarrow \varphi(\vec{x}, y) \wedge \forall u < y \neg \varphi(\vec{x}, u)] \wedge [\neg \exists z \varphi(\vec{x}, z) \rightarrow y = 0]$ , then clearly  $\psi$  also represents  $f$  and  $\mathbf{PA} \vdash \forall \vec{x} \exists! y \psi(\vec{x}, y)$ .

For metamathematical purposes it usually does not harm to consider the conservative extension by primitive recursive functions. E.g. with respect to decidability and completeness  $T'$  and  $T$  behave exactly alike.  $T'$  is decidable (complete)  $\Leftrightarrow T$  is decidable (complete).

Since the presence of function symbols for primitive recursive functions considerably streamlines the presentation of certain results, we will freely make use of the above *definitional* extension of arithmetic, which we, by abuse of notation, also call **PA**.

### 3.4 The First Incompleteness Theorem and the Undecidability of **PA**

Gödel formulated a sentence which had a certain analogy to the famous Liar Paradox. Paraphrased in everyday language, it states: ‘I am not provable in **PA**’. This kind of self-referential sentence makes full use of the ability of **PA** (and related systems) to formulate its own syntax and derivability relation. A convenient expedient is the

**THEOREM 33** (Fixed Point Theorem). *Let  $\varphi(x_0)$  have only the free variable  $x_0$  then there is a sentence  $\psi$  such that  $\mathbf{PA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$ .*

**Proof.** We will use the substitution function. Put  $s(m, n) = \text{Sub}(m, \ulcorner x_0 \urcorner, \text{Num}(n))$ , and let  $k := \ulcorner \varphi(s(x_0, x_0)) \urcorner$ , then  $\psi := \varphi(s(\bar{k}, \bar{k}))$  will do the trick.

For  $s(k, k) = s(\ulcorner \varphi(s(x_0, x_0)) \urcorner, k) = \ulcorner \varphi(s(\bar{k}, \bar{k})) \urcorner$ ; so by the representability of  $s$ ,  $\mathbf{PA} \vdash \psi \leftrightarrow \varphi(\ulcorner \psi \urcorner)$ . ■

Observe that Theorem 33 holds for any axiomatizable extension of **PA**.

We now quickly get the incompleteness result, by applying the fixed point theorem to  $\neg \exists y \text{Prov}(x, y)$ :

$$\mathbf{PA} \vdash \psi \leftrightarrow \neg \exists y \text{Prov}(\ulcorner \psi \urcorner, y) \text{ for a certain } \psi.$$

Note that our notation is slightly ambiguous. Since there is a primitive recursive predicate  $\text{Prov}(m, n)$  for provability, there is also a formula in the language of **PA** which represents  $\text{Prov}$ , we will commit a harmless abuse of language by calling this formula also  $\text{Prov}$ . The reader will always be able to see immediately what  $\text{Prov}$  stands for.

Now  $\mathbf{PA} \vdash \psi \rightarrow \mathbf{PA} \vdash \text{Prov}(\psi, \bar{k})$  where  $k$  is the Gödel number of the actual proof of  $\psi$ . So  $\mathbf{PA} \vdash \exists y \text{Prov}(\ulcorner \psi \urcorner, y)$ . But also  $\mathbf{PA} \vdash \neg \exists y \text{Prov}(\ulcorner \psi \urcorner, y)$ , so  $\mathbf{PA} \vdash \perp$ . Assuming consistency we get  $\mathbf{PA} \not\vdash \psi$ .

The negation is a bit more troublesome.

We assume that that **PA** is  $\omega$ -consistent, that is, it is not the case that  $\mathbf{PA} \vdash \sigma(n)$ , for all  $n$ , and  $\mathbf{PA} \vdash \exists x \neg \sigma(x)$ , for arbitrary  $\sigma$

Obviously,  $\omega$ -consistency implies consistency. The converse does not hold.

If  $\mathbf{PA} \vdash \neg \psi$ , then  $\mathbf{PA} \vdash \exists y \text{Prov}(\ulcorner \psi \urcorner, y)$  ..... (1)

Since **PA** is consistent, we have  $\mathbf{PA} \not\vdash \psi$ , so  $\neg \text{Prov}(\ulcorner \psi \urcorner, n)$  for all  $n$ , hence  $\mathbf{PA} \vdash \neg \text{Prov}(\ulcorner \psi \urcorner, n)$  for all  $n$  ..... (2)

(1) and (2) contradict the  $\omega$ -inconsistency of **PA**. Conclusion:  $\mathbf{PA} \not\vdash \neg \psi$ . This establishes the incompleteness of **PA**.

By appealing to the fact that  $\mathbb{N}$  is a model of  $\mathbf{PA}$  we can avoid those technicalities mentioning consistency or  $\omega$ -consistency:  $\mathbf{PA} \vdash \exists y \text{Prov}(\ulcorner \psi \urcorner, y) \Rightarrow \mathbb{N} \models \exists y \text{Prov}(\ulcorner \psi \urcorner, y)$ , so there is an  $n$  such that  $\text{Prov}(\ulcorner \psi \urcorner, \bar{n})$  holds, but this implies that  $\mathbf{PA} \vdash \psi$ . *Contradiction.*

Note that the above Gödel sentence  $\psi$  is  $\Pi_1^0$ .

### Remarks

1. Since  $\mathbf{PA}$  extends  $\mathbf{Q}$  and  $\mathbf{N}$  we also have established the incompleteness of  $\mathbf{Q}$  and  $\mathbf{N}$  (although for just incompleteness a simple model-theoretic argument would suffice, e.g. in  $\mathbf{Q}$  we cannot even prove addition to be commutative). By means of a careful rewording of the Gödel sentences, such that no use is made of primitive recursive functions *in* the system one can present a similar argument which shows that all consistent axiomatizable extensions  $T$  of  $\mathbf{Q}$  (or  $\mathbf{N}$ ) are incomplete. Cf. [Mendelson, 1979; Smorynski, 1991].
2. Rosser eliminated the appeal to  $\omega$ -consistency by applying the fixed-point theorem to another formula: “there is a proof of my negation and no proof of me precedes it.”

In symbols:  $\exists y[\text{Prov}(\text{neg}(x), y) \wedge \forall z < y \neg \text{Prov}(x, z)]$ . Here  $\text{neg}$  is a primitive recursive function such that  $\text{neg}(\ulcorner \varphi \urcorner) = \ulcorner \neg \varphi \urcorner$ . the whole formula is formulated in the language of a conservative extension of  $\mathbf{PA}$ . Call this formula  $R(x)$  and apply the fixed point theorem:  $\mathbf{PA} \vdash \psi \leftrightarrow R(\ulcorner \psi \urcorner)$ . For better readability we suppress the reference to  $\mathbf{PA}$ .

$$(1) \quad \vdash \psi \leftrightarrow \exists y(\text{Prov}(\ulcorner \neg \psi \urcorner, y) \wedge \forall z < y \neg \text{Prov}(\ulcorner \psi \urcorner, z))$$

Suppose  $\vdash \psi$ , then  $\text{Prov}(\ulcorner \psi \urcorner, \bar{n})$  is true for some  $n$ .

$$(2) \quad \text{So } \vdash \text{Prov}(\ulcorner \psi \urcorner, \bar{n})$$

$$(3) \quad \text{Also } \vdash \exists y(\text{Prov}(\ulcorner \neg \psi \urcorner, y) \wedge \forall z < y \neg \text{Prov}(\ulcorner \psi \urcorner, z))$$

$$(4) \quad (2), (3) \Rightarrow \vdash \exists y < \bar{n} \text{Prov}(\ulcorner \neg \psi \urcorner, y)$$

Using  $\vdash y < \bar{n} \leftrightarrow y = \bar{0} \vee y = \bar{1} \vee \dots \vee y = \overline{n-1}$  we find  $\vdash \text{Prov}(\ulcorner \neg \psi \urcorner, \bar{0}) \vee \dots \vee \text{Prov}(\ulcorner \neg \psi \urcorner, \overline{n-1})$ . Now one easily establishes  $\vdash \sigma \vee \tau \Rightarrow \vdash \sigma$  or  $\vdash \tau$  for sentences  $\sigma$  and  $\tau$  with only bounded quantifiers, this yields  $\vdash \text{Prov}(\ulcorner \neg \psi \urcorner, \bar{m})$  for some  $m < n$ .

Hence,  $\vdash \neg \psi$ . but this contradicts the consistency of  $\mathbf{PA}$ .

$$(5) \quad \text{Suppose now } \vdash \neg \psi \text{ then } \not\vdash \psi, \text{ i.e. } \vdash \neg \text{Prov}(\ulcorner \psi \urcorner, \bar{n}) \text{ for all } n,$$

From  $\vdash \neg\psi$  we get  $\vdash \forall y(\text{Prov}(\overline{\neg\psi}, y) \rightarrow \exists z < y \text{Prov}(\overline{\psi}, z))$  and  $\vdash \text{Prov}(\overline{\neg\psi}, \bar{m})$  for some  $m$ .

Hence  $\vdash \exists z < \bar{m} \text{Prov}(\overline{\psi}, z)$ , so as before we get  $\vdash \text{Prov}(\overline{\psi}, \bar{k})$  for some  $k < m$ . Together with (5) this contradicts the consistency of **PA**.

*Conclusion:*  $\not\vdash \psi$  and  $\not\vdash \neg\psi$ .

3. Interpreting the Gödel sentence  $\psi$  in the standard model  $\mathbb{N}$  we see that it is true. So the Gödel sentence provides an instance of a true but unprovable sentence.
4. From the above incompleteness theorem we can easily conclude that  $\{\ulcorner \varphi \urcorner \mid \mathbb{N} \models \varphi\}$ , i.e. the set of (Gödel numbers of) true sentences of arithmetic is not decidable (i.e. recursive). For suppose it were, then we could use the true sentences as axioms and repeat the incompleteness theorem for this system **Tr**: there is a  $\varphi$  such that **Tr**  $\not\vdash \varphi$  and **Tr**  $\not\vdash \neg\varphi$ . but this is impossible since the true and false sentences exhaust all sentences. Hence, **Tr** is not recursive.
5. By means of the provability predicate one can immediately show that in axiomatizable extensions of **Q** or **N** representable predicates and functions are recursive. To be precise, in any axiomatizable theory with 0 and  $S$  such that  $\vdash \bar{m} = \bar{n} \Rightarrow n = m$  the representable predicates and functions are recursive. This provides another characterization of the recursive functions: the class of recursive functions coincides with that of the functions representable in **PA**. Theories satisfying the above conditions are called *numerical*.
6. The technique of arithmetization and the representability of recursive functions allows us to prove the undecidability of extensions of **Q** and **N** (and hence **PA**) (35). We say that **Q** and **N** are *essentially undecidable*.

Exactly the same method allows us to prove the general result: numerical theories in which all recursive functions are representable are undecidable. We can also derive the following theorem of Tarski on the undefinability of truth in arithmetic. We say that a formula  $\tau$  is a *truth definition* in a numerical theory  $T$  if  $T \vdash \varphi \leftrightarrow \tau(\overline{\ulcorner \varphi \urcorner})$ .

**THEOREM 34** (Tarski's Theorem). *No consistent axiomatizable extension of **Q** or **N** has a truth definition.*

**Proof.** Suppose a truth definition  $\tau$  exists, then we can formulate the Liar Paradox. By the fixed point theorem there is a sentence  $\lambda$  such that  $T \vdash \lambda \leftrightarrow \neg\tau(\overline{\ulcorner \lambda \urcorner})$ . Since  $\tau$  is a truth definition, we get  $T \vdash \lambda \leftrightarrow \neg\lambda$ . This conflicts with the consistency of  $T$ . ■



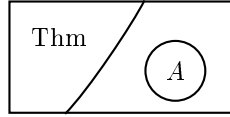


Figure 6.

For the undecidability of **Q**, **N** or **PA** we use the set  $K$  from 28. Or, what comes to the same thing, we diagonalize once more.

Consider the set  $\text{Thm}$  and an RE-set  $A$  disjoint from  $\text{Thm}$  (Figure 6), for any axiomatizable extension  $T$  of **Q** (or **N**).

Define

$$\varphi(k) = \begin{cases} 0 & \text{if } \vdash \exists z T(\bar{k}, \bar{k}, z) \wedge Uz \neq 0 \\ 1 & \text{if } \ulcorner \exists z T(\bar{k}, \bar{k}, z) \wedge Uz \neq 0 \urcorner \in A \\ \uparrow & \text{otherwise.} \end{cases}$$

Clearly  $\varphi$  is a partial recursive function, say with index  $e$ .

(i) Then  $\{e\}e = 0 \Rightarrow T(e, e, n) \wedge Un = 0$  for some  $n \Rightarrow$

$$(6) \quad \vdash T(\bar{e}, \bar{e}, \bar{n}) \wedge U\bar{n} = \bar{0} \Rightarrow \vdash \exists z (T(\bar{e}, \bar{e}, z) \wedge Uz = 0),$$

(7) Also, by definition,  $\{e\}e = 0 \rightarrow \vdash \exists z (T(\bar{e}, \bar{e}, z) \wedge Uz \neq 0)$ .

From the definition the  $T$ -predicate it easily follows that

$$(8) \quad \vdash T(\bar{e}, \bar{e}, z) \wedge T(\bar{e}, \bar{e}, z') \rightarrow z = z'$$

So (4), (7), (8) imply that  $T$  is inconsistent.

(ii)  $\{e\}e = 1 \Rightarrow T(e, e, n) \wedge Un = 1$  for some  $n \Rightarrow \vdash T(\bar{e}, \bar{e}, \bar{n}) \wedge U\bar{n} = 1 \rightarrow \vdash \exists z T(\bar{e}, \bar{e}, z) \wedge Uz \neq 0$ . also  $\{e\}e = 1 \Rightarrow \ulcorner \exists z T(\bar{e}, \bar{e}, z) \wedge Uz \neq 0 \urcorner \in A$ , but that contradicts the disjointness of  $\text{Thm}$  and  $A$ .

So we have shown that  $\text{Thm}^c$  is productive (cf. 30, corollary) and, hence, undecidable. A slight adaptation of the argument shows that the set of theorems and the set of refutable sentences (i.e.  $\sigma$  with  $T \vdash \neg\sigma$ ) are effectively inseparable. The above proof established

**THEOREM 35.** *If  $T$  is a consistent extension of **Q** (or **N**) then  $T$  is undecidable.*

As a corollary we get

**THEOREM 36** (Church's Theorem). *First-order predicate logic is undecidable.*

**Proof.** Consider a first-order language containing at least the language of arithmetic. Since **Q** is finitely axiomatized we can write  $\mathbf{Q} \vdash \varphi \Rightarrow \vdash \sigma \rightarrow \varphi$ , where  $\sigma$  is the conjunction of the axioms of **Q**. Clearly any decision procedure for first-order predicate logic would also provide a decision procedure for **Q**. ■

The undecidability of  $\mathbf{PA}$  yields another proof of its incompleteness, for there is a fairly obvious theorem of model theory that states

**THEOREM 37.** *A complete, axiomatizable theory is decidable (Vaught).*

Observe that it is standard practice to identify ‘decidable’ and ‘recursive’, i.e. to rely on Church’s Thesis. In a more cautious approach one would, of course, use phrases such as ‘the set of Gödel numbers of theorems of  $\mathbf{PA}$  is not recursive’.

There are various paths that lead from recursion theory to the area of decidable and undecidable theories. Usually one starts from a suitable class of undecidable (sometimes also called *unsolvable*) problems in one of the many approaches to the notion of algorithm, e.g. the Halting problem for Turing machines or register machine, or Post’s correspondence problem and somehow ‘interprets’ them in a convenient logical form. The actual role of recursion theory by then is modest or often even nil. Most of the techniques in the particular field of undecidable theories are of a logical nature, e.g. the construction of suitable translations.

### 3.5 *Decidable and Undecidable Theories*

In a large number of cases there are reductions of the decision problem of certain theories to well-known theories.

We list a few undecidable theories below:

1. Peano’s arithmetic,
2. Theory of rings [Tarski, 1951],
3. Ordered fields [Robinson, 1949],
4. Theory of lattices [Tarski, 1951],
5. Theory of a binary predicate [Kalmar, 19336],
6. Theory of a symmetric binary predicate [Church and Quine, 1952],
7. Theory of partial order [Tarski, 1951],
8. Theory of two equivalence relations [Rogers, 1956],
9. Theory of groups,
10. Theory of semi groups,
11. Theory of integral domains.

One should note that the underlying logic does play a role in decidability results, e.g. whereas classical monadic predicate logic is decidable, intuitionistic monadic predicate logic is *not* [Kripke, 1968], cf. [Gabbay, 1981, p. 234]).

The decidability aspects of predicate logic have been widely studied. One of the oldest results is the decidability of monadic predicate calculus [Löwenheim, 1915; Behmann, 1922], and the high-water mark is the undecidability of predicate logic [Church, 1936]. In this particular area there has been much research into solvable and unsolvable cases of the decision problem. A comprehensive treatment of decidability and undecidability results can be found in [Börger *et al.*, 1997].

There are special classes  $\Gamma$  of formulas, given by syntactical criteria, for which undecidability has been established by a reduction procedure that effectively associates to each  $\varphi$  a formula  $\varphi^*$  of  $\Gamma$  such that  $\vdash \varphi \Leftrightarrow \vdash \varphi^*$ . Such a class  $\Gamma$  is called a *reduction class* with respect to provability) (or validity). Analogously one has reduction classes with respect to satisfiability.

Among the syntactic criteria that are used we distinguish, e.g. (i) the number of arguments of the predicates, (ii) the number of predicates, (iii) the length of the quantifier prefix for the prenex normal form, (iv) the number of quantifier changes in the same.

For convenience we introduce some notation:  $Q_1^{n_1}, \dots, Q_m^{n_m}$  stands for the class of all prenex formulas with prefixes of  $n_1$  quantifiers  $Q_1, n_2$  quantifiers  $Q_2, \dots, n_m$  quantifiers  $Q_m$ . A superscript  $\infty$  indicates a quantifier block of arbitrary finite length.

Restrictions on the nature of predicate symbols is indicated by finite sequences, e.g.  $(0, 2, 1)$  indicates a predicate logic language with no unary predicates, two binary and one ternary predicate. Combining the two notations we get the obvious classes, such as  $\forall^\infty \exists^2(0, 1), \exists^1 \forall^1 \exists^3(2, 1)$ , etc.

An immediate simple but not trivial example is furnished by the *Skolem normal form* for satisfiability: the class of  $\forall^\infty \exists^\infty$  formulas is a reduction class for satisfiability. So this class consists of prenex formulas with universal quantifiers followed by existential quantifiers.

Of course, this can be improved upon since the undecidability proofs for predicate logic provide more information.

We list some examples of reduction classes. There will be no function symbols or constants involved.

$\exists\forall\exists\forall(0, 3)$	(Büchi 1962)
$\forall\exists\forall(0, \infty)$	(Kahr, Moore, Wang 1962)
$\forall^3\exists(0, \infty)$	(Gödel 1933)
$\exists\forall\exists\forall(\infty, 1)$	(Rödding 1969)
$\forall\exists\forall(\infty, 1)$	(Kahr 1962)
$\exists^\infty\forall^2\exists^2\forall^\infty(0, 1)$	(Kalmar 1932)
$\forall^\infty\exists(0, 1)$	(Kalmar, Suranyi 1950)
$\forall\exists\forall^\infty(0, 1)$	(Denton 1963)
$\forall\exists\forall\exists^\infty(0, 1)$	(Gurevich 1966)

One can also consider prenex formulas with the matrix in conjunctive or disjunctive normal form, and place restrictions on the number of disjuncts, conjuncts, etc.

$\mathcal{D}_n$  is the class of disjunctive normal forms with at most  $n$  disjuncts:  $\mathcal{C}_n$  is the class of conjunctive normal forms with at most  $n$  disjuncts per conjunct. *Krom formulas* are those in  $\mathcal{D}_2$  or  $\mathcal{C}_2$ ; *Horn formulas* those in  $\mathcal{C}_\infty$  with at most one negated disjunct per conjunct.

Prefix classes of Krom and Horn formulas have been investigated by Aanderla, Dreben, Börger, Goldfarb, Lewis, Maslov and others. For a thorough treatment of the subject the reader is referred to [Lewis, 1979], [Börger *et al.*, 1997].

The reader should not get the impression that logicians deal exclusively with undecidable theories. There is a considerable lore of decidable theories, but the actual decision methods usually employ little or no recursion theory. This is in accordance with the time-honoured practice: one recognizes a decision method when one sees one.

The single most important decision method for showing the decidability of theories is that of *quantifier elimination*. Briefly, a theory  $T$  is said to have (or allow) quantifier elimination if for each formula  $\varphi(x_1, \dots, x_n)$ , with all free variables shown, there is an open (i.e. quantifier free) formula  $\psi(x_1, \dots, x_n)$  such that  $T \vdash \varphi(x_1, \dots, x_n) \leftrightarrow \psi(x_1, \dots, x_n)$ . So for theories with quantifier elimination one has only to check derivability for open formulas. This problem usually is much simpler than the full derivability problem, and it yields a decision procedure in a number of familiar cases.

An early, spectacular result was that of [Presburger, 1930] who showed that the theory of arithmetic with only successor and addition has a quantifier elimination and is decidable. However, additive number theory is not the most impressive theory in the world, so Presburger's result was seen as a curiosity (moreover, people hoped at that time that full arithmetic was decidable, so this was seen as an encouraging first step).

The result that really made an impression was Tarski's famous *Decision Method for Elementary Algebra and Geometry* [1951], which consisted of a quantifier elimination for real closed (and algebraically closed) fields.

By now quantifier elimination is established for a long list of theories, among which are *linear dense ordering*, *Abelian groups* (Szmielew 1955), *p-*

*adic fields* (Cohen 1969), *Boolean algebras* (Tarski 1949. For a survey of decidable theories, including some complexity aspects, cf. [Rabin, 1977].

We will give a quick sketch of the method for the theory of equality.

1. Since one wants to eliminate step by step the quantifier in front of the matrix of a prenex formula it clearly suffices to consider formulas of the form  $\exists y\varphi(y, x_1, \dots, x_n)$ .
2. We may suppose  $\varphi$  to be in disjunctive normal form  $\bigvee_i \varphi_i$  and, hence, we can distribute the  $\exists$ -quantifier. So it suffices to consider formulas of the form  $\exists y\psi(y, x_1, \dots, x_n)$  where  $\psi$  is a conjunction of atoms and negations of atoms.
3. After a bit of rearranging, and eliminating trivial parts (e.g.  $x_i = x_i$  or  $\neg y = y$ ), we are left with  $\exists y(y = x_{i_1} \wedge \dots \wedge y = x_{i_k} \wedge y \neq x_{j_1} \wedge \dots \wedge y \neq x_{j_l} \wedge \delta)$ , where  $\delta$  does not contain  $y$ . By ordinary logic we reduce this to  $\exists y(\text{---}) \wedge \delta$ . Now the formula  $\exists y(\text{---})$  is logically equivalent to  $(x_{i_1} = x_{i_2} \wedge \dots \wedge x_{i_1} = x_{i_k} \wedge x_{i_1} \neq x_{j_1} \wedge \dots \wedge x_{i_1} \neq x_{j_l})$ .

So we eliminated one quantifier. However, there are a number of special cases to be considered, e.g. there are no atoms in the range of  $\exists y$ . Then we cannot eliminate the quantifier. So we simply introduce a constant  $c$  and replace  $\exists y(y \neq x_{j_a})$  by  $c \neq x_{j_a}$  (the Henkin constants, or witnesses), i.e. we consider conservative extensions of the theory.

In this way we finally end up with an open formula containing new constants. If we started with a sentence then the result is a Boolean combination of sentences which express conditions on the constants, namely which ones should be unequal. Such conditions can be read as conditions of cardinality: there are at least  $n$  elements.

From the form of the resulting sentence we can immediately see whether it is derivable or not. Moreover, the method shows what completions the theory of identity has (cf. [Chang and Keisler, 1973, p. 55]).

Whereas the actual stepwise elimination of quantifiers is clearly algorithmic, there are a number of decidability results that have no clear algorithmic content. Most of these are based on the fact that a set is recursive iff it and its complement are recursive enumerable. An example is the theorem: if  $T$  is axiomatizable and complete then  $T$  is decidable.

### 3.6 The Arithmetical Hierarchy

We have seen that decision problems can be considered to be of the form 'does  $n$  belong to a set  $X$ ?', so, up to a coding, the powerset of  $\mathbb{N}$  presents us with all possible decision problems. Put in this way, we get too many unrealistic decision problems. For, a reasonable decision problem is usually presented in the form: test if an element of a certain effectively generated

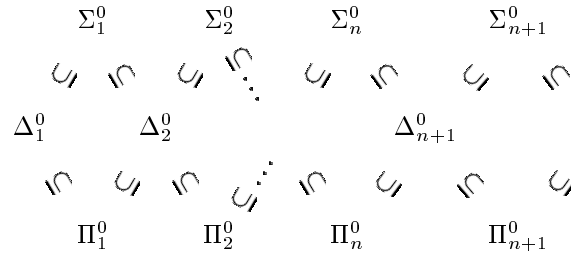


Figure 7.

(or described) set is an element of another such set, e.g. if a formula is a theorem. So among the subsets of  $\mathbb{N}$ , we are interested in certain sets that are somehow effectively described.

We have already met such sets, e.g. the primitive recursive, the recursive, and the recursive enumerable sets. It seems plausible to consider sets of natural numbers definable in a theory or in a structure. The very first candidate that comes to mind is the standard model of arithmetic,  $\mathbb{N}$ . Sets definable in  $\mathbb{N}$  are called *arithmetical*, they are of the form  $\{n \mid \mathbb{N} \models \varphi(\bar{n})\}$ .

We already know that recursive sets and RE sets are arithmetical, and we know at least one particular set that is not arithmetical: the set of (Gödel numbers of) true sentences of *arithmetic* (Tarski's theorem).

It is natural to ask if the collection of arithmetical sets has some structure, for example with respect to a certain measure of complexity. The answer, provided by Kleene and Mostowski, was 'yes'. We can classify the arithmetical sets according to their syntactic form. The classification, called the *arithmetical hierarchy*, is based on the prenex normal forms of the defining formulas.

The hierarchy is defined as follows:

a  $\Sigma_1^0$ -formula is of the form  $\exists \vec{y} \varphi(\vec{x}, \vec{y})$

a  $\Pi_1^0$ -formula is of the form  $\forall \vec{y} \varphi(\vec{x}, \vec{y})$

where  $\varphi$  contains only bounded quantifiers,

if  $\varphi(\vec{x}, \vec{y})$  is a  $\Sigma_n^0$  formula, then  $\forall \vec{y} \varphi(\vec{x}, \vec{y})$  is a  $\Pi_{n+1}^0$  formula,

if  $\varphi(\vec{x}, \vec{y})$  is a  $\Pi_n^0$  formula, then  $\exists \vec{y} \varphi(\vec{x}, \vec{y})$  is a  $\Sigma_{n+1}^0$  formula,

$\Sigma_n^0$  and  $\Pi_n^0$  sets are defined as extensions of corresponding formulas, and a set is  $\Delta_n^0$  if it is both  $\Sigma_n^0$  and  $\Pi_n^0$ .

The inclusions indicated in figure 7 are easily established,

e.g. if  $X \in \Sigma_1^0$ , then it is defined by a formula  $\exists \vec{y} \varphi(x, \vec{y})$ , by adding dummy variables and quantifiers we can get an equivalent  $\Pi_2^0$ -formula  $\forall z \exists \vec{y} \varphi(x, \vec{y})$ .

The above classification is based on syntactic criteria; there is, however, a 'parallel' classification in terms of recursion theory. At the lower end of the hierarchy we have the correspondence

- $\Delta_1^0$  recursive sets,
- $\Sigma_1^0$  RE sets,
- $\Pi_1^0$  complements of RE sets.

We know, by theorem 26, that  $\Sigma_1^0 \cap \Pi_1^0 = \Delta_1^0$ .

The predicate  $\exists zT(x, y, z)$  is universal for the RE sets in the sense that each RE set is of the form  $\exists zT(e, y, z)$  for some  $E$ . Via the above correspondence,  $\exists zT(x, y, z)$  is universal for the class  $\Sigma_1^0$ . Likewise  $\neg\exists zT(x, y, z)$  is universal for  $\Pi_1^0$  (for proofs of the facts stated here see any of the standard texts in the bibliography). To progress beyond the RE sets and their complements, we use an operation that can be considered as an infinite union, let  $U_n(x, y)$  be universal for  $\Pi_n^0$ , then  $\exists zU_n(x, \langle z, y \rangle)$  is universal for  $\Sigma_{n+1}^0$ . The basic technique here is in complete analogy to that in the so-called *hierarchy of Borel sets*, one can make this precise and show that the arithmetical hierarchy is the finite part of the effective Borel hierarchy (cf. [Shoenfield, 1967; Hinman, 1978]).

Since one goes from  $\Pi_n^0$  to  $\Sigma_{n+1}^0$  by adding an existential quantifier in front of the defining  $\Pi_n^0$  formula, we can also express the relation between  $\Pi_n^0$  and  $\Sigma_{n+1}^0$  in another recursion-theoretic way.

We have already defined the notion of relative recursiveness:  $f$  is recursive in  $g$  if  $f(x) = \{e\}^g(x)$  for some  $e$  (where the superscript  $g$  indicates that  $g$  is an extra initial function), or  $A$  is recursive in  $B$  if  $K_A(x) = \{e\}^B(x)$  for some  $e$ .

We use this definition to introduce the notion: *A is recursively enumerable in B*:  $A$  is RE in  $B$  if  $A$  is the domain of  $\{e\}^B$  for some  $e$ .

Generalising the  $T$ -predicate in an obvious way to a relativized version we get  $A$  is RE in  $B$  if  $n \in A \Leftrightarrow \exists zT^B(e, n, z)$  for some  $E$ .

Now the relation between  $\Pi_n^0$  and  $\Sigma_{n+1}^0$  can be stated as

$$A \in \Sigma_{n+1}^0 \Leftrightarrow A \text{ is RE in some } B \in \Pi_n^0.$$

There is another way to go up in the hierarchy. The operation that one can use for this purpose is the *jump*. The *jump of a set A*, denoted by  $A'$ , is defined as

$$\{x \mid \{x\}^A(x) \downarrow\}, \text{ or } \{x \mid \exists zT^A(x, x, z)\},$$

e.g. the jump  $\emptyset'$  of  $\emptyset$  is  $K$ .

The successive jumps of  $\emptyset : \emptyset', \emptyset'', \dots, \emptyset^{(n)}, \dots$  are the paradigmatic examples of  $\Sigma_n^0$  sets, in the sense that

$$A \in \Sigma_{n+1}^0 \Leftrightarrow A \text{ is RE in } \emptyset^{(n)}.$$

What happens if we take a set that is recursive in  $\emptyset^{(n)}$ ? The answer is given by *Post's theorem*:

$$A \in \Delta_{n+1}^0 \Leftrightarrow A \leq_T \emptyset^{(n)}.$$

An important special case is  $A \in \Delta_2^0 \Leftrightarrow A$  is recursive in the Halting Problem (i.e.  $K$ ).

A diagonal argument, similar to that of 28 shows that  $\Sigma_n^0 - \Pi_n^0 \neq \emptyset$  and  $\Pi_n^0 - \Sigma_n^0 \neq \emptyset$ , hence all the inclusions in the diagram above are proper (this is called the *Hierarchy Theorem*) and the hierarchy is infinite.

For arithmetical predicates one can simply compute the place in the arithmetical hierarchy by applying the reduction to (almost) prenex form, i.e. a string of quantifiers followed by a recursive matrix.

Example: ‘ $e$  is the index of a total recursive function’. We can reformulate this as ‘ $\forall x \exists y T(e, x, y)$ ’ and this is a  $\Pi_2^0$  expression.

The more difficult problem is to find the lowest possible place in the hierarchy. So in the example: could  $\{e \mid \{e\} \text{ is total}\}$  be  $\Sigma_1^0$ ?

There are various ways to show that one has got the best possible result. One such technique uses *reducibility* arguments.

**DEFINITION 38.** Let  $A, B \subseteq \mathbb{N}$ .  $A$  is many-one (one-one) reducible to  $B$  if there is a (one-one) recursive function  $f$  such that  $\forall x (x \in A \Leftrightarrow f(x) \in B)$ . Notation:  $A \leq_m B$  ( $A \leq_1 B$ ).

**EXAMPLE 39.**

1.  $W_e \leq_1 \{\langle x, y \rangle \mid x \in W_y\} = \{\langle x, y \rangle \mid \{y\}x \downarrow\} = \{\langle x, y \rangle \mid \exists z T(y, x, z)\}$ . Define  $f(n) = \langle n, e \rangle$ , then clearly  $n \in W_e \Leftrightarrow \langle n, e \rangle \in \{\langle x, y \rangle \mid x \in W_y\}$  and  $f$  is injective.
2.  $\{\langle x, y \rangle \mid x \in W_y\} \leq_m K = \{x \mid x \in W_x\} = \{x \mid \{x\}x \downarrow\}$ . Consider the function  $\{(z)_1\}(z)_0$ , and define

$$\{e\}(z, x) = \begin{cases} 1 & \text{if } \{(z)_1\}(z)_0 \downarrow \\ \text{divergent} & \text{otherwise} \end{cases}$$

By the  $S_n^m$  theorem  $\{e\}(z, x) = \{f(z)\}(x)$  and

$$\begin{aligned} a \in W_b &\Leftrightarrow \{b\}a \downarrow \Leftrightarrow \{(\langle a \ b \rangle)_1\}(\langle a \ b \rangle)_0 \downarrow \Leftrightarrow \\ &\Leftrightarrow \{f\langle a \ b \rangle\}f\langle a \ b \rangle \downarrow \Leftrightarrow f\langle a \ b \rangle \in K. \end{aligned}$$

an extra argument shows that  $\leq_m$  can be replaced by  $\leq_1$ .

Now we call  $A \in \Sigma_n^0(\Pi_n^0)\Sigma_n^0$  *complete* ( $\Pi_n^0$  *complete*) if for all  $B \in \Sigma_n^0(\Pi_n^0)$   $B \leq_m A$  (equivalently  $\leq_1$ ).

One can show that the  $n$ th jump of  $\emptyset$ ,  $\emptyset^{(n)}$ , is  $\Sigma_n^0$  complete.

Therefore, we use the sets  $\emptyset^{(n)}$  to establish, e.g. that  $A \in \Sigma_n^0$  does not belong to  $\Pi_{n-1}^0$  or  $\Sigma_{n-1}^0$  by showing  $\emptyset^{(n)} \leq_m A$ . For, otherwise  $\emptyset^{(n)} \leq_m A \leq_1 \emptyset^{(n-1)}$  (or its complement), which is not the case. For more applications and examples see [Rogers, 1967, Section 14.8].

The complexity of models can also be measured by means of the arithmetical hierarchy, i.e. we call a model  $\Sigma_n^0(\Pi_n^0$  or  $\Delta_n^0)$  if its universe is the set of all natural numbers and the relations are  $\Sigma_n^0(\Pi_n^0$  or  $\Delta_n^0)$ .



We mention several results:

1. *The Hilbert-Bernays completeness theorem.* If a theory (in a countable language) has an RE set of axioms, and it is consistent then it has a  $\Delta_2^0$  model (Kleene, Hasenjaeger).
2. A decidable theory has a recursive model.
3. The only recursive model of arithmetic is the standard model (Tenenbaum).

In 1970, Matijasevič gave a negative solution to Hilbert's tenth problem: *there is no algorithm that decides if any given Diophantine-equation* (that is of the form  $p(x_1, \dots, x_n) = 0$ , for a polynomial  $p$  with integer coefficients) *has a solution.* To be precise, he established that every RE subset of  $\mathbb{N}$  is Diophantine, that is for each RE set  $A$  we can find a polynomial  $p$  with integer coefficients such that  $A = \{n \mid \exists x_1, \dots, x_n p(n, x_1, \dots, x_n) = 0\}$ .

Using results of Davis, Julia Robinson and Putnam, Matijasevič solved the problem by a purely number-theoretic argument, [Matijasevič, 1973].

*Basis theorem, or do decent sets have decent elements?* Let us agree for a moment that decent subsets of the Euclidean plane are open sets and that decent points are points with rational coordinates, then each non-empty decent set contains a decent point. For, a non-empty open set  $A$  is a union of open discs, so if  $p \in A$  then  $p \in C \subseteq A$ , where  $C$  is an open disc. Now a bit of geometry tells us that  $C$  contains decent points. We express this by saying that the rational points form a basis for the open sets. Had we taken points with integer coordinates, then we would not have found a basis.

The problem of finding a basis for a given family of sets is of general interest in logic. One usually considers families of subsets of  $\mathbb{N}$ , or of functions from  $\mathbb{N}$  to  $\mathbb{N}$ . 'Decency' is mostly expressed in terms of the arithmetical (or analytical) hierarchy.

Since most basis-theorems deal with classes of functions, we will classify functions by means of their graphs. To be precise, a function is  $\Sigma_n^0$  ( $\Pi_n^0$  or  $\Delta_n^0$ ) if its graph is so.

Unfortunately the majority of basis theorems deal with the analytical hierarchy instead of the arithmetical hierarchy, so we restrict ourselves to a theorem that fits into the arithmetical hierarchy, and also has nice applications in logic:

**THEOREM 40** (Kreisel's Basis Theorem). *The  $\Delta_2^0$  functions form a basis of the  $\Pi_1^0$  sets of characteristic functions.*

Here the notions of  $\Sigma_n^0$ ,  $\Pi_n^0$  generalize in a natural way to classes of functions or sets. Consider a language for arithmetic that also contains set variables  $X_i$  and a relation symbol  $\in$  (for 'element of'), i.e. a language for second-order arithmetic. For this language we define the notions  $\Sigma_n^0$

and  $\Pi_n^0$  exactly as in first-order arithmetic. Now a class  $A$  of sets is  $\Sigma_n^0$  if  $A = \{X \mid \varphi(X)\}$  for a  $\Sigma_n^0$  formula  $\varphi$ . One may just as well consider a language for second-order arithmetic with function variables.

For a detailed treatment of basis theorems, cf. [Shoenfield, 1967] and [Hinman, 1978].

## APPENDIX

### *Proof of the Normal Form Theorem*

The normal form theorem states, roughly speaking, that there is a primitive recursive relation  $T(e, u, z)$  that formalizes the heuristic statement ‘ $z$  is a (coded) computation that is performed by a partial recursive function with index  $e$  on input  $u$ ’ (i.e.  $\langle \vec{x} \rangle$ ). The ‘computation’ has been arranged in such a way that its first projection is its output.

The proof is a matter of clerical perseverance—not difficult, but not exciting either.

We have tried to arrange the proof in a readable manner by providing a running commentary.

We have displayed below the ingredients for, and conditions on, computations. The index contains the information given in the clauses  $R_i$ . The computation codes the following items:

1. the output,
2. the input,
3. the index
4. subcomputations.

	<i>Index</i>	<i>Input</i>	<i>Step</i>	<i>Conditions on Subcomputations</i>
	$e$	$u$	$z$	
R <sub>1</sub>	$\langle 0, n, q \rangle$	$\langle \vec{x} \rangle$	$\langle q, u, e \rangle$	
R <sub>2</sub>	$\langle 1, n, i \rangle$	$\langle \vec{x} \rangle$	$\langle x_i, u, e \rangle$	
R <sub>3</sub>	$\langle 2, n, i \rangle$	$\langle \vec{x} \rangle$	$\langle x_i + 1, u, e \rangle$	
R <sub>4</sub>	$\langle 3, n + 4 \rangle$	$\langle p, q, r, s, \vec{x} \rangle$	$\langle p, u, e \rangle$ if $r = s$ $\langle q, u, e \rangle$ if $r \neq s$	
R <sub>5</sub>	$\langle 4, n, b, c_1, \dots, c_k \rangle$	$\langle \vec{x} \rangle$	$\langle (z')_1, u, e, z' \rangle$ , $\langle z''_1, \dots, z''_k \rangle$	$z', z''_1, \dots, z''_k$ are computations with indices $b, c_1, \dots, c_k$ . $z''_i$ has input $\langle (z''_1)_1, \dots, (z''_k)_1 \rangle$ . (cf. 14)
R <sub>6</sub>	$\langle 5, n + 2 \rangle$	$\langle p, q, \vec{x} \rangle$	$\langle s, u, e \rangle$	
R <sub>7</sub>	$\langle 6, n + 1 \rangle$	$\langle b, \vec{x} \rangle$	$\langle (z')_1, u, e, z' \rangle$	$z'$ is a computation with input $\langle \vec{x} \rangle$ and index $b$ .

Note that  $z$  is the ‘master number’, i.e. we can read off the remaining data from  $z$ , e.g.  $e = (z)_3$ ,  $\text{lth}(u) = (z)_{3,2}$  but in particular the ‘master numbers’

of the subcomputations. So, by decoding the code for a computation, we can effectively find the codes for the subcomputations, etc. This suggests a primitive recursive algorithm for the extraction of the total ‘history’ of a computation from its code. As a matter of fact, that is essentially the content of the normal form theorem.

We will now proceed in a (slightly) more formal manner, by defining a predicate  $C(z)$  (for  $z$  is a computation), using the information of the preceding table. For convenience, we assume that in the clauses below, sequences  $u$  (in  $Seq(u)$ ) have positive length.

$C(z)$  is defined by cases as follows:

$$\begin{aligned}
C(z) := & \left\{ \begin{array}{l}
\exists q, u, e < z [z = \langle q, u, e \rangle \wedge Seq(u) \wedge e = \langle 0, lth(u), q \rangle] \quad (1) \\
\text{or} \\
\exists u, e, i < z [z = \langle (u)_i, u, e \rangle \wedge Seq(u) \wedge e = \langle 1, 1h(u), i \rangle] \quad (2) \\
\text{or} \\
\exists u, e, i < z [z = \langle (u)_i + 1, u, e \rangle \wedge Seq(u) \wedge e = \langle 2, lth(u), i \rangle] \quad (3) \\
\text{or} \\
\exists u, e < z [Seq(u) \wedge e = \langle 3, lth(u) \rangle \wedge lth(u) > 4 \wedge ([z = \langle (u)_1, u, e \rangle \wedge \\
\wedge (u)_3 = (u)_4] \vee [z = \langle (u)_2, u, e \rangle \wedge (u)_3 \neq (u)_4])] \quad (4) \\
\text{or} \\
Seq(z) \wedge lth(z) = 5 \wedge Seq((z)_3) \wedge Seq((z)_5) \wedge lth((z)_3) = \\
= 3 + lth((z)_5) \wedge (z)_{3,1} = 4 \wedge C((z)_4) \wedge (z)_{4,1} = (z)_1 \wedge (z)_{4,2} = \\
= \langle (z)_{5,1,1}, \dots, (z)_{5, lth((z)_5), 1} \rangle \wedge (z)_{4,3} = (z)_{3,3} \wedge \\
\wedge \bigwedge_{i=1}^{lth((z)_5)} [C((z)_{5,i}) \wedge (z)_{5,i,3} = (z)_{1,3+i} \wedge (z)_{5,i,2} = (z)_2] \quad (5) \\
\text{or} \\
\exists s, u, e < z [z = \langle s, u, e \rangle \wedge Seq(u) \wedge e = \langle 5, lth(u) \rangle \wedge \\
s = \langle 4, (u)_{1,2} - 1, (u)_1, \langle 0, (u)_{1,2} - 1, (u)_2 \rangle, \langle 1, (u)_{1,2} - 1, 1 \rangle, \dots \\
\dots, \langle 1, (u)_{1,2} - 1, (e)_2 - 2 \rangle], \quad (6) \\
\text{or} \\
\exists u, e, w < z [Seq(u) \wedge e = \langle 6, lth(y) \rangle \wedge z = \langle (w)_1, u, e, w \rangle \wedge C(w) \wedge \\
\wedge (w)_3 = (u)_1 \wedge (w)_2 = \langle (u)_2, \dots, (u)_{lth(u)} \rangle] \quad (7)
\end{array} \right.
\end{aligned}$$

Now observe that each clause of the disjunction only refers to  $C$  for smaller numbers. Furthermore, each disjunct clearly is primitive recursive. By changing to the characteristic function of  $C$ , we see that (1)-(7) present us (via definition by cases) with a course of value recursion. Hence,  $K_C$  is primitive recursive, and so is  $C$ .

This brings us to the end of the proof; define  $T(e, u, z) := C(z) \wedge u = (z)_2 \wedge e = (z)_3$  (i.e. ‘ $z$  is a computation with index  $e$  and input  $u$ ’), then  $T$  is primitive recursive, and  $\{e\}(u) = (\mu z T(e, u, z))_1$ . ■

## HISTORICAL NOTES

Recursion theory was the offspring of Gödel’s famous investigation into the completeness of arithmetic [1931]. In the course of his proof of the incom-

pleteness theorem he introduced the primitive recursive functions (under the name ‘recursive’). In a subsequent paper [1934] he introduced, following a suggestion of Herbrand, a wider class of the so-called Herbrand–Gödel recursive functions. In the following years a number of approaches to the theory of algorithms were worked out:  *$\lambda$ -calculus* — Church, Kleene, cf. [Barendregt, 1981]; *Turing machines* — [Turing, 1936], cf. [Davis, 1965]; *Combinatory Logic* — [Schönfinkel, 1924; Curry, 1929], cf. [Barendregt, 1981]; *Post Systems* [Post, 1947], cf. [Hopcroft and Ullman, 1969]; *Register machines* — [Minsky, 1961; J.C. Shepherdson, 1963], cf. [Schnorr, 1974; Minsky, 1967]; *Markov algorithms* [Markov, 1954], cf. [Mendelson, 1979].

The theory of recursive functions as a discipline in its own right was developed by Kleene, who proved all the basic theorems: the  $S_n^m$  theorem, the recursion theorem, the normal form theorem, and many others. Turing developed the theory of Turing machines, constructed a Universal Turing machine and showed the unsolvability of the Halting Problem. After Gödel’s pioneering work, Rosser modified the independent statement so that only the consistency of arithmetic was required for the proof [Rosser, 1936], cf. [Smoryński, 1977].

Following Post [1944], a study of subsets of  $\mathbb{N}$  was undertaken, leading to notions as *creative*, *productive*, *simple*, etc. At the same time Post initiated the classification of sets of natural numbers in terms of ‘Turing reducibility’, i.e. a set  $A$  is Turing reducible to a set  $B$  if the characteristic function of  $A$  is recursive when the characteristic function of  $B$  is given as one of the initial functions— in popular terms if we can test membership of  $A$  given a test for membership of  $B$ . The theory of *degrees of unsolvability* has grown out of this notion, cf. [Shoenfield, 1971; Soare, 1980].

The applications to logic are of various sorts. In the first place there is the refinement of (un)decidability results of Gödel, Rosser, Turing and others, now known as the theory of *reduction types*, in which syntactical classes with unsolvable decision problems are studied, cf. [Lewis, 1979], on the other hand a study of solvable cases of the decision problem has been carried out, cf. [Drebden and Goldfarb, 1979].

The theory of (arithmetical and other) *hierarchies* started with papers of Kleene [1943] and Mostowski [1947], the subject has been extensively explored and generalized in many directions, cf. [Hinman, 1978; Griffor, 2000].

Generalizations of recursion theory to objects other than natural numbers have been studied extensively. To mention a few approaches: *recursion in higher types*, [Kleene, 1959]; *recursion on ordinals*, Kripke, Platek; *Admissible sets*, Kripke, Barwise, a.o.; *Definability Theory*, Moschovakis, a.o.; *Axiomatic recursion theory*, Wagner–Strong, Friedman, a.o.; *Recursion on continuous functionals*, Kleene, Kreisel. The reader is referred to [Barwise, 1975; Hinman, 1978; Fenstad, 1980; Moschovakis, 1974; Norman, 1980].

The connection between recursion theory and intuitionism was first es-

tablished by Kleene [1945], since then the subject has proliferated, cf. [Troelstra, 1973]. For historical accounts of the subject cf. [Kleene, 1976; Kleene, 1981; Mostowski, 1966; Heijenoort, 1967; Odifreddi, 1989].

*Utrecht University, The Netherlands.*

## BIBLIOGRAPHY

- [Börger, 1989] E. Börger. *Computability, Complexity, Logic*. North-Holland, Amsterdam, 1989.
- [Börger et al., 1997] E. Börger, E. Grädel and Y. Gurevich. *The Classical Decision Problem*, Springer-Verlag, Berlin, 1997.
- [Barendregt, 1981] H.P. Barendregt. *Lambda Calculus: Its Syntax and Semantics*. North-Holland, Amsterdam, 1981.
- [Barwise, 1975] J. Barwise. *Admissible Sets and Structures*. Springer-Verlag, Berlin, 1975.
- [Behmann, 1922] H. Behmann. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Mathematische Annalen*, 86:163–229, 1922.
- [Carnap, 1937] R. Carnap. *The Logical Syntax of Language*. Routledge and Kegan Paul, London, 1937.
- [Chang and Keisler, 1973] C.C. Chang and H.J. Keisler. *Model Theory*. North-Holland, Amsterdam, 1973.
- [Church, 1936] A. Church. A note on the entscheidungsproblem. *The Journal of Symbolic logic*, 1:40–41, 1936.
- [Church and Quine, 1952] A. Church and W. V. O. Quine. Some theorems on definability and decidability. *Journal of Symbolic Logic*, 17:179–187, 1952.
- [Curry, 1929] H.B. Curry. An analysis of logical substitution. *American Journal of Mathematics*, 51:363–384, 1929.
- [Davis, 1958] M. Davis. *Computability and Unsolvability*. McGraw-Hill, New York, 1958.
- [Davis, 1965] M. Davis, editor. *The Undecidable*. Raven Press, New York, 1965.
- [Drebden and Goldfarb, 1979] B. Drebden and W.D. Goldfarb. *The Decision Problem. Solvable Classes of Quantificational Formulas*. Addison-Wesley, Reading, MA, 1979.
- [Fenstad, 1980] J. E. Fenstad. *Generalized Recursion Theory: An Axiomatic Approach*. Springer-Verlag, Berlin, 1980.
- [Fraenkel et al., 1973] A. A. Fraenkel, Y. Bar-Hillel, A. Levy, and D. Van Dalen. *Foundations of Set Theory*. North-Holland, Amsterdam, 1973.
- [Gabbay, 1981] D.M. Gabbay. *Semantical Investigations in Heyting's Intuitionistic Logic*. Reidel, Dordrecht, 1981.
- [Gandy, 1980] R. Gandy. Church's thesis and principles for mechanisms. In J. Barwise, H.J. Keisler, and K. Kunen, editors, *Kleene Symposium*. North-Holland, Amsterdam, 1980.
- [Gödel, 1931] K. Gödel. Über formal unentscheidbare Sätze des Principia Mathematica und verwandter Systeme I. *Monatshefte Math. Phys.*, **38**, 173–198, 1931.
- [Gödel, 1934] K. Gödel. On undecidable propositions of formal mathematical systems. Mimeographed notes, 1934. Also in [Davis, 1965] and [Gödel, 1986].
- [Gödel, 1965] K. Gödel. Remarks before the Princeton Bicentennial Conference. In M. Davis, editor, *The Undecidable*. Raven Press, New York, 1965.
- [Gödel, 1986] K. Gödel. *Collected Works I, II, III*, edited by S. Feferman et al. Oxford University Press, 1986, 1990, 1995.
- [Griffor, 2000] E. Griffor, ed. *The Handbook of Recursion Theory*, Elsevier, Amsterdam, 2000.
- [Grzegorzczuk, 1961] A. Grzegorzczuk. *Fonctions Récurives*. Gauthier-Villars, Paris, 1961.
- [Heijenoort, 1967] J. Van Heijenoort. *From Frege to Gödel. A Source Book in Mathematical Logic 1879–1931*. Harvard University Press, Cambridge, MA, 1967.

- [Hinman, 1978] P.G. Hinman. *Recursion-Theoretic Hierarchies*. Springer-Verlag, Berlin, 1978.
- [Hopcroft and Ullman, 1969] J.E. Hopcroft and J.D. Ullman. *Formal Languages and Their Relations to Automata*. Addison-Wesley, Reading, MA, 1969.
- [Hyland, 1982] J.M.E. Hyland. The effective topos. In D. van Dalen A.S. Troelstra, editors, *The L.E.J. Brouwer Centenary Symposium*, pages 165–216. North-Holland, Amsterdam, 1982.
- [J.C. Shepherdson, 1963] H.E. Sturgis and J.C. Shepherdson. Computability of recursive functions. *Journal of the Association of Computing Machines*, 10:217–255, 1963.
- [Jeroslow, 1972] R.G. Jeroslow. *On the Encodings Used in the Arithmetization of Metamathematics*. University of Minnesota, 1972.
- [Kalmar, 19336] L. Kalmar. Zurückführung des Entscheidungsproblems auf binären Funktionsvariablen. *Comp. Math.*, 4:137–144, 1936.
- [Kleene, 1943] S.C. Kleene. Recursive predicates and quantifiers. *Transactions of the American Mathematical Society*, 53:41–73, 1943.
- [Kleene, 1945] S.C. Kleene. On the interpretation of intuitionistic number theory. *The Journal of Symbolic Logic*, 10:109–124, 1945.
- [Kleene, 1952] S.C. Kleene. *Introduction to Metamathematics*. North-Holland, Amsterdam, 1952.
- [Kleene, 1959] S.C. Kleene. Recursive functionals and quantifiers of finite type 1. *Transactions of the American Mathematical Society*, 91:1–52, 1959.
- [Kleene, 1976] S.C. Kleene. The work of Kurt Gödel. *Journal of Symbolic Logic*, 41:761–778, 1976.
- [Kleene, 1981] S.C. Kleene. Origins of recursive function theory. *Annals of History of Computing Science*, 3:52–67, 1981.
- [Kripke, 1968] S. Kripke. *Semantical Analysis for Intuitionistic Logic II*. (unpublished), 1968.
- [Lewis, 1979] H.R. Lewis. *Unsolvable Classes of Quantificational Formulas*. Addison-Wesley, Reading, MA, 1979.
- [Löwenheim, 1915] L. Löwenheim. Über Möglichkeiten im Relativkalkül. *Math. Ann.*, 76:447–470, 1915.
- [Markov, 1954] A.A. Markov. Theory of algorithms. *AMS translations (1960)*, 15:1–14. Russian original 1954.
- [Matijasevič, 1973] Y. Matijasevič. Hilbert's tenth problem. In P. Suppes, L. Henkin, A. Joyal, and G.C. Moisil, editors, *Logic, Methodology and Philosophy of Science*, pages 89–110. North-Holland, Amsterdam, 1973.
- [McCarty, 1986] D. McCarty. Realizability and recursive set theory. *Annals of Pure and Applied Logic*, 32:153–183, 1986.
- [Mendelson, 1979] E. Mendelson. *Introduction to Mathematical Logic*. Van Nostrand, New York, 1979.
- [Minsky, 1961] M. Minsky. Recursive unsolvability of Post's problem of 'tag' and other topics in the theory of turing machines. *Annals of Mathematics*, 74:437–455, 1961.
- [Minsky, 1967] M. Minsky. *Computation: Finite and Infinite Machines*. Prentice-Hall, Englewood Cliffs, NJ, 1967.
- [Moschovakis, 1974] Y. Moschovakis. *Elementary Induction on Abstract Structures*. North-Holland, Amsterdam, 1974.
- [Mostowski, 1947] A. Mostowski. On definable sets of positive integers. *Fundamenta mathematicae*, 34:81–112, 1947.
- [Mostowski, 1966] A. Mostowski. *Thirty Years of Foundational Studies*. Blackwell, Oxford, 1966.
- [Norman, 1980] D. Norman. *Recursion on the Countable Functionals*. Springer-Verlag, Berlin, 1980.
- [Odifreddi, 1989] P. Odifreddi. *Classical Recursion Theory. The Theory of Functions and Sets of Natural Numbers*. North-Holland, Amsterdam, 1989.
- [Papadimitriou, 1994] C.H. Papadimitriou. *Computational Complexity*. Addison-Wesley, Reading, MA, 1994.
- [Pèter, 1959] R. Pèter. Rekursivität und konstruktivität. In A. Heyting, editor, *Constructivity in Mathematics*. North-Holland, Amsterdam, 1959.

- [Post, 1947] E.L. Post. Recursive unsolvability of a problem of Thue. *Journal of Symbolic Logic*, **12**, 1–11, 1947.
- [Post, 1944] E.L. Post. Recursively enumerable sets of positive integers and their decision problems. *Bull. Am. Math. Soc.*, **50**, 284–316, 1944.
- [Presburger, 1930] M. Presburger. Über die Vollständigkeit eines gewissen Systems der Arithmetik ganzer Zahlen, in welchem die Addition als einzige Operation hervortritt. In *Comptes-rendus du I Congrès des Mathématiciens des Pays Slaves*, Warsaw, 1930.
- [Rabin, 1977] M.O. Rabin. Decidable theories. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 595–629. North-Holland, Amsterdam, 1977.
- [Robinson, 1949] J. Robinson. Definability and decision problems in arithmetic. *Journal of Symbolic Logic*, 14:98–114, 1949.
- [Rogers, 1956] H. Rogers, jnr. Certain logical reductions and decision problems. *Ann. Math.*, 64:264–284, 1956.
- [Rogers, 1967] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, New York, 1967.
- [Rosser, 1936] J.B. Rosser. Extensions of some theorems of Gödel and Church. *Journal of Symbolic Logic*, 1:87–91, 1936.
- [Schönfinkel, 1924] M. Schönfinkel. Über die Bausteine der mathematischen Logik. *Mathematische Annalen*, 92:305–316, 1924.
- [Schnorr, 1974] C.P. Schnorr. *Rekursive Funktionen und ihre Komplexität*. Teubner, Stuttgart, 1974.
- [Shoenfield, 1967] J.R. Shoenfield. *Mathematical Logic*. Addison-Wesley, Reading, MA, 1967.
- [Shoenfield, 1971] J.R. Shoenfield. *Degrees of Unsolvability*. North-Holland, Amsterdam, 1971.
- [Smoryński, 1977] C. Smoryński. The incompleteness theorems. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 821–866. North-Holland, Amsterdam, 1977.
- [Smoryński, 1991] C. Smoryński. *Logical Number Theory I. An Introduction*. Springer-Verlag, Berlin.
- [Soare, 1980] R.I. Soare. *Recursively Enumerable Sets and Degrees*. Springer, Berlin, 1980.
- [Soare, 1987] R.I. Soare. *Recursively Enumerable Sets and Degrees*. Springer, Berlin, 1987.
- [Tarski, 1951] A. Tarski. *A Decision Method for Elementary Algebra nad Geometry*. Berkeley, 2nd, revised edition, 1951.
- [Troelstra, 1973] A.S. Troelstra. *Metamathematical Investigations of Intuitionistic Arithmetic and Analysis*. Springer-Verlag, Berlin, 1973.
- [Turing, 1936] A. Turing. On computable numbers, with an application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society*, 42:230–265, 1936. Also in [Davis, 1965].
- [van Dalen, 1973] D. van Dalen. Lectures on intuitionism. In H. Rogers A.R.D. Mathias, editors, *Cambridge Summer School in Mathematical Logic*, pages 1–94. Springer-Verlag, Berlin, 1973.
- [van Dalen, 1997] D. van Dalen. *Logic and Structure (3rd ed.)*. Springer-Verlag, 1997.
- [Wang, 1974] Hao Wang. *From Mathematics to Philosophy*. Routledge and Kegan Paul, London, 1974.





## MATHEMATICS OF LOGIC PROGRAMMING

## INTRODUCTION

Consider a set  $\Phi$  of first-order sentences and a first-order sentence  $\psi$ . If  $\Phi \models \psi$ , i.e., if  $\psi$  is a consequence of  $\Phi$ , then this can be established by a formal proof using some complete first-order calculus. However, there is no universal program that, given any  $\Phi$  and  $\psi$  as inputs, decides whether  $\Phi \models \psi$ . This general fact does not exclude to ask for such a program for “simple”  $\Phi$  or  $\psi$ . As an important example, let  $\psi$  be restricted to existential statements of the form  $\exists x\varphi(x)$ , where  $\varphi(x)$  is quantifier-free. One might think of  $\varphi(x)$  as an equation in  $x$ . Then  $\exists x\varphi(x)$  states that  $\varphi(x)$  has a solution, and  $\Phi \models \exists x\varphi(x)$  means that  $\Phi$  guarantees a solution for  $\varphi(x)$ . Given  $\Phi$  and  $\exists x\varphi(x)$  one might not only wish to decide whether

$$(1) \quad \Phi \models \exists x\varphi(x),$$

but in the positive case to produce one solution (or all solutions) as terms  $t$  of the underlying language, i.e. those terms  $t$  such that

$$(2) \quad \Phi \models \varphi(t),$$

and, for practical purposes, it should be possible to produce these solutions quickly.

In general, if  $\Phi \models \exists x\varphi(x)$ , there is no term  $t$  such that  $\Phi \models \varphi(t)$ . An example is given by  $\Phi = \{\exists xRx\}$  and  $\varphi(x) = Rx$  with a unary relation symbol  $R$ . To give a further example, let  $\Phi$  be a set of sentences axiomatizing the class of real closed fields (in the notions  $+$ ,  $-$ ,  $\times$ ,  $\div$ ,  $0$ ,  $1$ ) and let  $\varphi(x)$  be  $x^3 + (1 + 1)x^2 = x + 1$ . Then “ $\Phi \models \exists x\varphi(x)$ ?” asks whether the polynomial  $x^3 + 2x^2 - x - 1$  has a root in all real closed fields or, equivalently, in the field of reals (the equivalence follows from a well-known theorem of the theory of models). In general, we are unable to give the roots of a polynomial as terms in the arithmetic operations.

Hence, in order to realize the extended expectations concerning concrete solutions, we have to impose further restrictions on  $\Phi$  or  $\psi$ . As it turns out, there are fairly general conditions under which the existence of solutions can be witnessed by terms and that even allow to create all solutions  $t$ . For example, one assumes that  $\Phi$  consists of so-called universal Horn formulas. The theory concerned is known as *logic programming*. The central idea here, going back mainly to [Kowalski, 1974] and [Colmerauer, 1970], is that quantifier-free Horn formulas can be given a procedural interpretation;

for instance, an implication of the form  $(\psi_1 \wedge \dots \wedge \psi_k) \rightarrow \psi$  can be viewed as a rule that allows to pass from  $\psi_1, \dots, \psi_k$  to  $\psi$ . The *ideas* rest on work of Herbrand [Herbrand, 1968] from the thirties; the *methods* refer to the so-called resolution developed in [Robinson, 1965].

Programming languages based on the theory of logic programming have widely been used, e.g. for knowledge based systems, the best known ones being those of the PROLOG (= programming in logic) family. There is a bulk of methods and results with respect to larger applicability and efficiency of the procedures. These aspects, however, will not be considered here; the interested reader is referred to [Apt, 1990], [Lloyd, 1984], [Sterling and Shapiro, 1986] and to [Gabbay, Hogger and Robinson, 1993f]. In particular, we will treat negation only marginally (see [Shepherdson, 1988]).

To give an example of how basic features of knowledge based systems can be modelled in the framework addressed by (1) and (2), think of a system  $S$  to check cars.  $S$  consists of a certain experience or knowledge about possible damages recorded as “rules” and formalized by the axioms in  $\Phi$ . A simple example of such a rule could be: “If the battery is empty, the starter does not operate”. Let  $\varphi(x)$  express that  $x$  is a possible reason for a car not to operate correctly in a specific manner (e.g., a reason for inefficient brakes). Then we may ask whether

$$\Phi \cup \Phi_D \models \exists x \varphi(x),$$

where  $\Phi_D$  contains the results from a diagnosis of a concrete car with the damage, and we expect that we are provided with a list of all possible reasons, i.e., of all  $t$  with

$$\Phi \cup \Phi_D \models \varphi(t).$$

The situation in (1) and (2) refers to all models of  $\Phi$  (in case of  $S$  to all cars of the type  $S$  is designed for). In practice, say, in connection with databases, it often is desirable to consider specific structures  $\mathcal{A}$  and to ask whether

$$(3) \quad \mathcal{A} \models \exists x \varphi(x),$$

and – in the positive case – to quickly produce one (or all) solution(s) in  $\mathcal{A}$ . As a typical example, which we will present in more detail in Section 2, let  $\mathcal{A}$  be a description of the schedules of a national bus company and let  $\exists x \varphi(x)$  ask for a connection between two towns (also involving a change of buses). The aim then could be to provide a customer with a full list of all connections. As it turns out, the methods that were developed in logic programming, can also be applied to structures, as there is a close connection between the general problem as mirrored by (1) and specific problems as mirrored by (3). Section 2 is concerned with these aspects.

The quantifier-free formulas  $\varphi(x)$  are propositional combinations of atomic first-order formulas. Therefore, important aspects of the theory of logic programming will be concerned with propositional logic and hence, in a first

approximation, may be treated in the framework of this logic. We will do so in Section 1, where we describe the main tool of logic programming, the resolution method, on the propositional level. The first-order case will be presented in Section 4.

The structure  $\mathcal{A}$  mentioned above containing the schedule of a national bus company may be viewed as a (relational) database. In Section 3 we will give a short description of DATALOG, a language designed to serve as a query language for databases, and investigate its relationship to logic programming.

Finally, in Section 5, we analyze the computational complexity of the methods in question. The main result we state says that the feasible database queries (where feasibility is made precise by the notion of polynomial time computability) just coincide with the queries that can be formulated in DATALOG.

In part 2 of Section 5, we come back to the undecidability of first order logic as mentioned at the beginning and give an even stronger result, namely the undecidability of the Horn part of first-order logic. In particular, we get that there is no uniform procedure to decide questions of the form “ $\Phi \models \exists x\varphi(x)$ ?” even in the framework of Horn formulas. We thus will have gained a principal borderline of efficiency that we cannot surpass.

We assume that the reader is acquainted with the basics of propositional logic and of first-order logic as given, for example, in [Ebbinghaus, Flum and Thomas, 1992] or [Hodges, 1983]. Our terminology and notations will deviate a little bit from those in [Hodges, 1983]. In any case, we hope that the short descriptions we usually give, will clarify their use.

## 1 PROPOSITIONAL RESOLUTION

We start by fixing our notation for propositional logic.

*Propositional formulas* are built up from the propositional variables  $p_1, p_2, p_3, \dots$  by means of the propositional connectives  $\neg$  (not),  $\wedge$  (and), and  $\vee$  (or) using the parantheses  $), ($ . We denote propositional formulas by  $\alpha, \beta, \gamma, \dots$  and use  $p, q, r, \dots$  for propositional variables. Then the propositional formulas are just the strings that can be obtained by means of the following rules:

$$\frac{}{p}; \quad \frac{\alpha}{\neg\alpha}; \quad \frac{\alpha, \beta}{(\alpha \wedge \beta)}; \quad \frac{\alpha, \beta}{(\alpha \vee \beta)}.$$

We also write  $(\alpha_0 \wedge \dots \wedge \alpha_n)$  for  $(\dots ((\alpha_0 \wedge \alpha_1) \wedge \alpha_2) \dots \wedge \alpha_n)$  and  $(\alpha_0 \vee \dots \vee \alpha_n)$  for  $(\dots ((\alpha_0 \vee \alpha_1) \vee \alpha_2) \dots \vee \alpha_n)$ ; furthermore,  $(\alpha_0 \wedge \dots \wedge \alpha_n \rightarrow \alpha)$  stands for  $(\neg\alpha_0 \vee \dots \vee \neg\alpha_n \vee \alpha)$ .

To provide the semantics, we consider *assignments*

$$b : \{p_1, p_2, \dots\} \rightarrow \{T, F\}$$

where T stands for the truth value “true” and F for the truth value “false”. For each  $b$  we define the truth value  $\|\alpha\|_b$  inductively by

$$\|q\|_b := b(q)$$

$$\|\neg\alpha\|_b := \begin{cases} F & \text{if } \|\alpha\|_b = T \\ T & \text{if } \|\alpha\|_b = F \end{cases}$$

$$\|(\alpha \wedge \beta)\|_b := \begin{cases} T & \text{if } \|\alpha\|_b = \|\beta\|_b = T \\ F & \text{else} \end{cases}$$

$$\|(\alpha \vee \beta)\|_b := \begin{cases} F & \text{if } \|\alpha\|_b = \|\beta\|_b = F \\ T & \text{else.} \end{cases}$$

If  $\|\alpha\|_b = T$  we say that  $b$  is a *model* of  $\alpha$  and also write  $b \models \alpha$ . For a set  $\Delta$  of propositional formulas,  $b$  is a *model* of  $\Delta$ , written  $b \models \Delta$ , if  $b \models \alpha$  for all  $\alpha \in \Delta$ .  $\Delta$  is *satisfiable*, if  $b \models \Delta$  for some  $b$ . Furthermore,  $\alpha$  is *satisfiable* if  $\{\alpha\}$  is, and  $\alpha$  is *valid* if  $b \models \alpha$  for all  $b$ . Finally,  $\alpha$  is a *consequence* of  $\Delta$ , written  $\Delta \models \alpha$ , if every model of  $\Delta$  is a model of  $\alpha$ . Note that  $\emptyset \models \alpha$  (also written  $\models \alpha$ ) iff  $\alpha$  is valid.

Clearly, the truth value  $\|\alpha\|_b$  is determined by the values  $b(q)$  of the variables  $q$  that occur in  $\alpha$ . Hence, to check whether  $\alpha$  with variables among  $q_1, \dots, q_n$  is satisfiable (or valid) we only need to calculate the truth values of  $\alpha$  under the  $2^n$  “assignments” of values to  $q_1, \dots, q_n$ , a number exponential in the number of variables in  $\alpha$ .

Of course, even for relatively small  $n$ , this is an unfeasible number of steps. So the question remains whether there is a feasible procedure to test satisfiability. According to the common model of complexity theory, the class of problems solvable with a feasible algorithm is identified with the class PTIME of problems that have an algorithm whose number of steps (or running time) is polynomially bounded in the length of the input. Hence, for a procedure that checks satisfiability, feasibility means that there is a polynomial  $f$  over the set  $\mathbb{N}$  of natural numbers such that for any input  $\alpha$ , the procedure needs at most  $f(\text{length of } \alpha)$  steps.

The question whether the satisfiability problem for propositional formulas belongs to PTIME is still open; it is equivalent to one of the most prominent questions of complexity theory, namely to the question whether PTIME = NPTIME, where NPTIME denotes the class of problems solvable by polynomially bounded nondeterministic algorithms. For more information cf.

[Hopcroft and Ullman, 1979] and also Section 5.

After having reviewed normal forms for propositional logic, we will present a syntactically defined class of propositional formulas, for which the satisfiability problem is polynomially bounded. The class and the algorithm will be of importance later.

*Normal forms.* A formula  $\alpha$  is in *conjunctive normal form* (CNF), if it is a conjunction of disjunctions of *literals* (that is, of propositional variables or negated propositional variables), i.e.,

$$(1) \quad \alpha = (\lambda_{11} \vee \dots \vee \lambda_{1m_1}) \wedge \dots \wedge (\lambda_{n1} \vee \dots \vee \lambda_{nm_n}),$$

where the  $\lambda_{ij}$  are literals. Dually,  $\alpha$  is in *disjunctive normal form* (DNF), if  $\alpha$  is a disjunction of conjunctions of literals,

$$(2) \quad \alpha = (\lambda_{11} \wedge \dots \wedge \lambda_{1m_1}) \vee \dots \vee (\lambda_{n1} \wedge \dots \wedge \lambda_{nm_n}).$$

Every propositional formula is *logically equivalent* to (i.e., has the same models as) both a formula in CNF and a formula in DNF (see, e.g., [Ebbinghaus, Flum and Thomas, 1992]). We mention that the satisfiability problem for formulas in CNF is as hard (in a precise sense) as it is for arbitrary propositional formulas. However, formulas in DNF can be tested quickly: Given a formula as in (2), one simply has to check whether there is an  $i$  such that the literals  $\lambda_{i1}, \dots, \lambda_{im_i}$  do not include a propositional variable and its negation. Therefore, it may be hard to translate formulas into logically equivalent formulas in DNF. In fact, the usual proofs lead to translation procedures of exponential time complexity.

*Horn formulas.* Our aim is to exhibit a feasible algorithm that decides satisfiability for formulas in conjunctive normal form where each conjunct contains at most one unnegated propositional variable. Thus, such a formula has the form  $(\alpha_0 \wedge \dots \wedge \alpha_n)$  where each  $\alpha_i$  is a Horn formula in the sense of the following definition.

DEFINITION 1. A *Horn formula*<sup>1</sup> is a formula of one of the forms

$$(H1) \quad q$$

$$(H2) \quad (\neg q_0 \vee \dots \vee \neg q_k \vee q), \quad i. \quad (q_0 \wedge \dots \wedge q_k \rightarrow q)$$

$$(H3) \quad (\neg q_0 \vee \dots \vee \neg q_k).$$

Horn formulas according to (H1) and (H2) are called *positive* (or *strict*), those according to (H3) are called *negative*.<sup>2</sup>

<sup>1</sup>After the logician Alfred Horn.

<sup>2</sup>In the literature, Horn formulas in our sense are usually called *basic* Horn formulas, whereas Horn formulas are conjunctions of basic Horn formulas.

Clearly, the satisfiability of  $(\alpha_0 \wedge \dots \wedge \alpha_n)$  is equivalent to that of  $\{\alpha_0, \dots, \alpha_n\}$ . Next we present a quick algorithm to decide whether a finite set of Horn formulas is satisfiable, an algorithm important for the applications we have in mind.

Whenever  $\Delta$  is a set of Horn formulas, we denote by  $\Delta^+$  and  $\Delta^-$  the subset of positive and negative Horn formulas in  $\Delta$ , respectively.

Let  $\Delta$  be a set of Horn formulas and let  $b$  be a model of  $\Delta$ . Then  $b(q) = \text{T}$  for  $q \in \Delta$ , and whenever  $(q_0 \wedge \dots \wedge q_k \rightarrow q) \in \Delta$  and  $b(q_0) = \dots = b(q_k) = \text{T}$ , then  $b(q) = \text{T}$ . Therefore, we have  $b(q) = \text{T}$  at least for those variables  $q$  that are underlined by applying the *underlining algorithm* consisting of the following rules (U1) and (U2):

(U1) Underline in  $\Delta$  all occurrences of propositional variables that themselves are elements of  $\Delta$ .

(U2) If  $(q_0 \wedge \dots \wedge q_k \rightarrow q) \in \Delta$  and  $q_0, \dots, q_k$  are already underlined, then underline all occurrences of  $q$  in formulas of  $\Delta$ .

The algorithm terminates when none of the two rules can be applied any more. If  $\Delta$  contains  $r$  propositional variables, this happens after at most  $r$  applications.

EXAMPLE 2. *Let*

$$\Delta := \{(r \rightarrow q), (s \wedge q \rightarrow t), (\neg r \vee \neg t), (p \rightarrow q), s, (s \wedge p \wedge r \rightarrow p), r\}.$$

*Then (U1) leads to*

$$\{\underline{r} \rightarrow q, (\underline{s} \wedge q \rightarrow t), (\neg \underline{r} \vee \neg t), (p \rightarrow q), \underline{s}, (\underline{s} \wedge p \wedge \underline{r} \rightarrow p), \underline{r}\}$$

*and (U2) to*

$$\{\underline{r} \rightarrow \underline{q}, (\underline{s} \wedge \underline{q} \rightarrow t), (\neg \underline{r} \vee \neg t), (p \rightarrow \underline{q}), \underline{s}, (\underline{s} \wedge p \wedge \underline{r} \rightarrow p), \underline{r}\}.$$

*Again, (U2) can be applied, yielding*

$$\{\underline{r} \rightarrow \underline{q}, (\underline{s} \wedge \underline{q} \rightarrow \underline{t}), (\neg \underline{r} \vee \neg \underline{t}), (p \rightarrow \underline{q}), \underline{s}, (\underline{s} \wedge p \wedge \underline{r} \rightarrow p), \underline{r}\}$$

*where the algorithm terminates.*

Now, let  $b^\Delta$  be the following assignment associated with  $\Delta$ :

$$b^\Delta(q) := \begin{cases} \text{T} & \text{if } q \text{ is underlined} \\ \text{F} & \text{else.} \end{cases}$$

Then the remarks leading to the underlining algorithm show:

$$\text{Whenever } b \models \Delta \text{ and } b^\Delta(q) = \text{T} \text{ then } b(q) = \text{T};$$

this is part (a) of the lemma below. Note that the set of underlined variables only depends on the set  $\Delta^+$  of positive Horn formulas in  $\Delta$ ; hence,  $b^\Delta = b^{\Delta^+}$  (this is part (b) of the lemma).

LEMMA 3.

(a) For all assignments  $b$  and propositional variables  $q$ :

$$\text{if } b \models \Delta \text{ and } b^\Delta(q) = \text{T then } b(q) = \text{T}.$$

(b)  $b^\Delta = b^{\Delta^+}$ .

(c)  $b^{\Delta^+} \models \Delta^+$ .

(d)  $b^\Delta \models \Delta$  iff  $\Delta$  is satisfiable  
 iff for all  $\alpha \in \Delta^-, \Delta^+ \cup \{\alpha\}$  is satisfiable.

**Proof.** (c) Formulas in  $\Delta^+$  have the form (H1) or (H2). Note that  $b^{\Delta^+}(q) = \text{T}$  for  $q \in \Delta^+$ , because such  $q$ 's are underlined by (U1). Now, let  $(q_0 \wedge \dots \wedge q_k \rightarrow q)$  be in  $\Delta^+$ . If  $b^{\Delta^+}(q_0) = \dots = b^{\Delta^+}(q_k) = \text{T}$  then  $q_0, \dots, q_k$  are underlined and hence, by (U2), also  $q$  is underlined; thus,  $b^{\Delta^+}(q) = \text{T}$ . Therefore we have  $b^{\Delta^+} \models (q_0 \wedge \dots \wedge q_k \rightarrow q)$ .

(d) Clearly, if  $b^\Delta \models \Delta$  then  $\Delta$  is satisfiable, and if  $\Delta$  is satisfiable then so is  $\Delta^+ \cup \{\alpha\}$  for  $\alpha \in \Delta^-$ . Now assume that

$$\text{for all } \alpha \in \Delta^- : \Delta^+ \cup \{\alpha\} \text{ is satisfiable.}$$

We have to show that  $b^\Delta \models \Delta$  or equivalently (by (b)) that  $b^{\Delta^+} \models \Delta$ . By (c),  $b^{\Delta^+} \models \Delta^+$ . Let  $\alpha \in \Delta^-$ , say,  $\alpha = (\neg q_0 \vee \dots \vee \neg q_k)$ . By our assumption there is  $b$  such that  $b \models \Delta^+ \cup \{\alpha\}$ ; in particular,  $b(q_i) = \text{F}$  for  $i = 0, \dots, k$ . By (a) applied to  $\Delta := \Delta^+$ ,  $b^{\Delta^+}(q_i) = \text{F}$  for  $i = 0, \dots, k$  and hence,  $b^{\Delta^+} \models \alpha$ . ■

Since negative Horn formulas have the form  $(\neg q_0 \vee \dots \vee \neg q_k)$ , we obtain:

COROLLARY 4. *The following are equivalent:*

(i)  $\Delta$  is satisfiable.

(ii) For no  $\alpha$  in  $\Delta^-$  all propositional variables in  $\alpha$  are marked by the underlining algorithm. ■

The corollary shows that the underlining algorithm gives us a feasible test for satisfiability of finite sets  $\Delta$  of Horn formulas: We use the rules (U1) and (U2) for  $\Delta$  and finally check whether in all negative Horn formulas of

$\Delta$  there is at least one propositional variable that is not underlined. Thus, the set  $\Delta$  of Example 2 is not satisfiable, since both variables of  $(\neg r \vee \neg t)$  get underlined.

REMARK 5. (1) In case  $\Delta$  is satisfiable, parts (a) and (d) of the lemma show that  $b^\Delta$  is a minimal model of  $\Delta$ , minimal in the sense that a variable gets the value  $\top$  only if necessary. Even, by (a)–(c),

$$b^\Delta(p) = \top \quad \text{iff} \quad \Delta^+ \models p.$$

Further reformulations of this minimality are contained in (3) and (4).

(2) Part (c) of the lemma shows that every set  $\Delta$  of positive Horn formulas is satisfiable. While  $b^\Delta$  is a minimal model, the assignment mapping all propositional variables to  $\top$  is a “maximal” model of  $\Delta$ .

(3) Given  $\Delta$ , set

$$CWA(\Delta) := \{\neg p \mid \text{not } \Delta \models p\}.$$

*CWA stands for closed world assumption: To assume CWA means that we regard  $\Delta$  as a full description of a world in the sense that  $\neg p$  must hold in case  $\Delta$  does not yield that  $p$  is true. If  $\text{not } \Delta \models p$ , then  $\Delta \cup \{\neg p\}$  is satisfiable, hence  $b^{\Delta \cup \{\neg p\}}$  ( $= b^\Delta$ ) is a model of  $\Delta \cup \{\neg p\}$  by the lemma. Thus,*

$$\text{if } \Delta \text{ is satisfiable then } b^\Delta \models \Delta \cup CWA(\Delta).$$

(The concept of closed world assumption goes back to [Reiter, 1978].)

(4) For a variable  $q$  occurring on the right side of an implication in  $\Delta$ , consider all such implications in  $\Delta$ , say

$$\begin{aligned} &(\beta_1 \rightarrow q) \\ &\quad \vdots \\ &(\beta_s \rightarrow q) \end{aligned}$$

and set

$$\alpha_q := q \leftrightarrow (\beta_1 \vee \dots \vee \beta_s).$$

Let

$$CDB(\Delta) := \{\alpha_q \mid q \notin \Delta, q \text{ occurs on the right side of an implication in } \Delta\}.$$

*CDB stands for completed database (here,  $\Delta$  is viewed as a database containing information in terms of propositional formulas). Using (a) and (d) of the lemma, one easily shows:*

$$\text{If } \Delta \text{ is satisfiable then } b^\Delta \models \Delta \cup CDB(\Delta).$$



(The concept of completion of a database goes back to [Clark, 1978].)

In the procedure for checking (un-)satisfiability which we will study later under the name “Horn resolution”, the underlining algorithm is run upside down. For example, let  $\Delta$  be a set of Horn formulas with only one negative element  $(\neg q_0 \vee \dots \vee \neg q_k)$ . If we want to prove the unsatisfiability of  $\Delta$  by use of the underlining algorithm, we have to show that all variables in  $\{\neg q_0, \dots, \neg q_k\}$  will finally be underlined. If  $(r_0 \wedge \dots \wedge r_n \rightarrow q_i) \in \Delta$  (or  $q_i \in \Delta$ ), by rule (U2) (or (U1)) it suffices to show that each variable in

$$(*) \quad \{\neg q_0, \dots, \neg q_{i-1}, \neg r_0, \dots, \neg r_n, \neg q_{i+1}, \dots, \neg q_k\}$$

$$(\text{or } \{\neg q_0, \dots, \neg q_{i-1}, \neg q_{i+1}, \dots, \neg q_k\})$$

ends up being underlined.

Now this argument can be repeated and applied to the set in (\*). It will turn out that  $\Delta$  is not satisfiable if in this way one can reach the empty set in finitely many steps (then none of the variables remains to be shown to be underlined). In the case of

$$\Delta_0 := \{(\neg p \vee \neg q \vee \neg s), (r \rightarrow p), r, q, (u \rightarrow s), u\}$$

we can reach the empty set at follows:

$$\begin{array}{ll} \{\neg p, \neg q, \neg s\} & \\ \{\neg p, \neg q, \neg u\} & (\text{since } (u \rightarrow s) \in \Delta_0.) \\ \{\neg p, \neg q\} & (\text{since } u \in \Delta_0) \\ \{\neg p\} & (\text{since } q \in \Delta_0) \\ \{\neg r\} & (\text{since } (r \rightarrow p) \in \Delta_0) \\ \emptyset & (\text{since } r \in \Delta_0). \end{array}$$

The idea underlying this procedure can be extended to arbitrary formulas in CNF; in this way one arrives at the *resolution method* due to A. Blake [1937] and J.A. Robinson [1965]. There, formulas in CNF are given in set-theoretic notation. For instance, one identifies a disjunction  $(\alpha_0 \vee \dots \vee \alpha_k)$  with the set  $\{\alpha_0, \dots, \alpha_k\}$  of its members. In this way the formulas  $(\neg p_0 \vee p_1 \vee \neg p_0)$ ,  $(\neg p_0 \vee \neg p_0 \vee p_1)$ , and  $(p_1 \vee \neg p_0)$  coincide with the set  $\{\neg p_0, p_1\}$ . Obviously, disjunctions which lead to the same set are logically equivalent. We introduce the notation in a more precise way.

For literals we write  $\lambda, \lambda_1, \dots$ . A finite, possibly empty set of literals is called a *clause*. We use the letters  $C, L, M, \dots$  for clauses and  $\mathcal{C}, \dots$  for (not necessarily finite) sets of clauses.

The transition from a disjunction  $(\lambda_0 \vee \dots \vee \lambda_k)$  of literals to the clause  $\{\lambda_0, \dots, \lambda_k\}$  motivates the following definitions:

**DEFINITION 6.** *Let  $b$  be an assignment,  $C$  a clause, and  $\mathcal{C}$  a set of clauses.*

- (a)  $b$  satisfies  $C$ , if there is  $\lambda \in C$  with  $b \models \lambda$ .

- (b)  $C$  is satisfiable, if there is an assignment which satisfies  $C$ .
- (c)  $b$  satisfies  $C$ , if  $b$  satisfies  $C$  for all  $C \in \mathcal{C}$ .
- (d)  $\mathcal{C}$  is satisfiable, if there is an assignment which satisfies  $\mathcal{C}$ .

Thus a formula in CNF and the set of clauses that corresponds to its conjuncts hold under the same assignments. The empty clause is not satisfiable. Therefore, if  $\emptyset \in \mathcal{C}$ ,  $\mathcal{C}$  is not satisfiable. On the other hand, the empty set of clauses is satisfiable.

With the *resolution method* one can check whether a set  $\mathcal{C}$  of clauses (and therefore, whether a formula in CNF) is satisfiable. The method is based on a single rule; it allows the formation of so-called resolvents.

For a literal  $\lambda$  let  $\lambda^F = \neg p$  if  $\lambda = p$ , and  $\lambda^F = p$  if  $\lambda = \neg p$ .

**DEFINITION 7.** Let  $C, C_1, C_2$  be clauses.  $C$  is called a resolvent of  $C_1$  and  $C_2$ , if there is a literal  $\lambda$  with  $\lambda \in C_1$  and  $\lambda^F \in C_2$  such that

$$(C_1 \setminus \{\lambda\}) \cup (C_2 \setminus \{\lambda^F\}) \subseteq C \subseteq C_1 \cup C_2.^3$$

The transition to resolvents preserves truth in the following sense:

**LEMMA 8 (Resolution Lemma).** Let  $C$  be a resolvent of  $C_1$  and  $C_2$ . Then for every assignment  $b$ ,

$$\text{if } b \models C_1 \text{ and } b \models C_2 \text{ then } b \models C.$$

**Proof.** As  $C$  is a resolvent of  $C_1$  and  $C_2$ , there is a literal  $\lambda$  with  $\lambda \in C_1, \lambda^F \in C_2$ , and  $(C_1 \setminus \{\lambda\}) \cup (C_2 \setminus \{\lambda^F\}) \subseteq C \subseteq C_1 \cup C_2$ . There are two cases:

$b \not\models \lambda$ : Since  $C_1$  holds under  $b$ , there is  $\lambda' \in C_1, \lambda \neq \lambda'$ , with  $b \models \lambda'$ . Since  $\lambda' \in C$ ,  $C$  is satisfied by  $b$ .

$b \models \lambda$ : Then  $b \not\models \lambda^F$ , and we argue similarly with  $C_2$  and  $\lambda^F$ . ■

We show in the appendix to this section that an arbitrary set  $\mathcal{C}$  of clauses is not satisfiable if and only if starting from the clauses in  $\mathcal{C}$  and forming resolvents, one can obtain the empty clause in finitely many steps. Here we show that for unsatisfiable sets of Horn clauses there is a more “direct” way leading to the empty clause.

Horn clauses are clauses stemming from Horn formulas. *Positive Horn*

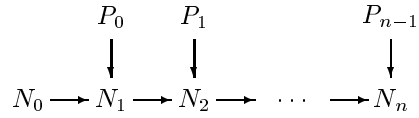
<sup>3</sup>The results that follow below remain valid if, in addition, we require that  $C = (C_1 \setminus \{\lambda\}) \cup (C_2 \setminus \{\lambda^F\})$ . For the purposes of logic programming, however, it is better to give the definition as done above.

clauses are clauses of the form  $\{p\}$  or  $\{\neg q_0, \dots, \neg q_k, p\}$  with  $k \geq 0$ , while negative Horn clauses are of the form  $\{\neg q_1, \dots, \neg q_k\}$  with  $k \geq 0$ . Thus the empty set is a negative clause ( $k = 0$ ). If  $\mathcal{C}$  is a set of Horn clauses, we denote by  $\mathcal{C}^+$  and  $\mathcal{C}^-$  the subset of its positive and negative Horn clauses, respectively.

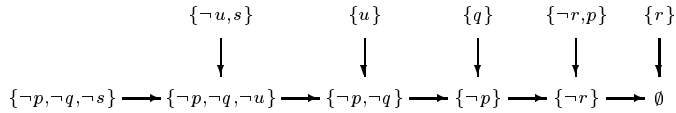
**DEFINITION 9.** *Let  $\mathcal{C}$  be a set of Horn clauses.*

- (a) *A sequence  $N_0, N_1, \dots, N_n$  is a Horn resolution (short: H-resolution) from  $\mathcal{C}$ , if there are  $P_0, \dots, P_{n-1} \in \mathcal{C}^+$  such that*
  - (1)  $N_0, \dots, N_n$  are negative Horn clauses;
  - (2)  $N_0 \in \mathcal{C}^-$ ;
  - (3)  $N_{i+1}$  is a resolvent of  $N_i$  and  $P_i$  for  $i < n$ .
- (b) *A negative Horn clause  $N$  is H-derivable from  $\mathcal{C}$ , if there is an H-resolution  $N_0, \dots, N_n$  from  $\mathcal{C}$  with  $N = N_n$ .*

We represent the ‘‘H-resolution via  $P_0, \dots, P_{n-1}$ ’’ of (a) by



In particular, the steps on page 321 leading to the unsatisfiability of  $\Delta_0$  correspond to the H-resolution



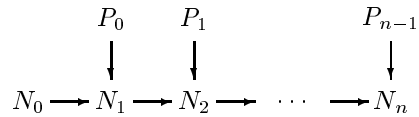
of  $\emptyset$  from the set of clauses corresponding to  $\Delta_0$ .

This relationship holds in general as shown by

**THEOREM 10 (Theorem on the H-Resolution).** *Let  $\mathcal{C}$  be a set of Horn clauses. Then the following are equivalent:*

- (i)  $\mathcal{C}$  is satisfiable.
- (ii)  $\emptyset$  is not H-derivable from  $\mathcal{C}$ .

**Proof.** First, let  $b$  be an assignment satisfying  $\mathcal{C}$  and let



be an H-resolution from  $\mathcal{C}$ . As  $N_0 \in \mathcal{C}$  and  $P_0 \in \mathcal{C}$ , and as  $N_1$  is a resolvent of  $N_0$  and  $P_0$ ,  $b$  is a model of  $N_1$  by the Resolution Lemma 8. Going on in this way, one gets  $b \models N_2, \dots, b \models N_n$ . In particular,  $N_n \neq \emptyset$ . Hence,  $\emptyset$  is not H-derivable from  $\mathcal{C}$ .

The direction from (ii) to (i): The clauses in  $\mathcal{C}^+$  correspond to a set  $\Delta$  of positive Horn formulas. We show:

(\*) If  $k \in \mathbb{N}$  and  $b^\Delta(q_0) = \dots = b^\Delta(q_k) = \text{T}$ , then  $\emptyset$  is H-derivable from  $\mathcal{C}^+ \cup \{\{\neg q_0, \dots, \neg q_k\}\}$ .

Then we are done: Assume (ii). By Lemma 3(c) it suffices to show that  $b^\Delta$  is a model of all clauses in  $\mathcal{C}^-$ . So let  $N \in \mathcal{C}^-$ . By (ii),  $\emptyset$  is not H-derivable from  $\mathcal{C}^+ \cup \{N\} (\subseteq \mathcal{C})$ , in particular,  $N \neq \emptyset$ , say  $N = \{\neg q_0, \dots, \neg q_k\}$ . Thus (\*) shows that there is an  $i \leq k$  with  $b^\Delta(q_i) = \text{F}$ . So  $b^\Delta \models N$ .

To show (\*), we prove by induction on  $l$  that (\*) holds provided each  $q_i$  is underlined during the first  $l$  steps, when applying the underlining algorithm to  $\Delta$ . For  $l = 1$ , the variables  $q_0, \dots, q_k$  are underlined in the first step, hence  $q_0, \dots, q_k \in \Delta$  and therefore,  $\{q_0\}, \dots, \{q_k\} \in \mathcal{C}^+$ . Thus,

$$\begin{array}{ccccccc} & & \{q_0\} & & \{q_{k-1}\} & \{q_k\} & \\ & & \downarrow & & \downarrow & \downarrow & \\ \{\neg q_0, \dots, \neg q_k\} & \longrightarrow & \{\neg q_1, \dots, \neg q_k\} & \longrightarrow & \dots & \longrightarrow & \{\neg q_k\} \longrightarrow \emptyset \end{array}$$

is an H-resolution of  $\emptyset$  from  $\mathcal{C}^+ \cup \{\{\neg q_0, \dots, \neg q_k\}\}$ .

Suppose  $l = m + 1$ , where  $m \geq 1$ . For simplicity, let  $q_0, q_1$  be all the variables among the  $q_i$ 's that are underlined in the  $l$ -th step (the general case being only notationally more complicated). Then, for  $i = 0, 1$ , there is a clause  $(r_{i0} \wedge \dots \wedge r_{im_i} \rightarrow q_i) \in \Delta$  such that  $r_{i0}, \dots, r_{im_i}$  are underlined in the first  $m$  steps. Set

$$N_0 := \{\neg r_{00}, \dots, \neg r_{0m_0}, \neg r_{10}, \dots, \neg r_{1m_1}, \neg q_2, \dots, \neg q_k\}.$$

By induction hypothesis, there is an H-resolution of  $\emptyset$  from  $\mathcal{C}^+ \cup \{N_0\}$ , say

$$\begin{array}{ccccccc} & & P_0 & & P_1 & & P_{n-1} \\ & & \downarrow & & \downarrow & & \downarrow \\ N_0 & \longrightarrow & N_1 & \longrightarrow & N_2 & \longrightarrow & \dots \longrightarrow N_n \end{array}$$

with  $N_n = \emptyset$ . Then

$$\begin{array}{ccccccc} & & \{\neg r_{00}, \dots, \neg r_{0m_0}, q_0\} & & \{\neg r_{10}, \dots, \neg r_{1m_1}, q_1\} & P_0 & & P_{n-1} \\ & & \downarrow & & \downarrow & \downarrow & & \downarrow \\ \{\neg q_0, \dots, \neg q_k\} & \longrightarrow & \{\neg r_{00}, \dots, \neg r_{0m_0}, \neg q_1, \dots, \neg q_k\} & \longrightarrow & N_0 & \longrightarrow & N_1 & \longrightarrow \dots \longrightarrow N_n \end{array}$$

is an H-resolution of  $\emptyset$  from  $\mathcal{C}^+ \cup \{\{\neg q_0, \dots, \neg q_k\}\}$ . ■

As indicated above, Horn resolution (for first-order logic) is essential for logic programming. We turn to it in Section 4.

### 1.1 Appendix

The Theorem on the H-Resolution has a generalization to arbitrary sets of clauses. This appendix is addressed to the reader interested in it.

For an arbitrary set  $\mathcal{C}$  of clauses we let  $\text{Res}_\infty(\mathcal{C})$  be the smallest set of clauses that contains  $\mathcal{C}$  and is closed under the formation of resolvents. Thus, if  $C_1, C_2 \in \text{Res}_\infty(\mathcal{C})$  and  $C$  is a resolvent of  $C_1$  and  $C_2$ , then  $C \in \text{Res}_\infty(\mathcal{C})$ .

**THEOREM 11 (Resolution Theorem).** *For any set  $\mathcal{C}$  of clauses, the following are equivalent:*

- (i)  $\mathcal{C}$  is satisfiable.
- (ii)  $\emptyset \notin \text{Res}_\infty(\mathcal{C})$ .

**Proof.** (i)  $\Rightarrow$  (ii): Let  $b$  be a model of  $\mathcal{C}$ . Then the set

$$\mathcal{C}_b := \{C \mid C \text{ a clause, } b \models C\}$$

contains  $\mathcal{C}$  and is closed under the formation of resolvents (by the Resolution Lemma). Hence,  $\text{Res}_\infty(\mathcal{C}) \subseteq \mathcal{C}_b$  and therefore,  $b$  is a model of  $\text{Res}_\infty(\mathcal{C})$ . In particular,  $\emptyset \notin \text{Res}_\infty(\mathcal{C})$ .

(ii)  $\Rightarrow$  (i): As by the compactness theorem for propositional logic

$\mathcal{C}$  is satisfiable iff each finite subset of  $\mathcal{C}$  is satisfiable

and as

$$\text{Res}_\infty(\mathcal{C}) = \bigcup_{\mathcal{C}_0 \subseteq \mathcal{C}, \mathcal{C}_0 \text{ finite}} \text{Res}_\infty(\mathcal{C}_0),$$

we may assume that  $\mathcal{C}$  is finite. For a contradiction suppose that  $\mathcal{C}$  is a counterexample, i.e.,

- (\*) <sub>$\mathcal{C}$</sub>   $\emptyset \notin \text{Res}_\infty(\mathcal{C})$  and  $\mathcal{C}$  is not satisfiable.

In addition, let  $\mathcal{C}$  be minimal with respect to the number of variables occurring in it. Since the empty set of clauses is satisfiable, we see that  $\mathcal{C} \neq \emptyset$ , and since  $\emptyset \in \text{Res}_\infty(\{\emptyset\})$ , we see that  $\mathcal{C} \neq \{\emptyset\}$ . Thus, there is at least one variable  $p$  occurring in  $\mathcal{C}$ . Without loss of generality we may assume that

- (+) no clause in  $\mathcal{C}$  contains both  $p$  and  $\neg p$ .

Otherwise, we may remove such clauses from  $\mathcal{C}$  without destroying the validity of (\*) <sub>$\mathcal{C}$</sub> .

Let  $\mathcal{C}'$  consist of

- (a) the clauses  $C \in \mathcal{C}$  with  $p \notin C, \neg p \notin C$ ;
- (b) the clauses of the form  $C = C_1 \cup C_2$  such that  $p \notin C_1, C_1 \cup \{p\} \in \mathcal{C}$ , and  $\neg p \notin C_2, C_2 \cup \{\neg p\} \in \mathcal{C}$  (thus,  $C$  is a resolvent of  $C_1 \cup \{p\}$  and  $C_2 \cup \{\neg p\}$ ).

As  $\mathcal{C}' \subseteq \text{Res}_\infty(\mathcal{C})$ , we get  $\text{Res}_\infty(\mathcal{C}') \subseteq \text{Res}_\infty(\mathcal{C})$  and thus, by  $(*)_{\mathcal{C}}$ , that  $\emptyset \notin \text{Res}_\infty(\mathcal{C}')$ . By  $(+)$ ,  $p$  does not occur in  $\mathcal{C}'$ . Hence, the minimality of  $\mathcal{C}$  yields that  $\mathcal{C}'$  is satisfiable. Let  $b$  be a model of  $\mathcal{C}'$  and, say,  $b(p) = \text{T}$ . We show that  $b$  or  $b \stackrel{\text{F}}{p}$  (where  $b \stackrel{\text{F}}{p}$  differs from  $b$  only in assigning F to  $p$ ) is a model of  $\mathcal{C}$ , a contradiction. We distinguish two cases.

*Case 1.* For all  $C' \in \mathcal{C}$ , if  $C' = C_1 \cup \{p\}$  with  $p \notin C_1$ , then  $b \models C_1$ . We show that  $b \stackrel{\text{F}}{p}$  is a model of  $\mathcal{C}$ . Let  $C \in \mathcal{C}$ . If  $p \in C$  then, by assumption,  $b \models C \setminus \{p\}$ ; hence,  $b \stackrel{\text{F}}{p} \models C$ . If  $\neg p \in C$ , then trivially  $b \stackrel{\text{F}}{p} \models C$ . If  $p \notin C, \neg p \notin C$ , then  $C \in \mathcal{C}'$  and hence,  $b \models C$  and therefore,  $b \stackrel{\text{F}}{p} \models C$ .

*Case 2.* For some  $C' = C_1 \cup \{p\} \in \mathcal{C}$  with  $p \notin C_1$  we have  $b \not\models C_1$ . We show that  $b$  is a model of  $\mathcal{C}$ . Let  $C \in \mathcal{C}$ . If  $\neg p \notin C$ , then  $b \models C$ . If  $\neg p \in C$ , say,  $C = C_2 \cup \{\neg p\}$  with  $\neg p \notin C_2$ , we have  $C_1 \cup C_2 \in \mathcal{C}'$  by (b) and hence,  $b \models C_1 \cup C_2$ . As  $b \not\models C_1$ , we get  $b \models C_2$  and therefore,  $b \models C$ . ■

The resolution method provides a further test for satisfiability of clauses (and, hence, of propositional formulas in CNF). However, in contrast to H-resolution for Horn clauses, this test may have exponential running time. For details and a comparison with other methods for checking satisfiability, see e.g. [Urquhart, 1995].

## 2 TERM MODELS

As we have mentioned in the introduction, we are exploring questions of two kinds: questions that ask whether an assertion  $\exists x\varphi(x)$  *follows from a set  $\Phi$  of sentences* and questions that ask whether such an assertion *is true in a particular structure*. In this section we exhibit a relationship between both kinds that is based on models of  $\Phi$  which reflect just the information expressed by  $\Phi$ , so-called term models (Subsections 2.3, 2.4). The relationship lives in the framework of universal sentences. For such sentences the problem of (consequence and) satisfaction can be reduced to propositional logic (Subsection 2.2), thus opening it for the methods of the preceding section. We start by fixing our notation for first-order logic.

### 2.1 First-Order Logic

*Vocabularies* are sets that consist of *relation symbols*  $P, Q, R, \dots$ , *function symbols*  $f, g, h, \dots$ , and *constants*  $c, d, e, \dots$ . Each relation symbol and each function symbol has a positive integer assigned to it, its *arity*.

Fix a vocabulary  $\sigma$ . We let  $\sigma$ -terms be the *variables*  $v_1, v_2, \dots$  (indicated by  $x, y, z, \dots$ ), the constants  $c \in \sigma$ , and the “composed” terms  $f(t_1, \dots, t_n)$  for  $n$ -ary  $f \in \sigma$  and  $\sigma$ -terms  $t_1, \dots, t_n$ .  $T^\sigma$  is the set of  $\sigma$ -terms,  $T_0^\sigma$  the set of  $\sigma$ -terms that contain no variables, so-called *ground terms*.

*First-order  $\sigma$ -formulas* comprise the *atomic formulas*  $Rt_1 \dots t_n$  for  $n$ -ary  $R \in \sigma$  and  $t_1, \dots, t_n \in T^\sigma$  and the  $\sigma$ -formulas  $\neg\varphi, (\varphi \wedge \psi), (\varphi \vee \psi), \forall x\varphi, \exists x\varphi$  for  $\sigma$ -formulas  $\varphi, \psi$  and variables  $x$ .<sup>4</sup>

$L^\sigma$  is the set of  $\sigma$ -formulas,  $L_0^\sigma$  the set of  $\sigma$ -sentences, i.e., of  $\sigma$ -formulas where every variable  $x$  is in the scope of a quantifier  $\forall x$  or  $\exists x$ . We use  $\varphi, \psi, \dots$  as variables for  $\sigma$ -formulas and  $\Phi, \Psi, \dots$  as variables for sets of  $\sigma$ -formulas.

For any quantifier-free  $\sigma$ -formula  $\varphi$ , any variable  $x$ , and any  $\sigma$ -term  $t$ , the  $\sigma$ -formula  $\varphi(x|t)$  arises from  $\varphi$  by replacing  $t$  for  $x$  throughout  $\varphi$ . If  $x$  is clear from the context, say by writing  $\varphi(x)$  for  $\varphi$ , we also use  $\varphi(t)$  for  $\varphi(x|t)$ . With  $\vec{x}$  for  $x_1, \dots, x_n$  and  $\vec{t}$  for  $t_1, \dots, t_n$  we use  $\varphi(\vec{x}|\vec{t})$  to denote the result of simultaneously replacing  $x_1$  by  $t_1, \dots, x_n$  by  $t_n$  in  $\varphi$ .

A  $\sigma$ -structure  $\mathcal{A}$  consists of a nonempty set  $A$ , the *universe* or *domain* of  $\mathcal{A}$ , of an  $n$ -ary relation  $R^A$  over  $A$  for every  $n$ -ary  $R \in \sigma$ , of an  $n$ -ary function  $f^A$  over  $A$  for every  $n$ -ary  $f \in \sigma$ , and of an element  $c^A$  of  $A$  for every constant  $c \in \sigma$ .

For a  $\sigma$ -structure  $\mathcal{A}$  and an  $A$ -assignment  $\beta$ , i.e., a map from  $\{v_1, v_2, \dots\}$  into  $A$ , we write

$$(*) \quad (\mathcal{A}, \beta) \models \varphi$$

if  $(\mathcal{A}, \beta)$  *satisfies* (or is a *model* of)  $\varphi$ . The satisfaction relation  $(*)$  depends only on how  $\mathcal{A}$  interpretes the symbols of  $\sigma$  that actually appear in  $\varphi$  and on the values  $\beta(x)$  for those  $x$  *free* in  $\varphi$ . So for  $\varphi \in L_0^\sigma$  we may write

$$\mathcal{A} \models \varphi$$

instead of  $(*)$ .

We say that  $\varphi$  *follows* from (or is a *consequence* of)  $\Phi$ , written  $\Phi \models \varphi$ , if every model of  $\Phi$  is a model of  $\varphi$ .

A  $\sigma$ -formula  $\varphi$  is *universal* if it is a formula of the form  $\forall \vec{x}\chi$ , where  $\chi$  is quantifier-free. We call  $\chi$  the *kernel* of  $\varphi$ . Quantifier-free formulas are logically equivalent to both formulas in *conjunctive normal form* (CNF) and formulas in *disjunctive normal form* (DNF),

$$\bigwedge_{i=1}^n \bigvee_{j=1}^{m_i} \lambda_{ij} \quad \text{and} \quad \bigvee_{i=1}^n \bigwedge_{j=1}^{m_i} \lambda_{ij},$$

---

<sup>4</sup>Note that we do not include equality. We treat formulas with equality in 2.4.

respectively, where the  $\lambda_{ij}$  are atomic or negated atomic formulas, so-called (first-order) *literals*.

Henceforth, we often omit the prefix “ $\sigma$ -” in connection with formulas and structures when it will be clear from the context or inessential. Moreover, *we always assume that the vocabulary  $\sigma$  contains a constant*. This assumption is not essential, as a variable could serve the same purpose; however, it facilitates the presentation.

## 2.2 Universal Sentences and Propositional Logic

We reduce the problem of satisfiability of universal sentences to that of quantifier-free formulas. This allows us to pass to propositional logic and to translate the results on propositional logic of the previous section to first-order logic.

DEFINITION 12. *A  $\sigma$ -structure  $\mathcal{A}$  is named if for every  $a \in A$  there is a term  $t \in T_0^\sigma$  such that  $a = t^{\mathcal{A}}$ .*

The structures introduced in the following definition are named. They will play a major role later.

DEFINITION 13. *A  $\sigma$ -structure  $\mathcal{A}$  is a Herbrand structure if*

- (a)  $A = T_0^\sigma$ .
- (b) For  $n$ -ary  $f \in \sigma$  and  $t_1, \dots, t_n \in T_0^\sigma$  :  $f^{\mathcal{A}}(t_1, \dots, t_n) = f(t_1, \dots, t_n)$ .
- (c) For  $c \in \sigma$  :  $c^{\mathcal{A}} = c$ .

Clearly, in every Herbrand structure  $\mathcal{A}$  we have  $t^{\mathcal{A}} = t$  for every  $t \in T_0^\sigma$ .

LEMMA 14. *Assume that  $\forall \vec{x} \psi \in L_0^\sigma$  and that  $\psi$  is quantifier-free. Then for every named structure  $\mathcal{A}$ ,*

- (a)  $\mathcal{A} \models \forall \vec{x} \psi$  iff for all  $t_1, \dots, t_n \in T^\sigma$ ,  $\mathcal{A} \models \psi(\vec{x}|\vec{t})$ .
- (b)  $\mathcal{A} \models \exists \vec{x} \psi$  iff there are  $t_1, \dots, t_n \in T^\sigma$  such that  $\mathcal{A} \models \psi(\vec{x}|\vec{t})$ .

The proof is immediate.

THEOREM 15. *Let  $\Phi \subseteq L_0^\sigma$  be a set of universal sentences. Then the following are equivalent:*

- (i)  $\Phi$  is satisfiable.
- (ii) The set  $\text{GI}(\Phi)$  of ground instances of sentences in  $\Phi$ ,

$$\text{GI}(\Phi) := \{ \psi(\vec{x}|\vec{t}) \mid \forall \vec{x} \psi \in \Phi, \psi \text{ quantifier-free, } \vec{t} \in T_0^\sigma \},$$

*is satisfiable.*



**Proof.** The implication (i)  $\Rightarrow$  (ii) is trivial, as  $\forall x^n \psi \models \psi(x|t)$  for all  $t$ .

(ii)  $\Rightarrow$  (i): Let  $\mathcal{A} \models \text{GI}(\Phi)$ . The subset  $B := \{t^A \mid t \in T_0^\sigma\}$  of values in  $\mathcal{A}$  of the ground terms is not empty (because  $\sigma$  contains a constant), contains the values of the constants in  $\sigma$ , and is closed under the functions  $f^A$  for  $f \in \sigma$  (as for  $n$ -ary  $f \in \sigma$  and  $t \in T_0^\sigma$ , we have  $f^A(t_1^A, \dots, t_n^A) = f(t_1, \dots, t_n)^A$ ). Hence,  $B$  is the domain of a substructure  $\mathcal{B}$  of  $\mathcal{A}$ . Clearly,  $\mathcal{B} \models \text{GI}(\Phi)$  (since  $\mathcal{A} \models \text{GI}(\Phi)$ ) and hence,  $\mathcal{B} \models \Phi$  by the preceding lemma, as  $\mathcal{B}$  is named. ■

As an immediate corollary we get:

**THEOREM 16.** (Herbrand's Theorem) *Let  $\Phi$  be a set of universal sentences and  $\exists y^m \psi$  a sentence with quantifier-free  $\psi$ . Assume that*

$$\Phi \models \exists y^m \psi.$$

*Then there are  $l \geq 1$  and ground terms  $t_1^m, \dots, t_l^m$  such that*

$$\Phi \models (\psi(y^m|t_1^m) \vee \dots \vee \psi(y^m|t_l^m)).$$

**Proof.** If  $\Phi \models \exists y^m \psi$ , then

$$\Phi \cup \{\forall y^m \neg \psi\} \text{ is not satisfiable,}$$

hence, by the preceding theorem,

$$\Phi \cup \{\neg \psi(y^m|t) \mid t \in T_0^\sigma\} \text{ is not satisfiable,}$$

so the compactness theorem for first-order logic yields terms  $t_1^m, \dots, t_l^m \in T_0^\sigma$  such that

$$\Phi \cup \{\neg \psi(y^m|t_i^m) \mid 1 \leq i \leq l\} \text{ is not satisfiable,}$$

i.e.

$$\Phi \models (\psi(y^m|t_1^m) \vee \dots \vee \psi(y^m|t_l^m)).$$

■

**REMARK 17.** *For  $\Phi$  and  $\exists y^m \psi$  as in Herbrand's Theorem we get that in case  $\psi$  has a solution, i.e., in case  $\Phi \models \exists y^m \psi$ , there are finitely many tuples  $t_1^m, \dots, t_l^m$  such that in every model  $\mathcal{A}$  of  $\Phi$  at least one of the tuples  $t_1^m, \dots, t_l^m$  is a solution. This is a first step towards our aim to witness the existence of solutions by terms. However, the example*

$$(Rc \vee Rd) \models \exists x Rx$$

shows that, in general, we cannot expect  $l$  to be one. Later we shall see that under further restrictions on  $\Phi$  and  $\exists^m \mathcal{Y}\psi$  we can choose  $l = 1$ .

By Theorem 15 the satisfiability problem for sets of universal sentences is reduced to the satisfiability problem for sets of quantifier-free sentences. Such sentences behave like propositional formulas. We start by clarifying this point.

Let  $\sigma$  be a countable vocabulary that contains a relation symbol (otherwise  $L^\sigma$  is empty). Then the set

$$\text{At}^\sigma := \{Rt_1 \dots t_n \mid R \in \sigma \text{ } n\text{-ary, } t_i \in T^\sigma\}$$

is countably infinite. Let

$$\pi_0 : \text{At}^\sigma \rightarrow \{p_1, p_2, \dots\}$$

be a bijection from the atomic  $\sigma$ -formulas onto the propositional variables. We extend  $\pi_0$  to a map  $\pi$  which is defined on the set of quantifier-free  $\sigma$ -formulas by setting inductively

$$\begin{aligned} \pi(\varphi) &:= \pi_0(\varphi) \text{ for } \varphi \in \text{At}^\sigma \\ \pi(\neg\varphi) &:= \neg\pi(\varphi) \\ \pi((\varphi \wedge \psi)) &:= (\pi(\varphi) \wedge \pi(\psi)) \\ \pi((\varphi \vee \psi)) &:= (\pi(\varphi) \vee \pi(\psi)). \end{aligned}$$

It can easily be seen that  $\pi$  is a bijection between quantifier-free  $\sigma$ -formulas and propositional formulas. The proof of the following lemma is immediate.

LEMMA 18. *Let  $\mathcal{A}$  be a  $\sigma$ -structure and  $b$  be a propositional assignment. Assume that  $\mathcal{A}$  and  $b$  agree on atomic sentences, i.e.*

$$\text{for all sentences } \varphi \in \text{At}^\sigma : \mathcal{A} \models \varphi \text{ iff } b \models \pi(\varphi).$$

*Then  $\mathcal{A}$  and  $b$  agree on quantifier-free sentences, i.e.*

$$\text{for all quantifier-free } \varphi \in L_0^\sigma : \mathcal{A} \models \varphi \text{ iff } b \models \pi(\varphi).$$

PROPOSITION 19.

- (a) *For every  $\sigma$ -structure  $\mathcal{A}$  there is an assignment  $b$  such that  $\mathcal{A}$  and  $b$  agree on quantifier-free sentences.*
- (b) *For every assignment  $b$  there is a Herbrand structure  $\mathcal{A}$  such that  $\mathcal{A}$  and  $b$  agree on quantifier-free sentences.*

**Proof.** (a) Given  $\mathcal{A}$ , define the assignment  $b$  by

$$b(p) = \text{T} \text{ iff } \pi^{-1}(p) \text{ is a sentence and } \mathcal{A} \models \pi^{-1}(p).$$

Then  $\mathcal{A}$  and  $b$  agree on atomic sentences and hence, by the preceding lemma, on quantifier-free sentences.

(b) Given  $b$ , let  $\mathcal{A}$  be the Herbrand structure with

$$R^A t_1 \dots t_n \text{ iff } b \models \pi(Rt_1 \dots t_n).$$

Again,  $\mathcal{A}$  and  $b$  agree on atomic sentences and hence, on quantifier-free sentences. ■

**COROLLARY 20.** *Let  $\Phi \cup \{\psi\}$  be a set of quantifier-free sentences. Then*

(a)  $\Phi$  is satisfiable iff  $\pi(\Phi)$  is satisfiable (iff  $\Phi$  has a Herbrand model).

(b)  $\Phi \models \psi$  iff  $\pi(\Phi) \models \pi(\psi)$ . ■

So with respect to satisfiability we may view quantifier-free sentences as propositional formulas. This allows to apply methods and results of propositional logic. For this purpose we consider universal sentences whose kernels correspond to a propositional Horn formula.

**DEFINITION 21.** *Universal Horn sentences are universal first-order sentences whose kernel is of the form (a), (b), or (c):*

(a)  $\psi$

(b)  $(\neg\psi_0 \wedge \dots \wedge \neg\psi_k \rightarrow \psi)$

(c)  $(\neg\psi_0 \vee \dots \vee \neg\psi_k)$

*with atomic  $\psi, \psi_0, \dots, \psi_k$ . They are positive (or strict) in cases (a), (b), and negative in case (c).*

Let  $\Phi$  be a set of universal Horn sentences. In order to check whether  $\Phi$  is satisfiable, Theorem 15 shows that we may check whether the set  $\text{GI}(\Phi) = \{\psi(x|t) \mid \forall x \psi \in \Phi, \psi \text{ quantifier-free, } t \in T_0^\sigma\}$  is satisfiable. By Corollary 20, its satisfiability is equivalent to that of the set  $\pi(\text{GI}(\Phi))$  of propositional formulas. If  $\Phi$  and  $T_0^\sigma$  are finite,  $\pi(\text{GI}(\Phi))$  is finite and we therefore can use propositional Horn resolution to quickly check satisfiability of  $\Phi$ .

$T_0^\sigma$  is finite if  $\sigma$  contains no function symbols (and only finitely many constants). In Section 5 we will see that the satisfiability problem for finite sets of universal Horn sentences is undecidable if function symbols are allowed. On the other hand, function symbols are often needed in applications (cf. the example at the end of this section). The problem of how to go through infinitely many terms in an “efficient” manner, will be discussed in Section 4.

### 2.3 Minimal Herbrand Models

For sets  $\Delta$  of propositional Horn formulas the minimal assignment  $b^\Delta$  played a special role. For a set  $\Phi$  of universal Horn sentences we are now going to define a minimal Herbrand structure  $\mathcal{H}^\Phi$  that will be of similar importance. For instance, similarly as with  $\Delta$  and  $b^\Delta$ , we can test the satisfiability of  $\Phi$  by just looking at  $\mathcal{H}^\Phi$ .

For the rest of this section, let  $\Phi$  be a set of universal Horn sentences and let  $\Phi^+$  and  $\Phi^-$  be the subsets of  $\Phi$  consisting of the positive and the negative Horn sentences, respectively.

Recall that for a set  $\Delta$  of propositional Horn formulas and a propositional variable  $p$  we have

$$b^\Delta \models p \quad \text{iff} \quad \Delta^+ \models p$$

(cf. Remark 5(1)). This motivates how to fix the relations in the first-order case.

**DEFINITION 22.** *The Herbrand structure  $\mathcal{H}^\Phi$  associated with  $\Phi$  is given by setting for  $n$ -ary  $R \in \sigma$  and  $t_1, \dots, t_n \in T_0^\sigma$ :*

$$R^{\mathcal{H}^\Phi} t_1 \dots t_n \quad : \text{ iff} \quad \Phi^+ \models R t_1 \dots t_n.$$

Note that the atomic sentences in  $L_0^\sigma$  are just the formulas of the form  $R t_1 \dots t_n$  with  $t_1, \dots, t_n \in T_0^\sigma$ . Now the following proposition parallels Lemma 3.

**PROPOSITION 23.**

(a) *For all structures  $\mathcal{A}$  and all  $\overset{n}{t} \in T_0^\sigma$ :*

$$\text{if } \mathcal{A} \models \Phi \text{ and } \mathcal{H}^\Phi \models R t_1 \dots t_n \text{ then } \mathcal{A} \models R t_1 \dots t_n.$$

(b)  $\mathcal{H}^\Phi = \mathcal{H}^{\Phi^+}$ .

(c)  $\mathcal{H}^{\Phi^+} \models \Phi^+$ .

(d)  $\mathcal{H}^\Phi \models \Phi$  iff  $\Phi$  is satisfiable  
iff for all  $\varphi \in \Phi^-$ :  $\Phi^+ \cup \{\varphi\}$  is satisfiable.

**Proof.** (a) If  $\mathcal{A} \models \Phi$  and  $\mathcal{H}^\Phi \models R t_1 \dots t_n$  then, by Definition 22,  $\Phi^+ \models R t_1 \dots t_n$  and thus,  $\mathcal{A} \models R t_1 \dots t_n$ .

(b) is immediate from Definition 22.

(c) Let  $\varphi \in \Phi^+$ , say,  $\varphi = \forall \overset{n}{x} (\psi_0 \wedge \dots \wedge \psi_k \rightarrow \psi)$  (the proof for  $\varphi$  of the form  $\forall \overset{n}{x} \psi$  is even simpler). Let  $t_1, \dots, t_n \in T_0^\sigma$  and assume that

$$(*) \quad \mathcal{H}^{\Phi^+} \models (\psi_0 \wedge \dots \wedge \psi_k)(\overset{n}{x} | \overset{n}{t}).$$

We have to show that  $\mathcal{H}^{\Phi^+} \models \psi(x|t)^n$ . By (\*) and Definition 22,

$$\Phi^+ \models \psi_0(x|t)^n, \dots, \Phi^+ \models \psi_k(x|t)^n$$

and thus, by  $\varphi \in \Phi^+$ ,  $\Phi^+ \models \psi(x|t)^n$ ; hence,  $\mathcal{H}^{\Phi^+} \models \psi(x|t)^n$ .

(d) Clearly, it suffices to show  $\mathcal{H}^\Phi \models \Phi$  in case

(+) for all  $\varphi \in \Phi^-$  :  $\Phi^+ \cup \{\varphi\}$  is satisfiable.

So assume (+). By (b) and (c),  $\mathcal{H}^\Phi \models \Phi^+$ . Let  $\varphi = \forall x^n(\neg\psi_0 \vee \dots \vee \neg\psi_k) \in \Phi^-$ . To prove  $\mathcal{H}^\Phi \models \varphi$ , let  $\mathcal{A}$  be a model of  $\Phi^+ \cup \{\varphi\}$ . Then

$$\text{for all } t \in T_0^\sigma \text{ there is } i \leq k \text{ such that } \mathcal{A} \models \neg\psi_i(x|t)^n$$

and hence, by (a),

$$\text{for all } t \in T_0^\sigma \text{ there is } i \leq k \text{ such that } \mathcal{H}^\Phi \models \neg\psi_i(x|t)^n,$$

i.e.,

$$\text{for all } t \in T_0^\sigma, \mathcal{H}^\Phi \models (\neg\psi_0(x|t)^n \vee \dots \vee \neg\psi_k(x|t)^n)$$

and thus, by Lemma 14,  $\mathcal{H}^\Phi \models \forall x^n(\neg\psi_0 \vee \dots \vee \neg\psi_k)$ . ■

**COROLLARY 24.** *Let  $\Phi = \Phi^+$  and let  $\varphi$  be a negative universal Horn sentence. Then*

$$\Phi \models \neg\varphi \quad \text{iff} \quad \mathcal{H}^\Phi \models \neg\varphi.$$

**Proof.**  $\Phi \models \neg\varphi$  iff  $\Phi \cup \{\varphi\}$  is not satisfiable  
 iff  $\mathcal{H}^{\Phi \cup \{\varphi\}} \not\models \Phi \cup \{\varphi\}$  (by Proposition 23(d))  
 iff  $\mathcal{H}^\Phi \models \neg\varphi$  (by 23(b),(c),  
 as  $(\Phi \cup \{\varphi\})^+ = \Phi$ ). ■

Since  $\neg\forall x^n(\neg\psi_0 \vee \dots \vee \neg\psi_k)$  is equivalent to  $\exists x^n(\psi_0 \wedge \dots \wedge \psi_k)$ , we immediately get the equivalences of (i) and (ii), of (iii) and (iv), and the equivalence in (\*) of the following corollary. The equivalence of (ii) and (iv) holds by Lemma 14(b).

**COROLLARY 25.** *Let  $\Phi$  be a set of positive universal Horn sentences and let  $\exists x^n(\psi_0 \wedge \dots \wedge \psi_k)$  be a sentence with atomic  $\psi_i$ . Then the following are equivalent:*

(i)  $\Phi \models \exists x^n(\psi_0 \wedge \dots \wedge \psi_k)$ .

(ii)  $\mathcal{H}^\Phi \models \exists x^n(\psi_0 \wedge \dots \wedge \psi_k)$ .

(iii) There are  $\overset{n}{t} \in T_0^\sigma$  such that  $\Phi \models (\psi_0(\overset{n}{x}|\overset{n}{t}) \wedge \dots \wedge \psi_k(\overset{n}{x}|\overset{n}{t}))$ .

(iv) There are  $\overset{n}{t} \in T_0^\sigma$  such that  $\mathcal{H}^\Phi \models (\psi_0(\overset{n}{x}|\overset{n}{t}) \wedge \dots \wedge \psi_k(\overset{n}{x}|\overset{n}{t}))$ .

Moreover, the equivalence of (iii) and (iv) is termwise in the sense that for all  $\overset{n}{t} \in T_0^\sigma$ :

(\*)  $\Phi \models (\psi_0(\overset{n}{x}|\overset{n}{t}) \wedge \dots \wedge \psi_k(\overset{n}{x}|\overset{n}{t}))$  iff  $\mathcal{H}^\Phi \models (\psi_0(\overset{n}{x}|\overset{n}{t}) \wedge \dots \wedge \psi_k(\overset{n}{x}|\overset{n}{t}))$ .

REMARK 26. (1) In the situation of the preceding corollary the validity of the implication  $\Phi \models \exists \overset{n}{x} \varphi$ , where  $\varphi = (\psi_0 \wedge \dots \wedge \psi_k)$ , can be tested by looking just at one structure, namely  $\mathcal{H}^\Phi$ . This implies that in case  $\Phi \models \exists \overset{n}{x} \varphi$  there is a single tuple  $\overset{n}{t}$  of terms such that  $\Phi \models \varphi(\overset{n}{x}|\overset{n}{t})$  (in Herbrand's Theorem 16 we needed finitely many tuples).

(2) Assume that  $\Phi$  is satisfiable. Then the model  $\mathcal{H}^\Phi$  of  $\Phi$  is minimal, that is:

if  $\mathcal{A}$  is a Herbrand structure and  $\mathcal{A} \models \Phi$ , then for all  $R \in \sigma$ :  $R^{\mathcal{H}^\Phi} \subseteq R^{\mathcal{A}}$ .

(3) (cf. Remark 5(3)) Let  $\Phi$  be satisfiable. Set

$$\text{CWA}(\Phi) := \{ \neg\psi \mid \psi \text{ an atomic sentence and not } \Phi \models \psi \},$$

the closed world assumption for  $\Phi$ . Then (see Proposition 23) we have that  $\mathcal{H}^\Phi \models \Phi \cup \text{CWA}(\Phi)$ . The closed world assumption goes back to [Reiter, 1978]. It decides to accept  $\neg\psi$  in case  $\psi$  is not deducible. However, one should pay attention to the following point: Consider  $\text{CWA}(\Phi)$  as a logical rule (extending the system of usual first-order rules) that allows to deduce  $\neg\psi$  from  $\Phi$ , if  $\psi$  is not deducible from  $\Phi$ . If  $\neg\psi \in \text{CWA}(\Phi)$ ,  $\neg\psi$  is deducible from  $\Phi$  in the new sense, however, it is not deducible from  $\Phi \cup \{\psi\}$ . So the new calculus lacks one of the fundamental properties of classical logical calculi, namely monotonicity according to which deducibility is preserved under the addition of axioms. For more information on non-monotonic reasoning cf. volume 3 of [Gabbay, Hogger and Robinson, 1993f]; for the role of negation in logic programming, see [Shepherdson, 1988].

(4) The preceding results Proposition 23 and Corollary 24 can also be derived directly from those for propositional logic. We encourage the reader to do so by noting that  $\mathcal{H}^\Phi$  is the Herbrand structure associated with the assignment  $b^\Delta$  for  $\Delta = \pi(\text{GI}(\Phi))$ .

As already pointed out in Remark 26(1), Corollary 25 emphasizes the specific role of the Herbrand structure  $\mathcal{H}^\Phi$ . However, in practice one usually starts with an arbitrary structure. So, the question arises which structures

$\mathcal{A}$  are of the form  $\mathcal{H}^\Phi$  for some set  $\Phi$  of positive universal Horn sentences. In case  $\sigma$  contains a function symbol we have infinitely many terms and  $T_0^\sigma$  will be infinite, so finite structures will not even be Herbrand structures. Once we have explained in the next subsection how to deal with equality, we will see that, to a certain extent, every structure can be viewed as a structure of the form  $\mathcal{H}^\Phi$ .

### 2.4 First-Order Logic with Equality

Denote by  $L^{\sigma,=}$  the set of formulas that we obtain by adding to  $L^\sigma$  atomic formulas of the form

$$s = t$$

where  $s$  and  $t$  are terms. In any  $\sigma$ -structure,  $=$  is interpreted by the equality relation.

To deal with  $L^{\sigma,=}$  in the framework of first-order logic without equality, we take a binary relation symbol  $E_0$  not in  $\sigma$ . Then we let  $\text{Eq}(\sigma)$  be the set consisting of the following sentences (the sentences under (1) say that  $E_0$  is an equivalence relation, those under (2) and (3) say that  $E_0$  is compatible with the relations and functions, respectively):

1.  $\forall x E_0 x x, \forall xy (E_0 xy \rightarrow E_0 yx), \forall xyz (E_0 xy \wedge E_0 yz \rightarrow E_0 xz)$
2.  $\forall \bar{x} \forall \bar{y} (R x_1 \dots x_n \wedge E_0 x_1 y_1 \wedge \dots \wedge E_0 x_n y_n \rightarrow R y_1 \dots y_n)$  for  $n$ -ary  $R \in \sigma$
3.  $\forall \bar{x} \forall \bar{y} (E_0 x_1 y_1 \wedge \dots \wedge E_0 x_n y_n \rightarrow E_0 f(x_1, \dots, x_n) f(y_1, \dots, y_n))$  for  $n$ -ary  $f \in \sigma$ .

Note that  $\text{Eq}(\sigma)$  is a set of positive universal Horn sentences.

If the  $\sigma \cup \{E_0\}$ -structure  $\mathcal{A}$  is a model of  $\text{Eq}(\sigma)$ , we define the  $\sigma$ -structure  $\mathcal{A}/E_0$ , the *quotient structure* of  $\mathcal{A}$  modulo  $E_0^A$ , as follows:

$$A/E_0 := \{\bar{a} \mid a \in A\},$$

where  $\bar{a}$  denotes the equivalence class of  $a$  modulo  $E_0^A$ ;

$$R^{A/E_0} := \{(\bar{a}_1, \dots, \bar{a}_n) \mid a_1, \dots, a_n \in A, \bar{a} \in R^A\};$$

$$f^{A/E_0}(\bar{a}_1, \dots, \bar{a}_n) := \overline{f^A(a_1, \dots, a_n)}.$$

As  $\mathcal{A} \models \text{Eq}(\sigma)$ , this definition makes sense.

For  $\varphi \in L^{\sigma,=}$  let  $\varphi^{E_0}$  be the  $L^{\sigma \cup \{E_0\}}$ -formula that arises from  $\varphi$  if one replaces the equality sign by  $E_0$ . One easily shows by induction on formulas:

For every  $\varphi \in L_0^{\sigma,=}$  and every  $\sigma \cup \{E_0\}$ -structure  $\mathcal{A}$  with  $\mathcal{A} \models \text{Eq}(\sigma)$ ,

$$(1) \quad \mathcal{A}/E_0 \models \varphi \quad \text{iff} \quad \mathcal{A} \models \varphi^{E_0},$$

and one concludes for  $\Phi \cup \{\psi\} \subseteq L_0^{\sigma,=}$ :

$$(2) \quad \Phi \models \psi \quad \text{iff} \quad \Phi^{E_0} \cup \text{Eq}(\sigma) \models \psi^{E_0},$$

where  $\Phi^{E_0} := \{\varphi^{E_0} \mid \varphi \in \Phi\}$ . Using these equivalences, we can translate problems of first-order logic with equality into problems of first-order logic without equality. In particular, for a set  $\Phi \subseteq L_0^{\sigma,=}$  of universal Horn sentences they tell us how to define the Herbrand structure  $\mathcal{H}^\Phi$  associated with  $\Phi$  (by definition,  $\psi \in L_0^{\sigma,=}$  is a universal Horn sentence, if  $\psi^{E_0}$  is): By (2), the role of  $\Phi$  is taken over by the set  $\Phi^{E_0} \cup \text{Eq}(\sigma)$ . Since the sentences in  $\text{Eq}(\sigma)$  are positive, its Herbrand structure  $\mathcal{H}^{\Phi^{E_0} \cup \text{Eq}(\sigma)}$  is a model of  $\text{Eq}(\sigma)$ ; hence,  $\mathcal{H}^{\Phi^{E_0} \cup \text{Eq}(\sigma)}/E_0$  is well-defined. Now, we set

$$(3) \quad \mathcal{H}^\Phi \quad := \quad \mathcal{H}^{\Phi^{E_0} \cup \text{Eq}(\sigma)}/E_0$$

Then the results on universal Horn sentences given above survive in the presence of equality. We illustrate this by proving the analogue of Corollary 24.

**COROLLARY 27.** *Let  $\Phi \subseteq L_0^{\sigma,=}$  be a set of positive universal Horn sentences and let  $\varphi \in L_0^{\sigma,=}$  be a negative universal Horn sentence. Then*

$$\Phi \models \neg\varphi \quad \text{iff} \quad \mathcal{H}^\Phi \models \neg\varphi.$$

**Proof.**

$$\begin{aligned} \Phi \models \neg\varphi & \quad \text{iff} \quad \Phi^{E_0} \cup \text{Eq}(\sigma) \models \neg\varphi^{E_0} && \text{(by (2))} \\ & \quad \text{iff} \quad \mathcal{H}^{\Phi^{E_0} \cup \text{Eq}(\sigma)} \models \neg\varphi^{E_0} && \text{(by Corollary 24)} \\ & \quad \text{iff} \quad \mathcal{H}^{\Phi^{E_0} \cup \text{Eq}(\sigma)}/E_0 \models \neg\varphi && \text{(by (1))} \\ & \quad \text{iff} \quad \mathcal{H}^\Phi \models \neg\varphi && \text{(by (3)).} \quad \blacksquare \end{aligned}$$

**REMARK 28.** *If  $\Phi$  is a set of universal Horn sentences that does not contain the equality sign, it does not matter whether we view  $\Phi$  as a subset of  $L_0^\sigma$  or as a subset of  $L_0^{\sigma,=}$ : The minimal Herbrand models  $\mathcal{H}^\Phi$  that we get by viewing  $\Phi$  as a subset of  $L_0^\sigma$  and as a subset of  $L_0^{\sigma,=}$  are the same (up to isomorphism). To show this, note that*

$$\Phi^+ \models t_1 = t_2 \quad \text{iff} \quad t_1 = t_2.$$

We come back to the problem of how a given structure can be viewed as the Herbrand structure  $\mathcal{H}^\Phi$  for a suitable  $\Phi$ .



Let  $\mathcal{A}$  be a  $\sigma$ -structure and set

$$\sigma(A) := \sigma \cup \{c_a \mid a \in A\}$$

where the  $c_a$  are new constants. The *positive diagram*  $D(\mathcal{A})$  of  $\mathcal{A}$  consists of the following  $\sigma(A)$ -sentences:

- (1)  $Rc_{a_1} \dots c_{a_n}$  for  $n$ -ary  $R \in \sigma$  and  $\overset{n}{a} \in R^A$ ;
- (2)  $f(c_{a_1}, \dots, c_{a_n}) = c_a$  for  $n$ -ary  $f \in \sigma$ ,  $\overset{n}{a}, a \in A$ , and  $f^A(a_1, \dots, a_n) = a$ ;
- (3)  $c_a = c$  for  $c \in \sigma$ ,  $a \in A$ , and  $c^A = a$ .

A simple induction on terms using the sentences in (2) and (3) shows:

- (4) For every  $t \in T_0^{\sigma(A)}$  there is an  $a \in A$  such that  $D(\mathcal{A}) \models t = c_a$ .

If we denote by  $(\mathcal{A}, (a)_{a \in A})$  the  $\sigma(A)$ -structure where  $c_a$  is interpreted by  $a$ , we therefore get:

**PROPOSITION 29.**  $\mathcal{H}^{D(\mathcal{A})} \cong (\mathcal{A}, (a)_{a \in A})$ , and hence, we have for the  $\sigma$ -reduct  $\mathcal{H}^{D(\mathcal{A})}|_\sigma$  of  $\mathcal{H}^{D(\mathcal{A})}$ :

$$\mathcal{H}^{D(\mathcal{A})}|_\sigma \cong \mathcal{A}.$$

The next example will show how we may apply the “ubiquity” of Herbrand structures in a concrete situation.

### 2.5 An Example

Using the results of the preceding subsections and the concept of diagram we analyze one of the examples indicated in the introduction.

Consider a directed graph, i.e., a structure  $\mathcal{G} = (G, E^G)$  with binary  $E$  such that  $\mathcal{G} \models \forall x \neg Exx$ . Imagine that the elements of  $G$  are the towns of a country and that  $(a \ b) \in E^G$  means that a certain bus company offers a direct connection from  $a$  to  $b$ . Then the question whether two persons living in towns  $a_1, a_2$ , respectively, can meet in some town  $b$  getting there by buses of the company, comes up to the question:

- (1) Is there a town  $b$  s.t. there are  $E^G$ -paths from  $a_1$  to  $b$  and from  $a_2$  to  $b$ ?

To give a first-order formulation, we introduce a new binary relation symbol  $C$  for connections possibly requiring a change of buses and set

$$\mathcal{G}' := (G, E^G, C^G),$$

where

$$C^G := \{(a \ b) \in G \times G \mid \text{there is an } E^G\text{-path from } a \text{ to } b\}$$

(by definition, there is an  $E^G$ -path from  $a$  to  $a$ ). Then (1) is equivalent to

$$(2) \quad \mathcal{G}' \models \exists z(Cxz \wedge Cyz)[a_1, a_2]?^5$$

Since  $\mathcal{H}^{D(\mathcal{G}')} \cong (\mathcal{G}', (a)_{a \in G})$ , by Corollary 25 we get that (2) is equivalent to

$$(3) \quad D(\mathcal{G}') \models \exists z(Cc_{a_1}z \wedge Cc_{a_2}z)?$$

Thus we have arrived at a formulation of (1) that falls under the “entailment form” we have been considering so far. Of course, once the data of  $\mathcal{G}'$  are available, one only has to go through them in an obvious way to obtain an answer to (2). However, in practice it may happen that only the data of the original  $\mathcal{G}$  are stored. Then, to the “data”  $D(\mathcal{G})$  corresponding to  $\mathcal{G}$  we add the “production rules” defining  $C^G$ . More precisely, we set

$$\Phi := \{\forall x Cxx, \forall x \forall y \forall z (Cxy \wedge Eyz \rightarrow Cxz)\}$$

and convince ourselves that (1) is equivalent to

$$(4) \quad D(\mathcal{G}) \cup \Phi \models \exists z(Cc_{a_1}z \wedge Cc_{a_2}z)?$$

(It suffices to show that (3) and (4) are equivalent; for this purpose prove that  $\mathcal{H}^{D(\mathcal{G}) \cup \Phi} \cong \mathcal{H}^{D(\mathcal{G}')}.$ )

The framework we have established so far does not suffice to give us paths or even a list of paths leading from  $a_1$  and  $a_2$  to a meeting point  $b$ . The reason simply is that we are missing adequate means to name connections. We therefore revise our model, replacing the relation symbol  $C$  by a ternary relation symbol  $P$  together with a binary function symbol  $f$ . Intuitively,

$$\begin{aligned} f(x, y) &\text{ represents a hypothetical path from } x \text{ to } y, \\ f(f(x, y), z) &\text{ represents a hypothetical path from } x \text{ via } y \text{ to } z, \\ Pxyz &\text{ says that } z \text{ is a “real” path from } x \text{ to } y. \end{aligned}$$

Hence, the hypothetical path  $a \rightarrow b \rightarrow d \rightarrow a \rightarrow e$  is represented by the term

$$t := f(f(f(f(c_a, c_b), c_d), c_a), c_e)$$

---

<sup>5</sup>Clearly, if the variables free in  $\varphi(x_1, \dots, x_n)$  are among  $x_1, \dots, x_n$ , then  $\mathcal{A} \models \varphi[a_1, \dots, a_n]$  means that  $\varphi$  holds in  $\mathcal{A}$  if  $x_1$  is interpreted by  $a_1$ ,  $x_2$  by  $a_2$ ,  $\dots$ , and  $x_n$  by  $a_n$ .

and

$$Pc_a c_e t$$

means that  $a \rightarrow b \rightarrow d \rightarrow a \rightarrow e$  is a real path from  $a$  to  $e$  in  $\mathcal{G}$ , that is,  $(a b), (b, d), (d, a), (a e) \in E^{\mathcal{G}}$ .

The “production rules” defining  $P$  are

$$\Phi' := \{\forall x P x x x, {}^6 \forall u \forall x \forall y \forall z (P x y u \wedge E y z \rightarrow P x z f(u, z))\}.$$

For the Herbrand structure  $\mathcal{H}^{D(\mathcal{G}) \cup \Phi'}$  we have

$$\mathcal{H}^{D(\mathcal{G}) \cup \Phi'} \models Pc_a c_b t \quad \text{iff} \quad t \text{ represents a path from } a \text{ to } b.$$

But then our original question (1) is equivalent to

$$\mathcal{H}^{D(\mathcal{G}) \cup \Phi'} \models \exists z \exists u \exists v (Pc_{a_1} z u \wedge Pc_{a_2} z v)?$$

and hence, by Corollary 25, to

$$D(\mathcal{G}) \cup \Phi' \models \exists z \exists u \exists v (Pc_{a_1} z u \wedge Pc_{a_2} z v)?$$

And

$$D(\mathcal{G}) \cup \Phi' \models Pc_{a_1} t t_1 \wedge Pc_{a_2} t t_2$$

is equivalent to the statement

$t$  is a town,  $t_1$  a path from  $a_1$  to  $t$ , and  $t_2$  a path from  $a_2$  to  $t$ .

### 3 DATALOG

Some of the notions, methods, and tools we have developed so far, play a role in the analysis of query languages for databases. In this section we consider an example of such a language, DATALOG, and point out similarities and differences. Sometimes, query languages are designed with the aim in mind to capture all queries which can be answered by algorithms of a given complexity. In Section 5 we show that DATALOG captures PTIME in this sense.

So far we mainly analyzed relations between

$$\Phi \models \exists \overset{n}{x} (\psi_0 \wedge \dots \wedge \psi_k) \quad \text{and} \quad \Phi \models (\psi_0 \wedge \dots \wedge \psi_k) (\overset{n}{x} | \overset{n}{t})$$

where  $\Phi$  is a set of positive universal Horn sentences, the  $\psi_i$  are atomic, and  $\overset{n}{t} \in T_0^\sigma$ . We know that

$$\Phi \models (\psi_0 \wedge \dots \wedge \psi_k) (\overset{n}{x} | \overset{n}{t}) \quad \text{iff} \quad \mathcal{H}^\Phi \models (\psi_0 \wedge \dots \wedge \psi_k) (\overset{n}{x} | \overset{n}{t}),$$

---

<sup>6</sup>So  $x$  represents the “empty path” from  $x$  to  $x$ .

where  $\mathcal{H}^\Phi$  is the Herbrand structure associated with  $\Phi$  (cf. Definition 22). The set

$$(*) \quad \{(t_1, \dots, t_n) \mid \Phi \models (\psi_0 \wedge \dots \wedge \psi_k)(x|t)\}$$

gives a new  $n$ -ary relation on the universe of  $\mathcal{H}^\Phi$ . Let  $R$  be a new  $n$ -ary relation symbol and set

$$\Phi_1 := \Phi \cup \{\forall x^n (\psi_0 \wedge \dots \wedge \psi_k \rightarrow Rx)\}.$$

One easily shows that

$$\Phi_1 \models R^n t \quad \text{iff} \quad \Phi \models (\psi_0 \wedge \dots \wedge \psi_k)(x|t)$$

and that

$$\mathcal{H}^{\Phi_1} = (\mathcal{H}^\Phi, R^{H^{\Phi_1}}),$$

where  $R^{H^{\Phi_1}}$  is the relation given by (\*).

It is this aspect of defining new relations from given ones (we already encountered in the example at the end of the previous section) that is important for DATALOG. However, it comes with several generalizations:

- we may define several new relations (instead of a single one) which, in addition, are allowed to occur in the bodies  $(\psi_0 \wedge \dots \wedge \psi_k)$ ;
- the old relation symbols and the equality sign may also occur negated in  $(\psi_0 \wedge \dots \wedge \psi_k)$ ;
- we consider arbitrary structures, not only Herbrand structures.

Now the precise notions.

**DEFINITION 30.** Fix a vocabulary  $\sigma$ . A DATALOG program  $\Pi$  over  $\sigma$  is a finite set of formulas of  $L^{\sigma,=}$  of the form

$$(+)$$

$$(\lambda_1 \wedge \dots \wedge \lambda_l \rightarrow \lambda)$$

where  $l \geq 0$ ,  $\lambda_1, \dots, \lambda_l$  are literals, and  $\lambda$  is atomic of the form  $Rt^n$  (so  $\lambda$  does not contain the equality sign). We call  $\lambda$  the head and  $(\lambda_1 \wedge \dots \wedge \lambda_l)$  the body of (+). The relation symbols occurring in the head of some formula of  $\Pi$  are intentional; the remaining symbols of  $\sigma$  are extensional. We denote the set of intentional symbols by  $\sigma_{\text{int}}$  ( $= \sigma_{\text{int}}^\Pi$ ) and the set of extensional symbols by  $\sigma_{\text{ext}}$ . Hence,  $\sigma_{\text{ext}} = \sigma \setminus \sigma_{\text{int}}$ . Finally, we require that no intentional symbol occurs negated in the body of any formula of  $\Pi$ . The formulas of  $\Pi$  are often called rules or clauses, and (+) is often written in the form  $\lambda_1, \dots, \lambda_l \rightarrow \lambda$  (or in the form  $\lambda \leftarrow \lambda_1, \dots, \lambda_l$ ).

Before giving a precise definition of the semantics of DATALOG programs, we consider a concrete example.

EXAMPLE 31. Let  $\sigma = \{E, C, P\}$  with binary  $E, C$  and unary  $P$ . Let  $\Pi_0$  be the DATALOG program which consists of the rules

- (1)  $Exy \rightarrow Cxy$
- (2)  $Cxy, \neg Py, Eyz \rightarrow Cxz$ .

Hence,  $\sigma_{\text{int}} = \{C\}$  and  $\sigma_{\text{ext}} = \{E, P\}$ .

Given an  $\{E, P\}$ -structure or “relational database”  $\mathcal{A} = (A, E^A, P^A)$ , the program  $\Pi_0$  defines a relation  $C^A$  on  $A$ .  $C^A$  is the union of “levels”  $C_0^A, C_1^A, \dots$  that are successively generated by viewing the formulas of  $\Pi_0$  as rules:

$$C_0^A := \emptyset$$

and

$$(a \ b) \in C_{i+1}^A \quad \text{iff} \quad (a \ b) \in E^A \text{ (cf. (1))}$$

$$\text{or there is } d \in A \text{ such that } (a \ d) \in C_i^A,$$

$$d \notin P^A, \text{ and } (d, b) \in E^A \text{ (cf. (2)).}$$

Then

$$C^A := \bigcup_{i \geq 0} C_i^A.$$

Note that  $C_0^A \subseteq C_1^A \subseteq C_2^A \subseteq \dots$

Obviously,  $C_i^A$  contains those pairs  $(a \ b)$  such that there is an  $E^A$ -path from  $a$  to  $b$  of length  $\leq i$  that does not pass through  $P^A$ . So  $C^A$  consists of those pairs  $(a \ b)$  for which there is an  $E^A$ -path from  $a$  to  $b$  that does not pass through  $P^A$ .

There is a different way to define (the same)  $C^A$  that is more in the spirit of the preceding sections: We form the vocabulary  $\sigma(A) := \sigma \cup \{c_a \mid a \in A\}$ , where the  $c_a$  are new constants, and let  $\text{GI}(\Pi, A)$  be the set of ground instances of  $\Pi$  in this vocabulary, i.e.,  $\text{GI}(\Pi, A)$  consists of the sentences of the form (1') or (2'):

- (1')  $Ec_a c_b \rightarrow Cc_a c_b$
- (2')  $Cc_a c_b, \neg Pc_b, Ec_b c_d \rightarrow Cc_a c_d$

for  $a, b, d \in A$ . Suppose that  $b_0 \in P^A$ . Then, for  $b = b_0$  (and arbitrary  $a, d \in A$ ), the rule in (2') never can “fire”, since  $\neg Pc_{b_0}$  gets the value F (false) in  $(\mathcal{A}, (a)_{a \in A})$ . This example shows that we can omit from  $\text{GI}(\Pi, A)$  all the ground instances which contain literals false in  $(\mathcal{A}, (a)_{a \in A})$ . Now,

suppose that  $b_0 \notin P^A$ . Then, the literal  $\neg P c_{b_0}$  always gets the value T (true); so we can delete such true literals in ground instances. Altogether, we obtain from  $\text{GI}(\Pi, A)$  a modified set  $\text{GI}(\Pi, \mathcal{A})$  that only contains positive literals and no extensional symbols, namely

$$(1'') \quad C c_a c_b \text{ if } (a \ b) \in E^A$$

$$(2'') \quad C c_a c_b \rightarrow C c_a d_a \text{ if } b \notin P^A \text{ and } (b, d) \in E^A.$$

Now we can apply the underlining algorithm (cf. Section 1) to  $\text{GI}(\Pi, \mathcal{A})$ , viewing the formulas in  $\text{GI}(\Pi, \mathcal{A})$  as propositional ones. It is easy to see that  $(a \ b) \in C^A$  iff  $C c_a c_b$  gets underlined this way.

We give a precise definition of the semantics of DATALOG that follows this approach.

Let  $\Pi$  be a DATALOG program over  $\sigma$ . Fix a  $\sigma_{\text{ext}}$ -structure  $\mathcal{A}$  and consider the set  $\text{GI}(\Pi, A)$  of ground instances in the vocabulary  $\sigma(A) := \sigma \cup \{c_a \mid a \in A\}$ . Pass from  $\text{GI}(\Pi, A)$  to  $\text{GI}(\Pi, \mathcal{A})$  by successively

- replacing every term  $t$  by  $c_b$  if  $b = t^{(\mathcal{A}, (a)_{a \in A})}$ ;
- deleting all instances that contain a literal false in  $(\mathcal{A}, (a)_{a \in A})$ ;
- deleting literals that are true in  $(\mathcal{A}, (a)_{a \in A})$ .

Note that the clauses in  $\text{GI}(\Pi, \mathcal{A})$  are of the form  $\gamma_1, \dots, \gamma_m \rightarrow \gamma$  where the atomic parts are of the form  $R c_{a_1} \dots c_{a_n}$  with  $R \in \sigma_{\text{int}}$  and  $\overset{n}{a} \in A$ . Now apply the underlining algorithm to  $\text{GI}(\Pi, \mathcal{A})$ . For an  $n$ -ary  $R \in \sigma_{\text{int}}$  set

$$R^A := \{(a_1, \dots, a_n) \mid R c_{a_1} \dots c_{a_n} \text{ has been underlined}\}$$

and, if  $\sigma_{\text{int}} = \{R_1, \dots, R_l\}$ , let

$$\mathcal{A}(\Pi) := (\mathcal{A}, R_1^A, \dots, R_l^A).$$

A DATALOG *formula* or DATALOG *query* has the form  $(\Pi, R) \overset{n}{x}$  where  $\Pi$  is a DATALOG program and  $R$  is an  $n$ -ary intentional relation symbol.  $(\Pi, R) \overset{n}{x}$  is a formula of vocabulary  $\sigma_{\text{ext}}$ . Its meaning is given by setting for a  $\sigma_{\text{ext}}$ -structure  $\mathcal{A}$  and  $\overset{n}{a} \in A$

$$\mathcal{A} \models (\Pi, R) \overset{n}{x} [\overset{n}{a}] \text{ iff } \overset{n}{a} \in R^{\mathcal{A}(\Pi)}.$$

To compare the expressive power of DATALOG with that of other logics, it is desirable to have something like DATALOG sentences. For this purpose one also admits zero-ary relation symbols  $R$ . Then  $(\Pi, R)$  is a DATALOG sentence. When evaluating  $\Pi$  in a  $\sigma_{\text{ext}}$ -structure  $\mathcal{A}$ , the value of  $R^A$  will be T if  $R$  is finally underlined, and F otherwise. So,

$$\mathcal{A} \models (\Pi, R) \text{ iff } R \text{ gets the value T.}$$

EXAMPLE 32. Let  $R$  be zero-ary and extend the DATALOG program of Example 31 to

$$\Pi' := \{(1), (2), Ccd \rightarrow R\}.$$

Then  $\sigma_{\text{int}}^{\Pi'} = \{C, R\}$  and  $\sigma_{\text{ext}}^{\Pi'} = \{E, P, c, d\}$ , and for any  $\sigma_{\text{ext}}^{\Pi'}$ -structure  $(\mathcal{A}, a \ b)$ ,

$$(\mathcal{A}, a \ b) \models (\Pi', R) \quad \text{iff} \quad \begin{array}{l} \text{there is an } E^A\text{-path from } a \text{ to } b \\ \text{which does not pass through } P^A. \end{array}$$

The relationship between DATALOG and the framework that we have developed in the preceding sections is illustrated by the following easy facts:

REMARK 33. (1) Let  $\Pi$  be a DATALOG program that contains only formulas  $\lambda_1, \dots, \lambda_n \rightarrow \lambda$  of vocabulary  $\sigma$  where the  $\lambda_i$  are atomic. Let  $\Phi(\Pi)$  consist of the positive universal Horn sentences

$$\forall \vec{x} (\lambda_1 \wedge \dots \wedge \lambda_l \rightarrow \lambda)$$

where  $\lambda_1, \dots, \lambda_l \rightarrow \lambda \in \Pi$  and  $\vec{x}$  are the variables in  $\lambda_1, \dots, \lambda_l \rightarrow \lambda$  (in some fixed order). Then, for every  $\sigma_{\text{ext}}$ -structure  $\mathcal{A}$ ,  $n$ -ary  $R \in \sigma_{\text{int}}$  and  $\vec{t} \in T_0^\sigma$ ,

$$R^{\mathcal{A}(\Pi)} \vec{t} \quad \text{iff} \quad D(\mathcal{A}) \cup \Phi(\Pi) \models R \vec{t}$$

(recall that  $D(\mathcal{A})$  denotes the positive diagram of  $\mathcal{A}$  (cf. Remark 28)). Hence,

$$\mathcal{A}(\Pi) = \mathcal{H}^{D(\mathcal{A}) \cup \Phi(\Pi)} |_\sigma.$$

To a certain extent the restriction on the  $\lambda_i$ 's is not essential, as negated  $\lambda_i$ 's can be replaced by their complements. For example, for any  $\{P\}$ -structure  $\mathcal{A} = (\mathcal{A}, P^A)$ , the program  $\Pi = \{Px, \neg Py \rightarrow Rxy\}$  gives the same meaning to  $R$  in  $\mathcal{A}$  as the program  $\Pi' = \{Px, Qy \rightarrow Rxy\}$  gives to  $R$  in  $(\mathcal{A}, Q^A)$  where  $Q^A$  is the complement  $A \setminus P^A$  of  $P^A$ .

(2) As in the introduction to this section, let  $\Phi$  be a set of positive universal Horn sentences from  $L_0^\sigma$ , let  $\psi_0(\vec{x}), \dots, \psi_k(\vec{x}) \in L^\sigma$  be atomic,  $\vec{t} \in T_0^\sigma$ , and  $R$  a new  $n$ -ary relation symbol. For the DATALOG program  $\Pi := \{\psi_0, \dots, \psi_k \rightarrow R\vec{x}\}$  (hence,  $\sigma_{\text{int}} = \{R\}$ ) one easily gets the equivalence of  $\Phi \models (\psi_0 \wedge \dots \wedge \psi_k)(\vec{x} | \vec{t})$  and  $\mathcal{H}^\Phi \models (\Pi, R)\vec{t}$ .

Part (2) of this remark shows how questions concerning the entailment relation can be treated within DATALOG, whereas (1) aims at the other

direction by showing us that the evaluation of  $R^{A(\Pi)} \overset{n}{t}$  can be reduced to the entailment relation  $D(\mathcal{A}) \cup \Phi(\Pi) \models R \overset{n}{t}$ . Altogether, we see a close relationship between the kind of entailment relations studied in the previous section and the kind of database queries addressed in this section. However, the two approaches stand for different aspects: resolution first aims at consequence relations of the form  $\Phi \models (\psi_0 \wedge \dots \wedge \psi_k) \overset{n}{x} \overset{n}{t}$ , whereas DATALOG first aims at a quick and uniform evaluation of queries of the form “ $\mathcal{A} \models (\Pi, R) \overset{n}{x} \overset{n}{[a]}$ ?”, uniform in  $\mathcal{A}$ ,  $\overset{n}{a}$ , and also in  $(\Pi, R) \overset{n}{x}$ . For fixed  $(\Pi, R) \overset{n}{x}$  these queries can be evaluated in time polynomial in the cardinality  $|A|$  of  $A$ :

**THEOREM 34.** *DATALOG queries can be evaluated in polynomial time, that is, given a DATALOG formula  $(\Pi, R) \overset{n}{x}$ , there is an algorithm  $\mathbf{A}$  and a polynomial  $f$  such that  $\mathbf{A}$  applied to (the coding <sup>7</sup> of) a finite  $\sigma_{\text{ext}}$ -structure  $A$  and any  $\overset{n}{a} \in A$  decides in  $\leq f(|A|)$  steps whether  $\mathcal{A} \models (\Pi, R) \overset{n}{x} \overset{n}{[a]}$ .*

**Proof.** Let  $\mathcal{A}$  be a finite  $\sigma_{\text{ext}}$ -structure. Recall the definition of the semantics of DATALOG programs. Note that we can pass from  $\mathcal{A}$  and  $\Pi$  to the set  $\text{GI}(\Pi, \mathcal{A})$  in a number of steps polynomial in  $|A|$ . Now it suffices to show that we obtain the values  $R_1^{A(\Pi)}, \dots, R_l^{A(\Pi)}$  of the intentional symbols in time polynomial in  $|A|$ . Let  $R_i$  be  $r_i$ -ary and set  $r := \max\{r_1, \dots, r_l\}$ . For  $s \geq 1$  let

$$R_i^s := \{(a_1, \dots, a_{r_i}) \mid R c_{a_1} \dots c_{a_{r_i}} \text{ is underlined during the first } s \text{ steps of the underlining algorithm}\}.$$

Clearly,

- $R_i^1 \subseteq R_i^2 \subseteq R_i^3 \subseteq \dots$
- $R_i^{A(\Pi)} = \bigcup_{s \geq 1} R_i^s$
- if for some  $m$

$$(*) \quad R_1^m = R_1^{m+1}, \dots, R_l^m = R_l^{m+1},$$

then for all  $s \geq 1$

$$R_1^m = R_1^{m+s} = R_1^{A(\Pi)}, \dots, R_l^m = R_l^{m+s} = R_l^{A(\Pi)}.$$

Since in the disjoint union of  $A^{r_1}, \dots, A^{r_l}$  there are  $\leq l \cdot |A|^r$  tuples, we see that (\*) must hold for some  $m \leq l \cdot |A|^r$ . ■

In Section 5 we prove the converse of the theorem: Queries evaluable in polynomial time can be expressed by DATALOG formulas.

<sup>7</sup>An explicit coding of finite structures is given in Section 5.



REMARK 35. *The precise semantics for DATALOG that we have introduced above provides an efficient way for evaluating the intentional predicates. We sketch another equivalent way of introducing the semantics that follows the first approach illustrated in Example 31. For a DATALOG program  $\Pi$  this approach makes more visible the uniform character of the rules of  $\Pi$  that in the definitions given above lies somewhat hidden under the (modified) set of ground instances.*

Let  $\Pi$  be a DATALOG program over  $\sigma$ . We assume that all heads in  $\Pi$  that belong to the same symbol  $R$  have the form  $R\bar{x}$  with a fixed tuple  $\bar{x}$  of distinct variables. (Otherwise, we replace, for instance,  $Tz, Px_1 \rightarrow Rz$  by  $Tx_1, x_1 = x_2, Pz \rightarrow Rx_1x_2$ .) Then we set

$$\varphi_R(\bar{x}) := \bigvee \{ \exists \bar{y} (\lambda_1 \wedge \dots \wedge \lambda_k) \mid \lambda_1, \dots, \lambda_k \rightarrow R\bar{x} \in \Pi \},$$

where  $\bar{y}$  is the tuple of those variables in  $\lambda_1, \dots, \lambda_k$  that are different from  $x_1, \dots, x_n$ .

Let  $R_1, \dots, R_l$  be the intentional symbols of  $\Pi$ ,  $R_i$  of arity  $r_i$ . For  $s \geq 0$  define the  $r_i$ -ary relation  $R_i^s$  on  $A$  by

$$(1) \quad \begin{aligned} R_i^0 &:= \emptyset \\ R_i^{s+1} &:= \{ \bar{a} \in A^{r_i} \mid (\mathcal{A}, R_1^s, \dots, R_l^s) \models \varphi_{R_i}[\bar{a}] \}. \end{aligned}$$

As the  $R_j$  occur only unnegated in  $\varphi_{R_i}(\bar{x})$ , we have

$$R_i^0 \subseteq R_i^1 \subseteq R_i^2 \subseteq \dots$$

It is not hard to show that

$$R_i^{A(\Pi)} = \bigcup_{s \geq 0} R_i^s$$

(in fact,  $R_i^s$  is the set of tuples  $\bar{a} \in A^{r_i}$  such that  $R_i c_{a_1} \dots c_{a_{r_i}}$  is underlined in the first  $s$  steps of the underlining algorithm applied to  $\text{GI}(\Pi, \mathcal{A})$ ).

In particular, for  $i = 1, \dots, l$  we have

$$(2) \quad \mathcal{A}(\Pi) \models \forall \bar{x} (R_i \bar{x} \leftrightarrow \varphi_{R_i}(\bar{x})).$$

The transition from  $(R_1^s, \dots, R_l^s)$  to  $(R_1^{s+1}, \dots, R_l^{s+1})$  given by  $(\varphi_{R_1}(\bar{x}^1), \dots, \varphi_{R_l}(\bar{x}^l))$  in (1) above, can be considered for arbitrary "arguments"  $(M_1, \dots, M_l)$  with  $M_i \subseteq A^{r_i}$ , thus yielding an operation  $Op$  given by the formulas  $\varphi_{R_i}(\bar{x}^i)$ . Obviously,  $Op(\bigcup_{s \geq 0} R_1^s, \dots, \bigcup_{s \geq 0} R_l^s) = (\bigcup_{s \geq 0} R_1^s, \dots, \bigcup_{s \geq 0} R_l^s)$ . So  $(R_1^{A(\Pi)}, \dots, R_l^{A(\Pi)})$  is a fixed point of  $Op$  (in fact, it is the least fixed point). There are extensions of first-order logic, so-called fixed point logics,

that are designed to speak about fixed points of such definable operations. The most prominent logic of this kind is least fixed point logic. There are close connections between fixed point logics and (variants of) DATALOG (cf. [Ebbinghaus and Flum, 1995]). For DATALOG and its relatives we refer to [Abiteboul and Vianu, 1991], [Abiteboul, Hull and Vianu, 1995], and [Chandra and Harel, 1985].

#### 4 FIRST-ORDER RESOLUTION

Let us come back to the questions concerning the entailment relation that we have addressed in Section 2: Given a set  $\Phi$  of positive universal Horn sentences and a sentence  $\exists \bar{x}(\psi_0 \wedge \dots \wedge \psi_k)$  with atomic  $\psi_i$ , we ask whether

$$(1) \quad \Phi \models \exists \bar{x}(\psi_0 \wedge \dots \wedge \psi_k).$$

By Corollary 25 we know that the answer is positive just in case

$$(2) \quad \text{there are terms } \bar{t} \in T_0^\sigma \text{ such that}$$

$$\Phi \models (\psi_0(\bar{x}|\bar{t}) \wedge \dots \wedge \psi_k(\bar{x}|\bar{t})).$$

We are interested in finding such terms  $\bar{t}$  or even in generating all such terms  $\bar{t}$ . By Theorem 15, (2) is equivalent to

$$(3) \quad \text{there are } \bar{t} \in T_0^\sigma \text{ such that}$$

$$\text{GI}(\Phi) \cup \{(\neg\psi_0(\bar{x}|\bar{t}) \vee \dots \vee \neg\psi_k(\bar{x}|\bar{t}))\}$$

is not satisfiable.

Clearly, (3) can be answered by systematically checking for all  $\bar{t} \in T_0^\sigma$  whether  $\text{GI}(\Phi) \cup \{(\neg\psi_0(\bar{x}|\bar{t}) \vee \dots \vee \neg\psi_k(\bar{x}|\bar{t}))\}$  is not satisfiable, thereby applying propositional Horn resolution. However, this way of handling things does not take into consideration that, say, for  $\Phi = \{\varphi_1, \dots, \varphi_n\}$  and infinite  $T_0^\sigma$ , the infinitely many ground instances in  $\text{GI}(\Phi)$  stem from the finitely many  $\varphi_1, \dots, \varphi_n$ . Taking into account this aspect, we are led to a more goal-oriented procedure.

So far we defined  $\text{GI}(\Phi)$  only for sets  $\Phi$  of universal sentences. We extend this definition to formulas:

**DEFINITION 36.** *Let  $\varphi$  be a formula of the form  $\forall \bar{x}\psi$  with quantifier-free  $\psi$ . Then for arbitrary pairwise distinct variables  $y_1, \dots, y_l$  and terms  $t_1, \dots, t_l$ , the formula  $\psi(\bar{y}|\bar{t})$  is called an instance of  $\varphi$ . If  $\psi(\bar{y}|\bar{t})$  is a sentence, we also call it a ground instance. For a set  $\Phi$  of formulas  $\varphi$  of the form above,  $\text{GI}(\Phi)$  is the set of its ground instances.*

Recall the function  $\pi$  mapping in a one-to-one way atomic formulas onto propositional variables and quantifier-free formulas onto propositional formulas. It allows to freely use notations such as literal, clause, Horn clause, resolvent also in the framework of first-order logic. Moreover, we freely pass from formulas to clauses and vice versa.

#### 4.1 An Example.

The following example serves to explain the idea underlying the goal-oriented procedure we have in mind.

Assume  $\sigma = \{P, R, f, g, c\}$  with ternary  $P$ , binary  $R$ , and unary  $f, g$ . Let

$$\Phi := \{\forall x\forall y(Pxyc \rightarrow Ryg(f(x))), \forall x\forall yPf(x)yc\}.$$

We want to check whether

$$\Phi \models \exists x\exists yRf(x)g(y),$$

i.e., equivalently, whether for some  $s, t \in T_0^\sigma$

$$\text{GI}(\{\neg Pxyc, Ryg(f(x))\}, \{Pf(x)yc\}) \cup \{\neg Rf(s)g(t)\}$$

is not satisfiable. Set

$$\begin{aligned} C_1 &:= \{\neg Pxyc, Ryg(f(x))\}, \\ C_2 &:= \{Pf(x)yc\}, \\ N_1 &:= \{\neg Rf(x)g(y)\}. \end{aligned}$$

By the Theorem on the H-Resolution 10 our problem is equivalent to the existence of a ground instance  $N'_1$  of  $N_1$  and of a set  $\mathcal{C}$  of ground instances of  $C_1$  and  $C_2$  such that the empty clause is H-derivable from  $\mathcal{C} \cup \{N'_1\}$ . Now, when forming resolvents, the idea is to use instances of  $C_1, C_2$ , and  $N_1$  not by substituting appropriate *ground* terms for the variables, but by choosing terms from  $T^\sigma$  as general as possible. In our case, a closer look at  $C_1, C_2$ , and  $N_1$  shows that there is at most one possibility for a resolution (i.e., for obtaining a resolvent) with  $N_1$ , namely a resolution involving  $N_1$  and  $C_1$ . To avoid a collision of variables, we first rename  $x$  and  $y$  in  $C_1$  by new variables  $u$  and  $v$  (recall that  $x, y$  are quantified) getting

$$C'_1 := \{\neg Puv, Rvg(f(u))\}.$$

Comparing  $N_1$  and  $C'_1$  we see that a replacement of  $v$  by  $f(x)$  and of  $y$  by  $f(u)$  leads to the “simplest” instances of  $N_1$  and  $C'_1$  that can be resolved. In fact, this replacement leads to

$$N'_1 := \{\neg Rf(x)g(f(u))\}$$

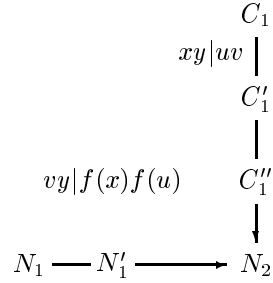
and

$$C_1'' := \{\neg Puf(x)c, Rf(x)g(f(u))\},$$

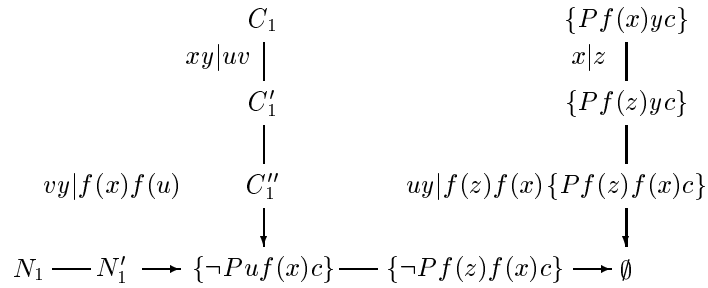
and we obtain the resolvent  $N_2$  of  $N_1'$  and  $C_1''$ ,

$$N_2 := \{\neg Puf(x)c\}.$$

This process can be pictured as



Now we can treat  $N_2$  and  $C_2$  similarly, arriving at the empty clause. The whole derivation is pictured by



When taking all renamings and substitutions together, the variable  $y$  of the negative clause  $N_1 = \{\neg Rf(x)g(y)\}$  has finally been replaced by  $f(f(z))$ , whereas the variable  $x$  of  $N_1$  has been kept unchanged. So it is intuitively clear that there is a set  $\mathcal{C}_0$  of instances of  $C_1$  and  $C_2$  such that

$$\mathcal{C}_0 \cup \{\neg Rf(x)g(y)\}(xy \mid xf(f(z)))$$

is not satisfiable, and therefore,

$$\Phi \models Rf(x)g(f(f(z))).$$

As we have chosen the substitutions in the derivation above as general as possible, it is plausible that we thus get all solutions, i.e.,

$$\{(s, t) \in T_0^\sigma \times T_0^\sigma \mid \Phi \models Rf(x)g(y)(xy|st)\} = \{(s, f(f(t))) \mid s, t \in T_0^\sigma\}.$$

The precise considerations that follow will show that this is true. We hope that the reader will have no difficulties to view the preceding example as a special case of the general theory. Our considerations take place in first-order logic without equality.

#### 4.2 Unification and U-Resolution.

We start with a systematic treatment of substitutions.

**DEFINITION 37.** A substitutor is a map  $\mu : \{v_1, v_2, \dots\} \rightarrow T^\sigma$  such that  $\mu(x) = x$  for almost all  $x$ .

For a substitutor  $\mu$ , let  $x_1, \dots, x_n$  be distinct variables such that  $\mu(x) = x$  for  $x \neq x_1, \dots, x \neq x_n$ . Setting  $t_i := \mu(x_i)$  for  $i = 1, \dots, n$ , we often denote  $\mu$  by  $(\bar{x}|\bar{t})$  and extend  $\mu$  to arbitrary terms and arbitrary quantifier-free formulas in the natural way by defining (with  $t\mu$  for  $\mu(t)$  and  $\varphi\mu$  for  $\mu(\varphi)$ )

$$t\mu := t(\bar{x}|\bar{t}), \quad \varphi\mu := \varphi(\bar{x}|\bar{t}).$$

Let  $\iota$  be the substitutor with  $\iota(x)(=x\iota) = x$  for all  $x$  and define the composition  $\mu\nu$  of substitutors  $\mu$  and  $\nu$  by

$$x(\mu\nu) := (x\mu)\nu$$

for all variables  $x$ . Then it is easy to check:

**LEMMA 38.** For all  $t \in T^\sigma$ , quantifier-free  $\varphi$ , and substitutors  $\mu, \nu, \rho$ :

- (a)  $t\iota = t$  and  $\varphi\iota = \varphi$ .
- (b)  $t(\mu\nu) = (t\mu)\nu$  and  $\varphi(\mu\nu) = (\varphi\mu)\nu$ .
- (c)  $(\mu\nu)\rho = \mu(\nu\rho)$ .

Part (c) justifies parenthesis-free notations such as  $t\mu\nu\rho$  or  $\varphi\mu\nu\rho$  that we will use later.

**DEFINITION 39.**

- (a) A renaming is a substitutor that is a bijection of the set  $\{v_1, v_2, v_3, \dots\}$  of variables.
- (b) Let  $C_1, C_2$  be clauses and  $\xi$  a renaming. We call  $\xi$  a separator of  $C_1$  and  $C_2$  if no variables occur both in  $C_1$  and in  $C_2\xi$  ( $:= \{\lambda\xi \mid \lambda \in C_2\}$ ).

In our example in Subsection 4.1 we can view the first step as applying the renaming  $\xi = (xyuv|uvxy)$  as a separator of  $N_1 (= \{\neg Rf(x)g(y)\})$  and  $C_1 (= \{\neg Pxyz, Ryg(f(x))\})$ . Note that  $C_1\xi = \{\neg Puvz, Rvg(f(u))\}$  is the clause which we denoted by  $C'_1$ . We then have chosen a “simplest” substitutor  $\mu$  such that we were able to form a resolvent of  $N_1\mu$  and  $C'_1\mu$ . The role of  $\mu$  can be described as to “unify in the simplest way” the literals  $\neg Rf(x)g(y)$  ( $\in N_1$ ) and  $Rvg(f(u))$  ( $\in C'_1$ ) in the sense that the clause  $\{Rf(x)g(y), Rvg(f(u))\}\mu$  consists of a single element.

**DEFINITION 40.** *A clause  $C$  is unifiable if there is a substitutor  $\mu$  such that  $C\mu$  consists of a single element. Such a substitutor  $\mu$  is called a unifier of  $C$ . A unifier of  $C$  is a general unifier of  $C$  if for any unifier  $\mu'$  of  $C$  there is a substitutor  $\nu$  such that  $\mu' = \mu\nu$ .*

Note that the empty clause is not unifiable. – We now establish an algorithm that, applied to a clause  $C$ , decides whether  $C$  is unifiable and, in the positive case, yields a general unifier of  $C$ .

**DEFINITION 41.** *The unification algorithm, applied to a clause  $C$ , is given by the following rules (u1) to (u9) which are applied step by step, starting with rule (u1).*

- (u1) *If  $C$  is empty or  $C$  contains atomic as well as negated atomic formulas or if the formulas in  $C$  do not all contain the same relation symbol, then stop with the answer “ $C$  is not unifiable”.*
- (u2) *Set  $i := 0$  and  $\mu_0 := \iota$ .*
- (u3) *If  $C\mu_i$  contains a single element, stop with the answer “ $C$  is unifiable and  $\mu_i$  is a general unifier”.*
- (u4) *If  $C\mu_i$  contains more than one element, let  $\lambda_1$  and  $\lambda_2$  be two distinct literals in  $C\mu_i$  (say, the first two distinct ones with respect to a fixed order, e.g. the lexicographic order). Determine the first place where the words  $\lambda_1$  and  $\lambda_2$  differ. Let  $\xi_1$  and  $\xi_2$  be the letters at this place in  $\lambda_1$  and  $\lambda_2$ , respectively.*
- (u5) *If the (different) letters  $\xi_1$  and  $\xi_2$  are function symbols or constants, stop with the answer “ $C$  is not unifiable”.*
- (u6) *One of the letters  $\xi_1, \xi_2$  is a variable  $x$ , say  $\xi_1$ . Determine the term  $t$  which starts with  $\xi_2$  in  $\lambda_2$ .<sup>8</sup>*
- (u7) *If  $x$  occurs in  $t$ , stop with the answer “ $C$  is not unifiable”.*
- (u8) *Set  $\mu_{i+1} := \mu_i(x|t)$  and  $i := i + 1$ .*
- (u9) *Go to (u3).*

---

<sup>8</sup> $t$  may be a variable; it is easy to show that  $t$  exists.

LEMMA 42. *Applied to any clause  $C$ , the unification algorithm stops and yields the right answer to the question whether  $C$  is unifiable, in the positive case also providing a general unifier.*

Before the proof we give some examples. We start with the clause discussed before Definition 40.

(1) Let  $C := \{Rf(x)g(y), Rvg(f(u))\}$ . The unification algorithm successively yields

$$\mu_0 = \iota, \quad \mu_1 = (v|f(x)), \quad \mu_2 = (v|f(x))(y|f(u)) (= (vy|f(x)f(u)))$$

together with the answer “ $C$  is satisfiable and  $\mu_2$  is a general unifier”.

(2) Let  $C := \{Ryf(y), Rzz\}$ . The unification algorithm yields  $\mu_0 = \iota$  and  $\mu_1 = (y|z)$  (or  $\mu_1 = (z|y)$ ) and then, going back to (u3) with  $C' := \{Rzf(z), Rzz\}$ , stops by (u7) with the answer “ $C$  is not unifiable”.

**Proof** [of Lemma 42]. Let  $C$  be a clause. We have to show that the unification algorithm stops when applied to  $C$  and gives the right answer to the question “Is  $C$  unifiable?”, and, in the positive case, yields a general unifier.

If the algorithm stops at (u1) then obviously  $C$  is not unifiable. Therefore we may assume that  $C$  is a nonempty clause whose literals are all atomic or all negated atomic formulas that, moreover, contain the same relation symbol.

The algorithm will stop for  $C$  after finitely many steps: Since applying (u8) causes the variable  $x$  to disappear ( $x$  does not occur in  $t!$ ), the only possible loop (u3)–(u9) can be passed through only as often as there are different variables in  $C$ .

If the algorithm stops at (u3),  $C$  is unifiable. Therefore, if  $C$  is not unifiable, it can stop only by (u5) or (u7). Thus the algorithm yields the right answer in case  $C$  is not unifiable.

Now let  $C$  be unifiable. We will show:

- (\*) If  $\nu$  is a unifier of  $C$  then for every value  $i$  reached by the algorithm there is  $\nu_i$  with  $\mu_i\nu_i = \nu$ .

Then we are done: If  $k$  is the last value of  $i$  then the clause  $C\mu_k$  is unifiable since  $C\mu_k\nu_k = C\nu$ ; so the algorithm cannot end with (u5) or (u7). (If it would end, e.g., with (u7), there would be two different literals in  $C\mu_k$  of the form  $\dots x \sim$  and  $\dots t_-$  where  $t \neq x$  and  $x$  occurs in  $t$ ; after any substitutions are carried out, there would always be terms of different length starting at the places corresponding to  $x$  and  $t$ , respectively, hence,  $C\mu_k$  would not be unifiable, and, by (\*), the same would hold for  $C$ .) Therefore the algorithm must end with (u3), i.e.,  $\mu_k$  is a unifier and by (\*) a general unifier of  $C$ .

We prove (\*) by induction on  $i$ . For  $i = 0$  we set  $\nu_0 := \nu$ . Then  $\mu_0\nu_0 = \nu\nu = \nu$ . In the induction step let  $\mu_i\nu_i = \nu$  and suppose the value  $i + 1$  has

been reached. By (u8) we have  $\mu_{i+1} = \mu_i(x|t)$  for some  $x, t$ , with  $x$  not occurring in  $t$ . Next, we observe ( $C\mu_i\nu_i$  has a single element!):

$$(1) \quad x\nu_i = t\nu_i.$$

We define  $\nu_{i+1}$  by

$$y\nu_{i+1} := \begin{cases} y\nu_i & \text{if } y \neq x, \\ x & \text{if } y = x. \end{cases}$$

Since  $x$  does not occur in  $t$ , we have

$$(2) \quad t\nu_{i+1} = t\nu_i.$$

Now  $(x|t)\nu_{i+1} = \nu_i$ : namely, if  $y \neq x$ , then  $y((x|t)\nu_{i+1}) = y\nu_{i+1} = y\nu_i$ , and  $x((x|t)\nu_{i+1}) = t\nu_{i+1} = t\nu_i = x\nu_i$ . Altogether:

$$\mu_{i+1}\nu_{i+1} = (\mu_i(x|t))\nu_{i+1} = \mu_i((x|t)\nu_{i+1}) = \mu_i\nu_i = \nu,$$

and we have finished the induction step. ■

The issue of the computational complexity of the unification algorithm is important for concrete implementations; it is addressed e.g. in [Baader and Siekmann, 1994; Börger, Grädel and Gurevich, 1997]).

For a clause  $C$ ,  $C^F$  stands for  $\{\lambda^F \mid \lambda \in C\}$ , where for a literal  $\lambda$  we set  $\lambda^F = \neg\lambda$  if  $\lambda$  is atomic, and  $\lambda^F = \gamma$  if  $\lambda = \neg\gamma$ . The following notion of U-resolution (U stands for “unification”) comprises the steps “renaming - substitution - forming a resolvent” as contained in the picture on page 348.

DEFINITION 43. *Let  $C, C_1, C_2$  be clauses.  $C$  is a U-resolvent of  $C_1$  and  $C_2$  if there are a separator  $\xi$  of  $C_1$  and  $C_2$  and clauses  $D_1, E_1 \subseteq C_1$  and  $D_2, E_2 \subseteq C_2$  such that*

$$(i) \quad E_1, E_2 \neq \emptyset.$$

$$(ii) \quad E_1^F \cup E_2\xi \text{ is unifiable.}$$

$$(iii) \quad C_1 = D_1 \cup E_1, \quad C_2 = D_2 \cup E_2, \quad \text{and } C = (D_1 \cup D_2\xi)\eta,$$

where  $\eta$  is the general unifier of  $E_1^F \cup E_2\xi$ , that is, the general unifier yielded by the unification algorithm.

Schematically, we represent this U-resolution by

$$\begin{array}{ccc} & C_2 & \\ & \xi \downarrow & \\ \eta & C_2\xi & \text{or even shorter by} & C_2 \\ & \downarrow & & \xi \downarrow \\ C_1 \longrightarrow & C & & C_1 \xrightarrow{\eta} C \end{array}$$



The reader may check that the resolution instances in the example above are really U-resolutions in the precise sense.

If  $C_1$  and  $C_2$  are ground clauses (i.e., clauses without variables) then, since a unifiable ground clause has only one element (with  $\iota$  as the general unifier), we have:

LEMMA 44. *For ground clauses  $C$ ,  $C_1$ , and  $C_2$ , clause  $C$  is a (propositional) resolvent of  $C_1$  and  $C_2$  iff  $C$  is a U-resolvent of  $C_1$  and  $C_2$ .*

The relationship between (propositional) resolution and U-resolution is even stronger; both forms are compatible in the following sense:

LEMMA 45. (Compatibility Lemma) *Let  $C_1$  and  $C_2$  be clauses. Then:*

- (a) *Every resolvent of a ground instance of  $C_1$  and of a ground instance of  $C_2$  is a ground instance of a U-resolvent of  $C_1$  and  $C_2$ .*
- (b) *Every ground instance of a U-resolvent of  $C_1$  and  $C_2$  is a resolvent of a ground instance of  $C_1$  and a ground instance of  $C_2$ .*

The following technical proof may be skipped in a first reading.

**Proof.** (a) Let  $C_i\mu_i$  be a ground instance of  $C_i$  ( $i = 1, 2$ ) and  $C$  a resolvent of  $C_1\mu_1$  and  $C_2\mu_2$ , i.e., for suitable  $M_1, M_2$ , and  $\lambda_0$

$$C_1\mu_1 = M_1 \cup \{\lambda_0\}, \quad C_2\mu_2 = M_2 \cup \{\lambda_0^F\}, \quad C = M_1 \cup M_2.$$

We set

$$\begin{aligned} M'_i &:= \{\lambda \in C_i \mid \lambda\mu_i \in M_i\} \quad (i = 1, 2), \\ L_1 &:= \{\lambda \in C_1 \mid \lambda\mu_1 = \lambda_0\}, \quad L_2 := \{\lambda \in C_2 \mid \lambda\mu_2 = \lambda_0^F\}. \end{aligned}$$

Then we have:

$$(1) \quad \begin{aligned} C_i &= M'_i \cup L_i \quad (i = 1, 2), \\ M'_i\mu_i &= M_i \quad (i = 1, 2), \\ L_1^F\mu_1 &= L_2\mu_2 = \{\lambda_0^F\}. \end{aligned}$$

Let  $\xi$  be a separator of  $C_1$  and  $C_2$  and  $\mu$  a substitutor with

$$x\mu := \begin{cases} x\xi^{-1}\mu_2 & \text{if } x \text{ appears in } C_2\xi \\ x\mu_1 & \text{otherwise.} \end{cases}$$

As no variable appears both in  $C_1$  and in  $C_2\xi$ , we obtain

$$(2) \quad \lambda\mu = \lambda\mu_1 \text{ for } \lambda \in C_1 \quad \text{and} \quad \lambda\xi\mu = \lambda\mu_2 \text{ for } \lambda \in C_2.$$

Therefore,

$$(L_1^F \cup L_2\xi)\mu = L_1^F\mu_1 \cup L_2\mu_2 = \{\lambda_0^F\},$$

hence  $\mu$  is a unifier of  $L_1^F \cup L_2\xi$ . Let  $\eta$  be the general unifier and  $\mu = \eta\nu$ . Then  $C^* := (M_1' \cup M_2'\xi)\eta$  is a U-resolvent of  $C_1$  and  $C_2$ . Finally,  $C$  is a ground instance of  $C^*$ ; namely  $C^*\nu = (M_1' \cup M_2'\xi)\mu \stackrel{(2)}{=} M_1'\mu_1 \cup M_2'\mu_2 \stackrel{(1)}{=} M_1 \cup M_2 = C$ .

So we proved (a). For later purposes we note the following strengthening: Since, for a given finite set  $Y$  of variables, we can choose the separator  $\xi$  of  $C_1$  and  $C_2$  such that no variable from  $Y$  appears in  $C_2\xi$ , we have shown:

- (†) If  $C_1$  and  $C_2$  are clauses and  $C_1\mu_1$  and  $C_2\mu_2$  are ground instances of  $C_1$  and  $C_2$ , respectively, and if

$$C \text{ is a resolvent of } C_1\mu_1 \text{ and } C_2\mu_2,$$

then for every finite set  $Y$  of variables there are  $C^*$ ,  $\xi$ ,  $\eta$ , and  $\nu$  such that

$$\begin{array}{ccc} & C_2 & \\ & \xi \downarrow & \\ C_1 & \xrightarrow{\eta} & C^* \end{array}$$

is a U-resolution and  $C = C^*\nu$  as well as  $y\eta\nu = y\mu_1$  for  $y \in Y$ .

(b) Let  $C$  be a U-resolvent of  $C_1$  and  $C_2$ , say  $C = (M_1 \cup M_2\xi)\eta$ ,  $C_i = M_i \cup L_i$  ( $i = 1, 2$ ), and  $(L_1^F \cup L_2\xi)\eta = \{\lambda_0\}$ , where  $\xi$  is a separator of  $C_1$  and  $C_2$ , and  $\eta$  the general unifier of  $L_1^F \cup L_2\xi$ .

Furthermore, let  $C\mu$  be a ground instance of  $C$ . We set

$$\mu_1 := \eta\mu \quad \text{and} \quad \mu_2 := \xi\eta\mu.$$

We can assume that  $C_1\mu_1$  and  $C_2\mu_2$  are ground clauses (otherwise we replace  $\mu$  by  $\mu\nu$  where  $\nu(x) \in T_0^\sigma$  if  $x$  appears in  $C_1\mu_1 \cup C_2\mu_2$ , and note that  $C\mu\nu = C\mu$ , since  $C\mu$  is a ground instance). Hence, it suffices to show

$$C\mu \text{ is a resolvent of } C_1\mu_1 \text{ and } C_2\mu_2.$$

For this, we only have to note that

$$C_1\mu_1 = M_1\mu_1 \cup L_1\mu_1 = M_1\mu_1 \cup \{\lambda_0^F\mu\},$$

$$C_2\mu_2 = M_2\mu_2 \cup L_2\mu_2 = M_2\mu_2 \cup \{\lambda_0\mu\},$$

and

$$M_1\mu_1 \cup M_2\mu_2 = (M_1 \cup M_2\xi)\eta\mu = C\mu.$$

■

We now adopt the propositional Horn resolution to our framework. For a set  $\mathcal{C}$  of (first-order) Horn clauses we let  $\mathcal{C}^+$  and  $\mathcal{C}^-$  be the set of positive and negative Horn clauses in  $\mathcal{C}$ , respectively.

**DEFINITION 46.** *Let  $\mathcal{C}$  be a set of (first-order) Horn clauses.*

(a) *A sequence  $N_0, N_1, \dots, N_m$  is a UH-resolution from  $\mathcal{C}$ , if there are  $P_0, \dots, P_{m-1} \in \mathcal{C}^+$  such that*

- (1)  $N_0, \dots, N_m$  are negative Horn clauses;
- (2)  $N_0 \in \mathcal{C}^-$ ;
- (3)  $N_{i+1}$  is a U-resolvent of  $N_i$  and  $P_i$  for  $i < m$ .

(b) *A negative Horn clause  $N$  is UH-derivable from  $\mathcal{C}$ , if there is a UH-resolution  $N_0, \dots, N_m$  from  $\mathcal{C}$  with  $N = N_m$ .*

If a “UH-resolution via  $P_0, \dots, P_{m-1}$ ” as in (a) uses the separators  $\xi_i$  and the substitutors  $\eta_i$  to form the corresponding U-resolvents  $N_{i+1}$ , we represent it as

$$\begin{array}{ccccccc}
 & P_0 & & P_1 & & & P_{m-2} & & P_{m-1} \\
 & \xi_0 \downarrow & & \xi_1 \downarrow & & & \xi_{m-2} \downarrow & & \xi_{m-1} \downarrow \\
 N_0 & \xrightarrow{\eta_0} & N_1 & \xrightarrow{\eta_1} & N_2 & \xrightarrow{\eta_2} & \dots & \xrightarrow{\eta_{m-2}} & N_{m-1} & \xrightarrow{\eta_{m-1}} & N_m
 \end{array}$$

Then we have the following connection between H-derivability and UH-derivability:

**LEMMA 47.** *For a set  $\mathcal{C}$  of Horn clauses and a negative ground clause  $N$  the following are equivalent:*

- (i)  $N$  is H-derivable from  $\text{GI}(\mathcal{C})$ .
- (ii)  $N$  is a ground instance of a clause that is UH-derivable from  $\mathcal{C}$ .

*In particular,  $\emptyset$  is H-derivable from  $\text{GI}(\mathcal{C})$  iff  $\emptyset$  is UH-derivable from  $\mathcal{C}$ .*

**Proof.** We prove the direction from (i) to (ii) by induction on the length  $m$  of an H-resolution of  $N$  from  $\text{GI}(\mathcal{C})$ .

If  $m = 0$ ,  $N$  belongs to  $\text{GI}(\mathcal{C}^-)$ , that means,  $N$  is a ground instance of a clause in  $\mathcal{C}^-$  and hence, a ground instance of a clause that is UH-derivable from  $\mathcal{C}$ . In the induction step, let  $N_0, \dots, N_m, N$  be an H-resolution from  $\text{GI}(\mathcal{C})$  which ends with

$$\begin{array}{ccc}
 & P_m & \\
 & \downarrow & \\
 N_m & \longrightarrow & N
 \end{array}$$

where  $P_m$  is a ground instance of some clause  $C \in \mathcal{C}^+$ . By induction hypothesis, there is a negative clause  $N'$  which is UH-derivable from  $\mathcal{C}$ , such that  $N_m$  is a ground instance of  $N'$ . In particular,  $N$  is an H-resolvent of a ground instance of  $N'$  and of a ground instance of  $C$ . Hence, by part (a) of the Compatibility Lemma 45,  $N$  is a ground instance of a U-resolvent, say  $N''$ , of  $N'$  and  $C$ . As  $N''$  is UH-derivable from  $\mathcal{C}$ , (ii) follows.

The other direction has a similar proof, using part (b) of the Compatibility Lemma. ■

As a corollary we have:

**THEOREM 48** (Theorem on the UH-Resolution). *Let  $\Phi$  be a set of universal Horn sentences and let  $\mathcal{C}(\Phi)$  denote the set of clauses which correspond to the kernels of the formulas in  $\Phi$ . Then the following are equivalent:*

- (i)  $\Phi$  is satisfiable.
- (ii)  $\emptyset$  is not UH-derivable from  $\mathcal{C}(\Phi)$ .

**Proof.** By Theorem 15 we have that  $\Phi$  is satisfiable iff  $\text{GI}(\mathcal{C}(\Phi))$  is satisfiable, that is, by the Theorem on the H-Resolution 10, iff  $\emptyset$  is not H-derivable from  $\text{GI}(\mathcal{C}(\Phi))$ . By the preceding lemma, the last statement is true iff  $\emptyset$  is not UH-derivable from  $\mathcal{C}(\Phi)$ . ■

Finally, we come to questions of the form “ $\Phi \models \exists x^n(\psi_0 \wedge \dots \wedge \psi_k)$ ?” where  $\Phi$  is a set of positive universal Horn sentences. The following theorem shows that in case “ $\Phi \models \exists x^n(\psi_0 \wedge \dots \wedge \psi_k)$ ” the method of UH-resolution provides all “solutions”  $t \in T_0^\sigma$ , thus being correct and complete for our purposes. Of course, any other adequate first-order calculus would do the same job, but the UH-resolution does it in a goal-oriented manner.

**THEOREM 49** (Theorem on Logic Programming). *Let  $\Phi$  be a set of positive universal Horn sentences and  $\exists x^n(\psi_0 \wedge \dots \wedge \psi_k)$  a sentence with atomic  $\psi_0, \dots, \psi_k$ . Set*

$$N := \{\neg\psi_0, \dots, \neg\psi_k\}.$$

*Then the following holds:*

- (a) **A d e q u a c y:**

$$\Phi \models \exists x^n(\psi_0 \wedge \dots \wedge \psi_k) \text{ iff } \emptyset \text{ is UH-derivable from } \mathcal{C}(\Phi) \cup \{N\}.$$

- (b) **C o r r e c t n e s s:** *If*

$$\begin{array}{ccccccc}
 & & P_1 & & P_{m-1} & & P_m \\
 & & \xi_1 \downarrow & & \xi_{m-1} \downarrow & & \xi_m \downarrow \\
 N = N_1 & \xrightarrow{\eta_1} & N_2 & \xrightarrow{\eta_2} & \cdots & \xrightarrow{\eta_{m-1}} & N_m & \xrightarrow{\eta_m} & \emptyset
 \end{array}$$

is a UH-resolution from  $\mathcal{C}(\Phi) \cup \{N\}$  then

$$\Phi \models (\psi_0 \wedge \dots \wedge \psi_k)\eta_1 \dots \eta_m.$$

(c) C o m p l e t e n e s s: If for  $t_1, \dots, t_m \in T_0^\sigma$

$$\Phi \models (\psi_0 \wedge \dots \wedge \psi_k)(\overset{n}{x}|\overset{n}{t})$$

then there is a UH-resolution of  $\emptyset$  from  $\mathcal{C}(\Phi) \cup \{N\}$  of the form given in (b) and a substitutor  $\nu$  with

$$t_i = x_i \eta_1 \dots \eta_m \nu \text{ for } i = 1, \dots, m.$$

If in part (b) exactly the variables  $z_1, \dots, z_s$  occur in the formula  $(\psi_0 \wedge \dots \wedge \psi_k)\eta_1 \dots \eta_m$  then  $\Phi \models \forall z_1 \dots \forall z_s (\psi_0 \wedge \dots \wedge \psi_k)\eta_1 \dots \eta_m$ ; therefore,  $\Phi \models (\psi_0 \wedge \dots \wedge \psi_k)\eta_1 \dots \eta_m \nu$  for every substitutor  $\nu$ . Thus, (b) and (c) show that the ground terms  $\overset{n}{t}$  with  $\Phi \models (\psi_0 \wedge \dots \wedge \psi_k)(\overset{n}{x}|\overset{n}{t})$  are exactly the instances of the “solutions”  $x_1 \eta_1 \dots \eta_m, \dots, x_n \eta_1 \dots \eta_m$  given by the UH-resolution.

**Proof.** Since  $\Phi \models \exists \overset{n}{x} (\psi_0 \wedge \dots \wedge \psi_k)$  iff  $\Phi \cup \{\forall \overset{n}{x} (\neg \psi_0 \vee \dots \vee \neg \psi_k)\}$  is not satisfiable, (a) follows immediately from the preceding theorem.

(b) The proof is by induction on  $m$ . For  $m = 1$  we have

$$\begin{array}{c}
 P_1 \\
 \xi_1 \downarrow \\
 N = N_1 \xrightarrow{\eta_1} \emptyset
 \end{array}$$

Therefore,  $N^F \eta_1 = P_1 \xi_1 \eta_1$ , so there must be a sentence  $\forall \overset{l}{y} \psi \in \Phi$  with quantifier-free  $\psi$  such that  $P_1 = \{\psi\}$  and  $\psi \xi_1 \eta_1 = \psi_i \eta_1$  for  $i = 0, \dots, k$ .

Since  $\Phi \models \forall \overset{l}{y} \psi$ , we have  $\Phi \models \psi \xi_1 \eta_1$  and hence,  $\Phi \models \psi_i \eta_1$  for  $i = 0, \dots, k$ , i.e.  $\Phi \models (\psi_0 \wedge \dots \wedge \psi_k)\eta_1$ .

For the induction step let  $m > 1$  and, say,  $N_2 = \{\neg \chi_0, \dots, \neg \chi_r\}$  ( $N_2$  is not empty!). The induction hypothesis, applied to the resolution starting with  $N_2$  and  $P_2$ , gives

$$(1) \quad \Phi \models (\chi_0 \wedge \dots \wedge \chi_r)\eta_2 \dots \eta_m.$$

Let  $i \leq k$ . We show

$$(*) \quad \Phi \models \psi_i \eta_1 \dots \eta_m,$$

thus getting our claim  $\Phi \models (\psi_0 \wedge \dots \wedge \psi_k) \eta_1 \dots \eta_m$ . We distinguish two cases: If  $\neg \psi_i \eta_1 \in N_2$ , we get (\*) immediately from (1).

Now suppose  $\neg \psi_i \eta_1 \notin N_2$ . Then we “lost”  $\neg \psi_i \eta_1$  in the resolution step leading to  $N_2$ . So in  $\Phi$  there is a sentence  $\forall y_1 \dots \forall y_l (\varphi_1 \wedge \dots \wedge \varphi_s \rightarrow \varphi)$  with  $P_1 = \{\neg \varphi_1, \dots, \neg \varphi_s, \varphi\}$  and

$$(2) \quad \varphi \xi_1 \eta_1 = \psi_i \eta_1,$$

$$(3) \quad \neg \varphi_j \xi_1 \eta_1 \in N_2 \text{ for } 1 \leq j \leq s,$$

therefore by (3) and (1):

$$(4) \quad \Phi \models \varphi_j \xi_1 \eta_1 \eta_2 \dots \eta_m \text{ for } 1 \leq j \leq s.$$

Since  $\Phi \models \forall y_1 \dots \forall y_l (\varphi_1 \wedge \dots \wedge \varphi_s \rightarrow \varphi)$  we get

$$\Phi \models (\neg \varphi_1 \vee \dots \vee \neg \varphi_s \vee \varphi) \xi_1 \eta_1 \eta_2 \dots \eta_m,$$

thus by (4)

$$\Phi \models \varphi \xi_1 \eta_1 \eta_2 \dots \eta_m,$$

and with (2) this leads to (\*).

(c): For  $t \in T^\sigma$  set  $\rho_1 := (\bar{x}|t)$  and let  $N_1 := N (= \{\neg \psi_0, \dots, \neg \psi_k\})$ ; suppose that  $\Phi \models (\psi_0 \wedge \dots \wedge \psi_k) \rho_1$  and that  $N' := N_1 \rho_1$  is a ground clause. Then, by Theorem 15,  $\mathcal{C}(\text{GI}(\Phi)) \cup \{N_1 \rho_1\}$  is not satisfiable. So, by the Theorem on the H-Resolution 10,  $\emptyset$  is H-derivable from  $\mathcal{C}(\text{GI}(\Phi)) \cup \{N_1 \rho_1\}$ , say, as pictured in

$$\begin{array}{ccccccc} & & P'_1 & P'_2 & & P'_{m-1} & P'_m \\ & & \downarrow & \downarrow & & \downarrow & \downarrow \\ N_1 \rho_1 = N'_1 & \longrightarrow & N'_2 & \longrightarrow & N'_3 & \longrightarrow & \dots & \longrightarrow & N'_m & \longrightarrow & \emptyset \end{array}$$

Here, the  $P'_j$  and the  $N'_j$  are ground clauses and, say,  $P'_j = P_j \mu_j$  with suitable clauses  $P_j \in \mathcal{C}(\Phi)$ .

We show: For every finite set  $X$  of variables there is a UH-resolution from  $\mathcal{C}(\Phi) \cup \{N\}$  as pictured below such that there exists a substitutor  $\nu$  with  $x \eta_1 \dots \eta_m \nu = x \rho_1$  for  $x \in X$ .

$$\begin{array}{ccccccc} & & P_1 & P_2 & & P_{m-1} & P_m \\ & & \xi_1 \downarrow & \xi_2 \downarrow & & \xi_{m-1} \downarrow & \xi_m \downarrow \\ N = N_1 & \xrightarrow{\eta_1} & N_2 & \xrightarrow{\eta_2} & N_3 & \xrightarrow{\eta_3} & \dots & \xrightarrow{\eta_{m-1}} & N_m & \xrightarrow{\eta_m} & \emptyset \end{array}$$

Then, for  $X := \{x_1, \dots, x_n\}$ , we get

$$x_i \eta_1 \dots \eta_m \nu = t_i \quad (1 \leq i \leq m),$$

and we are done.

We show the existence of a corresponding UH-resolution by induction on  $m$ .

For  $m = 1$  we have the resolution

$$\begin{array}{c} P'_1 \\ \downarrow \\ N_1 \rho_1 = N'_1 \longrightarrow \emptyset \end{array}$$

The claim follows immediately from (†) in the proof of the Compatibility Lemma 45 by setting

$$C_1 := N_1, \mu_1 := \rho_1, C := \emptyset \text{ and } Y := X.$$

In the induction step, let  $m \geq 2$ . For the first step of the H-resolution in the figure above we choose, again with (†) from Lemma 45,  $\xi_1, \eta_1, N_2$ , and  $\rho_2$  so that

$$\begin{array}{c} P_1 \\ \xi_1 \downarrow \\ N_1 \xrightarrow{\eta_1} N_2 \end{array}$$

and

$$(*) \quad x \eta_1 \rho_2 = x \rho_1 \text{ for } x \in X$$

as well as  $N'_2 = N_2 \rho_2$ . We apply the induction hypothesis to the part of the H-resolution above starting with  $N'_2$  and  $P'_2$  and to

$$Y := \text{the set of variables in } \{x \eta_1 \mid x \in X\}.$$

Then we get an UH-resolution as pictured by

$$\begin{array}{ccccccc} & P_2 & & P_{m-1} & & P_m & \\ & \xi_2 \downarrow & & \xi_{m-1} \downarrow & & \xi_m \downarrow & \\ N_2 & \xrightarrow{\eta_2} & N_3 & \xrightarrow{\eta_3} & \dots & \xrightarrow{\eta_{m-1}} & N_m & \xrightarrow{\eta_m} & \emptyset \end{array}$$

and a substitution  $\nu$  for which

$$y \eta_2 \dots \eta_m \nu = y \rho_2 \text{ for } y \in Y,$$

hence, by (\*) and the definition of  $Y$ ,

$$x \eta_1 \eta_2 \dots \eta_m \nu = x \eta_1 \rho_2 = x \rho_1 \text{ for } x \in X.$$

Thus, everything is proved. ■

### 4.3 Appendix

In the appendix to Section 1 we have generalized the Theorem on the H-Resolution to the Resolution Theorem of propositional logic. In the following, we give an analogous generalization of the Theorem on the UH-Resolution.

For an arbitrary set  $\mathcal{C}$  of (first-order) clauses we let  $\text{Res}_\infty(\mathcal{C})$  be the smallest set of clauses that contains  $\mathcal{C}$  and is closed under the formation of U-resolvents. Then we have:

**THEOREM 50 (U-Resolution Theorem).** *For any set  $\Phi$  of sentences of the form  $\forall \bar{x}\psi$ , where  $\psi$  is a disjunction of literals, the following are equivalent:*

- (i)  $\Phi$  is satisfiable.
- (ii)  $\emptyset \notin \text{Res}_\infty(\mathcal{C}(\Phi))$ .

**Proof.** As a generalization of Lemma 47 we show:

- (\*) For all ground clauses  $C$ :  $C \in \text{Res}_\infty(\text{GI}(\mathcal{C}(\Phi)))$  iff  $C$  is a ground instance of a clause in  $\text{Res}_\infty(\mathcal{C}(\Phi))$ .

Then we are done, because we have:

$$\begin{aligned} \Phi \text{ is satisfiable} & \text{ iff } \text{GI}(\mathcal{C}(\Phi)) \text{ is satisfiable} && \text{(by Theorem 15)} \\ & \text{ iff } \emptyset \notin \text{Res}_\infty(\text{GI}(\mathcal{C}(\Phi))) && \text{(by Theorem 11)} \\ & \text{ iff } \emptyset \notin \text{Res}_\infty(\mathcal{C}(\Phi)) && \text{(by (*))}. \end{aligned}$$

Concerning the direction from left to right in (\*), we set

$$\mathcal{C} := \{C \mid C \text{ is a ground instance of a clause in } \text{Res}_\infty(\mathcal{C}(\Phi))\}.$$

Then  $\text{GI}(\mathcal{C}(\Phi)) \subseteq \mathcal{C}$  and, by the Compatibility Lemma 45,  $\mathcal{C}$  is closed under resolvents. Hence,  $\text{Res}_\infty(\text{GI}(\mathcal{C}(\Phi))) \subseteq \mathcal{C}$ .

For the other direction we argue similarly with the set  $\mathcal{C}'$  of (first-order) clauses  $C'$  all ground instances of which belong to  $\text{Res}_\infty(\text{GI}(\mathcal{C}(\Phi)))$ , thus obtaining  $\text{Res}_\infty(\mathcal{C}(\Phi)) \subseteq \mathcal{C}'$ . ■

For extended representations of the resolution method besides those mentioned in the introduction, we refer the reader to [Eisinger and Ohlbach, 1993; Leitsch, 1996].

## 5 DECIDABILITY AND FEASIBILITY

In the preceding sections we often talked about quick or feasible procedures; however, we did so only in an intuitive way. The model commonly accepted



as a precise version of feasibility is that of polynomial complexity, i.e., the number of steps needed for the procedure is polynomial in the length of the input data. Here, the procedure is performed by some precisely defined computing device. As it turns out, the notion does not depend on the device as long as we refer to one of the universal machine models that are used to adequately define the basic notions of computability. The computing device we shall refer to will be register machines [Ebbinghaus, Flum and Thomas, 1992; Minsky, 1967]. We introduce them in Subsection 5.1. In 5.2, relying upon the undecidability of the halting problem for register machines, we show that the satisfiability problem for finite sets of universal Horn sentences is undecidable. Finally, in 5.3 we prove that the queries that are of polynomial complexity coincide with those that can be formalized in DATALOG, that is, DATALOG queries just coincide with the feasible ones. Subsections 5.2 and 5.3 can be read independently of each other.

### 5.1 Register Machines

Let  $\mathbb{A}$  be an alphabet, i.e., a non-empty finite set of symbols such as  $\{\}$  or  $\{0, 1\}$ .  $\mathbb{A}^*$  denotes the set of words over  $\mathbb{A}$ .

A *register machine* over  $\mathbb{A}$  is a computing device with a memory that consists of *registers* or *storing units*  $R_0, R_1, R_2, \dots$ . In each step of a computation each register contains a word over  $\mathbb{A}$ ; up to finitely many registers this is the empty word  $\square$ . The machine is able to perform so-called (*register*) *programs* that are built up by certain *instructions*. The instructions are preceded by a natural number  $L$ , their *label*. They are of the following form:

- (1) For  $L, i \in \mathbb{N}$  and  $a \in \mathbb{A}$ :

$$L \text{ LET } R_i = R_i + a$$

(“ $L$  Add the letter  $a$  at the end of the word in  $R_i$ ”)

- (2) For  $L, i \in \mathbb{N}$  and  $a \in \mathbb{A}$ :

$$L \text{ LET } R_i = R_i - a$$

(“ $L$  If the word in  $R_i$  ends with the letter  $a$ , delete this letter; else leave the word unchanged”)

- (3) For  $L, i, L', L'' \in \mathbb{N}$ :

$$L \text{ IF } R_i = \square \text{ THEN } L' \text{ ELSE } L''$$

(“ $L$  If  $R_i$  contains the empty word, continue with instruction  $L'$ , else with instruction  $L''$ ”)

(4) For  $L \in \mathbb{N}$  :

$L$  HALT

(“ $L$  Halt”).

A program  $\mathbb{P}$  is a finite sequence  $(\iota_0, \dots, \iota_k)$  of instructions with the following properties:

- (i)  $\iota_j$  has label  $j$ .
- (ii) Every instruction in  $\mathbb{P}$  of type (3) refers to labels  $\leq k$  (i.e.,  $L', L'' \leq k$ ).
- (iii) Only  $\iota_k$  is of type (4).

Note that each program addresses only finitely many  $R_i$ . A register machine that is programmed with a program  $\mathbb{P} = (\iota_0, \dots, \iota_k)$  and contains certain words in its registers, starts with instruction  $\iota_0$  and, stepwise, always performs the next instruction, only jumping if this is required by an instruction of type (3) and stopping if instruction  $\iota_k$  is performed. Of course, it may happen that the machine runs for ever.

Let  $\mathbb{P}$  be a program over  $\mathbb{A}$ ,  $n \in \mathbb{N}$ , and  $w_0, \dots, w_n \in \mathbb{A}^*$ . We write

$$\mathbb{P} : (w_0, \dots, w_n) \rightarrow \text{halt}$$

if  $\mathbb{P}$ , started with  $w_i$  in  $R_i$  for  $i \leq n$  and  $\square$  in the remaining registers, finally stops, and we write

$$\mathbb{P} : (w_0, \dots, w_n) \rightarrow \text{yes}$$

if  $\mathbb{P}$ , started with  $w_i$  in  $R_i$  for  $i \leq n$  and  $\square$  in the remaining registers, finally stops,  $R_0$  then containing the empty word.

A subset  $A$  of  $(\mathbb{A}^*)^{n+1}$  is *decidable* (in the precise sense) if there is a program  $\mathbb{P}$  over  $\mathbb{A}$  such that

- for all  $w_0, \dots, w_n \in \mathbb{A}^*$ ,  $P : (w_0, \dots, w_n) \rightarrow \text{halt}$
- $A = \{(w_0, \dots, w_n) \mid \mathbb{P} : (w_0, \dots, w_n) \rightarrow \text{yes}\}$ .

According to the *Church–Turing Thesis* decidability (in the precise sense) coincides with decidability in the intuitive sense, i.e., decidability by register machines exactly captures the intuitive counterpart.

One can code programs over  $\mathbb{A}$  as words over  $\mathbb{A}$  (cf. e.g. [Ebbinghaus, Flum and Thomas, 1992]). Let  $\mathbb{P} \mapsto \text{code}(\mathbb{P})$  be such a coding. Then we have the following well-known theorem:

**THEOREM 51** (Undecidability of the Halting Problem). *For any  $n \geq 2$  and any alphabet  $\mathbb{A}$ , the set*

$$\text{HALT}(\mathbb{A}) := \{\text{code}(\mathbb{P}) \mid \mathbb{P} \text{ a program over } \mathbb{A} \text{ which only addresses } n \text{ registers and with } \mathbb{P} : \square \rightarrow \text{halt}\}$$

is not decidable.

A subset  $A$  of  $(\mathbb{A}^*)^{n+1}$  is *polynomially decidable*, if  $A$  is decidable via a program  $\mathbb{P}$  over  $\mathbb{A}$  that, for any input  $(w_0, \dots, w_n)$  over  $\mathbb{A}$ , *stops after polynomially many steps*, i.e., there is a polynomial  $p$  with integer coefficients such that  $\mathbb{P}$ , started with  $(w_0, \dots, w_n)$ , stops after at most  $p(l)$  steps where  $l$  is the length of the word  $w_0 \dots w_n$ .

We let PTIME be the set of all polynomially decidable sets, regardless of the underlying alphabet. By naturally coding symbols of one alphabet by words over another alphabet, say  $\{0, 1\}$ , one can, without loss of generality, restrict oneself to the alphabet  $\{0, 1\}$ .

### 5.2 Undecidability of the Horn Part of First-Order Logic

As mentioned earlier, it is undecidable whether, for a finite set  $\Phi$  of universal Horn sentences and a sentence  $\exists \vec{x}(\psi_0 \wedge \dots \wedge \psi_k)$  with atomic  $\psi_i$ , we have  $\Phi \models \exists \vec{x}(\psi_0 \wedge \dots \wedge \psi_k)$ . In the following we give a proof of an even stronger result by a reduction to the undecidability of the halting problem for register machines.

Below we introduce a finite vocabulary  $\sigma_0$  and show (recall that  $L_0^{\sigma_0}$  denotes the set of first-order sentences of vocabulary  $\sigma_0$  without equality):

**THEOREM 52.** *The set*

$$\{(\Phi, \varphi) \mid \Phi \subseteq L_0^{\sigma_0} \text{ a finite set of positive universal Horn sentences,} \\ \varphi \in L_0^{\sigma_0} \text{ of the form } \exists \vec{x}\psi \text{ with atomic } \psi, \text{ and } \Phi \models \varphi\}$$

is undecidable.

**COROLLARY 53.** *It is undecidable whether a finite set of universal Horn sentences is satisfiable.*

**COROLLARY 54.** *It is undecidable whether a first-order sentence is satisfiable.*

Corollary 54, the undecidability of first-order logic, goes back to Church 1936; it contains the negative solution of the so-called *Entscheidungsproblem*; Theorem 52 is essentially due to Aandera 1971 and Börger 1971 (see [Börger, Grädel and Gurevich, 1997]). Clearly, Theorem 52 and its corollaries remain true for vocabularies containing  $\sigma_0$ . Corollary 54 even holds for vocabularies containing one at least binary relation symbol. However, Theorem 52 gets wrong if the vocabulary does not contain function symbols (cf. Corollary 20).

**Proof** [of Theorem 52]. We establish an effective procedure that assigns a pair  $(\Phi_{\mathbb{P}}, \varphi_{\mathbb{P}})$  to each register program  $\mathbb{P}$  over the alphabet  $\{\}$  which addresses only  $R_0, R_1$  such that

$$(1) \quad \Phi_{\mathbb{P}} \models \varphi_{\mathbb{P}} \quad \text{iff} \quad \mathbb{P}, \text{ started with empty registers, halts.}$$

Then we are done: If there would be an effective procedure to decide questions “ $\Phi \models \varphi$ ?” then, using (1), we could effectively decide whether a program  $\mathbb{P}$  of the form in question stops when started with empty registers, a contradiction to the undecidability of the halting problem (cf. Theorem 51).

Let  $\mathbb{P}$  be a program with instructions  $\iota_0, \dots, \iota_k$  which addresses only  $R_0$  and  $R_1$ . A triple  $(L, m_0, m_1)$  with  $L \leq k$  is called a *configuration* of  $\mathbb{P}$ . We say that  $(L, m_0, m_1)$  is the *configuration of  $\mathbb{P}$  after  $s$  steps* if  $\mathbb{P}$ , started with empty registers, runs for at least  $s$  steps and after  $s$  steps instruction  $L$  is to be executed next, while the numbers (i.e., the lengths of the words) in  $R_0, R_1$  are  $m_0, m_1$ , respectively. In particular,  $(0, 0, 0)$  is the configuration after 0 steps, the *initial configuration*. Since only  $\iota_k$  is a halt instruction, we have

- (2)  $\mathbb{P}$ , started with empty registers, halts iff for suitable  $s, m_0, m_1$ ,  $(k, m_0, m_1)$  is the configuration after  $s$  steps.

We set

$$\sigma_0 := \{C, f, \min\},$$

where  $C$  is 4-ary,  $f$  unary, and  $\min$  a constant. With  $\mathbb{P}$  we associate the following  $\sigma_0$ -structure  $\mathcal{A}_{\mathbb{P}}$  which is designed to describe the run of  $\mathbb{P}$ , started with empty registers:

$$\begin{aligned} A_{\mathbb{P}} &:= \mathbb{N} \\ C^{\mathcal{A}_{\mathbb{P}}} &:= \{(s, L, m_0, m_1) \mid (L, m_0, m_1) \text{ is the configuration of } \mathbb{P} \\ &\quad \text{after } s \text{ steps}\} \\ f^{\mathcal{A}_{\mathbb{P}}} &:= \text{the successor function on } \mathbb{N} \\ \min^{\mathcal{A}_{\mathbb{P}}} &:= 0. \end{aligned}$$

We abbreviate  $\min$  by  $\bar{0}$ ,  $f(\min)$  by  $\bar{1}$ ,  $f(f(\min))$  by  $\bar{2}$ , etc. Then we set

$$\varphi_{\mathbb{P}} := \exists x \exists y_0 \exists y_1 C x \bar{k} y_0 y_1,$$

and we define the set  $\Phi_{\mathbb{P}}$  of positive universal Horn sentences such that it has the following properties:

- (3)  $\mathcal{A}_{\mathbb{P}} \models \Phi_{\mathbb{P}}$ .  
 (4) If  $\mathcal{A}$  is a  $\sigma_0$ -structure which satisfies  $\Phi_{\mathbb{P}}$  and  $(L, m_0, m_1)$  is the configuration of  $\mathbb{P}$  after  $s$  steps, then  $\mathcal{A} \models C \bar{s} \bar{L} \bar{m}_0 \bar{m}_1$ .

Both (3) and (4) will be obvious from the definition.

$\Phi_{\mathbb{P}}$  consists of the following positive universal Horn sentences:

- (i)  $C \bar{0} \bar{0} \bar{0} \bar{0}$ ;

(ii) for each instruction  $L$  LET  $R_0 = R_0 + |$ :

$$\forall xy_0y_1(Cx\bar{L}y_0y_1 \rightarrow Cf(x)\overline{L+1}f(y_0)y_1),$$

and similarly for instructions  $L$  LET  $R_1 = R_1 + |$ ;

(iii) for each instruction  $L$  LET  $R_0 = R_0 - |$ :

$$\forall xy_1(Cx\bar{L}\bar{0}y_1 \rightarrow Cf(x)\overline{L+1}\bar{0}y_1),$$

$$\forall xy_0y_1(Cx\bar{L}f(y_0)y_1 \rightarrow Cf(x)\overline{L+1}y_0y_1),$$

and similarly for instructions  $L$  LET  $R_1 = R_1 - |$ ;

(iv) for each instruction  $L$  IF  $R_0 = \square$  THEN  $L'$  ELSE  $L''$ :

$$\forall xy_1(Cx\bar{L}\bar{0}y_1 \rightarrow Cf(x)\overline{L'}\bar{0}y_1),$$

$$\forall xy_0y_1(Cx\bar{L}f(y_0)y_1 \rightarrow Cf(x)\overline{L''}f(y_0)y_1),$$

and similarly for instructions  $L$  IF  $R_1 = \square$  THEN  $L'$  ELSE  $L''$ .

$\Phi_{\mathbb{P}}$  and  $\varphi_{\mathbb{P}}$  satisfy (1). To prove this, assume first that  $\Phi_{\mathbb{P}} \models \varphi_{\mathbb{P}}$ . Then, as  $\mathcal{A}_{\mathbb{P}} \models \Phi_{\mathbb{P}}$  (by (3)),  $\mathcal{A}_{\mathbb{P}} \models \varphi_{\mathbb{P}}$  and hence (cf. (2)),  $\mathbb{P}$  stops when started with empty registers. Conversely, if  $\mathbb{P}$  stops when started with empty registers, there are  $s, m_0, m_1$  such that  $(k, m_0, m_1)$  is the configuration after  $s$  steps. Then, if  $\mathcal{A}$  is a model of  $\Phi_{\mathbb{P}}$ , (4) yields that  $\mathcal{A}$  is a model of  $C\bar{s}k\overline{m_0}\overline{m_1}$  and, hence, of  $\varphi_{\mathbb{P}}$ . ■

### 5.3 PTIME and DATALOG

To prove our final result stating that DATALOG captures PTIME we must deal with structures as inputs for register machines. In logic, structures are abstract objects, and there is no canonical way of coding structures by words. Any reasonable coding of structures will rely on a naming of the elements and thus implicitly on an ordering of the structures. In general, different orderings will lead to different codes, and different codes, when serving as inputs for calculations, may lead to different results. To check independence, one would have to take into consideration the codes for all possible orderings. As their number is exponential in the size of the structures, we would be lost when considering questions of polynomial complexity. To overcome these difficulties, we restrict ourselves to *ordered* structures and then use their ordering to define the codes in a canonical way.

We let  $\sigma$  contain a binary relation symbol  $S$  and constants min and max. A finite  $\sigma$ -structure  $\mathcal{A}$  is *ordered*, if there is an ordering  $<$  on  $A$  such that  $S^{\mathcal{A}}$  is its *successor relation*, i.e.,  $S^{\mathcal{A}} = \{(a\ b) \mid a < b \text{ and for all } b', \text{ if } a < b' \text{ then } b \leq b'\}$ , and  $\text{min}^{\mathcal{A}}$  and  $\text{max}^{\mathcal{A}}$  are the minimal element and the maximal

element, respectively. Clearly,  $<$  is uniquely determined by  $S^A$  and is called the ordering *induced* by  $S^A$ .

To define the code of an ordered structure, we consider the special case

$$\sigma = \{S, \min, \max, \} \cup \{E, c\}$$

with binary  $E$ . (It will be clear how to handle the general case.) Let  $\mathcal{A}$  be an ordered structure. The code of  $\mathcal{A}$ ,  $\text{code}(\mathcal{A})$ , is the triple  $(w_{\text{dom}}, w_E, w_c)$  of words over  $\{0, 1\}$ , where

$$\begin{aligned} w_{\text{dom}} &:= \underbrace{1 \dots 1}_{|A| \text{ times}} \\ w_E &:= w_0 \dots w_{|A|^2-1}, \text{ where} \\ w_i &:= \begin{cases} 0, & \text{if the } i\text{-th pair in the lexicographic ordering} \\ & \text{induced by } S^A \text{ on } A \times A \text{ belongs to } E^A \\ 1, & \text{else} \end{cases} \\ w_c &:= \underbrace{1 \dots 1}_{i \text{ times}}, \text{ if } c^A \text{ is the } i\text{-th element in the induced ordering of } A.^9 \end{aligned}$$

Note that for ordered structures  $\mathcal{A}$  and  $\mathcal{B}$ ,

$$\mathcal{A} \cong \mathcal{B} \quad \text{iff} \quad \text{code}(\mathcal{A}) = \text{code}(\mathcal{B}).$$

Let  $\mathcal{K}$  be a class of finite ordered  $\sigma$ -structures closed under isomorphisms. We say that  $\mathcal{K} \in \text{PTIME}$ , if  $\{\text{code}(\mathcal{A}) \mid \mathcal{A} \in \mathcal{K}\} \in \text{PTIME}$ . The main result now is:

**THEOREM 55.** *Let  $\mathcal{K}$  be a class of finite ordered structures closed under isomorphisms. Then  $\mathcal{K}$  belongs to PTIME iff it is DATALOG-definable, that is, there exists a DATALOG sentence  $(\Pi, R)$  such that  $\mathcal{K} = \{\mathcal{A} \mid \mathcal{A} \models (\Pi, R)\}$ .*

**Proof.** At the end of the preceding section we have shown that the class of finite models of a DATALOG sentence is decidable (in the intuitive sense) in polynomial time. It is only a matter of patience to represent the algorithm given there as a register program.

Concerning the other direction, assume that  $\mathcal{K} \in \text{PTIME}$  via a program  $\mathbb{P}$  over  $\mathbb{A} = \{0, 1\}$  that, given (the code of) an ordered structure as input, has running time polynomial in the length of  $\text{code}(\mathcal{A})$  and, hence, polynomial in  $|A|$ . Say, the number of steps is  $\leq |A|^n$ . (This can be assumed without loss of generality; there are difficulties at most in case  $|A| = 1$ ; however, structures of cardinality one may be treated separately.) Moreover, we assume that  $n$  is greater than the arities of the relation symbols in  $\sigma$ .

<sup>9</sup>If  $c^A$  is the 0-th element,  $w_c$  is the empty word.

Again, as a typical example, let  $\sigma = \{S, \min, \max, \} \cup \{E, c\}$  with binary  $E$ . We define a DATALOG program  $\Pi$  which is designed to reflect the performance of the program  $\mathbb{P}$  and its outcome. We give  $\Pi$  in several steps.

*Part 1.* Let  $S^n$  be a  $(2n)$ -ary new (intentional) relation symbol. Let  $\Pi_1$  consist of the following rules that generate the  $n$ -ary lexicographic successor relation induced by  $S$ .

$$Suv \rightarrow S^n \overset{r}{x} u \overset{s}{\max} \overset{r}{x} v \overset{s}{\min} \quad \text{for } r + s = n - 1.$$

*Part 2.* For each register  $R_i$  addressed in  $\mathbb{P}$ , including the registers  $R_0, R_1,$  and  $R_2$  (that, in the beginning, store the components of the code of the input structure), we choose  $(2n)$ -ary (intentional) relation symbols  $Z_i, O_i, V_i$  where

- $Z_i \overset{n}{x} \overset{n}{y}$  means that in the  $\overset{n}{x}$ -th step (counted in the ordering given by  $S^n$ ) the word stored in  $R_i$  has an  $\overset{n}{y}$ -th letter, and this letter is 0;
- $O_i$  has a similar meaning, the letter now being 1;
- $V_i \overset{n}{x} \overset{n}{y}$  means that in the  $\overset{n}{x}$ -th step the word stored in  $R_i$  has length  $\overset{n}{y}$ .

The second part  $\Pi_2$  of  $\Pi$  serves to generate the relations  $Z_i, O_i, V_i$  before starting, that is at “time point”  $\overset{n}{x} = \min$ . It consists of the following rules:

$$\begin{aligned} & \rightarrow O_0 \overset{n}{\min} \overset{n-1}{\min} y \\ S \overset{n}{\min} x & \rightarrow V_0 \overset{n}{\min} \overset{n-2}{\min} x \overset{2}{\min} \\ Exy & \rightarrow Z_1 \overset{n}{\min} \overset{n-2}{\min} xy \\ \neg Exy & \rightarrow O_1 \overset{n}{\min} \overset{n-2}{\min} xy \\ S \overset{n}{\min} x & \rightarrow V_1 \overset{n}{\min} \overset{n-3}{\min} x \overset{2}{\min} \\ c \neq \overset{n}{\min} & \rightarrow O_2 \overset{n}{\min} \overset{n-1}{\min} \min \\ O_2 \overset{n}{\min} \overset{n-1}{\min} x, Sxy, y \neq c & \rightarrow O_2 \overset{n}{\min} \overset{n-1}{\min} y \\ & \rightarrow V_2 \overset{n}{\min} \overset{n-1}{\min} c \\ & \rightarrow V_i \overset{n}{\min} \overset{n}{\min} \quad \text{for } i \neq 0, 1, 2 \text{ and} \\ & \quad R_i \text{ addressed in } \mathbb{P}. \end{aligned}$$

*Part 3.* We now turn to a DATALOG description of how  $\mathbb{P}$  works. The last intentional symbols we need are  $n$ -ary symbols  $\text{Lab}_0, \dots, \text{Lab}_k$  for the labels  $0, \dots, k$  of  $\mathbb{P}$  and a zero-ary symbol  $P$ .  $P$  will code “success” and  $\text{Lab}_j \overset{n}{x}$  means that  $\mathbb{P}$  performs (at least)  $\overset{n}{x}$  steps and that after the  $\overset{n}{x}$ -th step the

instruction with label  $j$  has to be performed. The last part  $\Pi_3$  consists of the following rules:

- (i)  $\text{Lab}_0 \overset{n}{\text{min}}$
- (ii) Clauses describing a step according to the instructions of  $\mathbb{P}$  different from the halting instruction.

For any instruction with label  $L$  addressing  $R_i$ , and for any  $j \neq i$  such that  $R_j$  is addressed in  $\Pi$ , we take the rules

$$\begin{aligned} \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, O_j \overset{n}{x} \overset{n}{u} &\rightarrow O_j \overset{n}{y} \overset{n}{u} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, Z_j \overset{n}{x} \overset{n}{u} &\rightarrow Z_j \overset{n}{y} \overset{n}{u} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, V_j \overset{n}{x} \overset{n}{u} &\rightarrow V_j \overset{n}{y} \overset{n}{u}; \end{aligned}$$

they describe that nothing happens with  $R_j$ .

For  $L$  LET  $R_i = R_i + 0$  in  $\mathbb{P}$  we add the rules

$$\begin{aligned} \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y} &\rightarrow \text{Lab}_{L+1} \overset{n}{y} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, V_i \overset{n}{x} \overset{n}{u} &\rightarrow Z_i \overset{n}{y} \overset{n}{u} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, V_i \overset{n}{x} \overset{n}{u}, S^n \overset{n}{u} \overset{n}{v} &\rightarrow V_i \overset{n}{y} \overset{n}{v} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, O_i \overset{n}{x} \overset{n}{u} &\rightarrow O_i \overset{n}{y} \overset{n}{u} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, Z_i \overset{n}{x} \overset{n}{u} &\rightarrow Z_i \overset{n}{y} \overset{n}{u} \end{aligned}$$

and similarly for instructions  $L$  LET  $R_i = R_i + 1$ ,  $L$  LET  $R_i = R_i - 0$ ,  $L$  LET  $R_i = R_i - 1$ .

And for  $L$  IF  $R_i = \square$  THEN  $L'$  ELSE  $L''$  in  $\mathbb{P}$  we add the rules

$$\begin{aligned} \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, V_i \overset{n}{x} \overset{n}{\text{min}} &\rightarrow \text{Lab}_{L'} \overset{n}{y} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, O_i \overset{n}{x} \overset{n}{\text{min}} &\rightarrow \text{Lab}_{L''} \overset{n}{y} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, Z_i \overset{n}{x} \overset{n}{\text{min}} &\rightarrow \text{Lab}_{L''} \overset{n}{y} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, O_i \overset{n}{x} \overset{n}{u} &\rightarrow O_i \overset{n}{y} \overset{n}{u} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, Z_i \overset{n}{x} \overset{n}{u} &\rightarrow Z_i \overset{n}{y} \overset{n}{u} \\ \text{Lab}_L \overset{n}{x}, S^n \overset{n}{x} \overset{n}{y}, V_i \overset{n}{x} \overset{n}{u} &\rightarrow V_i \overset{n}{y} \overset{n}{u}. \end{aligned}$$

- (iii) Finally, for  $k$  HALT in  $\mathbb{P}$  we add the “success rule”

$$\text{Lab}_k \overset{n}{x}, V_0 \overset{n}{x} \overset{n}{\text{min}} \rightarrow P.$$



Setting  $\Pi := \Pi_1 \cup \Pi_2 \cup \Pi_3$ , it is easy to see that  $(\Pi, P)$  axiomatizes  $\mathcal{K}$  (with possible mistakes only for structures of cardinality one; to eliminate these mistakes, one can restrict  $\Pi$  to structures with at least two elements by adding  $S \min z$  to all bodies and treating structures of cardinality one separately by rules that have  $\min = \max$  in their body). ■

REMARK 56. *In Theorem 34 we not only considered queries defined by DATALOG sentences, but also queries defined by DATALOG formulas; in fact, we showed that for any such formula  $(\Pi, R)^n_x$  the corresponding query can be evaluated in polynomial time; hence,*

$$\{(A, \overset{n}{a}) \mid A \models (\Pi, R)^n_x[\overset{n}{a}]\} \in PTIME.$$

*This is the way we can reduce any query to a class of structures: We identify a query asking whether elements  $x_1, \dots, x_n$  in a  $\sigma$ -structure have property  $P$  (say, asking whether elements  $x_1, x_2$  in a graph are connected by a path) with the class of  $(\sigma \cup \{c_1, \dots, c_n\})$ -structures*

$$\{(A, \overset{n}{a}) \mid \overset{n}{a} \text{ have property } P \text{ in } A\}.$$

*The coincidence of DATALOG and PTIME goes back to [Immerman, 1987] and [Vardi, 1982], cf. [Papadimitriou, 1985], too; it is a typical result of descriptive complexity theory, a theory that relates logical descriptions to descriptions by machines, cf. [Ebbinghaus and Flum, 1995].*

*Institut für math. Logik, Universität Freiburg, Germany.*

#### BIBLIOGRAPHY

- [Abiteboul and Vianu, 1991] S. Abiteboul and V. Vianu. Datalog extensions for database queries and updates. *Journal Comp. System Sciences*, **43**, pp. 62–124, 1991.
- [Abiteboul, Hull and Vianu, 1995] S. Abiteboul, R. Hull and V. Vianu. *Foundations of Databases*. Addison-Wesley Publ. Company, 1995.
- [Apt, 1990] K. R. Apt. Logic Programming. In *Handbook of Theoretical Computer Science, Vol. B: Formal Models and Semantics*, J. van Leeuwen, ed. pp. 493–574. Elsevier, 1990.
- [Baader and Siekmann, 1994] F. Baader and J. Siekmann. Unification theory. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 2, Deduction Methodologies, D. Gabbay, C. Hogger and J. Robinson, eds. pp. 40–125, Oxford University Press, 1994.
- [Blake, 1937] A. Blake. *Canonical expressions in Boolean algebra*. Ph. D. dissertation. Dept. of Mathematics, Univ. of Chicago, 1937.
- [Börger, Grädel and Gurevich, 1997] E. Börger, E. Grädel and Y. Gurevich. *The Classical Decision Problem*. Perspectives in Mathematical Logic, Springer 1997.
- [Chandra and Harel, 1985] A. K. Chandra and D. Harel. Horn clause queries and generalizations. *Journal of Logic Programming*, **2**, 1–15, 1985.
- [Clark, 1978] K. L. Clark. Negation as failure. In *Logic and Data Bases*, H. Gallaire, J. Minker, eds. pp. 293–322. Plenum Press New York, 1978.
- [Colmerauer, 1970] A. Colmerauer. *Les systèmes-q ou un formalisme pour analyser et synthétiser des phrases sur ordinateur*. Intern Report 43, Département d'Informatique. Université de Montréal, 1970.

- [Colmerauer *et al.*, 1973] A. Colmerauer, H. Kanoui, P. Roussel and R. Pasero. *Un système de communication homme-machine en Français*. Technical Report. Groupe de Recherche en Intelligence Artificielle, Univ. d' Aix-Marseille, 1973.
- [Ebbinghaus and Flum, 1995] H.-D. Ebbinghaus and J. Flum. *Finite Model Theory*. Perspectives in Math. Logic, Springer 1995.
- [Ebbinghaus, Flum and Thomas, 1992] H.-D. Ebbinghaus, J. Flum and W. Thomas. *Mathematical Logic*. Springer 1992.
- [Eisinger and Ohlbach, 1993] N. Eisinger and H.-J. Ohlbach. Deduction systems based on resolution. In *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 1, Logical Foundations, D. Gabbay, C. Hogger and J. Robinson, eds. pp. 184–271. Oxford University Press, 1993.
- [Fitting, 1987] M. Fitting. *Computability Theory, Semantics, and Logic Programming*. Oxford Logic Guides 13. Oxford University Press, 1987.
- [Gabbay, Hogger and Robinson, 1993f] D. M. Gabbay, C. J. Hogger and J. A. Robinson. *Handbook of Logic in Artificial Intelligence and Logic Programming*, Vol. 1–3, Oxford University Press, 1993f.
- [Henkin, 1949] L. Henkin. The completeness of the first-order functional calculus. *Journal of Symbolic Logic*, **14**, 159–166, 1949.
- [Herbrand, 1968] J. Herbrand. *Ecrits logiques*, J. van Heijenoort, ed. Presses Univ. France Period., Paris, 1968. English translation: *Logical Writings by J. Herbrand*, W. D. Goldfarb, ed. Reidel, 1971 and Harvard Univ. Press, 1971.
- [Hodges, 1983] W. Hodges. First-Order Logic. In *Handbook of Philosophical Logic*, Vol. I, Elements of Classical Logic. D. Gabbay and F. Guenther, eds. pp. 1–131. Reidel, 1983.
- [Hopcroft and Ullman, 1979] J. E. Hopcroft and J. D. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley Publ. Comp. 1979.
- [Immerman, 1987] N. Immerman. Expressibility as a complexity measure. In *Second Structure in Complexity Conference*, pp. 194–202. Springer, 1987.
- [Kowalski, 1974] R. A. Kowalski. Predicate logic as a programming language. In *IFIP 74*, pp. 569–574. North-Holland, 1974.
- [Kowalski and van Emden, 1976] R. A. Kowalski and M. H. van Emden. The semantics of predicate logic as a programming language. *J. Ass. Comp. Mach.*, **23**, 713–742, 1976.
- [Lassaigne and de Rougemont, 1995] R. Lassaigne and M. de Rougemont. *Logique et Complexité*. Hermes Editions 1995.
- [Leitsch, 1996] A. Leitsch. *The Resolution Calculus*. Springer, 1996.
- [Lloyd, 1984] J. W. Lloyd. *Foundations of Logic Programming*. Springer, 1984.
- [Minsky, 1967] M. L. Minsky. *Computation: Finite and Infinite Machines*. Prentice Hall, 1967.
- [Papadimitriou, 1985] C. H. Papadimitriou. A note on the expressive power of PROLOG. *Bulletin EATCS*, **26**, 21–23, 1985.
- [Reiter, 1978] R. Reiter. On closed world data bases. In *Logic and Data Bases*, H. Gallaire and J. Minker, eds. pp. 55–76. Plenum Press, 1978.
- [Robinson, 1965] J. A. Robinson. A machine-oriented logic based on the resolution principle. *Journal Ass. Comp. Mach.*, **12**, 23–41, 1965.
- [Shepherdson, 1988] J. C. Shepherdson. Negation in logic programming. In *Foundations of Deductive Databases and Logic Programming*, J. Minker, ed. Morgan Kaufmann, 1988.
- [Siekman, 1988] J. H. Siekman. Unification Theory. *J. Symb. Comp.*, **7**, 207–274, 1988.
- [Sterling and Shapiro, 1986] L. Sterling and E. Y. Shapiro. *The Art of PROLOG*. The MIT Press, 1986.
- [Urquhart, 1995] A. Urquhart. The complexity of propositional proofs. *Bull. Symb. Logic*, **1**, 425–467, 1995.
- [Vardi, 1982] M. Y. Vardi. The complexity of relational query languages. In *Proc. 14th ACM Symp. on Theory of Computing*, pp. 137–146. 1982.

## INDEX

- $L_{\alpha\beta}$ , 192
- $L_{\infty\omega}$ , 238
- $L_{\omega_1\omega_1}$ , 192
- $L_{\omega_1\omega}$ , 192, 193
- $Pi_1^1$ , 212
- $S_n^m$  theorem, 270, 272
- $W_e^k$ , 277
- $\Delta$ -elementary, 213
- $\Delta\Sigma$ , 213
- $\Delta_1^1$ , 221
- $\Delta_m^n$ , 220
- $\Pi_1^0$ , 211
- $\Pi_n^0$ , 221
- $\Pi_1^1$ , 208–216, 227, 232
- $\Pi_n^1$ , 221
- $\Pi_m^n$ , 220
- $\Pi_1^1$ , 238
- $\Sigma$ -elementary, 213
- $\Sigma\Delta$ , 213
- $\Sigma\Delta$ -elementary, 213
- $\Sigma_1^0$ -completeness, 293
- $\Sigma_n^0$ , 221
- $\Sigma_1^1$ , 208–216, 218, 232
- $\Sigma_1^1$  arithmetical, 210
- $\Sigma_1^1$  sentences, 209
- $\Sigma_n^1$ , 221
- $\Sigma_m^n$ , 220
- $\lambda$ -calculus, 308
- $\omega$ -consistent, 294
- $\omega$ -logic, 154
- $\omega$ -model, 154, 162
- $\omega$ -logic, 154
- $\simeq$ , 269
- $\varepsilon$ -calculus, 55
- $\mathcal{L}\kappa\lambda$ , 176
- $\mathcal{L}(Q_1)$ , 166
- $\mathcal{L}(Q_0)$ , 153, 165
- $\mathcal{L}(Q_1)$ , 166
- $\mathcal{L}(Q_\alpha)$ , 165
- abbreviations, 14, 48
- abstract model theory, 189, 191, 192
- abstraction, 222
- absurdity, 6
- Ackermann's function, 266, 274
- Ackermann, W., 4, 205
- Ackermann, w., 76
- Aczel, P., 82
- admissible, 179
- admissible fragments, 193
- admissible langauges, 214
- Ajtai, M., 206
- algorithm, 248
- all, 190, 192
- almost all, 203
- Altham, J., 35
- ambiguous constants, 41
- ambiguous sign, 45
- analytic hierarchy, 132, 205, 221
- analytic hierarchy theorem, 221
- anaphora, 40
- ancestral, 154
- ancestral logic, 154
- Anderson, J., 28
- Anschauung, 10
- antecedent, 6
- anti-foundation axioms, 82
- Archimedean, 158
- Archimedean field, 158
- argument schema, 10, 42
- Aristotle, 1, 3
- arithmetic, 70, 108
- arithmetic truth, 216, 221

- arithmetical, 86, 205, 214, 216, 221, 302
- arithmetical definability, 205, 214
- arithmetical hierarchy, 211, 221
- arithmetical hierarchy, 302
- arithmetical truth, 214
- arithmetization, 288
- assignment, 10, 37, 48
  - first-order, 327
  - propositional, 316
  - suitalle, 49
- assignments, 50
- assumptions, 26, 54
- atomic formula, 8, 46
- Ax, J., 78
- axiom of choice, 88, 117, 135, 171, 230
- axiom of extensionality, 71
- axiom of regularity, 114
- axiomatise, 67
- axiomatizable theory, 291
- axioms, 29, 67, 71, 105, 113
- axioms of choice, 231
  
- Bäuerle, R., 37
- back-and-forth equivalence, 91
- back-and-forth equivalent, 92
- Barendregt, H. P., 234, 236, 238, 246
- Barwise, J., 35, 49, 77, 80, 82, 93, 94, 97, 98, 100, 101, 133, 192–194, 197, 203, 207, 209, 210, 212, 214, 219, 228, 236, 238
- basic, 12, 54
- Behmann, H., 76
- Bell, J. L., 18, 22, 47, 61, 69, 118, 213
- Belnap, N. D., 16, 28, 32
- Bencivenga, E., 53, 107
- Bentham, J. F. A. K., van, 77, 86, 95, 113, 211–213, 227, 231, 232, 235–238
- Bernays' axiom of choice, 230
  
- Bernays, P., 18, 55, 66, 69, 119, 209
- Beth theorem, 215
- Beth's Definability Theorem, 69, 99, 163
- Beth, E. W., 21, 61, 99
- biconditional, 6
- Bocheński, I., 53
- body, 340
- Bolzano, B., 41, 42, 52–54
- Bolzano-entail, 42
- Bolzano-valid, 42
- Boole, G., 1
- Bools, G., 86, 141, 237
- Bornat, R., 100
- bound in , 40
- bounded quantification, 263
- brackets, 14
- branching quantifiers, 168, 210, 235
- Brouwer, L., 86
- Burleigh, W., 53
- Byrne, R. M. J., 103
  
- cancelled, 26
- canonical codings, 289
- Cantor, G., 93, 117, 202, 207
- cardinalities, 148
- cardinality, 48, 117
- cardinality quantifier, 153
- cardinals, 117
- Carnap, R., 44, 51, 71
- Cartesian product, 269
- categorial grammar, 222, 238
- categorial, 132
- Chang logic, 168
- Chang quantifier, 167
- Chang, C. C., 69, 100, 118, 197, 207
- characterisations of first-order definability, 194
- Chastain, C., 40
- Cheng, P. W., 102
- Chomsky, N., 3
- Church's Theorem, 86, 151, 297

- Church's Thesis, 62, 284, 285, 287
- Church, A., 61, 72, 86, 119, 133, 221, 222
- Church–Turing Thesis, 284, 362
- Church-Rosser theorem, 234
- Clarke, M., 2
- clash of variables, 47
- class of models *elementary*, 197
- classes, 113
- classification of structures, 72
- clause
  - DATALOG, 340
  - Horn, 322
  - propositional, 321
  - unifiable, 350
- closed, 47
- closed classes, 78
- closed world assumption, 320, 334
- closure properties of RE-sets, 279
- coding, 258
- Coffa, J. A., 69
- Cohen, P. J., 6, 78
- combinatory logic, 308
- commutative groups, 68
- compact, 131, 136
- compactness, 189–194, 200, 203, 212, 236
- compactness property, 200
- Compactness theorem, 235
- compactness theorem, 63, 70, 95, 97, 117, 118, 190, 201, 202
- Compatibility Lemma, 353
- complements of RE sets, 303
- complete, 24, 54, 131, 193, 304
- complete axiomatisation, 205
- complete set of axioms, 71
- complete theory, 70
- completeness, 189, 192
- completeness theorem, 56, 201
- complexity, 10, 47
  - computational, 23
- composition, 261
- compound formulas, 8
- comprehension, 135, 145, 231
- comprehension axioms, 230
- computation, 251, 271
- computational complexity, 237
- computer science, 3, 237, 238
- computer scientists, 2
- conclusion, 1, 24, 26, 42, 106
- configuration, 364
- conjunction, 6, 15
- conjunctive normal form, 18
- conjuncts, 6, 15
- connective, 8
- consequence
  - first-order, 327
  - propositional, 316
- consequent, 6
- conservative, 293
- conservative extensions, 73
- consistent, 56
- constant, 66, 326
- constant function, 270
- constructing a model, 56
- contextual restrictions, 33
- contraction of quantifiers, 279
- contradictory opposite, 145
- converges, 271
- converges  $\downarrow$ , 277
- Cook, S. A., 23
- Cooper, R., 77, 194
- Corcoran, J., 144
- countably compact, 136
- counterexample, 22, 52
- course-of-value recursion, 266
- course-of-values induction, 10, 109
- Cox, R., 101
- Craig's Interpolation Lemma, 18, 99, 215
- Craig's Theorem, 230
- Craig's trick, 215
- Craig, W., 18, 99, 291
- Cresswell, M., 37
- cut elimination, 26, 61
- cut rule, 26
- cut-free sequent proofs, 23, 25

- cut-off subtraction, 262
- Dalen, D. van, 7, 28, 84, 104, 110
- database, 314
  - completed, 320
  - relational, 341
- DATALOG, 340
- Davis, M., 305
- De Morgan, A., 1
- decidable, 260, 277
- decision method, 19
- decision method for logic, 247
- Dedekind completeness, 202, 212
- Dedekind, R., 70, 108, 151, 201, 203, 260
- deducible from, 24
- deduction theorem, 16, 106, 146
- deep structures, 31
- definability, 208
- definable, 83
  - DATALOG, 366
- definite descriptions, 74, 92
- definition by cases, 272, 280
- definition by induction, 109
- definition of truth, 88
- definitional extension, 294
- definitions, 67, 100, 113
- degrees of unsolvability, 283, 308
- dependency prefix, 169
- derivable, 106
  - H-, 323
  - UH-, 355
- derivable formula, 24
- derivation, 106
- derivation rules, 29, 106
- descriptive complexity theory, 369
- diagonalization, 268
- diagram
  - positive, 337
- discharged, 26
- discriminator function, 270
- disjunction, 6, 15, 192
- disjunctive normal form, 18, 54
- disjuncts, 6
- diverges, 271
- diverges  $\uparrow$ , 277
- Došen, K., 3
- Doets, K., 77, 86, 95, 113, 189, 236
- Doherty, M., 238
- domain, 44, 48, 327
- domain of quantification, 33, 44
- downward Löwenheim–Skolem Theorem, 190
- downward Löwenheim–Skolem, 203
- downward Löwenheim–Skolem Theorem, 90, 93, 95, 97, 200
- downward Löwenheim–Skolem theorem, 202
- Dowty, D., 3
- Drake, F. R., 208
- Dummett, M., 6, 32
- Dunn, J. M., 16, 32
- Ebbinghaus, H.-D., 95, 236
- effective Borel hierarchy, 303
- effective topos, 286
- effectively inseparable sets, 282
- effectively regular, 137
- Ehrenfeucht, A., 95, 97
- Ehrenfeucht–Fraïssé game, 92, 95, 236
- elementarily equivalent, 89
- elementary, 213
- elementary class, 156
- elementary first-order predicate logic, 1
- elementary logic, 189
- elementary predicate logic, 131
- elements, 48
- elimination rule, 26, 54
- empty, 48
- empty domain, 52
- empty structures, 107
- encode, 112
- encoding, 80, 83
- end markers, 253
- Enderton, H. B., 52, 210, 231

- entail, 1
- entailment, 16
- Entscheidungsproblem, 363
- equality of RE sets, 281
- equivalence relations, 68
- Etchemendy, J., 52, 100, 101
- Euclid, 70
- Euclid's algorithm, 246
- Evans, G. , 40
- existence predicate, 269
- existential quantifier, 39
- existential second-order quantifiers
  - $\Sigma_1^1$ , 208
- expressive power (of second-order logic), 201
- expressive power of first-order logic, 77
- extension of first-order logic, 191
- extensionality, 114
- extensions, 43, 189
- extensions of first-order logic, 91, 194
  
- factorial function, 262
- falsehood, 5
- feasible, 316
- Feferman, S., 2, 35, 55, 80, 185
- Felscher, W., 87
- Field, H., 238
- Fine, K., 239
- finitary, 191
- finite, 48, 117
- finite intersection property, 118
- finite model theory, 236
- finite occurrence property, 137
- finite structures, 95
- first-order, 2, 202, 204, 217, 237
- first-order completeness, 215
- first-order definability, 194, 195, 198, 204, 208, 214
- first-order definable, 68, 197, 199
- first-order equivalent, 136
- first-order language, 46, 116
- first-order logic, 1, 91, 102, 131, 189, 191, 193, 194, 201, 202, 204, 225, 235
- first-order model theory, 236
- first-order Peano Arithmetic, 26, 70, 86, 90
- first-order Peano arithmetic, 111
- first-order schemas, 52
- first-order sentence, 197
- first-order syntax, 46
- first-order theory, 203
- first-order variable, 217
- Fitch, F. B., 28
- fixed point, 345
  - least, 345
- fixed point theorem, 294
- fixed-point logics, 238
- flow diagram, 250
- Flum, J., 94, 95, 98, 236
- Fong, G. T., 102
- formal, 62
- formal logic, 1
- formal proof, 24, 106
- formal proof calculus, 24
- formal system, 24
- formation tree, 8
- formula
  - atomic, 327
  - DATALOG, 342
  - first-order, 327
  - propositional, 315
  - universal, 327
- formulas, 7, 11, 46, 105
- Fraïssé, 93, 189, 198–200
- Fraïssé's Theorem, 197
- Fraenkel, A. A., 79, 113, 204
- fragments of the second-order language, 205
- Fraïssé, R., 95
- free for, 47
- free logic, 53
- free occurrence, 40, 105
- free variable, 37, 40

- Frege, G., 2, 5, 11, 29, 34, 35,  
 44, 45, 69, 100, 111, 113,  
 189, 201, 209  
 full type structure, 225  
 fully interpreted first-order language,  
 52  
 function, 66, 116  
 function constants, 66, 104  
 function quantifiers, 218  
 function symbol, 66, 326  
 functional, 76  
 functional application, 222  
 functional type theories, 222
- Gödel numbers, 289  
 Gödel's  $\beta$ -function, 267, 293  
 Gödel, K., 2, 4, 245, 261, 294, 308  
 Gabbay, D. M., 186, 239  
 Gallin, D., 223, 224, 227, 234  
 game, 86  
 Gandy, R. O., 80, 285  
 Garey, M., 23  
 Garland, S. J., 206  
 general completeness, 228  
 general model, 225, 226, 228, 230,  
 237  
 generalised continuum hypothesis,  
 168  
 generalised first-order definable, 68  
 generalised quantifier theory, 237  
 generalised quantifier *most*, 191  
 generalised quantifiers, 192, 203,  
 237  
 generalised sequents, 22  
 generative semanticists, 3  
 Gentzen, G., 2, 21, 23, 26, 28, 29,  
 54, 55  
 genuine logic, 78  
 geometry, 45  
 Girard, J.-Y., 3  
 Givant, S., 81  
 Gödel number, 85  
 Gödel's completeness theorem, 204
- Gödel, K., 56, 63, 70, 80, 81, 84–  
 86, 111, 112, 119, 204,  
 221  
 Goldblatt, R., 238  
 Goldfarb, W. D., 35  
 Goldson, D., 100  
 Gómez-Torrente, 42  
 grammar, 4  
 graph of a partial function, 279  
 Greenbaum, S., 78  
 Grice, H. P., 6  
 Groenendijk, J., 40  
 ground instance, 328, 346  
 ground term, 327  
 groups, 67  
 Guenther, F., 104  
 Gunter, C., 238  
 Gurevich, Y., 3
- Härtig quantifier, 168  
**HA**, 286  
 Halting Problem, 257, 298, 304  
 Hammer, E., 101  
 Hanf, 192  
 Hanf number, 206  
 Hanf numbers, 207  
 Harel, D., 238  
 Harnik, V., 80  
 Hasenjaeger, G., 62  
 Hausdorff, F., 93  
 head, 340  
 Heim, I., 40  
 Henkin models, 72  
 Henkin quantifier, 169  
 Henkin, L. A., 56, 59, 63, 72, 87,  
 88, 145, 210, 216, 222,  
 225, 230, 234, 235, 237  
 Henkin-style argument, 59  
 Herbrand structure, 328, 332, 336  
 Herbrand's Theorem, 329  
 Herbrand, J., 2, 4, 10, 26, 55, 261  
 Herbrand–Gödel, 308  
 hierarchies, 308  
 Hierarchy Theorem, 304



- higher-order, 203, 210  
 higher-order formulas, 216  
 higher-order logic, 131, 189, 192,  
     207, 216, 218, 225, 231,  
     232, 235–238  
 higher-order model theory, 214  
 higher-order proof theory, 239  
 higher-order sentences, 211  
 Hilbert, D., 2, 4, 10, 18, 29, 45,  
     52, 55, 66, 69, 84, 86,  
     149, 247  
 Hilbert–Bernays completeness the-  
     orem, 305  
 Hilbert-style proof calculus, 29, 104,  
     107  
 Hindley, J., 238  
 Hintikka sentences, 97  
 Hintikka set, 56  
 Hintikka, J., 21, 44, 56, 87, 90, 97,  
     175, 210, 218, 236  
 Hodges, W., 10, 46, 53, 81, 82, 91,  
     189, 193, 202, 226, 236  
 Horn clause, 322  
     negative, 323  
     positive, 323  
 Horn formula, 300, 317  
     negative, 317  
     positive, 317  
     strict, 317  
 Horn resolution, 323  
 Horn sentence, 331  
     negative, 331  
     positive, 331  
     strict, 331  
     universal, 331  
 Huntington, E. V., 93  
 Hyland, M., 286  
 hyperarithmetical, 221  
 hyperproof, 101  
  
 identity, 63, 73  
 Immermann, N., 237  
 implicit and explicit definitions, 100  
 implicit definition, 69  
  
 include, 156  
 incompleteness, 298  
 incompleteness theorem, 308  
 index, 269  
 indexical, 45  
 individual constants, 104  
 individual variable, 33  
 individuals, 119  
 induction, 135, 212  
 induction axiom, 108, 201  
 induction principle, 146  
 inductive definition, 11, 110  
 infinitary, 192  
 infinitary language, 175  
 infinitary logic, 90, 176, 238  
 infinite conjunctions, 192  
 infinity, 115  
 informal, 62  
 initial functions, 261  
 initial segment, 108  
 instance, 8, 346  
 instances, 1  
 intended interpretation, 70  
 intended models, 70  
 intensional logic, 212, 238  
 intensional type theory, 223  
 interpolation, 193  
 interpolation property, 163  
 interpolation theorem, 211, 215,  
     238  
 introduction rule, 26, 54  
 intuitionistic logic, 28, 239  
 isomorphic, 92  
 isomorphism property, 136  
  
 Jeffrey, R. C., 22, 61  
 Jeroslaw, R. G., 289  
 Johnson, D. S., 23  
 Johnson-Laird, P. N., 103  
 Johnstone, H. W., 28, 81  
 Johnstone, P. T., 81  
 jump, 303  
  
 Kalish, D., 10, 14, 28, 76

- Kalmár, L., 30  
 Kamp, H., 37, 40, 104, 227  
 Kanellakis, P., 237, 238  
 Kaplan, D., 76, 141  
 Karp property, 138  
 Karp, C., 93  
 Keenan, E., 237  
 Keisler's Theorem, 195, 197, 211  
 Keisler, H. J., 69, 100, 118, 189, 192, 197, 207  
 Kemeny, J., 219  
 Kempson, R., 4  
 kernel, 327  
 Kleene, S. C., 12, 18, 30, 62, 76, 84, 85, 106, 107, 209, 221, 269, 273, 286, 302, 308  
 Klenk, V., 82  
 Kneale, W., 29  
 knowledge based system, 314  
 Kochen, S., 78  
 Kowalski, R. A., 3  
 Koymans, K., 246  
 Krabbe, E., 246  
 Kreisel's Basis Theorem, 305  
 Kreisel, G., 62, 89, 285, 286  
 Kreisel, H. J., 207, 239  
 Kremer, P., 236, 239  
 Kripke, S., 32, 80  
 Krivine, J. L., 89  
 Krom formulas, 300  
 Kronecker, L., 56  
 Kunen, K., 206  
 König's tree lemma, 59  
  
 Löb, M. H., 239  
 Lönning, U., 238  
 Löwenheim number, 206  
 Löwenheim, L., 2, 139, 207  
 Löwenheim–Skolem, 189–194, 200, 228  
 Löwenheim–Skolem Theorem, 63  
 Löwenheim–Skolem theorem, 201, 235  
 label, 361  
  
 Ladd, C., 34  
 Lakoff, G., 3, 71  
 lambda abstraction, 222, 225, 233  
 lambda calculus, 233, 238  
 lambda-abstraction, 216  
 Landau, E., 261  
 Langford, C. H., 2, 4  
 language, 104  
 Lapierre, S., 238  
 Leblanc, H., 32, 50  
 Lehman, D. R., 102  
 Leibniz, 3, 200, 225  
 Leibniz' Law, 65  
 Leisenring, A. C., 55  
 Leivant, D., 237  
 Lemma  
     Compatibility, 353  
     Resolution, 322  
 Lemmon, E., 28  
 length, 265  
 Lepage, F., 238  
 Levy, A., 80, 114  
 Liar Paradox, 294  
 limitations, 189  
 Lindström's Theorem, 79, 95, 194  
 Lindström, S., 137, 189, 191–193, 199, 237  
 Lindström, P., 79, 91, 93, 95, 97  
 linguistic semantics, 194  
 linguistics, 3  
 Link, G., 236, 238  
 literal  
     first-order, 328  
     propositional, 317  
 logic, 91  
     first order, 30  
     first-order, 326  
     first-order with equality, 335  
     fixed point, 345  
     least fixed point, 346  
     propositional, 315  
 logic programming, 313  
 logical consequence, 52, 132, 151  
 logical forms, 3, 31

- logical implication, 19, 61  
 logical truth, 147, 151  
 logically equivalent, 16, 53, 76, 317  
 logically implies, 52  
 logically imply, 14  
 logically valid, 42, 52  
 logicist program of Frege and Russell, 189  
 Lorenzen, P., 87  
 Łoś Equivalence, 195–197  
 Łoś–Tarski theorem, 213  
 Łoś, J., 118  
 Łukasiewicz, J., 15  
 Łukasiewicz, J., 29, 55  
 Löwenheim, L., 63  
 Löwenheim–Skolem Theorem, 89, 131  
 Lévy, A., 84  
  
 $\mu$ -recursive function, 267  
 Machover, M., 18, 22, 47, 61  
 Magidor, M., 203, 207  
 Mal'tsev, A., 63  
 Malitz, J., 203  
 Manktelow, K. I., 103  
 many-one reducible, 304  
 many-sorted, 35  
 many-sorted first-order logic, 225  
 many-sorted language, 51  
 many-sorted logic, 216  
 Manzano, M., 236, 237  
 Markov algorithms, 308  
 Mason, I., 193  
 material implication, 6  
 Mates, B., 10, 29  
 mathematical induction, 10  
 Matijasevič, Y., 305  
 maximising argument, 59  
 McCarty, D., 286  
 McCawley, J. D., 3  
 meaning postulates, 71  
 mechanical, 285  
 Mendelson, E., 30, 61, 119  
 Menger, K., 63  
  
 mental mechanisms, 100, 103  
 metalanguage, 7  
 metavariables, 7, 8, 46, 52  
 minimal closure, 131, 147, 148  
 minimalization, 255, 267, 276  
 Mirimanoff, 113  
 Mitchell, J. C., 238  
 Mitchell, O. H., 34  
 modal definability theory, 238  
 modal logic, 239  
 model, 12, 46, 48, 49, 52  
     first-order, 327  
     minimal, 320, 334  
     propositional, 316  
 model of, 52  
 model theorist, 2, 31, 79  
 model theory, 131  
 model-theoretic, 135  
 modus ponens, 27, 106, 107  
 monadic, 139  
 monadic part, 208  
 monadic second-order theory, 205  
 monadic second-order variables, 203  
 Monk, D., 191  
 monotonicity of logical calculi, 334  
 Montague Grammar, 3, 210, 216  
 Montague, R., 3, 10, 14, 28, 76, 77, 216, 222, 223, 227  
 Moore, G., 119  
 Morrill, G., 4  
 Moschovakis, Y. N., 221  
 Moss, L., 82  
 most, 190, 192, 198, 203, 212  
 Mostowski, A., 63, 302  
 Mostowski, M., 221, 236  
 Muskens, R., 236, 237  
 Myhill, J., 208  
  
 $n$ -equivalent, 95  
 $n$ -place predicate, 37  
 N, 288  
 naive arithmetic, 108  
 name, 36, 53  
 natural deduction, 54

- natural deduction calculus, 26
- natural language, 237
- natural number, 115
- negation, 6
- Neumann, J. von, 80, 113, 115
- Nisbett, R. E., 102
- non-Archimedean models, 190
- non-axiomatisability, 204
- non-empty domains, 53
- non-standard analysis, 79
- non-standard models, 70, 141, 190
- non-standard models of ZFC, 82
- non-terminating computation, 271
- non-well-founded, 82
- non-well-founded sets, 82
- normal form
  - conjunctive, 317, 327
  - disjunctive, 317, 327
- normal form theorem, 275, 306
- notion of computability, 285
- noun phrases, 73
- NPTIME, 316
- null-set, 114
- numerals, 111, 291
  
- Oberlander, J., 101
- occurrences of the variable, 39
- Ohlbach, H. J., 238
- ontological commitment, 180
- open classes, 78
- oracle, 283
- ordinal, 115
- ordinary type theory, 220
- Orey, S., 227
- Over, D. E., 103
- overspill, 79
  
- Pèter, R., 287
- Padawitz, P., 81
- pair-set, 114
- palindrome, 247
- palindrome tester, 249
- paraphrased, 31
- Partee, B., 40
  
- partial function, 252
- partial isomorphism, 138, 197, 199
- partial isomorphisms, 197
- partial isomorphism, 199
- partial ordering, 68
- partial recursive function, 269
- partially isomorphic, 199
- partially ordered quantifier prefixes, 169
- Peano Arithmetic, 62, 108, 190, 205, 288
- Peano, G., 2, 108, 111
- Peirce, C. S., 1, 5, 29, 34, 35, 76
- Perry, J., 45
- philosophy of science, 238
- Pitts, A. M., 239
- Platek, 80
- plural quantifier, 141
- polynomial function, 23
- polynomial time, 344
- polynomially decidable, 363
- Popper, K. R., 28
- possible extensions, 189
- Post Systems, 308
- Post's correspondence problem, 298
- Post's Theorem, 211, 303
- Post, E., 4, 11, 17, 308
- power-set, 114
- Prawitz, D., 28
- predecessor function, 262
- predicate, 33, 34
- predicate constants, 104
- predicate logic, 30, 189
- predicative substitutions, 231
- premises, 1, 24, 26, 42, 106
- prenex form, 54
- Presburger, M., 300
- primitive recursion, 254, 261, 277
- primitive recursive functions, 110, 260, 261
- primitive recursive relation, 263
- principal, 118
- Principia Mathematica, 81, 216
- principle of completeness, 146

- principle of Identity of Indiscernibles, 200
- Prior, A., 15, 28
- production rules, 338
- productive set, 283
- program  
     DATALOG, 340  
     register, 361
- projection, 279
- projection function, 270
- projective class, 159
- projects, 159
- PROLOG, 3, 314
- pronouns, 40, 41
- proof, 24
- proof system, 104
- proof theorist, 2, 30, 62, 78
- proof theory, 131
- propositional language, 11
- propositional logic, 1, 4
- propositional quantification, 238
- $\text{Prov}(m, n)$ , 291
- psychologism, 100
- PSYCOP, 102
- PTIME, 316, 363, 366
- pure set, 119
- pure set structure, 61
- Putnam, H., 31, 82, 305
- Q**, 288
- quantification, 192, 201
- quantification over functions, 203, 209
- quantifier, 39, 73, 191  
     restricted, 35
- quantifier elimination, 300
- quantifier word, 34
- quantifiers, 32, 39, 190
- quasi-projects, 159
- quasi-satisfies, 145
- quasi-weak second-order logic, 164
- query  
     DATALOG, 342
- Quine, W. V. O., 7, 23, 28, 36, 73, 78, 83, 133, 189
- Quirk, R., 78
- Rabin's Theorem, 205
- Rabin, M. O., 205
- ramified analysis, 228
- ramified type theory, 220
- Ramsey quantifier, 168
- Rasiowa, H., 18, 56
- RE sets, 303
- realizability interpretation, 286
- recursion theorem, 273
- recursive axiomatisability, 192
- recursive definition, 109
- recursive relations, 221
- recursive saturation, 214, 215
- recursive set, 271, 278, 303
- recursively axiomatisable, 212, 230
- recursively enumerable, 205, 215, 221, 230
- recursively enumerable (RE), 277
- recursively enumerable in, 303
- recursively saturated, 215
- recursively separable sets, 282
- reducibility arguments, 304
- reductio ad absurdum, 27
- reduction class, 299
- reduction types, 308
- Reeves, S., 2, 100
- reflection, 270
- reflexivity of identity, 65
- register machine, 361
- regular ultrafilters, 118
- relation, 43, 115  
     successor, 365
- relation symbol, 326
- relative recursiveness, 283
- relativisation, 91
- relativisation predicate, 35
- relativization property, 137
- relevance logic, 239
- renaming, 349
- replacement, 117

- replacement principle, 147
- replendency, 215
- representability, 291
- Rescher quantifier, 168
- resolution
  - H-, 323
  - Horn, 323
  - UH-, 355
- Resolution Lemma, 322
- resolution method, 321, 322
- Resolution Theorem, 325
- resolvent, 322
  - U-, 352
- resplendency, 214, 215
- resplendent, 214, 215
- resplendent model, 214, 215
- Ressayre, J. P., 214, 215
- restricted quantifier, 48, 84
- restriction term, 33
- reverse mathematics, 135
- Rijke, M., de, 238
- Rips, L. J., 102
- Robinson, A., 79
- Robinson, J., 305
- Robinson, R., 288
- Robinson-consistency, 215
- Rogers, H., 272
- Rosser, J. B., 295, 308
- rule
  - DATALOG, 340
- Russell's Paradox, 221
- Russell's theory of finite types, 189
- Russell, B., 2–4, 74–76, 92, 113, 149, 189, 201, 220
  
- Sanchez Valencia, V., 237
- satisfaction, 36
- satisfiable, 89, 316
- satisfies, 36, 37, 43, 49
- satisfying, 91
- saturated models, 215
- schema, 1
- Schlipf, J., 214
- Schmidt, H., 55
  
- Schroeder-Heister, P., 3
- Schröder, E., 2, 34, 53
- Schwemmer, O., 87
- Schwichtenberg, H., 239
- Schütte, K., 26, 56
- scope, 40, 74
- Scott's theorem, 192
- Scott, D. S., 208
- Scott, W., 67
- second-order, 202–206, 210, 213, 217, 237
- second-order arithmetic, 113
- second-order cardinal characterization, 206
- second-order characterized, 207
- second-order definability, 208
- second-order definable, 202
- second-order definition, 203
- second-order logic, 71, 72, 131, 200–203, 207, 208, 238
- second-order monadic predicate logic, 205
- second-order Peano arithmetic, 208
- second-order quantification, 216
- second-order reduction, 232
- second-order sentences, 203, 212
- second-order set theory, 208
- second-order truth, 205, 207
- second-order validities, 205
- second-order validity, 205
- second-order variable, 209, 217, 228, 232
- Seldin, J., 238
- selection task, 103
- semantic tableaux, 21, 61
- semantics, 48
- sentence, 4, 37, 47, 107, 327
  - Horn, 331
- sentence letters, 7, 104
- sentence schema, 42
- separation, 115
- separator, 349
- sequence number, 265
- sequent, 14, 52

- proof, 19
- sequent calculus, 26
- set languages, 63
- set structures, 49
- set theoretic foundations of mathematics, 79
- set theory, 31, 113, 134, 206
- set-class theory, 119
- set-theoretic, 206
- Shapiro, S., 238
- Shelah, S., 141, 189, 197
- Shoenfield, J. R., 288
- Shoemith, D. J., 29
- signum function, 262
- Sikorski, R., 18, 56
- similarity type, 11, 46, 104
- situation, 4, 33, 39, 42
- Skolem functions, 86, 89, 170, 209, 231, 236
- Skolem normal form, 89, 209, 212, 299
- Skolem normal form for satisfiability, 89
- Skolem's Paradox, 81, 190
- Skolem, T., 2, 4, 55, 56, 58, 59, 70, 76, 79, 81, 82, 88–90, 93, 111, 113, 149, 204, 261
- Skolem–Behmann theorem, 77
- Skolemization, 170
- slash quantifiers, 90
- Slomson, A. B., 69, 118, 213
- Smiley, T. J., 29
- Smoryński, C., 86
- Smullyan, R., 22, 61
- Sneed, J. D., 69, 100
- solvable in polynomial time, 23
- some, 190
- sort, 51
- sortal, 35, 39
- sorted, 35, 39
- sound, 24, 54, 106
- spatial reasoning, 101
- standard identity, 64
- standard model, 70, 230
- state descriptions, 250
- stationary logic, 203
- Stegmuller, W., 69
- Steiner, M., 10
- Stenning, K., 101, 102, 104
- Stevenson, L., 32
- Stokhof, M., 40
- strategy, 88
- strongly complete, 24, 106
- structure, 12, 31, 44, 48
  - Herbrand, 328, 332, 336
  - named, 328
  - ordered, 365
  - quotient, 335
- Sturm's algorithm, 246
- subcomputation, 271
- subformulas, 46, 47
- subset, 114
- substitutable, 47
- substitution, 254, 261, 270
- substitution model, 181
- substitution operation, 290
- substitution property, 137
- substitution rule, 145
- substitutional quantification, 32, 180
- substitutional semantics, 180
- substitutor, 349
- successor function, 270
- suitable, 37
- Sundholm, G., 26, 28, 29
- Suppes, P., 28, 76, 80, 100, 114
- Svenonius' theorem, 212
- Svenonius, L., 212
- syllogism, 1
- symbol, 41
  - extensional, 340
  - intentional, 340
- syntactic turnstile, 24
- syntax, 112
- Tarski's Theorem, 158, 205, 214, 221, 296
- Tarski's theorem, 302

- Tarski's world, 100
- Tarski, A., 2, 42, 46, 51–55, 63, 69, 81, 84–86, 187, 205, 300
- tautology, 14
- Tennant, N. W., 28, 35
- term, 327
- terms, 46, 75, 104
- The Halting Problem, 281
- Theorem
  - Herbrand's, 329
  - on Logic Programming, 356
  - on the H-Resolution, 323
  - on the UH-Resolution, 356
  - Resolution, 325
  - U-Resolution, 360
- theory, 52
- third-order, 203
- Thm( $m$ ), 291
- Thomason, R., 10, 28
- tonk, 28
- topos, 81
- total, 271
- total ordering, 68
- total partial recursive function, 269
- traditional logician, 1, 30, 62, 78
- transfer, 79
- tree argument, 58
- Troelstra, A., 239
- true, 292
- true in, 12, 49
- truth, 2, 4, 49
- truth definition, 12, 49, 296
- truth-functors, 4
- truth-table, 5, 9, 10, 17
- truth-trees, 21
- truth-value, 5, 50
- Turing degrees, 283
- Turing machine, 248, 253, 268, 285, 308
- Turing reducible, 283
- Turing's Thesis, 248
- Turing, A., 221, 281, 308
- turnstile, 14, 24
- two-thirds quantifier, 76
- type structure, 225
- type theory, 221, 233, 238
- typed lambda calculus, 238
- types, 217
- ultrafilter, 117, 195, 196, 199
- ultraproduct, 69, 118, 139, 195–197, 199, 211–213, 232
- unbounded search, 255, 276
- undecidability, 297
  - of first-order logic, 363
  - of the halting problem, 362
  - of the Horn part of first-order logic, 363
- undecidability of predicate logic, 299
- undecidability of **PA**, 298
- undecidable, 297
- undecidable RE set, 281
- undefinability of truth, 84, 86
- underlining algorithm, 318
- unification algorithm, 350
- unifier, 350
  - the* general, 352
  - general, 350
- uniformity, 273
- union, 114
- universal, 89
- universal closure, 107
- universal Horn, 81
- universal machine, 270
- universal quantifier, 39
- universal quantifiers  $\Pi_1^1$ , 208
- universal Turing machine, 256, 268, 308
- universe, 327
- upward Löwenheim–Skolem Theorem, 95
- upward Löwenheim–Skolem, 97
- upward Löwenheim–Skolem Theorem, 202
- validity, 1, 62, 316



- validity of an argument, 1, 9, 10, 42
- valuations, 48
- van Benthem, J. F. A. K., 189, 194
- van Dalen, D., 62
- variable, 33
- variables, 46, 104
- Vaught, R. L., 2, 56, 63, 76, 298
- Venema, Y., 238
- Visser, A., 246
- vocabulary, 326
  
- Wang, H., 2, 55, 56
- Wason, P. C., 102, 103
- weak Keisler, 199
- weak second-order logic, 152, 191
- weakly compact, 136
- weakly complete, 24
- well-formed, 47
- well-founded, 202
- well-foundedness, 131, 213
- well-order, 131
- well-ordered, 109
- well-ordering, 117
- Westerståhl, D., 237
- Whitehead, A. N., 4
- winning strategy, 87, 88
- Wiredu, J. E., 55
- witnesses, 57
- Wittgenstein, L., 5
  
- yield, 24
  
- Zakharyashev, M., 236, 238
- Zermelo, 204
- Zermelo set theory, 116
- Zermelo, E., 79, 113, 175, 204
- Zermelo–Fraenkel axioms, 204
- Zermelo–Fraenkel set theory, 190, 208
- Zucker, J., 29

# Handbook of Philosophical Logic

2nd Edition

Volume 2

edited by Dov M. Gabbay and F. Guentner



## CONTENTS

Editorial Preface	vii
<b>Dov M. Gabbay</b>	
Systems of Deduction	1
<b>Goran Sundholm</b>	
Alternatives to Standard First-order Semantics	53
<b>Hugues Leblanc</b>	
Algebraic Logic	133
<b>Hajnal Andréka, Istvan Németi and Ildiko Sain</b>	
Basic Many-valued Logic	249
<b>Alisdair Urquhart</b>	
Advanced Many-valued Logics	297
<b>Reiner Hähnle</b>	
Index	396



## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good.!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London



<b>Logic</b>	<b>IT</b>			
	<b>Natural language processing</b>	<b>Program control specification, verification, concurrency</b>	<b>Artificial intelligence</b>	<b>Logic programming</b>
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs

<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical frag- ments</b>	Basic back- ground lan- guage	Program syn- thesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely use- ful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calcu- lus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syn- tax	Modules. Combining languages	Logics of space and time	Combining fea- tures
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game seman- tics gaining ground		
<b>Object level/ metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action- revision mod- els</b>			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model



GÖRAN SUNDHOLM

## SYSTEMS OF DEDUCTION

### 1 INTRODUCTION

Formal calculi of deduction have proved useful in logic and in the foundations of mathematics, as well as in metamathematics. Examples of some of these uses are:

1. The use of formal calculi in attempts to give a secure foundation for mathematics, as in the original work of Frege.
2. To generate syntactically an Already given semantical consequence relation, e.g. in some branches of technical modal logic.
3. Formal calculi can serve as heuristic devices for finding metamathematical properties of the consequence relation, as was the case, e.g. in the early development of infinitary logic via the use of cut-free Gentzen sequent calculi.
4. Formal calculi have served as the *objects* of mathematical study, as in traditional work on Hilbert's consistency programme.
5. Certain versions of formal calculi have been used in attempts to formulate philosophical insights into the nature of reasoning.

It goes without saying that a particular type of calculus which serves admirably for one of the above uses, which are just a small indication of the many uses to which calculi of deduction have been put, does not have to be at all suitable for some of the other. Thus a rich variety of different techniques have been developed for doing the 'book-keeping' of formal deduction, each one with its own advantages and disadvantages. It is the purpose of the present chapter to present a number of these techniques and to indicate some of the connections between the various versions.

More precisely, we shall concentrate on three main types of deductive systems known as (a) *Hilbert–Frege* style systems (b) *Natural deduction* systems and (c) *Sequent Calculi*. Under each of these headings we are going to study different variants of the main idea underlying the deductive technique in question. In particular, we shall relate the sequent calculus of Gentzen to currently fashionable '*tableaux*' systems of logic and show in what way they are essentially just a variant of the original Gentzen idea.

### A Remark about Notation

In the present chapter we largely follow the notation of Gentzen [1934], and in particular, we use ‘ $\supset$ ’ as implication and ‘ $\&$ ’ as conjunction, as well as the ‘*falsum*’ or *absurdity* symbol ‘ $\perp$ ’. The arrow ‘ $\rightarrow$ ’ we reserve for the *sequent arrow*.

Various versions of predicate logic can be formulated more conveniently by the use of a special category of free individual variables, ‘*parameters*’. As individual variables, free or bound, we use

$$x_0, x_1, \dots, y_0, y_1, \dots \quad x, y, z, \dots$$

and as parameters

$$a_0, a_1, \dots, b_0, b_1, \dots \quad a, b, \dots$$

Greek letters are used as schematic letters for formulae, cf. Hodges’ Chapter in Volume 1 of this *Handbook*.

we shall sometimes use subscripts on our turnstiles, e.g. ‘ $\vdash_{N\varphi}$ ’ will be used to indicate that  $\varphi$  is a theorem of a Natural Deduction system.

## 2 HILBERT–FREGE STYLE SYSTEMS

We begin by considering the historically earliest of the three main types of systems, viz. the Hilbert–Frege style systems. One of the main advantages of such systems in their contemporary versions is a certain neatness of formulation and typographical ease. It is therefore somewhat ironical that the first such system in [Frege, 1879] was typographically most cumbersome. The sort of presentation used here has become standard since the codification given in [Hilbert and Bernays, 1934].

The central notion in Hilbert–Frege style systems is *provability* (other terms used include *derivability* and *theoremhood*). One lays down that certain wffs are *axioms* and defines the *theorems* as the least class of wffs which contains all axioms and is closed under certain *rules of proof*. The most familiar of such rules is unquestionably *modus ponens*, MP, which states: If  $\varphi \supset \psi$  and  $\varphi$  are theorems, then so is  $\psi$ .

Sometimes the term ‘rule of inference’s is used where we use ‘rule of proof’. This is a question of terminological choice, but in order to emphasise the fact that, say, the premises and the conclusion of MP are all *theorems*, we prefer the present usage and reserve ‘rule of inference’ for those situations where the premises are allowed to depend on assumptions. In present terminology, therefore, the theorems are inductively defined by the axioms and rules of proof.

Let us begin by considering a very simple system for **CPC** based on  $\supset$  and  $\neg$  as primitives and with other connectives, e.g.  $\&$ , introduced by standard definitions, cf. Hodges (see Volume 1 of this *Handbook*).

The axioms are given in the form of *axiom schemata*, any instance of which is an axiom.

$$(A1) \quad \varphi \supset (\psi \supset \varphi)$$

$$(A2) \quad (\varphi \supset (\psi \supset \theta)) \supset ((\varphi \supset \psi) \supset (\varphi \supset \theta))$$

$$(A3) \quad (\neg\psi \supset \neg\varphi) \supset (\varphi \supset \psi).$$

Hence

$$(p_0 \supset p_1) \supset (\neg p_3 \supset (p_0 \supset p_1))$$

is an instance of A1, and if  $\varphi$  and  $\psi$  are wffs then

$$(\varphi \supset (\varphi \supset \psi)) \supset ((\varphi \supset \varphi) \supset (\varphi \supset \psi))$$

is a schema which gives a subclass of the axioms which are instances of A2.

We now define the theorems of the system:

DEFINITION 1. Every axiom is a theorem.

DEFINITION 2. If  $\varphi \supset \psi$  and  $\varphi$  both are theorems, then so is  $\psi$ .

The system so given may be called **HCPC**—‘**H**’ for Hilbert—and is known to be complete for tautologies. Note that **HCPC** has only three axiom schemata but *infinitely many axioms*, and that MP is the only rule of proof. Sometimes one sees formulations using three *axioms* only—the relevant instances of (A1)–(A3) for sentence variables  $p_0, p_1$  and  $p_2$ , but with an extra rule of proof instead. This is the

*Substitution rule*: The result of substituting a wff for a sentence variable in a theorem is a theorem.

In the case of predicate logic, the substitution rule becomes very complicated if one wants to formulate it in an exact way, cf. [Church, 1956, pp. 289–290], and formulations using axiom schemata have become almost universal.

We write ‘ $\vdash \varphi$ ’ to indicate that  $\varphi$  is a theorem of **HCPC**, and then the proper way to present MP becomes:

$$\frac{\vdash \varphi \supset \psi \quad \vdash \varphi}{\vdash \psi} \quad (\text{MP})$$

By the inductive generation of the theorems, to every theorem there corresponds a proof tree, or derivation, of  $\varphi$ . Such a tree  $D$  is a finite tree of wffs which is regulated by MP, has got  $\varphi$  at its root and axioms only as top formulae.

We are now going to give a proof tree for the schema  $\varphi \supset \varphi$ , thereby establishing that this is a *theorem schema*.

$$\frac{(\varphi \supset ((\varphi \supset \varphi) \supset \varphi)) \supset ((\varphi \supset (\varphi \supset \varphi)) \supset (\varphi \supset \varphi)) \quad \varphi \supset ((\varphi \supset \varphi) \supset \varphi)}{\frac{(\varphi \supset (\varphi \supset \varphi)) \supset (\varphi \supset \varphi) \quad \varphi \supset (\varphi \supset \varphi)}{\varphi \supset \varphi}}$$



This is a tree of the form

$$\frac{\frac{(1) \quad (2)}{(3) \quad (4)}}{(5)}$$

where (1) is a schematic instance of (A2), (2) is an instance of (A1) and (3) is a consequence by MP of (1) and (2). Likewise, (5) is an MP consequence of (3) and the (A1) instance (4). This is the shortest proof known to us of  $\varphi \supset \varphi$  in **HCPC**, and amply brings out one of the drawbacks of the Hilbert–Frege style systems. If one is interested in actually carrying out derivations in the systems, the work involved rapidly becomes enormous and quite unintuitive. If, on the other hand, we were allowed to use proofs from assumptions, then one could prove the above schema easily enough, provided that proofs from assumptions have the property that if  $\psi$  is provable from assumptions  $\varphi$  and  $\varphi_1, \varphi_2, \dots, \varphi_k$ , then  $\varphi \supset \psi$  is provable from assumption  $\varphi_1, \dots, \varphi_k$  only. We say that  $D$  is a *proof from assumptions*  $\varphi_1, \dots, \varphi_k$  of  $\varphi$ , if  $D$  is a finite tree of wffs regulated by MP and with  $\varphi$  as its end formula. All the top formulae of  $D$  are either axioms or one of  $\varphi_1, \dots, \varphi_k$ . Thus we may use the assumptions *as if they were axioms* in a proof from these assumptions. If there is a proof of  $\varphi$  from assumptions  $\varphi_1, \dots, \varphi_k$  we write

$$\varphi_1, \dots, \varphi_k \vdash \varphi.$$

This notion is extended to schemata in the obvious way. We can also define a consequence relation between possibly infinite sets  $\Gamma$  of assumptions and wffs by putting

$$\Gamma \vdash \varphi \text{ iff } \varphi_1, \dots, \varphi_k \vdash \varphi, \text{ for some } \{\varphi_1, \dots, \varphi_k\} \subseteq \Gamma.$$

For this notion of consequence from assumption by means of a proof tree from the assumptions, one is able to establish one of the central theorems of elementary metamathematics.

THE DEDUCTION THEOREM (Herbrand, Tarski). *If  $\Gamma, \varphi \vdash \psi$ , then  $\Gamma \vdash \varphi \supset \psi$ .*

**Proof.** (After Hilbert and Bernays [1934]).

By hypothesis, there is a proof tree ? of  $\psi$  from the assumptions  $\varphi$  and  $\Gamma$ . Such a  $D$  must, in principle, look like

$$\begin{array}{c} \varphi, \quad \varphi_1, \dots, \varphi_k, \quad \gamma_1, \dots, \gamma_m \\ \frac{\theta \supset \delta \quad \theta}{\delta} (MP) \\ \psi \end{array}$$

where the top formulae  $\varphi_1, \dots, \varphi_k$  are all assumptions from the set  $\Gamma$  and  $\gamma_1, \dots, \gamma_m$  are all axioms.

We need to find a proof tree for  $\varphi \supset \psi$  from assumption in  $\Gamma$  only. Consider first the ' $\varphi \supset$ ' transformation of  $D$ ; that is, in front of every wff in  $D$  we write ' $\varphi \supset$ '. This transformed tree, call it ' $\varphi \supset D$ ', is no longer a proof tree from assumptions but looks like:

$$\begin{array}{c} \varphi \supset, \varphi \supset_1, \dots, \varphi \supset_{\varphi_k} \quad \varphi \supset \gamma_1, \dots, \varphi \supset \gamma_m \\ \hline \varphi \supset (\theta \supset \delta) \quad \varphi \supset \theta; \quad \varphi \supset \text{MP}' \\ \hline \varphi \supset \psi \end{array}$$

Our task is thus to show that at each step in this transformed tree ' $\varphi \supset D$ ' we can restore provability from  $\Gamma$ . We begin by considering the three sorts of top formulae:

- (a) The top formula is  $\varphi \supset \varphi$ . We have already seen how to prove this without assumptions.
- (b) The top formula is one of the  $\varphi \supset \varphi_i$ , where  $\varphi_i$  is in  $\Gamma$ . Then we use (A1)—this is, in fact, the main *raison d'être* for the schema (A1). It is exactly what is needed to go from  $\varphi_i$  to  $\varphi \supset \varphi_i$ —given *modus ponens*—to construct a proof of  $\varphi \supset \varphi_i$  from  $\Gamma$ :

$$\frac{\varphi_i \supset (\varphi \supset \varphi_i) \quad \varphi_i}{\varphi \supset \varphi_i}$$

This proof tree uses only one assumption  $\varphi_i$  which we assume is a member of  $\Gamma$ . Thus, provability from  $\Gamma$  is also restored here.

- (c) The top formula is  $\varphi \supset \gamma_j$  where  $\gamma_j$  is an axiom. Then, in this case,

$$\frac{\gamma_j \supset (\varphi \supset \gamma_j) \quad \gamma_j}{\varphi \supset \gamma_j}$$

is actually a proof of  $\varphi \supset \gamma_j$  from no assumptions at all, and hence, *a fortiori*, provability from assumptions in  $\Gamma$  is also restored here.

This ends the discussion of the top formulae. It remains, however, to check that the transformation of instances of MP preserve derivability from assumptions in  $\Gamma$ . So assume that we are given proofs from assumptions in  $\Gamma$  of  $\varphi \supset \theta$  and  $\varphi \supset (\theta \supset \delta)$ . Using these given proof trees from assumptions

we continue via (A2):

$$(MP) \quad \frac{(\varphi \supset (\theta \supset \delta)) \supset ((\varphi \supset \theta) \supset (\varphi \supset \delta)) \quad \varphi \supset (\theta \supset \delta)}{\frac{(\varphi \supset \theta) \supset (\varphi \supset \delta) \quad \varphi \supset \theta}{\varphi \supset \delta} (MP)}$$

But, by hypothesis, we have obtained proofs from assumptions in  $\Gamma$  of  $\varphi \supset (\theta \supset \delta)$  and  $\varphi \supset \theta$ , and the remaining top formula is an axiom. Therefore, the here provability from  $\Gamma$  has also been secured and *the proof* of the Deduction Theorem *is completed*. ■

REMARK. The proof is very general and, in fact, shows that the Deduction Theorem holds for any system where the notion of consequence from sets of assumptions is introduced via proofs from assumptions, *provided* that

1. (A1) and (A2) are axiom schemata of the system, and
2. MP is the only rule of proof.

In order to see the importance of the second of these two conditions, we will consider a case where the Deduction Theorem does not hold, or better, where the notion of consequence from sets of assumptions cannot be introduced via proofs from assumptions, and where the latter are straightforwardly introduced, as in **HCPC** above. When proof trees are extended to those from assumptions, what in effect takes place is that MP is converted into a *rule of inference* rather than a rule of proof, because it now licenses the step from *inference* rather than a rule of proof, because it now licenses the step from  $\gamma \vdash \varphi \supset \theta$  and  $\Delta \vdash \varphi$  to  $\Gamma, \Delta \vdash \theta$ . (Here we use ‘ $\Gamma, \Delta$ ’ as an abbreviation of ‘ $\Gamma \cup \Delta$ ’ and similarly for ‘ $\Gamma, \varphi$ ’ and ‘ $\Gamma \cup \{\varphi\}$ ’.)

Consider now the modal logic **s4**, cf. Bull and Segerberg chapter in Volume D3 of this *Handbook*, where we have a further primitive connective  $\Box$  and  $\&$  is defined from  $\neg$  and  $\supset$ ), with the additional axiom schemata and extra rule of proof:

$$(A4) \quad \Box\varphi \supset \varphi$$

$$(A5) \quad \Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$$

$$(A6) \quad \Box\varphi \supset \Box\Box\varphi$$

$$\text{Necessitation (Nec)} \quad \frac{\vdash \varphi}{\vdash \Box\varphi}$$

If we were now to define proofs from assumptions in such a way that the proof trees have to be regulated by MP *and* Nec, then, as above, in the case of *modus ponens*, we would have converted *the rule of proof* (Nec) into a rule of inference, which licenses steps of the following form:

$$\frac{\Gamma \vdash_{\mathbf{S4}} \varphi}{\Gamma \vdash_{\mathbf{S4}} \Box \varphi}$$

Such a rule is not sound, however, for the standard semantics for  $\mathbf{S4}$ , and together with the Deduction Theorem, it leads to unacceptable consequences:

$$\frac{\frac{\varphi \vdash_{\mathbf{S4}} \varphi}{\varphi \vdash_{\mathbf{S4}} \Box \varphi} \text{ (Nec)}}{\vdash_{\mathbf{S4}} \varphi \supset \Box \varphi} \text{ (Deduction Theorem)}$$

In this case, one therefore introduces the notion of consequence from assumptions in another way:

$$\Gamma \vdash_{\mathbf{S4}} \varphi \text{ iff } \vdash_{\mathbf{S4}} \varphi_1 \& \dots \& \varphi_k \supset \varphi, \\ \text{for some } \varphi_1, \dots, \varphi_k \text{ in } \Gamma.$$

Note that this way of introducing consequences of assumptions has the same drawbacks as **HCPC** had before we introduced proofs from assumptions. The consequence from assumptions in **S4** is defined in terms of *provability* and hence, all the difficulties which adhere to straightforward provability also remain here. The Deduction Theorem holds for the turnstile, though; this is because in **HCPC** one can prove

$$\varphi \& \varphi_1 \& \dots \& \varphi_k \supset \psi$$

iff one can prove

$$\varphi_1 \& \dots \& \varphi_k \supset (\varphi \supset \psi).$$

We will give one more example of a system where the Deduction Theorem and its proof are of use, namely an axiomatic system for **CQC=**, classical predicate logic with identity. (Cf. Hodges in Volume 1 of this *Handbook*, in particular for the notions of term and free variable.) We use the universal quantifier  $\forall$  as a primitive.

The wff  $\psi$  is said to be a *generalisation* of  $\varphi$ , if for some variables  $x_1, \dots, x_k$ ,  $\psi$  is identical with  $\forall x_1 \dots \forall x_k \varphi$ , where the case  $k = 0$  is permitted.

The *axiom schemata* are:

(Q1)–(Q3) =<sub>def</sub> any generalisation of an instance of (A1)–(A3).

Any generalisation of the following:

(Q4)  $\forall x \varphi \supset \varphi_t^x$ , where  $t$  is a term *substitutable* for  $x$  in  $\varphi$ .

(Q5)  $\forall x (\varphi \supset \psi) \supset (\forall x \varphi \supset \forall x \psi)$

(Q6)  $\varphi \supset \forall x \varphi$ , provided that  $x$  does not occur free in  $\varphi$ .

(Q7)  $x = x$

(Q8)  $x = y \supset (\varphi \supset \varphi')$ , where  $\varphi$  is atomic and  $\varphi'$  results from  $\varphi$  by replacing  $x$  with  $y$  in zero or more (but not necessarily all) places in  $\varphi$ .

MP is the *only rule of proof*.

In (Q4) the notion ‘substitutable for  $x$ ’ needs to be explained as well as the substitution notation ‘ $\varphi_t^x$ ’. The latter stands for the expression which results from  $\varphi$  by replacing the variable  $x$ , wherever it occurs free in  $\varphi$ , by the term  $t$ . One can define this precisely by an induction:

0. For an atomic  $\varphi$ ,  $\varphi_t^x$  is the expression obtained by replacing every  $x$  in  $\varphi$  by the term  $t$ . (The use of ‘replacing’ can be replaced with another inductive definition.)
1.  $(\neg\varphi)_t^x =_{\text{def}} \neg(\varphi_t^x)$ ,
2.  $(\varphi \supset \psi)_t^x =_{\text{def}} \varphi_t^x \supset \psi_t^x$ ,
3.  $(\forall y\varphi)_t^x =_{\text{def}} \forall y\varphi$ , if  $x$  and  $y$  are the same variable,  
 $\forall y(\varphi_t^x)$  otherwise.

Hence,  $(x = y)_x^y$  is equal to  $(x = x)$  and  $\forall x(x = x)_t^x$  is equal to  $\forall x(x = x)$ .

Consider the wff  $\varphi =_{\text{def}} \neg\forall y(x = y)$ . Then  $\varphi_y^x =_{\text{def}} \neg\forall y(y = y)$ , and  $(\forall x\varphi \supset \varphi)_y^x =_{\text{def}} (\forall x\neg\forall y(x = y) \supset \neg\forall y(y = y))$ . This last sentence is not logically valid, because the antecedent is true whenever the individual domain has got more than one element and the consequence is never true. Thus, there is no lack of counter-models. In this phenomenon, sometimes known as ‘clashes between bound variables’, lies the reason for the restriction on schema Q4. One says that  $t$  is *substitutable for  $x$  in  $\varphi$* , if no free occurrence of  $x$  in  $\varphi$  lies within the scope of a quantifier which binds a variable of the term  $t$ . This notion can be precisely defined in the following way:

- (i) If  $\varphi$  is atomic, then  $t$  is substitutable for  $x$  in  $\varphi$ .
- (ii) If  $t$  is substitutable for  $x$  in  $\varphi$ , then  $t$  is substitutable for  $x$  in  $(\neg\varphi)$ .
- (iii) If  $t$  is substitutable for  $x$  in  $\varphi$  and in  $\psi$ , then  $t$  is substitutable for  $x$  in  $(\varphi \supset \psi)$ .
- (iv) (the crucial clause)  
 If either,  $x$  does not occur free in  $\forall y\varphi$ , or,  $y$  does not occur in  $t$  and  $t$  is substitutable for  $x$  in  $\varphi$ , then  $t$  is substitutable for  $x$  in  $\forall y\varphi$ .

In place of ‘substitutable for  $x$ ’ one sometimes sees the phrase ‘free for  $x$ ’.

We illustrate the use of the present formulation of **HCQC=** in an important

**Metatheorem.** *If  $x$  does not occur free in any wff of  $\Gamma$  and  $\Gamma \vdash \varphi$ , then  $\Gamma \vdash \forall x\varphi$ .*

(The notion of consequence from assumptions is defined via proof trees, just in the same way as before for **HCPC**. the proof of the Deduction Theorem works.)

**Proof.** Consider a proof tree  $D$  for  $\varphi$  from  $\Gamma$ . Then  $D$  must, in principle, be of the following form:

$$\begin{array}{c} \varphi_1, \dots, \varphi_k, \quad \gamma_1, \dots, \gamma_m \\ \vdots \qquad \qquad \qquad \vdots \\ \frac{\theta \supset \psi \quad \theta}{\psi} \text{(MP)} \\ \vdots \\ \varphi \end{array}$$

where  $\varphi_1, \dots, \varphi_k$  are members of the set of assumptions  $\Gamma$ , and  $\gamma_1, \dots, \gamma_m$  are axioms. We show that for each wff  $\delta$  which occurs in  $D$ ,  $\Gamma \vdash \forall x\delta$ . Consider first the cases where  $\delta$  is a top formula of  $D$ .

- (i)  $\delta$  is one of the axioms  $\gamma_1, \dots, \gamma_m$ , but any generalisation of an axiom is an axiom. Therefore,  $\forall x\delta$  is an axiom and, hence,  $\Gamma \vdash \forall x\delta$ .
- (ii)  $\delta$  is an element of  $\Gamma$ . By the hypothesis of the theorem,  $x$  does not occur free in  $\Gamma$  and, hence, *a fortiori*, also not in  $\delta$ . But then  $\delta \supset \forall x\delta$  is an instance of (Q6) and an application of MP proves  $\forall x\delta$  from assumptions in  $\Gamma$  only. This case provides us with the reason for the inclusion of axiom schema (Q6), just as the previous case gives the explanation for why every generalisation of an axiom is an axiom.

What remains to be considered is the case of *modus ponens*. So assume that we have already established that  $\Gamma \vdash \forall x(\theta \supset \psi)$  and  $\Gamma \vdash \forall x\theta$ . We must show that  $\Gamma \vdash \forall x\psi$ . This is accomplished by the use of a suitable instance of (Q5), viz.  $\forall x(\theta \supset \psi) \supset (\forall x\theta \supset \forall x\psi)$  and two applications of MP. Here, then, we see the reason for the inclusion of axiom schema (Q5).

Thus, the proof of our theorem is completed and this so-called *Generalisation rule* holds as a derivable rule of inference. ■

This sort of system for **CQC=**, which uses only MP as a rule of proof, is treated in great detail in a series of papers by Tarski [1965], Kalish and Montague [1965] and Monk [1965]. An elegant exposition is in Enderton

[2000], on which we have relied. This sort of system is sometimes of importance in quantified modal logic if one wishes to avoid the so called *Barcan formula*, cf. Kripke [1963].

We just want to remark that another common axiomatization is obtained by using (A1)–(A3) plus (Q4) with the same restriction and

(Q'5)  $\forall x(\varphi \supset \psi \supset (\varphi \supset \forall x\psi))$  if  $x$  does not occur free in  $\varphi$ .

Here, only the instances of the schemata, but *not their generalisations*, are axioms. There are, however, *two* rules of proof:

R'1: MP

R'2: If  $\varphi$  is a theorem, then so is  $\forall x\varphi$ .

This last rule is also known as '*generalisation*', but note that here it is a rule of *proof* and previously we showed that it was a (derived) rule of *inference* in **HCQC**=.

The equivalence between **HCQC** and the latter, primed, version (call it **H'CCQC**—for simplicity we leave = out) is readily established. We will not enter into details, but only note that the effect of the **HCQC** condition that every generalisation of an axiom is an axiom, is taken care of via the Generalisation rule of proof in **H'CCQC**. Detailed expositions of the **H'CCQC** type of system can be found in [Church, 1956; Mendelson, 1997].

The **HCQC** system is a system of *pure* predicate calculus. If we wish to deal with a specific first-order theory  $T$  we have to specify a language  $L_T$  and to define the 'non-logical axioms' of  $T$ , although the deductive machinery and the development of the theory remain, on the whole, unchanged.

In order to facilitate comparisons with the two other essentially different ways of presenting logical deduction, we find it convenient at this point to change the basic syntactic set-up used hitherto. As we hinted at in the *Remark on notation*, we shall use a separate category of *parameters*:  $a, b, \dots$  possibly with subscripts. The definition of *individual terms* then runs:

- (i) Individual constants are terms.
- (ii) Parameters are terms.
- (iii) If  $f^i$  is an  $i$ -place function symbol and  $t_1, \dots, t_i$  are all terms, then  $f^i(t_1, \dots, t_i)$  is a terms.
- (iv) Nothing is a term except by a finite number of (i)–(iii).

In future, we shall leave out the 'extremal clause' (iv) from our inductive definitions.

The language for **QC**, which we shall use in the sequel, is based on the full set of connectives;  $\&, \vee, \supset, \perp, \forall$  and  $\exists$ , where, however, we introduce negation by definition  $\neg\varphi =_{\text{def}} \varphi \supset \perp$ . The *definition of well-formed formula* runs as usual, except that we use a different clause for the quantifiers:

(+) If  $\varphi$  is a wff,  $b$  is a parameter and  $x$  is a variable *which does not occur in  $\varphi$* , then  $\forall x\varphi'$  and  $\exists x\varphi'$  are both wffs,

where  $\varphi'$  is the result of writing ' $x$ ' for ' $b$ ' wherever it occurs in  $\varphi$ .

The net effect of the two main changes—the use of parameters in place of free individual variables in the theory and the prohibition of quantifying with a variable over a wff, which already contains this variable—is to ensure that the same variable does not occur as both free and bound in a wff (this is given by the use of parameters in place of free occurrences of variables) and that no variable occurs bound 'twice over' in a wff. The properties are not important *per se* in the development of a Hilbert–Frege style system for **QC**, but they prove indispensable in the case of the sequent calculus. Note that although the restriction on quantification rules out such expressions as

$$\exists x(P(x)\&\forall xQ(x)),$$

where  $P$  and  $Q$  are predicate variables, from the class of wffs, this is not an impoverishment of the language because

$$\exists y(P(y)\&\forall xQ(x))$$

is a wff and has the same meaning as the forbidden expression.

We will use the same substitution notation as before, e.g. ' $\varphi_t^a$ ' denotes then the result of substituting the expression  $t$  everywhere for the expression  $a$  in the expression  $\varphi$ . Sometimes we wish to consider expressions which are just like wffs, except that they contain free variables in place of parameters. Such expressions are called *pseudo-wffs*. an example of a pseudo-wff is obtained by removing the quantifier prefix from a wff, e.g.  $P(y)\&\forall xQ(x)$  is a pseudo-wff. All the pseudo-wffs we shall have occasion to consider will be of this type.

We now give a version of intuitionistic predicate logic with identity, **HIQC=**, in a form which is particularly suited for establishing connections with the other main types of systems. The system is, essentially, due to Hilbert and Gentzen, cf. Gentzen [1934].

*Axiom schemata:* propositional part

$$(A \supset_1) \quad \varphi \supset (\psi \supset \varphi)$$

$$(A \supset_2) \quad (\varphi \supset (\psi \supset \theta)) \supset ((\varphi \supset \psi) \supset (\varphi \supset \theta))$$

$$(A \& I) \quad (\varphi \supset (\psi \supset (\varphi\&\psi)))$$

$$(A \& E_i) \quad \varphi_0\&\varphi_1 \supset \varphi_i, i = 0, 1.$$

$$(A \vee I_i) \quad \varphi_i \supset \varphi_0 \vee \varphi_1, i = 0, 1.$$



(A  $\vee$  E)  $(\varphi \supset \theta) \supset ((\psi \supset \theta) \supset (\varphi \vee \psi \supset \theta))$

This is the only complicated axiom so far; it says that given ways to reach  $\theta$  from  $\varphi$  and  $\psi$ , respectively, there is a way to go from  $\varphi \vee \psi$  to  $\theta$ .

these axiom schemata, together with MP, give *Minimal logic*. In intuitionistic logic we have one more axiom schema:

(A  $\perp$ )  $\perp \supset \varphi$ .

*Modus ponens* is the *only rule of proof* in the propositional part.

Hence, the Deduction Theorem holds and we establish two theorem schemata with its use:

$$\vdash (\varphi \& \psi \supset \theta) \supset (\varphi \supset (\psi \supset \theta))$$

and

$$\vdash (\varphi \supset (\psi \supset \theta)) \supset (\varphi \& \psi \supset \theta).$$

We reason informally:

By MP and (A  $\&$  I):  $\varphi, \psi \vdash \varphi \& \psi$ .

But then, by MP and A  $\&$  E:  $\varphi \& \psi \supset \theta, \varphi, \psi \vdash \theta$ .

So, by the Deduction Theorem (twice):  $\varphi \& \psi \supset \theta \vdash \varphi \supset (\psi \supset \theta)$ .

So, by the Deduction Theorem:  $\vdash (\varphi \& \psi \supset \theta) \supset (\varphi \supset (\psi \supset \theta))$ :

The other direction is left as an exercise.

There are further axioms for the quantifiers and the identity symbol.

(A  $\forall$  E)  $\forall x \varphi \supset \varphi_t^x$

(A  $\exists$  I)  $\varphi_t^x \supset \exists x \varphi$ .

Note that we have got ' $\varphi_t^x$ '. Hence, say,  $(a = a) \supset \exists x(a = x)$  is an instance of A  $\exists$  I, because

$$(a = x)_a^x =_{\text{def}} (a = a).$$

There are two rules of proof in the quantificational part:

$$(R \forall I) \frac{\vdash \varphi \supset \psi}{\vdash \varphi \supset \forall x \psi_x^a} \text{ provided that } a \text{ does not occur in } \varphi.$$

$$(R \exists E) \frac{\vdash \varphi \supset \psi}{\vdash \exists x \varphi_x^a \supset \psi} \text{ provided that } a \text{ does not occur in } \psi.$$

In these two rules, the parameter  $a$  is called the *eigen*-parameter, or the *proper* parameter, of the instance of the rule.

We note that the restrictions are necessary if the rules are to be sound. Clearly,  $\vdash P(a) \supset P(a)$ , but  $P(a) \supset \forall x P(x)$  is not logically valid. (Consider an interpretation with the domain of persons and interpretation of  $P$  as the

property of holding the world championship of chess. Over this particular interpretation, the assignment (at the moment of writing) of Anatoly Karpov to the parameter  $a$ , produces a counter-model. It is certainly not true that if Karpov is world champion, then anyone is. I, for one, am not.) It is also essential to grasp that (RVI) is a rule of proof. The corresponding rule of inference is not sound without further restrictions. Semantically, and also deductively in the *propositional* part, where  $\forall yQ(y)$  can be viewed just as another wff with no further structure,

$$P(a) \vdash \forall yQ(y) \supset P(a), \text{ say via } (A \supset_1) \text{ and MP}$$

if we define proof-trees in the usual way, but we cannot allow that

$$P(a) \vdash \forall yQ(y) \supset \forall xP(x)$$

as a similar counter-model to the Karpov one will show.

We therefore need to take particular care in the definition of proof trees from assumptions. The crucial restriction is this: If  $D$  is a proof tree for  $\varphi \supset \psi$  from certain top formulae, where  $\varphi_1, \dots, \varphi_k$  are all the top formulae in  $D$  which are not axioms, and the parameter  $a$  does not occur in  $\varphi$ , then

$$\frac{D \quad \varphi \supset \psi}{\varphi \supset \forall x\psi_x^a}$$

is a proof tree for  $\varphi \supset \forall x\psi_x^a - x$  from assumptions  $\varphi_1, \dots, \varphi_k$ , *provided that  $a$  does not occur in any of  $\varphi_1, \dots, \varphi_k$ .*

A similar *eigen*-parameter condition is imposed on applications of (R $\exists$ E) in proofs from assumptions. With these restrictions, the Deduction Theorem is valid and the ' $\varphi \supset$ ' transformation method of proof works.

Consider a proof tree  $D$

$$\frac{\varphi, \varphi_1, \dots, \varphi_k \quad \theta \supset \psi}{\theta \supset \forall x\psi_x^a} \text{ (RVI)}.$$

As (RVI) is permissible here, we know that (i) the *eigen*-parameter  $a$  does not occur in  $\theta$ , nor (ii) in any of the assumption formulae  $\varphi_1, \dots, \varphi_k$  and  $\varphi$ . The transformation gives a proof tree  $D'$  of  $\varphi \supset (\theta \supset \psi)$  from assumptions  $\varphi_1, \dots, \varphi_k$ , because in the restoration of provability from assumptions, we use only auxiliary proof trees of the form.

$$\frac{\varphi_1 \supset (\varphi \supset \varphi_1) \quad \varphi_1}{\varphi \supset \varphi_1} \text{ (MP)}$$

and this leaves only  $\varphi_1$  as an assumption formula. So we now continue  $D'$  as follows:

$$\begin{array}{c}
\frac{(\varphi \supset (\theta \supset \psi)) \supset (\varphi \& \theta \supset \psi) \quad \varphi \supset (\theta \supset \psi)}{\varphi \& \theta \supset \psi} \text{ (MP)} \\
\frac{(\varphi \& \theta \supset \forall x \psi_x^a) \supset (\varphi \supset (\theta \supset \forall x \psi_x^a)) \quad \varphi \& \theta \supset \forall x \psi_x^a}{\varphi \supset (\theta \supset \forall x \psi_x^a)} \text{ (R}\forall\text{I)} \\
\frac{(\varphi \supset (\theta \supset \forall x \psi_x^a)) \quad \varphi \& \theta \supset \forall x \psi_x^a}{\varphi \supset (\theta \supset \forall x \psi_x^a)} \text{ (MP)}
\end{array}$$

The application of (R $\forall$ I) is permissible because we know that the parameter  $a$  does not occur in  $\varphi \& \theta$ , nor in any of the assumption formulae of  $D'$ . The rest is just dotting the  $i$ 's and crossing the  $t$ 's using the two derivable schemata from the propositional part. The treatment of (R $\exists$ E) is similar.

The details of the entire development for rules of inference with parameters and the resulting Deduction Theorem are given meticulous treatment in Hilbert and Bernays [1934] and Kleene [1952, Sections 21–24].

We can now write (R $\forall$ I) as one condition of the consequence relation:

$$\text{If } \Gamma \vdash \theta \supset \psi, \text{ then } \Gamma \vdash \theta \supset \forall x \psi_x^a$$

provided that the *eigen*-parameter  $a$  does not occur in  $\Gamma, \theta$ . Note the similarity with the rule of inference which we showed was derivable in **HCQC**.

We still have to give the identity axioms:

$$\text{(A=I)} \quad (a = a)$$

$$\text{(A=E)} \quad (a = b) \supset (\varphi_a^x \supset \varphi_b^x), \text{ where } \varphi \text{ is } \textit{atomic}.$$

As these are *axioms*, the presence of parameters does not further complicate the proof trees.

Let us finally conclude our treatment of the Hilbert–Frege style systems by remarking that for all three systems just presented, propositional, quantificational and identity logic, the corresponding classical system results simply by adding either of the two axiom schemata

$$\text{(DN)} \quad \neg\neg\varphi \supset \varphi \text{ ('DN' for 'double negation')}$$

$$\text{(Excluded middle)} \quad \varphi \vee \neg\varphi$$

and that if we wish to use negation as a primitive the relevant intuitionistic rules are

$$\text{(A}\neg\text{I)} \quad (\varphi \supset \psi) \supset ((\varphi \supset \neg\psi) \supset \neg\varphi)$$

$$\text{(A}\neg\text{E)} \quad \varphi \supset (\neg\varphi \supset \psi)$$

The Hilbert–Frege style systems are particularly well suited for arithmetization of metamathematics, because the inductively defined objects have a very simple structure. Detailed treatment can be found in [Smoryński, 1977; Feferman, 1960].

## 3 NATURAL DEDUCTION

The Hilbert–Frege style systems have, as we have seen in Section 1, a reasonably smooth theory, but they suffer from one essential drawback: if one is interested in *actually carrying out* derivational work, they are hopelessly cumbersome, because even the simplest inferences have to be brought back to the fixed and settled axioms. The use of proofs from assumptions, and the ensuing Deduction Theorem, is an attempt to ease the derivational burden which is at least partially successful, particularly for the last of the formulations given above.

In Natural Deduction, on the other hand, one of the two main features is that all rules are *rules of inference* rather than rules of proof and, consequently, theoremhood is defined as the limiting case of derivability from the empty collection of assumptions. The other main feature of Natural Deduction is that the derivational use of each operator \$ — connective, quantifier, inductively defined predicate etc. — is regulated by two rules: one, the *introduction rule* for \$, (\$ I), which tells us how a sentence with \$ as its main operator may be inferred as a conclusion, how \$ may be introduced, and another rule, the *elimination rule* for \$, (\$ E), which tells us how further consequences may be drawn from a premise with \$ as its main operator, how \$ can be eliminated.

We now proceed directly to presenting these rules for a Natural Deduction version of **IQC**, which we henceforth call **NIQC**.

(A) **Assumption.** For any wff  $\varphi$ , the tree which consists of  $\varphi$  only is a derivation of  $\varphi$  which depends on the assumption  $\varphi$ .

If

(& I) 
$$\begin{array}{ccc} D_0 & & D_1 \\ \varphi_0 & \text{and} & \varphi_1 \\ \text{depend on assumptions } \varphi_1, \dots, \varphi_k & \text{and} & \psi_1, \dots, \psi_m, \text{ respectively,} \\ \text{then} & & \end{array}$$

$$\frac{\begin{array}{cc} D_0 & D_1 \\ \varphi \supset \psi & \varphi \end{array}}{\psi} (\supset E)$$

is a derivation of  $\varphi_0 \& \varphi_1$ , which depends on all of  $\varphi_1, \dots, \varphi_k, \psi_1, \dots, \psi_m$ .

(& E<sub>i</sub>)  $i = 0, 1$ . The elimination rule for conjunction is, properly speaking, not one, but two rules.

If

$$\frac{D}{\varphi_0 \& \varphi_1}$$

is a derivation of  $\varphi_0 \& \varphi_1$ , depending on  $\psi_1, \dots, \psi_m$ , then

$$\frac{D}{\frac{\varphi_0 \& \varphi_1}{(\& E_i)}} \varphi_i$$

is a derivation of  $\varphi_i$ , depending on the same assumptions.

( $\supset$  I) This is the most characteristic of Natural Deduction rules. It is also difficult to state precisely.

If

$$\frac{D}{\psi}$$

is a derivation of  $\psi$ , depending on assumptions  $\varphi_1, \dots, \varphi_m$ , then

$$\frac{D}{\psi} \quad (\supset I)$$

$$\varphi \supset \psi$$

is a derivation of  $\varphi \supset \psi$  depending on assumptions  $\varphi'_1, \dots, \varphi'_m$ , where this list results from  $\varphi_1, \dots, \varphi_m$  by removing some (all or no) occurrences of  $\varphi$ . We say that the removed occurrences have been *discharged* or *closed*.

( $\supset$  D) If

$$\frac{D_0}{\varphi \supset \psi} \quad \text{and} \quad \frac{D_1}{\varphi}$$

are derivations of  $\varphi \supset \psi$  and  $\varphi$ , respectively, depending on  $\varphi_1, \dots, \varphi_k$  and  $\psi_1, \dots, \psi_m$ , respectively, then

$$\frac{\frac{D_0}{\varphi \supset \psi} \quad \frac{D_1}{\varphi}}{\psi} \quad (\supset E)$$

is a derivation of  $\psi$  depending on all the assumptions  $\varphi_1, \dots, \varphi_k, \psi_1, \dots, \psi_m$ . The elimination-rule for  $\supset$  is nothing but MP construed as a *rule of inference*.

We now have enough rules to give a simple example:

$$\frac{\frac{\frac{\varphi^1 \quad \varphi^2}{(\varphi \& \psi)} (\& I)}{\psi \supset (\varphi \& \psi)^2} (\supset I)}{\varphi \supset (\psi \supset (\varphi \& \psi))^1} \supset I$$

This derivation tree is a *proof*, i.e. a derivation in which the assumptions have all been closed (there are no *open* assumptions left). It also illustrates how one sets out the derivations in practice. The assumptions are indexed with a numeral and, at the inference where an assumption is discharged, the numeral is written again to indicate closure. In practice, when the system is familiar, one does not always give the names of the rules, nor does one always indicate where the assumptions are discharged, but confines oneself just to crossing them out to indicate closure. The mechanism of the above derivation is thus: assume  $\varphi$  and  $\psi$ . By (& I) we get  $(\varphi \& \psi)$  depending on the assumptions  $\varphi$  and  $\psi$ . Therefore, by the use of ( $\supset$ I):  $\psi \supset (\varphi \& \psi)$ , now depending only on  $\varphi$ , and finally by one more use of ( $\supset$ I):  $\varphi \supset (\psi \supset (\varphi \& \psi))$ . The use of ( $\supset$ I) corresponds to the Deduction Theorem in **HIQC**. Note that we have here given a proof in **NIQC** of the **HIQC**-axiom (A& I). the reader may wish to try his hand at (A & E) as an exercise.

The above elementary example illustrates a point of principal importance. It is given as a derivation *schema* and we naturally wish that each instance thereof shall be a derivation. Consider then the, albeit somewhat extreme, choice of the propositional variable  $p$  both for  $\varphi$  and  $\psi$ . The result is

$$\frac{\frac{\frac{p^1 \quad p^2}{p \& p} (\& \text{I})}{p \supset (p \& p)^2} (\supset \text{I})}{p \supset (p \supset (p \& p))^1} (\supset \text{I})$$

We see that both assumptions  $p$  are struck out, but the discharge takes place at *different inferences*. The moral of this example is that not *all* assumptions of the form  $\varphi$  have to be discharged at an application of ( $\supset$ I) giving  $\varphi \supset \psi$  as a conclusion. In fact, no assumption needs to be discharged:

$$\frac{\frac{\varphi^1}{\psi \supset \varphi} (\supset \text{I})}{\varphi \supset (\psi \supset \varphi)^1} (\supset \text{I})$$

Here at the first application of ( $\supset$ I) *no* discharge takes place. We are given a derivation of  $\varphi$ , depending on certain assumptions—in fact only on  $\varphi$ —and we go on to a derivation of  $(\psi \supset \varphi)$  as we have the right to do by ( $\supset$ I). As we search for an occurrence to discharge, we see that there is none. Thus, with this permissiveness and the resulting liberal use of ( $\supset$ I) Natural Deduction is not suitable for *Relevant logic* (but cf. [Prawitz, 1965, Chapter VII]).

To sum up: *Discharge is a right, but not an obligation.*

as not all assumptions of the same form need to be discharged at the same place, one speaks of '*assumption classes*', where those assumptions of the same form which are discharged by the same inference belong to the

same assumption class. Leivant [1979] contains an exhaustive discussion of the need for assumption classes.

The rules given above were formulated in a rather elaborate way in order to be precise. In practice, one often sees the following sort of definition for the rule ( $\supset$ I): If

$$\begin{array}{c} \varphi \\ D \\ \psi \end{array}$$

is a derivation of  $\psi$ , depending on, among others, the assumption  $\varphi$ , then

$$\frac{\begin{array}{c} \varphi \\ D \\ \psi \end{array}}{\varphi \supset \psi} (\supset\text{I})$$

is a derivation of  $\varphi \supset \psi$ , where the indicated assumptions of the form  $\varphi$  have been *closed* (have been *discharged* or *cancelled*). This is a somewhat loose way of formulating the rule as the discussion of assumption-classes etc. shows. It has, however, a great intuitive appeal, and the rest of our rules will be set out in this fashion on the understanding that the precise versions of the rules have to be formulated in analogy with the exact statement of ( $\supset$ I).

We now give the rest of the rules:

(VI)  $i = 0, 1$ . (This is not one but two rules, just as in the case of ( $\&$  E).)  
If

$$\frac{D}{\varphi_i}$$

is a derivation of  $\varphi_i$ , depending on certain assumptions, then

$$\frac{\frac{D}{\varphi_i}}{\varphi_0 \vee \varphi_1} (\vee\text{I})_i$$

is a derivation of  $\varphi_0 \vee \varphi_1$ , depending on the same assumptions.

(VE) If

$$\frac{D}{\varphi \vee \psi}$$

is a derivation of  $\varphi \vee \psi$ , and

$$\begin{array}{ccc} \varphi & & \psi \\ D_1 & \text{and} & D_2 \\ \theta & & \theta \end{array}$$

are derivations of  $\theta$ , depending respectively on, among others, the assumptions  $\varphi$  and  $\psi$ , then

$$\frac{\begin{array}{ccc} & \varphi & \psi \\ & D & D_1 \quad D_2 \\ \varphi \vee \psi & \theta & \theta \end{array}}{\theta} \text{ (VE)}$$

is a derivation of  $\theta$ , where the indicated assumptions have been closed. (This could, and for exactness, should, be reformulated in a manner corresponding to the exact statement of ( $\supset$ I).)

The rule (VE) is a formal version of the type of reasoning known as ‘constructive dilemma’: we know that  $A$  or  $B$  is true. First case:  $A$  is true. Then  $C$  is also true. Second case:  $B$  is true. Again,  $C$  is also true. Therefore,  $C$  is true.

Using (VE) we can, of course, give a proof, i.e. a derivation without open assumptions, of the axiom (AVE):

$$\frac{\frac{\overline{\varphi \vee \psi^1} \quad \frac{\varphi \supset \theta^2 \quad \varphi^3}{\theta} \quad \frac{\psi \supset \theta^4 \quad \psi^5}{\theta}}{\theta} \text{ (VE)} \quad \frac{\theta}{\theta} \text{ (1)} \quad \frac{\theta}{\theta} \text{ (3, 5)} \quad \frac{\theta}{\theta} \text{ (4)} \quad \frac{\theta}{\theta} \text{ (2)} \quad \frac{\theta}{\theta} \text{ (2)}$$

(VI) If

$$\frac{D}{\varphi}$$

is a derivation of  $\varphi$ , depending on  $\psi_1, \dots, \psi_k$ , then

$$\frac{D}{\varphi} \quad \forall x \varphi_x^a$$

is a derivation of  $\forall x \varphi_x^a$ , depending on the same assumptions, *provided* that the parameter  $a$ , the *eigen*-parameter of the inference, does not occur in any of the assumptions  $\psi_1, \dots, \psi_k$ .

This rule is a codification of a type of reasoning very well-known to calculus students: ‘Pick an  $\varepsilon > 0$ . Then there is for this  $\varepsilon$  such and such a  $\delta > 0$ . But  $\varepsilon$  was chosen arbitrarily. Therefore, for *every*  $\varepsilon > 0$  there is such a  $\delta$ .’ The restriction on the *eigen*-parameter  $a$  is, of course, necessary. The Karpov counter-example given in Section 1 in a similar situation works here as well, and for the same reason.



( $\forall E$ ) If

$$\frac{D}{\forall x\varphi}$$

is a derivation of  $\forall x\varphi$ , depending on certain assumptions, then

$$\frac{\frac{D}{\forall x\varphi}}{\varphi_t^x} \quad (\forall E)$$

is a derivation of  $\varphi_t^x$ , depending on the same assumptions.

( $\exists I$ ) If

$$\frac{D}{\varphi_t^x}$$

is a derivation of  $\varphi_t^x$ , depending on certain assumptions, then

$$\frac{\frac{D}{\varphi_t^x}}{\exists x\varphi}$$

is a derivation of  $\exists x\varphi$ , depending on the same assumptions.

We now come to the most complicated of the rules, viz.

( $\exists E$ ) If

$$\frac{D}{\exists x\varphi}$$

is a derivation of  $\exists x\varphi$ , and

$$\frac{\varphi_a^x}{D_1}$$

$\theta$

is a derivation of  $\theta$  depending on, among others, the assumption  $\varphi_a^x$ , then

$$\frac{\frac{D}{\exists x\varphi} \quad \frac{\varphi_a^x}{D_1}}{\theta} \quad (\exists E)$$

is a derivation of  $\theta$  depending on all the assumptions used in  $D$  and  $D_1$ , except those of the indicated form  $\varphi_a^x$ , *provided* that the *eigen*-parameter  $a$  does not occur in  $\exists x\varphi$ , nor in  $\theta$  or any of the assumptions in  $D_1$ , except those of the form  $\varphi_a^x$ . For examples illustrating the need for the restrictions, we refer to [Tennant, 1978, Chapter 4.8].

(=I) For any term  $t, t = t$  (= I) is a derivation of  $t = t$ , depending on no assumptions.

(=E) If

$$\begin{array}{c} D \quad \text{and} \quad D_1 \\ t_0 = t_1 \quad \varphi_{t_0}^x \end{array}$$

are derivations of  $t_0 = t_1$  and  $\varphi_{t_0}^x$ , depending on certain assumptions, then

$$\frac{\begin{array}{c} D \quad D_1 \\ t_0 = t_1 \quad \varphi_{t_0}^x \end{array}}{\varphi_{t_1}^x} \text{ (= E)}$$

is a derivation of  $\varphi_{t_1}^x$  depending on all the assumption used in  $D$  and  $D_1$ .

Examples:

1.

$$\frac{(a = a)}{\exists x(a = x)} \text{ (= I)} \quad (\exists I)$$

and

2.

$$\frac{(a = a)}{\exists x(x = x)} \text{ (= I)} \quad (\exists I)$$

are correct proofs of their respective conclusions; in (1)  $(a = a) =_{\text{def}} (a = x)_a^x$  and in (2)  $(a = a) =_{\text{def}} (x = x)_a^x$

In order to complete the description of **HIQC=** we must add one more rule to the above set of rules for *Minimal* predicate logic with identity. The extra rule is of course

( $\perp$ ) If

$$\frac{D}{\perp}$$

is a derivation of  $\varphi$ , depending on the same assumptions.

This ends the presentation of the rules. In order to summarise the above description of the system of Natural Deduction we now give the rules as

inference figures:

$$\begin{array}{ll}
 (\& \text{I}) \quad \frac{\varphi \quad \psi}{\varphi \& \psi} & (\& \text{E}) \quad \frac{\varphi \& \psi \quad \varphi \& \psi}{\varphi \quad \psi} \\
 (\vee \text{I}) \quad \frac{\varphi}{\varphi \vee \psi} \quad \frac{\psi}{\varphi \vee \psi} & (\vee \text{E}) \quad \frac{\varphi \vee \psi \quad \begin{array}{c} \varphi \quad \psi \\ \vdots \quad \vdots \\ \theta \quad \theta \end{array}}{\theta} \\
 (\forall \text{I}) \quad \frac{\varphi \quad \vdots \quad \psi}{\psi} & (\supset \text{E}) \quad \frac{\varphi \supset \psi \quad \varphi}{\psi} \\
 (\forall \text{I}) \quad \frac{\varphi \supset \psi \quad \varphi}{\forall x \varphi_x^z} & (\forall \text{E}) \quad \frac{\forall x \varphi}{\varphi_t^x}
 \end{array}$$

provided that the *eigen*-parameter  $a$  does not occur in any assumptions on which  $\varphi$  depends

$$(\exists \text{I}) \quad \frac{\varphi_t^x}{\exists x \varphi} \quad (\exists \text{E}) \quad \frac{\varphi_a^x \quad \vdots \quad \theta}{\theta}$$

provided that the *eigen*-parameter  $a$  does not occur in any assumptions on which  $\theta$  depends (except  $\varphi_a^x$ ), nor in  $\exists x \varphi$  or  $\theta$ .

$$(= \text{I}) \quad t = t \quad (= \text{E}) \quad \frac{s = t \quad \varphi^x - s}{\varphi_t^x}$$

$$(\perp) \quad \frac{\perp}{\varphi}$$

The system thus given was introduced, essentially, by Gentzen [1934].

We write

$$\varphi_1, \dots, \varphi_k \vdash_N \varphi$$

if there is a derivation of  $\varphi$  where all open assumptions are among  $\varphi_1, \dots, \varphi_k$ , and similarly for sets  $\Gamma$ .

Above we gave a proof of (A & I), using the homonymous rule and the rules for  $\supset$ . Later we did the same for (A  $\vee$  E). The pattern thus presented is perfectly general and it should be clear to the reader how to prove every **HIQC**= axiom by using the corresponding Natural Deduction rule together with the implication rules. The Hilbert–Frege style axioms were chosen just because they are the linearisations of the Natural Deduction rules. When discussing the discharge of assumptions, we gave a proof of (A $\supset_1$ ). We now give a proof of (A $\supset_2$ ): in order to do this it is clearly enough to show that

$$\varphi \supset (\psi \supset \theta), (\varphi \supset \psi), \varphi \vdash_N \theta$$

because then a series of three ( $\supset$ I) will establish the axiom. This is readily done by means of the following derivation tree:

$$\frac{\frac{\varphi \supset (\psi \supset \theta) \quad \varphi}{\psi \supset \theta} \quad \frac{\varphi \supset \psi \quad \varphi}{\psi}}{\theta}$$

Here we have used a more relaxed style for setting out the derivation; the assumptions are not indexed and the names of the rules have not been indicated. ‘Officially’ this must be done, but in practice they are often left out.

We have seen that the Hilbert–Frege style axioms are all provable from our Natural Deduction rules. What about the rules of inference in the Hilbert–Frege style system? We consider (R $\forall$ I); then one is given a derivation  $D$  of  $\varphi \supset \psi$ , from assumptions  $\psi_1, \dots, \psi_k$  where  $a$  does not occur in any of the assumptions nor in  $\varphi$ , and is allowed to proceed to a derivation of  $\varphi \supset \forall x\psi_x^a$  from the same assumptions. Can this be effected using also Natural Deduction? One would expect the answer to be yes, and this is indeed found to be the case by means of the following derivation tree:

$$\frac{\frac{\frac{D}{\varphi \supset \psi} \quad \varphi^1}{\psi} (\supset E)}{\forall x\psi_x^a} (\forall I)}{\varphi \supset \forall x\psi_x^{a(1)}}$$

Here the use of ( $\forall$ I) is permitted because, by hypothesis, the *eigen*-parameter  $a$  does not occur in any of the assumptions of  $D$  nor in the assumption  $\varphi$ . Hence, the Natural Deduction system ‘is closed’ under the Hilbert–Frege style rule (R $\forall$ I). The other quantifier rule (R $\exists$ E) can be given an analogous treatment. We have therefore established the following

**THEOREM 3** ([Gentzen, 1934]). *If  $\Gamma \vdash_H \varphi$ , then  $\Gamma \vdash_N \varphi$ .*

For the other direction, the main work has already been carried out in the form of the Deduction Theorem. This theorem enables us to capture the effect of the discharge of assumptions within a Hilbert–Frege style system.

Consider the case of (& I). We must show that the system **HIQC**= is closed under this rule. To this effect assume that  $\varphi_1, \dots, \varphi_k \vdash_{\mathbf{H}} \varphi$  and  $\psi_1, \dots, \psi_m \vdash_{\mathbf{H}} \psi$ . We must show that  $\varphi_1, \dots, \varphi_k, \psi_1, \dots, \psi_m \vdash_{\mathbf{H}} \varphi \& \psi$ . But by use of (A& I) and MP we obtain first  $\varphi_1, \dots, \varphi_k \vdash_{\mathbf{H}} \psi \supset \varphi \& \psi$ ; and then by a use of the second given consequence relation and MP we reach our desired conclusion. The other rules are established in the same simple way by the use of the homonymous axiom and MP. The two eigen-parameter rules demand a separate treatment, however. We consider ( $\exists$ E). So let there be given **H**-derivations  $D$  and  $D_1$  which establishes that  $\Gamma \vdash_{\mathbf{H}} \exists x\varphi$  and  $\varphi^x - a, \delta \vdash_{\mathbf{H}} \theta$ . Assume further that the parameter  $a$  does not occur in  $\exists x\varphi, \theta$ , or  $\Delta$ . We have to show that  $\Gamma, \Delta \vdash_{\mathbf{H}} \theta$ , i.e. the desired conclusion of ( $\exists$ D) from the given premises. This is readily done, however:

$$\frac{\frac{\frac{\varphi_a^{x^1}, \Delta}{D_1} \theta}{\varphi_a^x \supset \theta^{(1)}} (\supset\text{I}) \quad (= \text{The Deduction Theorem})}{\frac{\Gamma \quad \varphi_a^x \supset \theta^{(1)}}{D \quad \exists x\varphi \supset \theta} (\text{R}\exists\text{E})} \frac{\exists x\varphi}{\theta} (\text{MP})$$

The use of rule (R $\exists$ E) is permitted, because by hypothesis the parameter  $a$  does not occur in  $\Delta$  (this justifies use of the Deduction Theorem as well), nor in  $\exists x\varphi$  or  $\theta$ . Hence, the Hilbert–Frege style system is closed under the quantifier rule ( $\exists$ E); the treatment of ( $\forall$ I) is similar. Thus, the second half of Gentzen’s

**THEOREM.** *If  $\Gamma \vdash_{\mathbf{N}} \varphi$ , then  $\Gamma \vdash_{\mathbf{H}} \varphi$  is established.*

The above system used intuitionistic logic; in order to obtain smooth systems for classical logic (whether, propositional or predicate, with or without identity) one can add any one of the following three rules:

- (i) The axiom *tertium non datur*:  $\varphi \vee \neg\varphi$  (TND)
- (ii) The rule of indirect proof:

$$\frac{\begin{array}{c} \neg\varphi \\ \vdots \\ \perp \end{array}}{\varphi} (\text{C}) \quad (\text{C for classical!})$$

1. The rule of Non-Constructive Dilemma:

$$\frac{\begin{array}{c} \varphi \quad \neg\varphi \\ \vdots \quad \vdots \\ \theta \quad \theta \end{array}}{\theta} \text{ (NCD)}$$

These all give classical logic when added to the intuitionistic rules for propositional calculus. We prove this by establishing a chain of implications:

If (ii), then (i). This shall be understood in such a way that given the schematic rule C, one can derive all instances of TND:

$$\frac{\frac{\frac{\neg(\varphi \vee \neg\varphi)^2}{\perp} \quad \frac{\frac{\varphi^1}{\varphi \vee \neg\varphi} \text{ (VI)}}{\varphi \vee \neg\varphi} \text{ (\supset E)}}{\perp}}{\frac{\frac{\neg(\varphi \vee \neg\varphi)^2}{\perp} \quad \frac{\frac{\neg\varphi^{(1)}}{\varphi \vee \neg\varphi} \text{ (VI)}}{\varphi \vee \neg\varphi^{(2)}} \text{ (C)}}{\perp} \text{ (C)}}$$

If (i), then (iii). This is immediate: given TND, the rule NCD simply becomes an instance of (VE).

If (iii), the (ii). Here we assume that we are given a derivation

$$\frac{\neg\varphi}{D} \perp$$

of  $\perp$  from the assumption  $\neg\varphi$ . Using only intuitionistic logic plus NCD we have to find a derivation of  $\varphi$ , *not depending on*  $\neg\varphi$ . first we note that

$$\frac{\frac{\frac{\neg\varphi^2}{D} \perp}{\varphi} \text{ (NCD)}}{\varphi} \text{ (1,2) (NCD)}$$

As  $\varphi$  is trivially derivable from the assumption  $\varphi$ , we have obtained a derivation of  $\varphi$  from the assumption  $\varphi$  and one of  $\varphi$  from the assumption  $\neg\varphi$ . The rule NCD gives us the right to discharge the assumptions  $\varphi$  and  $\neg\varphi$ .

With the given definition of negation in terms of the absurdity  $\perp$ , one is able to keep the idea that each symbol is governed by its introduction and elimination rules; the absurdity  $\perp$  has no *introduction* rules and only one elimination rule, viz. ( $\perp$ ). If one wishes to retain  $\neg$  as a primitive, the appropriate intuitionistic rules become:

( $\neg$ I)

$$\frac{\begin{array}{c} \varphi \quad \varphi \\ \vdots \quad \vdots \\ \psi \quad \neg\psi \end{array}}{\neg\varphi}$$

and

( $\neg$ E)

$$\frac{\varphi \quad \neg\varphi}{\theta}$$

In the classical case the elimination rule becomes:

( $\neg$ E<sub>c</sub>)

$$\frac{\neg\neg\varphi}{\varphi}$$

The ( $\neg$ I) and ( $\neg$ E) rules are less satisfactory than the other rules because the sign to be introduced occurs already in the premises of the rule.

The intuitionistic rules have a pleasing symmetry; each connective has its use regulated by two rules, one of which is, in a certain sense, the inverse of the other. To give this remark a bit more substance, consider a *maximum* formula, i.e. a formula occurrence in a derivation which is the conclusion of an I rule and the major premise of the corresponding E rule, e.g.  $\&$  in:

$$\frac{\varphi^1 \quad \psi^2}{\varphi \& \psi} \text{ (& I)}$$

$$\frac{\varphi \& \psi}{\varphi} \text{ (& E)}$$

This maximum can be removed, i.e. we can find a derivation of the same conclusion from (at most) the same assumptions, in a maximally simple way:  $\varphi^1$ .

This maximum-removing operation is called a *reduction* and similar reductions can also be given for other types of maxima, e.g. the  $\supset$  reduction is given by:

( $\supset$ I)

$$\frac{\begin{array}{c} \varphi^1 \\ \vdots \\ \psi \end{array}}{\varphi \supset \psi^{(1)}} \quad \begin{array}{c} \vdots \\ \vdots \\ \varphi \end{array} \quad \begin{array}{c} \vdots \\ \vdots \\ \psi \end{array}$$

$$\frac{\varphi \supset \psi^{(1)} \quad \varphi}{\psi} \text{ ( $\supset$  E) reduces to } \psi$$

and the  $\forall$ -reduction is given by:

$$\frac{\frac{\vdots}{\varphi} (\forall I)}{\forall x \varphi_x^a} (\forall I) \quad \frac{\vdots}{\varphi_t^a} \quad \begin{array}{l} a \\ \vdots \\ t \end{array} \quad \begin{array}{l} \text{(i.e. everywhere in the derivation of } \varphi \\ \text{replace } a \text{ with } t.) \end{array}$$

$$\frac{\frac{\vdots}{\varphi} (\forall I)}{(\varphi_x^a)_t} (\forall E) \text{ reduces to } \frac{\vdots}{\varphi_t^a}$$

For the other reductions, cf. [Prawitz, 1965, pp. 36–38] (where they were first introduced and [Prawitz, 1971, pp. 252–253].

Although each individual maximum can be removed via a reduction, the situation is not altogether as simple as one would wish because *the removal of one maximum can create a new maximum*, e.g. in:

$$\frac{\frac{\frac{\frac{\vdots}{\varphi^1}}{\vdots}}{\psi} (\supset I)}{\varphi \supset \psi^{(1)}} (\supset I) \quad \frac{\vdots}{\theta} \quad \frac{\frac{\frac{\frac{\vdots}{\varphi^1}}{\vdots}}{\psi} (\supset I)}{\varphi \supset \psi^{(1)}} (\supset I)}{\frac{(\varphi \supset \psi) \& \theta}{\varphi \supset \psi} (\& E)} (\& I) \quad \frac{\vdots}{\varphi}$$

$$\frac{\frac{(\varphi \supset \psi) \& \theta}{\varphi \supset \psi} (\& E)}{\psi} (\& E)$$

Here  $(\varphi \supset \psi) \& \theta$  is a maximum and an  $\&$  reduction gives:

$$\frac{\frac{\frac{\frac{\vdots}{\varphi^1}}{\vdots}}{\psi} (\supset I)}{\varphi \supset \psi^{(1)}} (\supset I) \quad \frac{\vdots}{\varphi} (\supset E)}{\psi} (\supset E)$$

Now the wff  $\varphi \supset \psi$ , which previously was not a maximum, has been turned into a maximum. A  $\supset$  reduction will, of course, suffice to remove *this* new maximum, but what about the general case? If we consider the complexities of the old and the new maximum, respectively, we find that the new one has a lower complexity than the old one, and this phenomenon holds in general. Hence, one has something to use induction on in a proof that every maximum can be removed by successive reductions (even though new ones may arise along the way). In fact Prawitz has proved the

**NORMALISATION THEOREM.** *Every derivation can be brought to maximum-free, or normal, form by means of successive reductions.*



For a proof and precise statements of this and related results, we refer to the works by Prawitz and Tennant just cited. Dummett [2000] also contains an exhaustive discussion of normalisation.

The reduction and the normalisation procedures have formed a basis for various attempts to give a ‘theory of meaning’ not using the Tarski truth definition as a key concept. Cf. [Prawitz, 1977] and this Sundholm’s chapter III.8 in this handbook for a description of this use of Natural Deduction and its metatheoretical properties.

Natural Deduction formulations can be given not only for pure propositional and predicate logic, as above, but also for, say, modal logic. Define:

- (0)  $\perp$  is essentially modal,
- (i)  $\Box\varphi$  is essentially modal,
- (ii) If  $\varphi$  and  $\psi$  are both essentially modal, then so are  $\varphi \& \psi$  and  $\varphi \vee \psi$ .

Using this concept of an essentially modal formula, one then gives an attractive version of classical **S4** by adding the following two rules to **CPC**

( $\Box$ E)  $\frac{\Box\varphi}{\varphi}$   
 (This rule is, of course, a rule version of the **T** axiom:  $\Box\varphi \supset \varphi$ ), and

( $\Box$ I)  $\frac{\varphi}{\Box\varphi}$   
*provided* that all assumptions on which  $\varphi$  depends are essentially modal.

This formulation of **S4** is given in [Prawitz, 1965, Chapter VI] We derive the **S4** axioms. for these axioms, cf. Bull and Segerberg’s chapter in Volume 3 of this *Handbook*.

**k:** We have to prove  $\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)$

( $\Box$ E)

$$\frac{\frac{\frac{\Box(\varphi \supset \psi)^1}{\varphi \supset \psi} \quad \frac{\Box\varphi^2}{\varphi} (\Box E)}{\varphi \supset \psi} (\supset E) \quad \frac{\psi}{\Box\psi} (\Box I)}{\Box\varphi \supset \psi^{(2)} (\supset I)} (\supset I)$$

$$\frac{\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)^{(1)}}{\Box(\varphi \supset \psi) \supset (\Box\varphi \supset \Box\psi)^{(1)}}$$

(All assumptions are essentially modal.)

**T:** We have to prove  $\Box\varphi \supset \varphi$ .

$$\frac{\frac{\Box\varphi^1}{\varphi} (\Box E)}{\Box\varphi \supset \varphi^{(1)}} (\supset I)$$

4. We have to prove  $\Box\varphi \supset \Box\Box\varphi$ .

$$\frac{\frac{\Box\varphi^1}{\Box\Box\varphi} (\Box E)}{\Box\varphi \supset \Box\Box\varphi^{(1)}} (\supset I)$$

The rule *Nec* is just a special case of ( $\Box I$ ), with the class of assumptions empty.

Thus we have proved that **NS4** includes the previously given **HS4**. For the other direction one uses the two readily verified facts:

- (i) If  $\varphi_1, \dots, \varphi_k \vdash_{\mathbf{HS4}} \varphi$ , then  $\Box\varphi_1, \dots, \Box\varphi_k \vdash_{\mathbf{HS4}} \Box\varphi$ .
- (ii) For every essentially modal  $\varphi$ ,  $\vdash_{\mathbf{HS4}} \varphi \supset \Box\varphi$ .

The details are left as an exercise. Finally we remark that if we change (ii) to ii') in the definition of essentially modal wffs, we obtain a formulation of **S5**, where (ii') says: If  $\varphi$  and  $\psi$  are essentially modal, then so are  $\varphi \& \psi$ ,  $\varphi \vee \psi$  and  $\varphi \supset \psi$ .

For another, final, example of a Natural Deduction version of a wider system, we consider a smooth version of Heyting Arithmetic, **HA**, cf. Van Dalen's chapter in Volume 7 of this *Handbook*. We use a new predicate  $N(x)$ —'x is a natural number', a constant 0 and a 'successor function'  $s(x)$  (plus some further function constants which we need not bother about now). The introduction rule for  $N$  gives an inductive definition of the natural numbers:

(NI):

$$N(0) \frac{N(a)}{N(s(a))} (NI)$$

By this rule, every natural numeral  $0, s(0), s(s(0)), \dots$  is in  $N$ . The elimination rule, on the other hand, says that nothing else is in  $N$ :

(NE):

$$\frac{N(t) \quad \varphi_0^x \quad \begin{array}{c} \varphi_a^{x^1} \\ \vdots \\ \varphi_{s(a)}^x \end{array}}{\varphi_t^x} \quad (1)(NE) \quad \textit{eigen-parameter condition on } a$$

This version of the induction axiom thus says that if  $t$  is in  $N$  and the property  $\varphi$  is closed under successor and contains 0, then  $t$  has the property  $\varphi$ . This form of spelling out inductive definitions can be applied not just to the inductively-generated natural numbers, but to a wide range of inductive definitions, cf. [Martin-Löf, 1971].

On the positive side of natural Deduction formulations there are, as we have seen, several advantages, the foremost of which is the great ease with which derivations can actually be *carried out*. On the minus score however, one must note that Natural Deductions is not suitable for all sorts of systems. In several modal systems, say between **S4** and **S5**, it is not at all clear that one can isolate the behaviour of  $\Box$  in the form of introduction and elimination rules. It was already noted that the conventions on assumptions made it difficult to find suitable systems for relevance logic. For some efforts in this area, cf. [Prawitz, 1965, Chapter VII].

The arithmetization of a Natural Deduction formulation is, *prima facie* also a little more cumbersome than in the case of an Hilbert–Frege style formulation. The Gödel number of a derivation has to contain the following information: (i) the end formula, (ii) the rule used at the last inference, (iii) Gödel numbers for derivation of the premises, (iv) the assumption class possibly discharged at the inference, (v) the still open assumption classes. Such Gödel numbers are best constructed as sequence numbers containing the above items, cf. [Kleene, 1952, Section 51– 52] and [Troelstra, 1973, Chapter IV]. (The latter reference contains much valuable material on various aspects on the derivability relation (for intuitionistic predicate logic):

- (0)  $\varphi \vdash \varphi$ —to make an assumption.
- (i) If  $\Gamma \vdash \varphi$ , then  $\Gamma, \Delta \vdash \varphi$ —we may add further assumptions.
- (ii) If  $\Gamma \vdash \varphi$  and  $\Delta \vdash \psi$ , then  $\Gamma, \Delta \vdash \varphi \& \psi$ —by (& I).
- (iii) If  $\Gamma \vdash \varphi_0 \& \psi_1$ , then  $\Gamma \vdash \varphi_i$ —by (& E) for  $i = 0, 1$ .
- (iv) If  $\varphi, \Gamma \vdash \psi$ , then  $\Gamma \vdash \varphi \supset \psi$ —by ( $\supset$ I).
- (v) If  $\Gamma \vdash \varphi \supset \psi$  and  $\Delta \vdash \varphi$ , then  $\Gamma, \Delta \vdash \psi$ —by ( $\supset$ E).
- (vi)–(vii) The corresponding clauses for (VI) and (VE) are left as exercises.

- (viii) If  $\Gamma \vdash \varphi$  and  $a$  does not occur in  $\Gamma$ , then  $\Gamma \vdash \forall x\varphi_x^a$ —by ( $\forall$ I).
- (ix) If  $\Gamma \vdash \forall x\varphi$ , then  $\Gamma \vdash \varphi_t^x$ —by ( $\forall$ E).
- (x) If  $\Gamma \vdash \varphi_t^x$ , then  $\Gamma \vdash \exists x\varphi$ —by ( $\exists$ I).
- (xi) If  $\Gamma \vdash \exists x\varphi$  and  $\varphi_a^x, \Delta \vdash \theta$  and  $a$  does not occur in any of  $\Delta, \theta$  and  $\exists x\varphi$ , then  $\Gamma, \Delta \vdash \theta$ —by ( $\exists$ E).
- (xii) If  $\Gamma \vdash \perp$ , then  $\Gamma \vdash \varphi$ —by ( $\perp$ ).

Define a *sequent* to be an expression of the form ' $\Gamma \vdash \varphi$ ' where  $\Gamma$  is a finite set of wffs—possibly empty—and  $\varphi$  a wff. Then the above clauses (0)–(xii) give the axioms and rules of a certain *Hilbert–Frege style system for deriving sequents*; the clause (0) gives the only axiom, which says that the sequent  $\varphi \vdash \varphi$  is derivable. (Here I adhere to the same connections about notation as before; strictly speaking ' $\{\varphi\} \vdash \varphi$ ' is the axiom.) The other clauses are *rules of proof*. The present system is just a notational variant of the tree arrangement version of Natural Deduction which we have used in the present section; in particular it is not to be confused with Gentzen's Sequent Calculi. The main feature of the sequential formulation of Natural Deduction is that there are both elimination rules and introduction rules. In the Sequent Calculus, on the other hand, there are *only introduction rules*. The sequential version was first given by Gentzen [1936] in his 'first consistency proof'. There is a certain formal difference between Gentzen's system and ours, because for him the sequent part  $\Gamma$  is not a set of wffs but of a *list* of wffs, and he therefore has rules to the effect that one may permute assumption formulae in the lists and contract two identical assumption formulae to one. The latter rule neglects the problem of assumption classes. The best way to treat that in the sequent version seems to be to let  $\Gamma$  be a set of *labelled* formulae, where a labelled formula is an order pair  $\langle \varphi, k \rangle$  where  $k$  is a natural number (and  $\varphi$  is a wff, of course). The details can be found in [Leivant, 1979].

One has several options as to the choice of rules for a sequential version—for an indication of the various possibilities, cf. [Dummett, 2000, Chapter IV] who treats the sequential formulations in considerable detail—and in particular the '*thinning*' rule (i) allows us to restrict the multi-premise rules to the special cases of always using the same set of assumptions. Thus, one can formulate the rule for conjunction- introduction as follows:

$$\frac{\Gamma \vdash \varphi \quad \Gamma \vdash \psi}{\Gamma \vdash \varphi \& \psi}$$

The effect of the old rule (ii) can be simulated. From  $\Gamma \vdash \varphi$ , by thinning, one obtains  $\Gamma, \Delta \vdash \varphi$  and from  $\Delta \vdash \psi$  one obtains  $\Gamma, \Delta \vdash \psi$ , also by thinning.

Finally, the (new) conjunction introduction gives the desired conclusion:  
 $\gamma, \Delta \vdash \varphi \& \psi$ .

Another variation is to formulate the axioms in a more general way:  
 $\Gamma \vdash \varphi$  is an axiom where  $\varphi$  is an element of  $\Gamma$ , or equivalently:  $\varphi, \Gamma \vdash \varphi$  is an axiom.

If we carry out both of these modifications the resulting calculus for deriving sequents will look like:

$$\begin{array}{l}
\text{Axiom} \quad \varphi, \Gamma \vdash \varphi \\
\text{Thinning} \quad \frac{\Gamma \vdash \varphi}{\Gamma, \Delta \vdash \varphi} \\
(\& \text{I}) \quad \frac{\Gamma \vdash \varphi \quad \Gamma \vdash \psi}{\Gamma \vdash \varphi \& \psi} \\
(\& \text{E}) \quad \frac{\Gamma \vdash \varphi_0 \& \varphi_1}{\Gamma \vdash \varphi_i} \quad i = 0, 1 \\
(\vee \text{I}) \quad \frac{\Gamma \vdash \varphi_i}{\Gamma \vdash \varphi_0 \vee \varphi_1} \quad i = 0, 1 \\
(\vee \text{E}) \quad \frac{\Gamma \vdash \varphi \vee \psi \quad \varphi, \Gamma \vdash \theta \quad \psi, \Gamma \vdash \theta}{\Gamma \vdash \theta} \\
(\supset \text{I}) \quad \frac{\varphi, \Gamma \vdash \psi}{\Gamma \vdash \varphi \supset \psi} \\
(\supset \text{E}) \quad \frac{\Gamma \vdash \varphi \supset \psi \quad \Gamma \vdash \varphi}{\Gamma \vdash \psi} \\
(\forall \text{I}) \quad \frac{\Gamma \vdash \varphi}{\Gamma \vdash \forall x \varphi_x^a} \\
(\forall \text{E}) \quad \frac{\Gamma \vdash \forall x \varphi}{\Gamma \vdash \varphi_t^x} \\
(\exists \text{I}) \quad \frac{\Gamma \vdash \varphi_t^x}{\Gamma \vdash \exists x \varphi} \\
(\exists \text{E}) \quad \frac{\Gamma \vdash \exists x \varphi \quad \varphi_a^x, \Gamma \vdash \theta}{\Gamma \vdash \theta}
\end{array}$$

<sup>1</sup>On these rules there is an eigen-parameter condition:  $a$  must not occur in  $\Gamma$ ,  $\exists x \varphi$  or  $\theta$ .

$$(\perp) \quad \frac{\Gamma \vdash \perp}{\Gamma \vdash \varphi}$$

We conclude our treatment of Natural Deduction by mentioning [Tennant, 1978] which treats the traditional metamathematics of first-order logic and arithmetic on the basis of Natural Deduction formulations of the systems concerned, and [van Dalen, 1997], which treats model theory in a similar way. Gentzen's treatment of Natural Deduction, which still seems the most natural to me, inspired many other attempts to keep the derivations from assumptions and the discharge of certain of these as a basic feature. In particular, many authors have experimented with *linear* arrangements of the derivations as opposed to Gentzen's tree formulations. A list of variant treatments of Natural Deduction is found in Hodges (Volume 1, Section 7).

#### 4 SEQUENT CALCULI

The (reformulated) sequential version of Natural Deduction, which we considered at the end of Section 2, was nothing but a direct linearisation of the introduction and elimination rules, where the latter were thought of as imposing conditions on the derivability relation  $\vdash$ . The pattern thus obtained is, however, not the only way of axiomatising the derivability relation on the basis of the Natural Deduction rules. Crucial in the above treatment was that the elimination rules were formulated as such and operated on the right-hand side of the turnstile  $\vdash$ . A completely different approach to the elimination rules is possible, though, and leads to Gentzen's [1934] *Sequent Calculi*.

Consider a derivable sequent

$$\varphi, \varphi_1, \dots, \varphi_k \vdash \theta;$$

hence if we read the derivation in the sequential system as a description of how to build a derivation tree for  $\theta$  from assumptions  $\varphi, \varphi_1, \dots, \varphi_k$ , we can find a derivation tree  $D$  such that

$$\frac{\varphi^1, \varphi_1, \dots, \varphi_k}{D} \theta$$

where, in particular,  $\varphi$  is an undischarged assumption. Therefore, using (one of) (& E) we obtain a derivation tree  $D'$  of  $\theta$  from assumptions  $\varphi_1, \dots, \varphi_k$  and  $\varphi \& \psi$ . In this tree  $D'$  the assumption  $\varphi$  has 'disappeared'.

$$D' =_{\text{def}} \quad (\&E) \quad \frac{\varphi \& \psi^1}{\varphi}, \quad \varphi_1, \dots, \varphi_k \quad \begin{array}{c} D \\ 0 \end{array}$$

So the sequent  $\varphi \& \psi, \varphi_1, \dots, \varphi_k \vdash \theta$  is derivable, and on the given interpretation the rule

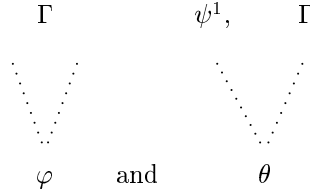
$$\frac{\varphi, \Gamma \vdash \theta}{\varphi \& \psi, \Gamma \vdash \theta}$$

is a sound *rule of proof*.

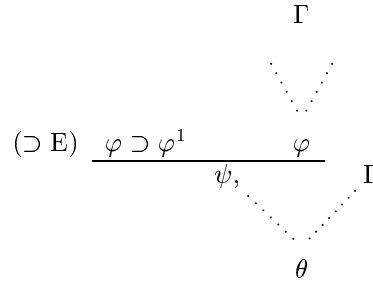
Likewise, the rule

$$\frac{\Gamma \vdash \varphi \quad \psi, \Gamma \vdash \theta}{\varphi \supset \psi, \Gamma \vdash \theta}$$

is a sound rule of proof, because given derivation trees



one readily finds a derivation tree in which the place of the Assumption  $\psi$  is taken by  $\varphi \supset \psi$ :



The use of ( $\vee E$ ) immediately justifies the step:

$$\frac{\varphi, \gamma \vdash \theta \quad \psi, \Gamma \vdash \theta}{\varphi \vee \psi, \Gamma \vdash \theta}$$

and ( $\forall E$ ) justifies:

$$\frac{\varphi_t^x, \Gamma \vdash \theta}{\exists x \varphi, \Gamma \vdash \theta}$$

Finally, if  $a$  does not occur in the conclusion, then trivially ( $\exists E$ ) justifies the step:

$$\frac{\varphi_t^x, \Gamma \vdash \theta}{\forall x \varphi, \Gamma \vdash \theta}$$

From the present point of view, the elimination rules are viewed as *left-hand side introduction rules*. Define a *sequent* to be an expression of the form

$$\Gamma \rightarrow \theta$$

where  $\Gamma$  is a finite set of wffs. This (re-) definition is only for historical reasons; by custom, one uses an arrow in the sequents.

The axioms and rules of Gentzen's Sequent Calculus are given by:

$$\text{(Axiom)} \quad \varphi, \Gamma \rightarrow \varphi$$

$$\text{(Thinning)} \quad \frac{\Gamma \vdash \varphi}{\Delta, \Gamma \vdash \varphi}$$

$$(\rightarrow \&) \quad \frac{\Gamma \rightarrow \varphi \quad \Gamma \rightarrow \psi}{\Gamma \rightarrow \varphi \& \psi}$$

$$(\& \rightarrow) \quad \frac{\varphi, \Gamma \rightarrow \theta \quad \psi, \Gamma \rightarrow \theta}{\varphi \& \psi, \Gamma \rightarrow \theta} \quad \frac{\psi, \Gamma \rightarrow \theta}{\varphi \& \psi, \Gamma \rightarrow \theta}$$

$$(\rightarrow \vee) \quad \frac{\Gamma \rightarrow \varphi}{\Gamma \rightarrow \varphi \vee \psi} \quad \frac{\Gamma \rightarrow \psi}{\Gamma \rightarrow \varphi \vee \psi}$$

$$(\vee \rightarrow) \quad \frac{\varphi, \gamma t \theta \quad \psi, \Gamma \rightarrow \theta}{\varpi \vee \psi, \Gamma \rightarrow \theta}$$

$$(\rightarrow \supset) \quad \frac{\varphi, \Gamma \rightarrow \psi}{\Gamma \rightarrow \varphi \supset \psi}$$

$$(\supset \rightarrow) \quad \frac{\Gamma \rightarrow \varphi \quad \psi, \Gamma \rightarrow \theta}{\varphi \supset \psi, \Gamma \rightarrow \theta}$$

$$(\rightarrow \forall) \quad \frac{\Gamma \rightarrow \varphi}{\Gamma \rightarrow \forall x \varphi_x^a} \quad \text{provided that } a \text{ does not occur in } \Gamma$$

$$(\forall \rightarrow) \quad \frac{\varphi_t^x, \Gamma \rightarrow \theta}{\forall x \varphi, \Gamma \rightarrow \theta}$$

$$(\rightarrow \exists) \quad \frac{\Gamma \rightarrow \varphi_t^x}{\Gamma \rightarrow \exists x \varphi}$$

$$(\exists \rightarrow) \quad \frac{\varphi_a^x, \Gamma \rightarrow \theta}{\exists x \varphi, \Gamma \rightarrow \theta} \quad \text{provided that } a \text{ does not occur in the conclusion}$$

$$(\perp) \quad \frac{\Gamma \rightarrow \perp}{\Gamma \rightarrow \theta}$$



A sequent  $\Gamma \rightarrow \varphi$  is provable in this ‘cut-free’ intuitionistic sequent calculus if it has got a proof tree regulated by the above rules and where all the top nodes are axioms. In symbols we write:

$$\vdash_{\mathbf{IS}} \Gamma \rightarrow \varphi$$

(**IS** for intuitionistic sequent calculus.)

The introductory discussion implicitly proved the following inclusion:

If  $\vdash_{\mathbf{IS}} \Gamma \rightarrow \varphi$ , then the sequent  $\Gamma \vdash \varphi$  is derivable in the sequential formulation of Natural Deduction.

A formal proof by induction on the length of the **IS** proof of  $\Gamma \rightarrow \varphi$  can safely be left to the patient reader.

Of course, one would like to establish the converse relation between the two axiomatizations of the Natural Deduction turnstile, i.e. that the idea of treating the elimination rules as left-hand side *introduction* is as strong as the original formulations where the eliminations operate to the right. This can, in fact, be done but a direct proof would be quite unwieldy. Instead, Gentzen introduced a rule:

$$(\text{CUT}) \frac{\Gamma \rightarrow \varphi \quad \varphi, \Gamma \rightarrow \theta}{\Gamma \rightarrow \theta}$$

such that in  $\mathbf{IS}^+ = \mathbf{IS} + \text{CUT}$  one readily shows that the effect of the full Natural Deduction eliminations can be simulated. Then, by means of his famous *Hauptsatz*, he established that if  $\vdash_{\mathbf{IS}^+} \Gamma \rightarrow \varphi$ , then  $\vdash_{\mathbf{IS}} \Gamma \rightarrow \varphi$ , and the equivalence is established.

We treat the case of ( $\& \text{E}$ ). Hence we are given an  $\mathbf{IS}^+$  proof of  $\Gamma \rightarrow \varphi \& \psi$  and we wish to find one for  $\Gamma \rightarrow \psi$ . This we do as follows:

$$\frac{\begin{array}{c} \vdots \\ \vdots \\ \Gamma \rightarrow \varphi \& \psi \end{array} \quad \frac{\psi, \Gamma \rightarrow \psi}{\varphi \& \psi, \Gamma \rightarrow \psi} \text{ (Axiom)} \quad (\& \rightarrow)}{\Gamma \rightarrow \psi} \text{ (CUT)}$$

The same technique using the CUT rule works uniformly for all the Natural Deduction elimination rules. To treat one more case, we choose the most complex, viz. ( $\supset \text{E}$ ): Here we are given  $\mathbf{IS}^+$  proofs of  $\Gamma \rightarrow \varphi$  and  $\Gamma \rightarrow \varphi \supset \psi$ , and have to find one for  $\Gamma \rightarrow \psi$ . This is done as follows:

$$\frac{\begin{array}{c} \vdots \\ \vdots \\ \Gamma \rightarrow \varphi \supset \psi \end{array} \quad \frac{\begin{array}{c} \vdots \\ \vdots \\ \Gamma \rightarrow \varphi \end{array} \quad \psi, \Gamma \rightarrow \psi}{\varphi \supset \psi, \Gamma \rightarrow \psi} \text{ (Axiom)} \quad (\supset \rightarrow)}{\Gamma \rightarrow \psi} \text{ (CUT)}$$

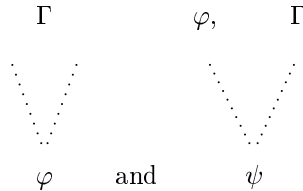
The same method proves the other cases and, hence, it is established that

If  $\Gamma \vdash \varphi$  is a provable Natural Deduction sequent, then  $\vdash_{\mathbf{IS}+} \Gamma \rightarrow \varphi$ .

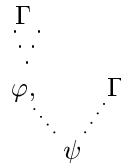
In order to get the inclusion in **IS** we have to discuss Gentzen's Hauptsatz that CUT can be eliminated. What is the significance of CUT for the Natural Deduction derivations? Here it licenses the step from provable sequents  $\Gamma \vdash \varphi$  and  $\varphi, \Gamma \vdash \psi$  to  $\Gamma \vdash \psi$ . It is simply a derivable rule of proof by means of ( $\supset$ I) and ( $\supset$ E), viz:

$$\begin{array}{c} \vdots \\ \varphi, \Gamma \vdash \psi \\ \hline \text{(\supset I)} \quad \Gamma \vdash \varphi \supset \psi \end{array} \quad \begin{array}{c} \vdots \\ \Gamma \vdash \varphi \\ \hline \text{(\supset E)} \quad \Gamma \vdash \psi \end{array}$$

Here we created a new maximum, viz.  $\varphi \supset \psi$  which is the conclusion of an introduction and the major premise of an elimination. Another way to view CUT in Natural Deduction contexts is that it expresses the closure under substituting a derivation of an assumption for that assumption, e.g. consider:



The result of putting the first derivation on top of the other looks like:



Here the *assumption*  $\varphi$  has disappeared, but a new *maximum* may have arisen, viz.  $\varphi$ . If the last rule of the derivation of  $\varphi$  from  $\Gamma$  is an introduction—say that  $\varphi = \varphi_0 \& \varphi_1$ , inferred by an ( $\&$  I)—and the first rule applied to  $\varphi$  in the derivation of  $\psi$  from assumptions  $\varphi$  and  $\Gamma$  is an elimination—in this case, then, an ( $\&$  E)—then  $\varphi$  has been turned into a maximum, although the derivations are closed under the rule of CUT. Hence, cut-free derivations in the Sequent Calculus correspond to normal derivations in the Natural Deduction systems, and if we may assume that the Natural Deduction derivation we have of a sequent is *normal*, one can directly find a cut-free **IS** proof of the corresponding sequent, cf. [Prawitz, 1965, Appendix

A, pp. 88–93]. Hence, one needs to prove the Normalisation Theorem, or the Hauptsatz, to have an easy proof of the inclusion of the Natural Deduction in the cut-free system **IS**. In view of the above discussion, this is hardly surprising, since there is a correspondence between cuts and maxima.

Gentzen’s proof of his Hauptsatz consists of a series of permutations of cuts in such a fashion that the complexity of the ‘cut formula’ is lowered. We can illustrate this by considering the simplest of the cases:

$$\begin{array}{c}
 \begin{array}{ccc}
 \vdots & & \vdots \\
 \Gamma \rightarrow \varphi & & \Gamma \rightarrow \psi \\
 \hline
 \Gamma \rightarrow \varphi \&\psi & & \varphi, \Gamma \rightarrow \theta \\
 \hline
 \Gamma \rightarrow \varphi \&\psi & & \varphi \&\psi, \Gamma \rightarrow \theta \\
 \hline
 & & \Gamma \rightarrow \theta
 \end{array}
 \end{array}
 \begin{array}{l}
 (\sup \&) \\
 \\
 \\
 (\& \rightarrow) \\
 (\text{CUT})
 \end{array}$$

Here we replace the cut on the wff  $\varphi \& \psi$  by a cut on the wff  $\varphi$  in this way:

$$\begin{array}{c}
 \begin{array}{ccc}
 \vdots & & \vdots \\
 \Gamma \rightarrow \varphi & & \varphi, \Gamma \rightarrow \theta \\
 \hline
 \Gamma \rightarrow \theta
 \end{array}
 \end{array}
 \quad (\text{CUT})$$

The situation is not always as simple as this but it gives the idea. The proof of the Hauptsatz can be found in many places. Apart from Gentzen’s own [1934], [Kleene, 1952; Takeuti, 1975] as well as [Dummett, 2000], contain good expositions of the proof.

Gentzen’s own Sequent Calculi used finite *lists* of formulae (in place of finite sets) and, as already remarked apropos his formulation of the sequential Natural Deduction system, he needs further ‘structural’ rules to permute and contract formulae in the finite lists. The presence of a contraction rule:

$$\frac{\varphi, \varphi, \Gamma \rightarrow \theta}{\varphi, \Gamma \rightarrow \theta}$$

again raises the question of assumption classes and mandatory discharge of assumptions (via the correspondence with Natural Deduction). The interested reader is referred to [Zucker, 1974] and [Pottinger, 1979] for a thorough treatment of such matters.

The Sequent Calculus just presented was set up so as to make the transition from Natural Deduction particularly easy and, hence, it was expedient to keep  $\perp$  as a primitive (with negation defined). In the Sequent Calculus, however, another more congenial treatment consists of dropping  $\perp$  and adding  $\neg$  as a primitive. The concept of sequent is widened to include sequents ‘ $\Gamma \rightarrow \Delta$ ’, where both the ‘*antecedent*’  $\Gamma$  and the ‘*succedent*’  $\Delta$  are *finite* sets of formulae, with the important restriction that  $\Delta$  *has got at most one element*. The sequent ‘ $\Gamma \rightarrow \{\varphi\}$ ’ will be written ‘ $\Gamma \rightarrow \varphi$ ’ and a sequent  $\Gamma \rightarrow \phi$ , with empty succedent, is written ‘ $\gamma \rightarrow$ ’.

In the calculus one drops the ( $\perp$ ) rule, of course, but in order to secure its effect, one also adds a *Thinning* for the succedent:

$$\frac{\Gamma \rightarrow}{\Gamma \rightarrow \theta}$$

This makes the empty succedent behave like absurdity, and the following two rules for negation suggest themselves:

$$(\rightarrow \neg) \frac{\varphi, \Gamma \rightarrow}{\Gamma \rightarrow \neg \varphi}$$

and

$$(\neg \rightarrow) \frac{\Gamma \rightarrow \varphi}{\neg \varphi, \Gamma \rightarrow}$$

The system thus modified we call **IS'**.

The underlying idea for our way of looking at the Sequent Calculus, up till this point, has been to regard the Sequent Calculus as a sort of 'meta-descriptive' system for how Natural Deduction derivations are put together. There is also another approach to the sequent apart from this consequence relation interpretation, which other approach is particularly well suited to effect the transition to the fully symmetric, classical Sequent Calculus.

On a naive semantical level, we may call a sequent  $\Gamma \rightarrow \Delta$  *valid*, if whenever all the members of  $\Gamma$  are true, then at least some member of  $\Delta$  is true. (Hence, a sequent  $\Gamma \rightarrow$  is valid iff  $\Gamma$  is inconsistent.) The rules of **IS'** then give rules of passage between valid sequents; in particular, every axiom is valid. On the other hand, let us call a sequent *falsifiable* if one can make all the members of  $\Gamma$  true and all the members of  $\Delta$  false. We note that these explanations of validity and falsifiability also work for sequents with more than one element in the succedent.

Now the other approach suggests itself; instead of reading the rules of **IS** as expressing validity conditions of the form, say 'if  $S_1$  and  $S_2$  are both valid sequents, then so is  $S$ ' we may read them as *expressing falsifiability conditions* of the form, say, 'if the sequent  $s$  is falsifiable, then  $S_1$  is falsifiable or  $S_2$  is falsifiable'. Let us work this out for the conjunction rules: Assume that the sequent ' $\Gamma \rightarrow \varphi \& \psi$ ' is falsifiable. Then one can make all of  $\Gamma$  true and yet make  $\varphi \& \psi$  false. Hence, one can make at least one of  $\varphi$  and  $\psi$  false, and therefore at least one of the premises  $\Gamma \rightarrow \varphi$  and  $\Gamma \rightarrow \psi$  must be a falsifiable sequent. Likewise, for the antecedent conjunction rule: Assume that  $\varphi \& \psi, \Gamma \rightarrow \Delta$  is a falsifiable sequent. Then one can make all of  $\varphi \& \psi, \Gamma$  true and  $\Delta$  false. But then both of  $\varphi, \Gamma \rightarrow \Delta$  and  $\psi, \Gamma \rightarrow \Delta$  are falsifiable sequents, so ( $\& \rightarrow$ ) expresses two falsifiability conditions.

The same sort of reasoning applies to the other rules of **IS'**. It is worthwhile considering the negation rules a bit more carefully. Take ( $\neg \rightarrow$ ) first.

If  $\neg\varphi, \Gamma \rightarrow$  is falsifiable, this means one can make all of  $\neg\varphi\Gamma$  true, so one can make  $\varphi$  false. Hence,  $\Gamma \rightarrow \varphi$  is falsifiable. As for  $(\rightarrow \neg)$ , assume that  $\Gamma \rightarrow \neg\varphi$  is falsifiable. Then all of  $\Gamma$  can be made true while  $\neg\varphi$  is false. Hence, all of  $\varphi, \Gamma$  can be made true.

Both of these rules thus express good falsifiability conditions, but they are too ‘weak’ to make use of full classical reasoning. The same argument would also justify the rules with *extra* members in the succedents:

$$\frac{\varphi, \Gamma \rightarrow \Delta}{\Gamma \rightarrow \Delta, \neg\varphi} \quad \frac{\Gamma \rightarrow \Delta, \varphi}{\neg\varphi, \Gamma \rightarrow \Delta}$$

The calculus which results is particularly nice because the rules express necessary and *sufficient* conditions for the falsifiability of the conclusion, i.e. in order that the conclusion be falsifiable it is necessary and sufficient that at least one of the premises be falsifiable. The axioms indicate unfalsifiability, because it is impossible to give  $\varphi$  both the value true *and* the value false. In general, we see that the complexity of the premises is lower—the quantifier rules form exceptions—than the complexity of the conclusion and, hence, the falsifiability condition for the conclusion is—in general—expressed in terms of falsifiability conditions of lower complexity. This leads to a practical way of systematically searching for a proof in the Sequent Calculus.

*Axiom*  $\varphi, \gamma \rightarrow \delta, \varphi$

$$(\rightarrow \&) \quad \frac{\Gamma \rightarrow \Delta, \varphi \quad \Gamma \rightarrow \Delta, \psi}{\Gamma \rightarrow \Delta, \varphi \& \psi}$$

$$(\& \rightarrow) \quad \frac{\varphi, \psi, \Gamma \rightarrow \Delta}{\varphi \& \psi, \Gamma \rightarrow \Delta}$$

$$(\rightarrow \vee) \quad \frac{\Gamma \rightarrow \Delta, \varphi, \psi}{\Gamma \rightarrow \Delta, \varphi \vee \psi}$$

$$(\vee \rightarrow) \quad \frac{\varphi, \Gamma \rightarrow \Delta \quad \psi, \Gamma \rightarrow \Delta}{\varphi \vee \psi, \Gamma \rightarrow \Delta}$$

$$(\rightarrow \supset) \quad \frac{\varphi, \Gamma \rightarrow \Delta, \psi}{\Gamma \rightarrow \Delta, \varphi \supset \psi}$$

$$(\supset \rightarrow) \quad \frac{\Gamma \rightarrow \Delta, \varphi \quad \psi, \Gamma \rightarrow \Delta}{\varphi \supset \psi, \Gamma \rightarrow \Delta}$$

$$(\rightarrow \neg) \quad \frac{\varphi, \Gamma \rightarrow \Delta}{\Gamma \rightarrow \Delta, \neg\varphi}$$

$$\begin{array}{l}
(\neg \rightarrow) \quad \frac{\Gamma \rightarrow \Delta, \varphi}{\neg \varphi, \Gamma \rightarrow \Delta} \\
(\rightarrow \forall) \quad \frac{\Gamma \rightarrow \Delta, \varphi}{\Gamma \rightarrow \Delta, \forall x \varphi_x^a} \text{ provided that } A \text{ does not occur in } \Gamma, \Delta \\
(\forall \rightarrow) \quad \frac{\varphi_t^x, \forall x \varphi, \Gamma \rightarrow \Delta}{\forall x \varphi, \Gamma \rightarrow \Delta} \\
(\rightarrow \exists)Q \quad \frac{\Gamma \rightarrow \Delta, \exists x \varphi, \varphi_t^x}{\Gamma \rightarrow \delta, \exists x \varphi} \\
(\exists \rightarrow) \quad \frac{\varphi, \Gamma \rightarrow \Delta}{\exists x \varphi_z^a, \Gamma \rightarrow \Delta} \text{ provided that } a \text{ does not occur in } \Gamma, \Delta.
\end{array}$$

This system we call **CS**—‘*C*’ for classical.

The reader should convince himself that **CS** is sound for the standard truth-value semantics for classical logic, i.e. that if  $\vdash_{\mathbf{CS}} \Gamma \rightarrow \Delta$ , then  $\models \Gamma \rightarrow \Delta$ , where  $\models \Gamma \rightarrow \Delta$  is defined to mean that if all the members of  $\Gamma$  are true in an interpretation, then at least one member of  $\Delta$  is true in the same interpretation.

Hence, **CS** does not give us more than classical logic, and as a matter of fact **CS** is *complete* and gives us *all* of classical logic. First one notes that **CS** does prove all instances of TND:

$$\begin{array}{l}
\frac{\varphi \rightarrow \varphi}{\rightarrow \varphi, \neg \varphi} (\rightarrow \neg) \quad (\text{Axiom}) \\
\frac{\rightarrow \varphi, \neg \varphi}{\rightarrow \varphi \vee \neg \varphi} (\rightarrow \vee)
\end{array}$$

Secondly, one notes that **CS** is closed under *Thinning*. Let  $D$  be a **CS** proof of the sequent  $\Gamma \rightarrow \Delta$ . One has to show that  $\Gamma, \Gamma' \rightarrow \Delta, \Delta'$  is **CS** provable for any finite  $\Gamma'$  and  $\Delta'$ . Inspect the parameters which occur in  $\Gamma', \Delta'$ ; these are finitely many. Therefore, every *eigen*-parameter of  $D$  which also occurs in  $\Gamma', \Delta'$ , can be exchanged for a new *eigen*-parameter which does not occur in  $D$  or in  $\Gamma', \delta'$  (as we have infinitely many parameters at our disposal). The result is still a derivation  $D'$  of the sequent  $\Gamma \rightarrow \Delta$ . If we now add  $\Gamma'$  and  $\Delta'$  as side formulae everywhere, the result is still a derivation  $D''$  in **CS** (as axioms go into axioms and applications of rules become applications of the same rules), *without violating any eigen-parameter conditions*. Hence, we have found the desired derivation of the sequent  $\Gamma, \Gamma' \rightarrow \Delta, \Delta'$  which is, thus **CS** provable.

Thirdly, observe that from the premise of an **IS** rule, the premise of the corresponding **CS** rule can be inferred by *Thinning*, and then the conclusion of the **IS** rule follows by the **CS** rule from the **CS** premise, e.g. from the

**IS** premise  $\varphi, \Gamma \rightarrow \Delta$  by thinning one obtains the **CS** premise  $\varphi, \psi, \Gamma \rightarrow \Delta$ , and the **CS** rule ( $\&\rightarrow$ ) yields the **IS** conclusion  $\varphi \& \psi, \Gamma \rightarrow \Delta$ . Hence, **CS** includes **IS**, and proves TND and is thus complete for classical logic.

The system **CS** is treated in [Kleene, 1967, Chapter VI] where, in particular, an elegant completeness-proof is given, using the technique of treating the sequent as expressing a falsifiability condition and applying the sequent rules backwards one searches for a counter-model effecting the falsifying interpretation. The whole treatment is very systematic and leads to a canonical search method. We will not enter into the details of the proof for the present calculus, but postpone the matter until we discuss the next system to be considered, viz. a ‘Tableaux system’ for *signed formulae*. Before we discuss this, however, we wish to consider one ore, fairly complicated, derivation in the system **CS**. For this we choose to derive in **CS** the schema CD—‘constant domains’, so called because it is valid in Kripke models with constant domains of individuals only, cf. van Dalen’s chapter in Volume 7 of this *Handbook*—which is not **IS**-provable.

$$\text{CD} =_{\text{def}} \forall x(\varphi \vee \psi) \supset \varphi \vee \forall x\psi, \text{ where } x \text{ does not occur in } \varphi.$$

We derive CD as follows: Pick a parameter  $a$  which does not occur in  $\varphi$ .

$$\begin{array}{l} (\vee \rightarrow) \quad \frac{\varphi, \forall x(\varphi \vee \psi) \rightarrow \varphi, \psi_a^x \quad \psi_a^x, \forall x(\varphi \vee \psi) \rightarrow \varphi, \psi_a^x}{\varphi \vee \psi_a^x, \forall x(\varphi \vee \psi) \rightarrow \varphi, \psi_a^x} \\ (\forall \rightarrow) \quad \frac{\varphi \vee \psi_a^x, \forall x(\varphi \vee \psi) \rightarrow \varphi, \psi_a^x}{\forall x(\varphi \vee \psi) \rightarrow \varphi, \psi_a^x} \\ (\rightarrow \forall) \quad \frac{\forall x(\varphi \vee \psi) \rightarrow \varphi, \psi_a^x}{\forall x(\varphi \vee \psi) \rightarrow \varphi, \forall x\psi} \\ (\rightarrow \vee) \quad \frac{\forall x(\varphi \vee \psi) \rightarrow \varphi, \forall x\psi}{\forall x(\varphi \vee \psi) \rightarrow \varphi \vee \forall x\psi} \\ (\rightarrow \supset) \quad \frac{\forall x(\varphi \vee \psi) \rightarrow \varphi \vee \forall x\psi}{\rightarrow \forall x(\varphi \vee \psi) \supset \varphi \vee \forall x\psi} \end{array}$$

The cut-elimination theorem holds for **CS** as well as by the same type of syntactic manipulations as in the case of **IS** and **IS**<sup>+</sup>. The fact that the cut-free calculus is complete is very useful, because cut-free derivations have the *sub-formula property*, i.e. that every formula which occurs in the derivation is a sub-formula of some formula in the conclusion (here one has to count  $\varphi$  as a sub-formula of  $\varphi$ ). This simple fact that the conclusion gives abound on the complexity of the derivation is the source of a host of metamathematical information. We refer to [Takeuti, 1975] for a comprehensive survey of how to use the cut-free derivations for metamathematical purposes.

Before we leave the system **CS**, we wish to mention what is basically a notational variant due to Schütte [1951]. The semantic interpretation of a sequent  $\varphi_1, \dots, \varphi_k \rightarrow \psi_1, \dots, \psi_m$  is given by the formula  $\varphi_1 \& \dots \& \varphi_k \supset \psi_1 \vee \dots \vee \psi_m$ , or equivalently, by classical logic,  $\neg\varphi_1 \vee \dots \vee \neg\varphi_k \vee \psi_1 \vee \dots \vee \psi_m$ . The sequent-rules of **CS** can therefore be thought of as giving rules

of passage between such disjunctions. The axioms, for instance, become  $\varphi \vee \psi \vee \neg\varphi$  (both sets of side formulae are joined into  $\psi$ ). The rule  $(\vee \rightarrow)$  becomes:

$$(\neg\vee) \frac{\neg\varphi \vee \theta \quad \neg\psi \vee \theta}{\neg(\varphi \vee \psi) \vee \theta}$$

whereas  $(\rightarrow \neg)$  disappears: the premise has the form  $\varphi, \Gamma \rightarrow \Delta$  and the conclusion  $\Gamma \rightarrow \Delta, \neg\varphi$ , but their ‘Schütte-translations’ are the same. We leave the other reformulations to the reader and just consider cut:

$$\frac{\psi \vee \varphi \quad \neg\varphi \vee \theta}{\psi \vee \theta}$$

NB. It is customary to restrict oneself to the connectives  $\&$ ,  $\vee$  and  $\neg$  in the Schütte-style formulation.

A readily available description of a Schütte-type system can be found in [Mendelson, 1997, Appendix], where an exposition of Schütte’s consistency proof using the infinitary  $\omega$ -rule also can be found. (Unfortunately, the appendix was dropped from later editions of the book.)

Another variant of the same idea is found in [Tait, 1968]. We now explain the mechanism for the propositional case. With each propositional letter  $p$  we associate another propositional letter  $\bar{p}$ , called the *complement* of  $p$ . As we work in classical logic, we confine ourselves to using  $\&$ ,  $\vee$  and  $\neg$  only. One then also defines the complements for complex formulae:

$$\begin{aligned} \underline{p} &=_{\text{def}} p \\ \overline{(p \vee q)} &=_{\text{def}} \bar{p} \& \bar{q} \\ \overline{(P \& q)} &=_{\text{def}} \bar{p} \vee \bar{q} \end{aligned}$$

The negation of a formula is identified with its complement. This has the advantage that all formulae are in negation-normal form, i.e. negation can be thought of as applying only to propositional letters.

The sequents are finite sets of formulae; on the intended interpretation they are read as disjunctions (so one has taken one step beyond Schütte; the disjunction signs are not written out). The system of axioms and rules is particularly easy to give:

*Axiom*  $\Gamma, \varphi, \bar{\varphi}$

$$(\vee) \frac{\Gamma, \varphi}{\Gamma, \varphi \vee \psi} \text{ and } \frac{\Gamma, \psi}{\Gamma, \varphi \vee \psi}$$

$$(\&) \frac{\Gamma, \varphi \quad \Gamma, \psi}{\Gamma, \varphi \& \psi}$$

$$(\text{CUT}) \frac{\Gamma, \psi \quad \Gamma, \neg\varphi}{\Gamma}$$



The rules ( $\vee$ ) and ( $\&$ ) correspond, in the first place, to ( $\rightarrow \vee$ ) and ( $\rightarrow \&$ ), respectively. But the original ( $\& \rightarrow$ ) rule corresponds to the inference of  $\neg(\varphi \& \psi), \Gamma$  from the premise  $\neg\varphi, \Gamma$ , say. As one now identifies the negation of a formula with the complement of that formula, this corresponds to the inference of  $\overline{\varphi \& \psi}, \Gamma$  from  $\overline{\varphi}, \Gamma$ , but by definition, the complement of a conjunction is the disjunction of the complements. Hence, we have an instance of the rule ( $\vee$ ).

The present approach thus manages to keep the number of rules at a minimum, and this can be put to metamathematical use. The present sort of Tait-like system is used by Schwichtenberg [1977] (for full predicate logic and various systems of arithmetic) to present Gentzen's cut elimination theorem in a very compact way. We give a derivation of CD within the Schütte-type formulation so as to bring out the essential equivalence with the original formulation:

$$\frac{\frac{\frac{\neg\varphi \vee \varphi \vee \psi_a^x \quad \neg\psi_a^x \vee \varphi \vee \psi_a^x}{\neg(\varphi \vee \psi_a^x) \vee \varphi \vee \psi_a^x} (\neg\vee)}{\neg\forall x(\varphi \vee \psi) \vee \varphi \vee \psi_a^x} (\forall)}{\neg\forall x(\varphi \vee \psi) \vee \varphi \vee \forall x\psi} (\forall)$$

The rest of the steps fall away since implication is no longer a connective.

We conclude our treatment of the systems **IS** and **CS** (and their notational variants) by drawing the reader's attention to the survey article Bernays [1965] in which a wealth of information as to various options in formulating the rules and systems is systematically presented. In particular, the connections between general consequence relations and conditions on sequents are examined. The reader interested in this area should also consult the series of papers by Scott, cf. [1974] and other references given therein.

There now remains only the last of the main variants of the sequent calculus, namely systems of *semantic tableaux*. These systems arose out of the completeness proofs for the cut-free Sequent Calculus which were independently discovered in the mid 1950s by Beth, Hintikka, Kanger and Schütte. (Cf. [Prawitz, 1975] for historical references and a lucid exposition of the completeness proof in question. Another good reference is, as already remarked [Kleene, 1967, Chapter VI, p. 285].) The method of tableaux can be applied both in the intuitionistic and classical settings, although the former strikes the author as being a bit artificial. Here we confine ourselves to the classical case, for which the *locus classicus* is [Smullyan, 1968] and whose presentation we follow. For the intuitionistic case refer, e.g. to [Fitting, 1969] or [Bell and Machover, 1977] (another good source for information about tableaux.)

We first need a

DEFINITION 4. If  $\varphi$  is a wff, then  $T\varphi$  and  $F\varphi$  are both *signed formulae*.

Note that signed formulae cannot be joined together by connectives etc. to form other signed formulae; their intended interpretations are, of course, ‘ $\varphi$  is true’ and ‘ $\varphi$  is false’, respectively.

The main idea is now to take the falsifiability condition interpretation of **CS** seriously. Thus, the *sequent*  $\{\varphi_1, \dots, \varphi_k\} \rightarrow \{\psi_1, \dots, \psi_m\}$  is now transformed into a *finite set of signed formulae*  $T\varphi_1, \dots, T\varphi_k, F\psi_1, \dots, F\psi_m$  and a further change is that the **CS** rules are *turned upside-down*. One therefore treats a derivation as a systematic search for a falsifying interpretation, and the branching rules, which in **CS** have got more than one premise, now indicate that a falsifying interpretation has to be sought along different possibilities.

We begin by presenting the derivation of CD (before we give the rules): Consider CD; we wish to prove it classically so we wish to show the absence of a falsifying interpretation. therefore, we begin by assuming that

$$F(\forall x(\varphi \vee \psi) \supset \varphi \vee \forall x\psi).$$

A necessary (and sufficient) condition for this to be the case, is:

$$T\forall x(\varphi \vee \psi), \quad F(\varphi \vee \forall x\psi)$$

because an implication is false precisely when its antecedent is true and the consequent is false. It is a necessary and sufficient condition for this to hold that

$$T\forall x(\varphi \vee \psi), \quad F\varphi, \quad F\forall x\psi$$

because a disjunction is false precisely when both disjuncts are false.

But in order to falsify  $\forall x\psi$ , we must falsify  $\psi_a^x$  for some  $a$ , *about which we have assumed nothing in particular* up till now, so:

$$T\forall x(\varphi \vee \psi), \quad F\varphi, \quad F\psi_a^x$$

is a falsifiability condition for CD. Now, in order to make  $\forall x(\varphi \vee \psi)$  true we also need to make, among other instances,  $\varphi \vee \psi_a^x$  true. Thus,

$$T\forall x(\varphi \vee \psi), \quad T(\varphi \vee \psi_a^x) \quad F\varphi, \quad F\psi_a^x \quad (\text{Here we use that } x \text{ does not occur in } \varphi.)$$

is a falsifiability condition for CD. Now, the condition splits into two possibilities for falsifying CD, as in order to make a disjunction true it is sufficient to make only one of the disjuncts true:

$$(\text{1st possibility}): \quad T\forall x(\varphi \vee \psi), \quad T\varphi, \quad F\varphi, \quad F\psi_a^x.$$

This possibility will not yield a falsifying interpretation though, because we cannot assign both the value true and the value false to the same wff. Hence, the search along this possibility may safely be abandoned or *closed*.

(2nd possibility):  $T\forall x(\varphi \vee \psi)$ ,  $T\psi_a^x$ ,  $F\varphi$ ,  $F\psi_a^x$ .

Also this possibility has to be closed, because  $\psi_a^x$  has to be made true and false. (It will not have escaped the attentive reader that the closure conditions correspond exactly to the *axioms* of **CS**.) But we have now exhausted *all* the possibilities for finding a falsifying interpretation; thus CD must be classically valid as none of the possible routes along which we could hope for a falsifying interpretation is able to yield one.

The above search can be set out more compactly:

$$\frac{\frac{\frac{\frac{F(\forall x(\varphi \vee \psi) \supset \varphi \vee \forall x\psi)}{T\forall x(\varphi \vee \psi), F\varphi \vee \forall x\psi}}{T\forall x(\varphi \vee \psi), F\varphi, F\forall x\psi}}{T\forall x(\varphi \vee \psi), F\varphi, F\psi_a^x}}{T\forall x(\varphi \vee \psi), T\varphi \vee \psi_a^x, F\varphi, F\psi_a^x}}{T\forall x(\varphi \vee \psi), T\varphi, F\varphi, F\psi_a^x \mid T\forall x(\varphi \vee \psi), T\psi_a^x, F\varphi, F\psi_a^x}}$$

This search tree, however, is nothing but the **CS** derivation of CD turned upside-down and rewritten using other notation.

We now give the rules for the tableaux system **T**. We will use ‘ $S$ ’ as a notation for finite sets of signed formulae.

$$\begin{array}{l} F\& \frac{S, F\varphi\&\psi}{S, F\varphi \mid S, F\psi} \\ T\& \frac{S, T\varphi\&\psi}{S, T\varphi, T\psi} \\ F\vee \frac{S, F\varphi \vee \psi}{S, F\varphi, F\psi} \\ T\vee \frac{S, T\varphi \vee \psi}{S, T\varphi \mid S, T\psi} \\ F\supset \frac{S, F\varphi \supset \psi}{S, T\varphi, F\psi} \\ T\supset \frac{S, T\varphi \supset \psi}{S, T\psi \mid S, F\varphi} \\ F\neg \frac{S, F\neg\varphi}{S, T\varphi} \end{array}$$

$$\begin{array}{l}
T\neg \frac{S, T\neg\varphi}{S, F\varphi} \\
F\forall \frac{S, F\forall x\varphi}{S, F\varphi_a^x} \text{ provided that } a \text{ does not occur in } S. \\
T\forall \frac{S, T\forall x\varphi}{S, T\forall x\varphi, T\varphi_t^x} \\
F\exists \frac{S, F\exists x\varphi}{S, F\exists x\varphi, F\varphi_t^x} \\
T\exists \frac{S, T\exists x\varphi}{s, T\varphi_a^x} \text{ provided that } a \text{ does not occur in } S.
\end{array}$$

Some of the rules give a *branching*, viz.  $T\forall$ ,  $F\&$  and  $T\supset$ . This is because of the fact that in these cases there is more than just one way of fulfilling the falsifiability condition above the line and all of these must be investigated.

A tableaux proof for  $\varphi$  is a finite tree of finite sets of signed formulae, regulated by the above rules such that (i) the top node of the tree is the set  $\{F\varphi\}$  and (ii) every branch in the tree is *closed*, i.e. ends with a set of the form  $S, T\varphi, F\varphi$ . Hence, the tableaux proofs should best be regarded as failed attempts at the construction of a counter-model.

We also show how one can define a counter-model to the top node set from an open branch in the tableaux, where one cannot apply the rules any longer.

Consider the propositional wff  $(p \supset q) \supset (p \& q)$ . This is not a tautology, and, hence, it is not provable. An attempt at a tableaux proof looks like:

$$\frac{\frac{F(p \supset q) \supset (p \& q)}{Tp \supset q, Fp \& q}}{\frac{Tp \supset q, Fp}{Fp, Fp \mid Tq, Fp} \mid \frac{Tp \supset q, Fp}{Fp, Fq \mid Tq, Fq}}$$

The last of the branches in the search tree is closed, but from the third, say, we can define a valuation  $\nu$  by

$$\nu(p) = F \text{ and } \nu(q) = F.$$

Then the value of the formula  $(p \supset q) \supset (p \& q)$  under  $\nu$  is  $F$ , so we have found our counter-model. This simple example contains the germ of the completeness proof for tableaux proofs; one describes a systematic procedure for generating a search tree (in particular, for how to choose the *eigen*-parameters) such that either the procedure breaks off in a tableaux

proof or it provides a counter-model. For the details we refer to [Smullyan, 1968].

The above formulation, using finite sets of signed formulae, was used mainly to show the equivalence with the classical Sequent Calculus **CS**. It is less convenient in actual practice, as one must duplicate the side formulae in  $S$  all the time. A more convenient arrangement of the tableaux consists of using trees where the nodes are not sets of signed formulae but just one signed formula. To transform a ‘set’ tableaux by a more convenient tableaux one takes the top node which has the form, say  $\{T\varphi_1, \dots, T\varphi_k, F\psi_1, \dots, F\psi_m\}$  and places it upright instead:

$$\begin{array}{c} T\varphi_1 \\ \vdots \\ T\varphi_k \\ F\psi_1 \\ \vdots \\ F\psi_m. \end{array}$$

Such an arrangement can then be continued by use of the rules:

$$\begin{array}{ccc} \begin{array}{c} F\varphi \& \psi \\ / \quad \backslash \\ F\varphi \quad F\psi \end{array} & & \begin{array}{c} T\varphi \& \psi \\ \hline T\varphi \\ T\psi \end{array} \\ \\ \begin{array}{c} F\varphi \vee \psi \\ \hline F\varphi \\ F\psi \end{array} & & \begin{array}{c} T\varphi \vee \psi \\ / \quad \backslash \\ T\varphi \quad T\psi \end{array} \\ \\ \begin{array}{c} F\varphi \supset \psi \\ \hline T\varphi \\ F\psi \end{array} & & \begin{array}{c} T\varphi \supset \psi \\ / \quad \backslash \\ T\psi \quad F\varphi \end{array} \end{array}$$

$$\begin{array}{c}
\frac{F\neg\varphi}{T\varphi} \\
\frac{F\forall x\varphi}{f\varphi_a^x} \text{ provided that } a \text{ is} \\
\text{new to the branch}
\end{array}
\qquad
\begin{array}{c}
\frac{T\neg\varphi}{F\varphi} \\
\frac{T\forall x\varphi}{T\varphi_t^x}
\end{array}$$
  

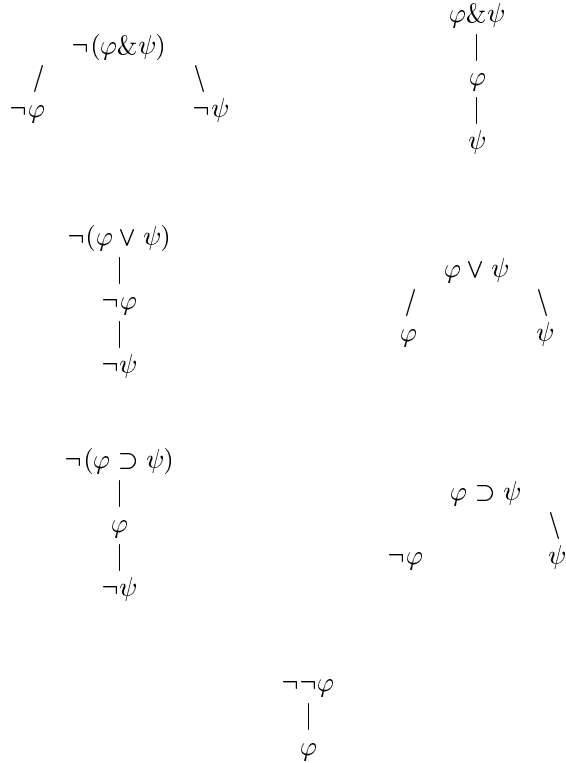
$$\begin{array}{c}
\frac{F\exists x\varphi}{F\varphi_t^x} \\
\frac{T\exists x\varphi}{T\varphi_a^x} \text{ provided that } a \text{ is} \\
\text{new to the branch}
\end{array}$$

We do not wish to trouble the reader with a more detailed description of the modified system, and so we confine ourselves to showing how the rules are used in practice for proving CD:

$$\begin{array}{l}
(1) \quad f\forall x(\varphi \vee \psi) \supset \varphi \vee \forall x\varphi \\
\quad \quad | \\
(2) \quad T\forall x(\varphi \vee \psi) \\
\quad \quad | \\
(3) \quad F\varphi \vee \forall x\psi \\
\quad \quad | \\
(4) \quad F\varphi \\
\quad \quad | \\
(5) \quad F\forall x\psi \\
\quad \quad | \\
(6) \quad F\psi_a^x \\
\quad \quad | \\
(7) \quad T\varphi \vee \psi_a^x \\
(8) \quad / \quad T\varphi \quad \backslash \quad T\psi_a^x \quad (9)
\end{array}$$

The tableaux proof begins with the signed formula  $FCD$ , as we try to show that there is no counter-model to CD. Lines (2) and (3) result from (1) by breaking down the implication. Lines (4) and (5) likewise result from (3) and, as the parameter  $a$  is new to the (only) branch the step from (5) to (6) is permitted. Line (7) results from (2) as we have the right to choose any term here. Finally, (8) and (9) come from (7) by breaking up a true disjunction. Both branches are closed at once; (8) closes off against (4) and (9) against (6).

Although this is very simple, there is yet another simplification which can be performed: *drop the signed formulae altogether* and in place of ' $F\varphi$ ' write ' $\neg\varphi$ ' and in place of ' $T\varphi$ ' write ' $\varphi$ '. The resulting rules look like:



the elucidating remarks after the previous example apply almost word for word also here. There remains only to point out that this last style of ‘unsigned’ tableaux is used in readily available elementary texts, e.g. [Jeffrey, 1990] and [Hodges, 1977].

The order of presentation used in the present Chapter was also used by the author in those sections of Scott *et al.* [1981, Volume II] which were drafted by him. Those notes, Volumes I and II, were intended as a supplement to [Hodges, 1977] and, in particular, Volume I contains a very detailed development of a tableaux system.

#### ACKNOWLEDGEMENT

I wish to thank my colleague Ton Weijters who read the manuscript of the present chapter and offered helpful comments.

*Universiteit Leiden, The Netherlands.*

## BIBLIOGRAPHY

- [Bell and Machover, 1977] J. Bell and M. Machover. *A Course in Mathematical Logic*. North-Holland, Amsterdam, 1977.
- [Bernays, 1965] P. Bernays. Betrachtungen zum sequenzen-kalkül. In A.-T. Tymieniecka, editor, *Logic and Methodology*, pages 1–44. North-Holland, Amsterdam, 1965.
- [Church, 1956] A. Church. *Introduction to Mathematical Logic*, volume 1. Princeton University Press, Princeton, 1956.
- [Dummett, 2000] M. Dummett. *Elements of Intuitionism*, 2nd edition. Oxford University Press, Oxford, (1st ed., 1977), 2000.
- [Enderton, 2000] H. B. Enderton. *A Mathematical Introduction to Logic*, 2nd edition. Academic Press, New York, (2st ed., 1972), 2000.
- [Feferman, 1960] S. Feferman. Arithmetisation of metamathematics in a general setting. *Fundamenta Mathematica*, 49:35–92, 1960.
- [Fitting, 1969] M. Fitting. *Intuitionistic Logic, Model Theory and Forcing*. North-Holland, Amsterdam, 1969.
- [Frege, 1879] G. Frege. *Begriffsschrift*. Nebert, Halle, 1879. Complete English translation in [van Heijenoort, 1967].
- [Gentzen, 1934] G. Gentzen. Untersuchungen über das logische schliessen. *Math. Zeitschrift*, 39:176–210, 405–431, 1934. Complete English translation in [Szabo, 1969].
- [Gentzen, 1936] G. Gentzen. *Die Widerspruchsfreiheit der reinen Zahlentheorie*, volume 112. 1936. Complete English translation in [Szabo, 1969].
- [Hilbert and Bernays, 1934] D. Hilbert and P. Bernays. *Grundlagen der Mathematik*, volume 1. Springer-Verlag, Berlin, 1934.
- [Hodges, 1977] W. Hodges. *Logic*. Penguin, Harmondsworth, 1977.
- [Jeffrey, 1990] R. C. Jeffrey. *Formal Logic: Its Scope and Limits*, 3rd edition. McGraw-Hill, New York, (1st ed., 1967), 1990.
- [Kalish and Montague, 1965] D. Kalish and R. Montague. On tarski's formalisation of predicate logic with identity. *Archiv für mathematische Logik und Grundlagenforschung*, 7:81–101, 1965.
- [Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics*. North-Holland, Amsterdam, 1952.
- [Kleene, 1967] S. C. Kleene. *Mathematical Logic*. John Wiley, New York, 1967.
- [Kripke, 1963] S. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963.
- [Leivant, 1979] D. Leivant. Assumption classes in natural deduction. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 25:1–4, 1979.
- [Martin-Löf, 1971] P. Martin-Löf. Hauptsatz for the intuitionistic theory of iterated inductive definitions. In J. E. Fenstad, editor, *Proceedings of the Second Scandinavian Logic symposium*, pages 179–216. North-Holland, Amsterdam, 1971.
- [Mendelson, 1997] E. Mendelson. *Introduction to Mathematical Logic*, 4th edition. Van Nostrand, New York, (1st ed., 1964), 1997.
- [Monk, 1965] D. Monk. Substitutionless predicate logic with identity. *Archiv für mathematische Logik und Grundlagenforschung*, 7:102–121, 1965.
- [Pottinger, 1979] G. Pottinger. Normalisation as a homomorphic image of cut-elimination. *Annals of Math. Logic*, 12:323–357, 1979.
- [Prawitz, 1965] D. Prawitz. *Natural Deduction*. Almqvist and Wiksell, Uppsala, 1965.
- [Prawitz, 1971] D. Prawitz. Ideas and results in proof theory. In J. E. Fenstad, editor, *Proceedings of the Second Scandinavian Logic symposium*, pages 235–307. North-Holland, Amsterdam, 1971.
- [Prawitz, 1975] D. Prawitz. Comments on gentzen-style procedures and the classical notion of truth. In J. Diller and G. H. Müller, editors, *Proof Theory Symposium*, pages 190–319. Lecture Notes in Mathematics 500, Springer-Verlag, Berlin, 1975.
- [Prawitz, 1977] D. Prawitz. Meaning of proofs. *Theoria*, 43:2–40, 1977.
- [Schütte, 1951] K. Schütte. Schlussweisen-kalküle er prädikatenlogik. *Math. Annalen*, 122:47–65, 1951.



- [Schwichtenberg, 1977] H. Schwichtenberg. Proof theory: some applications of cut-elimination. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 867–896. North-Holland, Amsterdam, 1977.
- [Scott et al., 1981] D. Scott, D. Bostock, G. Forbes, D. Isaacson, and G. Sundholm. *Notes on the Formalisation of Logic, Vols I, II*. Sub-Faculty of Philosophy, Oxford, 1981.
- [Scott, 1974] D. S. Scott. Rules and derived rules. In S. Stenlund, editor, *Logical Theory and Semantical Analysis*, pages 147–161. Reidel, Dordrecht, 1974.
- [Smoryński, 1977] C. Smoryński. The incompleteness theorems. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 821–865. North-Holland, Amsterdam, 1977.
- [Smullyan, 1968] R. Smullyan. *First-Order Logic*. Springer-Verlag, Berlin, 1968.
- [Szabo, 1969] M. Szabo. *The Collected Papers of Gerhard Gentzen*. North-Holland, Amsterdam, 1969.
- [Tait, 1968] W. W. Tait. Normal derivability in classical logic. In J. Barwise, editor, *The Syntax and Semantics of Infinitary Languages*, pages 204–236. Lecture Notes in Mathematics 72, Springer-Verlag, Berlin, 1968.
- [Takeuti, 1975] G. Takeuti. *Proof Theory*. North-Holland, Amsterdam, 1975.
- [Tarski, 1965] A. Tarski. A simplified formulation of predicate logic with identity. *Archiv für mathematische Logik und Grundlagenforschung*, 7:61–79, 1965.
- [Tennant, 1978] N. Tennant. *Natural Logic*. Edinburgh University Press, Edinburgh, 1978.
- [Troelstra, 1973] A. S. Troelstra, editor. *Metamathematical Investigations of Intuitionistic Arithmetic and Analysis*. Lecture Notes in Mathematics 344, Springer-Verlag, Berlin, 1973.
- [van Dalen, 1997] D. van Dalen. *Logic and Structure*, 3rd Edition. Springer-Verlag, Berlin, (1st ed., 1980), 1997.
- [van Heijenoort, 1967] J. van Heijenoort. *From Frege to Gödel*. Harvard University Press, Cambridge, MA, 1967.
- [Zucker, 1974] J. Zucker. The correspondence between cut-elimination and normalisation. *Annals of Math. Logic*, 7:1–112, 1974.

#### EDITOR'S NOTE

The following additional books are relevant to this and related chapters:

#### BIBLIOGRAPHY

- [Bostock, 1997] D. Bostock. *Intermediate Logics*, Clarendon Press, Oxford, 1997.
- [Gabbay, 1996] D. M. Gabbay. *Labelled Deductive Systems*, Oxford Logic Guides, Oxford University Press, Oxford, 1996.
- [Mackie and Goubault-Larrecq, 1997] I. Mackie and J. Goubault-Larrecq. *Proof Theory and Automated Deduction*, APLS, Kluwer, Dordrecht, 1997.
- [Mints, 2000] G. Mints. *A Short Introduction to Intuitionistic Logic*, Kluwer, Dordrecht, 2000.

# ALTERNATIVES TO STANDARD FIRST-ORDER SEMANTICS

## 0 INTRODUCTION

Alternatives to standard semantics are legion, some even antedating standard semantics. I shall study several here, among them: *substitutional semantics*, *truth-value semantics*, and *probabilistic semantics*. All three interpret the quantifiers *substitutionally*, i.e. all three rate a universal (an existential) quantification true if, and only if, every one (at least one) of its substitution instances is true.<sup>1</sup> As a result, the first, which retains models, retains only those which are to be called *Henkin models*. The other two dispense with models entirely, truth-value semantics using instead truth-value assignments (or equivalents thereof to be called *truth-value functions*) and probabilistic semantics using probability functions. So reference, central to standard semantics, is no concern at all of truth-value and probabilistic semantics; and truth, also central to standard semantics, is but a marginal concern of probabilistic semantics.

Each of these alternatives to standard semantics explicates logical entailment—and, hence, logical truth—in its own way.<sup>2</sup> In all three cases, however, it can and will be shown that

1. *A statement is logically entailed by a set of statements if, and only if, provable from the set,*

and hence that

2. *A statement is logically true if, and only if, provable.*

Statement (1) will legitimise each account of logical entailment, (2) will legitimise each account of logical truth, and (1) and (2) together will legitimise each alternative semantics treated here.

---

<sup>1</sup>The rival, and more generally accepted, interpretation of the quantifiers is of course the *objectual* one (that in (iv) on page 60). Goldfarb [1979] intimates that the first correct account of (first-order) logical truth, an objectual one, is in [Bernays, 1922]. By then a substitutional account of logical truth had already been sketched in [Wittgenstein, 1921]. To take the matter further back, Frege [1893–1903] interprets the quantifiers objectually, but Frege [1879] interprets them substitutionally (see [Stevenson, 1973]).

<sup>2</sup>In standard semantics and each alternative to it studied here, logical truth is but logical entailment by the null set, the way provability (i.e. theoremhood) is but provability from that set. As a result more space will generally be devoted to logical entailment than to logical truth, and by the same token to strong than to weak soundness and completeness.

I had two options: presenting each of substitutional semantics, truth-value semantics, and probabilistic semantics as a semantics *for CQC*, *the first-order quantificational calculus*, or as one *for an arbitrary first-order language L*. I chose the latter option for a simple reason. This essay deals largely with truth and probability, and to me truth and probability (the latter understood as a degree of rational belief) are features of the statements you meet in a language rather than the statement forms you meet in such a language form as **CQC**.

For use in alternatives to standard semantics, though *not* in standard semantics itself, I outfit  $L$  with term extensions. One of them,  $L^\infty$ , will serve to prove various completeness theorems in Sections 2–5. I could, instead, have outfitted each set of statements of  $L$  with term rewrites, which was the practice in [Leblanc, 1976] and earlier writings of mine. However, term extensions, exploited in [Dunn and Belnap, 1968] are handier and admittedly more natural. So I switched to them.<sup>3</sup>

Several matters are studied below, and in the process are bound together. I define for each semantics the notions of logical truth and logical entailment, plus, of course, such notions as they presuppose; and, as announced, I justify the definitions by showing that of logical entailment *strongly*—and hence, that of logical truth *weakly*—sound and complete. I further show that (i) substitutional semantics is a by-product of standard semantics, (ii) truth-value semantics is substitutional semantics done without models, and (iii) probabilistic semantics is a generalisation of truth-value semantics, a transcription of truth-value semantics into the idiom of sets, and model-set semantics, an intriguing and handy variant by Hintikka of truth-set semantics. And, occasionally in the main text, but more often in the footnotes and the Appendix, I supply pertinent names, dates, and references.

However, there is far more to truth-value and probabilistic semantics than the essay conveys. (i) When studying probabilistic semantics, I pay particular attention to *singular* (i.e. one-argument) probability functions. *Binary* (i.e. two-argument) ones also permit definition of logical truth and logical entailment, as writers from Popper on have shown. They receive some attention here, but deserve far more. (ii) When studying truth-value and probabilistic semantics, I largely restrict myself to matters of logical truth, logical entailment, soundness, and completeness. But, as Leblanc [1976] attests, a host of definitions, theorems, and proofs from standard semantics translate into the idiom of truth-values; and many—though, for sure, not all—of them translate as well into the idiom of probabilities. The translations, sometimes easy to come by but sometimes not, should figure in any full-fledged treatment of either semantics. (iii) I restrict myself throughout

---

<sup>3</sup>Term extensions—i.e. languages exactly like  $L$  except for having individual terms—have been in use for a good many years. They figure in [Henkin, 1949], in [Hintikka, 1955] (a passage from which is quoted in note 16), in [Gaifman, 1964], etc. The Dunn–Belnap extensions I use are the least extensions of  $L$  that serve all my purposes here.

the essay to *first-order logic without identity*. Yet other logics (first-order logic with identity, higher-order logics, many-valued logics, modal logics, tense logics, intuitionistic logic, conditional logic, etc.) have also been supplied with a truth-value semantics or a probabilistic one. Only study of these extensions of and alternatives to elementary logic would reveal the true scope of either semantics.

The alternatives to standard semantics studied here have, in my opinion, considerable interest and—possibly—merit. As noted above, they are *frugal*, (i) substitutional semantics discarding all but Henkin models, while truth-value semantics and probabilistic semantics discard all models, and (ii) truth-value-semantics—though it retains the notion of truth—discarding that of reference, while probabilistic semantics discards both notions. And they are *innovative*, truth-value semantics assigning truth-values to the atomic ‘substatements’ of all statements (those with quantifiers as well as those without), while probabilistic semantics assigns to statements degrees of credibility rather than truth-values.

I touch on these matters in Sections 2–5, and devote much of Section 6 to them. My main concern, though, is different: to provide the formal prerequisites to further study of and research in non-standard semantics.

## 1 FIRST-ORDER SYNTAX AND STANDARD SEMANTICS

The *primitive signs* of  $L$  will be (i) one or more predicates, each identified as being of a certain degree  $d$  ( $d \geq 1$ ), (ii)  $\aleph_0$  individual terms, presumed to come in some order known as their *alphabetic order*, (iii)  $\aleph_0$  individual variables, (iv) the three logical operators ‘ $\neg$ ’, ‘ $\wedge$ ’, and ‘ $\vee$ ’, (v) the two parentheses ‘(’ and ‘)’’, and (vi) the comma ‘,’. The *formulas* of  $L$  will be all the finite sequences of the primitive signs of  $L$ . I shall refer to the predicates of  $L$  by means of ‘ $Q$ ’; to its individual terms in general by means of ‘ $T$ ’, and—for each  $i$  from 1 on—to its alphabetically  $i$ th individual term by means of ‘ $t_i$ ’; to its individual variables by means of ‘ $x$ ’; to its individual *signs* (i.e. its individual terms and individual variables) by means of ‘ $T$ ’; to its formulas by means of ‘ $A$ ’, ‘ $B$ ’, and ‘ $C$ ’; and to sets of its formulas by means of ‘ $S$ ’. And, (i)  $A$  being a formula of  $L$ , (ii)  $I_1, I_2, \dots, I_n$  ( $n \geq 0$ ) being distinct individual signs of  $L$ , and (iii)  $I'_1, I'_2, \dots, I'_n$  being individual signs of  $L$  not necessarily distinct from one another nor from  $I_1, I_2, \dots, I_n$ , I shall refer by means of ‘ $((A)(I'_1, I'_2, \dots, I'_n / I_1, I_2, \dots, I_n))$ ’ to the result of simultaneously putting  $I'_1$  everywhere in  $A$  for  $I_1$ ,  $I'_2$  for  $I_2$ ,  $\dots$ ,  $I'_n$  for  $I_n$ . (When clarity permits I shall omit some of the parentheses in ‘ $((A)(I'_1, I'_2, \dots, I'_n / I_1, I_2, \dots, I_n))$ ’, thus writing ‘ $(A)(I'_1, I'_2, \dots, I'_n / I_1, I_2, \dots, I_n)$ ’, ‘ $(A(I'_1, I'_2, \dots, I'_n / I_1, I_2, \dots, I_n))$ ’, and ‘ $A(I'_1, I'_2, \dots, I'_n / I_1, I_2, \dots, I_n)$ ’.)

The *statements* of  $L$  will be all formulas of  $L$  of the following sorts: (i)  $Q(T_1, T_2, \dots, T_d)$ , where  $Q$  is a predicate of  $L$  of degree  $d$  ( $d \geq 1$ ) and

$T_1, T_2, \dots, T_d$  are (not necessarily distinct) individual terms of L, (ii)  $\neg A$ , where  $A$  is a statement of L, (iii)  $(A \wedge B)$ , where  $A$  and  $B$  are (not necessarily distinct) statements of L, and (iv)  $(\forall x)A$ , where  $A(T/x)$ — $T$  here being any term of L you please—is a statement of L.<sup>4</sup> As usual, the statements in (i) will be called *atomic*, and the rest *compound*, those in (ii) being *negations*, (iii) *conjunctions*, and (iv) (*universal*) *quantifications*. Statements that contain no individual terms will be called *termless* and statements that contain no ‘ $\forall$ ’ will be called *quantifierless*. At a few points the statements of L will be presumed to come in some definite order, to be known as their *alphabetic order* (that on p. 9–10 of [Leblanc, 1976] would do). For brevity’s sake, I shall write ‘ $(A \rightarrow B)$ ’ for ‘ $\neg(A \wedge \neg B)$ ’, ‘ $(A \vee B)$ ’ for ‘ $\neg(\neg A \wedge \neg B)$ ’, ‘ $(A \leftrightarrow B)$ ’ for ‘ $(\neg(A \wedge \neg B) \wedge \neg(B \wedge \neg A))$ ’, ‘ $(\exists x)A$ ’ for ‘ $\neg(\forall x)\neg A$ ’, and ‘ $\pi_{i=1}^n A_i$ ’ for ‘ $((\dots(A_1 \wedge A_2) \wedge \dots) \wedge A_n)$ ’.<sup>5</sup> I shall also drop outer parentheses whenever clarity permits, and I shall talk of terms and variables rather than individual terms and individual variables.

*Substatements* will behave like Gentzen’s subformulas. (i) A statement of L will count as one of its substatements; (ii)  $A$  will count as a substatement of a negation  $\neg A$  of L, each of  $A$  and  $B$  as a substatement of a conjunction  $A \wedge B$  of L, and  $A(T/x)$  for each term  $T$  of L as a substatement of a quantification  $(\forall x)A$  of L; and (iii) the substatements of any substatement of a statement of L will count as substatements of that statement. The *substatements* of a set of statements of L will be the substatements of the various members of the set. As for the *atomic substatements* of a statement or set of statements of L, they will, of course, be those among its substatements that are atomic. When further precision is needed, I shall refer to the foregoing as the substatements *in* L of a statement or set of statements of L.<sup>6</sup>

The substatements  $A(t_1/x), A(t_2/x), A(t_3/x)$  etc. of a universal quantification  $(\forall x)A$ —and, by extension, of an existential one  $(\exists x)A$ —of L are what I called above the *substitution instances* (*in* L) of the quantification. As suggested, they play a critical role in substitutional, truth-value, and probabilistic semantics. (Note that when a quantification  $(\forall x)A$  of L is vacuous, it has but one substitution instance,  $A$  itself.)

The length  $l(A)$  of an atomic statement  $A$  of L will be 1; that,  $l(\neg A)$ , of a negation  $\neg A$  of L will be  $l(A) + 1$ ; that,  $l(A \wedge B)$ , of a conjunction  $A \wedge B$  of L will be  $l(A) + l(B) + 1$ ; and that,  $l((\forall x)A)$ , of a quantification  $(\forall x)A$

<sup>4</sup>It follows from the account that (a) individual variables can occur only *bound* and (b) identical quantifiers cannot overlap in a statement of L. Because of (a) all statements of L are so-called *closed* statements. As regards (iv): When  $x$  does not occur in  $A$ ,  $(\forall x)A$  is known as a *vacuous* quantification.

<sup>5</sup>Statements of the sorts  $(A \rightarrow B), (A \vee B), (A \leftrightarrow B)$ , and  $(\exists x)A$  are, of course known as *conditionals*, *disjunctions*, *biconditionals*, and *existential quantifications*. The convention whereby  $(A \rightarrow B)$  is short for  $\neg(A \wedge \neg B)$  will be referred to by means of ‘ $D_{\rightarrow}$ ’, that whereby  $(A \vee B)$  is short for  $\neg(\neg A \wedge \neg B)$  by means of ‘ $D_{\vee}$ ’, etc.

<sup>6</sup>The (atomic) substatements of a *quantifierless* statement of L are also known as—and on page 56 will be called—its (atomic components).

of  $L$  will be  $l(A(T/x)) + 1$ , where  $T$  is any term of  $L$  you please.

Lastly, a term will be said to be *foreign* to a statement  $A$  of  $L$  if it does not occur in  $A$ ; to *occur* in a set  $S$  of statements of  $L$  if it occurs in at least one member of  $S$ ; and to be *foreign* to  $S$  if it is foreign to each member of  $S$ . and  $S$  will be held *infinitely extendible* in  $L$  if  $\aleph_0$  terms of  $L$  are foreign to  $S$ .

Borrowing from Quine [1940], Fitch [1948], Rosser [1953] and Leblanc [1979a], I shall take (i) the *axioms* of  $L$  to be all the statements of  $L$  of the following sorts:

- A1.  $A \rightarrow (A \wedge A)$
- A2.  $(A \wedge B) \rightarrow A$
- A3.  $(A \rightarrow B) \rightarrow (\neg(B \wedge C) \rightarrow \neg(C \wedge A))$
- A4.  $A \rightarrow (\forall x)A$
- A5.  $(\forall x)A \rightarrow A(T/x)$
- A6.  $(\forall x)(A \rightarrow B) \rightarrow ((\forall x)A \rightarrow (\forall x)B)$ ,

plus all those of the sort  $(\forall x)(A(x/T))$ , where  $A$  is an axiom of  $L$ , and (ii) the *potential* of two statements  $A$  and  $A \rightarrow B$  of  $L$  to be  $B$ . (Note as regards A4 that, with  $A \rightarrow (\forall x)A$  presumed here to be a statement,  $x$  is sure not to occur in  $A$ , and hence  $(\forall x)A$  is sure to be a vacuous quantification.)

It follows from this account of an axiom that:

THEOREM 1. *Every axiom of  $L$  is of the sort*

$$(\forall x_1)(\forall x_2) \dots (\forall x_n)(A(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n)),$$

where  $n \geq 0$  and  $A$  is one of the six sorts A1–A6;

and

THEOREM 2. *If  $(\forall x_1)(\forall x_2) \dots (\forall x_n)(A(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n))$  is an axiom of  $L$ , so for each  $i$  from 1 on is  $((\forall x_2) \dots (\forall x_n)(A(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n)))(t_i/x_1)$ .*

By a *proof* in  $L$  of a statement  $A$  of  $L$  from a set  $S$  of statements of  $L$ , I shall understand any finite column of statements of  $L$  such that (i) each entry in the column is a member of  $S$ , an axiom of  $L$ , or the potential of two earlier entries in the column, and (ii) the last entry in the column is  $A$ . I shall say that a statement  $A$  of  $L$  is *provable* in  $L$  from a set  $S$  of statements of  $L$ — $S \vdash A$ , for short—if there is a proof in  $L$  of  $A$  from  $S$ . I shall say that a statement  $A$  of  $L$  is *provable* in  $L$ — $\vdash A$ , for short—if  $\emptyset \vdash A$ . When all the axioms that turn up in a proof in  $L$  of a statement  $A$  of  $L$  are of the sorts

**A1–A3**, I shall say that  $A$  is *provable in L by means of just A1–A3*, and write ‘ $\vdash_0 A$ ’ in place of ‘ $\vdash A$ ’. And I shall say of a set  $S$  of statements of L, (i) when each statement of L provable in L from  $S$  belongs to  $S$ , that  $S$  is *deductively closed* in L or constitutes a *theory* of L (*in Tarski’s sense*),<sup>7</sup> (ii) when there is no (there is a) statement  $A$  of L such that  $S \vdash A \wedge \neg A$ , that  $S$  is (*in*)*consistent* in L; (iii) when  $S$  is consistent in L and—for any statement  $A$  of L not in  $S$ — $S \cup \{A\}$  is inconsistent in L, that  $S$  is *maximally consistent* in L, and (iv) when—no matter the quantification  $(\forall x)A$  of  $L$ — $S \vdash (\forall x)A$  if  $S \vdash A(T/x)$  for each term  $T$  of L, that  $S$  is  $\omega$ -*complete* in L.

Deductively closed sets (i.e. theories) will make one appearance below. Consistent, maximally consistent, and  $\omega$ -complete sets, on the other hand, will turn up everywhere and, hence, immediately rate a few extra words.

When a set  $S$  of statements of L is consistent and infinitely extendible in L, there is a way of extending  $S$  to a set—called here the *Henkin extension*  $\mathcal{H}(S)$  of  $S$  in L—that is maximally consistent and  $\omega$ -complete in L. The method, due essentially to Henkin, is as follows:

1. let  $S_0$  be  $S$  itself,
2.  $A_n$  being for each  $n$  from 1 on the alphabetically  $n$ th statement of L,

$$\text{let } S_n \text{ be } \begin{cases} S_{n-1} \text{ if } S_{n-1} \cup \{A_n\} \text{ is inconsistent in L} \\ S_{n-1} \cup \{A_n\} \text{ if } S_{n-1} \cup \{A_n\} \text{ is consistent in L and } A_n \\ \text{is not a negated universal quantification of L} \\ S_{n-1} \cup \{A_n, \neg B(T/x)\}, \text{ where } T \text{ is the alphabetically} \\ \text{earliest term of L foreign to } S_{n-1} \cup \{A_n\}, \text{ if} \\ S_{n-1} \cup \{A_n\} \text{ is consistent in L and } A_n \text{ is a negated} \\ \text{quantification } \neg(\forall x)B \text{ of L,}^8 \end{cases}$$

and

3. let  $\mathcal{H}(S)$  be  $\cup_{i=0}^{\infty} S_i$ .

(Henkin’s original instructions for extending a set appeared in [Henkin, 1949], a seminal paper for substitutional and truth-value semantics. I avail myself here of simplifications to those instructions due to Hasenjaeger and Henkin himself—see [Smullyan, 1968, pp. 93–97] on this matter. Lindenbaum had already shown in the late Twenties how to extend a consistent set into a maximal one, thus paving the way for Henkin’s result—see [Tarski, 1930, Section 7].)

Proofs of the following theorems are in numerous texts (e.g. [Leblanc, 1976, pp. 38–40]) and will be taken for granted. (To abridge things I write ‘iff’ for ‘if, and only if’.)

<sup>7</sup>Tarski himself talked of *deductive systems* rather than *theories* (see Tarski [1930; 1935–36]); but the application ‘theory in Tarski’s sense’ has prevailed.

<sup>8</sup>With  $S$  presumed to be infinitely extendible in L, there is sure to be—however large the  $n$ —a term  $T$  of L foreign to  $S_{n-1} \cup \{A_n\}$ ; not so, otherwise.

**THEOREM 3.** *If a set  $S$  of statements of  $L$  is maximally consistent in  $L$ , then (i) a negation  $\neg A$  of  $L$  belongs to  $S$  iff  $A$  does not, and (ii) a conjunction  $A \wedge B$  of  $L$  belongs to  $S$  iff each of  $A$  and  $B$  does. Further, if  $S$  is  $\omega$ -complete in  $L$  as well, then (iii) a quantification  $(\forall x)A$  of  $L$  belongs to  $S$  iff each substitution instance  $A(T/x)$  of  $(\forall x)A$  in  $L$  does.*

**THEOREM 4.** *If a set  $S$  of statements of  $L$  is consistent and infinitely extendible in  $L$ , then the Henkin extension  $\mathcal{H}(S)$  of  $S$  in  $L$  is maximally consistent and  $\omega$ -complete in  $L$  (= Henkin's Extension Lemma).*

With Dunn and Belnap I shall understand by a *term extension* of  $L$  any language that is exactly like  $L$  except for having *countably* many terms—i.e. finitely many or  $\aleph_0$  many terms—besides those of  $L$ . Note that by this definition  $L$ —having zero and, hence, finitely many terms besides its own—is one of its term extensions. I shall refer to the term extensions of  $L$  by means of ' $L^+$ '. And,  $L^+$  being an arbitrary term extension of  $L$ , I shall assume that the terms of  $L^+$  come in some *alphabetic order*, refer to the alphabetically  $i$ th ( $i = 1, 2, 3, \dots$ ) of them by means of ' $t_i^+$ ', and write ' $S \vdash^+ A$ ' for ' $A$  is provable in  $L^+$  from  $S$ '.

It will prove convenient, when  $(\forall x)A$  is a quantification of  $L$ , to talk of the substitution instances of  $(\forall x)A$  *in any term extension  $L^+$  of  $L$  one pleases*, not just in  $L$  itself. These will of course be

$$A(t_1^+/x), A(t_2^+/x), A(t_3^+/x), \dots$$

a list which in the case that  $L^+$  is  $L$  boils down to  $A(t_1/x), A(t_2/x), A(t_3/x), \dots$ , but otherwise will sport fresh entries—the substitution instances of  $(\forall x)A$  *peculiar to  $L^+$* . It will also prove convenient, when  $S$  is a set of statements of  $L$ , to talk of  $S$  as being (or failing to be) infinitely extendible, consistent, maximally consistent, and  $\omega$ -complete *in any term extension  $L^+$  of  $L$  one pleases*.  $S$ , for example, will be held infinitely extendible in  $L^+$  if  $\aleph_0$  terms of  $L^+$  are foreign to  $S$ , consistent in  $L^+$  if there is no statement  $A$  of  $L^+$  such that  $S \vdash^+ A \wedge \neg A$ , etc. And it will prove convenient, when  $S$  is as above, to talk of the Henkin extension of  $S$  *in any term extension of  $L$  one pleases*.

One term extension of  $L$  will play a special role below. It will have  $\aleph_0$  terms besides those of  $L$ , and for that reason will be known as  $L^\infty$ .

The  $\aleph_0$  terms of  $L^\infty$  *peculiar to  $L^\infty$*  are foreign of course to any set of statements of  $L$ . Hence:

**THEOREM 5.** *Each set of statements of  $L$  is infinitely extendible in  $L^\infty$ .*

Note further that:

**THEOREM 6.** *If  $S \vdash^\infty A$ , where  $S$  is a set of statements of  $L$  and  $A$  a statement of  $L$ , then  $S \vdash A$ ,*

and hence:



THEOREM 7. *If a set  $S$  of statements of  $L$  is consistent in  $L$ , then  $S$  is consistent in  $L^\infty$ .*

For a proof of Theorem 6, suppose the column made up of  $B_1, B_2, \dots, B_p$  constitutes a proof in  $L^\infty$  of  $A$  from  $S$ , and for each  $i$  from 1 through  $p$  let  $C_i$  be the result of putting  $t_1$  (the alphabetically first term of  $L$ ) for every term in  $B_i$  that is peculiar to  $L^\infty$ . The column made up of  $C_1, C_2, \dots, C_p$  will constitute a proof in  $L$  of  $A$  from  $S$ . For a proof of Theorem 7, suppose  $S$  is inconsistent in  $L^\infty$ . Then by a familiar result  $S \vdash^\infty A$  for every statement  $A$  of  $L^\infty$ ; hence, by Theorem 6,  $S \vdash A$  for every statement  $A$  of  $L$ ; and, hence,  $S$  is inconsistent in  $L$ . Hence, Theorem 7 by Contraposition.

Now for one kind of standard semantics that  $L$  might be outfitted with.<sup>9</sup>

Understand by a *domain* any non-empty set. Given a domain  $D$ , understand by a  *$D$ -interpretation* of (the terms and predicates of)  $L$  any result of assigning to each term of  $L$  a member of  $D$  and to each predicate of  $L$  of degree  $d$  ( $d = 1, 2, \dots$ ) a subset of  $D^d$  ( $D^d$  the Cartesian product of  $D$  with itself  $d$  times);  $D$  being a domain,  $I_D$  a  $D$ -interpretation of  $L$ , and  $T$  a term of  $L$ , understand by a  *$T$ -variant* of  $I_D$  any  $D$ -interpretation of  $L$  that is like  $I_D$  except for possibly assigning to  $T$  a member of  $D$  different from that assigned by  $I_D$ ; and understand by a *model* for  $L$  any pair of the sort  $\langle D, I_D \rangle$ , where  $D$  is a domain and  $I_D$  is a  $D$ -interpretation of  $L$ . A model  $\langle D, I_D \rangle$  will be termed *finite* if  $D$  is, *denumerably infinite* if  $D$  is *countable* if  $D$  is, etc. (Many writers, Hodges among them, talk of *structures* where I talk of models.)

This done, let  $\langle D, I_D \rangle$  be a model for  $L$ ,  $A$  a statement of  $L$ , and  $S$  a set of statements of  $L$ . I shall say that  $A$  is *true in  $\langle D, I_D \rangle$*  (for short, *true on  $I_D$* ) if (i) in the case that  $A$  is an atomic statement  $Q(T_1, T_2, \dots, T_d)$ , the  $d$ -tuple  $\langle I_D(T_1), I_D(T_2), \dots, I_D(T_d) \rangle$  belongs to  $I_D(Q)$ , (ii) in the case that  $A$  is a negation  $\neg B$ ,  $B$  is not true in  $\langle D, I_D \rangle$ , (iii) in the case that  $A$  is a conjunction  $B \wedge C$ , each of  $B$  and  $C$  is true in  $\langle D, I_D \rangle$ , and (iv) in the case that  $A$  is a quantification  $(\forall x)B$ ,  $B(T/x)$ — $T$  here any term of  $L$  *foreign* to  $A$ —is true in  $\langle D, I'_D \rangle$  for each  $T$ -variant  $I'_D$  of  $I_D$ ; and I shall say that  $S$  is *true in  $\langle D, I_D \rangle$*  if each member of  $S$  is. (Note: (ii)–(iv) are known and will be referred to as the *truth-conditions* for negations, conjunctions, and universal quantifications; as indicated in note 2, (iv) embodies the *objectual* interpretation of ‘All’.)

This done, I shall declare a statement  $A$  of  $L$  *logically true in the standard sense* if  $A$  is true in every model for  $L$ ; and, where  $S$  is a set of statements

<sup>9</sup>Standard (i.e. model-theoretic) semantics comes in various brands. The present brand (an early version of which appeared in [Leblanc and Wisdom, 1993]) especially suits languages such as  $L$  whose statements are all closed; the one propounded in [Tarski, 1936] and used in quite a number of recent texts, Hodges among them, especially suits languages with both open and closed statements. The beginnings of standard semantics are chronicled in [Goldfarb, 1979].

of  $L$ , I shall declare  $A$  *logically entailed by  $S$  in the standard sense* if  $A$  is true in every model for  $L$  in which  $S$  is true. Under these definitions,  $A$  is logically true in the standard sense iff logically entailed by  $\emptyset$  in the standard sense, the point made in note 3.

When a statement  $A$  (a set  $S$  of statements) of  $L$  is true in a model  $\langle D, I_D \rangle$ , (i)  $A(S)$  is often said *to have  $\langle D, I_D \rangle$  as a model* and (ii)  $\langle D, I_D \rangle$  is said *to be a model of  $A(S)$* . Given the second of these locutions,  $A$  is logically entailed by  $S$  if every model of  $S$  is one of  $A$ .

Prominent in Section 2 will be Henkin  $D$ -interpretations and Henkin models.  $I_D$  will constitute a *Henkin  $D$ -interpretation* of  $L$  if *each* member of  $D$  is assigned by  $I_D$  to a term of  $L$  (more formally, if for each  $d$  in  $D$  there is a term  $T$  of  $L$  such that  $I_D(T) = d$ ); and  $\langle D, I_D \rangle$  will constitute a *Henkin model* for  $L$  if  $I_D$  is a Henkin  $D$ -interpretation of  $L$ . Since  $L$  has only  $\aleph_0$  terms, the domain  $D$  must be countable in each case. So Henkin models are countable by definition. (The telling use to which Henkin  $D$ -interpretations and models were put in [Henkin, 1949] accounts for their names, bestowed upon them in [Leblanc and Wisdom, 1993].)

The foregoing definitions of a  $D$ -interpretation, a model, a Henkin  $D$ -interpretation, and a Henkin model are easily generalised to suit *any* term extension  $L^+$  of  $L$  (rather than just  $L$  itself): write ' $L^+$ ' everywhere for ' $L$ '. Since  $L^+$ —like  $L$  itself—has only  $\aleph_0$  terms, Henkin models for  $L^+$  are countable by definition.

## 2 SUBSTITUTION IN STANDARD SEMANTICS AND SUBSTITUTIONAL SEMANTICS

Substitutional semantics was characterised on page 53 as a semantics that interprets 'All' and 'Some' substitutionally and, hence, can own only Henkin models. It might also be characterised as a semantics that owns only Henkin models, and hence can interpret 'All' and 'Some' substitutionally. But, whichever characterisation can be favoured, substitutional semantics is an elaboration of the old, yet resilient, dictum: "A universal quantification is true iff all its substitution instances are (and an existential one true iff at least one of them is)."

The dictum, suitably rephrased, appears below as Theorem 13 and will be called the *Substitution Theorem*. Devotees of standard semantics often slight it—*unaccountably so*. Indeed, *their* proof of the completeness theorem

*A, if logically entailed by  $S$  in the standard sense, is provable  
from  $S$  in  $L$  (= Theorem 21)*

appeals to Theorem 13; and, slightly reworded, the half of that proof concerning infinitely extendible  $S$ 's yields the counterpart of Theorem 21 in substitutional semantics. Further, *their* proof of the soundness theorem

*A, if provable from S in L, is logically entailed by S in the standard sense (= Theorem 11)*

appeals to half of Theorem 13, a half recorded below as Theorem 9; and, reworded to concern all Henkin models for all term extensions of L, that proof yields the counterpart of Theorem 11 in substitutional semantics. So, Theorem 21 and Theorem 11, the theorems which together legitimise the standard account of logical entailment, hinge upon a substitution lemma; and editing half the proof of one theorem and all the proof of the other will legitimise the substitutional account of that notion.

Concerned to show substitutional semantics a by-product of standard semantics, I first substantiate the above claims, and stress as I go along the role that Henkin models play in standard semantics. Then I spell out and justify what substitutional semantics understands by logical truth and logical entailment. Last I comment on the substitutional handling of theories.

Demonstrating Theorem 11 is easy once you have shown that the axioms of L are true in all models for L. Showing that the axioms of L are true in all models for L is relatively easy once you have shown that those of sorts **A1–A6** are. And showing that the axioms of L of sorts **A1–A4** and **A6** are true in all models for L is quite easy. (In the case of **A4**, refer to [Leblanc, 1976, Theorem 4.1.2], which guarantees that if  $A$  is true on a  $D$ -interpretation  $I_D$  of L, then  $A$  is true on any  $T$ -variant of  $I_D$ .  $T$  here is an arbitrary term of L foreign to  $(\forall x)A$ .) But showing that the axioms of L of sort **A5** are true in all models for L calls for a lemma whose proof is surprisingly difficult. (Indeed, few texts undertake to prove Theorem 8. The proof here is borrowed from Leblanc and Wisdom [1993, pp. 313–314] and Leblanc [1976, pp. 86–88], with several simplifications and corrections due to Wisdom.)

**THEOREM 8.** *Let  $A$  be a statement of L,  $T$  and  $T_i$  be terms of L,  $D$  be a domain,  $I_D$  be a  $D$ -interpretation of L, and  $I'_D$  be the  $T$ -variant of  $I_D$  such that  $I'_D(T) = I_D(T')$ . Then  $A(T'/T)$  is true on  $I_D$  iff  $A$  is true on  $I'_D$ .*

**Proof** of Theorem 8 is by mathematical induction on the length  $l(A)$  of  $A$ .

*Basis:*  $l(A) = 1$ . Then  $A$  is of the sort  $Q(T_1, T_2, \dots, T_d)$ , and hence  $A(T'/T)$  is of the sort  $Q(T'_1, T'_2, \dots, T'_d)$ , where—for each  $i$  from 1 through  $d$ — $T'_i$  is  $T_i$  itself if  $T_i$  is distinct from  $T$ , otherwise  $T'_i$  is  $T'$ . But, by the construction of  $I'_D$ ,  $I'_D(T_i) = I_D(T'_i)$  for each  $i$  from 1 through  $d$ , and  $I'_D(Q) = I_D(Q)$ . Hence,  $\langle I'_D(T_1), I'_D(T_2), \dots, I'_D(T_d) \rangle$  belongs to  $I'_D(Q)$  iff  $\langle I_D(T'_1), I_D(T'_2), \dots, I_D(T'_d) \rangle$  belongs to  $I_D(Q)$ . Hence  $A(T'/T)$  is true on  $I_D$  iff  $A$  is true on  $I'_D$ .

*Inductive Step:*  $l(A) > 1$ .

*Case 1.*  $A$  is a negation  $\neg B$ . By the hypothesis of the induction  $B(T'/T)$  is (not) true on  $I_D$  iff  $B$  is (not) true on  $I'_D$ . Hence,  $\neg(B(T'/T))$ , i.e.  $(\neg B)(T'/T)$ , is true on  $I_D$  iff  $\neg B$  is true on  $I'_D$ .

*Case 2.*  $A$  is a conjunction  $B \wedge C$ . Proof similar to that of Case 1.

*Case 3.*  $A$  is a quantification  $(\forall x)B$ . (i) Suppose  $((\forall x)B)(T'/T)$ —i.e.  $(\forall x)(B(T'/T))$ —is not true on  $I_D$ . Then, with  $T''$  an arbitrary term of  $L$  foreign to  $(\forall x)(B(T'/T))$  and distinct from  $T$ , there is a  $T''$ -variant  $I_D''$  of  $I_D$  on which  $(B(T'/T))(T''/x)$  is not true. But, given the hypothesis on  $T''$ ,  $(B(T''/x))(T'/T)$  is the same as  $(B(T'/T))(T''/x)$ . Hence,  $(B(T''/x))(T'/T)$  is not true on  $I_D''$ . Now let  $I_D'''$  be the  $T$ -variant of  $I_D''$  such that  $I_D'''(T) = I_D''(T')$ . Then, by the hypothesis of the induction,  $B(T''/x)$  is not true on  $I_D'''$ . But  $I_D'''$  is a  $T''$ -variant of  $I_D''$ , a point I demonstrates two lines hence. So, there is a  $T''$ -variant of  $I_D''$  on which  $B(T''/x)$  is not true. So,  $(\forall x)B$  is not true on  $I_D$ . (For proof that  $I_D'''$  is a  $T''$ -variant of  $I_D''$ : Since  $I_D'''$  is a  $T$ -variant of  $I_D''$  and  $I_D''$  is a  $T''$ -variant of  $I_D$ ,  $I_D'''$  and  $I_D$  can differ only on  $T$ . So,  $I_D'''$  and  $I_D$  can differ only on  $T''$ .) (ii) Suppose  $(\forall x)B$  is not true on  $I_D$ . Then, with  $T''$  a term of  $L$  foreign to  $(\forall x)B$  and distinct from  $T$ , there is a  $T''$ -variant  $I_D''$  of  $I_D$  on which  $B(T''/x)$  is not true. Now let  $I_D'''$  be the  $T''$ -variant of  $I_D$  such that  $I_D'''(T'') = I_D''(T'')$ . Then  $I_D'''$  is the  $T$ -variant of  $I_D''$  such that  $I_D'''(T) = I_D''(T')$ , a point I demonstrate four lines hence. So, by the hypothesis of the induction  $(B(T''/x))(T'/T)$ —i.e.  $(B(T'/T))(T''/x)$ —is not true on  $I_D'''$ . So, there is a  $T''$ -variant of  $I_D$  on which  $(B(T'/T))(T''/x)$  is not true. So,  $(\forall x)(B(T'/T))$ —i.e.  $((\forall x)B)(T'/T)$ —is not true on  $I_D$ . (For proof that  $I_D'''$  is the  $T$ -variant of  $I_D''$  such that  $I_D'''(T) = I_D''(T')$ : Since  $I_D'''$  is a  $T''$ -variant of  $I_D''$  and  $I_D''$  is a  $T$ -variant of  $I_D$ ,  $I_D'''$  and  $I_D$  can differ only on  $T''$  and  $T$ . But  $I_D'''$ , being a  $T''$ -variant of  $I_D$ , can differ from  $I_D$  only on  $T''$ . But  $I_D'''(T'') = I_D''(T'')$ . So,  $I_D'''$  is the  $T$ -variant of  $I_D''$  such that  $I_D'''(T) = I_D''(T')$ .) ■

Hence, the following half of Theorem 13 which guarantees at once that every axiom of  $L$  of the sort  $(\forall x)A \rightarrow A(T/x)$  (**A5**) is true in all models for  $L$ :

**THEOREM 9.** *Let  $D$  be a domain and  $I_D$  be a  $D$ -interpretation of  $L$ . If  $(\forall x)A$  is true on  $I_D$ , so is each substitution instance  $A(T/x)$  of  $(\forall x)A$  in  $L$ .*

**Proof.** Suppose  $(\forall x)A$  is true on  $I_D$ , and let  $T'$  be an arbitrary term of  $L$  foreign to  $(\forall x)A$ . Then  $A(T'/x)$  is true on the  $T'$ -variant  $I_D'$  of  $I_D$  such that  $I_D'(T') = I_D(T)$  and, hence, by Theorem 8  $(A(T'/x))(T/T')$  is true on  $I_D$ . But, with  $T'$  foreign to  $(\forall x)A$ ,  $A(T/x)$  is the same as  $(A(T'/x))(T/T')$ . Hence,  $A(T/x)$  is true on  $I_D$ . ■

Proof can now be given that:

**THEOREM 10.** *Every axiom of  $L$  is true in all models for  $L$ .*

**Proof.** Suppose  $A$  is an axiom of  $L$ , in which case  $A$  is bound by Theorem 1 to be of the sort

$$(\forall x_1)(\forall x_2) \dots (\forall x_n)(B(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n)),$$

where  $n \geq 0$  and  $B$  is one of the sorts **A1–A6**. Proof that  $A$  is true in all models for  $L$  will be by mathematical induction on  $n$ .

*Basis:*  $n = 0$ . Then  $A(= B)$  is one of the sorts **A1–A6**, a case treated above.

*Inductive Step:*  $n > 0$ . Let  $\langle D, I_D \rangle$  be an arbitrary model for  $L$ , and  $T$  be an arbitrary term of  $L$  foreign to  $A$ .

$$((\forall x_2) \dots (\forall x_n)(B(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n)))(T/x_1)$$

is bound by Theorem 2 to be an axiom of  $L$ , and hence by the hypothesis of the induction to be true in  $\langle D, I'_D \rangle$  for every  $T$ -variant  $I'_D$  of  $I_D$ . Hence,  $A$  is true in  $\langle D, I_D \rangle$ . Hence,  $A$  is true in all models for  $L$ . ■

And, with Theorem 10 on hand, proof can be given that:

**THEOREM 11.** *Let  $S$  be an arbitrary set of statements and  $A$  be an arbitrary statement of  $L$ . If  $S \vdash A$ , then  $A$  is true in every model for  $L$  in which  $S$  is true, i.e.  $A$  is logically entailed by  $S$  in the standard sense (= The Strong Soundness Theorem for  $L$  in Standard Semantics).*

**Proof.** Suppose the column made up of  $B_1, B_2, \dots, B_p$  constitutes a proof in  $L$  of  $A$  from  $S$ , and let  $\langle D, I_D \rangle$  be an arbitrary model for  $L$  in which  $S$  is true. It is easily shown by mathematical induction on  $i$  that, for each  $i$  from 1 through  $p$ ,  $B_i$  is true in  $\langle D, I_D \rangle$ . Indeed, when  $B_i$  is a member of  $S$ ,  $B_i$  is true in  $\langle D, I_D \rangle$  by the hypothesis on  $\langle D, I_D \rangle$ ; when  $B_i$  is an axiom of  $L$ ,  $B_i$  is true in  $\langle D, I_D \rangle$  by Theorem 10; and, when  $B_i$  is the ponential of, say,  $B_g$  and  $B_h$ ,  $B_i$  is sure to be true in  $\langle D, I_D \rangle$  if—as the hypothesis of the induction guarantees— $B_g$  and  $B_h$  are. Hence,  $B_p(= A)$  is true in  $\langle D, I_D \rangle$ . Hence,  $A$  is true in every model for  $L$  in which  $S$  is true. Hence, Theorem 11. ■

Because of Theorem 11,  $A$ —when provable in  $L$  from  $S$ —is of course sure to be true in every *countable* model for  $L$  in which  $S$  is true, and—more particularly—to be true in every *Henkin* model for  $L$  in which  $S$  is true. With an eye to proving the counterpart of Theorem 11 for substitutional semantics, I record the second of these corollaries as a separate theorem:

**THEOREM 12.** *Let  $S$  and  $A$  be as in Theorem 11. If  $S \vdash A$ , then  $A$  is true in every Henkin model for  $L$  in which  $S$  is true.*

Theorem 11, as previously noted, is one of the two theorems that legitimise the standard account of logical entailment. Now for the other, the converse of Theorem 11. Proof of it for the case where  $S$  is infinitely extendible in  $L$  uses one extra notion, that of a model associate, and two extra lemmas, Theorems 13 and 14.

Let  $S$  be an arbitrary set of statements of  $L$ . By the *model associate* of  $S$  in  $L$ , I shall understand the pair  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$ , where (i)  $\mathcal{T}$  consists of the  $\aleph_0$  terms of  $L$ , (ii) for each term  $T$  of  $L$ ,  $I_{\mathcal{T}}(T) = T$ , and (iii) for each predicate  $Q$  of  $L$  of degree  $d$  ( $d = 1, 2, 3, \dots$ )  $I_{\mathcal{T}}(Q) = \{ \langle T_1, T_2, \dots, T_d \rangle : Q(T_1, T_2, \dots, T_d) \in S \}$ .  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$  is a Henkin model, and one of particular significance. Theorem 13 is what I called on p. 61 the *Substitution Theorem (for L)*. Together with Theorem 3, it yields Theorem 14, a result guaranteeing that if a set  $S$  of statements of  $L$  is maximally consistent and  $\omega$ -complete in  $L$ , then a statement of  $L$  will be true in the model associate of  $S$  in  $L$  iff it belongs to  $S$ . And Theorem 14 yields in turn the converse of Theorem 11 for infinitely extendible  $S$ , this in a mere seven lines.

**THEOREM 13.** *Let  $I_D$  be a Henkin  $D$ -interpretation of  $L$ . Then a quantification  $(\forall x)A$  of  $L$  is true on  $I_D$  iff each substitution instance  $A(T/x)$  of  $(\forall x)A$  in  $L$  is true on  $I_D$  (= The Substitution Theorem for  $L$ ).*

**Proof.** Suppose  $A(T/x)$  is true on  $I_D$  for every term  $T$  of  $L$ ; let  $T'$  be an arbitrary term of  $L$  foreign to  $(\forall x)A$ ; let  $I'_D$  be an arbitrary  $T'$ -variant of  $I_D$ ; and let  $d$  be the value of  $I'_D$  for  $T'$ .  $I_D$  being a Henkin  $D$ -interpretation of  $L$ , there is sure to be a term  $T''$  of  $L$  such that  $I_D(T'') = d$  and, hence,  $I_D(T'') = I'_D(T')$ . But, since  $A(T/x)$  is true on  $I_D$  for every term  $T$  of  $L$  and  $(A(T'/x))(T/T')$  is the same as  $A(T/x)$ ,  $(A(T'/x))(T''/T')$  is sure to be true on  $I_D$ . Hence, by Theorem 8  $A(T'/x)$  is true on  $I'_D$ . So, if  $A(T/x)$  is true on  $I_D$  for every term  $T$  of  $L$ , then  $A(T'/x)$  is true on every  $T'$ -variant of  $I_D$ , and hence  $(\forall x)A$  is true on  $I_D$ . Hence, Theorem 13 by Theorem 9. ■

**THEOREM 14.** *Let  $S$  be a set of statements of  $L$  that is maximally consistent and  $\omega$ -complete in  $L$ , and let  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$  be the model associate of  $S$  in  $L$ . Then a statement  $A$  of  $L$  belongs to  $S$  iff  $A$  is true in  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$ .*

**Proof.** of Theorem 14 is by mathematical induction on the length  $l(A)$  of  $A$ .

*Basis:*  $l(A) = 1$ , in which case  $A$  is of the sort  $Q(T_1, T_2, \dots, T_d)$ . By the construction of  $I_{\mathcal{T}}$ , (i)  $Q(T_1, T_2, \dots, T_d)$  belongs to  $S$  iff  $\langle T_1, T_2, \dots, T_d \rangle$  belongs to  $I_{\mathcal{T}}(Q)$  and (ii)  $\langle T_1, T_2, \dots, T_d \rangle$  belongs to  $I_{\mathcal{T}}(Q)$  iff  $\langle I_{\mathcal{T}}(T_1), I_{\mathcal{T}}(T_2), \dots, I_{\mathcal{T}}(T_d) \rangle$  does. Hence, Theorem 14.

*Inductive Step:*  $l(A) > 1$ . Proof of Theorem 14 in the case that  $A$  is a negation  $\neg B$  or a conjunction  $B \wedge C$  is by Theorem 3 (i)–(ii), the hypothesis of the induction, and the truth-conditions for ‘ $\neg$ ’ and ‘ $\wedge$ ’. Suppose then that  $A$  is a universal quantification  $(\forall x)B$ . By Theorem 3 (iii)  $(\forall x)B$  belongs to  $S$  iff each substitution instance  $B(T/x)$  of  $(\forall x)B$  in  $L$  does, hence by the hypothesis of the induction iff each substitution instance  $B(T/x)$  of  $(\forall x)B$  in  $L$  is true in  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$ , hence by Theorem 13 (and the fact that  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$  is a Henkin model) iff  $(\forall x)B$  is true in  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$ . ■

Suppose now that  $S \not\vdash A$ , where  $S$  is infinitely extendible in  $L$ .<sup>10</sup> Then by a familiar result  $S \cup \{\neg A\}$ , a set also infinitely extendible in  $L$ , is consistent in  $L$ ; hence, by Theorem 4 the Henkin extension  $\mathcal{H}(S \cup \{\neg A\})$  of  $S \cup \{\neg A\}$  in  $L$  is maximally consistent and  $\omega$ -complete in  $L$ ; and, hence by Theorem 14 every member of  $\mathcal{H}(S \cup \{\neg A\})$  is true in the model associate of  $\mathcal{H}(S \cup \{\neg A\})$  in  $L$ . But  $S \cup \{\neg A\}$  is a subset of  $\mathcal{H}(S \cup \{\neg A\})$ . Hence, (every member of)  $S$  is true in that model, but  $A$  is not. Hence:

**THEOREM 15.** *Let  $S$  be a set of statements of  $L$  that is infinitely extendible in  $L$ , and let  $A$  be an arbitrary statement of  $L$ . If  $S \not\vdash A$ , then there is a model for  $L$  in which  $S$  is true but  $A$  is not.*

Hence by Contraposition:

**THEOREM 16.** *Let  $S$  and  $A$  be as in Theorem 15. If  $A$  is true in every model for  $L$  in which  $S$  is true, i.e. if  $A$  is logically entailed by  $S$  in the standard sense, then  $S \vdash A$ .*

Since the model associate of  $\mathcal{H}(S \cup \{\neg A\})$  in  $L$  is a Henkin model for  $L$ , Theorems 15 and 16 can be sharpened to read:

**THEOREM 17.** *Let  $S$  and  $A$  be as in Theorem 15. If  $S \not\vdash A$ , then there is a Henkin model for  $L$  (and hence, there is a countable model for  $L$ ) in which  $S$  is true but  $A$  is not.*

**THEOREM 18.** *Let  $S$  and  $A$  be as in Theorem 15. If  $A$  is true in every Henkin model for  $L$  in which  $S$  is true, then  $S \vdash A$ .*

$\emptyset$  is, of course, infinitely extendible in  $L$ . So by Theorems 11, 12, 16 and 18:

**THEOREM 19.** *Let  $A$  be an arbitrary statement of  $L$ .*

- (a) *If  $\vdash A$ , then  $A$  is true in every model for  $L$ , i.e.  $A$  is logically true in the standard sense (= The Weak Soundness Theorem for  $L$  in Standard Semantics).*
- (b) *If  $A$  is true in every model for  $L$ , i.e. if  $A$  is logically true in the standard sense, then  $\vdash A$  (= The Weak Completeness Theorem for  $L$  in Standard Semantics).*
- (c)  *$\vdash A$  iff  $A$  is true in every Henkin model for  $L$ .*
- (d)  *$A$  is true in every model for  $L$  if—and, hence iff—true in every Henkin model for  $L$ .*

Theorem 19 (a)–(b) legitimises the standard account of logical truth.

<sup>10</sup>' $S \not\vdash A$ ' is short here for ' $A$  is not provable from  $S$  in  $L$ .' The proof I embark upon is essentially due to Henkin, and for this reason is often called *Henkin's Completeness Proof*.

The argument leading to Theorem 15 and hence to theorems 16–18, fails when  $S$  is *not* infinitely extendible in  $L$ .  $S \cup \{\neg A\}$ , if consistent in  $L$ , will extend to a set that is maximally consistent in  $L$ , but it need not extend to one that is *both* maximally consistent and  $\omega$ -complete in  $L$ . A set of the sort  $\{Q(t_1), Q(t_2), Q(t_3), \dots, \neg(\forall x)Q(x)\}$ , for example, will clearly not extend to one maximally consistent and  $\omega$ -complete in  $L$ . In point of fact, for many an  $S$  that is not infinitely extendible in  $L$ , Theorem 17 fails: for many an  $A$  such that  $S \not\vdash A$ , *there is no Henkin model for  $L$  in which  $S$  is true but  $A$  is not*. To exploit the same counterexample as above, though  $\{Q(t_1), Q(t_2), Q(t_3), \dots\} \vdash (\forall x)Q(x)$ , there is no Henkin model for  $L$  in which  $\{Q(t_1), Q(t_2), Q(t_3), \dots\}$  is true but  $(\forall x)Q(x)$  is not.<sup>11</sup> However, it can be shown that if  $S \not\vdash A$ , where  $S$  is not infinitely extendible in  $L$ , then there is a *non-Henkin* model for  $L$  (more specifically, a denumerably infinite non-Henkin model for  $L$ ) in which  $S$  is true but  $A$  is not.

Understand (i) by the *double rewrite*  $A^2$  of a statement  $A$  of  $L$  the result of substituting everywhere in  $At_{2i}$  for  $t_i$  ( $i = 1, 2, 3, \dots$ ), (ii) by the double rewrite  $S^2$  of a set  $S$  of statements of  $L$  the set  $\emptyset$  when  $S$  is empty, otherwise that consisting of the double rewrites of the various members of  $S$ , (iii) by the double rewrite of a  $D$ -interpretation  $I_D$  of  $L$  the  $D$ -interpretation  $I_D^2$  of  $L$  such that  $I_D^2(Q) = I_D(Q)$  for every predicate  $Q$  of  $L$ , and  $I_D^2(t_i) = I_D(t_{2i})$  for each  $i$  from 1 on, and (iv) by the double rewrite of a model  $\langle D, I_D \rangle$  for  $L$  the model  $\langle D, I_D^2 \rangle$ , where  $I_D^2$  is the double rewrite of  $I_D$ . It is easily verified that (a) whether or not  $S$  is infinitely extendible,  $S^2$  is, (b) if  $S$  is consistent in  $L$ , so is  $S^2$ , (c) if the double rewrite of a statement  $A$  of  $L$  is true in a model  $\langle D, I_D \rangle$  for  $L$ , then  $A$  itself is true in the double rewrite of  $\langle D, I_D \rangle$ , and (d) although the model associate  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$  in  $L$  of a set of statements of  $L$  is a Henkin model for  $L$ , the double rewrite  $\langle \mathcal{T}, I_{\mathcal{T}}^2 \rangle$  of  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$  is not (members  $t_1, t_3, t_5, \dots$ , of  $\mathcal{T}$  are not assigned in  $I_{\mathcal{T}}^2$  to any term of  $L$ ).

Suppose now that  $S \not\vdash A$ , where  $S$  is not infinitely extendible in  $L$ , and hence  $S \cup \{\neg A\}$  is consistent in  $L$ . By (a)–(b), the double rewrite  $S^2 \cup \{\neg A^2\}$  of  $S \cup \{\neg A\}$  is both consistent and infinitely extendible in  $L$ . Hence, as before,  $S^2 \cup \{\neg A^2\}$  is true in the model associate in  $L$  of  $\mathcal{H}(S^2 \cup \{\neg A^2\})$ . Hence, by (c)  $S$  is true in the double rewrite of that model but  $A$  is not. Hence, there is a model for  $L$  in which  $S$  is true but  $A$  is not.<sup>12</sup> That the model in question is *not* a Henkin one follows from (d).

Hence

<sup>11</sup>The counterexample in the text was noted in [Thomason, 1965]. I learned of it from Thomason and reported it in [Leblanc, 1968]. Dunn and Belnap, who had known of the counterexample for several years, reported it in [Dunn and Belnap, 1968].

<sup>12</sup>The result is proved in surprisingly few texts. The proof here stems from [Leblanc, 1966, pp. 177–78], but avoids that text's recourse to  $L^\infty$ . To illustrate matters,  $\{Q(t_1), Q(t_2), Q(t_3), \dots\}$  is true—but  $(\forall x)Q(x)$  is *not*—in  $\langle \mathcal{T}, I_{\mathcal{T}}^2 \rangle$ , the double rewrite of the model associate  $\langle \mathcal{T}, I_{\mathcal{T}} \rangle$  of  $\mathcal{H}(\{Q(t_2), Q(t_4), Q(t_6), \dots, \neg(\forall x)Q(x)\})$ .  $I_{\mathcal{T}}$  assigns  $t_i$  to  $t_i$  and, hence,  $I_{\mathcal{T}}^2$  assigns  $t_{2i}$  to  $t_i$ ; but both  $I_{\mathcal{T}}$  and  $I_{\mathcal{T}}^2$  assign  $\mathcal{T}$  to  $Q$ .



**THEOREM 20.** *Let  $S$  be a set of statements of  $L$  that is not infinitely extendible in  $L$ , and let  $A$  be an arbitrary statement of  $L$ . If  $S \not\vdash A$ , then there is a model for  $L$  (more specifically, a countable model for  $L$ ) in which  $S$  is true but  $A$  is not.*

Hence, by Contraposition on Theorem 20 and (for the case where  $S$  is infinitely extendible in  $L$ ) on Theorem 15:

**THEOREM 21.** *Let  $S$  be an arbitrary set of statements of  $L$  and  $A$  be an arbitrary statement of  $L$ . If  $A$  is true in every model for  $L$  in which  $S$  is true, i.e. if  $A$  is logically entailed by  $S$  in the standard sense, then  $S \vdash A$  (= The Strong Completeness Theorem for  $L$  in Standard Semantics).*

The argument yielding Theorem 11 and, hence, Theorem 12 can be made to yield:

**THEOREM 22.** *Let  $S$  be an arbitrary set of statements of  $L$  and  $A$  be an arbitrary statement of  $L$ . If  $S \vdash A$ , then—no matter the term extension  $L^+$  of  $L$ — $A$  is true in every Henkin model for  $L^+$  in which  $S$  is true.*

Substituting ‘ $L^+$ ’ for ‘ $L$ ’ at places that the very phrasing of Theorem 12 dictates will do the trick, as the reader may verify.

Theorem 17 fails, we just say, for  $S$  *not* infinitely extendible in  $L$ . So the converse of Theorem 12 fails. Thanks, however, to the term extension  $L^\infty$  of  $L$ , the converse of Theorem 22 does hold. Recall indeed that a set of statements of  $L$ , either infinitely extendible in  $L$  or not, is infinitely extendible in  $L^\infty$  (= Theorem 5), and is sure to be consistent in  $L^\infty$  if consistent in  $L$  (= Theorem 7). So, suppose  $S \not\vdash A$ , where  $S$  and  $A$  are as in Theorem 22. Then  $S \cup \{\neg A\}$  is consistent in  $L^\infty$ ; hence, by the analogue of Theorem 4, for  $L^\infty$ , the Henkin extension  $\mathcal{H}^\infty(S \cup \{\neg A\})$  of  $S \cup \{\neg A\}$  in  $L^\infty$  is maximally consistent and  $\omega$ -complete in  $L^\infty$ ; hence, by the analogue of Theorem 14 for  $L^\infty$  every member of  $\mathcal{H}^\infty(S \cup \{\neg A\})$  is true in the model associate of  $\mathcal{H}^\infty(S \cup \{\neg A\})$  in  $L^\infty$ ; and, hence,  $S$  is true in that model but  $A$  is not.<sup>13</sup> But the model associate of  $\mathcal{H}^\infty(S \cup \{\neg A\})$  in  $L^\infty$  constitutes a Henkin model for  $L^\infty$ . So, there is a term extension  $L^+$  of  $L$  and a Henkin model for  $L^+$  in which  $S$  is true but  $A$  is not.<sup>14</sup> So, by a slight editing of the proof of Theorems 15 and 16:

**THEOREM 23.** *Let  $S$  and  $A$  be as in Theorem 22. If—no matter the term extension  $L^+$  of  $L$ — $A$  is true in every Henkin model for  $L^+$  in which  $S$  is true, then  $S \vdash A$ .*

<sup>13</sup>I assume in the text that sets of statements of  $L^\infty$  (as well as  $L$ ) have been provided with model associates.

<sup>14</sup>To illustrate matters,  $\{Q(t_1), Q(t_2), Q(t_3), \dots\}$  is true, but  $(\forall x)Q(x)$  is not, on the (Henkin)  $I_{\mathcal{T}^\infty}$ -interpretation of  $L^\infty$  ( $\mathcal{T}^\infty$  the set of all the terms of  $L^\infty$ ) that assigns each term of  $L^\infty$  to itself and  $\{t_1, t_2, t_3, \dots\}$  to  $Q$ . Though not true in any Henkin model for  $L$ , the set  $\{Q(t_1), Q(t_2), Q(t_3), \dots, \neg(\forall x)Q(x)\}$  is thus true in one for  $L^\infty$ .

As their proofs attest, Theorems 21 and 23 would hold with  $L^\infty$  the only term extension of  $L$  (besides  $L$  itself), but would thereby lose much of their intuitive appeal.

Now for truth, logical truth and logical entailment in substitutional semantics. Theorem 25 (a)–(b), the results of writing ‘true<sub>*S*</sub>’ for ‘true’ in Theorems 22 and 23, will legitimise the substitutional account of the third—and hence, of the second—of these notions.

Let  $L^+$  be an arbitrary term extension of  $L$ ,  $A$  be an arbitrary statement of  $L^+$ , and  $S$  be an arbitrary set of statements of  $L^+$ . I shall say that  $A$  is *substitutionally true*—or, for short, *true<sub>*S*</sub>*—in a *Henkin model*  $\langle D, I_D \rangle$  for  $L^+$  if (i) in the case that  $A$  is an atomic statement  $Q(T_1, T_2, \dots, T_d)$ , the  $d$ -tuple  $\langle I_D(T_1), I_D(T_2), \dots, I_D(T_d) \rangle$  belongs to  $I_D(Q)$ , (ii) in the case that  $A$  is a negation  $\neg B$ ,  $B$  is not true<sub>*S*</sub> in  $\langle D, I_D \rangle$ , (iii) in the case that  $A$  is a conjunction  $B \wedge C$ , each of  $B$  and  $C$  is true<sub>*S*</sub> in  $\langle D, I_D \rangle$ , and (iv) in the case that  $A$  is a quantification  $(\forall x)B$ , each substitution instance  $B(T/x)$  of  $(\forall x)B$  in  $L^+$  is true<sub>*S*</sub> in  $\langle D, I_D \rangle$ ; and I shall say that  $S$  is *true<sub>*S*</sub>* in  $\langle D, I_D \rangle$  if every member of  $S$  is.

This done, I shall declare a statement  $A$  of  $L$  *logically true in the substitutional sense* if—no matter the term extension  $L^+$  of  $L$ — $A$  is true<sub>*S*</sub> in every Henkin model for  $L^+$ ; and, where  $S$  is a set of statements of  $L$ , I shall declare  $A$  *logically entailed by  $S$  in the substitutional sense* if—no matter the term extension  $L^+$  of  $L$ — $A$  is true<sub>*S*</sub> in every Henkin model for  $L^+$  in which  $S$  is true<sub>*S*</sub>. Equivalently, but more simply,  $A$  may be declared logically true in the substitutional sense if true<sub>*S*</sub> in all Henkin models for  $L$ . (By supplying two definitions of logical truth I set here a precedent to be repeatedly heeded in the essay. In each case it is, of course, the infinite extendibility of  $\emptyset$  in  $L$  which allows for the simpler of the two definitions, the one using just  $L$ .)

It follows at once from Theorem 13 that:

**THEOREM 24.** *Let  $L^+$  be a term extension of  $L$ ,  $A$  be a statement of  $L^+$ ,  $S$  be a set of statements of  $L^+$ , and  $\langle D, I_D \rangle$  be a Henkin model for  $L^+$ .*

- (a)  *$A$  is true<sub>*S*</sub> in  $\langle D, I_D \rangle$  iff true in  $\langle D, I_D \rangle$ .*
- (b)  *$S$  is true<sub>*S*</sub> in  $\langle D, I_D \rangle$  iff true in  $\langle D, I_D \rangle$ .*

Hence by Theorems 22 and 23:

**THEOREM 25.**

*Let  $S$  be an arbitrary set of statements and  $A$  be an arbitrary statement of  $L$ .*

- (a) *If  $S \vdash A$ , then—no matter the term extension of  $L^+$  of  $L$ — $A$  is true<sub>*S*</sub> in every Henkin model for  $L^+$  in which  $S$  is true<sub>*S*</sub>, i.e.  $A$  is logically entailed by  $S$  in the substitutional sense (= The Strong Soundness Theorem for  $L$  in Substitutional Semantics).*

- (b) *If—no matter the term extension  $L^+$  of  $L$ — $A$  is  $\text{true}_S$  in every Henkin model for  $L^+$  in which  $S$  is  $\text{true}_S$ , i.e. if  $A$  is logically entailed by  $S$  in the substitutional sense, then  $S \vdash A$  (= The Strong Completeness Theorem for  $L$  in Substitutional Semantics).*

And hence:

**THEOREM 26.** *Let  $A$  be an arbitrary statement of  $L$ .*

- (a) *If  $\vdash A$ , then—no matter the term extension  $L^+$  of  $L$ — $A$  is  $\text{true}_S$  in every Henkin model for  $L^+$ , i.e.  $A$  is logically true in the substitutional sense  
(= The Weak Soundness Theorem for  $L$  in Substitutional Semantics).*
- (b) *If—no matter the term extension  $L^+$  of  $L$ — $A$  is  $\text{true}_S$  in every Henkin model for  $L^+$ , i.e. if  $A$  is logically true in the substitutional sense, then  $\vdash A$  (= The Weak Completeness Theorem for  $L$  in Substitutional Semantics).*

Theorem 26 legitimises the first account of logical truth in the preceding paragraph. Writing ‘ $\text{true}_S$ ’ for ‘true’ in Theorem 19 (c) legitimises the second and simpler one:

**THEOREM 27.** *Let  $A$  be as in Theorem 26. Then  $\vdash A$  iff  $A$  is  $\text{true}_S$  in every Henkin model for  $L$ .*

Suppose there is a model for  $L$  in which a set  $S$  of statements of  $L$  is true. Then there is one in which  $S$  is true but, say  $Q(t_1) \wedge \neg Q(t_1)$  is not; hence, by Theorem 11  $Q(t_1) \wedge \neg Q(t_1)$  is not provable in  $L$  from  $S$ ; and, hence, by Theorems 17 and 20 there is a countable model for  $L$  in which  $S$  is true (and  $Q(t_1) \wedge \neg Q(t_1)$  is not). So,

- (1) *A set  $S$  of statements of  $L$ , if true in a model for  $L$ , is sure to be true in a countable model for  $L$ .*

The result, known as the *Löwenheim–Skolem Theorem* (or *Skolem’s Generalisation of Löwenheim’s Theorem*) can be edited some (use Theorem 25 (b) in lieu of Theorems 25 and 20):

- (2) *A set  $S$  of statements of  $L$ , if true in a model for  $L$ , is sure—for some term extension  $L^+$  of  $L$ —to be  $\text{true}_S$  in a Henkin model for  $L^+$ .*

(1), from which the so-called ‘Skolem Paradox’ issues, disturbed many when reported in [Skolem, 1920]. But it comforted others who, uneasy over large infinite cardinals (possibly over all cardinals beyond  $\aleph_0$ ), welcomed word that any consistent (first-order) theory can be trusted to have a countable model. In view of (2) such a theory will even have a Henkin

model, the kind of model permitting the quantifiers in the theory to be understood *substitutionally*.

On closer inspection, though, (2) affords only slight comfort to devotees of substitutional semantics. The ‘intended model’ of a (first-order) theory may very much matter, the theory often owing to its intended model whatever consideration it enjoys. Yet  $\langle \mathcal{T}^\infty, I_{\mathcal{T}^\infty} \rangle$ , the Henkin model whose existence the above proof of (2) guarantees, is generally no kin of *that* model. Indeed,  $t_1^\infty, t_2^\infty, t_3^\infty, \dots$ , rarely figure in the ‘intended domain’ (= the domain of the ‘intended model’) of a (first-order) theory. To be sure, some proofs of Theorem 23 (the theorem from which Theorem 25(b) issues) use  $\{1, 2, 3, \dots\}$  in lieu of  $\mathcal{T}^\infty$ , but the positive integers—though figuring in more domains than  $t_1^\infty, t_2^\infty, t_3^\infty, \dots$  do—hardly figure in the intended domain of *every* (first-order) theory. So, one might hesitate to turn in the intended model of a (first-order) theory, even if uncountable, for  $\langle \mathcal{T}^\infty, I_{\mathcal{T}^\infty} \rangle$ .

Well-informed readers will at this point appeal to a sharper version of the Löwenheim–Skolem Theorem. Given a model  $\langle D, I_D \rangle$  for L, acknowledge as a *submodel* of  $\langle D, I_D \rangle$  any pair  $\langle D', I_{D'} \rangle$  such that (i)  $D'$  is a subset of  $D$  to which belong all of  $I_D(t_1), I_D(t_2), I_D(t_3), \dots$ , and (ii)  $I_{D'}$  is the restriction of  $I_D$  to  $D'$ . What indeed [Skolem, 1920] shows is that

- (3) *A set  $S$  of statements of L, if true in a model  $\langle D, I_D \rangle$  for L, is sure to be true in a countable model for L that is a submodel of  $\langle D, I_D \rangle$ .*

The model in question, call it  $\langle D', I_{D'} \rangle$ , need not constitute a Henkin model for L. But it easily extends to a Henkin model  $\langle D', I'_{D'} \rangle$  for  $L^\infty$  in which  $S$  is sure to be true <sub>$S$</sub> . If the members of  $D'$  that are not assigned by  $I_{D'}$  to any term of L are finitely many in number, let them be  $d_1, d_2, \dots, d_n$  (Case 1); otherwise, let them be  $d_1, d_2, d_3, \dots$  (Case 2); and let  $I'_{D'}$  (i) agree with  $I_{D'}$  on all the terms and predicates of L, and (ii) assign in Case 1  $d_i$  to  $t_i^\infty$  for each  $i$  from 1 through  $n - 1$  and  $d_n$  to the remaining terms of  $L^\infty$ , and in Case 2  $d_i$  to  $t_i^\infty$  for each  $i$  from 1 on. It is easily verified that  $S$ , if true in  $\langle D', I_{D'} \rangle$  will be true <sub>$S$</sub>  in  $\langle D', I'_{D'} \rangle$ .

$\langle D', I'_{D'} \rangle$  escapes the criticism levelled at the Henkin model in (2). It bears to the intended model  $\langle D, I_D \rangle$  of a (first-order) theory the closest relationship that logical propriety allows,  $D'$  being a subset of  $D$  and the interpretation that  $I'_{D'}$  places upon the terms and predicates of L being the restriction of  $I_D$  to  $D'$ . So, some might actually turn in  $\langle D, I_D \rangle$  for  $\langle D', I'_{D'} \rangle$ . Others would not, to be sure. But they might well cite  $\langle D', I'_{D'} \rangle$ —a ‘fall-back model’ of the theory, so to speak—as their reason for retaining  $\langle D, I_D \rangle$ , and retaining it with a clear conscience.

There are, however, practitioners of substitutional semantics who have no qualms over cardinals beyond  $\aleph_0$  and consequently provide room in their texts for Henkin models of *any* size. For example, in [Robinson, 1951] and [Shoenfield, 1967] no bound is placed on the number of terms such a language as L might have; and in [Hintikka, 1955] none of it placed on the

number of terms the extensions of  $L$  might have.<sup>15</sup> Under these circumstances the intended model of any theory of  $L$ , if not a Henkin model for  $L$ , extends to one for some extension or other of  $L$ ; and the quantifiers in the theory can as a result be understood substitutionally, something Robinson and Shoenfield both proceed to do on page 19 of their respective texts and which Hintikka does (in a beautifully sly way) on page 24 of his.<sup>16</sup>

(The Appendix has further information on the history of substitutional semantics.)

### 3 TRUTH-VALUE SEMANTICS

What I called on page 56 a *quantifierless* statement of  $L$  would normally be counted *logically true* (also *truth-functionally true*, *tautologous*, etc.) if true on all truth-value assignments to *its atomic components*. Beth in 1959, Schütte in 1962, Dunn and Belnap in 1968, etc. showed in effect that an *arbitrary* statement of  $L$  is logically true in the standard sense iff (substitutionally) true on all truth-value assignments to *the atomic statements of*  $L$ ; and I showed, also in 1968, that such a statement is logically true in the standard sense iff (substitutionally) true on all truth-value assignments to *its atomic substatements*. These may have been the first contributions to what, at Quine's suggestion, I term *truth-value semantics*.

Truth-value semantics is a non-referential kind of semantics: it dispenses with models and with condition (i) on page 69, the one truth condition of a model-theoretic sort that substitutional semantics retains. Instead, atomic statements are assigned a truth-value each, and compound statements are then assigned truth-values via counterparts of conditions (ii)–(iv) on page 69. Though modelless, truth-value semantics is nonetheless a congener of substitutional semantics, a claim I made on page 54 and will substantiate in Theorem 28 below.

There are several versions of truth-value semantics. I begin with one sketched in [Leblanc, 1968], subsequently developed in [Leblanc and Wisdom, 1993; Leblanc, 1976], and various papers in Part 2 of [Leblanc, 1982b], and recently amended to exploit Dunn and Belnap's term extensions.

Let  $L^+$  be an arbitrary term extension of  $L$ . By a *truth-value assignment* to a non-empty set  $\Sigma^+$  of atomic statements of  $L^+$ , I shall understand any function from the members of  $\Sigma^+$  to  $\{\mathbf{T}, \mathbf{F}\}$ . Where  $A$  is a statement of  $L^+$ ,  $\Sigma^+$  a set of atomic statements of  $L^+$  to which belong all the atomic

<sup>15</sup>To quote from p. 52 of [Hintikka, 1955]: "We assume that on any particular occasion we can choose the cardinal number of the individual constants and of the individual variables as large as we wish, by constructing a new, more comprehensive calculus, if necessary." The passage certainly allows what I call the term extensions of  $L$ , and it might be construed as allowing  $L$  itself, to have *any* number of terms.

<sup>16</sup>Hintikka deals principally with *model sets*, a topic I cover in the next section. Robinson and Shoenfield, on the other hand, deal with (*Henkin*) *models*.

substatements of  $A$ , and  $\alpha^+$  a truth value assignment to  $\Sigma^+$ , I shall say that  $A$  is *true* on  $\alpha^+$  if (i) in the case that  $A$  is atomic,  $\alpha^+(A) = \top$ , (ii) in the case that  $A$  is a negation  $\neg B$ ,  $B$  is not true on  $\alpha^+$ , (iii) in the case that  $A$  is a conjunction  $B \wedge C$ , each of  $B$  and  $C$  is true on  $\alpha^+$ , and (iv) in the case that  $A$  is a quantification  $(\forall x)B$ , each substitution instance of  $(\forall x)B$  in  $L^+$  is true on  $\alpha^+$ . finally, where  $S$  is a set of statements of  $L^+$ ,  $\Sigma^+$  a set of atomic statements of  $L^+$  to which belong all the atomic substatements of  $S$ , and  $\alpha^+$  a truth-value assignment to  $\Sigma^+$ , I shall say that  $S$  is *true* on  $\alpha^+$  if each member of  $S$  is true on  $\alpha^+$ .

This done, I shall declare a statement  $A$  of  $L$  *logically true in the truth-value sense* if—no matter the term extension  $L^+$  of  $L$  and truth-value assignment  $\alpha^+$  to the atomic substatements of  $A$  in  $L^+$ — $A$  is true on  $\alpha^+$ ; and, where  $S$  is a set of statements of  $L$ , I shall declare  $A$  *logically entailed by  $S$  in the truth-value sense* if—no matter the term extension  $L^+$  of  $L$  and truth-value assignment  $\alpha^+$  to the atomic substatements of  $S \cup \{A\}$  in  $L^+$ — $A$  is true on  $\alpha^+$  if  $S$  is. Equivalently but more simply,  $A$  may be declared logically true in the truth-value sense if (as the matter was put above)  $A$  is true on every truth-value assignment to the atomic substatements of  $A$  in  $L$ .

Proof that, where  $S$  is a set of statements and  $A$  a statement of  $L$ ,

1.  $S \vdash A$  iff  $A$  is logically entailed by  $S$  in the foregoing sense

and

2.  $\vdash A$  iff  $A$  is logically true in either of the foregoing senses,

can be retrieved from Chapter 2 of [Leblanc, 1976]. So I turn at once to another version of truth-value semantics, a version using *total* rather than *partial* assignments.

Again, let  $L^+$  be an arbitrary term extension of  $L$ . By a *truth-value assignment for  $L^+$*  I shall understand any function from *all* the atomic statements of  $L^+$  to  $\{\top, \text{F}\}$ . Where  $A$  is a statement of  $L^+$ ,  $S$  a set of statements of  $L^+$ , and  $\alpha^+$  a truth-value assignment for  $L^+$ , I shall say that  $A$  is *true* on  $\alpha^+$  if conditions (i)–(iv) two paragraphs back are met, and say that  $S$  is *true* on  $\alpha^+$  if each member of  $S$  is. And, where  $A$  is a statement of  $L$  and  $S$  a set of statements of  $L$ , I shall declare  $A$  *logically true in the truth-value sense* if—no matter the term extension  $L^+$  of  $L$ — $A$  is true on every truth-value assignment for  $L^+$  (equivalently, but more simply, if  $A$  is true on every truth-value assignment for  $L$ ); and I shall declare  $A$  *logically entailed by  $S$  in the truth-value sense* if—no matter the term extension  $L^+$  of  $L$ — $A$  is true on every truth-value assignment for  $L^+$  on which  $S$  is true.

Proof that, where  $S$  is a set of statements and  $A$  is a statement of  $L$ ,

3.  $S \vdash A$  iff  $A$  is logically entailed by  $S$  in this truth-value sense

and

4.  $\vdash A$  iff  $A$  is logically true in either of these truth-value senses,

can be had in two different ways.

Arguing (3) (and hence (4)) directly, understand by the *truth-value associate* in  $L^\infty$  of a set  $S^\infty$  of statements of  $L^\infty$  the function  $\alpha^\infty$  that assigns  $\mathbb{T}$  to the atomic statements of  $L^\infty$  in  $S^\infty$  and  $\mathbb{F}$  to the rest. It follows from Theorem 3 that if  $S^\infty$  is maximally consistent and  $\omega$ -complete in  $L^\infty$ , then a statement of  $L^\infty$  will belong to  $S^\infty$  iff true on  $\alpha^\infty$ . But, if  $S \not\vdash A$ , then by Theorems 7 and 4 the Henkin extension  $\mathcal{H}^\infty(S \cup \{\neg A\})$  of  $S \cup \{\neg A\}$  in  $L^\infty$  is maximally consistent and  $\omega$ -complete in  $L^\infty$ , and hence  $S$  is true on  $\alpha^\infty$  but  $A$  is not. Hence, if—no matter the term extension  $L^+$  of  $L-A$  is true on every truth-value assignment for  $L^+$  on which  $S$  is true, then  $S \vdash A$ . But the proof of the converse is routine. Hence, (3) and, letting  $S$  be  $\emptyset$ , (4).<sup>17</sup>

You can also argue (3)–(4) by appealing to Theorems 25–27, the counterparts of (3)–(4) in substitutional semantics, and to equivalence theorems that bind Henkin models and truth-value assignments. (The definitions used are from [Leblanc, 1976, pp. 92–93].)

With  $\langle D, I_D \rangle$  a Henkin model for  $L^+$ , understand by the *truth-value counterpart* of  $\langle D, I_D \rangle$  the function  $\alpha^+$  that assigns  $\mathbb{T}$  to the atomic statements of  $L^+$  true in  $\langle D, I_D \rangle$  and  $\mathbb{F}$  to the rest; and,  $\alpha^+$  being a truth-value assignment for  $L^+$ , understand by the *model counterpart* of  $\alpha^+$  the pair  $\langle \mathcal{T}^+, I_{\mathcal{T}^+} \rangle$ , where (i)  $\mathcal{T}^+$  consists as usual of the terms of  $L^+$ , (ii) for each term  $T$  of  $L^+$ ,  $I_{\mathcal{T}^+}(T) = T$ , and (iii) for each predicate  $Q$  of  $L^+$  of degree  $d$ ,  $I_{\mathcal{T}^+}(Q) = \{ \langle T_1, T_2, \dots, T_d \rangle : \alpha^+(Q(T_1, T_2, \dots, T_d)) = \mathbb{T} \}$ . The truth-value counterpart of  $\langle D, I_D \rangle$  constitutes, of course, a truth-value assignment for  $L^+$ , and the model counterpart of  $\alpha^+$  a Henkin model for  $L^+$ .

Proof that:

**THEOREM 28.** *Let  $A$  be an arbitrary statement of  $L^+$ .*

- (a)  *$A$  is true in a Henkin model  $\langle D, I_D \rangle$  for  $L^+$  iff  $A$  is true on the truth-value counterpart  $\alpha^+$  of the model (= Equivalence Theorem One for Truth-value Semantics);*
- (b)  *$A$  is true on a truth-value assignment  $\alpha^+$  for  $L^+$  iff  $A$  is true in the model counterpart  $\langle \mathcal{T}^+, I_{\mathcal{T}^+} \rangle$  of the assignment (= Equivalence Theorem Two for Truth-value Semantics);*

is immediate. Suppose indeed that  $A$  is atomic. The foregoing definitions see to it that  $A$  is true in  $\langle D, I_D \rangle$  iff true *on* its counterpart, and that  $A$  is true on  $\alpha^+$  iff true *in* its counterpart. Or suppose  $A$  is a universal quantification. Since  $\langle D, I_D \rangle$  is a Henkin model by hypothesis and  $\langle \mathcal{T}^+, I_{\mathcal{T}^+} \rangle$  is one

<sup>17</sup>Because it uses the truth-value associate of  $\mathcal{H}^\infty(S \cup \{\neg A\})$ , this proof of (3) will be recalled in Section 6.

by construction,  $A$  will be true in either model iff each of its substitution instances in  $L^+$  is. So Henkin models and truth-value assignments match one-to-one.

Tackling soundness first, suppose that  $S \vdash A$  ( $S$  a set of statements and  $A$  a statement of  $L$ ); let  $L^+$  be an arbitrary term extension of  $L$ ; and let  $\alpha^+$  be an arbitrary truth-value assignment for  $L^+$  on which  $S$  is true. Then by Theorem 28 (b)  $S$  will be true in the model counterpart of  $\alpha^+$ ; hence, by Theorem 25 (a),  $A$  will be true in that model; and, hence, by Theorem 28 (b) again,  $A$  will be true on  $\alpha^+$ . So:

**THEOREM 29.** *Let  $S$  be an arbitrary set of statements and  $A$  be an arbitrary statement of  $L$ . If  $S \vdash A$ , then  $A$  is logically entailed by  $S$  in the truth-value sense of page 73 (= The Strong Soundness Theorem for  $L$ , in Truth-Value Semantics).*

Tackling completeness next, suppose that—no matter the term extension  $L^+$  of  $L$ — $A$  is true on every truth-value assignment for  $L^+$  on which  $S$  is true; let  $L^+$  be an arbitrary term extension of  $L$ ; and let  $\langle D, I_D \rangle$  be an arbitrary Henkin model for  $L^+$  in which  $S$  is true. Then by Theorem 28 (a)  $S$  will be true on the truth-value counterpart of  $\langle D, I_D \rangle$ ; hence, by hypothesis,  $A$  will be true on that assignment; and hence, by Theorem 28 (a) again,  $A$  will be true in  $\langle D, I_D \rangle$ . Hence,  $A$  will be true in every Henkin model for  $L^+$  in which  $S$  is true. Hence, by Theorem 25 (b),  $S \vdash A$ . So:

**THEOREM 30.** *Let  $S$  and  $A$  be as in Theorem 29. If  $A$  is logically entailed by  $S$  in the truth-value sense of page 73, then  $S \vdash A$  (= The Strong Completeness Theorem for  $L$  in Truth-Value Semantics).*

The same arguments, but using Theorems 12 and 18 rather than Theorem 25, will guarantee that:

**THEOREM 31.** *Let  $S$  be a set of statements of  $L$  that is infinitely extendible in  $L$ , and let  $A$  be an arbitrary statement of  $L$ . Then  $S \vdash A$  iff  $A$  is true on every truth-value assignment for  $L$  on which  $S$  is true.*

Hence:

**THEOREM 32.** *Let  $A$  be an arbitrary statement of  $L$ . Then  $\vdash A$  iff  $A$  is true on every truth-value assignment for  $L$ .*

Theorems 29 and 30 are in [Dunn and Belnap, 1968]; and variants thereof, using isomorphisms in lieu of term extensions, are in [Leblanc, 1968]. A special case of Theorem 31— $S$  a set of termless statements of  $L$ —is in [Beth, 1959]; and, as reported earlier, versions of Theorem 32 are in these three texts and in [Schütte, 1962].

Yet another version of truth-value semantics assigns truth-values to *all* the statements of  $L^+$  (rather than just the atomic statements of  $L^+$  or just the atomic substatements of a statement or set of statements of  $L^+$ ). I expound it in some detail because of its kinship to truth-set semantics and



to probability theory.

Let  $L^+$  be an arbitrary term extension of  $L$ . By a *truth-value function* for  $L^+$  I shall understand any function  $\alpha^+$  from the statements of  $L^+$  to  $\{\mathbf{T}, \mathbf{F}\}$  that meets the following three constraints (reminiscent of (i)–(iii) in Theorem 3):

**B1.**  $\alpha^+(\neg A) = \mathbf{T}$  iff  $\alpha^+(A) \neq \mathbf{T}$

**B2.**  $\alpha^+(A \wedge B) = \mathbf{T}$  iff  $\alpha^+(A) = \mathbf{T}$  and  $\alpha^+(B) = \mathbf{T}$

**B3.**  $\alpha^+(\forall x)A = \mathbf{T}$  iff  $\alpha^+(A(T/x)) = \mathbf{T}$  for each term  $T$  of  $L^+$ .

I shall then declare a statement  $A$  of  $L$  *logically true in the truth-value sense* if—no matter the term extension  $L^+$  of  $L$  and truth-value function  $\alpha^+$  for  $L^+$ — $\alpha^+(A) = \mathbf{T}$  (equivalently, but more simply, if  $\alpha(A) = \mathbf{T}$  for every truth-value function  $\alpha$  for  $L$ ); and, where  $S$  is a set of statements of  $L^+$ —I shall declare  $A$  *logically entailed by  $S$  in the truth-value sense* if—no matter the term extension  $L^+$  of  $L$  and truth-value function  $\alpha^+$  for  $L^+$ — $\alpha^+(A) = \mathbf{T}$  if  $\alpha^+(B) = \mathbf{T}$  for each member  $B$  of  $S$ .

It is clear that *each truth-value function for  $L^+$  is an extension to all the statements of  $L^+$  of a truth-value assignment for  $L^+$ , and each truth-value assignment for  $L^+$  the restriction to just the atomic statements of  $L^+$  of a truth-value function for  $L^+$* . So, truth-value assignments and truth-value functions match one-to-one, the way Henkin models and truth-value assignments did. the present accounts of logical truth are therefore sure in light of Theorems 29, 30 and 32 to be weakly sound and complete, and the present account of logical entailment is sure in light of Theorems 29 and 30 to be strongly sound and complete.

When showing in Section 5 the truth-value functions for  $L$  to issue into probability functions, I shall refer to them by means of ‘ $P$ ’ rather than ‘ $\alpha$ ’, and I shall use as truth-values the reals 1 and 0 rather than  $\mathbf{T}$  and  $\mathbf{F}$ .

Truth-set semantics, my next concern, is but truth-value semantics in set-theoretic disguise. And like truth-value semantics, which—interestingly enough—it antedates, it comes in several versions. One, due to Carnap<sup>18</sup> but recast here in the Dunn and Belnap manner, matches the semantics of page 73 with its total truth-value assignments.

Where  $L^+$  is an arbitrary term extension of  $L$ , understand by a *basic pair* for  $L^+$  any set of the sort  $\{A, \neg A\}$ , where  $A$  is an atomic statement of  $L^+$ ; understand by a *basic truth set* for  $L^+$ —or, as Carnap would put it, a *state-description* for  $L^+$ —any set consisting of one statement from each basic pair for  $L^+$ ; take a statement  $A$  of  $L^+$  to *hold* in a state-description  $SD^+$  for  $L^+$  if (i) in the case that  $A$  is atomic,  $A$  belongs to  $SD^+$ , (ii) in

<sup>18</sup>To be more exact, sketched in [Wittgenstein, 1921] and formalised in [Carnap, 1950]. The definitions in the next paragraph are borrowed from [Carnap, 1950, Section 18].

the case that  $A$  is a negation  $\neg B$ ,  $B$  does not hold in  $SD^+$ , (iii) in the case that  $A$  is a conjunction  $B \wedge C$ , each of  $B$  and  $C$  holds in  $SD^+$ , and (iv) in the case that  $A$  is a quantification  $(\forall x)B$ , each substitution instance of  $A$  in  $L^+$  holds in  $SD^+$ ; and take a set of statements of  $L^+$  to *hold* in  $SD^+$  if each member of  $S$  does.

This done, declare a statement  $A$  of  $L$  *logically true in the truth-set sense* if—no matter the term extension  $L^+$  of  $L$ — $A$  holds in every state-description for  $L^+$  (equivalently, but more simply, if  $A$  holds in every state-description for  $L$ ); and where  $S$  is a set of statements of  $L$ , declare  $A$  *logically entailed by  $S$  in the truth-set sense* if—no matter the term extension  $L^+$  of  $L$ — $A$  holds in every state-description for  $L^+$  in which  $S$  does.

As truth-value assignments and state-descriptions for  $L^+$  match one-to-one, a statement of  $L^+$  will be true on a truth-value assignment for  $L^+$  iff it holds *in* the matching state-description, and the statement will hold in a state-description for  $L^+$  iff it is true *on* the matching truth-value assignment. So, by Theorems 29, 30 and 32:

**THEOREM 33.** *Let  $S$  be an arbitrary set of statements and  $A$  be an arbitrary statement of  $L$ .*

- (a)  $S \vdash A$  iff  $A$  is logically entailed by  $S$  in the truth-set sense.
- (b)  $\vdash A$  iff  $A$  is logically true in the truth-set sense.

That  $A$ , if provable in  $L$ , holds in every state-description for  $L$  was first shown in [Carnap, 1950]. (See Section 22, which should be consulted for more information on this particular brand of truth-set semantics.)

A second version of truth-set semantics can be retrieved from [Quine, 1940], [Smullyan, 1968] and doubtless several other sources. It matches the semantics of pages 76–77 with its truth-value functions, boasts the tersest accounts yet of logical truth and entailment, and can be legitimised in the swiftest manner yet.

With  $L^+$  an arbitrary term extension of  $L$ , understand by a *truth set* for  $L^+$  any set  $S^+$  of statements of  $L^+$  that meets the following three conditions (patterned after (i)–(iii) in Theorem 3):

- (a)  $\neg A$  belongs to  $S^+$  iff  $A$  does not,
- (b)  $A \wedge B$  belongs to  $S^+$  if each of  $A$  and  $B$  does,
- (c)  $(\forall x)A$  belongs to  $S^+$  iff  $A(T/x)$  does for each term  $T$  of  $L^+$ .

Declare a statement  $A$  of  $L$  *logically true in the truth-set sense* if—no matter the term extension  $L^+$  of  $L$ — $A$  belongs to every truth set for  $L^+$  (equivalently, but more simply, if  $A$  belongs to every truth set for  $L$ ); and, where  $S$  is a set of statements of  $L$ , declare  $A$  *logically entailed by  $S$  in the truth-set sense* if—no matter the term extension  $L^+$  of  $L$ — $A$  belongs to every truth set for  $L^+$  of which  $S$  is a subset.<sup>19</sup>

<sup>19</sup>Quine's truth sets consist of components of a quantifierless statement, and are ac-

Truth-value functions and truth sets for  $L^+$  match one-to-one. Indeed, the set of all statements of  $L^+$  evaluating to  $T$  on a truth-value function for  $L^+$  constitutes a truth set for  $L^+$ ; and the function on which all the statements of  $L^+$  in a truth set for  $L^+$  evaluate to  $T$  and the rest evaluate to  $F$  constitutes a truth-value function for  $L^+$ . So, the accounts of logical truth in the foregoing paragraph are sure to be weakly sound and complete since those of page 76 were, and the account of logical entailment in that paragraph is sure to be strongly sound and complete since the one of page 76 was. However, more direct and compact proofs of these matters can be had. Suppose that  $S \vdash A$ , where  $S$  is a set of statements and  $A$  is a statement of  $L$ ; let  $L^+$  be an arbitrary term extension of  $L$ ; and let  $S^+$  be an arbitrary truth set for  $L^+$  of which  $S$  is a subset. It is readily verified that each axiom of  $L$  belongs to each truth set for  $L^+$ , and hence to  $S^+$ ; and that the ponential of two statements of  $L$  belongs to a truth set for  $L^+$ , say  $S^+$ , if the two statements themselves do. So each entry in any proof of  $A$  from  $S$  in  $L$  will belong to  $S^+$ . So:

**THEOREM 34.** *Let  $S$  be an arbitrary set of statements and  $A$  be an arbitrary statement of  $L$ . If  $S \vdash A$ , then—no matter the term extension  $L^+$  of  $L$ — $A$  belongs to every truth set for  $L^+$  of which  $S$  is a subset (= The Strong Soundness Theorem for  $L$  in Truth-Set Semantics).*

As for the converse of Theorem 34, recall that if a set of statements of  $L$  is consistent and infinitely extendible in  $L$ , then the Henkin extension of the set is sure by Theorem 4 to be maximally consistent and  $\omega$ -complete in  $L$ . But by virtue of Theorem 3 any set of statements of  $L$  that is maximally consistent and  $\omega$ -complete in  $L$  constitutes a truth set for  $L$  (and vice-versa), truth sets being sometimes called, as a result, *Henkin sets*. So:

**THEOREM 35.** *Let  $S$  be a set of statements of  $L$  that is infinitely extendible in  $L$ . If  $S$  is consistent in  $L$ , then there is a truth set for  $L$ —and, hence, there is for some term extension  $L^+$  of  $L$  a truth set for  $L^+$ —of which  $S$  is a subset.*

But, if  $S \vdash A$ , then  $S \cup \{\neg A\}$  is consistent in  $L$ . So:

**THEOREM 36.** *Let  $S$  be as in Theorem 35, and let  $A$  be an arbitrary statement of  $L$ . If  $S \not\vdash A$ , then there is a truth set for  $L$ —and, hence, there is for some term extension  $L^+$  of  $L$  a truth set for  $L^+$ —of which  $S$  is a subset but  $A$  is not a member.*

Hence, thanks to  $L^\infty$ , when  $S$  is not infinitely extendible:

**THEOREM 37.** *Let  $S$  be an arbitrary set of statements and  $A$  be an arbitrary statement of  $L$ . If—no matter the term extension  $L^+$  of  $L$ — $A$  belongs to every truth set for  $L^+$  of which  $S$  is a subset, then  $S \vdash A$  (= The Strong*

---

*cordingly called the truth set of that statement.* As one would expect, a quantifierless statement is declared tautologous if it belongs to every one of its truth sets.

*Completeness Theorem for L in Truth-Set Semantics).*

Hence also:

**THEOREM 38.** *Let  $A$  be an arbitrary statement of  $L$ . Then  $\vdash A$  iff  $A$  belongs to every truth set for  $L$ .*

Proof of Theorem 38 is in [Smullyan, 1968], a text to be consulted on this particular brand of truth-set semantics.

Note that  $Q(t_1) \wedge \neg Q(t_1)$  cannot belong to any truth set for  $L^+$  since at most one of  $Q(t_1)$  and  $\neg Q(t_1)$  does. But a set  $S$  of statements of  $L$  is consistent in  $L$  iff  $Q(t_1) \wedge \neg Q(t_1)$  is not provable in  $L$  from  $S$ . Hence, by Theorem 34:

**THEOREM 39.** *Let  $S$  be an arbitrary set of statements of  $L$ . If there is for some term extension  $L^+$  of  $L$  a truth set for  $L^+$  of which  $S$  is a subset, then  $S$  is consistent in  $L$ .*

Hence as a transition to our last topic in this section:

**THEOREM 40.** *Let  $S$  be as in Theorem 39. Then  $S$  is consistent in  $L$  iff there is for some term extension  $L^+$  of  $L$  a truth set for  $L^+$  of which  $S$  is a subset (i.e. to which  $S$  extends).*

Besides extending to truth sets, consistent sets also extend to *model sets*, a kind of set first investigated in [Hintikka, 1955]. These are intriguing in several respects. Though true on a truth-value assignment, a model set need not comprise—as a truth set would—all the statements true on that assignment, and hence it is not a truth-value function in disguise. A model set may even be finite, in which case it (plus of course all sets extending to it) has a finite model. And routines have been devised which (i) when a finite set  $S$  is inconsistent, will invariably apprise us of the fact and (ii) when  $S$  is consistent, will frequently—though not invariably—extend it to a finite model set and hand us a model of  $S$ .

Model sets also permit definitions of logical truth and logical entailment which, though reminiscent of those on page 77, cunningly differ from them. It is, of course, these definitions which particularly interest us and which we will legitimise by means of suitable soundness and completeness theorems. But some of the results in the previous paragraph are readily had and will be recorded as well.

$L^+$  being an arbitrary term extension of  $L$ , I shall understand by a *model set* for  $L^+$  any set  $S$  of statements of  $L^+$  such that (i) at least one term of  $L^+$  occurs in  $S$  and (ii)  $S$  meets the following constraints:

- (a) where  $A$  is an atomic statement of  $L^+$ , at most one of  $A$  and  $\neg A$  belongs to  $S$ ,
- (b) if  $\neg\neg A$  belongs to  $S$ , then so does  $A$ ,

- (c) if  $A \wedge B$  belongs to  $S$ , then so does each of  $A$  and  $B$ ,
- (d) if  $\neg(A \wedge B)$  belongs to  $S$ , then at least one of  $\neg A$  and  $\neg B$  does,
- (e) if  $(\forall x)A$  belongs to  $S$ , then so does  $A(T/x)$  for every term  $T$  of  $L^+$  that occurs in  $S$ ,
- (f) if  $\neg(\forall x)A$  belongs to  $S$ , then so does  $\neg A(T/x)$  for at least one term  $T$  of  $L^+$ .

Note: Because of requirement (i)  $S$  must, of course, be non-empty and—more importantly—at least one substitution instance of any universal quantification belonging to  $S$  must belong to  $S$  (see clause (e)). The latter is tantamount, model-theoretically, to requiring that domains be non-empty, a point first made in [Hintikka, 1955, pp. 34–35]. And, because of the qualification ‘that occurs in  $S$ ’ in clause (e), not every substitution instance of  $(\forall x)A$  need belong to  $S$  (though, as we just saw, at least one must). This is the most distinctive feature of model sets, and the one which most notably allows (some) finite sets of statements of  $L^+$  to qualify as model sets for  $L^+$ .

It is easily verified that each truth set for  $L$  constitutes a model set for  $L$ , a fact which with an eye to later developments I record separately:

**THEOREM 41.** *Each truth set for  $L$  constitutes a model set for  $L$ .*

(And each model set for  $L$  is a subset—though not necessarily more than a subset—of a truth set for  $L$ : as noted above and proved on pages 82–83, each model set for  $L$  is true on a truth-value assignment for  $L$  and, hence, is a subset of the truth-set associate of that assignment.)

This done, I declare a statement  $A$  of  $L$  *logically true in the model-set sense* if—no matter the term extension  $L^+$  of  $L$ — $\neg A$  does not belong to any model set for  $L^+$  (equivalently, but more simply, if  $\neg A$  does not belong to any model set for  $L$ ); and, where  $S$  is a set of statements of  $L$ , I declare  $A$  *logically entailed by  $S$  in the model-set sense* if—no matter the term extension  $L^+$  of  $L$ — $\neg A$  does not belong to any model set for  $L^+$  of which  $S$  is a subset. (The definitions are suggested by results in [Hintikka, 1955].)

Proof that the foregoing definition of logical entailment is strongly sound calls for four lemmas (Theorems 42–45) and one definition. Theorem 42 is crucial to the whole enterprise. As noted above, the substitution instances of a quantification  $(\forall x)A$  that belongs to a model set need not all belong to the set. Truth-value assignment  $\alpha$  in theorem 42 ensures that the substitution instances not in the set behave exactly like those in it. (Clause (a) of Theorem 42 stands to clause (b) and eventually to Theorem 44 somewhat as Theorem 8 stands to Theorem 13.)

**THEOREM 42.** *Let  $\mathcal{T}$  be a non-empty set of terms of  $L$ ; let  $\mathcal{T}'$  be the complement of  $\mathcal{T}$ ;  $T_1$  being the alphabetically earliest member of  $\mathcal{T}$ , let the  $T_1$ -rewrite  $T_1(A)$  of a statement  $A$  of  $L$  be the result of putting  $T_1$  for each*

member of  $\mathcal{T}'$  in  $A$ ,<sup>20</sup> and let  $\alpha$  be any truth-value assignment for  $L$  such that, for each atomic statement  $A$  of  $L$ ,  $\alpha(A) = \alpha(T_1(A))$ .

- (a)  $T'_1, T'_2, T'_3, \dots$ , being in alphabetic order the various members (if any) of  $\mathcal{T}'$ , let the  $\mathcal{L}$ -rewrite  $\mathcal{L}(A)$  of a statement  $A$  of  $L$  be the result of simultaneously putting  $T_1$  for  $T'_1$  in  $A$ ,  $T'_1$  for  $T'_2$ ,  $T'_2$  for  $T'_3$ , etc.<sup>21</sup> Then  $A$  is true on  $\alpha$  iff  $\mathcal{L}(A)$  is.
- (b) Let  $(\forall x)A$  be a quantification of  $L$  in which no member of  $\mathcal{T}'$  occurs. If  $A(T/x)$  is true on  $\alpha$  for each term  $T$  in  $\mathcal{T}$ , then  $A(T'/x)$  is true on  $\alpha$  for each term  $T'$  in  $\mathcal{T}'$ .

**Proof.** (a) That  $A$  is true on  $\alpha$  iff  $\mathcal{L}(A)$  is true on  $\alpha$  is shown by mathematical induction on the length  $l(A)$  of  $A$ .

*Basis:*  $l(A) = 1$ , in which case  $A$  is atomic. By construction  $A$  and  $\mathcal{L}(A)$  have the same  $T_1$ -rewrite. Hence,  $\alpha(A) = \alpha(\mathcal{L}(A))$ . Hence, (a).

*Inductive Step:*  $l(A) > 1$ .

*Case 1:*  $A$  is a negation  $\neg B$ . By the hypothesis of the induction  $B$  is true on  $\alpha$  iff  $\mathcal{L}(B)$  is. Hence,  $\neg B$  is true on  $\alpha$  iff  $\neg(\mathcal{L}(B))$  is, i.e. iff  $\mathcal{L}(\neg B)$  is. Hence, (a).

*Case 2:*  $A$  is a conjunction  $B \wedge C$ . Proof similar to that of Case 1.

*Case 3:*  $A$  is a quantification  $(\forall x)B$ . (i) for each term  $T$  in  $\mathcal{T}$ ,  $(\mathcal{L}(B))(T/x)$  is the same as  $\mathcal{L}(B(T/x))$ ; and, for each  $i$  from 1 on,  $(\mathcal{L}(B))(T'_i/x)$  is the same as  $\mathcal{L}(B(T'_{i+1}/x))$ . So, each substitution instance of  $(\forall x)(\mathcal{L}(B)) (= \mathcal{L}((\forall x)B))$  is the  $\mathcal{L}$ -rewrite of a substitution instance for  $(\forall x)B$ . (ii) For each member  $T$  of  $\mathcal{T}$ ,  $\mathcal{L}(B(T/x))$  is the same as  $(\mathcal{L}(B))(T/x)$ ;  $\mathcal{L}(B)T'_1/x$  is the same as  $(\mathcal{L}(B))(T_1/x)$ ; and, for each  $i$  from 2 on,  $\mathcal{L}(B(T'_i/x))$  is the same as  $(\mathcal{L}(B))(T'_{i-1}/x)$ . So, the  $\mathcal{L}$ -rewrite of each substitution instance of  $(\forall x)B$  is a substitution instance of  $(\forall x)(\mathcal{L}(B))$ . (iii) Suppose first that  $(\forall x)B$  is true on  $\alpha$ , and hence that each substitution instance of  $(\forall x)B$  is true on  $\alpha$ . Then, by the hypothesis of the induction, the  $\mathcal{L}$ -rewrite of each substitution instance of  $(\forall x)B$  is true on  $\alpha$ , and hence by (i) so is each substitution instance of  $\mathcal{L}((\forall x)B)$ . Hence,  $\mathcal{L}((\forall x)B)$  is true on  $\alpha$ . Suppose next that  $(\forall x)B$  is not true on  $\alpha$ , and hence that at least one substitution instance of  $(\forall x)B$  is not true on  $\alpha$ . Then, by the hypothesis of the induction, the  $\mathcal{L}$ -rewrite of at least one substitution instance of  $(\forall x)B$  is not true on  $\alpha$ , and hence by (ii) at least one substitution instance of  $\mathcal{L}((\forall x)B)$  is not either. hence,  $\mathcal{L}((\forall x)B)$  is not true on  $\alpha$ . Hence, (a).

(b) (i) Since no member of  $\mathcal{T}'$  occurs in  $A$ ,  $\mathcal{L}(A(T'_1/x))$  is the same as  $A(t_1/x)$ , and for each  $i$  from 2 on  $\mathcal{L}(A(T'_i/x))$  is the same as  $A(T'_{i-1}/x)$ . (ii) Suppose  $A(T/x)$  is true on  $\alpha$  for each member  $T$  of  $\mathcal{T}$ . Then in particular  $A(T_1/x)$  is true on  $\alpha$ , hence by (i) so is  $\mathcal{L}(A(T'_1/x))$ , hence by (a) so is

<sup>20</sup>When ambiguity threatens, I shall enclose ' $T_1(A)$ ' within parentheses.

<sup>21</sup>When ambiguity threatens, I shall enclose ' $\mathcal{L}(A)$ ' within parentheses.

$A(T'_1/x)$ , hence by (i) so is  $\mathcal{L}(A(T'_2/x))$ , hence, by (a) so is  $A(T'_2/x)$ , hence by (i) so is  $\mathcal{L}(A(T'_3/x))$ , hence by (a) so is  $A(T'_3/x)$ , etc. Hence, if  $A(T/x)$  is true on  $\alpha$  for each member  $T$  of  $\mathcal{T}$ , so is  $A(T'/x)$  for each member  $T'$  of  $\mathcal{T}'$ . ■

Now let  $S$  be a model set for  $L$ , and let  $T_1$  be the alphabetically earliest term of  $L$  that occurs in  $S$ .<sup>22</sup> By the *truth-value associate* of  $S$  in  $L$ , I shall understand the function  $\alpha_S$  from the atomic statements of  $L$  to  $\{\mathbb{T}, \mathbb{F}\}$  such that, for each atomic statement  $A$  of  $L$ ,

$$\alpha_S(A) = \begin{cases} \mathbb{T} & \text{if } T_1(A) \text{ belongs to } S \\ \mathbb{F} & \text{otherwise.} \end{cases}$$

(The function stems from [Leblanc and Wisdom, 1993, p. 191], and the proofs of Theorems 43–45 below are simplifications of proofs in that text, pp. 300–304.)

**THEOREM 43.** *Let  $S$  be a model set for  $L$ , and  $\alpha_S$  be the truth-value associate of  $S$  in  $L$ . Then:*

- (a)  $\alpha_S$  constitutes a truth-value assignment for  $L$ , and
- (b) for each atomic statement  $A$  of  $L$ ,  $\alpha_S(A) = \alpha_S(T_1(A))$ .

As regards (b), note that any term of  $L$  that occurs in  $T_1(A)$  is sure to occur in  $S$ . So,  $T_1(A)$  and  $T_1(T_1(A))$  are the same. So,  $T_1(A)$  belongs to  $S$  iff  $T_1(T_1(A))$  does. So, by the construction of  $\alpha_S$ ,  $\alpha_S(A) = \alpha_S(T_1(A))$ .

Hence, by Theorem 42 (b) (with the set of all the terms that occur in  $S$  serving as  $\mathcal{T}$ , and the truth-value associate of  $S$  serving as  $\alpha$ ):

**THEOREM 44.** *Let  $S$  be a model set for  $L$ , and  $(\forall x)A$  be a quantification of  $L$  that belongs to  $S$ . If—for each and every term  $T$  of  $L$  that occurs in  $S$ — $A(T/x)$  is true on the truth-value associate of  $S$ , then  $(\forall x)A$  is true on that associate.*

I am now in a position to show that each model set for  $L$  is true on its truth-value associate, hence is true in the model associate of that associate, and hence has a (Henkin) model.

**THEOREM 45.** *Each model set for  $L$  is true on its truth-value associate.*

**Proof.** Let  $S$  be an arbitrary model set for  $L$ ,  $A$  be an arbitrary member of  $S$ , and  $\alpha_S$  be the truth-value associate of  $S$ . That  $A$  is sure to be true on  $\alpha_S$  is shown by mathematical induction on the length  $l(A)$  of  $A$ .

*Basis:*  $l(S) = 1$ , in which case  $A$  is atomic. Since  $A$  belongs to  $S$ ,  $A$  is its own  $T_1$ -rewrite ( $T_1$  the alphabetically earliest term of  $L$  to occur in  $S$ ); hence,  $\alpha_S(A) = \mathbb{T}$ ; and, hence,  $A$  is true on  $\alpha_S$ .

<sup>22</sup>Here, as in Theorem 42, any other term of  $L$  occurring in  $S$  could of course substitute for  $T_1$ .

*Inductive Step:*  $l(A) > 1$ , in which case  $A$  is a negation, or a conjunction or a universal quantification (and, when a negation, one of an atomic statement or of a negation, or of a conjunction, or of a universal quantification).<sup>23</sup>

*Case 1:*  $A$  is of the sort  $\neg B$ , where  $B$  is an atomic statement. Then by the definition of a model set  $B$  does not belong to  $S$ . But, since  $\neg B$  belongs to  $S$ ,  $\neg B$  is its own  $T_1$ -rewrite and, hence, so is  $B$ . Hence,  $T_1(B)$  does not belong to  $S$ . Hence,  $\alpha_S(B) = F$ . Hence,  $B$  is not true on  $\alpha_S$ . Hence,  $\neg B$  is.

*Case 2:*  $A$  is of the sort  $\neg\neg B$ . Then by the definition of a model set  $B$  belongs to  $S$ ; hence, by the hypothesis of the induction,  $B$  is true on  $\alpha_S$ ; and hence, so is  $\neg\neg B$ .

*Case 3:*  $A$  is of the sort  $\neg(B \wedge C)$ . Then by the definition of a model set at least one of  $\neg B$  and  $\neg C$  belongs to  $S$ ; hence, by the hypothesis of the induction, at least one of  $\neg B$  and  $\neg C$  is true on  $\alpha_S$ ; and, hence,  $\neg(B \wedge C)$  is true on  $\alpha_S$ .

*Case 4:*  $A$  is of the sort  $\neg(\forall x)B$ . Then by the definition of a model set  $\neg B(T/x)$  belongs to  $S$  for some term  $T$  of  $L$ ; hence, by the hypothesis of the induction,  $\neg B(T/x)$  is true on  $\alpha_S$  for some term  $T$  of  $L$ ; and, hence,  $\neg(\forall x)B$  is true on  $\alpha_S$ .

*Case 5:*  $A$  is of the sort  $B \wedge C$ . Then by the definition of a model set each of  $B$  and  $C$  belongs to  $S$ ; hence, by the hypothesis of the induction, each of  $B$  and  $C$  is true on  $\alpha_S$ ; and, hence, so is  $B \wedge C$ .

*Case 6:*  $A$  is of the sort  $(\forall x)B$ . Then by the definition of a model set  $B(T/x)$  belongs to  $S$  for each term  $T$  of  $L$  that occurs in  $S$ ; hence by the hypothesis of the induction,  $B(T/x)$  is true on  $\alpha_S$  for each term  $T$  of  $L$  that occurs in  $S$ ; and, hence, by Theorem 44,  $(\forall x)B$  is true on  $\alpha_S$ . ■

Now for the Strong Soundness Theorem for  $L$  in model-set semantics.

**THEOREM 46.** *Let  $S$  be an arbitrary set of statements and  $A$  be an arbitrary statement of  $L$ . If  $S \vdash A$ , then  $\neg A$  does not belong to any model set for  $L$  of which  $S$  is a subset.*

**Proof.** Suppose  $S \vdash A$ , and suppose there is no model set for  $L$  of which  $S$  is a subset. Then, trivially,  $\neg A$  does not belong to any model set for  $L$  of which  $S$  is a subset. Suppose, on the other hand, there is at least one model set for  $L$ , say,  $S'$ , of which  $S$  is a subset, and let  $\alpha_{S'}$  be the truth-value associate of  $S'$ . Then, by Theorem 45,  $S'$  is true on  $\alpha_{S'}$ ; hence, so is  $S$ ; and hence by Theorem 29, so is  $A$ . But, if  $A$  is true on  $\alpha_{S'}$ , then  $\neg A$  is not; and, if  $\neg A$  is not true on  $\alpha_{S'}$ , then by Theorem 45 again  $\neg A$  does not belong to  $S'$ . Hence, again,  $\neg A$  does not belong to any model set for  $L$  of which  $S$  is a subset. ■

<sup>23</sup>An induction of this sort, with each kind of negation treated separately, is called in [Leblanc and Wisdom, 1993] a *Hintikka induction*.



But the argument leading to Theorem 46 holds with ‘ $L^+$ ’ in place of ‘ $L$ ’. Hence:

**THEOREM 47.** *Let  $S$  and  $A$  be as in Theorem 46. If  $S \vdash A$ , then—no matter the term extension  $L^+$  of  $L$ — $\neg A$  does not belong to any model set for  $L^+$  of which  $S$  is a subset (= The Strong Soundness Theorem for  $L$  in Model-Set Semantics).*

The converse of Theorem 46 for infinitely extendible  $S$  and that of Theorem 47 for arbitrary  $S$  readily issue from Theorems 41, 36 and 37:

**THEOREM 48.** *Let  $S$  be a set of statements of  $L$  that is infinitely extendible in  $L$ , and  $A$  be an arbitrary statement of  $L$ . If  $\neg A$  does not belong to any model set for  $L$  of which  $S$  is a subset, then  $S \vdash A$ .*

**THEOREM 49.** *Let  $S$  be an arbitrary set of statements and  $A$  be an arbitrary statement of  $L$ . If—no matter the term extension  $L^+$  of  $L$ — $\neg A$  does not belong to any model set for  $L^+$  of which  $S$  is a subset, then  $S \vdash A$  (= The Strong Completeness Theorem for  $L$  in Model-Set Semantics).*

Hence:

**THEOREM 50.** *Let  $S$  and  $A$  be as in Theorem 49. Then  $S \vdash A$  iff—no matter the term extension  $L^+$  of  $L$ — $\neg A$  does not belong to any model set for  $L^+$  of which  $S$  is a subset.*

And hence:

**THEOREM 51.**  *$\vdash A$  iff—no matter the term extension  $L^+$  of  $L$ — $\neg A$  does not belong to any model set for  $L^+$ .*

Theorem 50 legitimises the account of logical entailment and Theorem 51 the first account of logical truth, on page 79.

Theorems 46 and 49 combine, of course, to read:

**THEOREM 52.** *Let  $S$  be a set of statements of  $L$  that is infinitely extendible in  $L$ , and  $A$  be an arbitrary statement of  $L$ . Then  $S \vdash A$  iff  $\neg A$  does not belong to any model set for  $L$  of which  $S$  is a subset.*

Hence the following theorem, which legitimises the second (and simpler) account of logical truth on page 79:

**THEOREM 53.**  *$\vdash A$  iff  $\neg A$  does not belong to any model set for  $L$ .*

Now for some of the results reported on page 79. Note that (i) a set  $S$  of statements of  $L$  is consistent in  $L$  iff  $Q(T_1) \wedge \neg Q(t_1)$ , for example, is not provable from  $S$  in  $L$ , and (ii) any model set for  $L^+$  is a subset of one with  $\neg(Q(t_1) \wedge \neg Q(t_1))$  as a member. (For proof of (ii) note that any model set for  $L^+$  is a subset of a truth set for  $L^+$  (page 81),  $\neg(Q(t_1) \wedge \neg Q(t_1))$  belongs by Theorem 33 (b) to every truth set for  $L^+$ , and any truth set for  $L^+$  is a model set for  $L^+$ . So, any model set for  $L^+$  is sure to be a subset of one with  $\neg(Q(t_1) \wedge \neg Q(t_1))$  as a member). Hence, by Theorem 49:

**THEOREM 54.** *Let  $S$  be an arbitrary set of statements of  $L$ . Then  $S$  is consistent in  $L$  iff there is, for some term extension  $L^+$  of  $L$ , a model set for  $L^+$  to which  $S$  extends (i.e. of which  $S$  is a subset).*

Similarly, but with Theorem 52 substituting for Theorem 49:

**THEOREM 55.** *Let  $S$  be a set of statements of  $L$  that is infinitely extendible in  $L$ . Then  $S$  is consistent in  $L$  iff there is a model set for  $L$  to which  $S$  extends (i.e. of which  $S$  is a subset).*

The two results stem from [Hintikka, 1955], where—*thanks to first-order languages having as many terms as the occasion calls for*—sets of first-order statements were shown to be consistent iff they extend to model sets. Since a set of first-order statements is consistent iff it has a model, Hintikka's result reads, in effect: 'Sets of first-order statements have *models* iff they extend to *model* sets'.

As suggested earlier, one model in which a model set for  $L^+$  is sure to be true can be obtained from pages 74 and 82. Indeed, let  $S$  be a model set for  $L^+$ ; let  $\alpha_S^+$  be the counterpart for  $L^+$  of the truth-value assignment  $\alpha_S$  on page 82; and let  $\langle \mathcal{T}^+, I_{\mathcal{T}^+} \rangle$  be the model counterpart of  $\alpha_S^+$  as per page 74. By the counterpart of Theorem 45 for  $L^+$ ,  $S$  is sure to be true on  $\alpha_S^+$ . Hence,  $S$  is sure by Theorem 28 to have  $\langle \mathcal{T}^+, I_{\mathcal{T}^+} \rangle$  as a model. Hence, so are all sets of statements of  $L$  extending to  $S$ . The model, one will recall, is a Henkin one.

**THEOREM 56.** *Let  $L^+$  be an arbitrary term extension of  $L$ , and  $S$  be an arbitrary model set for  $L^+$ . Then there is a (Henkin) model for  $L^+$  in which  $S$  (and, hence, each set of statements of  $L$  that extends to  $S$ ) is true.*

Since the set  $\mathcal{T}^+$  of all the terms of  $L^+$  is infinite, model  $\langle \mathcal{T}^+, I_{\mathcal{T}^+} \rangle$  is infinite as well, a point noted in [Smullyan, 1968, p. 62]. However, when  $S$  is a finite model set for  $L$ ,  $S$  is sure to have a finite model as well. Let (i)  $\mathcal{T}_S$  consist of the various terms of  $L$  that occur in  $S$ , (ii) for each term  $T$  of  $L$ , let  $I_{\mathcal{T}_S}(T)$  be  $T$  itself if  $T$  belongs to  $\mathcal{T}_S$ , otherwise the alphabetically earliest member of  $\mathcal{T}_S$ , and (iii) for each predicate  $Q$  of  $L$  of degree  $d$ , let  $I_{\mathcal{T}_S}(Q)$  be  $\{ \langle I_{\mathcal{T}_S}(T_1), I_{\mathcal{T}_S}(T_2), \dots, I_{\mathcal{T}_S}(T_d) \rangle : Q(T_1, T_2, \dots, T_d) \in S \}$ . Theorems 45 and 28 are easily edited to show that  $S$  has finite model  $\langle \mathcal{T}_S, I_{\mathcal{T}_S} \rangle$  as a model. Hence, so do all sets of statements of  $L$  extending to  $S$ . The model, again, is a Henkin one.

**THEOREM 57.** *Let  $S$  be a finite model set for  $L$ . Then there is a finite (Henkin) model for  $L$  in which  $S$  (and, hence, each set of statements of  $L$  that extends to  $S$ ) is true.*

It can further be shown of any finite set  $S$  of statements of  $L$  that if  $S$  has a finite model, then  $S$  extends to a finite model set for  $L$ . *A finite set of statements of  $L$  thus has a finite model if it extends to a finite model set for  $L$ .* The result is particularly interesting as regards the routine in [Leblanc

and Wisdom, 1993] (and that in [Jeffrey, 1990] from which it stems) for making consistency trees. As pointed out on p. 298 of [Leblanc and Wisdom, 1993], any set of statements declared CONSISTENT by the routine of p. 189 is a subset of a finite model set (as that text puts it, a subset of a finite *Hintikka set*). Any such set thus has a finite model. The routine in question declares CONSISTENT only some of the sets of statements of  $L$  that extend to finite model sets. However, I have since found another routine which *does* declare CONSISTENT *all* sets of statements of  $L$  that extend to finite model sets.<sup>24</sup>

So much, however, for accounts of logical truth and entailment which—be it overtly or covertly—hinge upon the notion of truth.

#### 4 PROBABILISTIC SEMANTICS (1)

My concern in this section and the next is with probabilities, i.e. with degrees of rational belief. The probability functions I consider first are *singular* real-valued functions. They will thus take *single* statements of  $L^+$  as their arguments, the way truth-value functions do; but they will take as their values reals from the entire interval  $[0,1]$ , unlike truth-value functions, which merely take the end-points 0 and 1. I shall place on the functions seven constraints ensuring that ‘for any such function  $P^+$  for  $L^+$ , a rational agent might *simultaneously* believe the various statements of  $L^+$ —say  $A_1, A_2, A_3$ , etc.<sup>25</sup>—to the respective extents  $P^+(A_1), P^+(A_2), P^+(A_3)$ , etc.’<sup>26</sup> These constraints, adopted in some cases and adapted in others from [Kolmogorov, 1933; Popper, 1955] and [Gaifman, 1964], are of considerable interest.<sup>27</sup> Not only do they eschew all semantic notions, thus being what one calls *autonomous*; they in fact permit definition of several of them—in particular, logical truth and logical entailment. So, besides freeing probability theory (hence, to some extent, inductive logic) of its past dependence upon deductive logic, they make for a brand-new semantics: *probabilistic semantics*.

<sup>24</sup>For further results concerning model (or Hintikka) sets, see [Hintikka, 1955], [Jeffrey, 1990], [Smullyan, 1968], [Leblanc and Wisdom, 1993] and [Leblanc, 1976].

<sup>25</sup> $A_1$  here the alphabetically first statement of  $L^+$ ,  $A_2$  the second,  $A_3$  the third, etc.

<sup>26</sup>I.e. ensuring that any such function  $P^+$  for  $L^+$  is *coherent* in the sense of [De Finetti, 1937]. The literature on belief functions is considerable. For a survey of early results on coherent belief (particularly betting) functions, see [Carnap and Jeffrey, 1971, p. 105–116]; for a recent study of belief systems, see [Ellis, 1979]. I owe the phrasing in the text to Kent Bendall; hence the quotation marks.

<sup>27</sup>As the reader doubtless knows, Kolmogorov’s functions take sets rather than statements as their arguments. However, they convert into statement-theoretic functions once you think of his complements and intersections as negations and conjunctions. Popper’s functions take what he calls *elements* as their arguments; these, as pointed out in [Popper, 1959, p. 319], may be understood either as sets or as statements. Gaifman’s functions do take statements as their arguments.

Thanks in good part to Popper, the functions studied here thus have a fresh look and a new thrust. They nonetheless are of the most orthodox and in the present context most welcome sort. They accord to negations and conjunctions the very values that the functions in [Kolmogorov, 1933] would to complements and intersections. and they accord to universal quantifications values complying with the substitution interpretation of ‘ $\forall$ ’ (and, quite serviceably, accord such values to *all* universal quantifications).

Popper’s interest eventually shifted from the present functions to the binary ones studied in [Popper, 1959]. It is the latter which figure most prominently in recent contributions to probabilistic semantics. I shall investigate them in the second half of Section 5. For novelty’s sake, however, and because of the close relationship they bear to truth-value functions, I shall devote this section and half the next to singular functions.

Formal details, borrowed from [Leblanc, 1982c], are as follows.

Let  $L^+$  be an arbitrary term extension of  $L$ . By a (*singular*) *probability function* for  $L^+$  I shall understand any function  $P^+$  from the statements of  $L^+$  to the reals that meets the following constraints:

$$\mathbf{C1.} \quad 0 \leq P^+(a)$$

$$\mathbf{C2.} \quad P^+(\neg(A \wedge \neg A)) = 1$$

$$\mathbf{C3.} \quad P^+(A) = P^+(A \wedge B) + P^+(A \wedge \neg B)$$

$$\mathbf{C4.} \quad P^+(A) \leq P^+(A \wedge A)$$

$$\mathbf{C5.} \quad P^+(A \wedge B) \leq P^+(B \wedge A)$$

$$\mathbf{C6.} \quad P^+(A \wedge (B \wedge C)) \leq P^+((A \wedge B) \wedge C)$$

$$\mathbf{C7.} \quad P^+(A \wedge (\forall x)B) = \text{Limit}_{j \rightarrow \infty} P^+(A \wedge \prod_{i=1}^j B(t_i^+ / x)).^{28}$$

---

<sup>28</sup>Information concerning the provenance of **C1–C7** may be welcome.

(i) **C1** and **C2** are statement-theoretic counterparts of two axioms in [Kolmogorov, 1933]; they are known as *Non-negativity* and *Unit Normalisation*. **C3–C6** are borrowed from [Popper, 1955]; they are known as *Complementation*, *Idempotence*, *Commutativity*, and *Associativity*. And **C7** is an adaptation by Bendall of an axiom in [Gaifman, 1964]. Both Gaifman and Bendall, incidentally, use minima where I use limits, but the difference is immaterial here.

(ii) Kolmogorov had two additional axioms, known as *Additivity* and *Continuity*. The statement-theoretic counterpart of the latter calls for infinite disjunctions and, hence, does not belong here. The statement-theoretic counterpart of the former runs:

**C8** If two statements  $A$  and  $B$  of  $L^+$  are logically incompatible in the standard sense (i.e. if  $\neg(A \wedge B)$  is logically true in that sense), then  $P^+(A \vee B) = P^+(A) + P^+(B)$ .

Switching from sets to statements forces adoption of yet another constraint:

**C9**. If  $A$  and  $B$  are logically equivalent in the standard sense (i.e. if  $A \equiv B$  is logically true), then  $P^+(A) = P^+(B)$ . (= *Substitutivity*).

It is readily shown, given Theorem 107 below, that **C1–C2** and **C8–C9** pick out exactly

I shall declare a statement  $A$  of  $L$  *logically true in the probabilistic sense* if—no matter the term extension  $L^+$  of  $L$  and probability function  $P^+$  for  $L^+$ — $P^+(A) = 1$ ; and, where  $S$  is a set of statements of  $L$ , I shall declare  $A$  *logically entailed by  $S$  in the probabilistic sense* if—no matter the term extension  $L^+$  of  $L$  and probability function  $P^+$  for  $L^+$ — $P^+(A) = 1$  if  $P^+(B) = 1$  for each member  $B$  of  $S$ . Equivalently, but more simply,  $A$  may be declared logically true in the probabilistic sense if  $A$  evaluates to 1 on every probability function for  $L$ .

I shall legitimise the foregoing account of logical entailment by showing it strongly sound (= Theorem 100) and strongly complete (= Theorem 105). That the first account of logical truth is legitimate follows, of course, from Theorems 100 and 105; and that the second is follows from the theorems leading to Theorems 100 and 105.

Preparatory to proving Theorem 100, the Strong Soundness Theorem for  $L$  in probabilistic semantics, I establish that (i) each axiom of  $L$  evaluates to 1 on an arbitrary function  $P$  for  $L$  (= Theorem 98) and (ii) the ponential  $B$  of two statements  $A$  and  $A \rightarrow B$  of  $L$  evaluates to 1 on  $P$  if  $A$  and  $A \rightarrow B$  do (= Theorem 87). Proof of (i) is in three steps. I first show that each axiom of  $L$  of sorts **A1–A3** evaluates to 1 on  $P$  (= Theorems 75, 76, and 86). Given this first result and Theorem 87, I next show that if  $\vdash_0 A$  (i.e. if  $A$  is provable in  $L$  by means of just **A1–A3**), then  $A$  evaluates to 1 on  $P$  (= Theorem 88). And, given this second result, I then show that each axiom of  $L$  of sorts **A4–A6**—and, more generally, each axiom of  $L$ —evaluates to 1 on  $P$  (= Theorems 95–98).

the same probability functions as **C1–C6** do.

(iii) Popper uses in place of **C1–C2** the following three constraints:

**C10.**  $P^+(A \wedge B) \leq P^+(A)$  (= *Monotony*)

**C11.** There are at least two distinct statements  $A$  and  $B$  of  $L^+$  such that  $P^+(A) \neq P^+(B)$  (= *Existence*)

**C12.** For each statement  $A$  of  $L^+$  there is a statement  $B$  of  $L^+$  such that  $P^+(A) \leq P^+(B)$  and  $P^+(A \wedge B) = P^+(A) \times P^+(B)$  (= *Multiplication*).

It can be shown that **C3–C6** and **C10–C12** pick out exactly the same probability functions as **C1–C6** do; see [Leblanc, 1982c] on the matter.

(iv) That  $P^+(\neg A) = 1 - P^+(A)$  (= Theorem 68) would not do as a substitute for **C3** is readily shown. Let  $P^+(\neg A)$  be  $1 - P^+(A)$ , and  $P^+(A \wedge B)$  be the smaller of  $P^+(A)$  and  $P^+(B)$ . All of **C1–C2**, Theorem 68, and **C4–C7** will then 'pan out', but **C3** will not. Whether  $P^+(\forall x)A = \text{Limit}_{j \rightarrow \infty} P^+(\prod_{i=1}^j A(t_i^+/x))$  (= Theorem 90) would do as a substitute for **C7** is an open question. It does in [Gaifman, 1964]; but, as Bendall pointed out to me, the non- autonomous constraint that Gaifman would use to pass from Theorem 90 to **C7** is not available here.

Note that the two limits  $\text{Limit}_{j \rightarrow \infty} P^+(A \wedge \prod_{i=1}^j B(t_i^+/x))$  and  $\text{Limit}_{j \rightarrow \infty} P^+(\prod_{i=1}^j A(t_i^+/x))$  exist for every function  $P^+$  for  $L^+$  and all quantifications  $(\forall x)B$  and  $(\forall x)A$  of  $L^+$ . Carnap would rather construe  $P^+(\forall x)A$  as  $\text{Limit}_{j \rightarrow \infty} P^+(A(t_j^+/x))$  (see [Carnap, 1950, p. 302]). But, since this third limit does not always exist, the ensuing function  $P^+$  would not accord values to all universal quantifications, a serious drawback this.

As the reader will notice, the proofs of Theorems 58–98 hold with ‘ $P^+$ ’ everywhere for ‘ $P$ ’. Hence, so do Theorems 59–98 themselves. Hence, so do such among Theorem 58–98 as are needed to prove Theorem 100 (and, further on, Theorem 105).

**THEOREM 58.**  $P(A \wedge B) \leq P(A)$ .

**Proof.**  $P(A \wedge \neg B) \geq 0$  by **C1**. Hence,  $P(A) \geq P(A \wedge B)$  by **C3**. Hence Theorem 58. ■

**THEOREM 59.**  $P(A) = P(A \wedge A)$ .

**Proof.** By **C4** and Theorem 58. ■

**THEOREM 60.**  $P(A \wedge B) = P(B \wedge A)$ .

**Proof.** By **C5**. ■

**THEOREM 61.**  $P((A \wedge B) \wedge C) = P(A \wedge (B \wedge C))$ .

**Proof.**

$$\begin{aligned} P((A \wedge B) \wedge C) &= P(C \wedge (A \wedge B)) && \text{(Theorem 60)} \\ &\leq P((C \wedge A) \wedge B) && \text{(C6)} \\ &\leq P(B \wedge (C \wedge A)) && \text{(Theorem 60)} \\ &\leq P((B \wedge C) \wedge A) && \text{(C6)} \\ &\leq P(A \wedge (B \wedge C)) && \text{(Theorem 60)}. \end{aligned}$$

Hence Theorem 61 by **C6**. ■

**THEOREM 62.**  $P(A) = P(B \wedge A) + P(\neg B \wedge A)$ .

**Proof.** By **C3** and Theorem 60. ■

**THEOREM 63.** *If  $P(A) \leq 0$ , then  $P(A) = 0$ .*

**Proof.** By **C1**. ■

**THEOREM 64.**  $P(A \wedge B) \leq P(B)$ .

**Proof.** By Theorems 58 and 60. ■

**THEOREM 65.**  $P(A \wedge \neg A) = 0$ .

**Proof.**

$$\begin{aligned} P(A \wedge \neg A) &= P(A) - P(A \wedge A) && \text{(C3)} \\ &= 0 && \text{(Theorem 59)}. \end{aligned}$$

■

THEOREM 66.  $P((A \wedge \neg A) \wedge B) = 0$ .

**Proof.**

$$\begin{aligned} P((A \wedge \neg A) \wedge B) &\leq P(A \wedge \neg A) && \text{(Theorem 58)} \\ &\leq 0 && \text{(Theorem 65)} \\ &= 0 && \text{(Theorem 63)}. \end{aligned}$$

■

THEOREM 67.  $P(A) = P(\neg(B \wedge \neg B) \wedge A)$ .

**Proof.**

$$\begin{aligned} P(A) &= P(\neg(B \wedge \neg B) \wedge A) + P((B \wedge \neg B) \wedge A) && \text{(Theorem 62)} \\ &= P(\neg(B \wedge \neg B) \wedge A) && \text{(Theorem 66)}. \end{aligned}$$

■

THEOREM 68.  $P(\neg A) = 1 - P(A)$ .

**Proof.**

$$\begin{aligned} P(\neg A) &= P(\neg(A \wedge \neg A) \wedge \neg A) && \text{(Theorem 67)} \\ &= P(\neg(A \wedge \neg A)) - P(\neg(A \wedge \neg A) \wedge A) && \text{(C3)} \\ &= 1 - P(\neg(A \wedge \neg A) \wedge A) && \text{(C2)} \\ &= 1 - P(A) && \text{(Theorem 67)}. \end{aligned}$$

■

THEOREM 69.  $P(A) \leq 1$ .

**Proof.**  $P(A) = 1 - P(\neg A)$  by Theorem 68. but  $P(\neg A) \geq 0$  by **C1**. Hence, Theorem 69. ■

THEOREM 70. *If  $P(A) \geq 1$ , then  $P(A) = 1$ .*

**Proof.** By Theorem 69. ■

THEOREM 71.  $P(\neg A \wedge (A \wedge B)) = 0$ .

**Proof.**

$$\begin{aligned} P(\neg A \wedge (A \wedge B)) &= P((\neg A \wedge A) \wedge B) && \text{(Theorem 61)} \\ &\leq P(\neg A \wedge A) && \text{(Theorem 58)} \\ &\leq P(A \wedge \neg A) && \text{(Theorem 60)} \\ &\leq 0 && \text{(Theorem 65)} \\ &= 0 && \text{(Theorem 63)}. \end{aligned}$$

■

**THEOREM 72.**  $P(A \wedge B) = P(A \wedge (A \wedge B))$ .

**Proof.**

$$\begin{aligned} P(A \wedge B) &= P(A \wedge (A \wedge B)) + P(\neg A \wedge (A \wedge B)) && \text{(Theorem 62)} \\ &= P(A \wedge (A \wedge B)) && \text{(Theorem 71)}. \end{aligned}$$

■

**THEOREM 73.**  $P(A \rightarrow B) = 1$  iff  $P(A \wedge \neg B) = 0$ .

**Proof.** By  $D_{\rightarrow}$  and Theorems 68.

■

**THEOREM 74.**  $P(A \rightarrow B) = 1$  iff  $P(A) = P(A \wedge B)$ .

**Proof.** By Theorems 73 and **C3**.

■

**THEOREM 75.**  $P(A \rightarrow (A \wedge A)) = 1$  (= Axiom Schema **A1**).

**Proof.**

$$\begin{aligned} P(A) &= P(A \wedge A) && \text{(Theorem 59)} \\ &= P(A \wedge (A \wedge A)) && \text{(Theorem 72)}. \end{aligned}$$

Hence, Theorem 75 by Theorem 74.

■

**THEOREM 76.**  $P((A \wedge B) \rightarrow A) = 1$  (= Axiom Schema **A2**).

**Proof.**

$$\begin{aligned} P(A \wedge B) &= P(A \wedge (A \wedge B)) && \text{(Theorem 72)} \\ &= P((A \wedge B) \wedge A) && \text{(Theorem 60)}. \end{aligned}$$

Hence, Theorem 76 by Theorem 74.

■

**THEOREM 77.**  $P((A \wedge B) \rightarrow B) = 1$ .

**Proof.**

$$\begin{aligned} P(A \wedge B) &= P((A \wedge B) \wedge B) + P((A \wedge B) \wedge \neg B) && \text{(C3)} \\ &= P((A \wedge B) \wedge B) + P(A \wedge (B \wedge \neg B)) && \text{(Theorem 61)} \\ &= P((A \wedge B) \wedge B) + ((B \wedge \neg B) \wedge A) && \text{(Theorem 60)} \\ &= P((A \wedge B) \wedge B) && \text{(Theorem 66)}. \end{aligned}$$

Hence, Theorem 77 by Theorem 74.

■

**THEOREM 78.** If  $P(A \rightarrow B) = 1$  and  $P(B \rightarrow C) = 1$ , then  $P(A \rightarrow C) = 1$ .



**Proof.** Suppose first that  $P(A \rightarrow B) = 1$ . Then

$$\begin{aligned}
 0 &= P(A \wedge \neg B) && \text{(Theorem 73)} \\
 &\geq P(\neg C \wedge (A \wedge \neg B)) && \text{(Theorem 64)} \\
 &= P(\neg C \wedge (A \wedge \neg B)) && \text{(Theorem 63)} \\
 &= P((\neg C \wedge A) \wedge \neg B) && \text{(Theorem 61)}.
 \end{aligned}$$

Suppose next that  $P(B \rightarrow C) = 1$ . Then

$$\begin{aligned}
 0 &= P(B \wedge \neg C) && \text{(Theorem 73)} \\
 &\geq P((B \wedge \neg C) \wedge A) && \text{(Theorem 58)} \\
 &= P((B \wedge \neg C) \wedge A) && \text{(Theorem 63)} \\
 &= P(B \wedge (\neg C \wedge A)) && \text{(Theorem 61)} \\
 &= P((\neg C \wedge A) \wedge B) && \text{(Theorem 60)}.
 \end{aligned}$$

Hence,  $P(\neg C \wedge A) = 0$  by **C3**; hence,  $P(A \wedge \neg C) = 0$  by Theorem 60; and, hence,  $P(A \rightarrow C) = 1$  by Theorem 73. Hence, Theorem 78. ■

**THEOREM 79.** *If  $P(A \rightarrow (B \wedge C)) = 1$ , then  $P(A \rightarrow B) = 1$ .*

**Proof.** Suppose  $P(A \rightarrow (B \wedge C)) = 1$ . Then

$$\begin{aligned}
 P(A) &= P(A \wedge (B \wedge C)) && \text{(Theorem 74)} \\
 &= P((A \wedge B) \wedge C) && \text{(Theorem 61)} \\
 &\leq P(A \wedge B) && \text{(Theorem 58)}.
 \end{aligned}$$

Hence,  $P(A) = P(A \wedge B)$  by Theorem 58; and, hence,  $P(A \rightarrow B) = 1$  by Theorem 74. Hence, Theorem 79. ■

**THEOREM 80.**  $P(A \wedge \neg \neg B) = P(A \wedge B)$ .

**Proof.** By **C3**

$$P(A) = P(A \wedge \neg B) + P(A \wedge \neg \neg B)$$

and

$$P(A) = P(A \wedge \neg B) + P(A \wedge B).$$

Hence, Theorem 80. ■

**THEOREM 81.** *If  $P(A \rightarrow (B \rightarrow C)) = 1$  and  $P(A \rightarrow B) = 1$ , then  $P(A \rightarrow C) = 1$ .*

**Proof.** Suppose first that  $P(A \rightarrow (B \rightarrow C)) = 1$ . Then

$$\begin{aligned}
 0 &= P(A \wedge \neg \neg (B \wedge \neg C)) && \text{(Theorem 73 and } D_{\rightarrow}\text{)} \\
 &= P(A \wedge (B \wedge \neg C)) && \text{(Theorem 80)} \\
 &= P((A \wedge B) \wedge \neg C) && \text{(Theorem 61)} \\
 &= P(\neg C \wedge (A \wedge B)) && \text{(Theorem 60)} \\
 &= P((\neg C \wedge A) \wedge B) && \text{(Theorem 61)}.
 \end{aligned}$$

Suppose next that  $P(A \rightarrow B) = 1$ . Then

$$\begin{aligned}
0 &= P(A \wedge \neg C) && \text{(Theorem 73)} \\
&\geq P(\neg C \wedge (A \wedge \neg B)) && \text{(Theorem 64)} \\
&= P(\neg C \wedge (A \wedge \neg B)) && \text{(Theorem 63)} \\
&= P((\neg C \wedge A) \wedge \neg B) && \text{(Theorem 61)}.
\end{aligned}$$

Hence,  $P(\neg C \wedge A) = 0$  by **C3**; hence,  $P(A \wedge \neg C) = 0$  by Theorem 60; and, hence,  $P(A \rightarrow C) = 1$  by Theorem 73. Hence, Theorem 81. ■

**THEOREM 82.** *If  $P(A \rightarrow B) = 1$  and  $P(A \rightarrow C) = 1$ , then  $P(A \rightarrow (B \wedge C)) = 1$ .*

**Proof.** Suppose first that  $P(A \rightarrow B) = 1$ . Then

$$\begin{aligned}
P(A) &= P(A \wedge B) && \text{(Theorem 74)} \\
&= P((A \wedge B) \wedge C) + P((A \wedge B) \wedge \neg C) && \text{(C3)} \\
&= P(A \wedge (B \wedge C)) + P((A \wedge B) \wedge \neg C) && \text{(Theorem 61)} \\
&= P(A \wedge (B \wedge C)) + P(\neg C \wedge (A \wedge B)) && \text{(Theorem 60)}.
\end{aligned}$$

Suppose next that  $P(A \rightarrow C) = 1$ . Then

$$\begin{aligned}
0 &= P(A \wedge \neg C) && \text{(Theorem 73)} \\
&= P(\neg C \wedge A) && \text{(Theorem 60)} \\
&\geq P((\neg C \wedge A) \wedge B) && \text{(Theorem 58)} \\
&= P((\neg C \wedge A) \wedge B) && \text{(Theorem 63)} \\
&= P(\neg C \wedge (A \wedge B)) && \text{(Theorem 61)}.
\end{aligned}$$

Hence,  $P(A) = P(A \wedge (B \wedge C))$ ; and, hence,  $P(A \rightarrow (B \wedge C)) = 1$  by Theorem 74. Hence, Theorem 82. ■

**THEOREM 83.** *If  $P(A \rightarrow (B \wedge C)) = 1$ , then  $P(A \rightarrow C) = 1$ .*

**Proof.** Suppose that  $P(A \rightarrow (B \wedge C)) = 1$ . Then

$$\begin{aligned}
P(A) &= P(A \wedge (B \wedge C)) && \text{(Theorem 74)} \\
&= P((A \wedge B) \wedge C) && \text{(Theorem 61)} \\
&= P(C \wedge (A \wedge B)) && \text{(Theorem 60)} \\
&= P((C \wedge A) \wedge B) && \text{(Theorem 61)} \\
&\leq P(C \wedge A) && \text{(Theorem 58)} \\
&\leq (A \wedge C) && \text{(Theorem 60)}.
\end{aligned}$$

Hence,  $P(A \rightarrow C) = 1$  by the same reasoning as in the proof of Theorem 79. Hence, Theorem 83. ■

**THEOREM 84.** *If  $P(A \rightarrow B) = 1$  and  $P(A \rightarrow \neg B) = 1$ , then  $P(A) = 0$ .*

**Proof.** Suppose  $P(A \rightarrow B) = 1$ . Then  $P(A \wedge \neg B) = 0$  by Theorem 73. Suppose further that  $P(A \rightarrow \neg B) = 1$ . Then  $P(A \wedge \neg\neg B) = 0$  by Theorem 73 again. Hence,  $P(A) = 0$  by **C3**. Hence, Theorem 84. ■

**THEOREM 85.** *If  $P((A \wedge B) \rightarrow C) = 1$ , then  $P(A \rightarrow (B \rightarrow C)) = 1$ .*

**Proof.** Suppose  $P((A \wedge B) \rightarrow C) = 1$ . Then

$$\begin{aligned} 0 &= P((A \wedge B) \wedge \neg C) && \text{(Theorem 73)} \\ &= P(A \wedge (B \wedge \neg C)) && \text{(Theorem 61)} \\ &= P(A \wedge \neg\neg(B \wedge \neg C)) && \text{(Theorem 80)}. \end{aligned}$$

Hence,  $P(A \rightarrow (B \rightarrow C)) = 1$  by Theorem 74 and  $D_{\rightarrow}$ . Hence, Theorem 85. ■

**THEOREM 86.**  $P((A \rightarrow B) \rightarrow (\neg(B \wedge C) \rightarrow \neg(C \wedge A))) = 1$ .  
(= *Axiom Schema A3*).

**Proof.** (i) By Theorem 77

$$P((((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow (C \wedge A)) = 1.$$

But by Theorem 76

$$P((C \wedge A) \rightarrow C) = 1.$$

Hence, by Theorem 78

$$P((((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow C) = 1.$$

(ii) Similarly, but with Theorem 77 substituting for Theorem 76.

$$P((((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow A) = 1.$$

But by Theorem 76

$$P((((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow ((A \rightarrow B) \wedge \neg(B \wedge C))) = 1,$$

and, hence, by Theorem 79

$$P((((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow (A \rightarrow B)) = 1.$$

Hence, by Theorem 81

$$P((((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow B) = 1.$$

(iii) By (i)–(ii) and Theorem 82

$$P((((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow (B \wedge C)) = 1.$$

But by Theorem 76

$$P(((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow ((A \rightarrow B) \wedge \neg(B \wedge C)) = 1,$$

and, hence, by Theorem 83

$$P(((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) \rightarrow \neg(B \wedge C) = 1.$$

Hence, by Theorem 84

$$P(((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge (C \wedge A)) = 0,$$

hence, by Theorem 80

$$P(((A \rightarrow B) \wedge \neg(B \wedge C)) \wedge \neg\neg(C \wedge A)) = 0,$$

hence, by Theorem 73

$$P(((A \rightarrow B) \wedge \neg(B \wedge C)) \rightarrow \neg(C \wedge A)) = 1.$$

and, hence, by Theorem 85

$$P((A \rightarrow B) \rightarrow (\neg(B \wedge C) \rightarrow \neg(C \wedge A))) = 1.$$

■

**THEOREM 87.** *If  $P(A) = 1$  and  $P(A \rightarrow B) = 1$ , then  $P(B) = 1$  (= Modus Ponens).*

**Proof.** Suppose  $P(A) = 1$  and  $P(A \rightarrow B) = 1$ . Then by Theorem 74  $P(A \wedge B) = 1$ ; hence, by Theorem 64  $P(B) \geq 1$ ; and, hence, by Theorem 70  $P(B) = 1$ . ■

Hence:

**THEOREM 88.** *If  $\vdash_0 A$ , then  $P(A) = 1$ .*

**Proof.** Suppose the column made up of  $B_1, B_2, \dots, B_p$  constitutes a proof of  $A$  in  $L$  by means of just **A1–A3**. It is easily shown by mathematical induction on  $i$  that, for each  $i$  from 1 through  $p$ ,  $P(B_i) = 1$ . For in the case that  $B_i$  is an axiom,  $P(B_i) = 1$  by Theorems 75, 76 and 86; and in the case that  $B_i$  is the ponential of two pervious entries in the column,  $P(B_i) = 1$  by Theorem 87 and the hypothesis of the induction. Hence,  $P(B_p) = 1$ , i.e.  $P(A) = 1$ . Hence, Theorem 88. ■

And hence:

**THEOREM 89.** *If  $\vdash_0 A \leftrightarrow B$ , then  $P(A) = P(B)$ .*

**Proof.** If  $\vdash_0 A \leftrightarrow B$ , then  $\vdash_0 A \rightarrow B$  and  $\vdash_0 B \rightarrow A$ ; hence, by Theorem 74,

$$P(A) = P(A \wedge B)$$

and

$$P(B) = P(B \wedge A),$$

and, hence, by Theorem 60

$$P(A) = P(B).$$

■

With Theorem 89 on hand, I am ready to show that all axioms of  $L$  of sorts **A4–A6**—and, more generally, all axioms of  $L$ — evaluate to 1 on any probability function for  $L$ .

**THEOREM 90.**  $P((\forall x)A) = \text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j A(t_i/x))$ .

**Proof.** By **C7**

$$P(\neg(A \wedge \neg A) \wedge (\forall x)A) = \text{Limit}_{j \rightarrow \infty} P(\neg(A \wedge \neg A) \wedge \prod_{i=1}^j A(t_i/x)).$$

Hence, Theorem 90 by Theorem 67 and the definition of a limit. ■

**THEOREM 91.** *If  $P(A) = 1$  and  $P(B) = 1$ , then  $P(A \wedge B) = 1$ .*

**Proof.** Suppose  $P(A) = 1$ . Then  $P(A \wedge B) + P(A \wedge \neg B) = 1$  by **C3**. Suppose also that  $P(B) = 1$ . then  $P(\neg B) = 0$  by Theorem 68; hence,  $P(A \wedge \neg B) \leq 0$  by Theorem 64; and, hence,  $P(A \wedge \neg B) = 0$  by Theorem 63. Hence, Theorem 91. ■

**THEOREM 92.** *If  $P(A_i) = 1$  for each  $i$  from 1 through  $j$ , then  $P(\prod_{i=1}^j A_i) = 1$ .*

**Proof.** By Theorem 91 and mathematical induction on  $j$ . ■

**THEOREM 93.** *If  $P(A(t_i/x)) = 1$  for each  $i$  from 1 on, then  $P((\forall x)A) = 1$ .*

**Proof.** By Theorems 92 and 90, and the definition of a limit. ■

**THEOREM 94.** *If  $P(A \rightarrow B(t_i/x)) = 1$  for each  $i$  from 1 on, then  $P(A \rightarrow (\forall x)B) = 1$ .*

**Proof.** Suppose

$$P(A \rightarrow B(t_i/x)) = 1$$

for each  $i$  from 1 on. Then by Theorem 82 and mathematical induction on  $j$

$$P(A \rightarrow \prod_{i=1}^j B(t_i/x)),$$

for each  $j$  from 1 on, hence, by Theorem 74

$$P(A) = P(A \wedge \prod_{i=1}^j B(t_i/x)),$$

hence, by the definition of a limit

$$P(A) = \text{Limit}_{j \rightarrow \infty} P(A \wedge \prod_{i=1}^j B(t_i/x)),$$

hence, by **C7**

$$P(A) = P(A \wedge (\forall x)B),$$

and, hence, by Theorem 74

$$P(A \rightarrow (\forall x)B) = 1.$$

Hence, Theorem 94. ■

**THEOREM 95.**  $P(A \rightarrow (\forall x)A) = 1$  (= *Axiom Schema A4*).

**Proof.** Since  $x$  here is sure to be foreign to  $a$ ,  $A(t_i/x)$  is sure to be the same as  $A$ . Hence,

$$\vdash_0 (A \wedge \prod_{i=1}^j A(t_i/x)) \leftrightarrow A,$$

hence, by Theorem 89 and the definition of a limit

$$\text{Limit}_{j \rightarrow \infty} P(A \wedge \prod_{i=1}^j A(t_i/x)) = P(A),$$

hence, by **C7**

$$P(A \wedge (\forall x)A) = P(A),$$

and, hence, by Theorem 74

$$P(A \rightarrow (\forall x)A) = 1. ■$$

**THEOREM 96.**  $P((\forall x)A \rightarrow A(T/x)) = 1$  (= *Axiom Schema A5*).

**Proof.** Let  $T$  be the  $k$ th term of  $L$ . So long as  $j \geq k$ ,

$$\vdash_0 (A(T/x) \wedge \prod_{i=1}^j A(t_i/x)) \leftrightarrow \prod_{i=1}^j A(t_i/x).$$

Hence, by Theorem 89 and the definition of a limit

$$\text{Limit}_{j \rightarrow \infty} P(A(T/x) \wedge \prod_{i=1}^j A(t_i/x)) = \text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j A(t_i/x)),$$

hence, by **C7** and Theorem 90

$$P(A(T/x) \wedge (\forall x)A) = P((\forall x)A),$$

hence, by Theorem 60

$$P((\forall x)A \wedge A(t/x)) = P((\forall x)A),$$

and, hence, by Theorem 74

$$P((\forall x)A \rightarrow A(T/x)) = 1.$$

■

**THEOREM 97.**  $P((\forall x)(A \rightarrow B) \rightarrow ((\forall x)A \rightarrow (\forall x)B)) = 1$  (*Axiom Schema A6*).

**Proof.** Let  $T$  be an arbitrary term of  $L$ . By Theorem 76

$$P(((\forall x)(A \rightarrow B) \wedge (\forall x)A) \rightarrow (\forall x)(A \rightarrow B)) = 1.$$

But by Theorem 96

$$P((\forall x)(A \rightarrow B) \rightarrow (A \rightarrow B)(T/x)) = 1.$$

Hence, by Theorem 78

$$P(((\forall x)(A \rightarrow B) \wedge (\forall x)A) \rightarrow (A \rightarrow B)(T/x)) = 1.$$

Similarly, but using Theorem 77 in place of Theorem 76,

$$P(((\forall x)(A \rightarrow B) \wedge (\forall x)A) \rightarrow A(T/x)) = 1.$$

Hence, by Theorem 81

$$P(((\forall x)(A \rightarrow B) \wedge (\forall x)A) \rightarrow B(T/x)) = 1,$$

hence, by Theorem 94 and the hypothesis on  $T$

$$P(((\forall x)(A \rightarrow B) \wedge (\forall x)A) \rightarrow (\forall x)B) = 1,$$

and, hence, by Theorem 85

$$P((\forall x)(A \rightarrow B) \rightarrow ((\forall x)A \rightarrow (\forall x)B)) = 1.$$

■

**THEOREM 98.** *If  $A$  is an axiom of  $L$ , then  $P(A) = 1$ .*

**Proof.** Suppose  $A$  is an axiom of  $L$ , in which case  $A$  is bound by Theorem 1 to be of the sort

$$(\forall x_1)(\forall x_2) \dots (\forall x_n)(B(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n)),$$

where  $n \geq 0$  and  $B$  is one of the sorts **A1–A6**. Proof that  $P(A) = 1$  will be by mathematical induction on  $n$ .

*Basis:*  $n = 0$ . Then  $A(= B)$  is of one of the sorts **A1–A6** and, hence,  $P(A) = 1$  by Theorems 75, 76, 86 and 95–97.

*Inductive Step:*  $n > 0$ . By Theorem 2

$$((\forall x_2) \dots (\forall x_n)(B(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n)))(t_i/x_1)$$

constitutes an axiom of  $L$  for each  $i$  from 1 on. Hence, by the hypothesis of the induction

$$P(((\forall x_2) \dots (\forall x_n)(B(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n)))(t_i/x_i)) = 1$$

for each  $i$  from 1 on. Hence, by Theorem 93,

$$P((\forall x_1)(\forall x_2) \dots (\forall x_n)(B(x_1, x_2, \dots, x_n/T_1, T_2, \dots, T_n))) = 1.$$

Hence,  $P(A) = 1$ .

■

With Theorems 87 and 98 on hand, an obvious induction readily delivers Theorem 99, which in turn delivers the Strong Soundness Theorem for  $L$  in probabilistic semantics. Suppose indeed that the column made up of  $B_1, B_2, \dots, B_p$  constitutes a proof of  $A$  from  $S$  in  $L$ , and suppose each member of  $S$  evaluates to 1 on  $P$ . In the case that  $B_i$  belongs to  $S$ ,  $P(B_i) = 1$  by Theorem 98; and in the case that  $B_i$  is the ponential of two earlier entries in the column,  $P(B_i) = 1$  by Theorem 87 and the hypothesis of the induction. Hence,  $P(B_i) = 1$  for each  $i$  from 1 through  $p$ . Hence  $P(B_p) = 1$ . hence,  $P(A) = 1$ .

**THEOREM 99.** *If  $S \vdash A$ , then  $A$  evaluates to 1 on  $P$  if all the members of  $S$  do.*



But Theorems 87 and 98 hold with ‘ $P^+$ ’ for ‘ $P$ ’. Hence:

**THEOREM 100.** *If  $S \vdash A$ , then—no matter the term extension  $L^+$  of  $L$  and probability function  $P^+$  for  $L^+$ — $A$  evaluates to 1 on  $P^+$  if all the members of  $S$  do, i.e.  $A$  is logically entailed by  $S$  in the probabilistic sense (= The Strong Soundness Theorem for  $L$  in Probabilistic Semantics).*

Proof of the converse of Theorem 100 calls for an additional definition and two additional lemmas.

Let  $S$  be an arbitrary set of statements of  $L$ . By the *probability associate* of  $S$  in  $L$ , I shall understand the function  $P_S$  such that for any statement  $A$  of  $L$ :

$$P_S(A) = \begin{cases} 1 & \text{if } S \vdash A \\ 0 & \text{otherwise.} \end{cases}$$

I first establish that the probability associate of  $S$  in  $L$  meets Constraints **C1–C2** and **C4–C6** under all circumstances, meets Constraint **C3** as well when  $S$  is maximally consistent in  $L$ , and meets Constraint **C7** as well when  $S$  is  $\omega$ -complete in  $L$ . I then establish that when  $S$  is maximally consistent in  $L$ , a statement  $A$  of  $L$  belongs to  $S$  iff it evaluates to 1 on the probability associate of  $S$ .

**THEOREM 101.** *Let  $S$  be a set of statements of  $L$ , and  $P_S$  be the probability associate of  $S$  in  $L$ .*

- (a)  $P_S$  meets Constraints **C1–C2** and **C4–C6**.
- (b) If  $S$  is maximally consistent in  $L$ , then  $P_S$  meets Constraint **C3** as well.
- (c) If  $S$  is  $\omega$ -complete in  $L$ , then  $P_S$  meets Constraint **C7** as well.
- (d) If  $S$  is maximally consistent and  $\omega$ -complete in  $L$ , then  $P_S$  constitutes a probability function for  $L$ .<sup>29</sup>

**Proof.**

(a) follows from the definition of  $P_S$  and elementary facts about provability in  $L$ .

(b) Suppose  $S$  is maximally consistent in  $L$ , and suppose first that  $P_S(A) = 1$ . Then  $S \vdash A$  (by the definition of  $P_S$ ). Hence  $S \vdash A \wedge B$  or  $S \vdash A \wedge \neg B$ , but not both. Hence, one of  $P_S(A \wedge B)$  and  $P_S(A \wedge \neg B)$  equals 1 and the other 0. Hence,  $P_S(A) = P_S(A \wedge B) + P_S(A \wedge \neg B)$ . Suppose next that

<sup>29</sup>Note that if a set of statements of  $L$  is infinitely extendible in  $L$ , then  $S$  is sure to be  $\omega$ -complete and hence its probability associate is sure to meet **C7**. The result, exploited in [Morgan and Leblanc, 1983c], is of no use here: the probability associate of  $S$  cannot meet **C3** unless  $S$  is maximally consistent, in which case  $S$  sports *every* term of  $L$ .

$P_S(A) = 0$ . Then  $S \not\vdash A$  (by the definition of  $P_S$ ). Hence neither  $S \vdash A \wedge B$  nor  $S \vdash A \wedge \neg B$ . Hence, both  $P_S(A \wedge B)$  and  $P_S(A \wedge \neg B)$  equal 0. Hence, again,  $P_S(A) = P_S(A \wedge B) + P_S(A \wedge \neg B)$ .

(c) Suppose  $S$  is  $\omega$ -complete in  $L$ , and suppose first that  $P_S(A \wedge (\forall x)B) = 1$ . Then  $S \vdash A \wedge (\forall x)B$  (by the definition of  $P_S$ ). Hence,  $S \vdash A \wedge \prod_{i=1}^j B(t_i/x)$  for every  $j$  from 1 on. Hence,  $P_S(A \wedge \prod_{i=1}^j B(t_i/x)) = 1$  for every  $j$  from 1 on. Hence,  $\text{Limit}_{j \rightarrow \infty} P_S(A \wedge \prod_{i=1}^j B(t_i/x)) = 1$ . Suppose next that  $P_S(A \wedge (\forall x)B) = 0$ . Then  $S \not\vdash A \wedge (\forall x)B$  (by the definition of  $P_S$ ). Hence, either  $S \not\vdash A$  or  $S \not\vdash (\forall x)B$  (or both). Now, if  $S \not\vdash A$ , then  $S \not\vdash A \wedge \prod_{i=1}^j B(t_i/x)$  for any  $j$ , hence  $P_S(A \wedge \prod_{i=1}^j B(t_i/x)) = 0$  for every  $j$ , and hence  $\text{Limit}_{j \rightarrow \infty} P_S(A \wedge \prod_{i=1}^j B(t_i/x)) = 0$ . If, on the other hand,  $S \not\vdash (\forall x)B$ , then  $S \not\vdash B(t/x)$  for at least one term  $T$  of  $L$ , this because  $S$  is  $\omega$ -complete in  $L$ . Hence, there is a  $k$  such that, for every  $j$  from  $k$  on,  $S \not\vdash A \wedge \prod_{i=1}^j B(t_i/x)$ . Hence, there is a  $k$  such that, for every  $j$  from  $k$  on,  $P_S(A \wedge \prod_{i=1}^j B(t_i/x)) = 0$ ; and, hence,  $\text{Limit}_{j \rightarrow \infty} P_S(A \wedge \prod_{i=1}^j B(t_i/x)) = 0$ . Hence,  $P_S$  meets Constraint **C7**. ■

**THEOREM 102.** *Let  $S$  be a set of statements of  $L$  that is maximally consistent and  $\omega$ -complete in  $L$ , and let  $P_S$  be the probability associate of  $S$  in  $L$ . Then a statement  $A$  of  $L$  belongs to  $S$  iff  $P_S(A) = 1$ .*

**Proof.** If  $A$  belongs to  $S$ , then  $S \vdash A$  and, hence,  $P_S(A) = 1$  by the definition of  $P_S$ . If  $A$  does not belong to  $S$ , then by Theorem 3  $\neg A$  belongs to  $S$ , hence,  $P_S(\neg A) = 1$  by the definition of  $P_S$ , hence,  $P_S(A) = 0$  by Theorems 101 (d) and 68, and, hence,  $P_S(A) \neq 1$ . Hence, Theorem 102. ■

Now suppose, as on earlier occasions, that  $S \not\vdash A$  where  $S$  is infinitely extendible in  $L$ . Then by Theorem 4 the Henkin extension  $\mathcal{H}(S \cup \{\neg A\})$  of  $S \cup \{\neg A\}$  in  $L$  is sure, as usual, to be maximally consistent and  $\omega$ -complete in  $L$ ; hence, by Theorem 102, every member of  $\mathcal{H}(S \cup \{\neg A\})$  is sure to evaluate to 1 on the probability associate of  $\mathcal{H}(S \cup \{\neg A\})$  in  $L$ . But by Theorem 101 the probability associate in question constitutes a probability function for  $L$ . Hence, by Theorem 68:

**THEOREM 103.** *Let  $S$  be a set of statements of  $L$  that is infinitely extendible in  $L$ , and let  $A$  be an arbitrary statement of  $L$ . If  $S \not\vdash A$ , then there is a probability function for  $L$  on which all members of  $S$  evaluate to 1 but  $A$  does not.*

But all of Theorems 4, 102, 101 and 68 hold with ' $L^+$ ' in place of ' $L$ ' and ' $P^+$ ' in place of ' $P$ '. Hence:

**THEOREM 104.** *Let  $S$  and  $A$  be as in Theorem 103. If—no matter the term extension  $L^+$  of  $L$  and probability function  $P^+$  for  $L^+$ — $A$  evaluates to 1 on  $P^+$  if all members of  $S$  do, then  $S \vdash A$ .*

When  $S$  is *not* infinitely extendible in  $L$ ,  $S \cup \{\neg A\}$  is sure—as usual—to be infinitely extendible in  $L^\infty$ . Hence, if  $S \not\vdash A$ , then by the analogues of Theorems 4, 101, 102 and 68 for  $L^+$  there is sure to be a term extension  $L^+$  of  $L$  ( $L^\infty$ ) and a probability function  $P^+$  for  $L^+$  (the probability associate in  $L^\infty$  of the Henkin extension of  $S \cup \{\neg A\}$  in  $L^\infty$ ) such that all members of  $S$  evaluate to 1 on  $P^+$  but  $A$  does not. Hence, given Theorem 104:

**THEOREM 105.** *Let  $S$  be an arbitrary set of statements of  $L$  and  $A$  be an arbitrary statement of  $L$ . If—no matter the term extension  $L^+$  of  $L$  and probability function  $P^+$  for  $L^+$ — $A$  evaluates to 1 on  $P^+$  if all members of  $S$  do, i.e. if  $A$  is logically entailed by  $S$  in the probabilistic sense, then  $S \vdash A$  (= The Strong Completeness Theorem for  $L$  in Probabilistic Semantics).*

Hence, appealing to Theorem 100 and taking  $S$  in each case to be  $\emptyset$ :

**THEOREM 106.** *Let  $A$  be an arbitrary statement of  $L$ . Then  $\vdash A$  iff—no matter the term extension  $L^+$  of  $L$  and probability function  $p^+$  for  $L^+$ — $A$  evaluates to 1 on  $P^+$ , i.e. iff  $A$  is logically true in the probabilistic sense.*

As mentioned earlier, Theorems 100 and 105 legitimise the account on page 87 of logical entailment; Theorem 106 legitimises the first account there of logical truth; and the following corollary of Theorems 99 and 103 legitimises the second:

**THEOREM 107.** *Let  $A$  be as in Theorem 106. Then  $\vdash A$  iff  $A$  evaluates to 1 on every probability function for  $L$ .*

## 5 PROBABILISTIC SEMANTICS II

Getting on with the agenda of page 87, I establish here that (singular) probability theory is but a generalisation of truth-value theory, and then turn to *binary* probability functions. These too, the reader will recall, permit definition of logical truth, logical entailment, and such.

As documented in Section 3, truth-value semantics can be couched in either of two idioms: that of truth-value assignments and that of truth-value functions. The former idiom makes for a smooth and easy transition from truth-functional truth to logical truth, truth-functional entailment to logical entailment, etc. (write ‘substatements’ for ‘components’). However, when it comes to investigating the relationship between truth-value theory and probability theory, the latter idiom is the handier one, and I accordingly switch to it.

I indicated on page 76 that I would eventually refer to both the truth-value functions for  $L$  and the probability ones by means of a single letter ‘ $P$ ’, and use 0 and 1 as truth values. That time has come. From here on a *truth-value function* for  $L$  will thus be any function  $P$  from the statements of  $L$  to  $\{0, 1\}$  that meets the following three constraints:

**B1.**  $P(\neg A) = 1 - P(A)$

**B2.**  $P(A \wedge B) = P(A) \times P(B)$

**B3.**  $P((\forall x)A) = 1$  iff  $P(A(T/x)) = 1$  for each term  $T$  of  $L$ .

Thus understood, a truth-value function for  $L$  is obviously a *two-valued* function from the statements of  $L$  to the reals that meets Constraints **B1–B3**. That, conversely, a two-valued function from the statements of  $L$  to the reals is sure, if it meets Constraints **B1–B3**, to have 0 and 1 as its two values and hence be a truth-value function for  $L$ , follows from clause (a) in Theorem 108 ( $\prod_{i=1}^n P(A_i)$  in the proof of that clause is short of course for  $P(A_1) \times P(A_2) \times \cdots \times P(A_n)$ ).

**THEOREM 108.** *Let  $P$  be a two-valued function from the statements of  $L$  to the reals.*

(a) *If  $P$  meets Constraint **B1–B2**, then  $P$  has 0 and 1 as its values.*

(b) *If  $P$  meets Constraints **C1–C4** then  $P$  again has 0 and 1 as its values.*

**Proof.**

(a) Suppose  $P$  meets Constraint **B2**, in which case

$$P\left(\prod_{i=1}^n A_i\right) = \prod_{i=1}^n P(A_i)$$

for each  $n$  from 1 on, and suppose at least one value of  $P$  were some real  $r$  other than 0 and 1. Then  $P$  would number all of  $r, r^2, r^3$ , etc. among its values and, hence, would not—as supposed—be two-valued. Hence, (a).

(b)  $P(\neg(A \wedge \neg A))$  equals 1 by virtue of Constraint **C2**, and—as the proof of Theorem 65 attests— $P(A \wedge \neg A)$  equals 0 by virtue of Constraints **C1, C3, and C4**. Hence, (b). ■

So:

**THEOREM 109.**  *$P$  constitutes a truth-value function for  $L$  iff  $P$  is a two-valued function from the statements of  $L$  to the reals that meets Constraints **B1–B3**.*

Theorem 109 is the first, and more trite, step in my proof of Theorem 115.

I next show that:

**THEOREM 110.** *Let  $P$  be a two-valued function from the statements of  $L$  to the reals. If  $P$  meets Constraints **B1–B3**, then  $P$  meets Constraints **C1–C7** as well.*

**Proof.** Because of the hypothesis on  $P$  and Theorem 108 (a),  $P$  has but 0 and 1 as values. (i) Proof that under them  $P$  meets Constraints **C1–C2** and **C4–C6** is immediate. (ii) Proof that under these circumstances  $P$  meets Constraint **C3** is as follows:

$$\begin{aligned}
P(A) &= (P(A) \times P(B)) + P(A) - (P(A) \times P(B)) \\
&= P(A \wedge B) + P(A) - (P(A) \times P(B)) && \text{(B2)} \\
&= P(A \wedge B) + P(A)(1 - P(B)) \\
&= P(A \wedge B) + (P(A) \times P(\neg B)) && \text{(B1)} \\
&= P(A \wedge B) + P(A \wedge \neg B) && \text{(B2)}
\end{aligned}$$

(iii) Proof that under these circumstances  $P$  meets Constraint **C7** is as follows. Suppose first that  $P(A \wedge (\forall x)B) = 1$ . Then by Constraint **B2**  $P(A) \times P((\forall z)B) = 1$ ; hence,  $P(A) = P((\forall x)B) = 1$ ; hence, by **B3**  $P(B(T/x)) = 1$  for every term  $T$  of  $L$ ; hence, by Constraint **B2**  $P(A \wedge \prod_{i=1}^j B(t_i/x)) = 1$  for each  $j$  from 1 on; and, hence,  $\text{Limit}_{j \rightarrow \infty} P(A \wedge \prod_{i=1}^j B(t_i/x)) = 1$ . Suppose next that  $P(A \wedge (\forall x)B) = 0$ . Then by Constraint **B2**  $P(A) \times P((\forall x)B) = 0$  and, hence,  $P(A) = 0$  or  $P((\forall x)B) = 0$ . Now, if  $P(A) = 0$ , then by Constraint **B2**  $P(A \wedge \prod_{i=1}^j B(t_i/x)) = 0$  for each  $j$  from 1 on, and hence  $\text{Limit}_{j \rightarrow \infty} P(A \wedge \prod_{i=1}^j B(t_i/x)) = 0$ . If, on the other hand,  $P((\forall x)B) = 0$ , then by Constraint **B3**  $P(B(T/x)) = 0$  for at least one term  $T$  of  $L$ , hence by Constraint **B2** there is a  $k$  such that for every  $j$  from  $k$  on  $P(A \wedge \prod_{i=1}^j B(t_i/x)) = 0$ , and hence  $\text{Limit}_{j \rightarrow \infty} P(A \wedge \prod_{i=1}^j B(t_i/x)) = 0$ . ■

Proof of the converse of Theorem 110 calls for two lemmas. Note that in the first  $P$  may have anywhere from two to  $\aleph_0$  values, but in the second it is understood to have but two.

**THEOREM 111.** *Let  $P$  be a function from the statements of  $L$  to the reals that meets Constraints **C1–C5**.*

- (a)  $P(\neg A) = 1 - P(A)$  (= **B1**).
- (b) If  $P(A \wedge B) = 1$ , then  $P(A) = P(B) = 1$ .
- (c) If  $P(A \wedge B) = 1$ , then  $P(A \wedge B) = P(A) \times P(B)$ .

**Proof.** (a) is Theorem 68. (b) Suppose  $P(A \wedge B) = 1$ . Then  $P(A) = 1$  by Theorems 58 and 70, and  $P(B) = 1$  by Theorems 64 and 70. (c) By (b). ■

**THEOREM 112.** *Let  $P$  be a two-valued function from the statements of  $L$  to the reals that meets Constraints **C1–C5** and **C7**.*

- (a) If  $P(A \wedge B) = 0$ , then  $P(A) = 0$  or  $P(B) = 0$ .

- (b) If  $P(A \wedge B) \neq 1$ , then  $P(A \wedge B) = P(A) \times P(B)$ .
- (c)  $P(A \wedge B) = P(A) \times P(B)$  (= **B2**)
- (d)  $P((\forall x)A) = 1$  iff  $P(A(T/x)) = 1$  for each term  $T$  of  $L$  (= **B3**).

**Proof.** (a) Suppose  $P(A \wedge B) = 0$ , in which case  $P(A) = P(A \wedge \neg B)$  by **C3**. Suppose further that  $P(A) \neq 0$ , and hence  $P(A) = 1$  by Theorem 108 (b). then  $P(A \wedge \neg B) = 1$ ; hence,  $P(\neg B) = 1$  by Theorems 64 and 70; and, hence,  $P(B) = 0$  by Theorem 68. Hence,  $P(A) = 0$  or  $P(B) = 0$ . hence, (a).

(b) By Theorem 108 (b) and clause (a) of the present theorem.

(c) By (b) and Theorem 111 (c).

(d) Suppose first that  $P((\forall x)A) = 1$ . Then by Theorem 90  $\text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j A(t_i/x)) = 1$ . But by Theorem 108 (b) each of  $P(\prod_{i=1}^1 A(t_i/x))$ ,  $P(\prod_{i=1}^2 A(t_i/x))$ , etc. equals 0 or 1, and by Theorem 58 each one of them is equal to or smaller than the preceding one. Hence,  $P(\prod_{i=1}^j A(t_i/x)) = 1$  for each  $j$  from 1 on and, hence, by Theorem 111 (b)  $P(A(T/x)) = 1$  for each term  $T$  of  $L$ . Suppose next that  $P((\forall x)A) \neq 1$ . Then by Theorem 108 (b)  $P((\forall x)A) = 0$ ; hence by Theorem 90  $\text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j A(t_i/x)) = 0$ ; hence, by Theorem 108 (b) there is a  $k$  such that, for every  $j$  from  $k$  on,  $P(\prod_{i=1}^j A(t_i/x)) = 0$ ; hence; by (a)  $P(A(T/x)) = 0$  for at least one term  $T$  of  $L$ ; and, hence,  $P(A(T/x)) \neq 1$  for at least one such term. Hence, (d). ■

Hence,

**THEOREM 113.** *Let  $P$  be a two-valued function from the statements of  $L$  to the reals. If  $P$  meets Constraints **C1–C7**, then  $P$  meets Constraints **B1–B3** as well.<sup>30</sup>*

Hence, the second step in my proof of Theorem 115:

**THEOREM 114.** *Let  $P$  be a two-valued function from the statements of  $L$  to the reals. Then  $P$  meets Constrains **B1–B3** iff  $P$  meets Constrains **C1–C7**.*

Hence by Theorem 109:

**THEOREM 115.**  *$P$  constitutes a truth-value function for  $L$  iff  $P$  constitutes a two-valued probability function for  $L$ .*

Hence, as claimed earlier, probability theory is but a generalisation of truth-value theory: allow truth-value functions—understood as meeting

<sup>30</sup>Since Constraint **C6** does not figure in Theorems 111 and 112, two-valued functions that meet Constraints **C1–C5** are sure by Theorem 110 to meet Constraint **C6** as well. I doubt, however, that this holds true of functions with more than two values.

Constraints **C1–C7**—to have anywhere from two to  $\aleph_0$  values, and truth-value theory expands into probability theory.<sup>31</sup> Or, to view it the other way round, truth-value theory is but a restriction of probability theory: require probability functions to have but two values, and probability theory reduces to truth-value theory. (To press the point further, probability functions are *measure functions* with reals from the interval  $[0, 1]$  as their only values, and truth-value functions are *probability functions* with the end-points in this interval as their only values.)

However, two-valued probability functions (i.e. truth-value functions) differ from the rest in one fundamental respect: the former are extensional, but the latter are not. Indeed, call a probability function  $P$  for  $L$  *extensional* if the value of  $P$  for a statement of  $L$  depends exclusively upon the value(s) of  $P$  for the immediate substatement(s) of that statement—more formally if

- (i) given that  $P(A) = P(A')$ ,  $P(\neg A) = P(\neg A')$ ,
- (ii) given that  $P(A) = P(A')$  and  $P(B) = P(B')$ ,  $P(A \wedge B) = P(A' \wedge B')$ ,
- (iii) given that  $P(A(T/x)) = P(A'(T/x))$  for each term  $T$  of  $L$ ,  $P((\forall x)A) = P((\forall x)A')$ .

Since  $P(\neg A) = 1 - P(A)$  by Theorem 68, (i) holds however many values  $P$  may have. When  $P$  is two-valued, (ii) and (iii) also hold: (ii) because  $P(A \wedge B)$  is then equal by Theorem 112(c) to  $P(A) \times P(B)$ , and (iii) because  $P((\forall x)A)$  is equal by Theorem 90 to  $\text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j A(t_i/x))$  and, hence, behaves as  $P(A(t_1/x))$ ,  $P(A(t_1/x) \wedge A(t_2/x))$ ,  $P((A(t_1/x) \wedge A(t_2/x)) \wedge A(t_3/x))$ , etc. do. However, when  $P$  has more than two values, (ii) and hence (iii) fail. Suppose, for example, that for some statement  $A$  of  $L$ ,  $P(A)$ —and, hence,  $P(\neg A)$ —equals  $\frac{1}{2}$ . Then by Theorem 59  $P(A \wedge A)$  equals  $P(A)$  and hence  $\frac{1}{2}$ , whereas by Theorem 65  $P(A \wedge \neg A)$  equals 0. So those among the probability functions for  $L$  that have just two values are extensional, but the rest are not.<sup>32</sup> The two results were to be expected.  $P$  is understood here as a measure of rational belief, and by all accounts rational belief is an intensional matter. When  $P$  has just two values, though,  $P$  is but a truth-value function, and truth-value functions *are* extensional.

Binary probability functions, also known as *conditional probability functions*, are often defined in terms of singular ones. For example, in [Kolmogorov, 1933; Carnap, 1950], etc.,  $P(A/B)$  is set at  $P(A \wedge B)/P(B)$  when

<sup>31</sup>A similar conclusion is reached in [Popper, 1959, p. 356]: ‘In its logical interpretation, the probability calculus is a genuine generalisation of the logic of derivation.’ It is based, though, on a different result, that reported in the Appendix.

<sup>32</sup>The point was brought to my attention by Bas C. van Fraassen. The counterexample in the text is a simplification by Michael E. Levin and myself of one originally suggested by van Fraassen.

$P(B) \neq 0$ , but otherwise is left without a value (recall that, when dealing with Kolmogorov, I think of his intersections as conjunctions). Partial functions, however, are unwieldy and—perforce—of limited service.

An alternative approach, favoured by Keynes as early as 1921 and, hence, antedating Kolmogorov's, has now gained wide currency, thanks to such diverse writers as Reichenbach, Jeffreys, Von Wright, Rényi, Carnap (in post-1950 publications), Popper, etc. Handling binary probability functions as you would singularly ones, you adopt constraints suiting your understanding of  $P(A/B)$  and then own as your binary probability functions all functions meeting these constraints.

Heeding the precedent of Section 4, I shall think of  $P^+(A_1/B)$ ,  $P^+(A_2/B)$ ,  $P^+(A_3/B)$ , etc., as *degrees to which a rational agent might—in the light of (or relative to) statement  $B$  of  $L^+$ —simultaneously believe the various statements  $A_1, A_2, A_3$ , etc., of  $L^+$* .<sup>33</sup> Constraints particularly suiting this understanding of  $P^+(A/B)$  can be found in [Von Wright, 1957]. Slightly edited for the occasion, they run:

1.  $0 \leq P^+(A/B)$
2.  $P^+(A/A) = 1$
3. If  $B$  is not logically false in the standard sense, then  $P^+(\neg A/B) = 1 - P^+(A/B)$
4.  $P^+(A \wedge B/C) = P^+(A/B \wedge C) \times P^+(B/C)$
5. If  $A$  and  $B$  are logically equivalent in the standard sense, then  $P^+(A/C) = P^+(B/C)$
6. If  $B$  and  $C$  are logically equivalent in the standard sense, then  $P^+(A/B) = P^+(A/C)$ .<sup>34</sup>

<sup>33</sup>I thus construe the present  $P^+$ 's as what note 27 called *coherent* belief functions. Since singularly probability functions are—in effect—binary ones relative to a logical truth (say,  $\neg(Q(t_1) \wedge \neg Q(t_1))$ ), just one notion of coherence is involved here.

<sup>34</sup>Von Wright weakens (2) to read:

(2\*) If  $A$  is not logically false, then  $P^+(A/A) = 1$ ,

and strengthens (3) to read:

(3\*)  $P^+(\neg A/B) = 1 - P^+(A/B)$ ,

a course few have adopted. That one or the other of (2\*) and (3\*) must carry a restriction is obvious. Suppose  $B$  is logically false, in which case both  $B \rightarrow A$  and  $B \rightarrow \neg A$  are logically true. (2) will then permit proof of  $P^+(A/B) = P^+(\neg A/B) = 1$ , and (3\*) must on pain of contradiction be weakened to read like (3). However, weaken (2) to read like Von Wright's (2\*). Then proofs of  $P^+(A/B) = 1$  and  $P^+(\neg A/B) = 1$  are blocked, and (3) may be strengthened to read like Von Wright's (3\*). Carnap, who accords  $P^+(A/B)$  a value only when  $B$  is not logically false, can make do with (2) and (3\*). But, as suggested in the text, partial probability functions are wanting.



Note: A statement is held *logically false* in the standard sense if its negation is logically true in that sense; and, as in note 29, two statements are held *logically equivalent* in the standard sense if their biconditional is logically true in that sense.

Substitutes for (3), (5) and (6) must of course be found, as these constraints are not autonomous. A number are available in the literature, among them:

$$(3') \text{ If } P^+(C/B) \neq 1 \text{ for at least one statement } C \text{ of } L^+, \text{ then } P^+(\neg A/B) = 1 - P^+(A/B),$$

$$(5') P^+(A \wedge B/C) = P^+(B \wedge A/C),$$

and

$$(6') P^+(A/B \wedge C) = P^+(A/C \wedge B).$$

The last two of these come from [Leblanc, 1981]. (Like **C5** they could sport ' $\leq$ ' rather than ' $=$ ', but the present formulations are more commonly employed.) (3') stems from [Popper, 1957]. It is weaker than (3) (in that certain functions meeting (3') will not meet (3)), but welcome so. To compare the two constraints, declare a statement  $B$  of  $L^+$   $P^+$ -*absurd* if  $P^+(C/B) = 1$  for every statement  $C$  of  $L^+$ , i.e. if in light of  $B$  a rational agent might believe any such  $C$ . Given either of (3) and (3') you can prove that any logical falsehood is  $P^+$ -*absurd*. And, *given* (3), you can go on and prove that, conversely, any statement  $P^+$ -*absurd* is logically false. *Not so, however given* (3'). Several writers, Carnap one of them, favour (3) as a result. Equally many, Popper one of them, do not—agreeing with Wellington that there are statements which are not logically false, to be sure, but in light of which one might believe anything.<sup>35</sup> I side with Popper on this and adopt Constraint (3').<sup>36</sup>

<sup>35</sup>Occasions when (Wellington) exhibited his other peculiarity, the crashing retort, were no doubt multiplied by the wit and inventiveness of his contemporaries; but the celebrated exchange between the Duke and some minor official from a government office is authentic.

"Mr. Jones, I believe," said the official blandly, accosting the great man in Pall Mall and mistaking him for the secretary of the Royal Academy. The world-famous profile froze.

"If you believe that, you'll believe anything." (Elizabeth Longford, *Wellington Pillar of State*, Harper & Rowe, New York, 1972.)

Because of this anecdote  $P^+$ -*absurd* statements are occasionally called *Jones statements*.

<sup>36</sup>To my knowledge, Popper is the first to have devised autonomous constraints for binary probability functions. The ones he finally settled on appeared in [Popper, 1957], and are extensively studied in [Popper, 1959, Appendices \*iv-\*v]. They run (in the symbolism of this essay):

A1. For any two statements  $A$  and  $B$  of  $L^+$  there are statements  $A'$  and  $B'$  of  $L^+$  such that  $P^+(A/B) \neq P^+(A'/B')$  (*Existence*)

Popper placed on his binary probability functions an extra constraint, requiring each of them to have at least two values. I adopt a simplified but equivalent version of it, requiring that—for each binary probability function  $P^+$  for  $L^+$ —at least one statement is not  $P^+$ -absurd. The constraint guarantees, among other things, that logical truths are not  $P^+$ -absurd, a matter of some comfort, and that singulary and binary probability functions bear to each other the relationship reported in my last theorem Theorem 141.

Formal details are as follows, with the more commonly used ' $P^+$ -abnormal' substituting for ' $P^+$ -absurd'.

Let  $L^+$  be an arbitrary term extension of  $L$ . By a *binary probability function* for  $L^+$  I shall understand any function  $P^+$  that takes each pair of statements for  $L^+$  into a real and meets the following constraints:

- D0.** There is a statement  $A$  and a statement  $B$  of  $L^+$  such that  $P^+(A/B) \neq 1$
- D1.**  $0 \leq P^+(A/B)$
- D2.**  $P^+(A/A) = 1$
- D3.** If there is a statement  $C$  of  $L^+$  such that  $P^+(C/B) \neq 1$ , then  $P^+(\neg A/B) = 1 - P^+(A/B)$
- D4.**  $P^+(A \wedge B/C) = P^+(A/B \wedge C) \times P^+(B/C)$
- D5.**  $P^+(A \wedge B/C) = P^+(B \wedge A/C)$
- D6.**  $P^+(A/B \wedge C) = P^+(A/C \wedge B)$
- D7.**  $P^+(\forall x)A/B = \text{Limit}_{j \rightarrow \infty} P^+(\prod_{i=1}^j A(t_i^+/x)/B)$ .

Where  $L^+$  is a term extension of  $L$  and  $P^+$  a binary probability function for  $L^+$ , I shall declare a statement  $B$  of  $L^+$   $P^+$ -normal ( $P^+$ -abnormal) if there is a (there is no) statement  $A$  of  $L^+$  such that  $P^+(A/B) \neq 1$ . I shall declare a statement  $A$  of  $L$  *logically true in the binary probabilistic sense* if—no

- A2. If  $P^+(A/C) = P^+(B/C)$  for every statement  $C$  of  $L^+$ , then  $P^+(C/A) = P^+(C/B)$  (*Substitutivity*)
- A3.  $P^+(A/A) = P^+(B/B)$  (*Reflexivity*)
- B1.  $P^+(A \wedge B/C) \leq P^+(A/C)$  (*Monotony*)
- B2.  $P^+(A \wedge B/C) = P^+(A/B \wedge C) \times P^+(B/C)$  (*Multiplication*)
- C. If  $P^+(C/B) \neq P^+(B/B)$  for at least one statement  $C$  of  $L^+$ , then  $P^+(\neg A/B) = P^+(B/B) - P^+(A/B)$  (*Complementation*).

That **D0–D6** pick out exactly the same binary probability functions as these six constraints do was undoubtedly known to Popper. The result is formally established in [Harper *et al.*, 1983]. (For a proof of A2 that uses only **D1**, **D2**, **D4** and Theorem 116 see [Leblanc, 1981].) In effect, Popper only dealt with quantifierless statements and hence had no analogue of **D7**.

matter the term extension  $L^+$  of  $L$ , binary probability function  $P^+$  for  $L^+$ , and statement  $B$  of  $L^+$ — $P^+(A/B) = 1$ ; and, where  $S$  is a set of statements of  $L$ , I shall declare  $A$  *logically entailed by  $S$  in the binary probabilistic sense* if—no matter the term extension  $L^+$  of  $L$ , binary probability function  $P^+$  for  $L^+$ , and statement  $B$  of  $L^+$ — $P^+(A/B) = 1$  if  $P^+(C/B) = 1$  for each member  $C$  of  $S$ . Equivalently, but more simply,  $A$  may be declared logically true in the binary probabilistic sense if  $P(A/B) = 1$  for every binary probability function  $P$  for  $L$  and every statement  $B$  of  $L$ . (In the idiom of rational belief  $A$  is thus logically true if in the light of *any* statement  $B$  of  $L$  a rational agent could not but believe  $A$ .)

I first argue for the foregoing definitions by showing that of logical entailment strongly sound and complete, and hence that of logical truth weakly sound and complete. I then study the relationship between singulary and binary probability functions.

Thanks to Harper, proof of my last Strong Soundness Theorem can be breathtakingly simple:

**THEOREM 116.**  $P(A/B) \leq 1$ .

**Proof.** When  $B$  is  $P$ -abnormal,  $P(A/B) = 1$  by definition and, hence,  $P(A/B) \leq 1$ . Suppose then that  $B$  is  $P$ -normal. By **D1**  $P(\neg A/B) \geq 0$  and, hence, by **D3**  $P(A/B) \leq 1$ . Hence, Theorem 116. ■

**THEOREM 117.**  $P(A/B \wedge A) = 1$ .

**Proof.** By **D2**  $P(B \wedge A/B \wedge A) = 1$ . Hence, by **D4**,  $P(B/A \wedge (B \wedge A)) \times P(A/B \wedge A) = 1$ . But by **D1** and Theorem 116 each of  $P(B/A \wedge (B \wedge A))$  and  $P(A/B \wedge A)$  lies in the interval  $[0, 1]$  and, hence, each of them here must equal 1. Hence, Theorem 117. ■

**THEOREM 118.**  $P(A/A \wedge B) = 1$ .

**Proof.** By Theorem 117 and **D6**. ■

**THEOREM 119.** *Let  $B$  be  $P$ -normal. Then:*

$$(a) P(\neg B/B) = 0.$$

$$(b) P(\neg B \wedge A/B) = 0.$$

**Proof.** (a) By **D2**, **D3**, and the hypothesis on  $B$ . (b) By **D5**  $P(\neg B \wedge A/B)$  equals  $P(A \wedge \neg B/B)$ , which by **D4** equals  $P(A/\neg B \wedge B) \times P(\neg B/B)$ . Hence (b) by (a). ■

**THEOREM 120.** *If  $P(A/B) = 0$  for no statement  $B$  of  $L$ , then  $P(A/B) = 1$  for every such  $B$  (= Harper's Lemma).*

**Proof.** Suppose  $P(A/B) \neq 1$  for some statement  $B$  of  $L$ . Then  $B$  is  $P$ -normal; hence,  $P(\neg B \wedge \neg A/B) = 0$  by Theorem 119 (b); and, hence,  $P(\neg B/\neg A \wedge B) \times P(\neg A/B) = 0$  by **D4**. But, since  $P(A/B) \neq 1$  and  $B$  is  $P$ -normal,  $P(\neg A/B) \neq 0$ . Hence,  $P(\neg B/\neg A \wedge B) = 0$ ; and, hence,  $\neg A \wedge B$  is  $P$ -normal. But  $P(\neg A/\neg A \wedge B) = 1$  by Theorem 118. Hence,  $P(A/\neg A \wedge B) = 0$  by **D3**; and, hence, there is a statement  $B$  of  $L$  such that  $P(A/B) = 0$ . Hence, Theorem 120. ■

**THEOREM 121.**  $P(A \wedge B/C) \leq P(B/C)$ .

**Proof.** By **D4**  $P(A \wedge B/C) = P(A/B \wedge C) \times P(B/C)$ . But by **D1** and Theorem 116, each of  $P(A/B \wedge C)$  and  $P(B/C)$  lies in the interval  $[0, 1]$  and, hence, neither can be less than  $P(A \wedge B/C)$ . Hence, Theorem 121. ■

**THEOREM 122.**  $P(A \wedge B/C) \leq P(A/C)$ .

**Proof.** By Theorem 121 and **D5**. ■

**THEOREM 123.** *If  $P(A \wedge B/C) = 1$  then  $P(A/C) = P(B/C) = 1$ .*

**Proof.** Suppose  $P(A \wedge B/C) = 1$ . Then  $P(A/C) = 1$  by Theorems 122 and 116, and  $P(B/C) = 1$  by Theorems 121 and 116. Hence, Theorem 123. ■

**THEOREM 124.** *If  $P(A \rightarrow B/C) = 0$ , then  $P(A/C) = 1$  and  $P(B/C) = 0$ .*

**Proof.** Suppose  $P(A \rightarrow B/C) = 0$ , in which case  $C$  is  $P$ -normal. Then  $P(A \wedge \neg B/C) = 1$  by  $D_{\rightarrow}$  and **D3**; hence,  $P(A/C) = P(\neg B/C) = 1$  by Theorem 123 and, hence,  $P(A/C) = 1$  and  $P(B/C) = 0$  by **D3**. ■

**THEOREM 125.**  $P(A \wedge A/B) = P(A/B)$ .

**Proof.** By **D4**  $P(A \wedge A/B) = P(A/A \wedge B) \times P(A/B)$ . Hence, Theorem 125 by Theorem 118. ■

**THEOREM 126.** *If  $C$  is  $P$ -normal, then  $P(A/C) = P(A \wedge B/C) + P(A \wedge \neg B/C)$ .*

**Proof.** Suppose  $C$  is  $P$ -normal. By definition when  $A \wedge C$  is  $P$ -abnormal but otherwise by **D3**,

$$P(B/A \wedge C) + P(\neg B/A \wedge C) = P(C/A \wedge C) + P(\neg C/A \wedge C),$$

hence

$$\begin{aligned} & P(B/A \wedge C) \times P(A/C) + P(\neg B/A \wedge C) \times P(A/C) \\ &= P(C/A \wedge C) \times P(A/C) + P(\neg C/A \wedge C) \times P(A/C), \end{aligned}$$

hence by **D4**

$$\begin{aligned} & P(B \wedge A/C) + P(\neg B \wedge A/C) \\ &= P(C/A \wedge C) \times P(A/C) + P(\neg C \wedge A/C), \end{aligned}$$

and, hence, by Theorems 117 and 119 (b)

$$P(B \wedge A/C) + P(\neg B \wedge A/C) = P(A/C).$$

Hence, Theorem 126 by **D5**. ■

**THEOREM 127.** *If  $P(A \rightarrow B/C) = 1$  and  $P(A/C) = 1$ , then  $P(B/C) = 1$  (= Modus Ponens).*

**Proof.** When  $C$  is  $P$ -abnormal,  $P(B/C) = 1$  by definition. So suppose that  $C$  is  $P$ -normal and  $P(A \rightarrow B/C) = 1$ . Then by  $D_{\rightarrow}$  and **D3**,  $P(A \wedge \neg B/C) = 0$ ; and, hence, by Theorem 126,  $P(A \wedge B/C) = P(A/C)$ . Suppose further that  $P(A/C) = 1$ . Then  $P(A \wedge B/C) = 1$ ; and, hence,  $P(B/C) = 1$  by Theorems 121 and 116. Hence, Theorem 127. ■

**THEOREM 128.** *If  $P(A/C) = P(B/C) = 1$ , then  $P(A \wedge B/C) = 1$ .*

**Proof.** When  $C$  is  $P$ -abnormal,  $P(A \wedge B/C) = 1$  by definition. So suppose that  $C$  is  $P$ -normal and  $P(A/C) = 1$ . Then by Theorem 126  $P(A \wedge B/C) + P(A \wedge \neg B/C) = 1$ . Suppose further that  $P(B/C) = 1$ . Then by **D3**  $P(\neg B/C) = 0$ ; hence, by Theorem 121 and **D1**  $P(A \wedge \neg B/C) = 0$ ; and, hence, by Theorem 126  $P(A \wedge B/C) = 1$ . Hence, Theorem 128. ■

That, no matter the statement  $B$  of  $L$ ,  $P(A/B) = 1$  for any axiom  $A$  of  $L$  of sorts **A1–A6** follows by six applications of Harper's Lemma:

**THEOREM 129.** *If  $A$  is an axiom of  $L$  of sorts **A1–A6**, then  $P(A/B) = 1$  for every statement  $B$  of  $L$ .*

**Proof.** Let  $A$  be an axiom of  $L$  of sorts **A1–A6**. I show that if there were a statement  $B$  of  $L$  such that  $P(A/B) = 0$ , a contradiction would ensue. So  $P(A/B) = 0$  for no statement  $B$  of  $L$  and hence, by Theorem 120  $P(A/B) = 1$  for every such  $B$ .

*Case 1:*  $A$  is of the sort  $A' \rightarrow (A' \wedge A')$ . Suppose  $P(A/B) = 0$  for some statement  $B$  of  $L$ . Then by Theorem 124

1.  $P(A'/B) = 1$

and

2.  $P(A' \wedge A'/B) = 0$ .

But by (2) and Theorem 125

$$3. P(A'/B) = 0.$$

So a contradiction ensues. So  $P(A/B) = 1$  for every  $B$ .

*Case 2:*  $A$  is of the sort  $(A' \wedge B') \rightarrow A'$ . Suppose  $P(A/B) = 0$  for some statement  $B$  of  $L$ . Then by Theorem 124

$$1. P(A' \wedge B'/B) = 1$$

and

$$2. P(A'/B) = 0.$$

But by (1) and Theorem 123

$$3. P(A'/B) = 1.$$

So a contradiction ensues. So  $P(A/B) = 1$  for every  $B$ .

*Case 3:*  $A$  is of the sort  $(A' \rightarrow B') \rightarrow (\neg(B' \wedge C') \rightarrow \neg(C' \wedge A'))$ . Suppose  $P(A/B) = 0$  for some statement  $B$  of  $L$ . Then by Theorem 124

$$1. P(A' \rightarrow B'/B) = 1$$

and

$$2. P(\neg B' \wedge C' \rightarrow \neg(C' \wedge A')/B) = 0.$$

But by (2) and Theorem 124

$$3. P(\neg(B' \wedge C')/B) = 1.$$

and

$$4. P(\neg(C' \wedge A)/B) = 0.$$

Hence by **D3** (and the fact that, since  $P(A/B) = 0$ ,  $B$  is  $P$ -normal).

$$5. P(C' \wedge A'/B) = 1,$$

and, hence, by Theorem 123

$$6. P(A'/B) = 1.$$

But by (1),  $D_{\rightarrow}$ , and **D3**

$$7. P(A' \wedge \neg B'/B) = 0.$$

Hence, by (6) and Theorem 126

$$8. P(A' \wedge B'/B) = 1,$$

and, hence, by Theorem 123

$$9. P(B'/B) = 1.$$

But by (3) and **D3**

$$10. P(B' \wedge C' / B) = 0,$$

and, hence, by **D5**

$$11. P(C' \wedge B' / B) = 0,$$

whereas by (5) and Theorem 123

$$12. P(C' / B) = 1.$$

Hence, by (11) and Theorem 126

$$13. P(C' \wedge \neg B' / B) = 1,$$

hence, by Theorem 123

$$14. P(\neg B' / B) = 1,$$

and, hence, by **D3**

$$15. P(B' / B) = 0.$$

So a contradiction (= (9) and (15)) ensues. So  $P(A/B) = 1$  for every  $B$ .

*Case 4:*  $A$  is of the sort  $A' \rightarrow (\forall x)A'$ . Suppose  $P(A/B) = 0$  for some statement  $B$  of L. Then by Theorem 124

$$1. P(A'/B) = 1$$

and

$$2. P((\forall x)A' / B) = 0.$$

But by (2) and **D7**

$$3. \text{Limit}_{j \rightarrow \infty} P(\underbrace{(\dots (A' \wedge A') \wedge \dots)}_{j \text{ times}} \wedge A' / B) = 0$$

and, hence, by Theorem 125 and the definition of a limit

$$4. (A'/B) = 0.$$

So a contradiction ensues. So  $P(A/B) = 1$  for every  $B$ .

*Case 5:*  $A$  is of the sort  $(\forall x)A' \rightarrow A'(T/x)$ . Suppose  $P(A/B) = 0$  for some statement  $B$  of L. Then by Theorem 124

$$1. P((\forall x)A' / B) = 1$$

and

$$2. P(A'(T/x) / B) = 0 \text{ for some term } T \text{ of L.}$$

But by (1) and **D7**

$$3. \text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j A(t_i/x)/B) = 1,$$

hence, by Theorem 122 and familiar considerations

$$4. P(\prod_{i=1}^j A(t_i/x)/B) = 1 \text{ for } j \text{ from } 1 \text{ on,}$$

and hence, by Theorem 123

$$5. P(A(T/x)/B) = 1 \text{ for each term } T \text{ of } \mathbb{L}.$$

So a contradiction ensues. So  $P(A/B) = 1$  for every  $B$ .

*Case 6:*  $A$  is of the sort  $(\forall x)(A' \rightarrow B') \rightarrow ((\forall x)A' \rightarrow (\forall x)B')$ . Suppose  $P(A/B) = 0$  for some statement  $B$  of  $\mathbb{L}$ . Then by Theorem 124

$$1. P((\forall x)(A' \rightarrow B')/B) = 1$$

and

$$3. P((\forall x)A' \rightarrow (\forall x)B'/B) = 0.$$

But by (2) and Theorem 124

$$3. P((\forall x)A'/B) = 1$$

and

$$4. P((\forall x)B'/B) = 0.$$

Hence by (1), (3), and **D7**

$$5. \text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j (A' \rightarrow B')(t_i/x)/B) = 1$$

and

$$6. \text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j A'(t_i/x)/B) = 1,$$

and, hence, by Theorem 122 and familiar considerations

$$7. P((A' \rightarrow B')(T/x)/B) = 1 \text{ for each term } T \text{ of } \mathbb{L}$$

and

$$8. P(A'(T/x)/B) = 1 \text{ for each term } T \text{ of } \mathbb{L}.$$

Hence, by Theorem 127

$$9. P(B'(T/x)/B) = 1 \text{ for each term } T \text{ of } \mathbb{L},$$

hence, by Theorem 128 and the definition of a limit

$$10. \text{Limit}_{j \rightarrow \infty} P(\prod_{i=1}^j B'(t_i/x)/B) = 1,$$



and, hence, by **D7**

$$11. P((\forall x)B'/B) = 1.$$

So a contradiction (= (4) and (11)) ensues. So  $P(A/B) = 1$  for every  $B$ . ■

So by an induction like the one on page 99 (with Theorem 128 substituting for Theorem 91):

**THEOREM 130.** *If  $A$  is an axiom of  $L$ , then  $P(A/B) = 1$  for every statement  $B$  of  $L$ .*

So by an induction like the one on page 100 (with Theorem 130 substituting for Theorem 98 and Theorem 127 for Theorem 87):

**THEOREM 131.** *If  $S \vdash A$ , then—no matter the statement  $B$  of  $L$ — $P(A/B) = 1$  if  $P(C/B) = 1$  for every member  $C$  of  $S$ .*

But Theorems 127 and 130 hold with ‘ $P^+$ ’ for ‘ $P$ ’. Hence, the following counterpart of Theorem 100 for binary probability functions:

**THEOREM 132.** *If  $S \vdash A$ , then—no matter the term extension  $L^+$  of  $L$ , binary probability function  $P^+$  for  $L^+$ , and statement  $B$  of  $L^+$ — $P^+(A/B) = 1$  if  $P^+(C/B) = 1$  for every member  $C$  of  $S$ , i.e.  $A$  is logically implied by  $S$  in the binary probabilistic sense.*

Slight editing of the material on pages 100–102 yields proof of the counterpart of Theorem 103 for binary probability functions.

Where  $S$  is an arbitrary set of statements of  $L$ , understand by the *binary probability associate* of  $S$  in  $L$  the function  $P_S$  such that, for any statement  $A$  and any statement  $B$  of  $L$ :

$$P_S(A/B) = \begin{cases} 1 & \text{if } S \vdash B \rightarrow A \\ 0 & \text{otherwise.} \end{cases}$$

As the reader may wish to verify.

**THEOREM 133.** *Let  $S$  be a set of statements of  $L$ , and  $P_S$  be the binary probability associate of  $S$  in  $L$ .*

- (a)  $P_S$  meets Constraints **D1–D2** and **D4–D6**; and, when  $S$  is consistent in  $L$ ,  $P_S$  meets constraint **D0** as well. .
- (b) If  $S$  is maximally consistent in  $L$ , then  $P_S$  meets Constraint **D3** as well.
- (c) If  $S$  is  $\omega$ -complete in  $L$ , then  $P_S$  meets Constraint **D7** as well.
- (d) If  $S$  is maximally consistent and  $\omega$ -complete in  $L$ , then  $P_S$  constitutes a binary probability function for  $L$ .

**THEOREM 134.** *Let  $S$  be a set of statements of  $L$  that is maximally consistent in  $L$ , and let  $P_S$  be the binary probability associate of  $S$  in  $L$ .*

- (a) *A conditional  $B \rightarrow A$  of  $L$  belongs to  $S$  iff  $P_S(A/B) = 1$ .*
- (b)  *$P_S(A/\neg(B \wedge \neg B))$  equals 1 if  $A$  belongs to  $S$ , but 0 if  $\neg A$  does.*

Note for proof of (b) that, with  $S$  maximally consistent in  $L$ , (i)  $\neg(B \wedge \neg B) \rightarrow A$  belongs to  $S$  if  $A$  does, but (ii)  $\neg(B \wedge \neg B) \rightarrow A$  does not if  $\neg A$  does (and, hence,  $A$  does not). So, (b) by (a).

Now suppose, for the last time, that  $S \not\vdash A$ , where  $S$  is infinitely extendible in  $L$ ; let  $\mathcal{H}(S \cup \{\neg A\})$  be the Henkin extension of  $S \cup \{\neg A\}$  in  $L$ ; let  $P_{\mathcal{H}(S \cup \{\neg A\})}$  be the binary probability associate of  $\mathcal{H}(S \cup \{\neg A\})$  in  $L$ ; and let  $B'$  be an arbitrary statement of  $L$ . Since by Theorem 4  $\mathcal{H}(S \cup \{\neg A\})$  is maximally consistent in  $L$ , we know by Theorem 134 (b) that relative to  $\neg(B' \wedge \neg B')$  each member  $C$  of  $S$  evaluates to 1 on  $P_{\mathcal{H}(S \cup \{\neg A\})}$ , whereas  $A$  evaluates to 0. But, since by Theorem 4  $\mathcal{H}(S \cup \{\neg A\})$  is  $\omega$ -complete as well as maximally consistent in  $L$ , we know by Theorem 133(b) that  $P_{\mathcal{H}(S \cup \{\neg A\})}$  constitutes a binary probability function for  $L$ . Hence:

**THEOREM 135.** *Let  $S$  be a set of statements of  $L$  that is infinitely extendible in  $L$ , and  $A$  be an arbitrary statement of  $L$ . If  $S \not\vdash A$ , then there is a binary probability function  $P$  for  $L$  and a statement  $B$  of  $L$  such that  $P(C/B) = 1$  for each member  $C$  of  $S$  but  $P(A/B) \neq 1$ .*

But all of Theorems 4, 133, and 134 hold with ' $P^+$ ' in place of ' $P$ '. Hence:

**THEOREM 136.** *Let  $S$  and  $A$  be as in Theorem 135. If—no matter the binary probability function  $P$  of  $L$  and statement  $B$  of  $L$ — $P(A/B) = 1$  if  $P(C/B) = 1$  for each member  $C$  of  $S$ , then  $S \vdash A$ .*

When  $S$  is not infinitely extendible in  $L$ , resort to  $L^\infty$  will do the trick as usual. Hence, given Theorem 136, the following counterpart of Theorem 105 for binary probability functions:

**THEOREM 137.** *Let  $S$  be an arbitrary set of statements of  $L$ , and  $A$  be an arbitrary statement of  $L$ . If—no matter the term extension  $L^+$  of  $L$ , binary probability function  $P^+$  for  $L^+$ , and statement  $B$  of  $L^+$ — $P^+(A/B) = 1$  if  $P^+(C/B) = 1$  for each member  $C$  of  $S$ , i.e. if  $A$  is logically entailed by  $S$  in the binary probabilistic sense, then  $S \vdash A$ .*

Hence, appealing to Theorem 132 and taking  $S$  in each case to be  $\emptyset$ :

**THEOREM 138.** *Let  $A$  be an arbitrary statement of  $L$ . Then  $\vdash A$  iff—no matter the term extension  $L^+$  of  $L$ , binary probability function  $P^+$  for  $L^+$ , and statement  $B$  of  $L^+$ — $P^+(A/B) = 1$ , i.e. iff  $A$  is logically true in the binary probabilistic sense.*

And, with  $S$  again set at  $\emptyset$ , Theorems 131 and 136 yield:

**THEOREM 139.** *Let  $A$  be as in Theorem 138. Then  $\vdash A$  iff—no matter the probability function  $P$  for  $L$  and statement  $B$  of  $L$ — $P(A/B) = 1$ .*

The binary account of logical entailment on page 109 is thus strongly sound and complete, and the two binary accounts of logical truth on that page weakly sound and complete.

No appeal has been made so far to **D0**: the soundness and completeness theorems I have just proved thus hold with  $P^+$  presumed to meet only Constraints **D1–D7**. However, **D0** will now do some work, namely: ensure that  $\neg(Q(t_1) \wedge \neg Q(t_1))$  ( $Q$  a predicate of  $L$  of degree 1 and  $t_1$  the alphabetically first term of  $L$ ) is  $P$ -normal, and thereby allow for an important connection between singular and binary probability functions. To abridge matters I refer to  $\neg(Q(t_1) \wedge Q(t_1))$  by means of ' $\top$ '.

**D0**, as stressed earlier, demands that at least one statement of  $L$  be  $P$ -normal. After proving one ancillary result, I establish by means of just **D1–D7** (indeed, just **D1–D6**) that  $\top$  is  $P$ -normal if any statement of  $L$  is, and hence by dint of **D0** that  $\top$  is  $P$ -normal.

**THEOREM 140.** *Let  $P$  be an arbitrary binary probability function for  $L$ .*

(a)  $P(A/B \wedge \top) = P(A/B)$ .

(b) *If any statement of  $L$  is  $P$ -normal, then  $\top$  is.*

(c)  $\top$  is  $P$ -normal.

**Proof.** (a) By **D4**  $P(\top \wedge A/B) = P(\top/A \wedge B) \times P(A/B)$ . But by Theorem 131  $P(\top/A \wedge B) = 1$ . Hence,  $P(\top \wedge A/B) = P(A/B)$ . But by the same reasoning  $P(A \wedge \top/B) = P(A/\top \wedge B)$ . Hence, by **D5**,  $P(A/\top \wedge B) = P(A/B)$  and, hence, by **D6**,  $P(A/B \wedge \top) = P(A/B)$ .

(b) Suppose  $P(A/\top) = 1$  for every statement  $A$  of  $L$ , and let  $B$  be an arbitrary statement of  $L$ . Then  $P(A \wedge B/\top) = 1$ ; hence, by **D4**  $P(A/B \wedge \top) \times P(B/\top) = 1$ ; hence, by **D1** and theorem 116  $P(A/B \wedge \top) = 1$ ; and, hence, by (a)  $P(A/B) = 1$ . Hence,  $P(A/B) = 1$  for every statement  $A$  and every statement  $B$  of  $L$ . Hence, by Contraposition, if  $P(A/B) \neq 1$  for any statement  $A$  and any statement  $B$  of  $L$ , then  $P(A/\top) \neq 1$  for some statement  $A$  of  $L$ . Hence, (b).

(c) By **D0** and (b). ■

The argument is easily edited to show that *any* logical truth of  $L$ , not just  $\top$ , is  $P$ -normal. (It follows, incidentally, from Theorem 140 (c) and **D2** that  $P(\neg\top/\top) = 0$  and, hence, that each binary probability function  $P$  for  $L$  has at least two values. So **D0** does the same work as Popper's constraint A1.)

The reader will note that by virtue of Theorems 140 (c) and 126

$$P(A/\top) = P(A \wedge B/\top) + P(A \wedge \neg B/\top),$$

a  $\top$ -version—so to speak—of **C3**. But similar  $\top$ -versions of **C1–C2** and **C4–C7** can be had as well. Indeed, the  $\top$ -versions of **C1**, **C5**, and **C7**:

$$0 \leq P(A/\top), \quad P(A \wedge B/\top) \leq P(B \wedge A/\top),$$

and

$$P((\forall x)A/\top) = \text{Limit}_{j \rightarrow \infty} P\left(\prod_{i=1}^j A(t_i/x)/\top\right),$$

are special cases of **D1**, **D5** and **D7**. The  $\top$ -versions of **C2** and **C4**:

$$P(\neg(A \wedge \neg A)/\top) = 1$$

and

$$P(A/\top) \leq P(A \wedge A/\top),$$

hold by Theorem 131 and Theorem 125. And the  $\top$ -version of **C6**:

$$P(A \wedge (B \wedge C)/\top) \leq P((A \wedge B) \wedge C/\top),$$

can be had by repeated applications of **D4** and Theorem 140 (a). What is known as the *restriction of  $P$  to  $\top$* , i.e. the function  $P_\top$  such that—for each statement  $A$  of  $L$ — $P_\top(A) = P(A/\top)$ , thus constitutes a singular probability function for  $L$ .

On the other hand, let  $P$  be a singular probability function for  $L$ ; let  $P'$  be the function such that, for any statement  $A$  and any statement  $B$  of  $L$ ,

$$P'(A/B) = \begin{cases} P(A \wedge B)/P(B) & \text{if } P(B) = 0 \\ 1 & \text{otherwise;} \end{cases}$$

and let  $P'_\top$  be the restriction of  $P'$  to  $\top$ . It is easily verified that (i)  $P'_\top$  constitutes a binary probability function for  $L$  and (ii)  $P'_\top(A) = P(A)$  for any statement  $A$  of  $L$ . For proof of (ii) note that by definition  $P'_\top(A)$  equals  $P'_\top(A/\top)$ . But by **C2**  $P(\top) = 0$ . Hence by definition  $P'_\top(A)$  equals  $P(A \wedge \top)/P(\top)$ , which by Theorem 67 and **C2** equals  $P(A)$ . So there is a binary probability function for  $L$ , to wit:  $P'_\top$ , of which  $P$  is the restriction to  $\top$ .

So:

**THEOREM 141.** (a) *Each binary probability function for  $L$  has as its restriction to  $\top$  a singular probability function for  $L$ .*

(b) *Each singular probability function for  $L$  is the restriction to  $\top$  of a binary probability function for  $L$ .*

Like results hold of course with ' $L^+$ ' for ' $L$ ', thus binding together all the probability functions treated in Sections 4 and 5.<sup>37</sup>

<sup>37</sup>Binary probability functions, like their singularity brethren, bear a close relationship to truth-value functions. See [Gumb, 1983] for preliminary results on the matter.

## 6 IN SUMMARY AND CONCLUSION

The non-standard accounts of logical entailment (and, hence, of logical truth) in this essay were justified thusly—with truth sets and model ones treated first to vary the perspective slightly:

Supposing  $S \not\vdash A$ , I formed the Henkin extension  $\mathcal{H}^\infty(S \cup \{\neg A\})$  of  $S \cup \{\neg A\}$  in  $L^\infty$  (in  $L$  itself when  $S$  is infinitely extendible). That set proved to be a truth set for  $L^\infty$  (for  $L$ ), which ensured that if  $S \vdash A$ , then there is—for some term extension  $L^+$  of  $L$ —a truth set for  $L^+$  of which  $S$  is a subset but  $A$  is not a member. Hence, by dint of a suitable soundness theorem,  $S \vdash A$  iff—no matter the term extension  $L^+$  of  $L$ — $A$  belongs to every truth set for  $L^+$  of which  $S$  is a subset (Theorems 34 and 37). But any truth set for  $L^+$  is a model set for  $L^+$ , which ensured that if  $S \not\vdash A$ , then there is—for some term extension  $L^+$  of  $L$ —a model set for  $L^+$  of which  $S$  is a subset and  $\neg A$  is a member. Hence, by dint again of a suitable soundness theorem,  $S \vdash A$  iff—no matter the term extension  $L^+$  of  $L$ — $\neg A$  does not belong to any model set for  $L^+$  of which  $S$  is a subset (Theorems 47 and 50). Two alternatives to the standard account of logical entailment, the truth-set account of page 77 and the model-set account on pages 79 were thereby legitimised. So, consequently, were two alternatives to the standard account of logical truth.

But any truth set for  $L^\infty$  generates a Henkin model for  $L^\infty$ , the model associate of that set defined on page 64. The fact ensured that if  $S \not\vdash A$ , then there is—for some extension  $L^+$  of  $L$ —a Henkin model for  $L^+$  in which  $S$  is  $\text{true}_S$  but  $A$  does not. Hence, by dint of a suitable soundness theorem,  $S \vdash A$  iff—no matter the term extension  $L^+$  of  $L$ — $A$  is  $\text{true}_S$  in every Henkin model for  $L^+$  in which  $S$  is  $\text{true}_S$  (Theorem 25). But any truth set for  $L^\infty$  also generates a truth-value assignment (equivalently, a truth-value function) for  $L^\infty$ , the truth-value associate of that set defined on page 73. The fact ensured that if  $S \not\vdash A$ , then there is—for some extension  $L^+$  of  $L$ —a truth-value assignment (a truth-value function) for  $L^+$  on which each member of  $S$  is true (evaluates to  $\top$ ) but  $A$  is not (does not). Hence, by dint again of a suitable soundness theorem,  $S \vdash A$  iff—no matter the term extension  $L^+$  of  $L$ — $A$  is true on every truth-value assignment ( $A$  evaluates to  $\top$  on every truth-value function) for  $L^+$  on which all members of  $S$  are (do) (Theorems 29 and 30).<sup>38</sup> Two extra alternatives to the standard account of logical entailment, the substitutional account of page 69 and the truth-value one of page 73, were thereby legitimised. So, consequently, were two extra alternatives to the standard account of logical truth.

But any truth set for  $L^\infty$  also generates a singulary probability function for  $L^\infty$ , the probability associate of that set defined on page 100. The fact ensured that if  $S \vdash A$ , then there is—for some term extension  $L^+$  of  $L$ —a

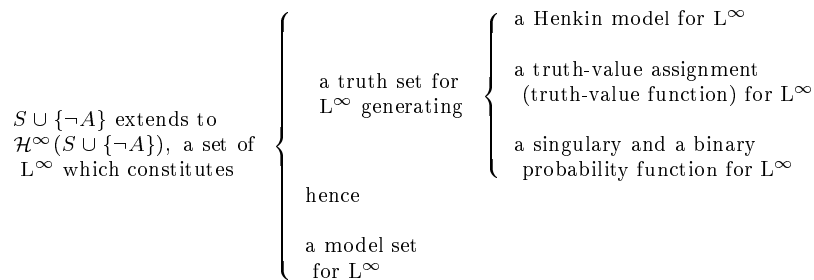
<sup>38</sup>The proof of Theorem 30 I just rehearse is that on page 74 (see note 18).

singular probability function for  $L^+$  on which all members of  $S$  evaluate to 1 but  $A$  does not. Hence, by dint of a suitable soundness theorem,  $S \vdash A$  iff—no matter the term extension  $L^+$  of  $L$ — $A$  evaluates to 1 on every singular probability function for  $L^+$  on which all members  $s$  of  $S$  do (Theorems 100 and 105). But any truth set for  $L^\infty$  also generates a binary probability function for  $L^\infty$ , the probability associate of that set defined on page 118. The fact ensured that if  $S \vdash A$ , then there is—for some term extension  $L^+$  of  $L$  and some statement  $B$  of  $L^+$ —a binary probability function for  $L^+$  on which all members of  $S$  evaluate to 1 relative to  $B$  but  $A$  does not. Hence, by dint of a suitable soundness theorem,  $S \vdash A$  iff—no matter the term extension  $L^+$  of  $L$  and statement  $B$  of  $L^+$ — $A$  evaluates to 1 relative to  $B$  on every binary probability function for  $L^+$  on which all members of  $S$  do (Theorems 132 and 137). Two extra alternatives to the standard account of logical entailment, the probability account of page 87 and that of page 109, were thereby legitimised. So, consequently, were two extra alternatives to the standard account of logical truth.<sup>39</sup>

The definitions of pages 69, 73, 77–78, 79, 87 and 109, thus known to capture exactly the same logical truths and exactly the same logical entailments as the standard ones do, merit attention on further counts. I largely limit myself in what follows to logical truths. Like remarks apply *mutatis mutandis* to logical entailments.

Logical truths are defined quite grandly in standard semantics as statements true in *all* models or—to spell it out once more—true in  $\langle D, I_D \rangle$ , no matter the domain  $D$  and  $D$ -interpretation  $I_D$  of the terms and predicates in those statements. Unfortunately, domains are sets and since 1895, the year Cantor discovered the Burali–Forti paradox, sets have been very much a concern, one indeed that may never be alleviated.<sup>40</sup> So the standard account of logical truth, with its mention—be it overt or covert—of *all* domains and hence of *all* sets, may be injudicious: it rests logic on very

<sup>39</sup>Schematically:



<sup>40</sup>The Zermelo–Fraenkel axiomatisation of set theory is mathematically serviceable, to be sure, but ontologically and epistemologically how much of a case has ever been made for it?

shaky foundations.<sup>41</sup>

That the models figuring in the above account can be presumed to be countable was therefore major news (though, as noted earlier, not welcome news to all). Since finite cardinals and  $\aleph_0$  are unproblematic (to put it more judiciously, are less problematic than larger cardinals), logic studies could be pursued with little apprehension. As conceded on page 56, one might still deal in uncountable models when attending to certain theories (though countable submodels of those models would always do the trick, we saw). But attending to theories whose intended models are uncountable is one thing, explicating logical truth (and logical entailment) is quite another; and discharging the latter assignment without courting disaster or at the very least embarrassment seems the wiser course.

But, as later developments showed, one can do far better than the Löwenheim Theorem and Skolem's generalisation thereof guaranteed. Since countable models will do and first-order languages have  $\aleph_0$  terms, Henkin models should do as well, with all items in their domains provided with names. What, given just  $\aleph_0$  names, uncountable models ruled out, should now be feasible. And indeed statements true in all countable models (hence, in all models) proved to be, as Henkins' completeness proof intimated, those very statements true *in all Henkin models*. The result buoyed those who like things to bear names and quantifiers to be construed substitutionally, a natural inclination as Wittgenstein admitted.

True, logical entailment does pose a problem. As sets are not all infinitely extendible, one may occasionally run out of terms, even though one has  $\aleph_0$  of them and only  $\aleph_0$  items to name. But, following Henkins' precedent, one can always send out for extra terms by extending one's original language to such a language as  $L^\infty$ . The Henkin models of  $L^\infty$  will serve all of one's needs, yielding a definition of logical entailment to the same effect as the standard one. The resulting semantics is, of course, that on page 69: *substitutional semantics*.

However, even such minimal use of sets as substitutional semantics makes is unnecessary, as Beth, Schütte, and others showed. With the quantifiers substitutionally construed, sets (*qua* domains) turn up at only one point: the truth condition for *atomic statements* on page 69. That condition, incidentally, is not to everyone's taste: some (Frege, Church, and possibly Whitehead and Russell, among them) prefer one that uses propositional functions and, hence, has an intensional rather than extensional flavour. But, whatever one's preference in these matters, is it really logic's charge to spell out the circumstances under which atomic statements are true? The recounting plays no role in substitutional semantics. So, instead of engaging in such controversial and (in this context) inconsequential business,

---

<sup>41</sup>When teaching standard semantics, how many warn students not to ask whether the set of all domains is itself a domain?

why not assume with Beth, Schütte, and others that atomic statements have truth-values, however they come by them, and proceed with matters of truly logical import?<sup>42</sup> Thus was *truth-value semantics* born, a semantics that dispenses with domains and, hence, with reference (crucial though that notion may be elsewhere). And, dispensing with reference, truth-value semantics can focus on a single notion: truth. In one version of it, truth-value assignments (to *atomic* statements) and a recounting of when *compound* statements are true on them share the work; in another and even sparer version truth-functions do it all. And if future studies prove as rewarding as past ones, most (if not all) of standard semantics might soon admit of translation into truth-value idiom.

As it turns out, though, even truth can be dispensed with, the truth sets introduced in [Quine, 1940] permitting definition (we saw) of logical truth and logical entailment, as do the model sets introduced in [Hintikka, 1955]. Some have objected that for  $A$  to belong to *every* truth set (for  $\neg A$  to belong to *no* model set) is a purely syntactical feature of  $A$ , and consequently that truth-set semantics (model set semantics) is but mislabelled syntax.

I shall not meet the objection *here*, except for recalling that truth sets, for example, automatically convert into truth-value functions, the pre-requisites for belonging to a truth set being those for evaluating to  $\top$  on a truth-value function (**B1–B3** on page 103). Rather, I will draw attention to probabilistic semantics, where the notion of truth yields to that of rational (more specifically, coherent) belief, and accounts of logical truth and logical entailment matching the standard ones are also available. Probabilistic semantics is as spare as truth-value semantics is—probability functions (be they singular ones or binary ones) doing *all* the work. And it offers a whole new perspective on logical truth (and logical entailment): traditionally viewed as statements true in all models (more judiciously, in all countable models) and more recently as statements evaluating to  $\top$  on all truth-value functions, logical truths can now be seen as statements than which none are more (coherently) believable.<sup>43</sup>

Probabilistic semantics, the new intensional semantics studied here, has yet to match standard semantics in breadth. But it already boasts significant results, some presented in Sections 4–5; and, feeding as it does on probability theory, belief theory, game theory, etc. it should soon boast others.

In any event there are accounts of logical truth and logical entailment other than the standard one and no less legitimate. *This much* the essay *has* shown.

---

<sup>42</sup>I borrow at this point from a conversation with Alex Orenstein.

<sup>43</sup>Note indeed that  $\vdash A$  iff—no matter the probability function  $P$  for  $L$ — $P(A) = 1$ . But  $P(B) \leq 1$  for every statement  $B$  of  $L$ . Hence  $\vdash A$  iff—no matter the probability function  $P$  for  $L$  and statement  $B$  of  $L$ — $P(A) \not\leq P(B)$ .



## ACKNOWLEDGEMENTS

The essay was written while the author held a research grant (SES 8007179) from the National Science Foundation and was on a partial research leave from Temple University. Some of the material was used in lectures delivered at City College of CUNY in 1980–1981 and to be published by D. Reidel Publishing Company under the title *Probabilistic Semantics*.

Thanks are due to Ermanno Bencivenga, Kent Bendall, John Corcoran, Melvin C. Fitting, Raymond D. Gumb, John T. Kearns, Jack Nelson, Muffy E. A. Siegel, George Weaver, William A. Wisdom, and graduate students at Temple University for reading an earlier version of this essay; to Elizabeth Lazeski and Anne D. Ward for assisting in the preparation of the manuscript; to Thomas McGuinness and John Serembus for reading the proofs with me; and to my son Stephen for urging me, at a difficult time, to nonetheless write this essay.

## APPENDIX

I supplement here historical information provided in the main text.

A THE SUBSTITUTION INTERPRETATION AND  
SUBSTITUTIONAL SEMANTICS

The substitution interpretation of the quantifiers and, more particularly, substitutional semantics have a long and still unchronicled history. As remarked in note 2, Frege understood the quantifiers substitutionally in 1879, but switched later on to the objectual interpretation. Wittgenstein is quoted in [Moore, 1959, p. 297] as saying in the course of a 1932 lecture that “there was a temptation, to which he had yielded in the *Tractatus*, to say that  $(x) \cdot f(x)$  is identical with the logical product ‘ $fa \cdot fb \cdot fc \dots$ ’, and  $(\exists x) \cdot fx$  identical with the logical sum ‘ $fa \vee fb \vee fc \dots$ ’; but that this was in both cases a mistake.” In papers written in 1925 and 1926 F. P. Ramsey reported on the interpretation as Wittgenstein’s, defended it against an objection of Hilbert’s, and stated that, with the quantifiers interpreted Wittgenstein’s way, all axioms of Whitehead and Russell [1910–13] other than the Axiom of Reducibility hold true (see Ramsey [1926a] and [1926b]). Ramsey’s remark on Whitehead and Russell’s work may well be the first formal contribution to substitutional semantics.

To my knowledge little was heard (in any event, little was made) of the substitution interpretation of the quantifiers for the next fifteen years. It then turned up in a succession of books and papers by Carnap, beginning with [Carnap, 1942] and culminating in [Carnap, 1950]. The latter book is quite instructive. The account of truth you find there is essentially the

substitutional one I gave on page 69, but the account of logical truth, logical entailment, etc. is the truth-set one that I gave on pages 76–77.

Henkin's completeness paper, to which reference has frequently been made, appeared in 1949; the Robinson book mentioned on page 56 appeared in 1951; the first of several papers in which Ruth Marcus champions the substitution interpretation of the quantifiers, and particularly urges its adoption in modal logic, appeared in 1963; the Shoenfield book mentioned on page 56, appeared in 1967; etc. By then, of course, what I call *truth-value semantics* had come into its own, and it eventually captured some of the interest earlier accorded to substitutional semantics. The beginnings of truth-value semantics were reported in Section 3. since substitutional and truth-value semantics are close relatives, some of the information supplied there belongs here as well.

As the substitution interpretation gained greater currency, it was subjected—expectedly enough—to intense and often critical scrutiny. Davidson and some of his students wondered, for example, whether the substitutional account of truth satisfies Tarski's celebrated Convention *T*, a matter of considerable importance reviewed in [Kripke, 1976]. Other concerns were voiced by other writers (see [Quine, 1969], for instance), but in my opinion have been or could be met. Admittedly, the pros and cons of the substitution interpretation demand further study. As suggested at the outset, though, my attention in this essay goes to the novel accounts of logical truth, logical entailment, etc. that the interpretation allows.

## B TRUTH-VALUE SEMANTICS

Truth-value semantics, the reader may recall, dates back to 1959, the year that saw the publication of Beth's *Foundations of Mathematics*. Several of the contributions to truth-value semantics between 1959 and 1975 are recapitulated in [Leblanc, 1976]; they concern first-order logic without and with identity, second-order logic, a variety of modal logic's, three-valued logic, and presupposition-free variants of most of these. Further contributions will be found in [Leblanc, 1973] (the proceedings of a Temple University conference on alternative semantics) and in Part 2 of [Leblanc, 1982a] (a set of papers on truth-value semantics I authored or co-authored from 1968 onwards). The papers in the bibliography of [Kripke, 1976] deal primarily with the substitution interpretation of the quantifiers, but many of them touch on truth-value semantics as well. A few additional titles will be found in the bibliography of this essay: [Kearns, 1978], which distinguished three brands of substitution interpretation and—hence—of truth-value semantics; [Garson, 1979], which investigates which logics are susceptible of a substitutional semantics and—hence—of a truth-value one; [Parsons, 1971] and [Gottlieb and McCarthy, 1979], which attend to substitutional and truth-

value semantics in the particular context of set theory; [McArthur, 1976], [McArthur and Leblanc, 1976] and [Barnes and Gumb, 1979], which provide a truth-value semantics for tense logic, [Leblanc and Gumb, 1983] and [Leblanc, 1982b] which provide one for intuitionistic logic; and [Gumb, 1978; Gumb, 1979], which sport an important variant of truth-value semantics and also deal with model sets.

Some of these texts are explicitly labelled contributions to truth-value semantics; others are not. A few, anxious to show logical truth and entailment outgrowths of truth-functional truth and entailment, employ the truth-value assignments of page 73; others employ the truth-value assignments of page 76, or—following the precedent of [Schütte, 1960]—truth-value functions.

## C PROBABILISTIC SEMANTICS

Popper devised autonomous constraints for singular probability functions as early as 1938 (see [Popper, 1959, Appendix\*ii]), and attended to binary ones only eighteen years later. Yet contributions to singular probabilistic semantics are comparatively recent and few in number: [Stalnaker, 1970; Bendall, 1979; Leblanc, 1982c], portions of this essay, etc. The first two study the relationship between singular probability functions and what I call truth-value functions (Stalnaker talks instead of *truth-valuation functions* and limits himself to quantifierless statements); the third defines logical truth and logical entailment as on page 87, proves some of the theorems in Sections 4 and 5 (referring the reader to this text for proof of Theorem 100), and shows Constraints **C1–C6** to pick out the same functions as Kolmogorov’s constraints and Popper’s do.

Binary probabilistic semantics has a more eventful history, and one spreading over more than two decades.

The earliest account of *truth-functional truth* to employ binary probability functions may be that in [Leblanc, 1960]. As pointed out in [Stalnaker, 1970], it is unfortunately too broad, suiting only the binary probability functions attributed to Carnap in [Harper *et al.*, 1983]. A correct account of the notion can, incidentally, be retrieved from Stalnaker’s paper.

The earliest *published* accounts of *logical truth* and *logical entailment* to employ binary probability functions may be those in [Field, 1977]. Dissatisfied with Field’s resort to limits, I proposed in [Leblanc, 1979b] substitute—but equivalent—accounts of the two notions, those used in this essay (page 109). However, while my paper was in press, I learned that Harper had used the very same accounts in his unpublished doctoral dissertation of 1974. A summary of the dissertation is to appear in [Leblanc, 1983]. Further, but again equivalent, accounts will be found in [Adams, 1981; Van Fraassen, 1981; Morgan and Leblanc, 1983a], and doubtless many other texts. Incidentally, [Van Fraassen, 1981] uses in lieu of **D7** constraints which

render the term extensions of  $L$  unnecessary but rule out some of the functions acknowledged here. Like substitutes for **C7** have yet to be found.

The earliest theorem that probabilistic semantics boasts of is in [Popper, 1959, Appendix \*v]. A soundness theorem, it is roughly to the effect that if a Boolean identity  $A = B$  ( $A$  and  $B$  here either sets or statements) is provable by means of the ‘fourth set’ of postulates in [Huntington, 1933], then  $P(A/C) = P(B/C)$  for any set or statement  $C$  and any binary probability function  $P$  meeting Popper’s constraints. I extended the result in [Leblanc, 1960] showing in effect that if  $\vdash_0 A$ , where  $A$  is a quantifierless statement of  $L$ , then  $P(A/B) = 1$  for any quantifierless statement  $B$  of  $L$  and any binary probability function  $P$  meeting Popper’s constraints. Soundness and completeness theorems essentially like Theorem 132 and Theorem 137 in this essay were proved in [Harper, 1974; Field, 1977; Leblanc, 1979b] and most probably other texts as well. The strategy used on pages 112–116 to prove Theorem 130 is due to Harper and already figures in [Harper *et al.*, 1983]; details, however are simpler here. The strategy used on page 118 to prove Theorem 135 is essentially that in [Leblanc, 1979b], but again details are simpler here.

As announced on page 55, I confined myself in this essay to first-order logic without identity. [Gaifman, 1964; Gumb, 1983] and [Seager, 1983] attend to first-order logic with identity; [Van Fraassen, 1981] attends to intuitionistic logic, as do Morgan and Leblanc [1983a; 1983b]; Morgan [1982b; 1982a] attend to modal logic, as does [Schotch and Jennings, 1981]; several papers, [Stalnaker, 1970] possibly the earliest of them, attend the conditional logic (see Section 9 of Nute’s essay in this Handbook [II,9] for further information); etc. The majority of these results will be surveyed in *Probabilistic Semantics*.

*Temple University*

## BIBLIOGRAPHY

- [Adams, 1981] E. W. Adams. Transmissible improbabilities and marginal essentialness of premises in inferences involving indicative conditionals. *Journal of Philosophical Logic*, 10:149–178, 1981.
- [Barnes and Gumb, 1979] R. F. Barnes, Jr. and R. D. Gumb. The completeness of presupposition-free tense logics. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 25:192–208, 1979.
- [Behmann, 1922] H. Behmann. Beiträge zur Algebra der Logik, in besondere zum Entscheidungsproblem. *Math Annalen*, 86:163–229, 1922.
- [Bendall, 1979] K. Bendall. Belief-theoretic formal semantics for first-order logic and probability. *Journal of Philosophical Logic*, 8:375–394, 1979.
- [Bergmann *et al.*, 1980] M. Bergmann, J. Moore, and J. C. Nelson. *The Logic Book*. Random House, New York, 1980.
- [Bernays, 1922] P. Bernays. Review of [Behmann, 1922]. *Jahrbuch über die Fortschritte der Mathematik*, 48:1119, 1922.

- [Beth, 1959] E. W. Beth. *The Foundations of Mathematics*. North-Holland, Amsterdam, 1959.
- [Carnap and Jeffrey, 1971] R. Carnap and R. C. Jeffrey. *Studies in Inductive Logic and Probability*, volume I. University of California Press, Berkeley and Los Angeles, CA, 1971.
- [Carnap, 1942] R. Carnap. *Introduction to Semantics*. Harvard University Press, Cambridge, MA, 1942.
- [Carnap, 1950] R. Carnap. *Logical Foundations of Probability*. University of Chicago Press, Chicago, IL, 1950.
- [Carnap, 1952] R. Carnap. *The Continuum of Inductive Methods*. University of Chicago Press, Chicago, IL, 1952.
- [De Finetti, 1937] B. De Finetti. La prévision: Ses lois logiques, ses sources subjectives. *Annales de l'Institut Henri Poincaré*, 7:1–68, 1937.
- [Dunn and Belnap, 1968] J. M. Dunn and Jr. N. D. Belnap. The substitution interpretation of the quantifiers. *Noûs*, 2:177–185, 1968.
- [Ellis, 1979] B. Ellis. *Rational Belief Systems*. APQ Library of Philosophy, Rowmans and Littlefield, Totowa, NJ, 1979.
- [Field, 1977] H. H. Field. Logic, meaning and conceptual role. *Journal of Philosophy*, 74:379–409, 1977.
- [Fitch, 1948] F. B. Fitch. Intuitionistic modal logic with quantifiers. *Portugaliae Mathematica*, 7:113–118, 1948.
- [Frege, 1879] G. Frege. *Begriffsschrift*. Halle, 1879.
- [Frege, 1893–1903] G. Frege. *Grundgesetze der Arithmetik*. Jena, 1893–1903.
- [Gaifman, 1964] H. Gaifman. Concerning measures on first-order calculi. *Israel Journal of Mathematics*, 2:1–18, 1964.
- [Garson, 1979] J. W. Garson. The substitution interpretation and the expressive power of intensional logic. *Notre Dame Journal of Formal Logic*, 20:858–864, 1979.
- [Gentzen, 1934–35] G. Gentzen. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1934–35.
- [Goldfarb, 1979] W. D. Goldfarb. Logic in the twenties: the nature of the quantifier. *Journal of Symbolic Logic*, 44:351–68, 1979.
- [Gottlieb and McCarthy, 1979] D. Gottlieb and T. McCarthy. Substitutional quantification and set theory. *Journal of Philosophical Logic*, 8:315–331, 1979.
- [Gumb, 1978] R. D. Gumb. Metaphor theory. *Reports on Mathematical Logic*, 10:51–60, 1978.
- [Gumb, 1979] R. D. Gumb. *Evolving Theories*. Haven Publishing, NY, 1979.
- [Gumb, 1983] R. D. Gumb. Comments on probabilistic semantics. In H. Leblanc *et al.*, editor, *Essays in Epistemology and Semantics*. Haven Publishing, NY, 1983.
- [Harper *et al.*, 1983] W. L. Harper, H. Leblanc, and B. C. Van Fraassen. On characterising popper and carnap probability functions. In H. Leblanc *et al.*, editor, *Essays in Epistemology and Semantics*. Haven Publishing, NY, 1983.
- [Harper, 1974] W. L. Harper. *Counterfactuals and representations of rational belief*. PhD thesis, University of Rochester, NY, 1974.
- [Hasenjaeger, 1953] G. Hasenjaeger. Eine bemerkung zu henkin's beweis für die vollständigkeit des prädikatenkalküls der ersten stufe. *Journal of Symbolic Logic*, 18:42–48, 1953.
- [Henkin, 1949] L. Henkin. The completeness of the first-order functional calculus. *Journal of Symbolic Logic*, 14:159–166, 1949.
- [Hintikka, 1955] J. Hintikka. Two papers on symbolic logic. *Acta Philosophica Fennica*, 8, 1955.
- [Huntington, 1933] E. V. Huntington. New sets of independent postulates for the algebra of logic. *Transactions of the American Mathematical Society*, 35:274–304, 1933.
- [Jeffrey, 1990] R. C. Jeffrey. *Formal Logic: Its Scope and Limits*, 3rd edition. McGraw-Hill, NY, (1st edition, 1967), 1990.
- [Jeffreys, 1939] H. Jeffreys. *Theory of Probability*. Oxford University Press, Oxford, 1939.
- [Kearns, 1978] J. T. Kearns. Three substitution-instance interpretations. *Notre Dame Journal of Formal Logic*, 19:331–354, 1978.

- [Keynes, 1921] J. M. Keynes. *A Treatise on Probability*. Macmillan, London, 1921.
- [Kolmogorov, 1933] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*. Berlin, 1933.
- [Kripke, 1976] S. Kripke. Is there a problem about substitutional quantification? In G. Evans and J. McDowell, editors, *Truth and Meaning*, pages 325–419. Clarendon Press, Oxford, 1976.
- [Leblanc and Gumb, 1983] H. Leblanc and R. D. Gumb. Soundness and completeness proofs for three brands of intuitionistic logic. In H. Leblanc, R. D. Gumb, and R. Stern, editors, *Essays in Epistemology and Semantics*. Haven Publishing, NY, 1983.
- [Leblanc and Morgan, 1984] H. Leblanc and C. G. Morgan. Probability functions and their assumption sets: The binary case. *Synthese*, **60**, 91–106, 1984.
- [Leblanc and Wisdom, 1993] H. Leblanc and W. A. Wisdom. *Deductive Logic*, 3rd edition. Allyn and Bacon, Boston, MA, (1st edition, 1972), Prentice-Hall, 1993.
- [Leblanc et al., 1983] H. Leblanc, R. D. Gumb, and R. Stern, editors. *Essays in Epistemology and Semantics*. Haven Publishing, NY, 1983.
- [Leblanc, 1960] H. Leblanc. On requirements for conditional probability functions. *Journal of Symbolic Logic*, **25**:171–175, 1960.
- [Leblanc, 1966] H. Leblanc. *Techniques of Deductive Inference*. Prentice-Hall, Englewood Cliffs, NJ, 1966.
- [Leblanc, 1968] H. Leblanc. A simplified account of validity and implication for quantificational logic. *Journal of Symbolic Logic*, **33**:231–235, 1968.
- [Leblanc, 1973] H. Leblanc. *Truth, Syntax and Modality*. North-Holland, Amsterdam, 1973.
- [Leblanc, 1976] H. Leblanc. *Truth-Value Semantics*. North-Holland, Amsterdam, 1976.
- [Leblanc, 1979a] H. Leblanc. Generalization in first-order logic. *Notre Dame Journal of Formal Logic*, **20**:835–857, 1979.
- [Leblanc, 1979b] H. Leblanc. Probabilistic semantics for first-order logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **25**:497–509, 1979.
- [Leblanc, 1981] H. Leblanc. What price substitutivity? A note on probability theory. *Philosophy of Science*, **48**:317–322, 1981.
- [Leblanc, 1982a] H. Leblanc. *Existence, Truth and Provability*. SUNY Press, Albany, NY, 1982.
- [Leblanc, 1982b] H. Leblanc. Free intuitionistic logic: A formal sketch. In J. Agassi and R. Cohen, editors, *Scientific Philosophy Today: Essays in Honor of Mario Bunge*, pages 133–145. D. Reidel, Dordrecht, 1982.
- [Leblanc, 1982c] H. Leblanc. Popper's 1955 axiomatization of absolute probability. *Pacific Philosophical Quarterly*, **63**:133–145, 1982.
- [Leblanc, 1983] H. Leblanc. Probability functions and their assumption sets: The singular case. *Journal of Philosophical Logic*, **12**, 1983.
- [Löwenheim, 1915] L. Löwenheim. Über möglichkeiten im relativkalkül. *Mathematischen Annalen*, **76**:447–470, 1915.
- [Marcus, 1963] R. B. Marcus. Modal logics I: Modalities and international languages. In M. W. Wartofsky, editor, *Proceedings of the Boston Colloquium for the Philosophy of Science, 1961–1962*. D. Reidel, Dordrecht, 1963.
- [McArthur and Leblanc, 1976] R. P. McArthur and H. Leblanc. A completeness result for quantificational tense logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **22**:89–96, 1976.
- [McArthur, 1976] R. P. McArthur. *Tense Logic*. D. Reidel, Dordrecht, 1976.
- [Moore, 1959] G. E. Moore. *Philosophical Papers*. Allen and Unwin, London, 1959.
- [Morgan and Leblanc, 1983a] C. G. Morgan and H. Leblanc. Probabilistic semantics for intuitionistic logic. *Notre Dame Journal of Formal Logic*, **23**:161–180, 1983.
- [Morgan and Leblanc, 1983b] C. G. Morgan and H. Leblanc. Probability theory, intuitionism, semantics and the Dutch Book argument. *Notre Dame Journal of Formal Logic*, **24**:289–304, 1983.
- [Morgan and Leblanc, 1983c] C. G. Morgan and H. Leblanc. Satisfiability in probabilistic semantics. In H. Leblanc, R. D. Gumb, and R. Stern, editors, *Essays in Epistemology and Semantics*. Haven Publishing, NY, 1983.

- [Morgan, 1982a] C. G. Morgan. Simple probabilistic semantics for propositional K, T, B, S4 and S5. *Journal of Philosophical Logic*, 11:442–458, 1982.
- [Morgan, 1982b] C. G. Morgan. There is a probabilistic semantics for every extension of classical sentence logic. *Journal of Philosophical Logic*, 11:431–442, 1982.
- [Morgan, 1983] C. G. Morgan. Probabilistic semantics for propositional modal logics. In H. Leblanc, R. D. Gumb, and R. Stern, editors, *Essays in Epistemology and Semantics*. Haven Publishing, NY, 1983.
- [Orenstein, 1979] A. Orenstein. *Existence and the Particular Quantifier*. Temple University Press, Philadelphia, PA, 1979.
- [Parsons, 1971] C. Parsons. A plea for substitutional quantification. *Journal of Philosophy*, 68:231–237, 1971.
- [Popper, 1955] K. R. Popper. Two autonomous axiom systems for the calculus of probabilities. *British Journal of the Philosophy of Science*, 6:51–57, 176, 351, 1955.
- [Popper, 1957] K. R. Popper. Philosophy of science: a personal report. In A. C. Mace, editor, *British Philosophy in Mid-Century*, pages 155–191. Allen and Unwin, London, 1957.
- [Popper, 1959] K. R. Popper. *The Logic of Scientific Discovery*. Basic Books, New York, 1959.
- [Quine, 1940] W. V. Quine. *Mathematical Logic*. Norton, NY, 1940.
- [Quine, 1969] W. V. Quine. *Ontological Relativity and Other Essays*. Columbia University Press, New York and London, 1969.
- [Ramsey, 1926a] F. P. Ramsey. The foundations of mathematics. *Proceedings of the London Mathematical Society, Series 2*, 25:338–384, 1926.
- [Ramsey, 1926b] F. P. Ramsey. Mathematical logic. *The Mathematical Gazette*, 13:185–194, 1926.
- [Reichenbach, 1935] R. Reichenbach. *Wahrscheinlichkeitslehre*. Leiden, 1935.
- [Rényi, 1935] A. Rényi. On a new axiomatic theory of probability. *Acta Mathematica Aca. Scient. Hungaricae*, 6:285–335, 1935.
- [Robinson, 1951] A. Robinson. *On the Mathematics of Algebra*. North-Holland, Amsterdam, 1951.
- [Rosser, 1953] J. B. Rosser. *Logic for Mathematicians*. McGraw Hill, NY, 1953.
- [Schotch and Jennings, 1981] P. K. Schotch and R. E. Jennings. Probabilistic considerations on modal semantics. *Notre Dame Journal of Formal Logic*, 22:227–238, 1981.
- [Schütte, 1960] K. Schütte. Syntactical and semantical properties of simple type theory. *Journal of Symbolic Logic*, 25:305–326, 1960.
- [Schütte, 1962] K. Schütte. *Lecture Notes in Mathematical Logic*, volume 1. Pennsylvania State University, 1962.
- [Seager, 1983] W. Seager. Probabilistic semantics, identity and belief. *Canadian Journal of Philosophy*, 12, 1983.
- [Shoenfield, 1967] J. R. Shoenfield. *Mathematical Logic*. Addison-Wesley, Reading, MA, 1967.
- [Skolem, 1920] T. A. Skolem. Logisch-Kombinatorische Untersuchungen über die Erfüllbarkeit und Beweisbarkeit mathematischer Sätze nebst einem Theorem über dichte Mengen. *Skrifter utgit av Videnskapsselskapet i Kristiania, I. Matematisk-naturvidenskabelig klasse 1920*, 4:1–36, 1920.
- [Smullyan, 1968] R. M. Smullyan. *First-Order Logic*. Springer-Verlag, NY, 1968.
- [Stalnaker, 1970] R. Stalnaker. Probability and conditionals. *Philosophy of Science*, 37:64–80, 1970.
- [Stevenson, 1973] L. Stevenson. Frege's two definitions of quantification. *Philosophical Quarterly*, 23:207–223, 1973.
- [Tarski, 1930] A. Tarski. Fundamentale Begriffe der Methodologie der deductiven Wissenschaften. I. *Monatshefte für Mathematik und Physik*, 37:361–404, 1930.
- [Tarski, 1935–36] A. Tarski. Grundzüge des systemkalkül. *Fundamenta Math*, 25, 26:503–526, 283–301, 1935–36.
- [Tarski, 1936] A. Tarski. Der Wahrheitsbegriff in den formalisierten Sprachen. *Studia Philosophica*, 1:261–405, 1936.
- [Thomason, 1965] R. H. Thomason. *Studies in the formal logic of quantification*. PhD thesis, Yale University, 1965.

- [Van Fraassen, 1981] B. C. Van Fraassen. Probabilistic semantics objectified. *Journal of Philosophical Logic*, 10:371–394, 495–510, 1981.
- [Van Fraassen, 1982] B. C. Van Fraassen. Quantification as an act of mind. *Journal of Philosophical Logic*, 11:343–369, 1982.
- [Von Wright, 1957] G. R. Von Wright. *The Logical Problem of Induction*, volume Second, revised edition. MacMillan, NY, 1957.
- [Whitehead and Russell, 1910–1913] A. N. Whitehead and B. Russell. *Principia Mathematica*. Cambridge University Press, Cambridge, 1910–1913.
- [Wittgenstein, 1921] L. Wittgenstein. *Tratatus logico-philosophicus* (Logisch-philosophische Abhandlung). *Annalen der Naturphilosophie*, 14:185–262, 1921.

#### EDITOR'S NOTE

The following additional books are relevant to this and related chapters:

#### BIBLIOGRAPHY

- [Adams and Skyrms, 1998] E. W. Adams and B. Skyrms. *A Primer of Probability Logic*. CSLI Publications, 1998.
- [Hailperin, 1996] T. Hailperin. *Sentential Probability Logic: Origins, Development, Current Status and Technical Applications*. Lehigh University Press, 1996.
- [Roeper and Leblanc, 1999] P. Roeper and H. Leblanc. *Probability Theory and Probability Logic*.





H. ANDRÉKA, I. NÉMETHI, I. SAIN

## ALGEBRAIC LOGIC

*We dedicate this work to  
J. Donald Monk  
who taught us algebraic logic and more.*

### INTRODUCTION

Algebraic logic can be divided into two main parts. Part I studies algebras which are relevant to logic(s), e.g. algebras which were obtained from logics (one way or another). Since Part I studies algebras, its methods are, basically, algebraic. One could say that Part I belongs to ‘Algebra Country’. Continuing this metaphor, Part II deals with studying and building the bridge between Algebra Country and Logic Country. Part II deals with the methodology of solving logic problems by (i) translating them to algebra (the process of algebraization), (ii) solving the algebraic problem (this really belongs to Part I), and (iii) translating the result back to logic. There is an emphasis here on step (iii), because without such a methodological emphasis one could be tempted to play the ‘enjoyable games’ (i) and (ii), and then forget about the ‘boring duty’ of (iii). Of course, this bridge can also be used backwards, to solve algebraic problems with logical methods. We will give some simple examples for this in the present work.

Accordingly, the present work consists of two parts, too. Parts I and II of the Paper deal with the corresponding parts of algebraic logic. More specifically, Part I deals with the algebraic theory in general, and with algebras of sets of sequences, or algebras of relations, in particular. Part II deals with the methodology of algebraization of logics and logical problems, equivalence theorems between properties of logics and properties of (classes of) algebras, and in particular, discusses concrete results about logics obtained via this methodology of algebraization. Since Part II deals with general connections between logics and algebras, a general definition of what we understand by a logic or logical system is needed. Of course, such a definition has to be broad enough to be widely applicable and narrow enough to support interesting theorems. The first Section of Part II is devoted to finding such a definition.

We need to make a disclaimer here. Algebraic logic, today, is an extremely broad subject. We could not cover all of it. In Part II we managed to be broader than in Part I. Even in Part II we could not come even close to discussing the important research directions, but the definitions in Part II are

general enough to render the results applicable to all those logics which W. J. Blok and D. L. Pigozzi call algebraizable (cf. e.g. [Blok and Pigozzi, 1989; Font and Jansana, 1994]). Most of what we say in Part II can be generalized even beyond this, e.g. to the equivalential logics of J. Czelakowski. Further possibilities of generalizing Part II beyond algebraizable logics are in recent works of Blok and Pigozzi, and others, cf. e.g. [Blok and Pigozzi, 1986; Pigozzi, 1991; Czelakowski and Pigozzi, 1999; Andréka *et al.*, to appear] and [Czelakowski, 1997].

In Part I we had to be more restrictive. We concentrated attention to those kinds of algebras which are connected to the idea of ‘relations’ (one way or another), the idea of sets of pairs, or sets of triples, sets of sequences or something related to these. An important omission is the theory of Boolean Algebras with Operators (BAO’s). BAO’s are related to algebras of relations, and they provide an important unifying theory of many of the algebras we discuss here. Another important omission is Category Theoretic Logic. That branch of algebraic logic is not (at all) unrelated to what we are discussing here, but for various reasons we could not include an appropriate discussion here. In this connection more references are given in the survey [Németi, 1991]. Here we mention only [Makkai, 1987; Makkai and Reyes, 1977; Zlatoš, 1985; Makkai and Paré, 1989]. We could not cover polyadic algebras, either. However, their (basic) theory is analogous to that of cylindric algebras which we do discuss in detail. There are a few exceptional points where the two theories wildly diverge, e.g. in [Németi and Sági, to appear] it was proved that the equational theory of representable polyadic algebras is highly non-computable (while that of cylindric algebras is recursively enumerable). We refer the reader to the survey paper [Németi, 1991] and to [Henkin, Monk and Tarski, 1985] for modern overviews of polyadic algebras. Cf. also [Sain and Thompson, 1991; Pigozzi and Salibra, 1993; Andréka *et al.*, 1998]. Further important omissions are: (i) the finitization problem (cf. [Németi, 1991, beginning with Remark 2], [Sain, 1995; Simon, 1993; Madarász, Németi and Sági, 1997]); (ii) propositional modal logics of quantification, and connections with the new research direction ‘Logic, Language and Information’ (cf. [Venema, 1995; Marx and Venema, 1997; Marx, Pólos and Masuch, 1996; Andréka, van Benthem and Németi, 1997; van Benthem and ter Meulen, 1997; van Benthem, 1997]); (iii) relativization as a methodology for turning negative results to positive (cf. [Németi, 1996; Monk, 1993; Marx, 1995; Mikulás, 1995; Marx and Venema, 1997; Andréka, van Benthem and Németi, 1996]). Also there are strong connections between algebraic logic and computer science, we do not discuss these here.

**On the history:** The invention of Boolean algebras belongs to the ‘pre-history’ of Part I. Algebras of sets of sequences (as in Part I) were studied by De Morgan, Peirce and Schröder in the last century; and the modern

form of their theory was created by Tarski and his school.<sup>1</sup> The history of Part II also goes back to Tarski and his followers, but is, in general, more recent. For more on history we refer to [Andréka *et al.*, to appear; Anellis and Houser, 1991; Blok and Pigozzi, 1991a; Blok and Pigozzi, 1989; Henkin, Monk and Tarski, 1985; Maddux, 1991; Pratt, 1992] and [Tarski and Givant, 1987].

## I: Algebras of Relations

### GETTING ACQUAINTED WITH THE SUBJECT OF PART I.

The algebraization of classical propositional logic, yielding Boolean algebras (in short BA's), was immensely successful. What happens then if we want to extend the original algebraization yielding BA's to other, more complex logics, among others, say, to predicate logic (first-order logic)?<sup>2</sup>

Boolean algebras can be viewed as algebras of unary relations. Indeed, the elements of a BA are subsets of a set  $U$ , i.e. unary relations over  $U$ , and the operations are the natural operations on unary relations, e.g. intersection, complementation. The problem of extending this approach to predicate logics boils down to the problem of expanding the natural algebras of unary relations to *natural algebras of relations of higher ranks*, i.e. of relations in general. The reason for this is, roughly speaking, the fact that the basic building blocks of predicate logics are predicates, and the meanings of predicates can be relations of arbitrary ranks.<sup>3</sup> Indeed, already in the middle of the last century, when De Morgan wanted to generalize algebras of propositional logic in the direction of what we would call today predicate logic, he turned to algebras of binary relations.<sup>4</sup> That was probably the

---

<sup>1</sup>Relation and cylindric algebras were introduced by Tarski, polyadic algebras were introduced by Halmos, algebras of sets of finite sequences were studied by Craig; for other kinds of algebras of sets of sequences cf. e.g. [Németi, 1991; Henkin, Monk and Tarski, 1985].

<sup>2</sup>The things we say here about predicate logic apply also to most logics having individual variables, hence to all quantifier logics. However, the present Paper need not be 'predicate logic centered' because our considerations apply also to many *propositional* logics, e.g. to Lambek Calculus, propositional dynamic logic, arrow logics, many-dimensional modal logics. C.f. e.g. [Marx and Venema, 1997; van Benthem, 1996; van Benthem and ter Meulen, 1997; Mikulás, 1995; Tarski and Givant, 1987].

<sup>3</sup>For more on this see Part II, Sections 4 and 7 of the present Paper.

<sup>4</sup>De Morgan illustrated the need for expanding the algebras of unary relations (i.e. BA's) to algebras of relations in general (the topic of Part I of the present Paper) by saying that the scholastics, after two millennia of Aristotelian tradition, were still unable to prove that if a horse is an animal, then a horse's tail is an animal's tail. (' $v_0$  is a tail of  $v_1$ ' is a binary relation.)

beginning of the quest for algebras of relations in general. Returning to this quest, the new algebras will, of course, have more operations than BA's, since between relations in general there are more kinds of connections than between unary relations (e.g. one relation might be the converse, sometimes called inverse, of the other). So, our algebras in most cases will be Boolean algebras with some further operations.

The *framework* for the quest for the natural algebras of relations is *universal algebra*. The reason for this is that universal algebra is the field which investigates classes of algebras in general, their interconnections, their fundamental properties etc. Therefore universal algebra can provide us for our search with a 'map and a compass' to orient ourselves. There is a further good reason for using universal algebra. Namely, universal algebra is not only a unifying framework, but it also contains powerful theories. E.g. if we know in advance some general properties of the kinds of algebras we are going to investigate, then universal algebra can reward us with a powerful machinery for doing these investigations. Among the special classes of algebras concerning which universal algebra has powerful theories are the so called *discriminator varieties* and the *arithmetical varieties*. At the same time, algebras originating from logic turn out to fall in one of these two categories, in most cases. More concretely, more than half of these algebras are in discriminator varieties and almost all are in arithmetical ones. Certainly, all the algebras studied in the present Paper are in arithmetical varieties. Therefore, awareness of these recent parts of universal algebra can be rewarding in algebraic logic. We will not assume familiarity with these theories of universal algebra, we will cite the relevant definitions and theorems when using them.<sup>5</sup>

Moreover, as we already said, most of our algebras will be BA's with some additional (extra-Boolean) operations. When these operations are distributive over the Boolean join, as will be the case most often, such algebras are called Boolean Algebras with Operators, in short BAO's. Many of our important classes of algebras will be discriminator varieties of BAO's. The theory of BAO's is well-developed.<sup>6</sup>

Let us return to our task of moving from BA's of unary relations to expanded BA's of relations in general. What are the elements of a BA? They

---

<sup>5</sup>Some good introductions to universal algebra and discriminator varieties are [Henkin, Monk and Tarski, 1971, Chapter 0], [Burris and Sankappanavar, 1981; Cohn, 1965; Grätzer, 1979; McKenzie, McNulty and Taylor, 1987; Werner, 1978].

<sup>6</sup>Distributivity of the extra-Boolean operations over join is used in the theory to build a well-working duality-theory for it (atom-structures or Kripke-frames, complex algebras). This duality theory is a quite central part of algebraic logic. Because of the limited size of the present Paper, we will not deal with this here. Some references are [Jónsson and Tarski, 1951], [Henkin, Monk and Tarski, 1971, section 2.5], [Jónsson, 1995; Goldblatt, 1990; Goldblatt, 1991; Venema, 1996; Venema, 1997; Andréka, Givant and Németi, 1995; Hodkinson, 1997; Andréka, Goldblatt and Németi, 1998].

are sets of ‘points’. What will be the elements of the expanded new algebras? One thing about them seems to be certain, they will be sets of sequences, because relations in general are sets of sequences. These sequences may be just pairs if the relation is binary, they may be triples if the relation is ternary, or they may be longer — or even more general kinds of sequences.<sup>7</sup> So, one thing is clear at this point, namely that the elements of our expanded BA’s of relations will be sets of sequences. Indeed, this applies to all known algebraizations of predicate logics or quantifier logics.<sup>8</sup>

At this point it might be useful to point out that the most obvious approach (to studying algebras of relations) based on the above observation (that the elements of the algebra are sets of sequences) leads to difficulties right at the start.<sup>9</sup> So, what is the most obvious approach? Consider some set  $U$ ; let  ${}^{<\omega}U$  denote the set of all finite sequences over  $U$ , and consider the BA  $\mathcal{P}({}^{<\omega}U)$  (the powerset of  ${}^{<\omega}U$  conceived as a BA the standard way). Now if we are given any finitary relation, say,  $R \subseteq U \times U$  over  $U$ , then  $R \in \mathcal{P}({}^{<\omega}U)$ . So  $\mathcal{P}({}^{<\omega}U)$  contains all relations over  $U$  independently of their ranks. Therefore it might be a candidate for being the universe of an algebra of relations. Before thinking about what the new, so called extra-Boolean operations on  $\mathcal{P}({}^{<\omega}U)$  should be, let us have another look at its Boolean structure: If  $R$  is a binary relation, we would like to obtain its complement  $(U \times U) \setminus R$  as a result of applying a Boolean operation to  $R$ . However, in our algebra  $\mathcal{P}({}^{<\omega}U)$ , the complement of  $R$  is not  $(U \times U) \setminus R$  but something infinitely bigger.

## 1 ALGEBRAS OF BINARY RELATIONS

The above difficulty with  $\mathcal{P}({}^{<\omega}U)$  motivates our concentrating first on the simplest nontrivial case, namely that of the *algebras of binary relations* (BRA’s). Actually, BRA’s will be strong enough to be called a truly first-order (as opposed to propositional) algebraic logic, namely the logic captured by BRA’s is strong enough to serve as a vehicle for set theory and

---

<sup>7</sup>There is another consideration pointing in the direction of sequences. Namely, the semantics of quantifier logics is defined via satisfaction of formulas in models, which in turn is defined via evaluations of variables, and these evaluations are sequences. The meaning of a formula in a model is the *set of those sequences* which satisfy the formula in that model. So we arrive again at sets of sequences. For more on this see Part II, Section 7 of the present Paper.

<sup>8</sup>As mentioned earlier, this also applies to the more complex propositional logics, like e.g. many-dimensional modal logic.

<sup>9</sup>With further work this approach can be turned into a fruitful approach to algebraizing logic, see [Németi, 1991, §7 (2–4) and the Section containing Facts 2, 3 at the end of §4]; see also [Henkin, Monk and Tarski, 1985, §5.6.(A survey).3, p. 265], and the references therein. The approach originates with Craig, but already the algebras in [Quine, 1936] consist of sets of finite sequences.

hence for ordinary metamathematics.<sup>10</sup>

Throughout this Paper,  $\mathcal{P}(U)$  denotes the powerset of  $U$ , and  $\mathfrak{B}(U)$  denotes the Boolean algebra (in short **BA**) with universe  $\mathcal{P}(U)$ , for any set  $U$ . Thus  $\mathcal{P}(U)$  is the set of all subsets of  $U$ , and

$$\mathfrak{B}(U) = \langle \mathcal{P}(U), \cup, - \rangle$$

where  $\cup$  is the binary operation of taking union of two subsets of  $U$ , and  $-$  is the unary operation of taking complement (w.r.t.  $U$ ) of a subset of  $U$ . Then  $\mathfrak{B}(U)$ , as well as any of its subalgebras, is a natural *algebra of unary relations* on  $U$ , because a unary relation on  $U$  is just a subset of  $U$ , hence an element of  $\mathcal{P}(U)$ .

A binary relation is a set of pairs. Thus the usual set-theoretic (or in other words, Boolean) operations of union and complementation can be performed on binary relations. First we consider two natural operations on binary relations that use the fact that we have sets of *pairs*, namely relation-composition and relation conversion. Let  $R, S$  be binary relations. Then their composition<sup>11</sup>  $R \circ S$  and the converse  $R^{-1}$  of  $R$  are defined as<sup>12</sup>

$$R \circ S = \{ \langle a, b \rangle : \exists c (aRc \text{ and } cSb) \}$$

$$R^{-1} = \{ \langle b, a \rangle : \langle a, b \rangle \in R \}.$$

By a *concrete algebra of binary relations*, a **cBRA**, we understand an algebra whose elements are binary relations having a greatest one among them, and whose operations are the Boolean ones: union and complementation (w.r.t. this greatest relation), relation-composition and relation conversion. Thus the universe of a **cBRA** is closed under these operations, e.g. the union and relation composition of any two elements of the algebra are also in the universe of the algebra.<sup>13</sup>

Formally, a **cBRA** is of the form

$$\mathfrak{A} = \langle A, \cup, -, \circ, {}^{-1} \rangle$$

where

<sup>10</sup>A very interesting class of algebras of relations which is halfway between **BA**'s and **BRA**'s is the class  $\text{RCA}_2$  of cylindric algebras of dimension 2. They will be discussed at the beginning of Section 2.

<sup>11</sup>This is denoted by  $R|S$  in part of the literature, e.g. in [Henkin, Monk and Tarski, 1985]. The reason for this is that in a large part of the literature,  $\circ$  is reserved for the case when  $R$  and  $S$  are functions and is written backwards, i.e. what we denote by  $R \circ S$  is denoted by  $S \circ R$ .

<sup>12</sup>Throughout this Paper we will use the convention that if  $R$  is a binary relation, then  $aRb$  means that  $\langle a, b \rangle \in R$ .

<sup>13</sup>To understand how (and why) the theory works, it would be enough to include only ' $\circ$ ' as 'extra-Boolean' operation. Inclusion of conversion is motivated by some of the applications. Cf. the discussion of **BSR** below Theorem 9 (this Section).

$A$  is a set of binary relations and  $A$  has a biggest element  $V$ ,

$\cup, -, \circ, {}^{-1}$  are total operations on  $A$ , which means that  
 $\{R \cup S, V - R, R \circ S, R^{-1}\} \subseteq A$  whenever  $R, S \in A$ .

An *algebra of binary relations*, a **BRA**, is an algebra isomorphic to a concrete algebra of binary relations. If  $\mathfrak{A}$  is a **BRA**, then  $1^{\mathfrak{A}}$  denotes the greatest element of  $\mathfrak{A}$ , which we shall sometimes call the *unit* of  $\mathfrak{A}$ .

Throughout, we use abbreviations like **BRA** also for denoting the corresponding class itself, e.g. **BRA** also denotes the class of all **BRA**'s, and **BA** also denotes the class of all **BA**'s MDNM.

The similarity type, or language, of our **BRA**'s should contain two binary function symbols for  $\cup$  and  $\circ$ , and two unary function symbols for  $-$  and  ${}^{-1}$ . In this Paper, for simplicity and suggestiveness, we use the symbols  $\cup, \circ, -, {}^{-1}$  for these. We hope, this will cause no confusion.<sup>14</sup> Typical equations holding in **BRA** are  $(x \cup y) \circ z = (x \circ z) \cup (y \circ z)$ ,  $(x \cup y)^{-1} = x^{-1} \cup y^{-1}$ . In the literature  $\vee, +$  are often used as function symbols for  $\cup$ , and likewise  $;\, \smile$  are used as function symbols for  $\circ, {}^{-1}$ . Using these symbols, the above equations look as  $(x \vee y); z = (x; z) \vee (y; z)$ ,  $(x \vee y)^{\smile} = x^{\smile} \vee y^{\smile}$ , or  $(x + y); z = (x; z) + (y; z)$ ,  $(x + y)^{\smile} = x^{\smile} + y^{\smile}$ .

So we will use the symbol  $\cup$  also in abstract Boolean algebras. Moreover, in abstract Boolean algebras we also will use  $\cap$  as derived operation:  $x \cap y \stackrel{\text{def}}{=} -(x \cup -y)$ .  $\leq, 0, 1$  will denote the ordering  $x \leq y \stackrel{\text{def}}{\iff} x \cup y = y$ , smallest element and biggest element, respectively. Thus  $\tau \leq \sigma$  is an equation.

Having a fresh look at our **BRA**'s with an abstract algebraic eye, we notice that they should be very familiar from the abstract algebraic literature. Namely, a **BRA**  $\mathfrak{A}$  consists of two well known algebraic structures, a Boolean algebra  $\langle A, \cup, - \rangle$  and an involuted semigroup  $\langle A, \circ, {}^{-1} \rangle$  sharing the same universe  $A$ . The two structures are connected so that they form a *normal Boolean algebra with operators*, in short a normal **BAO**, which means that each extra-Boolean operation is distributive over  $\cup$  (additivity) and takes the value 0 whenever at least one of the arguments is 0 (normality). Also  ${}^{-1}$  is a Boolean isomorphism and  $x \mapsto 1 \circ x$ , where 1 is the Boolean 1, defines a *complemented closure operation*<sup>15</sup> on  $A$ . The properties listed in

<sup>14</sup>When seeing, say ' $x \cup y$ ', the reader will have to decide whether this denotes a term of **BRA**'s built up from the variables  $x, y$ , or whether it denotes a set (the union of the sets  $x$  and  $y$ ).

<sup>15</sup>Closure operations are unary functions  $f : U \rightarrow U$ , where we have an ordering  $\leq$  on  $U$ .  $f$  is called a closure operation if it is order preserving, idempotent and increasing, i.e. if for all  $u, v \in U$  we have  $u \leq f(u) = ff(u)$  and  $u \leq v \Rightarrow f(u) \leq f(v)$ . Boolean orderings with closure operations on them are one of the central concepts of abstract algebra, for example topological spaces or subalgebras of an algebra are often represented as such.  $f$  is called a complemented closure operation if  $f(-fx) = -f(x)$ , i.e. the complement of a closed element is closed. For more on these see e.g. [Henkin, Monk and Tarski, 1971, p.38].



this paragraph define a nice variety **ARA** containing **BRA** and is a reasonable starting point for an axiomatic study of the algebras of binary relations.<sup>16</sup>

**DEFINITION 1.** (**ARA**, an abstract approximation of **BRA**) **ARA** is defined to be the class of all algebras of the similarity type of **BRA**'s which validate the following equations.

(1) The Boolean axioms<sup>17</sup>

$$\begin{aligned}x \cup y &= y \cup x, \\x \cup (y \cup z) &= (x \cup y) \cup z, \\-[-(x \cup y) \cup -(x \cup -y)] &= x.\end{aligned}$$

(2) The axioms of involuted semigroups, i.e.

$$\begin{aligned}(x \circ y) \circ z &= x \circ (y \circ z), \\(x \circ y)^{-1} &= y^{-1} \circ x^{-1}, \\x^{-1-1} &= x.\end{aligned}$$

(3) The axioms of normal **BAO**, i.e.<sup>18</sup>

$$\begin{aligned}(x \cup y) \circ z &= (x \circ z) \cup (y \circ z), \\(x \cup y)^{-1} &= x^{-1} \cup y^{-1}, \\0 \circ x = 0, \quad 0^{-1} &= 0.\end{aligned}$$

(4)  $^{-1}$  is a Boolean isomorphism and  $x \mapsto 1 \circ x$  is a complemented closure operation, i.e.

$$\begin{aligned}-(x^{-1}) &= (-x)^{-1}, \\x \leq 1 \circ x, \\-(1 \circ x) &= 1 \circ -(1 \circ x).\end{aligned}$$

If  $E$  is a set of equations, then  $\text{Mod}(E)$  denotes the class of all algebras (of a given similarity type) in which  $E$  holds. A class  $K$  of algebras is called a *variety*, or an *equational class*, if  $K = \text{Mod}(E)$  for some set  $E$  of equations. The following theorem is due to A. Tarski.

<sup>16</sup>Most of these postulates already appear in [De Morgan, 1964], and since then investigations of **ARA**'s have been carried on for almost 130 years.

<sup>17</sup>Problem 1.1 in [Henkin, Monk and Tarski, 1971, p.245], originating with H. Robbins, asks whether this is an axiom system for **BA**. This problem has recently been solved affirmatively (by the theorem prover program EQP developed at Argonne National Laboratory, USA). We will use this axiom system for **BA** in Part II, Section 7.1.

<sup>18</sup>We are omitting some axioms that follow from the already stated ones. E.g. here we omit  $x \circ (y \cup z) = (x \circ y) \cup (x \circ z)$ ,  $x \circ 0 = 0$ .

**THEOREM 2.** *BRA is an equational class, i.e. there is a set  $E$  of equations such that  $\text{BRA} = \text{Mod}(E)$ .*

To prove the above theorem, we will use the machinery of universal algebra. First we prove that BRA is closed under taking subalgebras and direct products. If  $\mathbf{K}$  is a class of algebras, then  $\mathbf{SK}$  denotes the class of all subalgebras of elements of  $\mathbf{K}$ ,  $\mathbf{PK}$ ,  $\mathbf{IK}$ ,  $\mathbf{HK}$  and  $\mathbf{UpK}$  denote the classes of all algebras isomorphic to direct products, isomorphic copies, homomorphic images, and ultraproducts of elements of  $\mathbf{K}$  respectively.<sup>19</sup> Thus  $\text{BRA} = \mathbf{IcBRA}$ .

**LEMMA 3.**  $\text{BRA} = \mathbf{SP}\{\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle : U \text{ is a set}\}$ .

**Proof.** Let  $V$  be a binary relation. We say that  $V$  is an *equivalence relation* if  $V$  is symmetric and transitive, i.e. if  $V^{-1} = V$  and  $V \circ V \subseteq V$ . The *field* of  $V$  is the smallest set  $U$  such that  $V \subseteq U \times U$ , i.e.  $U = \{u : (\exists v)[\langle u, v \rangle \in V \text{ or } \langle v, u \rangle \in V]\}$ . The following three statements (\*)–(\*\*\*) will not be difficult to check:

1. (\*) If  $\mathfrak{A} \in \text{cBRA}$ , then  $1^{\mathfrak{A}}$  is an equivalence relation.
- (\*\*) If  $V$  is an equivalence relation, then

$$\mathfrak{R}(V) \stackrel{\text{def}}{=} \langle \mathfrak{P}(V), \circ, {}^{-1} \rangle \in \text{cBRA}.$$

- (\*\*\*) Let  $I$  be a set and for all  $i \in I$  let  $V_i$  be an equivalence relation. Assume that the fields of the  $V_i$ s are pairwise disjoint. Then

$$\mathfrak{R}\left(\bigcup_{i \in I} V_i\right) \cong \text{P}_{i \in I} \mathfrak{R}(V_i).$$

Indeed, to see (\*), let  $\mathfrak{A} \in \text{cBRA}$  and  $V \stackrel{\text{def}}{=} 1^{\mathfrak{A}}$ . Then  $V \in A$ , hence  $V \circ V, V^{-1}$  are in  $A$  as well, hence  $V \circ V \subseteq V$  and  $V^{-1} \subseteq V$ , because  $V$  is the biggest element of  $A$ . But  $V^{-1} \subseteq V$  is equivalent to  $V^{-1} = V$ , hence  $V$  is an equivalence relation.

To show (\*\*), one has to check that for any  $R, S \subseteq V$  also  $R \circ S \subseteq V$  and  $R^{-1} \subseteq V$ . These follow from  $V \circ V \subseteq V, V^{-1} \subseteq V$ .

To show (\*\*\*), we define the function  $f : \mathcal{P}\left(\bigcup_{i \in I} V_i\right) \rightarrow \text{P}_{i \in I} \mathcal{P}(V_i)$  by letting for all  $X \subseteq \bigcup_{i \in I} V_i$ ,

$$f(X) \stackrel{\text{def}}{=} \langle X \cap V_i : i \in I \rangle.$$

---

<sup>19</sup>Note that  $\mathbf{S}$  is different from  $\mathbf{P}, \mathbf{I}, \mathbf{H}$  and  $\mathbf{Up}$  in that  $\mathbf{S} \neq \mathbf{IS}$ , while  $\mathbf{P} = \mathbf{IP}$  etc. For our reasons for defining  $\mathbf{S}$  this way see the remark after the definition of  $\mathbf{Alg}_m$  in Part II, Definition 42. We will use simple facts like  $\mathbf{IS} = \mathbf{SI}, \mathbf{IP} = \mathbf{PI}, \mathbf{SS} = \mathbf{S}$ , etc. without mentioning them.

Then it is easy to check that this  $f$  is the required isomorphism.

We are ready to prove the lemma. First we show that  $\mathbf{BRA} = \mathbf{SPBRA}$ . By definition,  $\mathbf{cBRA}$  is closed under taking subalgebras, so  $\mathbf{BRA}$  is also closed under taking subalgebras (because  $\mathbf{BRA} = \mathbf{IcBRA}$ ). Let  $I$  be a set, and let  $\mathfrak{A}_i \in \mathbf{cBRA}$  with unit  $V_i$  for each  $i \in I$ . We may assume that the  $V_i$ s have disjoint fields. Then  $\mathfrak{A}_i \subseteq \mathfrak{R}(V_i)$ , so  $\mathbf{P}\mathfrak{A}_i \subseteq \mathbf{P}\mathfrak{R}(V_i) \cong \mathfrak{R}(\bigcup V_i) \in \mathbf{cBRA}$  by  $(*)$ – $(***)$ . This shows that  $\mathbf{P}\mathfrak{A}_i$  is isomorphic to a  $\mathbf{cBRA}$ , i.e.  $\mathbf{BRA}$  is closed under taking direct products.

Now let  $\mathfrak{A} \in \mathbf{cBRA}$  with greatest element  $V$ . Then  $V$  is an equivalence relation, let  $U_i, i \in I$  be the blocks of this equivalence relation. Then  $U_i \times U_i$  are also equivalence relations with pairwise disjoint fields, and  $V$  is the union of these. Hence by  $(**)$ – $(***)$  we have that  $\mathfrak{A} \subseteq \mathbf{P}\langle \langle \mathfrak{P}(U_i \times U_i), \circ, {}^{-1} \rangle : U_i \text{ is a block of } 1^{\mathfrak{A}} \rangle$ . This completes the proof of Lemma 3. ■

To formulate our next lemma, we need the notions of a subdirect product and a discriminator term.

Subdirect products of algebras, and subdirectly irreducible algebras are defined in practically every textbook on universal algebra, cf. e.g. [Grätzer, 1979], or [Burris and Sankappanavar, 1981; Henkin, Monk and Tarski, 1971; McKenzie, McNulty and Taylor, 1987]. By a *subdirect product* we mean a subalgebra of a product such that the projections of the product restricted to the subalgebra remain surjective mappings. An algebra  $\mathfrak{A}$  is *subdirectly irreducible* if it is not (isomorphic to) a subdirect product of algebras different from  $\mathfrak{A}$ . We note that the one-element algebra is not subdirectly irreducible. By Birkhoff's classical theorem, every algebra is a subdirect product of some subdirectly irreducible ones. Therefore, the subdirectly irreducible algebras are often regarded as the basic building blocks of all the other algebras. In particular, when studying an algebra  $\mathfrak{A}$ , it is often enough to study its subdirectly irreducible building blocks. For a class  $\mathbf{K}$  of algebras,  $\mathbf{Sir}(\mathbf{K})$  denotes the class of subdirectly irreducible members of  $\mathbf{K}$ . For  $\mathbf{K} = \mathbf{BA}$ ,  $\mathbf{Sir}(\mathbf{BA})$  consists of the 2-element Boolean algebra only (up to isomorphisms).

We say that a class  $\mathbf{K}$  of algebras *has a discriminator term* iff there is a term  $\tau(x, y, z, u)$  in the language of  $\mathbf{K}$  such that in every member of  $\mathbf{K}$  we have

$$\tau(x, y, z, u) = \begin{cases} z, & \text{if } x = y, \\ u, & \text{if } x \neq y. \end{cases}$$

The term  $\tau$  above is called a *discriminator term*. Sometimes instead of the four-ary  $\tau$ , the ternary discriminator term  $t(x, y, z) = \tau(x, y, z, x)$  is used. They are interdefinable, since  $\tau(x, y, z, u) = t(t(x, y, z), t(x, y, u), u)$ . Therefore, it does not matter which one is used. Moreover, in classes of algebras which have a Boolean algebra reduct, like our  $\mathbf{BRAs}$  or  $\mathbf{ARAs}$ , the

discriminator term can be replaced with the so called *switching term*

$$c(x) = \begin{cases} 1, & \text{if } x \neq 0, \\ 0, & \text{if } x = 0. \end{cases}$$

By this we mean that in such a class of algebras, if  $\tau(x, y, z, u)$  is a discriminator term, then  $c(x) = \tau(x, 0, 0, 1)$  is a switching term, and vica versa, if  $c(x)$  is a switching term, then  $\tau(x, y, z, u) = [-c(x \oplus y) \cap z] \cup [c(x \oplus y) \cap u]$  is a discriminator term. Here, and later on,  $\oplus$  denotes symmetric difference, i.e.  $x \oplus y \stackrel{\text{def}}{=} (x \cap -y) \cup (-x \cap y)$ .

**LEMMA 4.**  $\mathbf{Sir}(\mathbf{BRA}) = \mathbf{IS}\{\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle : U \text{ is a nonempty set}\}$  and  $\mathbf{Sir}(\mathbf{BRA})$  has a discriminator term.

**Proof.** Let  $\mathbf{K} \stackrel{\text{def}}{=} \mathbf{IS}\{\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle : U \text{ is a nonempty set}\}$ . Let  $\mathfrak{A} \in \mathbf{BRA}$ . Then  $\mathfrak{A}$  is isomorphic to a subalgebra of  $\mathbf{P}\mathfrak{R}(U_i \times U_i)$  for some system  $\langle U_i : i \in I \rangle$  of sets, by Lemma 3. If  $U_i = \emptyset$ , then  $\mathfrak{R}(U_i \times U_i)$  is the one-element algebra which can be left out from any product, so we may assume that each  $U_i$  above is nonempty. But then  $\mathfrak{A}$  is a subdirect product of some  $\mathfrak{B}_i$ ,  $i \in I$  where each  $\mathfrak{B}_i$  is a subalgebra of  $\mathfrak{R}(U_i \times U_i)$ . This shows that  $\mathbf{Sir}(\mathbf{BRA}) \subseteq \mathbf{K}$ .

It is not difficult to check that

$$c(x) \stackrel{\text{def}}{=} 1 \circ x \circ 1$$

is a switching term on  $\mathfrak{R}(U \times U)$  for all  $U$ . Hence it is a switching term on  $\mathbf{K}$  also. Thus,  $\mathbf{K}$  has a discriminator term.

Finally, if  $\mathfrak{A}$  has a discriminator term, then  $\mathfrak{A}$  has no nontrivial congruences, i.e.  $\mathfrak{A}$  is simple. This is a basic fact of discriminator theory.<sup>20</sup> Clearly, any simple algebra is subdirectly irreducible, so  $\mathbf{K} \subseteq \mathbf{Sir}(\mathbf{BRA})$ . ■

We say that  $\mathbf{K}$  is a *pseudo-axiomatizable* class if there are an expansion  $\mathcal{L}$  of the language of  $\mathbf{K}$ , a set  $\Sigma$  of first-order formulas in this bigger language  $\mathcal{L}$  and a unary relation symbol  $U$  of  $\mathcal{L}$  such that

$$\mathbf{K} = \mathbf{Rd}_U \text{Mod} \Sigma,$$

where  $\text{Mod} \Sigma$  denotes the class of all models of  $\Sigma$ , and  $\mathbf{Rd}_U$  denotes the operator of taking reducts to the language of  $\mathbf{K}$  and restricting the universe to  $U$  at the same time. In more detail: Let  $\mathfrak{M}$  be a model of the language  $\mathcal{L}$ . Then  $\mathbf{Rd} \mathfrak{M}$  denotes the reduct of  $\mathfrak{M}$  to the language of  $\mathbf{K}$ , and  $\mathbf{Rd}_U \mathfrak{M}$

<sup>20</sup>The reason is the following. Assume that  $R$  is a nonidentity congruence of  $\mathfrak{A}$ . We will show that then  $R = A \times A$ . Let  $u, v \in A$ ,  $u \neq v$  be such that  $uRv$ , and let  $a, b \in A$  be arbitrary. Then  $a = \tau(u, u, a, b)R\tau(u, v, a, b) = b$ , so  $aRb$ .

denotes the restriction of the model  $\mathbf{Rd}\mathfrak{M}$  to the interpretation  $U^{\mathfrak{M}} \subseteq M$  of  $U$  in  $\mathfrak{M}$ . I.e. while  $\mathfrak{M}$  is a model of the bigger language  $\mathcal{L}$ ,  $\mathbf{Rd}_U\mathfrak{M}$  is a model of the smaller language of  $\mathbf{K}$ . If  $\mathbf{N}$  is a class of models of the language of  $\mathcal{L}$ , then  $\mathbf{Rd}_U\mathbf{N} \stackrel{\text{def}}{=} \{\mathbf{Rd}_U\mathfrak{M} : \mathfrak{M} \in \mathbf{N}\}$ .

It is known that pseudo-axiomatizable classes are closed under ultraproducts, this is easy to show.

LEMMA 5.  $\mathbf{Sir}(\mathbf{BRA})$  is a pseudo-axiomatizable class.

**Proof.** The expansion  $\mathcal{L}$  of the language of  $\mathbf{BRA}$  will be a many-sorted first-order language with three sorts:  $S, P$  and  $R$  (for set, pairs, and relations), two unary functions  $p_0, p_1$  from  $P$  to  $S$  (first and second projections), a binary relation  $\varepsilon$  between  $P$  and  $R$  (for ‘is an element of’), and binary functions  $\cup, \circ$  on  $R$ , unary functions  $-,^{-1}$  on  $R$ . The variables  $x, y, z$  are of sort  $S$ , the variables  $u, v, w$  are of sort  $P$ , and the variables  $a, b, c$  are of sort  $R$ . We also consider  $S, P, R$  as unary relations.<sup>21</sup> See Figure 1.

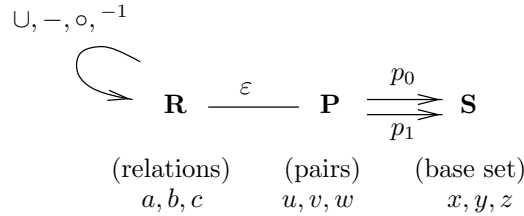


Figure 1.

The set  $\Sigma$  of axioms is as follows: In the following formulas we will write  $\varepsilon$  in infix mode, like  $u\varepsilon a$ . Also we will write comma in place of conjunction  $\wedge$ . There are free variables in the elements of  $\Sigma$ , validity of an open formula is meant in such a way that all the free variables are universally quantified at the beginning of the formula.  $\Sigma$  is defined to be  $\{(1a), (1b), \dots, (4)\}$ , where

The ‘pair-axioms’ are:

$$(1a) \quad (\exists u)(p_0(u) = x, p_1(u) = y).$$

$$(1b) \quad p_0(u) = p_0(v), p_1(u) = p_1(v) \rightarrow u = v.$$

<sup>21</sup>If one is not familiar with many-sorted models, then one can think of the above language as having  $S, P, R$  as unary relation symbols, and e.g.  $p_0$  as a binary relation. Then to our axioms we have to add statements like  $p_0(x, y) \rightarrow P(x)$ ,  $p_0(x, y) \rightarrow S(y)$ ,  $p_0(x, y), p_0(x, z) \rightarrow y = z$ ,  $P(x) \rightarrow (\exists y)p_0(x, y)$ . Then the fact that the variable  $x$  in the many-sorted language is of sort  $S$  while the variable  $u$  is of sort  $P$  means that one has to replace e.g. the formula  $\forall x \exists u p_0(x) = u$  with  $(\forall x)(S(x) \rightarrow \exists u(P(u) \wedge p_0(x, u)))$ .

Extensionality of sets of pairs:

$$(2) \quad \forall u(u\varepsilon a \leftrightarrow u\varepsilon b) \rightarrow a = b.$$

The definitions of the operations of cBRA:

$$(3a) \quad u\varepsilon(a \cup b) \leftrightarrow (u\varepsilon a \text{ or } u\varepsilon b).$$

$$(3b) \quad u\varepsilon(-a) \leftrightarrow \neg u\varepsilon a.$$

$$(3c) \quad u\varepsilon(a \circ b) \leftrightarrow (\exists vw)(v\varepsilon a, w\varepsilon b, p_0(u) = p_0(v), p_1(v) = p_0(w), \\ p_1(w) = p_1(u)).$$

$$(3d) \quad u\varepsilon(a^{-1}) \leftrightarrow (\exists v)(v\varepsilon a, p_0(u) = p_1(v), p_1(u) = p_0(v)).$$

There are at least two elements in the relations sort:

$$(4) \quad (\exists ab)a \neq b.$$

This finishes the definition of  $\Sigma$ . We will show that

$$\mathbf{Sir}(\mathbf{BRA}) = \mathbf{Rd}_R \mathbf{Mod} \Sigma.$$

Indeed, let  $\mathfrak{A} \in \mathbf{Sir}(\mathbf{BRA})$ , say  $\mathfrak{A}$  is isomorphic to a subalgebra of  $\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle$ . We may assume that  $\mathfrak{A} \subseteq \langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle$ . We define the three-sorted model  $\mathfrak{M}$  as follows.

$$\begin{aligned} S^{\mathfrak{M}} &\stackrel{\text{def}}{=} U, & P^{\mathfrak{M}} &\stackrel{\text{def}}{=} U \times U, & R^{\mathfrak{M}} &\stackrel{\text{def}}{=} A, \\ p_0^{\mathfrak{M}}(\langle u, v \rangle) &= u, & p_1^{\mathfrak{M}}(\langle u, v \rangle) &= v \text{ for all } u, v \in U, \\ \langle u, v \rangle \varepsilon^{\mathfrak{M}} a &\text{ iff } \langle u, v \rangle \in a, \text{ for all } u, v \in U, a \in A, \\ a \cup^{\mathfrak{M}} b &\stackrel{\text{def}}{=} a \cup b, & -^{\mathfrak{M}} a &\stackrel{\text{def}}{=} -a, & a \circ^{\mathfrak{M}} b &\stackrel{\text{def}}{=} a \circ b, & a^{-1 \mathfrak{M}} b &\stackrel{\text{def}}{=} a^{-1}. \end{aligned}$$

Then it is easy to check that  $\mathfrak{M} \models \Sigma$  and  $\mathbf{Rd}_R \mathfrak{M} = \mathfrak{A}$ . See Figure 2.

Conversely, let  $\mathfrak{M}$  be such that  $\mathfrak{M} \models \Sigma$ . Let  $U \stackrel{\text{def}}{=} S^{\mathfrak{M}}$ . Define the relation  $T$  between  $U \times U$  and  $P^{\mathfrak{M}}$  as follows:

$$\langle u, v \rangle T x \quad \text{iff} \quad p_0(x) = u, p_1(x) = v.$$

By (1a),(1b) in  $\Sigma$  then  $T$  is a bijection between  $U \times U$  and  $P^{\mathfrak{M}}$ . Therefore we will assume that

$$U \times U = P^{\mathfrak{M}} \quad \text{and} \quad u = p_0^{\mathfrak{M}}(\langle u, v \rangle), v = p_1^{\mathfrak{M}}(\langle u, v \rangle).$$

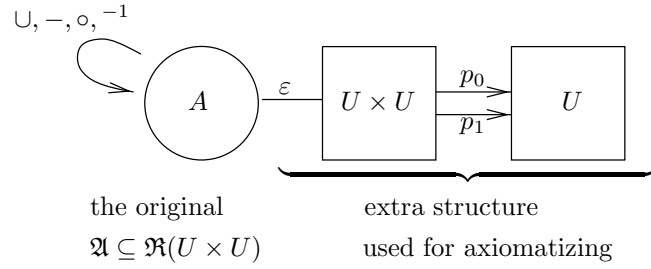


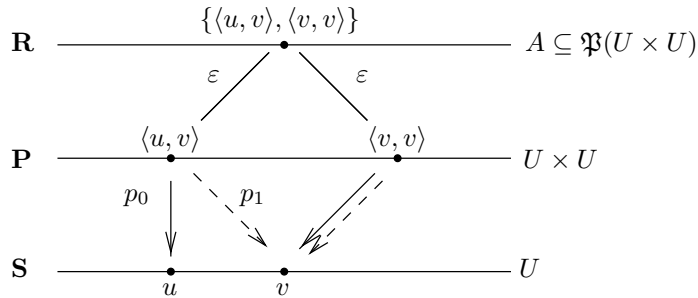
Figure 2.

We define now the function  $Q : R^{\mathfrak{M}} \rightarrow \mathcal{P}(U \times U)$  as

$$Q(a) = \{ \langle u, v \rangle : \langle u, v \rangle \varepsilon^{\mathfrak{M}} a \}.$$

See Figure 3. Then  $Q$  is one-to-one by (2) in  $\Sigma$ . Axioms (3a)–(3d) in  $\Sigma$  say that<sup>22</sup>

$$\begin{aligned} Q(a \cup^{\mathfrak{M}} b) &= Q(a) \cup Q(b), \\ Q(-^{\mathfrak{M}} a) &= (U \times U) \setminus Q(a), \\ Q(a \circ^{\mathfrak{M}} b) &= Q(a) \circ Q(b), \\ Q(a^{-1 \mathfrak{M}}) &= (Q(a))^{-1}. \end{aligned}$$



The three-story structure of  $\mathfrak{A}$

Figure 3.

This shows that  $Q$  is an isomorphism from  $\mathbf{Rd}_R \mathfrak{M}$  into  $\langle \mathfrak{P}(U \times U), \circ, -^1 \rangle$ . Finally, (4) in  $\Sigma$  implies that  $U$  is nonempty,  $\mathbf{Rd}_R \mathfrak{M}$  is nonempty. ■

**Proof of Theorem 2.** Now we are ready to prove Theorem 2, applying the following theorem of universal algebra (it follows easily from e.g. [Burris

<sup>22</sup>  $X \setminus Y \stackrel{\text{def}}{=} \{x \in X : x \notin Y\}$ .

and Sankappanavar, 1981, Thm. IV.9.4 (b,c)], and it is proved in detail in [Németi, 1991, Thm 9.1]).

**THEOREM** (Universal algebra). If  $\mathbf{SUpK}$  has a discriminator term, then  $\mathbf{SPUpK}$  is an equational class.

Indeed, let  $\mathbf{K} = \mathbf{Sir}(\mathbf{BRA})$ . Then  $\mathbf{K} = \mathbf{SK}$  by Lemma 4, and  $\mathbf{K} = \mathbf{UpK}$  by Lemma 5, thus  $\mathbf{K} = \mathbf{SUpK}$ . Also,  $\mathbf{K}$  has a discriminator term by Lemma 4. Thus  $\mathbf{SPUpK}$  is an equational class by the above theorem of universal algebra. But  $\mathbf{SPUpK} = \mathbf{SPK} = \mathbf{SPSir}(\mathbf{BRA}) = \mathbf{BRA}$  by Lemma 3, and we are done with proving that  $\mathbf{BRA}$  is an equational class. **QED** (Theorem 2)  $\blacksquare$

In universal algebra, an equational class  $\mathbf{K}$  such that  $\mathbf{SirK}$  has a discriminator term is called a *discriminator variety*. So we proved that  $\mathbf{BRA}$  is a discriminator variety.

The above proof of Theorem 2 uses techniques that can be applied in many cases in algebraic logic. E.g. these same techniques work for cylindric and polyadic algebras. See e.g. Theorems 10, 17.  $\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle$  is called the *full BRA* over the set  $U$ . By Lemma 3 we could have defined  $\mathbf{BRA}$  as

$$\begin{aligned} \mathbf{BRA} &= \mathbf{SP}\{\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle : U \text{ is a set}\}, & \text{or as} \\ \mathbf{BRA} &= \mathbf{SP}\{\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle : U \text{ is a nonempty set}\}. \end{aligned}$$

Set

$$\text{setBRA} \stackrel{\text{def}}{=} \mathbf{S}\{\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle : U \text{ is a nonempty set}\}.$$

Then  $\mathbf{BRA} = \mathbf{SPsetBRA}$ .<sup>23</sup> This fact, and the class  $\text{setBRA}$  will be used in Part II (Section 7.4) when translating our algebraic results to logic.

In the following, we will define our classes of algebras of relations in this style. So *when defining new kinds of algebras of relations, we will first define the simplest version* (e.g. the one with top element  $U \times U \times \dots \times U$ ), *and then take all subalgebras of all direct products of these*.

Let  $\mathbf{K} = \text{setBRA}$ . Then, as we have seen,  $\mathbf{BRA} = \mathbf{SPK}$  is a variety because  $\mathbf{K}$  has a discriminator term and  $\mathbf{K}$  is pseudo-axiomatizable.<sup>24</sup> In almost all

<sup>23</sup>Because  $\mathbf{SPS} = \mathbf{SP}$ , this is a basic theorem in universal algebra. See e.g. [Henkin, Monk and Tarski, 1971, 0.3.12].

<sup>24</sup>We could have proved  $\mathbf{K} = \mathbf{SUpK}$  more directly, as follows. An ultraproduct of full  $\mathbf{BRAs}$  on some sets  $U_i$  is isomorphic to the full  $\mathbf{BRA}$  on the ultraproduct of the  $U_i$ s, namely if  $F$  is an ultrafilter on  $I$ , and  $\mathfrak{F}(U)$  denotes  $\langle \mathfrak{P}(U \times U), \circ, {}^{-1} \rangle$ , then  $\mathbf{P}\mathfrak{F}(U_i)/F \cong \mathfrak{F}(\mathbf{PU}_i/F)$ , and the isomorphism  $h$  is given by  $a/F \mapsto \{\langle u/F, v/F \rangle : \{i \in I : \langle u_i, v_i \rangle \in a_i\} \in F\}$ . The reader is invited to check that  $h$  is indeed an isomorphism. This method also is applicable in many places. We chose the method of pseudo-axiomatizability for proving that  $\mathbf{K}$  is closed under ultraproducts, because we feel that this method reveals the real cause: our concrete algebras are usually pseudo-axiomatizable, because ‘concrete’ very much means this, i.e. ‘concrete’ means that there is some extra structure not coded in the operations, which means that this extra structure may disappear when taking isomorphic copies.



our cases,  $\mathbf{K}$ , where  $\mathbf{K}$  is the class of the corresponding set algebras, will be pseudo-axiomatizable because  $\mathbf{K}$  is defined to be a three-story structure like  $\mathbf{BRA}$ , only the operations on the third level will vary (and instead of  $U \times U$  we may have  $U \times U \times \dots \times U$ ), and in most cases  $\mathbf{K}$  will have a discriminator term.<sup>25</sup>

Theorem 2 indicates that  $\mathbf{BRA}$  is indeed a promising start for developing a nice algebraization of stronger logics (like e.g. quantifier logics), or in the non-logical perspective, for developing an algebraic theory of relations. After Theorem 2, the question comes up naturally whether we can strengthen the postulates defining  $\mathbf{ARA}$  to obtain a finite set  $E$  of equations describing the variety  $\mathbf{BRA}$ , i.e. such that  $\mathbf{BRA} = \text{Mod}(E)$  would be the case. The answer is due to J. D. Monk:

**THEOREM 6.**  *$\mathbf{BRA}$  is not finitely axiomatizable, i.e. for no finite set  $\Sigma$  of first-order formulas is  $\mathbf{BRA} = \text{Mod}(\Sigma)$ .*

The idea of one possible proof is explained in Remark 23 in Section 2 herein. This idea is based on the proof of Theorem 19 which is the reason why it is postponed to that part of the Paper. See [Monk, 1964], and also [Henkin, Monk and Tarski, 1985, 5.1.57, 4.1.3], for the original proof of Theorem 6 (in slightly different settings).

For a class  $\mathbf{K}$  of algebras, let  $\mathbf{EqK}$  denote the set of all equations valid in  $\mathbf{K}$ .

**THEOREM 7.**  *$\mathbf{Eq}(\mathbf{BRA})$  is recursively enumerable but not decidable.*

**Proof.** An equation holds in  $\mathbf{BRA}$  iff it holds in  $\mathbf{Sir}(\mathbf{BRA})$  by Lemma 3. Let  $\Sigma$  be the finite set of first-order formulas such that  $\mathbf{Sir}(\mathbf{BRA}) = \mathbf{Rd}_R \text{Mod}(\Sigma)$ , from the proof of Lemma 5. Thus an equation is valid in  $\mathbf{Sir}(\mathbf{BRA})$  iff it is valid in  $\text{Mod}(\Sigma)$  (when all the variables of the original equation are considered to be of sort  $R$ ). The consequences of any finite set of first-order formulas is recursively enumerable by the completeness theorem of first-order logic. Thus  $\mathbf{Eq}(\mathbf{BRA})$  is recursively enumerable (and an enumeration is given by the present proof).

The proof of undecidability of  $\mathbf{Eq}(\mathbf{BRA})$  goes via interpreting the quasi-equational theory of semigroups into  $\mathbf{Eq}(\mathbf{BRA})$ . The proof consists of two steps:

- (\*) An equational implication (i.e. a *quasi-equation*) about  $\circ$  is valid in all semigroups iff it is valid in  $\mathbf{BRA}$ .

---

<sup>25</sup>Even if  $\mathbf{K}$  would not have a discriminator term, then  $\mathbf{SPK}$  would still be a *quasi-variety*, i.e. definable by equational implications, because  $\mathbf{K}$  will be pseudo-axiomatizable, hence  $\mathbf{K} = \mathbf{UpK}$ , thus  $\mathbf{SPK} = \mathbf{SPUpK}$  will hold. It is a basic theorem of universal algebra that  $\mathbf{K}$  is a quasi-variety iff  $\mathbf{K} = \mathbf{SPUpK}'$  for some  $\mathbf{K}'$ .

(\*\*) To any equational implication  $q$  there is an equation  $e$  in the language of **BRA** such that  $\mathbf{BRA} \models q$  iff  $\mathbf{BRA} \models e$ .

*Proof of (\*):* If  $q$  is true in all semigroups, then it is true in **BRA** because  $\circ$  is associative in **BRA**. If  $q$  fails in a semigroup  $\langle S, \cdot \rangle$ , then take the Cayley-representation of this semigroup, this is an embedding of  $\langle S, \cdot \rangle$  into  $\langle \mathcal{P}(S' \times S'), \circ \rangle$  which is a reduct of  $\mathfrak{R}(S' \times S') \in \mathbf{BRA}$ . Thus  $q$  fails in **BRA**.

*Proof of (\*\*):* The reason is that **BRA** is a discriminator variety, and in every discriminator variety a quasi-equation  $q$  is equivalent to an equation  $e$  on the subdirectly irreducibles<sup>26</sup>. Now, by  $\mathbf{BRA} = \mathbf{SPSir}(\mathbf{BRA})$  we have that  $\mathbf{BRA} \models q$  iff  $\mathbf{Sir}(\mathbf{BRA}) \models q$  iff (by the above)  $\mathbf{Sir}(\mathbf{BRA}) \models e$  iff  $\mathbf{BRA} \models e$ .

Now (\*) and (\*\*) above give an interpretation of the quasi-equations valid in all semigroups into the equations valid in **BRA**. Since it is known that the former is undecidable, we also have that the latter,  $\mathbf{Eq}(\mathbf{BRA})$ , is undecidable. ■

The above method of proof for undecidability is also widely applicable in algebraic logic. The above proof e.g. is in [Crvenkovic and Madarász, 1992]. For more refined uses of this technique see e.g. [Maddux, 1980], [Németi, 1985a] (finite dimensional part) [Kurucz *et al.*, 1995], [Kurucz *et al.*, 1993], [Andréka, Givant and Németi, 1997, chapter II], [Kurucz, 1997]. For more on (*un*)decidability in algebraic logic we refer to the just quoted works together with [Jipsen, 1992], [Maddux, 1978], [Henkin, Monk and Tarski, 1985], [Németi, 1986], [Németi, 1987], [Németi, 1992], [Marx and Venema, 1997], [Mikulás, 1995], [Németi, 1991].

We turn to determining the logic ‘captured by’ **BRA**. We note that the connection with logic will be much more lucid in the case of cylindric (and polyadic) algebras of  $n$ -ary relations.

Let  $\mathcal{L}_{3,2}^\neq$  denote first order logic without equality and using only three variables  $x, y, z$ , with countably many *binary* relation symbols  $R_0, R_1, \dots$  (so, e.g., no ternary relation symbols are allowed), and the atomic formulas are  $R_i(u, v)$  with distinct variables  $u, v$  (so atomic formulas of the form  $R_i(u, u)$  are not allowed).

**THEOREM 8.**  $\mathcal{L}_{3,2}^\neq$  can be interpreted into  $\mathbf{Eq}(\mathbf{BRA})$ . I.e. there is a recursive function  $e$  mapping  $\mathcal{L}_{3,2}^\neq$  into the set of equations on the language of **BRA** such that for every  $\varphi \in \mathcal{L}_{3,2}^\neq$

$$\varphi \text{ is valid} \quad \text{iff} \quad \mathbf{BRA} \models e(\varphi).$$

<sup>26</sup>This is one of the basic facts of discriminator varieties. Assume that  $q$  is  $\tau_1 = \sigma_1 \wedge \dots \wedge \tau_n = \sigma_n \rightarrow \tau_0 = \sigma_0$ . Then  $c - (\tau_1 \oplus \sigma_1) \cap \dots \cap c - (\tau_n \oplus \sigma_n) \leq -(\tau_0 \oplus \sigma_0)$  can be chosen for  $e$ , where  $c$  denotes the switching term and  $\oplus$  denotes symmetric difference.

Theorem 8 will be a consequence of the following, stronger Theorem 9. We stated Theorem 8 because it states that  $\mathcal{L}_{3,2}^\neq$  can be interpreted into **Eq(BRA)**, thus **Eq(BRA)** is ‘at least as strong’ as  $\mathcal{L}_{3,2}^\neq$ . Set Theory can be interpreted in  $\mathcal{L}_{3,2}^\neq$ , this is proved in [Tarski and Givant, 1987, §4.6, pp.127–134]. Thus the logic captured by **BRA** is strong enough to serve as a vehicle for set theory, and hence for ordinary mathematics, as we mentioned at the beginning of this chapter.<sup>27</sup>

We can characterize the expressive power of **BRA** in terms of  $\mathcal{L}_{3,2}^\neq$ . This will be stated and proved as Theorem 9 below. We need some preparations for stating Theorem 9.

In the equational language of **BRA** let us use the variables  $v_i, i \in \omega$ , where  $\omega = \{0, 1, 2, \dots\}$  is the set of natural numbers. For any model  $\mathfrak{M} = \langle M, R_i^{\mathfrak{M}} \rangle_{i \in \omega}$  of  $\mathcal{L}_{3,2}^\neq$  let  $k_{\mathfrak{M}}$  denote the evaluation of the variables  $v_i, i \in \omega$  such that

$$k_{\mathfrak{M}}(v_i) = R_i^{\mathfrak{M}} \quad \text{for all } i \in \omega.$$

Recall that  $\mathfrak{R}(M \times M) = (\mathfrak{P}(M \times M), \circ, {}^{-1}) \in \mathbf{setBRA}$ .

If  $\mathfrak{A}$  is an algebra,  $k$  is an evaluation of the variables,  $\tau$  is a term, and  $e$  is an equation, then  $\mathfrak{A} \models e[k]$  denotes that the equation  $e$  is true in the algebra  $\mathfrak{A}$  under the evaluation  $k$  of the variables, and  $\tau^{\mathfrak{A}, k}$  denotes the element of  $A$  denoted by the term  $\tau$  when the variables are evaluated according to  $k$ .

Let  $u, v$  be distinct elements of  $\{x, y, z\}$ . Then  $\mathcal{L}_3^{uv}$  denotes the set of those elements of  $\mathcal{L}_{3,2}^\neq$  which contain only  $u, v$  as *free* variables. If  $\varphi \in \mathcal{L}_3^{xy}$  and  $\mathfrak{M}$  is a model, then  $\varphi^{\mathfrak{M}}$  denotes the following binary relation on  $M$ :

$$\varphi^{\mathfrak{M}} \stackrel{\text{def}}{=} \{(a, b) \in M \times M : \mathfrak{M} \models \varphi[a, b]\}.$$

The following Theorem 9 says<sup>28</sup> that, in a way, the *expressive power of BRA* is  $\mathcal{L}_3^{xy}$ . We included (i) for its simple content, and (ii) states a correspondence between meanings of formulas in  $\mathcal{L}_3^{xy}$  and denotation of terms in elements of **setBRA**. For more on the background ideas of this see Part II of the present Paper.

**THEOREM 9** (The expressive power of **Eq(BRA)**).

- (i) For any  $\varphi \in \mathcal{L}_{3,2}^\neq$  there is an equation  $e$  such that for all models  $\mathfrak{M}$  of  $\mathcal{L}_{3,2}^\neq$
- $$\mathfrak{M} \models \varphi \quad \text{iff} \quad \mathfrak{R}(M \times M) \models e[k_{\mathfrak{M}}].$$

<sup>27</sup>This also gives another proof for undecidability of **Eq(BRA)**, because Set Theory is undecidable.

<sup>28</sup>These statements and proofs are simplified versions of those in [Tarski and Givant, 1987]. Cf. also [Henkin, Monk and Tarski, 1985, §5.3, 4.3].

(ii) There are recursive functions  $t : \mathcal{L}_{3,2}^\neq \rightarrow \text{Terms}$  and  $f : \text{Terms} \rightarrow \mathcal{L}_{3,2}^\neq$  such that for any  $\varphi \in \mathcal{L}_3^{xy}$  and model  $\mathfrak{M}$

$$\varphi^{\mathfrak{M}} = t(\varphi)^{\langle \mathfrak{R}(M \times M), k_{\mathfrak{M}} \rangle}, \quad \text{and}$$

for any term  $\tau$ , set  $U$ , and evaluation  $k$ ,

$$\tau^{\langle \mathfrak{R}(U \times U), k \rangle} = f(\tau)^{\langle U, k(v_i)_{i \in \omega} \rangle}.$$

**Proof.** (i) follows from (ii), so it is enough to prove (ii).

The translation function  $f : \text{Terms} \rightarrow \mathcal{L}_{3,2}^\neq$  is not hard to give. Let  $u, v \in \{x, y, z\}$  be distinct, and let  $w$  be the third variable, i.e.  $\{u, v, w\} = \{x, y, z\}$ . We will simultaneously define the functions  $f_{uv} : \text{Terms} \rightarrow \mathcal{L}_{3,2}^\neq$  as follows:

$$\begin{aligned} f_{uv}(v_i) &\stackrel{\text{def}}{=} R_i(uv), \\ f_{uv}(\tau \cup \sigma) &\stackrel{\text{def}}{=} f(\tau) \vee f(\sigma), \quad f(-\tau) \stackrel{\text{def}}{=} \neg f(\tau), \\ f_{uv}(\tau \circ \sigma) &\stackrel{\text{def}}{=} \exists w (f_{uw}(\tau) \wedge f_{wv}(\sigma)), \\ f_{uv}(\tau^{-1}) &\stackrel{\text{def}}{=} f_{vu}(\tau). \end{aligned}$$

For the other direction, we want to define, by simultaneous recursion, a term  $\tau(\varphi, u, v)$  for all distinct variables  $u, v \in \{x, y, z\}$  and  $\varphi \in \mathcal{L}_3^{uv}$  such that for all models  $\mathfrak{M}$  we have

$$(*) \quad \{ \langle a, b \rangle \in M \times M : \mathfrak{M} \models \varphi(u/a, v/b) \} = \tau(\varphi, u, v)^{\langle \mathfrak{R}(M \times M), k_{\mathfrak{M}} \rangle}.$$

So let  $\varphi \in \mathcal{L}_3^{uv}$ .

*Case 1.* If  $\varphi$  is an atomic formula, then  $\varphi$  is  $R_i(uv)$  or  $R_i(vu)$  for some  $i \in \omega$  (by  $\varphi \in \mathcal{L}_3^{uv}$ ).

$$\tau(R_i(uv), u, v) \stackrel{\text{def}}{=} v_i, \quad \tau(R_i(vu), u, v) \stackrel{\text{def}}{=} v_i^{-1}.$$

*Case 2.* If  $\varphi$  is a disjunction of two formulas, say  $\varphi$  is  $\psi \vee \eta$ , then  $\psi, \eta \in \mathcal{L}_3^{uv}$ , and

$$\tau(\psi \vee \eta, u, v) \stackrel{\text{def}}{=} \tau(\psi, u, v) \cup \tau(\eta, u, v).$$

*Case 3.* If  $\varphi$  is a negation of another formula, then  $\varphi$  is  $\neg\psi$  for some  $\psi \in \mathcal{L}_3^{uv}$ , and we define

$$\tau(\neg\psi, u, v) \stackrel{\text{def}}{=} \neg\tau(\psi, u, v).$$

*Case 4.* If  $\varphi$  begins with  $\exists u$ , then  $\varphi$  is  $\exists u\psi$  for some  $\psi \in \mathcal{L}_3^{uv}$ , and then we define

$$\tau(\exists u\psi, u, v) \stackrel{\text{def}}{=} 1 \circ \tau(\psi, u, v).$$

Likewise we define

$$\tau(\exists v\psi, u, v) \stackrel{\text{def}}{=} \tau(\psi, u, v) \circ 1.$$

*Case 5.* Assume that  $\varphi$  begins with  $\exists w$ , i.e.  $\varphi$  is  $\exists w\psi$ . Then  $\psi \in \mathcal{L}_{3,2}^\neq$  can be arbitrary. It is easy to prove by induction that every element of  $\mathcal{L}_{3,2}^\neq$  is a Boolean combination of formulas in  $\mathcal{L}_3^{xy}$ ,  $\mathcal{L}_3^{xz}$  and  $\mathcal{L}_3^{yz}$ . Bring  $\psi$  into disjunctive normal form  $\psi_1 \vee \dots \vee \psi_n$  where each  $\psi_i$  is a conjunction of formulas with two free variables. Now  $\exists w\psi$  is equivalent to

$$(\exists w\psi_1) \vee \dots \vee (\exists w\psi_n),$$

so by Case 2 we may assume that  $\psi$  is of form  $\psi^{uv} \wedge \psi^{uw} \wedge \psi^{vw}$  where  $\psi^{uv} \in \mathcal{L}_3^{uv}$ , etc. Now  $\exists w\psi$  is equivalent to

$$\psi^{uv} \wedge \exists w(\psi^{uw} \wedge \psi^{vw}).$$

We now define

$$\tau(\exists w(\psi^{uw} \wedge \psi^{vw}), u, v) \stackrel{\text{def}}{=} \tau(\psi^{uw}, u, w) \circ \tau(\psi^{vw}, w, v).$$

It is not difficult to check that the so defined  $\tau(\varphi, u, v)$  satisfies our requirement (\*). ■

One can get very far in doing algebraic logic (for quantifier or predicate logics) via BRAs.<sup>29</sup> As we have seen, the natural logical counterpart of BRAs is classical first-order logic restricted to three individual variables and without equality. As shown in [Tarski and Givant, 1987, §5.3], this system is an adequate framework for building up set theory and hence metamathematics in it. One can illustrate most of the main results, ideas and problems of algebraic logic by using only BRAs.

We do not know how far BRAs can be simplified without losing this feature. In this connection, a natural candidate would be the class BSR of Boolean semigroups of relations defined as

$$\text{BSR} = \mathbf{SP} \{ \langle \mathfrak{P}(U \times U), \circ \rangle : U \text{ is a set} \}.$$

<sup>29</sup>If we want to investigate nonclassical quantifier logics, we can replace the Boolean reduct  $\mathfrak{B}$  of  $\mathfrak{A} = \langle \mathfrak{B}, \circ, -^1 \rangle \in \mathbf{BRA}$  with the algebras (e.g. Heyting algebras) corresponding to the propositional version of the nonclassical logic in question.

So we require only one extra-Boolean operation ‘ $\circ$ ’.

The question is, how far BSR could replace BRA as the simplest, ‘introductory’ example of Tarskian algebraic logic. We conjecture that the answer will be ‘very far’. BSR is a discriminator variety with a recursively enumerable but not decidable equational theory, and it is not finitely axiomatizable. Thus Theorems 2–7 remain true if BRA is replaced with BSR in them.<sup>30</sup> We conjecture that, following the lines of [Tarski and Givant, 1987, §5.3], set theory can be built up in BSR instead of BRA with basically the same positive properties (e.g. finitely many axioms) as the present version [Tarski and Givant, 1987] has.<sup>31</sup> It would be nice to know if this conjecture is true, and, more generally, to see a variant of algebraic logic elaborated on the basis of BSR. We do not know what natural fragment of first-order logic with three variables corresponds to BSR (if any). It certainly is difficult to simulate substitution of individual variables using only  $\circ$ . The converse operation,  $^{-1}$ , is the algebraic counterpart of substitution because, intuitively,  $R(v_0, v_1)^{-1} = R(v_1, v_0)$ . One can simulate quantification by  $\circ$ , and it is easily seen that  $\circ$  is stronger than quantification but without  $^{-1}$  it is not clear exactly how much stronger.<sup>32</sup> Curiously enough, these issues are better understood in the case of cylindric algebras to be discussed in Section 2.

If we want to algebraize *first-order logic with equality*, we have to add an extra constant  $\text{ld}$ , representing equality, to the operations. RRA denotes the class of algebras embeddable into direct products of algebras of the form

$$\langle \mathfrak{P}(U \times U), \circ, ^{-1}, \text{ld} \rangle$$

where  $\text{ld} = \text{ld} \upharpoonright U = \{ \langle u, u \rangle : u \in U \}$  is a constant of the expanded algebra. I.e.

$$\text{RRA} = \mathbf{SP} \{ \langle \mathfrak{P}(U \times U), \circ, ^{-1}, \text{ld} \rangle : U \text{ is a set} \}.$$

RRA abbreviates representable relation algebras. RRAs have been investigated more thoroughly than BRAs; actually, Theorems 2, 6 above were proved first for RRAs.

Let  $\mathcal{L}_{3,2}^=$  denote first-order logic with three individual variables  $x, y, z$ , with equality, and with infinitely many *binary* relation symbols. (Thus the

---

<sup>30</sup>The proofs of Theorems 2, 7 given here go through for BSR with the obvious modifications. Nonfinite axiomatizability of BSR will follow from the later Theorems 10, 11.

<sup>31</sup>Perhaps here [Németi, 1985], [Németi, 1986] can be useful, because an analogous task was carried through there. The last 12 lines of [Jónsson, 1982, p. 276], seem to be also useful here.

<sup>32</sup>For applications in propositional dynamic logic, BSR seems to be more relevant than BRA, because there converse (of programs or actions) is not an essential feature of the logic.

atomic formulas are  $R(uv)$ ,  $u = v$  for any variables  $u, v \in \{x, y, z\}$ , and the logical connectives are  $\vee, \neg, \exists x, \exists y, \exists z$ .)

**THEOREM 10** (Basic properties of RRA).

- (i) RRA is a nonfinitely axiomatizable discriminator variety with a recursively enumerable but undecidable equational theory.
- (ii) The logic captured by RRA is  $\mathcal{L}_{3,2}^-$ , i.e. there are recursive functions  $t : \mathcal{L}_{3,2}^- \rightarrow \text{Terms}$ , and  $f : \text{Terms} \rightarrow \mathcal{L}_{3,2}^-$  such that the ‘meanings’ of  $\varphi$  and  $t(\varphi)$  as well as those of  $\tau$  and  $f(\tau)$  coincide, i.e. for any model  $\mathfrak{M}$ ,  $\varphi \in \mathcal{L}_{3,2}^-$  with free variables  $x, y$ , term  $\tau$  and evaluation  $k$  of variables,

$$\varphi^{\mathfrak{M}} = t(\varphi)^{\langle \mathfrak{R}(M \times M), k_{\mathfrak{M}} \rangle} \quad \text{and} \quad \tau^{\langle \mathfrak{R}(U \times U), k \rangle} = f(\tau)^{\langle U, k(v_i)_{i \in \omega} \rangle}.$$

**Proof.** Obvious modifications of the proofs of Theorems 2, 7, 8 prove Theorem 10, except for nonfinite axiomatizability of RRA. For the proof of nonfinite axiomatizability of RRA see Remark 23. ■

The classes of algebras RRA, BRA, BSR have less operations in this order, they form a chain of subreduct classes. Note that  $\mathbf{Eq}(K)$  denotes the set of all equations in the language of  $K$  holding in  $K$ . Thus

$$\mathbf{Eq}(\text{BSR}) \subset \mathbf{Eq}(\text{BRA}) \subset \mathbf{Eq}(\text{RRA}).$$

The next theorem says that these classes are finitely axiomatizable over the bigger ones.

**THEOREM 11.** *Let  $E_0$  denote the following set of equations:*

$$\begin{aligned} (x \cup y)^{-1} &= x^{-1} \cup y^{-1}, & (x \circ y)^{-1} &= y^{-1} \circ x^{-1}, & x^{-1-1} &= x \\ x \circ -(x^{-1} \circ -y) &\leq y \\ x &\leq x \circ [-(y^{-1} \circ -y) \cap ((-y)^{-1} \circ y)^{-1}]. \end{aligned}$$

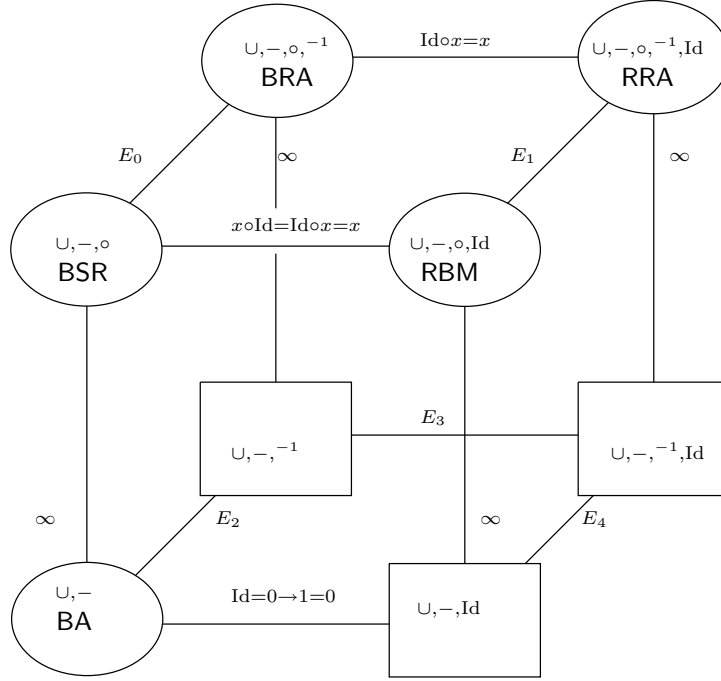
*Then  $\mathbf{Eq}(\text{BSR}) \cup E_0$  axiomatizes BRA, and  $\mathbf{Eq}(\text{BSR}) \cup E_0 \cup \{\text{Id} \circ x = x\}$  axiomatizes RRA.*

The *proof* can be found in [Andréka and Németi, 1993]. ■

Theorem 11 talks about interconnections between the operations of RRA. It says, in a way, that the sole cause of nonfinite axiomatizability of RRA is the operation  $\circ$ , it is so strong that the other operations,  $^{-1}$  and  $\text{Id}$ , are finitely axiomatizable with its help. This is in contrast with the case of

cylindric algebras of  $n$ -ary relations, where the strength of the operations are ‘evenly distributed’, see Figure 10.

The next figure, taken from [Andréka and Németi, 1993] describes completely the interconnections between the operations  $\circ, ^{-1}, \text{Id}$  (in the presence of the Boolean operations). On Figure 4, all classes represented by the nodes are varieties, except the ones inside a box (those are only quasi-varieties), and the classes inside a circle are not finitely axiomatizable, except BA.<sup>33</sup>



$E_0$  from Thm. 1.11

$$E_1 = \{x^{-1} \circ -x \leq -\text{Id}, -(x^{-1}) = (-x)^{-1}\}$$

$$E_2 = \{x^{-1} = -x \rightarrow 1 = 0, (x \cup y)^{-1} = x^{-1} \cup y^{-1}, x^{-1-1} = x\}$$

$$E_3 = \{\text{Id} = 0 \rightarrow 1 = 0, (x \cap \text{Id})^{-1} = x \cap \text{Id}\}$$

$$E_4 = \{(x \cup y)^{-1} = x^{-1} \cup y^{-1}, x^{-1-1} = x, (x \cap \text{Id})^{-1} = x \cap \text{Id}\}$$

Figure 4.

<sup>33</sup>In Pratt [1990], the class RBM of representable Boolean monoids is obtained from our BSRs by adding  $\text{Id}$  as an extra distinguished constant. So the extra-Boolean operations of the RBMs are  $\circ, \text{Id}$ , and thus BSRs are the  $\text{Id}$ -free subreducts of RBMs. All the results mentioned above for BSRs carry over to RBMs; e.g. RBM is a discriminator variety, hence the simple RBMs form a universally axiomatizable class, Theorems 2, 6 above apply to RBM. RRA, BRA, BSR, RBM, BA all occur as nodes on Figure 4.



### More on the equational theories of RRA, BRA and BSR:

Theorem 2 says that there is a set  $E$  of equations which defines BRA. Let  $E$  be an arbitrary set of equations defining BRA. What do we know about  $E$ ? Theorem 6 says that  $E$  is not finite, and Theorem 7 says that  $E$  can be chosen to be recursively enumerable. By using the fact that BRA is a discriminator variety and that  $\mathbf{Eq}(\mathbf{BRA})$  is recursively enumerable, and by using an argument of W. Craig, one can show that  $E$  can be chosen to be decidable<sup>34</sup>, i.e. there is a decidable set  $E$  defining BRA. On the other hand, we know that  $E$  has to be complex in the following sense: to any number  $k$ ,  $E$  must contain an equation that uses more than  $k$  variables and all of the operation symbols  $\cup, -, \circ$ . There is an  $E$  such that  $^{-1}$  occurs only in finitely many members of  $E$ , by Theorem 11. The analogous statements are true for BSR, RRA.<sup>35</sup>

Concrete decidable sets  $E$  defining RRA are known in the literature, cf. e.g. [Monk, 1969]. Lyndon [1956] outlines another recipe for obtaining a different such  $E$ . Hirsch–Hodkinson [1997] also contains such a set  $E$ . Some of these work for BSR, BRA. However, the structures of these  $E$ s are rather involved.<sup>36</sup> In this connection, we note that the following is still one of the most important open problems of algebraic logic:

PROBLEM 12. Find *simple*, mathematically transparent, decidable sets  $E$  of equations axiomatizing BSR, BRA, RRA. (A solution for this problem has to be considerably simpler than, or at least markedly different from the  $E$ s discussed above.)

Equational axiom systems for algebras of relations like for RRA, BRA, BSR are interesting not only because of purely aesthetical reasons, but also because such an axiom system gives an inference system for the corresponding logic. About this logical connections see e.g. Theorems 45, 46, 50 in Part II.

Since the classes RRA, BRA, BSR are not finitely axiomatizable, finitely axiomatizable approximations, or ‘computational cores’ are used for them. For BRA we can take ARA as such an approximation. For RRA, the variety RA of *relation algebras*, defined by Tarski, is used in the literature as such

<sup>34</sup>The idea is as follows. Let  $E$  be recursively enumerable, say  $E = \{e(1), e(2), \dots\}$  for a recursive function  $e$ . For each number  $n$ , let  $\eta(n)$  denote the conjunction of  $n$  copies of  $e(n)$ . Since BRA is a discriminator variety, there is an equation  $\varepsilon(\eta(n))$  which is equivalent to  $\eta(n)$  in  $\mathbf{SirBRA}$ . Moreover, from  $\varepsilon(\eta(n))$  we can compute back  $\eta(n)$ , see an earlier footnote. Then  $E' \stackrel{\text{def}}{=} \{\varepsilon(\eta(1)), \varepsilon(\eta(2)), \dots\}$  is equivalent to  $E$  and  $E'$  is decidable. The decision procedure for  $E'$  is as follows: Take any equation  $g$ . Decide whether  $g$  is  $\varepsilon(f)$  for some  $f$  or no, and if yes, compute the  $f$ . If we get an  $f$ , check whether  $f$  is the conjunction of some, say  $n$ , copies of an equation  $h$ . If yes, compute  $e(n)$  and check whether  $h$  is  $e(n)$ . If yes,  $g$  is in  $E'$ , otherwise not.

<sup>35</sup>The need of infinitely many variables in any axiom system for RRA was proved in [Jónsson, 1991], the need of all the operation symbols  $\cup, -, \circ$  in addition is proved in [Andréka, 1994]. By Theorem 11 then the same hold for BRA, BSR.

<sup>36</sup>Cf. [Henkin, Monk and Tarski, 1985, pp. 112–119], for an overview.

an approximation. We get the definition of RA from the definition of ARA by replacing (4) with one stronger equation (5), and by adding the equation  $\text{ld} \circ x = x$ .

DEFINITION 13 (RA, an abstract approximation of RRA). RA is defined to be the class of all algebras of the similarity type of RRAs which satisfy the equations (1)–(3) from the definition of ARA, together with (5), (6) below.

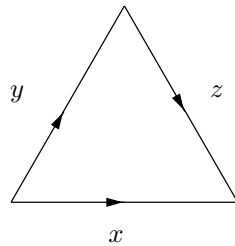
$$(5) \quad x^{-1} \circ [-(x \circ y)] \leq -y.$$

$$(6) \quad \text{ld} \circ x = x.$$

Equation (5) is equivalent, in the presence of the other RA-axioms<sup>37</sup> with the following so called *triangle-rule* (5')

$$(5') \quad x \cap (y \circ z) = 0 \quad \text{iff} \quad y \cap (x \circ z^{-1}) = 0 \quad \text{iff} \quad z \cap (y^{-1} \circ x) = 0.$$

Intuitively, (5') says that the three ways of telling that no triangle



exists, are equivalent.<sup>38</sup> Thus, a relation algebra is a Boolean algebra together with an involuted monoid sharing the same universe, and the interconnection between the two structures is that they form a normal BAO and the triangle rule (5') holds.

Equation (6) says that  $\text{ld}$  is the neutral element of the semigroup operation ' $\circ$ '. We note that in algebraic logic this translates to the so called *Leibniz law* of equality in logic which says that equals cannot be distinguished.<sup>39</sup>

<sup>37</sup>We note that  $0 \circ x = 0, 0^{-1} = 0$  are usually omitted from the axiomatization of RA, because they follow from the rest of the axioms.

<sup>38</sup>In a more algebraic language, (5') says that the maps  $x \mapsto a \circ x$  and  $y \mapsto a^{-1} \circ y$  are *conjugates* of each other, and likewise the maps  $x \mapsto x \circ a$  and  $y \mapsto y \circ a^{-1}$  are conjugates. We recall from [Jónsson and Tarski, 1951] that in a BAO the functions  $f, g$  are conjugates of each other means that  $x \cap f(y) = 0$  iff  $y \cap g(x) = 0$  for all  $x, y$ .

<sup>39</sup>For more on this see [Blok and Pigozzi, 1989, p.10].

RA is a very strong computational core for RRA, almost all natural equations about RRA hold also in RA.<sup>40</sup> An RA which is not in RRA is called a *nonrepresentable* RA. Equations holding in RRA and not in RA can be obtained from each finite nonrepresentable RA by using that RA is a discriminator variety of BAOs, as follows. Let  $\mathfrak{A} \in \text{RA} - \text{RRA}$  be finite. Then  $\mathfrak{A}$  cannot be embedded in any RRA, and this can be expressed with a universal formula because  $\mathfrak{A}$  is finite. Now using the switching function, this universal formula can be coded as an equation  $e$ . Then  $e$  does not hold in  $\mathfrak{A} \in \text{RA}$ , while it holds in RRA. Many finite nonrepresentable RAs are known in the literature. The smallest such has 16 elements.<sup>41</sup> so-called Lyndon algebras. A finite *Lyndon algebra* is a finite RA such that  $\text{Id}$  is an atom,  $a^{-1} = a$ ,  $a \circ a = a \cup \text{Id}$ , and  $a \circ b = 1 - (a \cup b)$  hold for all other distinct atoms  $a, b$ . Infinitely many of the finite Lyndon algebras are nonrepresentable (and infinitely many are representable). Another way of getting finite nonrepresentable RAs is to ‘distort’ a representable one. There are some known methods, like *splitting* and *dilating*<sup>42</sup> with which we can obtain nonrepresentable RAs from representable ones. Nonrepresentable RAs are almost as important as representable ones.<sup>43</sup>

Some special, interesting classes of RRAs turn out to be finitely axiomatizable, below we list two such classes. These finite axiomatizations give (non-standard) finitary inference systems for  $\mathcal{L}_{3,2}^-$ , cf. Mikulás [1996; 1995]. The elegant, purely algebraic proofs for the items in the next theorem are examples for significant applications of algebra to logic, via connections between algebra and logic indicated in Part II of this Paper.

**THEOREM 14.** *Let  $\varphi$  and  $\psi$  denote the following formulas, respectively.*

$$\begin{aligned} & \exists xy(x^{-1} \circ x \leq \text{Id} \wedge y^{-1} \circ y \leq \text{Id} \wedge x^{-1} \circ y = 1) \\ & \forall x \exists y(x \neq 0 \rightarrow [0 \neq y \wedge y \leq x \wedge y^{-1} \circ y \leq \text{Id}]). \end{aligned}$$

*Then  $\text{RRA} \cap \text{Mod}\Sigma = \text{RA} \cap \text{Mod}\Sigma$  for  $\Sigma = \{\varphi\}$  and  $\Sigma = \{\psi\}$ .*

For the *proofs* see [Maddux, 1978b] and [Tarski and Givant, 1987]. ■

An RA in which  $\varphi$  is true is called a *quasi-projective* RA, or a **QRA**, and an RA in which  $\psi$  is true is called a *functionally dense* RA.<sup>44</sup> We can look

<sup>40</sup>In other words, only complicated equations can distinguish RRA and RA.

<sup>41</sup>This was found by [McKenzie, 1966].

<sup>42</sup>For splitting in RA see [Andréka, Monk and Néméti, 1991], for dilation in RA see [Néméti, 1986], [Néméti and Simon, 1997], Simon [Simon, 1997].

<sup>43</sup>E.g. one proof of nonfinite axiomatizability of RRA goes by finding infinitely many nonrepresentable RAs whose ultraproduct is representable. Investigating the structure of possible axiom systems for RRA often boils down to finding suitable nonrepresentable RAs.

<sup>44</sup>The fact that any QRA is representable is a theorem of Tarski, an elegant algebraic proof was given in [Maddux, 1978b]. A different, illuminating proof is given in [Simon,

at Theorem 14 in two ways: on one hand it says that the class of quasi-projective RRAs is finitely axiomatizable (while RRA is not), and on the other hand it says that quasi-projective RAs are representable (while RAs in general are not). (And the same for functionally dense RAs.)

## 2 ALGEBRAS OF RELATIONS IN GENERAL

By this point we might have developed some vague picture of how algebras of binary relations are introduced, investigated etc. One might even sense that they give rise to a smooth, elegant, exciting and powerful theory. However, our original intention was to develop algebras of relations in general, which should surely incorporate not only binary but also ternary, and in general  $n$ -ary relations.

Let us see how to generalize our RRAs and BRAs to relations of higher ranks. Let us first fix  $n$  to be a *finite* ordinal. As we said, we would like the new algebras to be expansions of RRAs (and BRAs). However, defining composition of  $n$ -ary relations for  $n > 2$  is complicated.<sup>45</sup> Therefore the following sounds like a more attractive idea: We single out the simplest basic operations on  $n$ -ary relations, and hope that composition will be derivable as a term-function from these. Let us see how we could generalize our generic or full RRAs  $\langle \mathfrak{P}(U \times U), \circ, ^{-1}, \text{ld} \rangle$  to relations of rank  $n$ . The obvious part is that these algebras will begin with  $\langle \mathfrak{P}(U \times U \times \dots \times U), \text{ld}, \dots \rangle$ , where

$$\text{ld} = \{ \langle u, u, \dots, u \rangle : u \in U \}$$

is the  $n$ -ary identity relation. Again,  $\text{ld}$  is a constant, just as it was in the RRA case. Let  ${}^nU$  denote  $U \times U \times \dots \times U$ , e.g.  ${}^3U = U \times U \times U$ . The new operations (besides the Boolean ones and  $\text{ld}$ ) we will need are the algebraic counterparts of quantification  $\exists v_i$ , for  $i < n$ . So, we want an operation that sends the relation defined by  $R(v_0, v_1)$  to the one defined by  $\exists v_0 R(v_0, v_1)$ , and similarly for  $\exists v_1$ . For  $R \subseteq U \times U$  let  $\text{Dom}(R)$  and  $\text{Rng}(R)$  denote the usual domain and range of  $R$ . For  $n = 2$  we define

$$c_0(R) = U \times \text{Rng}(R) \quad \text{and} \quad c_1(R) = \text{Dom}(R) \times U.$$

Now

$$\langle \mathfrak{P}(U \times U), c_0, c_1, \text{ld} \rangle$$

is the *full cylindric set algebra* of binary relations over  $U$ , for short the full  $\mathbf{Cs}_2$ .

---

1996]. For logical applications of this area see [Tarski and Givant, 1987]. The proof that every functionally dense RA is representable is in [Maddux, 1978b]. See also [Andréka *et al.*, 1998].

<sup>45</sup>Composition for  $n$ -ary relations is studied in [Marx, Némethi and Sain, 1996] and [Marx, 1995]. The definition is  $\circ(R_1, \dots, R_n) = \{ \langle a_1, \dots, a_n \rangle : \exists x (\langle a_1, \dots, a_{n-1}, x \rangle \in R_1 \ \& \ \langle a_1, \dots, a_{n-2}, x, a_n \rangle \in R_2 \ \& \ \dots \ \& \ \langle x, a_2, \dots, a_n \rangle \in R_n) \}$ .

Before turning seriously to  $n$ -ary relations, we need the following:

CONVENTION 15. Throughout we will pretend that Cartesian products and Cartesian powers are associative such that  ${}^nU \times {}^mU = {}^{n+m}U$ , and if e.g.  $R \subseteq {}^3U$  then  ${}^2U \times R \subseteq {}^5U \supseteq R \times {}^2U$ .

The full  $\mathbf{Cs}_n$ , i.e. the *full cylindric set algebra* of  $n$ -ary relations, is the natural generalization of  $\mathbf{Cs}_2$  as follows. Let  $R \subseteq {}^nU$ . If  $\mathit{Rng}(R) = \{\langle b_1 \dots b_{n-1} \rangle : \langle b_0 b_1 \dots b_{n-1} \rangle \in R \text{ for some } b_0\}$ , then  $c_0(R) = U \times \mathit{Rng}(R)$  considered as a set of  $n$ -tuples. Similarly, let  $\mathit{Dom}(R) = \{\langle b_0 \dots b_{n-2} \rangle : \langle b_0 \dots b_{n-2} b_{n-1} \rangle \in R \text{ for some } b_{n-1}\}$ , and let  $c_{n-1}(R) = \mathit{Dom}(R) \times U$ . Generalizing this to  $c_i$  with  $i < n$  arbitrary, we obtain

$$c_i(R) = \{\langle b_0, \dots, b_{i-1}, a, b_{i+1}, \dots, b_{n-1} \rangle : \langle b_0, \dots, b_{n-1} \rangle \in R \text{ and } a \in U\}.$$

$c_i$  is one of the most natural operations on relations. It simply forgets the  $i$ -th argument of the relation, or in other words, deletes the  $i$ -th column. However, since deleting the  $i$ -th column would leave us with an  $(n-1)$ -ary relation,  $\mathit{Dom}(R)$  if  $i = n-1$ , we replace the  $i$ -th column with a dummy column i.e. in the  $i = n-1$  case we represent  $\mathit{Dom}(R)$  with the ‘pseudo  $n$ -ary relation’  $\mathit{Dom}(R) \times U$ . The ‘real rank’ of an  $R \subseteq {}^nU$  is always easy to recover, namely it is  $\Delta(R) = \{i < n : c_i(R) \neq R\}$ . So  $c_i$  is the natural operation of removing  $i$  from the (real) rank of a relation.

For example,  $c_{\text{father}}$  when applied to the ‘father, mother, child’ relation gives back the ‘mother, child’ relation coded as ‘anybody, mother, child’ (in which the anybody argument carries no information i.e. is dummy). By a *full  $\mathbf{Cs}_n$*  we understand an algebra

$$\mathfrak{Rel}_n(U) \stackrel{\text{def}}{=} \langle \mathfrak{P}({}^nU), c_0, \dots, c_{n-1}, \text{ld} \rangle$$

for some set  $U$ . By a  $\mathbf{Cs}_n$  we understand a subalgebra of a full  $\mathbf{Cs}_n$  with nonempty<sup>46</sup> base set  $U$  i.e.

$$\mathbf{Cs}_n \stackrel{\text{def}}{=} \mathbf{S}\{\mathfrak{Rel}_n(U) : U \text{ is a nonempty set}\}.$$

By a *representable cylindric algebra* of  $n$ -ary relations, (an  $\mathbf{RCA}_n$ ) we understand a subalgebra of a direct product of full  $\mathbf{Cs}_n$ s (up to isomorphism), formally:

$$\mathbf{RCA}_n = \mathbf{SP}\{\langle \mathfrak{P}({}^nU), c_0, \dots, c_{n-1}, \text{ld} \rangle : U \text{ is a set}\}.$$

Note that  $\mathbf{RCA}_n = \mathbf{SP}\{\mathfrak{Rel}_n(U) : U \text{ is a set}\} = \mathbf{SP}(\text{full } \mathbf{Cs}_n) = \mathbf{SP}\mathbf{Cs}_n$ . By the same argument as in the case of BRAs, every  $\mathbf{RCA}_n$  is directly representable as an algebra of  $n$ -ary relations (with the greatest relation a disjoint

<sup>46</sup>Excluding the empty base set here is not essential, it serves easier applicability in the second part of this Paper, in Section 7.

union of Cartesian spaces).  $\text{RCA}_n$  is one of the ‘leading candidates’ for being the natural algebra of  $n$ -ary relations.

The abstract algebraic picture is simple: an  $\text{RCA}_n$  is a BA together with  $n$  closure operations and an extra constant. Accordingly, an (abstract) *cylindric algebra of dimension  $n$* , a  $\text{CA}_n$ , is defined as a normal BAO with  $n$  self-conjugated and commuting closure operations, and with a constant satisfying two equations. In more detail:

**DEFINITION 16** ( $\text{CA}_n$ , an abstract approximation of  $\text{RCA}_n$ ).  $\text{CA}_n$  is defined to be the class of all algebras of the similarity type of  $\mathfrak{Rel}_n(U)$  which satisfy the following equations for all  $i, j < n$ .

(1) The axioms for normal BAO, i.e.

the Boolean axioms,

$$c_i 0 = 0, \quad c_i(x \cup y) = c_i(x) \cup c_i(y).$$

(2) Axioms expressing that  $c_i$ s are self-conjugated commuting closure operations, i.e.

(i)  $x \leq c_i x = c_i c_i x$ ,

(ii)  $y \cap c_i x = 0 \quad \text{iff} \quad c_i y \cap x = 0$ ,

(iii)  $c_i c_j x = c_j c_i x$ .

Because of the above axioms, the notation  $c_{(\Gamma)}x = c_{i_1} \dots c_{i_k}x$  where  $\Gamma = \{i_1, \dots, i_k\}$  makes sense. We will use that notation from now on. We will also use the convention<sup>47</sup> that  $n = \{0, 1, \dots, n-1\}$ .

(3) The constant  $\text{ld}$  has domain 1 and satisfies the ‘Leibniz-law’, i.e.

(i)  $c_{(n \setminus \{i\})} \text{ld} = 1$ ,

(ii)  $c_{(n \setminus \{i, j\})} \text{ld} \cap c_i x = x$  whenever  $x \leq c_{(n \setminus \{i, j\})} \text{ld}$  and  $i \neq j$ .

An equivalent form of saying that  $c_i$  is self-conjugated<sup>48</sup> is to say that the complement of a closed element is closed. Thus, in the above definition, (2)(ii) can be replaced with

$$c_i - c_i x = -c_i x.$$

<sup>47</sup>For more on this see e.g. [Henkin, Monk and Tarski, 1971, Part I, pp. 31–32].

<sup>48</sup>I.e. that it is the conjugate of itself, cf. the footnote to the triangle-rule (5') in Section 1.

Note that (2)(ii) is an analogue of the triangle rule (5') in the definition of RA. (3)(ii) expresses that the closure operator  $c_i$  is 'discrete', or is the identity, when relativized to

$$\text{ld}_{ij} \stackrel{\text{def}}{=} c_{(n \setminus \{i,j\})} \text{ld},$$

$i \neq j$ . Both (2)(ii) and (3)(ii) have equivalent equational forms, e.g. an equivalent form of (3)(ii) is

$$\text{ld}_{ij} \cap c_i(\text{ld}_{ij} \cap x) = \text{ld}_{ij} \cap x \quad \text{when } i \neq j$$

and an equivalent (together with the other axioms) form of (2)(ii) is

$$c_i(x \cap c_i y) = c_i x \cap c_i y.$$

**Connection with geometry:** The names in cylindric algebra theory come from connection with geometry. Namely, an  $n$ -ary relation is a set of  $n$ -tuples, while an  $n$ -tuple is a point of the  $n$ -dimensional space. E.g.  $\langle a, b \rangle$  is a point in the 2-dimensional space with coordinates  $a$  and  $b$ , while  $\langle a, b, c \rangle$  is a point in the 3-dimensional space with coordinates  $a, b$  and  $c$ . Thus a binary relation is a subset of the 2-dimensional space, while an  $n$ -ary relation is a subset of the  $n$ -dimensional space. Hence the name 'cylindric algebra of dimension  $n$ '.

If  $R$  is a subset of the  $n$ -dimensional space, then  $c_i(R)$  is the *cylinder* above  $R$  parallel to the  $i$ -th axis, and  $\text{ld}$  is the main *diagonal*. Hence the name 'cylindric algebra'. The operations  $c_i$  and  $\text{ld}$  are called '*cylindrifications*' and '*diagonals*' in CA-theory, and  $\text{ld}_{ij}$  is usually denoted by  $d_{ij}$  (for diagonal). Because of these geometrical meanings, also the operations of  $\text{RCA}_n$  are easy to draw. This is illustrated on Figure 5, see also Figure 7.

How can we draw the operations of RRA? Converse is easy to draw: the converse of  $R$  is the mirror image of  $R$  w.r.t. the diagonal. However, relation composition of two relations  $R, S$  is not so easy to draw. See Figure 6.

Thus cylindric algebras (CAs) are simpler than relation algebras RAs in two ways: CAs have only unary operations  $c_i$ , while the central operation of RA is the binary composition operation  $\circ$ ; and secondly, cylindrifications are easy to draw, while composition is not so easy to draw. There are further connections with geometry, e.g. via projective planes. We do not discuss these herein, but cf. [Monk, 1974; Givant, 1999; Némethi and Sági, 1997], and [Andréka, Givant and Némethi, 1997, Chapter II].

**Connection between CAs and RAs:** A natural question comes up: can these 'simple' CAs recapture the power of RAs? This was a requirement we expected to meet, namely we expected that the theory of  $n$ -ary relations should be an extension of that of binary relations. The answer is that  $\text{RCA}_n$

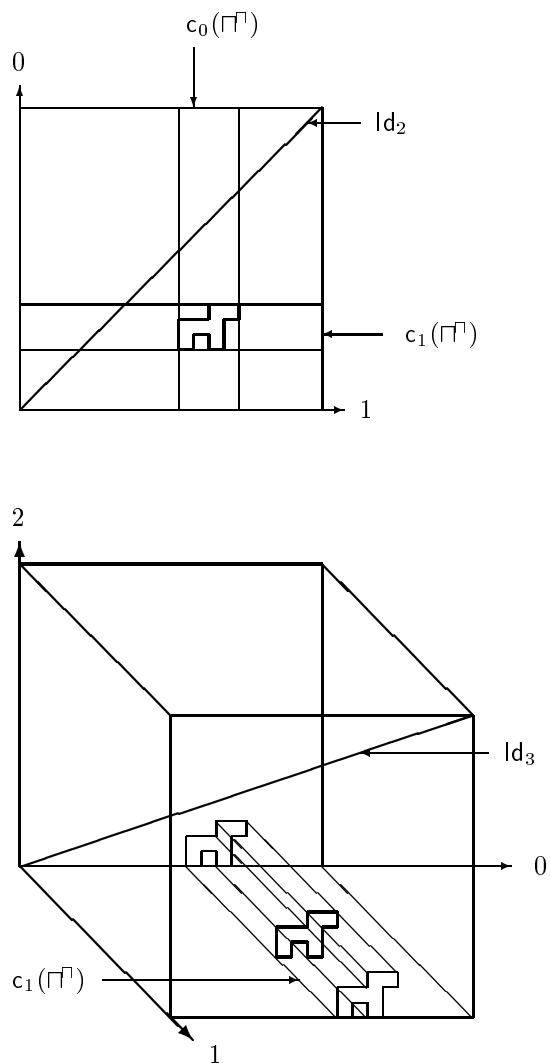
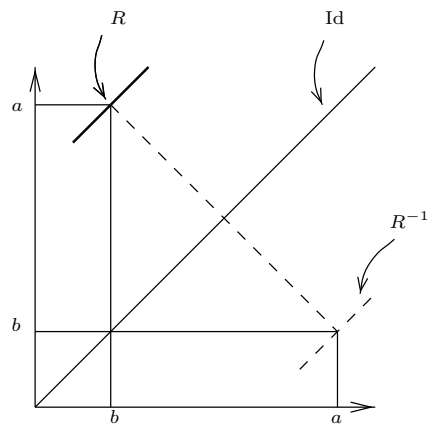
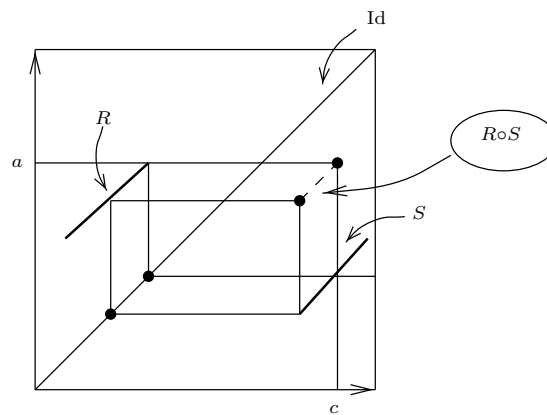


Figure 5.





converse is mirroring



composition is not so easy to draw

Figure 6.

with  $n \geq 3$  is strong enough to recapture RRA, while  $\text{RCA}_2$  is not strong enough. In more detail:

Mirroring cannot be expressed by the diagonal and the cylindrifications in the plane (i.e. in 2-dimensional space), but it can be expressed if we can move out to 3-dimensional space, see Figure 7.

Namely, by letting

$$s_j^i(x) \stackrel{\text{def}}{=} c_i(\text{Id}_{ij} \cap x) \quad \text{and} \quad P = U \times U \times \{u\}$$

for some fixed  $u \in U$ , we have

$$R^{-1} = P \cap s_1^2 s_0^1 s_2^0 c_2 R$$

for  $R \subseteq P$ . Here, we identified the binary  $R \subseteq U \times U$  with the ternary  $R \times \{u\}$ , and similarly for  $R^{-1}$ . Composition also can be expressed:

$$R \circ S = P \cap c_2(s_2^1 c_2 R \cap s_2^0 c_2 S).$$

A more natural approach is based on identifying a binary relation  $R \subseteq U \times U$  with the ternary relation

$$\mathbf{Dr}(R) \stackrel{\text{def}}{=} R \times U;$$

we call  $\mathbf{Dr}(R)$  the dummy representation of  $R$  as a ternary relation. Then  $\mathbf{Dr} : \mathcal{P}(U \times U) \rightarrow \mathcal{P}(U \times U \times U)$ ; and

$$\mathbf{Dr}(R^{-1}) = s_1^2 s_0^1 s_2^0 \mathbf{Dr}(R),$$

$$\mathbf{Dr}(R \circ S) = c_2(s_2^1 \mathbf{Dr}(R) \cap s_2^0 \mathbf{Dr}(S)).$$

So in a sense, RRAs form a kind of a reduct of  $\text{RCA}_n$ s for  $n \geq 3$ . In more detail: Let  $\mathfrak{A} \in \text{CA}_n$ ,  $n \geq 3$ . The *relation-algebra reduct*  $\mathfrak{Ra}\mathfrak{A}$  of  $\mathfrak{A}$  is defined as

$$\mathfrak{Ra}\mathfrak{A} \stackrel{\text{def}}{=} \langle \text{Ra}\mathfrak{A}, \cup^{\mathfrak{A}}, -^{\mathfrak{A}}, \circ, {}^{-1}, \text{Id}_{01}^{\mathfrak{A}} \rangle,$$

where

$$\text{Ra}\mathfrak{A} = \{a \in A : c_j^{\mathfrak{A}} a = a \text{ for all } 2 \leq j < n\}$$

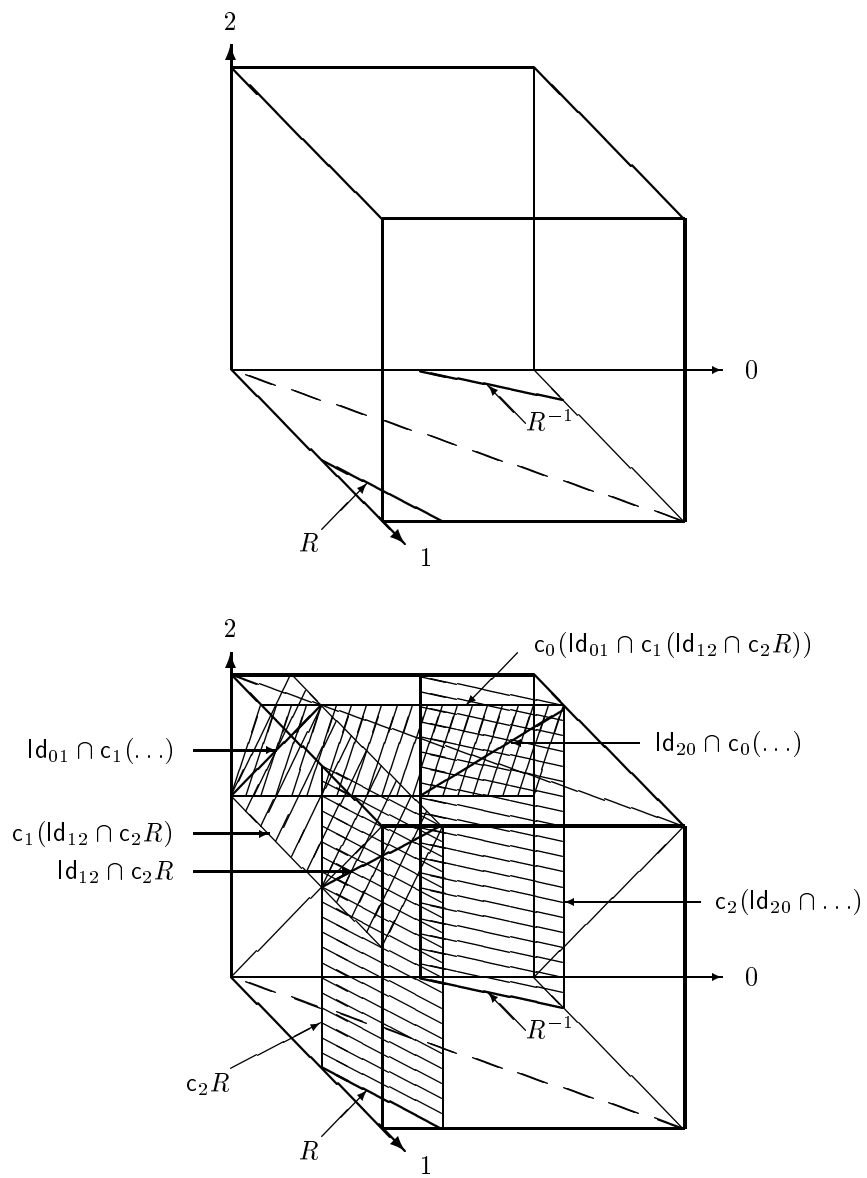
and if  $a, b \in \text{Ra}\mathfrak{A}$ , then

$$a^{-1} \stackrel{\text{def}}{=} s_1^2 s_0^1 s_2^0 a,$$

$$a \circ b \stackrel{\text{def}}{=} c_2(s_2^1 a \cap s_2^0 b).$$

Now,  $\mathbf{Dr} : \langle \mathfrak{P}(U \times U), \circ, {}^{-1}, \text{Id} \rangle \mapsto \mathfrak{Ra}\mathfrak{Rel}_n(U)$  is an isomorphism for  $n \geq 3$ . We define

$$\mathbf{RaCA}_n \stackrel{\text{def}}{=} \{\mathfrak{Ra}\mathfrak{A} : \mathfrak{A} \in \text{CA}_n\} \quad \text{for } 3 \leq n.$$



Mirroring in 3-dimensional space  
can be expressed by  $c_i, Id_{ij}, \cap$ .

Figure 7.

For  $K \subseteq CA_n$ ,  $\mathbf{Ra}K$  is defined similarly.

Then  $RRA = \mathbf{SRa}RCA_n$  for all  $n > 2$ . The classes  $\mathbf{SRa}CA_n$  ( $n > 3$ ) form a chain between  $RA$  and  $RRA$ , providing a ‘dimension-theory’ for<sup>49</sup>  $RA$ . In more detail,  $\mathbf{SRa}CA_3 \supseteq RA$  [Monk, 1961],  $\mathbf{SRa}CA_4 = RA$  [Maddux, 1978],  $RA \supset \mathbf{SRa}CA_5 \supseteq \dots \supseteq RRA$  [Monk, 1961], and<sup>50</sup>

$$RRA = \bigcap \{ \mathbf{SRa}CA_n : 3 \leq n < \omega \} = \mathbf{SRa}CA_\omega.$$

Investigating the connection between  $RRA$  ( $RA$ ) and  $RCA$  ( $CA$ ) is an interesting subject. Some of the references are [Monk, 1961; Maddux, 1978; Maddux, 1989; Henkin, Monk and Tarski, 1985; Németi and Simon, 1997; Simon, 1996; Simon, 1997; Hirsch and Hodkinson, 1997a]. Recent developments in the  $RA - CA$  (– polyadic algebras) connection are reported in [Simon, 1997; Németi and Simon, 1997].

Thus the answer is that  $RCA_n$ s,  $n > 2$ , do recapture the power of  $RRA$ s. (On the other hand,  $RCA_2$ s do not.<sup>51</sup>)

**Connection with logic:** Cylindric algebras have a very close and rich connection with logic. This connection is partly described in [Henkin, Monk and Tarski, 1985, §4.3], and in Examples 6, 8, 9 in Section 7 herein.

Summing up this connection very briefly:  $RCA$ -theory corresponds to model theory of first-order like languages (or quantifier logics), while abstract  $CA$ -theory corresponds to their proof theory. Individual  $CA$ s correspond to theories in such logics, homomorphisms between  $CA$ s correspond to interpretations between theories, while isomorphism of  $CA$ s corresponds to definitional equivalence of models and/or theories.  $CA$ -theoretic terms and equations correspond to first-order formula schemas, an equation  $e$  is valid in  $RCA$  if its corresponding formula schema is valid, an equational derivation of  $e$  corresponds to a proof of the formula (schema) corresponding to  $e$ . More on this is written in Section 7, Examples 6,8,9. Here, first-order like languages encompass finite-variable fragments of first-order language (FOL for short), usual FOL, FOL with infinitary relation symbols but with finitary logical connectives, FOL considered as a propositional multi-modal logic, FOL with several modified semantics etc.

Most of the above is discussed in [Henkin, Monk and Tarski, 1985, §4.3], especially when taken together with [van Benthem, 1996]. Some other references illustrating the rich connection of  $CA$ s with logic are the following. In [Monk, 1993] the connection with FOL is treated. In [Németi, 1987] and in [Rybakov, 1997] valid formula-schemas, in [Németi, 1990] model

<sup>49</sup>As later, in Theorem 28(ii), we will see, this is analogous to the chain  $\mathbf{SNr}_n CA_{n+k}$  ( $k \geq 0$ ) between  $CA_n$  and  $RCA_n$ .

<sup>50</sup>For the definition of  $CA_\omega$  see Def. 26.

<sup>51</sup>For example,  $\mathbf{Eq}RRA$  is undecidable, while  $\mathbf{Eq}RCA_2$  is decidable, see Theorem 10(i), Theorem 17(iii).

theory of FOL with infinitary relation symbols, in [Németi, 1996] FOL with generalized semantics, in [Amer, 1993], [Sayed Ahmed, 1997] algebras of sentences, in [Serény, 1985] and [Biró and Shelah, 1988] model theoretic notions like saturated, universal, atomic models are investigated with the help of CAs, respectively. [Andréka, van Benthem and Németi, 1996; van Benthem, 1996; Marx and Venema, 1997] connect CAs with modal logic, [Sain, 1995; Sain and Gyuris, 1994] use CAs for searching for a FOL with nicer behaviour. Further references on this line are e.g. [van Benthem, 1997; Venema, 1995a].

The connection of CAs with logic also sheds light on the above ways of expressing composition and converse of binary relations in CA. Namely,  $\text{RCA}_n$  is the algebraic counterpart of first-order logic with  $n$  variables (see Part II, example 6 in Section 7), and in particular  $n$ -variable first-order formulas and terms in the language of  $\text{RCA}_n$  are in strong correspondence with each other (see Corollary 46 in Part II). The RCA-terms in the definition of an RA-reduct are just the transcripts of the 3-variable formulas defining composition and conversion of binary relations. (On the intuitive meaning of the terms  $s_j^i$  see the remark after Example 7 in Section 7.)

At this point we can state the counterparts of Theorems 2–10.<sup>52</sup>

**THEOREM 17** (Basic properties of  $\text{RCA}_n$ ). *Let  $n$  be finite.*

- (i)  $\text{RCA}_n$  is a discriminator variety, with a recursively enumerable equational theory.
- (ii)  $\text{RCA}_n$  is not axiomatizable with a finite set of equations and its equational theory is undecidable if  $n > 2$ .
- (iii)  $\text{RCA}_2$  is axiomatizable with a finite set of equations, and its equational theory is decidable. The same is true for  $\text{RCA}_1$ . Any  $\text{CA}_2$  satisfying for all  $i, j < 2$ ,  $i \neq j$

$$c_i x \cdot s_i^j c_i x \leq \text{ld} \rightarrow x = c_0 x \cdot c_1 x, \quad \text{or}$$

$$c_0 x \cdot c_1 x - x \leq c_i (s_i^j c_i x - \text{ld})$$

is representable.<sup>53</sup>

- (iv) The logic captured by  $\text{RCA}_n$  is first-order logic with equality restricted to  $n$  individual variables.

<sup>52</sup>L. Henkin and A. Tarski proved that  $\text{RCA}_n$  is a variety, J. D. Monk [Monk, 1969] proved that  $\text{RCA}_n$  is not finitely axiomatizable, L. Henkin gave a finite equational axiom system for  $\text{RCA}_2$ , and D. Scott proved that  $\mathbf{Eq}(\text{RCA}_2)$  is decidable.

<sup>53</sup>The above quasi-equation and equation then are equivalent to the so-called Henkin-equation (see [Henkin, Monk and Tarski, 1985, 3.2.65]), which was further simplified in [Venema, 1991, §3.5.2]. On the intuitive meaning of these see the paragraph preceding Theorem 70 in this work.

**Proof.** The proof of (i) goes exactly as in the previous Section, cf. the proofs of Theorems 2, 7.: The subdirectly irreducible members of  $\mathbf{RCA}_n$  are exactly the isomorphic copies of the nontrivial  $\mathbf{Cs}_n$ s (i.e. those with nonempty base set  $U$ ), and a switching term is  $c_{(n)}x$ , i.e. in  $\mathbf{Cs}_n$  we have<sup>54</sup>

$$c_{(n)}x = \begin{cases} 1 & \text{if } x \neq 0 \\ 0 & \text{if } x = 0. \end{cases}$$

The proof of undecidability of  $\mathbf{BRA}$  can be adapted here, too, by using the above outlined connection between  $\mathbf{RA}$  and  $\mathbf{CA}$ . Namely, the quasi-equational theory of semigroups can be interpreted in  $\mathbf{RCA}_n$ , e.g. by using the term

$$x; y = c_2(s_2^1 c_{(n \setminus 2)} x \cap s_2^0 c_{(n \setminus 2)} y).$$

The proof for nonfinite axiomatizability of  $\mathbf{RCA}_n$ ,  $n > 2$  will be discussed in Remark 23. The proof of the first part of (iii) can be found in [Henkin, Monk and Tarski, 1985, 3.2.65, 4.2.9]. It is not hard to check that in all  $\mathbf{CA}_2$ s, the quasi-equation and the equation in (iii) are equivalent to each other. Also, the equation in (iii) is equivalent to

$$(*) \quad c_0 x \cdot c_1 x \leq x + c_i(s_i^j c_i x - \text{Id})$$

which is then preserved under taking perfect extensions<sup>55</sup> (because negation  $-$  occurs only in front of a constant). Thus it is enough to show that any simple atomic  $\mathfrak{A} \in \mathbf{CA}_2$  satisfying  $(*)$  is representable. Now,  $(*)$  implies that there are no defective atoms in  $\mathfrak{A}$ , in the sense of [Henkin, Monk and Tarski, 1985, 3.2.59], and then  $\mathfrak{A} \in \mathbf{RCA}_2$  by [Henkin, Monk and Tarski, 1985, 3.2.59]. For (iv) see Example 6 in Section 7 of Part II. ■

More on the fine-structure of the equational theory of  $\mathbf{RCA}_n$  will be said later, after Problem 25.

How far did we get in obtaining algebras of relations in general (binary, ternary,  $\dots$ ,  $n$ -ary,  $\dots$ )?  $\mathbf{RCA}_n$  is a smooth and satisfactory algebraic theory of  $n$ -ary relations. So, can our theory handle all finitary relations? The answer is both yes and no. Namely, since  $n$  is an arbitrary finite number, in a sense, we can handle all finitary relations. But, we cannot have them *all* in the same algebra or in the same variety. For any finite family of relations, we can pick  $n$  such that they are all in  $\mathbf{RCA}_n$ . But this does not extend to infinite families of relations. To alleviate this, we could try working in the system  $\langle \mathbf{RCA}_n : n \in \omega \rangle$  of varieties instead of using just one of these. To use them all together, we need a strong coordination between them. This coordination is easily derivable from the embedding function  $\mathbf{Dr}$  sending

<sup>54</sup>Recall that  $n = \{0, 1, 2, \dots, n-1\}$ . Thus  $n \setminus 2 = \{2, 3, \dots, n-1\}$ .

<sup>55</sup>Perfect extensions are called canonical embedding algebras in [Henkin, Monk and Tarski, 1971]. Cf. also [Goldblatt, 1991; Goldblatt, 2000; Jónsson, 1995; Venema, 1996].

$R$  to  $R \times U$  for  $R \subseteq {}^nU$  defined above. Let  $\mathfrak{A} \subseteq \mathfrak{Rel}_n(U) = \langle \mathfrak{P}({}^nU) \cdots \rangle$  be a  $\mathbf{Cs}_n$  and let  $\mathfrak{B}$  be the  $\mathbf{Cs}_{n+1}$  generated by the  $\mathbf{Dr}$  image of  $\mathfrak{A}$ , i.e.  $\mathfrak{B} \subseteq \mathfrak{Rel}_{n+1}(U) = \langle \mathfrak{P}({}^{n+1}U) \cdots \rangle$  is generated by  $\{\mathbf{Dr}(R) : R \in \mathfrak{A}\}$ . The biggest  $\mathfrak{A}$  yielding the same  $\mathfrak{B}$  is called the  $n$ -ary *neat-reduct* of  $\mathfrak{B}$ , formally  $\mathfrak{A} = \mathbf{Nr}_n(\mathfrak{B})$ . Then

$$\mathbf{Nr}_n(\mathfrak{B}) = \{b \in B : c_n(b) = b\}.$$

Intuitively,  $\mathbf{Nr}_n(\mathfrak{B})$  is the algebra of  $n$ -ary relations ‘living in’ the algebra  $\mathfrak{B}$  of  $n+1$ -ary relations. It is not hard to see that  $\mathbf{Nr}_n : \mathbf{RCA}_{n+1} \rightarrow \mathbf{RCA}_n$  is a functor, in the category theoretical sense, for every  $n$ . Now, we can use the collection of varieties  $\mathbf{RCA}_n$  for all  $n \in \omega$ , synchronized via the functors  $\langle \mathbf{Nr}_n : n \in \omega \rangle$ , as a single mathematical entity containing all finitary relations.

Another possibility is to insist that we want all finitary relations over  $U$  represented as elements of a *single* algebra. In other words, this goal means that instead of a system of varieties we want to consider a single variety that in some sense incorporates all the original varieties taken together. Indeed, each  $\mathbf{RCA}_n$  can be viewed as incorporating all the  $\mathbf{RCA}_k$ s for  $k \leq n$ , since the latter can be recovered from  $\mathbf{RCA}_n$  by using the functors  $\mathbf{Nr}_{n-1}$ ,  $\mathbf{Nr}_{n-2}$  etc. So as  $n$  increases,  $\mathbf{RCA}_n$  gets closer and closer to the variety we want. Indeed, we take the limit of this sequence. There are two ways of doing this, the naïve way we will follow here and the category theoretical way we only briefly mention. It is shown in the textbook [Adámek, Herrlich and Strecker, 1990] that the system or ‘diagram’

$$\mathbf{RCA}_1 \xleftarrow{\mathbf{Nr}_1} \mathbf{RCA}_2 \xleftarrow{\mathbf{Nr}_2} \cdots \mathbf{RCA}_n \xleftarrow{\mathbf{Nr}_n} \mathbf{RCA}_{n+1} \cdots$$

is ‘convergent’ in the category theoretic sense, i.e. that it has a limit  $\mathbf{L}$ . Indeed, it is this class  $\mathbf{L}$  of algebras that we will construct below in a naïve way that does not use category theoretic tools or concepts.

We first extend our Convention 15, stated at the beginning of the present Section concerning associativity of Cartesian products and powers. *In the sequel*,  $\omega$  is the smallest infinite ordinal, as well as the set of all finite numbers, and  ${}^\omega U$  is the set of  $\omega$ -sequences over  $U$ . Furthermore,  ${}^n U \times {}^\omega U = {}^\omega U$ , and if  $R \subseteq {}^n U$  then  $R \times {}^\omega U \subseteq {}^\omega U$ , for  $n < \omega$ . We will also have to distinguish the constant  $\text{Id}$  of  $\mathbf{RCA}_3$  from that of  $\mathbf{RCA}_4$ . Therefore we let

$$\text{Id}^n \stackrel{\text{def}}{=} \{\langle a, \dots, a \rangle : a \in U\}$$

denote the  $n$ -ary identity relation on  $U$ .

How do we obtain an algebra containing all finitary relations over  $U$ ? If  $R$  is binary, but we want to treat it together with a 5-ary relation, then we represent  $R$  by  $R \times U \times U \times U = R \times {}^3 U$  in a  $\mathbf{Cs}_5$ . Taking this procedure to the limit, if we want to treat  $R$  together with relations of arbitrary high ranks,

then we can represent  $R$  with  $R \times {}^\omega U$ . This way we can embed all finitary relations into relations of rank  $\omega$ , and relations of different ranks become ‘comparable’ and ‘compatible’.<sup>56</sup> We still haven’t obtained the definition of  $\mathbf{Cs}_\omega$ s from that of  $\mathbf{Cs}_n$ s because we do not know what to do with the constant  $\text{ld}$ . More specifically, we want to be able to use the neat reduct functor  $\mathbf{Nr}_n$ , as the inverse of  $R \mapsto R \times {}^\omega U$  for  $R \subseteq {}^n U$ , in order to recover the original  $\mathbf{Cs}_n$ s from the new  $\mathbf{Cs}_\omega$ . This means that for  $\text{ld}^n \subseteq {}^n U$  we want  $\text{ld}^n \times {}^\omega U$  to be a derived constant (distinguished element) in our algebra. Adding  $\text{ld}^\omega = \{\langle a, \dots, a, \dots \rangle : a \in U\}$  as an extra constant does *not* ensure this any more. One of the most natural solutions is letting

$$\text{ld}_{ij}^\omega = \text{ld}_{ij} = \{q \in {}^\omega U : q_i = q_j\}$$

and defining a full  $\mathbf{Cs}_\omega$  as

$$\mathfrak{Rel}_\omega(U) \stackrel{\text{def}}{=} \langle \mathfrak{P}({}^\omega U), c_i, \text{ld}_{ij} \rangle_{i,j < \omega},$$

where the  $\text{ld}_{ij}$ s are constants. The price we had to pay for replacing the finite bound  $n$  on the ranks of relations we can treat with the infinite bound  $\omega$  is that we had to break up our single constant  $\text{ld}$  to infinitely many constants  $\text{ld}_{ij}$  ( $i, j \in \omega$ ).

$\mathbf{RCA}_\omega$  is defined to consist of all subalgebras of direct products of full  $\mathbf{Cs}_\omega$ s (up to isomorphisms),

$$\mathbf{RCA}_\omega = \mathbf{SP}\{\langle \mathfrak{P}({}^\omega U), c_i, \text{ld}_{ij} \rangle_{i,j < \omega} : U \text{ is a set}\}.$$

Again, as it was the case with BRAs and  $\mathbf{RCA}_n$ s,  $\mathbf{RCA}_\omega$ s are directly representable as algebras whose elements are  $\omega$ -ary relations.

The elements of  $\mathcal{P}({}^\omega U)$  are all the  $\omega$ -ary relations over  $U$ , and not only the ‘representations’  $R \times {}^\omega U$  of finitary relations  $R$  over  $U$ . Can we actually recover the algebras of finitary relations from the huge full  $\mathbf{Cs}_\omega$ s? Let

$$\mathbf{Rf}(U) = \{R \times {}^\omega U : R \subseteq {}^n U \text{ for some } n \in \omega\}.$$

Then  $\mathbf{Rf}(U) \subseteq \mathcal{P}({}^\omega U)$ ; moreover it is a subalgebra of the full  $\mathbf{Cs}_\omega$   $\mathfrak{Rel}_\omega(U)$  with universe  $\mathcal{P}({}^\omega U)$ . We will denote this subalgebra by  $\mathfrak{Rf}(U)$ . Now, we set

$$\mathbf{Csf}_\omega \stackrel{\text{def}}{=} \mathbf{S}\{\mathfrak{Rf}(U) : U \text{ is a nonempty set}\}.$$

---

<sup>56</sup>In particular we avoid the problem we ran into at the end of the introduction to this part in connection with the Boolean algebra  $\mathfrak{P}(<^\omega U)$ , because e.g. the complement of  $R \times {}^\omega U$  is  $\overline{R} \times {}^\omega U$  where  $\overline{R}$  denotes  ${}^n U \setminus R$  if  $R$  is an  $n$ -ary relation. Instead of trying to tame the Boolean-like algebra  $\{R : R \subseteq {}^n U \text{ for some } n\} \subseteq \mathcal{P}(<^\omega U)$ , we simply represent  $R \subseteq {}^n U$  by  $R \times {}^\omega U$  which *is* an element of  $\langle \mathfrak{P}({}^\omega U), c_0, \dots, c_n, \dots \rangle_{n < \omega}$ .



In a sense,  $\text{Csf}_\omega$  is the narrowest reasonable class of algebras of finitary relations.<sup>57</sup>

If  $R \in \text{Rf}(U)$ , then  $\Delta(R) = \{i \in \omega : c_i R \neq R\}$  is finite, in short  $R$  is finite-dimensional. We note that the converse is not true, there are  $R \subseteq {}^\omega U$  with  $\Delta(R) = \emptyset$ , yet  $R \notin \text{Rf}(U)$ . Indeed, fix  $u \in U$  and set

$$R = \{s \in {}^\omega U : \{i < \omega : s_i \neq u\} \text{ is finite}\}.$$

Then  $R \notin \text{Rf}(U)$  if  $|U| \geq 2$ , while  $c_i R = R$  for all  $i < \omega$ . A relation  $R \subseteq {}^\omega U$  is called *regular* if

$$(s \in R \quad \text{iff} \quad z \in R) \quad \text{whenever} \quad s, z \in {}^\omega U, s \upharpoonright \Delta(R) = z \upharpoonright \Delta(R).$$

Then the elements of  $\text{Rf}(U)$  are exactly the finite-dimensional regular relations on  $U$ .

Now we turn to the connections between the classes  $\text{Csf}_\omega$ ,  $\text{Cs}_\omega$  and  $\text{RCA}_\omega$ . Intuitively, the elements of  $\text{Csf}_\omega$  are algebras of finitary relations, while the elements of  $\text{Cs}_\omega$  (as well as those of  $\text{RCA}_\omega$ ) are algebras of  $\omega$ -ary relations.<sup>58</sup>

**THEOREM 18** (Basic properties of  $\text{RCA}_\omega$ ).

- (i)  $\text{RCA}_\omega$  is a variety with recursively enumerable and undecidable equational theory.
- (ii)  $\mathbf{Eq}(\text{RCA}_\omega) = \bigcup \{\mathbf{Eq}(\text{RCA}_n) : n \in \omega\}$ . I.e. in the language of  $\text{RCA}_n$ , the same<sup>59</sup> equations are true in  $\text{RCA}_n$  and in  $\text{RCA}_\omega$ .
- (iii)  $\text{RCA}_\omega = \text{SPCs}_\omega = \text{SPUpCsf}_\omega \neq \text{SPCsf}_\omega$ . I.e.  $\text{RCA}_\omega$  is both the variety and quasi-variety generated by  $\text{Csf}_\omega$ ; the same equations and quasi-equations are true in  $\text{Csf}_\omega$  and in  $\text{Cs}_\omega$ , but there is an infinitary quasi-equation distinguishing  $\text{Csf}_\omega$  and  $\text{Cs}_\omega$ .

<sup>57</sup>The letters  $\text{Rf}$  (and  $\mathfrak{Rf}$ ) refer to ‘finitary relations’.  $\text{Csf}_\omega$  is the above mentioned category theoretical limit  $\mathbf{L}$ . More precisely, for this equality to be literally true, when forming the category theoretic limit  $\mathbf{L}$ , instead of the varieties  $\text{RCA}_n$  we have to start out from their subdirectly irreducible members, which are nothing but  $\text{Cs}_n$ s. So  $\text{Csf}_\omega$  is the limit of the sequence  $\text{Cs}_1, \dots, \text{Cs}_n, \dots$ . The class  $\text{Csf}_\omega$  and its relationship with  $\text{RCA}_\omega$  was systematically investigated in [Andréka, 1973; Andréka, Gergely and Néméti, 1973; Andréka, Gergely and Néméti, 1977; Henkin *et al.*, 1981] and [Henkin, Monk and Tarski, 1985]. In the first three works the class was denoted by  $\text{Lv}$  or  $\text{Lr}$ , while in the last two by  $\text{Cs}_\omega^{\text{reg}} \cap \text{Lf}_\omega$ , the latter being the standard notation today.

<sup>58</sup>Theorem 18(i) is due to L. Henkin and A. Tarski. For the rest of the credits in connection with Theorem 18 we refer the reader to [Henkin, Monk and Tarski, 1971, I, II] and [Henkin *et al.*, 1981].

<sup>59</sup>Here the language of  $\text{RCA}_n$  should be taken to be the one defined above Definition 27.

**Proof.** (ii) follows from [Henkin, Monk and Tarski, 1985, 3.1.126]. Recursive enumerability and undecidability of  $\mathbf{Eq}(\mathbf{RCA}_\omega)$  follows from (ii) and Theorem 17 (for recursive enumerability one also has to use the proof of Theorem 17, namely that the recursive enumerations of  $\mathbf{Eq}(\mathbf{RCA}_n)$  given there are ‘uniform’ in  $n$ ). That  $\mathbf{RCA}_\omega$  is a variety follows e.g. from [Henkin, Monk and Tarski, 1985, 3.1.103], (where it is proved directly that  $\mathbf{RCA}_\omega$  is closed under taking homomorphic images).  $\mathbf{RCA}_\omega = \mathbf{SPUpCsf}_\omega$  follows from [Henkin, Monk and Tarski, 1985, 3.2.8, 3.2.10, 2.6.52]. To show  $\mathbf{RCA}_\omega \neq \mathbf{SPCsf}_\omega$  consider<sup>60</sup> the following infinitary quasi-equation  $q$ :

$$\bigwedge \{c_i x = x : i < \omega\} \wedge \text{ld}_{01} \leq x \rightarrow x = 1.$$

Then  $q$  is valid in  $\mathbf{SPCsf}_\omega$  while it is not valid in  $\mathbf{RCA}_\omega$ . ■

We note that  $\mathbf{SUPCsf}_\omega \neq \mathbf{RCA}_\omega$ . Indeed, consider the following universal formula  $\eta$

$$\begin{aligned} \overline{\text{ld}}_{(3)} &\neq 0 \rightarrow -\text{ld}_{01} \leq c_2 \overline{\text{ld}}_{(3)}, \quad \text{where} \\ \overline{\text{ld}}_{(3)} &\stackrel{\text{def}}{=} -\text{ld}_{01} \cap -\text{ld}_{02} \cap -\text{ld}_{12}. \end{aligned}$$

(The intuitive content of  $\eta$  is that if there is a ‘subbase’ of cardinality  $\geq 3$ , then there are no subbases of size 2.) Then  $\mathbf{Csf}_\omega \models \eta$  and  $\mathbf{RCA}_\omega \not\models \eta$ .

$\mathbf{RCA}_\omega$  is not a discriminator variety, e.g. because there are subdirectly irreducible but not simple  $\mathbf{RCA}_\omega$ s. But it is still an arithmetical variety of BAOs, from which many properties of  $\mathbf{RCA}_\omega$  follow by using theorems of universal algebra. It is not true that  $\mathbf{Sir}(\mathbf{RCA}_\omega) = \mathbf{ICs}_\omega$ , in fact no intrinsic characterization of  $\mathbf{Sir}(\mathbf{RCA}_\omega)$  is known.<sup>61</sup> The variety  $\mathbf{RCA}_\omega$  is very well investigated, perhaps the most detailed study is in [Henkin *et al.*, 1981; Henkin, Monk and Tarski, 1985]. For more recent results see e.g. [Goldblatt, 1995; Monk, 1993; Shelah, 1991; Serény, 1985; Serény, 1997] and [Hodkinson, 1997].

The theorems which say that  $\mathbf{BRA}$ ,  $\mathbf{RRA}$ , and  $\mathbf{RCA}_n$  are not finitely axiomatizable, carry over to  $\mathbf{RCA}_\omega$  too. However, to avoid triviality, instead of non-finite axiomatizability we have to state something stronger, because  $\mathbf{RCA}_\omega$  has infinitely many operations and finitely many axioms can speak about only finitely many operations. Taking this into account, when trying to axiomatize  $\mathbf{RCA}_\omega$ , one could still hope for a finite ‘schema’ (in some sense) of equations treating the infinity of the  $\mathbf{RCA}_\omega$ -operations uniformly. A possible example for a finite schema is  $c_i c_j x = c_j c_i x$  ( $i, j \in \omega$ ). The

<sup>60</sup>Another proof, exporting logical properties to algebras, can be found at the end of Example 9 in Section 7.

<sup>61</sup>It is known that  $\mathbf{Sir}(\mathbf{RCA}_\omega)$  is a proper subclass of  $\mathbf{IW}_\omega$ . More on this see [Henkin, Monk and Tarski, 1985, 3.1.83–3.1.88], [Andréka, Némethi and Thompson, 1990].

following theorem<sup>62</sup> implies that it will be hard to find such a schema, and that certain kinds of schemata are ruled out to begin with.

**THEOREM 19** (Nonfinite axiomatizability of  $\text{RCA}_\omega$ ). *The variety  $\text{RCA}_\omega$  is not axiomatizable by any set  $\Sigma$  of universally quantified formulas such that  $\Sigma$  involves only finitely many variables.*

**Proof. Plan:** For all  $m < \omega$  we will construct an algebra  $\mathfrak{A}_m$  such that

- a)  $\mathfrak{A}_m \notin \text{RCA}_\omega$
- b) every  $m$ -generated subalgebra of  $\mathfrak{A}_m$  is in  $\text{RCA}_\omega$ .

This will prove the theorem because of the following. Assume that  $\Sigma$  is a set of quantifier-free formulas such that  $\Sigma$  involves at most  $m$  variables ( $|\text{var}(\Sigma)| \leq m < \omega$ ) and  $\text{RCA}_\omega \models \Sigma$ . Then  $\Sigma$  is valid in an algebra  $\mathfrak{B}$  iff  $\Sigma$  is valid in every  $m$ -generated subalgebra of  $\mathfrak{B}$ , because  $|\text{var}(\Sigma)| \leq m$  and  $\Sigma$  contains no quantifiers. Thus  $\mathfrak{A}_m \models \Sigma$  by b) and by  $\text{RCA}_\omega \models \Sigma$ . Then  $\mathfrak{A}_m \notin \text{RCA}_\omega$  shows that  $\Sigma$  does not axiomatize  $\text{RCA}_\omega$ .

**Construction of  $\mathfrak{A}_m$ :** Let  $\kappa \geq 2^m$  be finite, and let  $\langle U_i : i < \omega \rangle$  be a system of pairwise disjoint sets each of cardinality  $\kappa$ . Let

$$U = \bigcup \{U_i : i \in \omega\}, \quad \text{let}$$

$$q \in \text{P}_{i \in \omega} U_i \stackrel{\text{def}}{=} \{s \in {}^\omega U : (\forall i \in \omega) s_i \in U_i\} \text{ be arbitrary,}$$

$$R = \{z \in \text{P}_{i \in \omega} U_i : |\{i \in \omega : z_i \neq q_i\}| < \omega\}, \quad \text{and let}$$

$\mathfrak{A}'$  be the subalgebra of  $\langle \mathfrak{P}({}^\omega U), c_i, \text{Id}_{ij} \rangle_{i,j \in \omega}$  generated by the element  $R$ .

Then  $R$  is an atom of  $\mathfrak{A}'$  because of the following. For any two sequences  $s, z \in R$  there is a permutation  $\sigma : U \rightarrow U$  of  $U$  taking  $s$  to  $z$  and fixing  $R$ , i.e.  $s \circ \sigma = z$  and  $R = \{p \circ \sigma : p \in R\}$  (the obvious choice for  $\sigma$ , interchanging  $s_i$  and  $z_i$  for all  $i \in \omega$  and leaving everything else fixed, works). If  $\sigma$  is a permutation of  $U$  fixing  $R$ , then  $\sigma$  fixes all the elements generated by  $R$  because the operations of  $\text{RCA}_\omega$  are permutation invariant. Thus if  $\emptyset \neq a \in \mathfrak{A}'$  and  $s \in a \cap R$  then  $R \subseteq a$ , showing that  $R$  is an atom of  $\mathfrak{A}'$ .

We now ‘split  $R$  into  $\kappa + 1$  new atoms  $R_j$  each imitating  $R$ ’ obtaining a new, bigger algebra  $\mathfrak{A}$  from our old  $\mathfrak{A}'$ . i.e. we choose a larger algebra  $\mathfrak{A}$  such that  $\mathfrak{A}$  satisfies the conditions below and is otherwise arbitrary.

---

<sup>62</sup>Monk [1969] proves that  $\text{RCA}_\omega$  cannot be axiomatized by a finite number of schemas of equations like those in the definition of  $\text{CA}_\omega$ . See [Henkin, Monk and Tarski, 1985, 4.1.7]. Theorem 19, due to Andréka, is a generalization of that result and can be found in [Andréka, 1997] or in [Monk, 1993].

$\mathfrak{A}' \subseteq \mathfrak{A}$ , the Boolean reduct of  $\mathfrak{A}$  is a Boolean algebra,

$R_j$  are atoms of  $\mathfrak{A}$  and  $c_i R_j = c_i R$  for  $j \leq \kappa, i < \omega$ ,

each element of  $\mathfrak{A}$  is a Boolean join of an element of  $\mathfrak{A}'$  and of some  $R_j$ s

$c_i$  distributes over joins, for any  $i < \omega$ , i.e.  $\mathfrak{A} \models c_i(x \cup y) = c_i x \cup c_i y$ .

Note that in  $\mathfrak{A}$  '∪' is only an abstract algebraic operation and not necessarily set theoretical union. It is easy to see that such an extension  $\mathfrak{A}$  of  $\mathfrak{A}'$  exists. See Figure 8.

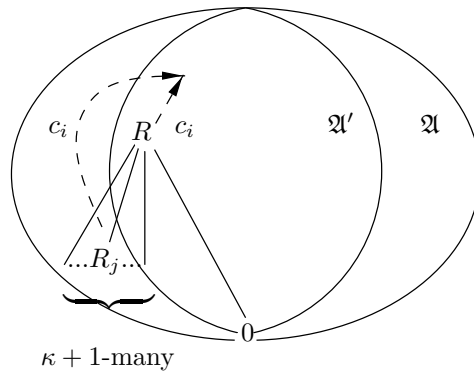


Figure 8.

By the above, we have constructed our algebra  $\mathfrak{A}_m$  which in the following we will denote just by  $\mathfrak{A}$ .

CONVENTION 20. In this proof we use the symbols  $\cap, \cup$  denoting the concrete operations of our set algebras ( $\mathbf{Cs}_\omega$ s) also as the corresponding *abstract algebraic* operation symbols (denoting themselves in  $\mathbf{Cs}_\omega$ s). So, if  $x, y$  are variable symbols, then  $x \cap y$  is a term. We hope context will help in deciding whether  $x \cap y$  is meant to be a term or a concrete set.  $x - y$  is the Boolean term  $(x \cap -y)$  denoting the set  $x \setminus y$  in  $\mathbf{Cs}_\omega$ s. It is especially important to note that since, for the algebra  $\mathfrak{A}$  constructed above,  $\mathfrak{A} \notin \mathbf{RCA}_\omega$  was not excluded, the operations denoted by  $\cup, \cap, c_i$  etc. in  $\mathfrak{A}$  are not assumed to be the real, set theoretic ones. They are just abstract operations despite of the notation '∪' etc. The Boolean ordering on  $\mathfrak{A}$  will be denoted by  $\leq$ .

CLAIM 21.  $\mathfrak{A} \notin \mathbf{RCA}_\omega$ .

**Proof.** For  $i, j < \omega$ ,  $i \neq j$ , let  $s_j^i(x) = c_i(\text{Id}_{ij} \cap x)$  and  $s_i^i(x) = x$ . Let the term  $\tau$  be defined by

$$\tau(x) \stackrel{\text{def}}{=} \bigcap_{i \leq \kappa} s_i^0 c_1 \dots c_\kappa x \cap \bigcap_{i < j \leq \kappa} -\text{Id}_{ij}.$$

Let  ${}^\omega U^{(q)} = \{z \in {}^\omega U : \{i \in \omega : z_i \neq q_i\} \text{ is finite}\}$ . ( ${}^\omega U^{(q)}$  is usually called the *weak Cartesian space determined by  $U$  and  $q$* .) Then, for our concrete choice of  $R$ ,  $\mathfrak{A}' \models \tau(R) = 0$  because of the following:

$$\begin{aligned} c_1 \dots c_\kappa R &= (U_0 \times {}^\kappa U \times U_{\kappa+1} \times \dots) \cap {}^\omega U^{(q)}, \\ s_i^0 c_1 \dots c_\kappa R &= ({}^i U \times U_0 \times ({}^{\kappa-i} U \times U_{\kappa+1} \times \dots)) \cap {}^\omega U^{(q)}, \\ \bigcap_{i \leq \kappa} s_i^0 c_1 \dots c_\kappa R &= ({}^{(\kappa+1)} U_0 \times U_{\kappa+1} \times \dots) \cap {}^\omega U^{(q)}. \end{aligned}$$

Then by  $|U_0| \leq \kappa$  we have that there is no repetition-free sequence in  ${}^{(\kappa+1)} U_0$ . Thus  $\mathfrak{A}' \models \tau(R) = 0$ .

Then  $\mathfrak{A} \models \tau(R) = 0$  by  $\mathfrak{A}' \subseteq \mathfrak{A}$  and  $R \in A'$ . Assume that  $\mathfrak{A} \in \text{RCA}_\omega$ . Then there is a homomorphism  $h : \mathfrak{A} \rightarrow \mathfrak{B} \in \text{Cs}_\omega$  such that  $h(R) \neq 0$ , for some  $\mathfrak{B}$ . By  $h(R) \neq 0$ , there is  $s \in h(R)$ . By  $R \leq c_0 R_j$  we have  $h(R) \subseteq c_0 h(R_j)$ , so there is  $u_j$  such that  $\langle u_j, s_1, s_2, \dots, s_i \dots \rangle \in h(R_j)$ , for all  $j \leq \kappa$ . These  $u_j$ s are different from each other since the  $R_j$ s are disjoint from each other, and so the  $h(R_j)$ s are disjoint from each other. Consider the sequence

$$z = \langle u_0, u_1, \dots, u_\kappa, s_{\kappa+1}, \dots \rangle.$$

Then  $z \in \tau(h(R))$  is easily seen as follows. Obviously  $z \in -\text{Id}_{ij}$ , if  $i < j \leq \kappa$ . Further  $\langle u_i, z_1, \dots, z_j \dots \rangle \in \text{Id}_{0i} \cap c_1 \dots c_\kappa h(R)$ , hence  $z \in s_i^0 c_1 \dots c_\kappa h(R)$  if  $i \leq \kappa$ . Thus  $z \in \tau(h(R))$ , a contradiction. ■

CLAIM 22. The  $m$ -generated subalgebras of  $\mathfrak{A}$  are in  $\text{RCA}_\omega$ .

**Proof.** Let  $G \subseteq A$ ,  $|G| \leq m$ . For all  $i, j \leq \kappa$  define

$$R_i \equiv R_j \text{ iff } (\forall g \in G)[R_i \leq g \leftrightarrow R_j \leq g].$$

Then  $\equiv$  is an equivalence relation on  $\{R_j : j \leq \kappa\}$  which has  $\leq 2^m$  blocks by  $|G| \leq m$ . Let  $p$  denote the number of blocks of  $\equiv$ , i.e.  $p = |\{R_j / \equiv : j \leq \kappa\}| \leq 2^m \leq \kappa$ . Define

$$B = \{a \in A : (\forall i, j \leq \kappa)([R_i \equiv R_j \text{ and } R_j \leq a] \Rightarrow R_i \leq a)\}.$$

We now show that  $B$  is closed under the operations of  $\mathfrak{A}$ .: Let  $k < l < \omega$ .

- 1)  $B$  clearly is closed under the Boolean operations.
- 2)  $\text{Id}_{kl} \in B$  since  $(\forall j \leq \kappa) R_j \not\leq \text{Id}_{kl}$ .
- 3) Clearly,  $A' \subseteq B$  (since  $R$  is an atom of  $\mathfrak{A}'$ ), and  $c_k a \in A'$  for all  $a \in A$ . Thus  $c_k b \in B$  (for all  $b \in A$ ).

Let  $\mathfrak{B} \subseteq \mathfrak{A}$  be the subalgebra of  $\mathfrak{A}$  with universe  $B$ . By  $G \subseteq B$ , it is enough to show that  $\mathfrak{B} \in \text{RCA}_\omega$ .

We will define an embedding  $h : \mathfrak{B} \rightarrow \langle \mathfrak{P}({}^\omega U), c_i, \text{Id}_{ij} \rangle_{i,j < \omega}$ . Let  $\{y_j : j < p\} = \{\sum(R_j / \equiv) : j \leq \kappa\}$ . Then  $\{y_j : j < p\}$  is a partition of  $R$  in  $\mathfrak{B}$ , i.e. they are pairwise disjoint and sum up to  $R$ ,  $c_i y_j = c_i R$  for all  $j < p$  and  $i < \omega$  and every element of  $\mathfrak{B}$  is a join of some element of  $A'$  and of some  $y_j$ s. So,  $\mathfrak{B}$  looks like the algebra on Figure 9.

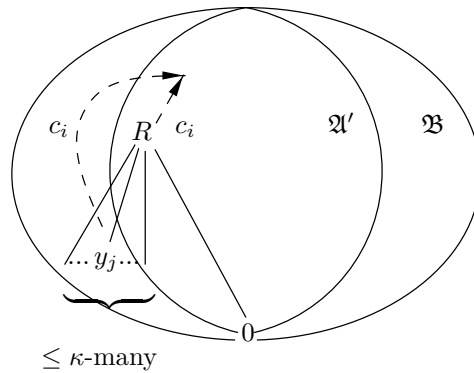


Figure 9.

First we define the images of the  $y_j$ s. Let  $Q = \{0, 1, \dots, \kappa - 1\}$  and let  $(Q, +, 0)$  be a commutative group. For each  $i < \omega$  let  $f_i : U_i \rightarrow Q$  be a bijection such that  $f_i(q_i) = 0$ . For  $j < \kappa$  define

$$R''_j = \left\{ z \in R : \sum \langle f_i(z_i) : i < \omega \rangle = j \right\},$$

where  $\sum$  denotes the group-theoretic sum in  $(Q, +, 0)$ . Then it is not difficult to check that the  $R''_j$ s are disjoint from each other and

$$c_i R''_j = c_i R \quad \text{for all } i < \omega$$

for the concrete set theoretic  $c_i$ s. Define for all  $j < p - 1$

$$R'_j = R''_j$$

$$R'_{p-1} = \bigcup \{R''_j : p - 1 \leq j < \kappa\}.$$

We are ready to define the embedding  $h$  of  $B$ .: We define for all  $b \in B$

$$h(b) = (b - R) \cup \bigcup \{R'_j : j < p, y_j \leq b\},$$

where  $b - R$  is computed in  $\mathfrak{A}$ , and since  $(b - R) \in A' \subseteq \mathcal{P}({}^\omega U)$ , the rest of the operations are the concrete set theoretic ones. Now it is not difficult to check that  $h$  is an embedding  $h : \mathfrak{B} \rightarrow \langle \mathfrak{P}({}^\omega U), c_i, \text{Id}_{ij} \rangle_{i,j < \omega}$  as follows.

Clearly  $h$  preserves  $\cup, -$ .  $h(b) = 0$  implies  $b = 0$ , hence  $h$  is one-one.  $h(\text{Id}_{kl}) = \text{Id}_{kl}$ . Now we check  $c_k h(b) = h(c_k b)$ .

$$\begin{aligned} c_k h(b) &= c_k [(b - R) \cup \bigcup \{R'_j : y_j \leq b\}] \\ &= c_k (b - R) \cup \bigcup \{c_k R'_j : y_j \leq b\} \\ &= c_k (b - R) \cup \bigcup \{c_k y_j : y_j \leq b\} \\ &= c_k [(b - R) \cup \bigcup \{y_j : y_j \leq b\}] \\ &= c_k b, \end{aligned}$$

where the operations in the first two lines are set-theoretic while those in the last three lines are understood in the abstract algebra  $\mathfrak{A}$ .  $h(c_k b) = (c_k b - R) \cup \bigcup \{R'_j : y_j \leq c_k b\} = c_k b$ , since  $(\exists j)y_j \leq c_k b$  iff  $R \leq c_k b$ , and  $R \not\leq c_k b$  iff  $c_k b = c_k b - R$ . **QED**(Claim 22)  $\blacksquare$

According to our Plan way above, the above two Claims complete the proof of Theorem 19.  $\blacksquare$

Remark 23 below describes the modifications needed for obtaining proofs for the analogous (with Theorem 19) non-finitizability theorems for  $\text{RCA}_n$  ( $n > 2$ ) and  $\text{RRA}$ .

**REMARK 23.** Here we outline the modifications of the above proof of Theorem 19 yielding proofs for non-finite axiomatizability of  $\text{RCA}_n$  and  $\text{RRA}$ .

Let  $\alpha$  be  $n$  or  $\omega$ . An algebra  $\mathfrak{A}$  similar to  $\text{RCA}_\alpha$ s is said to be *representable* if  $\mathfrak{A} \in \text{RCA}_\alpha$ . Thus representability means that  $\mathfrak{A}$  is isomorphic to an algebra  $\mathfrak{A}^+$  whose elements are  $\alpha$ -ary relations and whose greatest element is a disjoint union of Cartesian spaces.  $\mathfrak{A}^+$  is called a *representation* of  $\mathfrak{A}$  and sometimes the isomorphism  $h : \mathfrak{A} \rightarrow \mathfrak{A}^+$  too is called the representation of  $\mathfrak{A}$ . By a *homomorphic representation* we understand a homomorphism mapping  $\mathfrak{A}$  into some  $\text{Cs}_\alpha$ . This concept receives its importance from the simple but useful fact that representability of  $\mathfrak{A}$  is equivalent with the existence of a set  $H$  of homomorphic representations of  $\mathfrak{A}$  such that  $(\forall \text{nonzero } x \in A)(\exists h \in H)h(x) \neq 0$ .

The *intuitive idea* of the above proof of Theorem 19 was the following. We found two different ways of ‘counting’ the elements of the domain  $\{s_0 :$

$s \in R\}$  of the relation  $R$ . This counting was done by looking only at the abstract, i.e. isomorphism invariant properties of  $\mathfrak{A}$ . The two ways of counting were: (1) Looking at the number of the disjoint elements  $R_j$  below  $R$ . This allowed us to conclude that the domain of  $R$  must be big. (2) Using the  $\text{Id}_{i,j}$ s exactly as one uses equality in first-order logic to express that a certain finite set is smaller than some  $\kappa$ , we concluded that the domain of  $R$  must be small. (This was done by the term  $\tau(R)$  in the proof of Claim 21.)

We started out from an  $\mathfrak{A}' \in \text{RCA}_\omega$  in which the counting (2) said that ‘ $\text{Dom}(R)$ ’ is small. Then by splitting, we enlarged  $\mathfrak{A}'$  to  $\mathfrak{A}$ , such that in this bigger algebra  $\mathfrak{A}$  the counting (1) said that ‘ $\text{Dom}(R)$ ’ is big. Thus in  $\mathfrak{A}$  the two countings (1) and (2) contradict each other, ensuring  $\mathfrak{A} \notin \text{RCA}_\omega$ .

This is how we constructed one nonrepresentable algebra ( $\mathfrak{A}_m$ ). We were able to construct an infinite sequence of such algebras in such a way that as  $m$  increases, the contradiction between (1) and (2) becomes weaker and weaker. Actually, as  $m$  approaches infinity, the contradiction between (1) and (2) vanishes. So in the ultraproduct of the  $\mathfrak{A}_m$ s, (1) and (2) do not contradict each other any more, and this ultraproduct is in  $\text{RCA}_\omega$ . In our construction the conflict between (1) and (2) became weaker and weaker in the sense that more and more elements had to be inspected for discovering this contradiction.<sup>63</sup> This finishes the intuitive idea of the proof of Theorem 19.

Next we would like to repeat this proof for  $\text{RCA}_n$  in place of  $\text{RCA}_\omega$ , with  $2 < n < \omega$ . If we simply replace  $\omega$  everywhere with  $n$ , the proof does not go through because the counting in (2) needs an arbitrarily large number of  $\text{Id}_{i,j}$ s and we have only  $n \times n$  many.<sup>64</sup> So we need a new method for doing (2). This amounts to looking for an abstract algebra  $\mathfrak{A}$  together with its element  $R$  and concluding that in *any* (homomorphic) representation  $h : \mathfrak{A} \rightarrow \mathfrak{B} \in \text{Cs}_n$  of  $\mathfrak{A}$ , the domain  $U_0$  of  $h(R)$  must be of smaller size than a certain  $\kappa$ . (The difficulty is that we have to be able to repeat this for arbitrarily large  $\kappa \in \omega$ .) We also need to keep in mind that we will want to have a contradiction with (1), which means that we will want to split  $R$ . In order to be able to do this, we only need that  $R$  remain an atom. There are many natural ways for ensuring (by abstract properties) smallness of a set. Perhaps the simplest way is the following. If we could ‘see’ by looking at  $\mathfrak{A}$  ‘abstractly’ that  $U_0 \times U_0$  is a union of fewer than  $\kappa$  functions  $f_i$  ( $i < \kappa - 1$ ) each of which is coded by an element of  $\mathfrak{A}$ , then the domain  $U_0$  of  $R$  must be of smaller cardinality than  $\kappa$  in *any* representation of  $\mathfrak{A}$ . E.g. we can take these functions (the  $f_i$ s) to be powers of a single suitable permutation  $f$  of  $U_0$ ; say let  $U_0 = \kappa$ , and let  $f$  be the usual successor modulo  $\kappa$ . Let

<sup>63</sup>This allowed us to avoid ultraproducts in the final argument. We find it more natural to explain the intuitive idea in terms of ultraproducts, which incidentally happens to be the way the original proof of nonfinitizability went.

<sup>64</sup>We can construct the algebras  $\mathfrak{A}_m$  as in the proof of Thm. 19 for  $m < \log_2(n)$ . But for the contradiction to vanish, we need  $\mathfrak{A}_m$  for arbitrarily large  $m$ .



$F \stackrel{\text{def}}{=} f \times {}^{n-2}U$ . Then  $F \subseteq {}^nU$ . We include into our algebra  $\mathfrak{A}'$ , besides  $R$ , also  $F$  as a new generator element. It can be checked that  $R$  remains an atom (because no subset of  $U_0$  became ‘definable’). Now, similarly to the way we used the equation  $\tau(R) = 0$  in the proof of Claim 21, by studying the abstract properties of  $F$  and  $R$  in the new  $\mathfrak{A}'$  we can conclude that in any homomorphic representation  $h$  of  $\mathfrak{A}'$ , the Cartesian square of the domain of  $h(R)$  is contained in the union of fewer than  $\kappa$  powers of a function coded by  $h(F)$ . But then this domain must be of cardinality  $\leq \kappa$ . (Exactly what we proved in Claim 21 of the old proof. So we can prove our new Claim 21.) After this modification, the whole proof goes through by replacing all occurrences of  $\omega$  with  $n$ .<sup>65</sup> This completes the outline of the proof that  $\text{RCA}_n$  cannot be axiomatized with quantifier free formulas using finitely many variables, if  $n > 2$ .

Let us turn to the RRA case, i.e. to Theorems 6, 10. The idea is basically the same as in the above outlined  $\text{RCA}_n$  case. Exactly as in the  $\text{RCA}_n$  case, here too we use two counting principles (1), (2) and construct algebras  $\mathfrak{A}_m$  in which (1) and (2) contradict each other. Again we want a controllable contradiction such that as  $m$  approaches infinity the contradiction vanishes. Here we have to take a less obvious principle for counting in (2), because in RRA, functions interfere with splitting<sup>66</sup> elements  $R = U_0 \times U_1$ . E.g. we can use colorings of the full graph  $U_0 \times U_0$  with finitely many colors without monochromatic triangles, and then apply Ramsey’s theorem. This means that we arrange  $U_0 \times U_0$  to be a disjoint union of symmetric relations  $G_0, \dots, G_r$  such that  $G_r = \text{Id} \upharpoonright U_0$  (symmetric means  $G_i = G_i^{-1}$ ) and  $(G_i \circ G_j) \cap G_k = 0$ . To ensure splittability of  $R$  we also arrange that  $G_i \circ G_j \supseteq G_k$  whenever  $|\{i, j, k\}| > 1, i, j, k < r$ . We let our  $\mathfrak{A}'$  be generated in this case by  $\{G_0, \dots, G_r, R\}$ . All these properties of the  $G_i$ s were abstract, ‘equational’ ones.<sup>67</sup> This ensures that in every representation of  $\mathfrak{A}'$ , the domain  $U_0$  of  $R$  must be finite (by Ramsey’s theorem). We split  $R$  into  $\omega$  many  $R_i$ s obtaining  $\mathfrak{A}$  from  $\mathfrak{A}'$  as we did in the  $\text{RCA}_\omega, \text{RCA}_n$  cases before. The rest of the proof goes through as before with replacing  $\omega$  (or  $n$ ) everywhere by 2, except for the following change. In the RRA case we have to look at the ultraproduct of the  $\mathfrak{A}_m$ s and observe that it is representable (since the contradiction between (1) and (2) disappeared as both counting gives us continuum many elements). Therefore this proof gives only non-finite

<sup>65</sup>We need  $n \geq 3$  to be able to see abstractly that the  $f_i$ s are functions. This proof is worked out in detail in [Andréka, 1997, Thm.1].

<sup>66</sup>Splitting in RA is defined, and the conditions for splittability are described, in [Andréka, Maddux and Németi, 1991].

<sup>67</sup>In fact, it is an open problem in RA-theory (see e.g. [Andréka, Monk and Németi, 1991, section on open problems], whether there are such concrete relations  $G_0, \dots, G_r$  on some set  $U_0$  or not. What we should do here is that we state these properties abstractly on some abstract relations  $G_0, \dots, G_r$ . The only difference from the previous proof will then be that we do not know whether  $\mathfrak{A}' \in \text{RRA}$ . But this does not matter, what we need is that  $\mathfrak{A}_m \notin \text{RRA}$  and  $\text{P}\mathfrak{A}_m/F \in \text{RRA}$ .

axiomatizability of RRA (i.e. Monk's theorem) without proving (Jónsson's result saying) that infinitely many variables are needed. For the latter, one has to fine-tune the construction some more.<sup>68</sup>

In Section 1, Theorem 6 leads to Problem 12 in a natural way. Exactly the same way our present Theorem 19 leads to the following important open problem.

**PROBLEM 24.** Find *simple*, mathematically transparent, decidable sets  $E$  of equations axiomatizing  $\text{RCA}_\omega$ . The  $\text{RCA}_n$ ,  $2 < n < \omega$  version of this problem is open and interesting, too.

The  $\text{RCA}_n$  version is strongly related to Problem 12 in Section 1. On the other hand, the present,  $\text{RCA}_\omega$  version has a logical counterpart, cf. e.g. [Henkin, Monk and Tarski, 1985, Prob.4.16, p.180]. This is one of the central problems of Algebraic Logic, cf. [Henkin, Monk and Tarski, 1985, Prob.4.1], [Henkin and Monk, 1974, Prob.5], etc. For strongly related results (or for partial solutions) see [Henkin, Monk and Tarski, 1985; Hirsch and Hodkinson, 1997; Simon, 1991; Simon, 1993; Venema, 1991; Venema, 1995, pp.112–119].

**PROBLEM 25.** Is there a finite schema axiomatizable quasi-variety  $K$  such that  $\mathbf{Eq}(K) = \mathbf{Eq}(\text{RCA}_\omega)$ , i.e. the variety generated by  $K$  is  $\text{RCA}_\omega$ ? The same for  $\text{RCA}_n$  for  $n < \omega$ . I.e., is there a finitely axiomatizable quasi-variety  $K \subseteq \text{RCA}_n$  such that  $\text{RCA}_n = \mathbf{HK}$ ?

This problem is related to the existence of weakly sound Hilbert-style inference systems for first-order logic, see Part II, Theorem 52 and Open Problem 64.

### **On the structure of the equational axiomatizations of $\text{RCA}_n$ , $\text{RCA}_\omega$ :**

Let  $E$  be an arbitrary set of equations axiomatizing  $\text{RCA}_n$ . As in the RRA-case,  $E$  must be infinite, but it can be chosen to be decidable. Unlike the RRA-case, here every operation symbol has to occur infinitely many times in  $E$  (in the RRA-case, only the Booleans and  $\circ$  had to occur infinitely many times). A similar statement is true for  $\text{RCA}_\omega$  in place of  $\text{RCA}_n$ . For more on this see Figure 10 and [Andréka, 1994; Andréka, 1997]. Concrete decidable sets  $E$  are known, see e.g. [Henkin, Monk and Tarski, 1985, pp.112–119], cf. also [Hirsch and Hodkinson, 1997; Simon, 1991; Simon, 1993; Venema, 1991; Venema, 1995]. However, it would be important to find choices of  $E$  with more perspicuous structures, see Problem 24.

Let us turn to the relationship between  $\text{RCA}_\omega$  and its abstract approximation  $\text{CA}_\omega$ . These investigations yield information on proof theoretical

<sup>68</sup>This is done in [Andréka and Némethi, 1990], where we use projective geometries for the purposes of counting in (2).

properties of first-order logic and of some related logics. See Examples 6, 8, 9 in Section 7, especially Theorem 66 – Remark 69.

**DEFINITION 26** ( $\text{CA}_\omega$ , an abstract approximation of  $\text{RCA}_\omega$ ). A  $\text{CA}_\omega$  is a normal BAO of the same similarity type as  $\text{RCA}_\omega$  in which the  $c_i$ s are self-conjugated commuting closure operations, and in which the constants  $\text{ld}_{ij}$  satisfy the following equations:

(3') For all  $i, j, k < \omega$

$$(i) \quad c_i \text{ld}_{ij} = 1, \quad c_k \text{ld}_{ij} = \text{ld}_{ij} \text{ if } k \neq i, j, \quad \text{ld}_{ii} = 1, \quad \text{ld}_{ij} = \text{ld}_{ji}, \quad \text{and} \\ \text{ld}_{ij} \cap \text{ld}_{jk} \leq \text{ld}_{ik}.$$

$$(ii) \quad \text{ld}_{ij} \cap c_i x = x \quad \text{whenever } x \leq \text{ld}_{ij} \text{ and } i \neq j.$$

To treat  $\text{RCA}_n, \text{CA}_n$  and  $\text{RCA}_\omega, \text{CA}_\omega$  in a unified manner, we replace  $\omega$  in the definitions of  $\text{RCA}_\omega$  and  $\text{CA}_\omega$  with an arbitrary but fixed ordinal  $\alpha$ , obtaining  $\text{RCA}_\alpha, \text{CA}_\alpha$  (here  $\alpha = n$  and  $\alpha = \omega$  are of course permitted).<sup>69</sup>

If  $\alpha = n < \omega$ , then the newly defined  $\text{RCA}_n$  and  $\text{CA}_n$  are only definitionally equivalent with the previously defined ones, because in Definition 16 we had only one constant  $\text{ld}$  in place of the present  $n \times n$ -many constants  $\text{ld}_{ij}$ . This definitional equivalence is given by

$$\text{ld} = \bigcap_{i,j < n} \text{ld}_{ij} \quad \text{and} \quad \text{ld}_{ij} = c_{(n \setminus \{i,j\})} \text{ld}.$$

Since definitional equivalence is a very close connection between classes of algebras, we did not give new names for  $\text{RCA}_n$  and  $\text{CA}_n$ .

**DEFINITION 27.** (Locally finite, dimension-complemented CAs, and neat-reducts) Let  $\alpha, \beta$  be any ordinals.

(i) Let  $\mathfrak{A} \in \text{CA}_\alpha$ ,  $x \in A$ . Then  $\Delta(x) \stackrel{\text{def}}{=} \{i \in \alpha : c_i x \neq x\}$ .

$$\text{Lf}_\alpha \stackrel{\text{def}}{=} \{\mathfrak{A} \in \text{CA}_\alpha : (\forall x \in A) (\Delta(x) \text{ is finite})\}.$$

$$\text{Dc}_\alpha \stackrel{\text{def}}{=} \{\mathfrak{A} \in \text{CA}_\alpha : (\forall x \in A) (\alpha \setminus \Delta(x) \text{ is infinite})\}.$$

(ii) Assume  $\alpha \leq \beta$  and  $\mathfrak{A} \in \text{CA}_\beta$ . Then

$$\text{Nr}_\alpha \mathfrak{A} \stackrel{\text{def}}{=} \{x \in A : \Delta(x) \subseteq \alpha\}, \quad \mathfrak{Nr}_\alpha \mathfrak{A} \stackrel{\text{def}}{=} \langle \text{Nr}_\alpha \mathfrak{A}, c_i^\mathfrak{A}, \text{ld}_{ij}^\mathfrak{A} \rangle_{i,j < \alpha}.$$

In the above,  $c_i^\mathfrak{A}, \text{ld}_{ij}^\mathfrak{A}$  denote the corresponding operations of  $\mathfrak{A}$ . It can be checked that  $\mathfrak{Nr}_\alpha \mathfrak{A} \in \text{CA}_\alpha$ .

<sup>69</sup>This generalization will also be useful in algebraizing various quantifier logics different from classical first-order logic.

$$\mathbf{Nr}_\alpha \mathbf{CA}_\beta \stackrel{\text{def}}{=} \{\mathfrak{Nr}_\alpha \mathfrak{A} : \mathfrak{A} \in \mathbf{CA}_\beta\}.$$

The elements of  $\mathbf{Lf}_\alpha$  and  $\mathbf{Dc}_\alpha$  are called *locally finite* and *dimension-complemented*  $\mathbf{CA}_\alpha$ s respectively.  $\mathfrak{Nr}_\alpha \mathfrak{A}$  is called the  $\alpha$ -neat reduct of  $\mathfrak{A}$ . We note that

$$\mathbf{Dc}_\alpha = \{\mathfrak{A} \in \mathbf{CA}_\alpha : (\forall x \in A)(\alpha \setminus \Delta(x) \neq \emptyset)\}.$$

**THEOREM 28** (Relationships between  $\mathbf{Lf}_\alpha$ ,  $\mathbf{Dc}_\alpha$ ,  $\mathbf{Nr}_\alpha \mathbf{CA}_\beta$  and  $\mathbf{RCA}_\alpha$ <sup>70</sup>).

- (i)  $\mathbf{RCA}_\omega = \mathbf{SUPlf}_\omega \neq \mathbf{SPLf}_\omega$  and  $\mathbf{Dc}_\omega \subseteq \mathbf{RCA}_\omega$ . I.e. there is no universal formula distinguishing  $\mathbf{RCA}_\omega$  from  $\mathbf{Lf}_\omega$ , and every  $\mathbf{Dc}_\omega$  is representable. The same hold for all  $\alpha \geq \omega$  in place of  $\omega$ .
- (ii)  $\mathbf{RCA}_\alpha = \mathbf{SNr}_\alpha \mathbf{CA}_{\alpha+\omega} = \bigcap \{\mathbf{SNr}_\alpha \mathbf{CA}_{\alpha+m} : m \in \omega\} \neq \mathbf{SNr}_\alpha \mathbf{CA}_{\alpha+m}$  for all  $\alpha$  and for all finite  $m$ .  $\mathbf{SNr}_\alpha \mathbf{CA}_{\alpha+m}$  is a variety for all  $\alpha$  and  $m$ .

**Proof.** The positive statements follow from [Henkin, Monk and Tarski, 1985, 3.2.10, 2.6.32(ii), 2.6.50, 2.6.52, 3.2.11]. The negative statements are also proved in [Henkin, Monk and Tarski, 1985] taken together with [Henkin *et al.*, 1981]. ■

The above theorem gives information on the proof theory of first order logic (FOL) and on its  $n$ -variable fragment  $\mathbf{L}_n$ . Intuitively, it says (in several different forms) that the important feature of FOL is not that each formula involves only finitely many variables, but that given any formula, there are infinitely many variables it does not involve.<sup>71</sup>

Based on the above theorem, an inference system is given both for FOL and  $\mathbf{L}_n$  which uses the finite-schema axiomatization of  $\mathbf{CA}_\alpha$  together with a supply of variables which do not occur in our original formulas.<sup>72</sup> I.e. these variables can occur in a proof, but not in the final formula we want to prove.

<sup>70</sup>The classes  $\mathbf{SNr}_\alpha \mathbf{CA}_\beta$  were introduced by Henkin, and Monk [1961] proved that they are varieties. Theorem 28(i) is due to J. D. Monk. The equalities in (ii) are due to L. Henkin [1955], while the inequality in (ii) was proved by J. D. Monk [1969].

<sup>71</sup>The earlier mentioned theorem saying that quasi-projective RAs are representable, also speaks about this phenomenon: the projections are used for coding together the already involved variables, so that we get one more ‘unused’ variable. This idea comes through clearly in [Simon, 1996]. The same idea is used for proving finite schema axiomatizability (i.e. completeness of the corresponding logic) in [Sain, 1995], [Sain and Gyuris, 1994]. The same idea is used for obtaining an unorthodox completeness theorem in [Simon, 1991].

<sup>72</sup>Cf. e.g. [Henkin, Monk and Tarski, 1985, p.157], [Németi, 1996], [Andréka, Gergely and Németi, 1977, Thm. 3.15] and [Simon, 1991]. cf. also Section 7 herein.

## 3 ALGEBRAS FOR LOGICS WITHOUT IDENTITY

We start from  $\text{RCA}_\alpha$ , and would like to consider its ‘natural’ ld-free reducts. If we simply omit ld (or  $\text{ld}_{ij}$ ), then we lose not only identity (or equality), but also our ability to ‘algebraize’ substitution of individual variables like  $\varphi \mapsto \varphi[v_i/v_j]$  in the logic to be algebraized. Therefore, before dropping the  $\text{ld}_{ij}$ s, we first add our term functions  $s_j^i(x) = c_i(\text{ld}_{ij} \cap x)$  for  $i \neq j$ ,  $s_i^i(x) = x$ . Now,

$$\text{RSC}_\alpha \stackrel{\text{def}}{=} \mathbf{SP} \{ \langle \mathfrak{P}(\alpha U), c_i, s_j^i \rangle_{i,j < \alpha} : U \text{ is a set} \}.$$

$\text{RSC}_\alpha$ s are called *representable substitution-cylindric* algebras, cf. [Németi, 1991], [Andréka *et al.*, 1998]. They are the simplest kind in the family of polyadic-style algebras. The theory of  $\text{RSC}_\alpha$  is analogous with that of  $\text{RCA}_\alpha$ , in particular, if  $\alpha > 2$ , then  $\text{RSC}_\alpha$  is not finitely axiomatizable, cf. [Sain and Thompson, 1991].

Let  $\mathbf{Rd}_{sc}$  be the operator which associates to any cylindric-type algebra  $\mathfrak{A} = \langle A, \cup, -, c_i, \text{ld}_{ij} \rangle_{i,j < \alpha}$  the  $\text{RSC}_\alpha$ -type algebra

$$\mathbf{Rd}_{sc}(\mathfrak{A}) \stackrel{\text{def}}{=} \langle A, \cup, -, c_i, s_j^i \rangle_{i,j < \alpha}$$

where  $s_j^i$  is the derived operation of  $\mathfrak{A}$  defined above. Now,  $\text{RSC}_\alpha = \mathbf{SRd}_{sc} \text{RCA}_\alpha$ . The finitely axiomatizable approximation  $\text{SC}_\alpha$  of  $\text{RSC}_\alpha$  is defined analogously,

$$\text{SC}_\alpha \stackrel{\text{def}}{=} \mathbf{SRd}_{sc} \text{CA}_\alpha.$$

$\text{SC}_\alpha$ s are called *substitution-cylindric* algebras.

**THEOREM 29.**  *$\text{SC}_\alpha$  is a finite schema axiomatizable variety containing  $\text{RSC}_\alpha$ .* ■

For the simple set of axioms, and for information on the proof we refer the reader to [Andréka *et al.*, 1998]. Cf. also [Németi, 1991, §8].

The theory of the pair  $\text{SC}_\alpha, \text{RSC}_\alpha$  is almost completely analogous with that of  $\text{CA}_\alpha, \text{RCA}_\alpha$ . The connection between  $\text{SC}_\alpha$ -theory and first order logic without equality is analogous with the connection between  $\text{CA}_\alpha$ -theory and logic with equality. In particular, the logic counterpart of the algebraic operation  $s_j^i$  is the ‘substitution-modality’  $[v_i/v_j]$ , cf. Part II, Section 7. The logics  $\mathbf{L}_n^s$  and  $\mathbf{L}_n^{s=}$  are introduced in Section 7, Example 7. Let  $\mathbf{L}_n^z$  and  $\mathbf{L}_n^{z=}$  denote the fragments of these logics obtained by dropping the connective  $[v_i, v_j]$ . Then the  $\text{CA}_n, \text{RCA}_n$  pair is strongly connected to  $\mathbf{L}_n^{z=}$  while the  $\text{SC}_n, \text{RSC}_n$  pair is completely analogously connected to  $\mathbf{L}_n^z$ . More on the connection between  $\text{SC}_\alpha$  and logic can be found in [Sági and Németi, 1997; Marx and Venema, 1997; Venema, 1995a; van Benthem, 1996; Németi, 1991].

The classes  $CA_\alpha$ ,  $RCA_\alpha$  and  $SC_\alpha$ ,  $RSC_\alpha$  introduced so far constitute the hearts of the following two ‘worlds’: the algebraic counterpart of logics with equality (the ‘cylindric world’), and the algebraic counterpart of logics without equality (the ‘polyadic world’). In both worlds one can introduce natural extra operations like e.g. cardinality quantifiers, generalized cylindrifications, but what determines the most basic theorems (remaining true for the expanded algebras) remains the  $CA_\alpha$ -structure or the  $SC_\alpha$ -structure. Therefore it seems reasonable to pay somewhat more attention (say, as a default) to  $CA_\alpha$  and  $SC_\alpha$  than to their versions enriched with extra operators.<sup>73</sup>

$SC_\alpha$ s with extra operators (like e.g.  $[v_i, v_j]$ ) are discussed in the literature under the names *quasi-polyadic* algebras ( $QPA_\alpha$ s) and *polyadic* algebras respectively.<sup>74</sup> The most important extra operator  $p_{ij}$  is a substitution operator like  $s_j^i$  ( $p_{ij}$  corresponds to the logical connective  $[v_i, v_j]$  in Section 7, Example 7). Let  $X \subseteq {}^\alpha U$ . Then

$$p_{01}(X) \stackrel{\text{def}}{=} \{ \langle q_1, q_0, q_2, \dots \rangle : q \in X \}.$$

I.e.,  $p_{01}$  interchanges  $q_0$  and  $q_1$  in a sequence  $q$ . For  $i, j < \alpha$ ,  $p_{ij}$  is defined completely analogously. Now,  $RQPA_\alpha$ s are defined to be  $RSC_\alpha$ s enriched with the  $p_{ij}$ s ( $i, j < \alpha$ ):

$$RQPA_\alpha \stackrel{\text{def}}{=} \mathbf{SP} \{ \langle \mathfrak{B}({}^\alpha U), c_i, s_j^i, p_{ij} \rangle_{i, j < \alpha} : U \text{ is a set} \}.$$

The abstract class  $QPA_\alpha$  approximating  $RQPA_\alpha$  is defined by finitely many axiom-schemes analogously to the definition of  $CA_\alpha$  or  $SC_\alpha$ , cf. [Németi, 1991; Sain and Thompson, 1991; Andréka *et al.*, 1998].

*Polyadic* algebras ( $RPA_\alpha$  and  $PA_\alpha$ ) are obtained from  $RSC_\alpha$  and  $SC_\alpha$  by adding infinitary substitutions and infinitary cylindrifications denoted as  $s_\tau$ ,  $c_\Gamma$ , for  $\tau : \alpha \rightarrow \alpha$  and  $\Gamma \subseteq \alpha$ . For the theory of these algebras we refer to [Halmos, 1962; Henkin, Monk and Tarski, 1985; Németi, 1991]. Cf. also [Németi and Sági, to appear].

We note that the theory of  $QPAs$  seems to be very strongly analogous with that of  $SC_\alpha$ s. (However, in certain studies, e.g. when investigating the connection between  $RAs$  and  $CAs$ ,  $CAs$  enriched with the  $p_{ij}$ s play a very illuminating role (cf. [Németi and Simon, 1997]). The latter algebras are called  $QPAs$  with equality, or  $QPEAs$ . Cf. e.g. [Henkin, Monk and Tarski, 1985] for their theory.)

<sup>73</sup>Of course, no such rule is valid in general. Actually, pushing the above considerations further,  $SC_\alpha$ s seem to be at the heart of  $CA_\alpha$ -theory, too, therefore they could be considered as the core of (or basis for) the algebraizations of quantifier logics in general. This unified perspective for algebraic logic has not been elaborated yet.

<sup>74</sup> $PAs$ ,  $QPAs$ , and their versions with equality like  $QPEAs$ , originate with P. Halmos (cf. [Halmos, 1962]).  $SCs$  originate with C. Pinter, cf. e.g. [Andréka *et al.*, 1998] or [Németi, 1991].

For lack of space, we do not discuss further the theory of  $SC_\alpha$ s with extra operators (like QPAs, PAs, etc).

The next figure, taken from [Andréka, 1997] describes the interconnections between the operations  $c_i, ld_{ij}, s_j^i$  and  $p_{ij}$  (in the presence of the Boolean operations). (In the figure,  $ld_{ij}, s_j^i$  are denoted as  $d_{ij}, s_{ij}$ , respectively.) In Figure 10, nodes represent classes of algebras of relations where the units are Cartesian spaces  ${}^nU$  ( $3 \leq n < \omega$ ), and the operations are those along the path leading to the node. A broken edge between two nodes means that the second class is finitely axiomatizable over the first one, a bold edge means non-finite axiomatizability over, and a normal line means that it is unknown (to the authors) whether finite or non-finite axiomatizability holds.

## II: Bridge Between Logic and Algebra: Abstract Algebraic Logic

### INTRODUCTION TO PART II

Let us start by putting the subject matter of Part II of the present Paper into perspective.

The idea of solving problems in logic by first translating them to algebra, then using the well developed methodology of algebra for solving them, and then translating the solution back to logic, goes back to Leibnitz and Pascal. Papers on the history of Logic (e.g. [Anellis and Houser, 1991; Maddux, 1991]) point out that this method was fruitfully applied in the 19th century not only to propositional logics but also to quantifier logics (De Morgan, Peirce etc. applied it to quantifier logics too). The number of applications grew ever since. (Though some of these remained unnoticed, e.g. the celebrated Kripke–Lemmon completeness theorem for modal logic w.r.t. Kripke models was first proved by Jónsson and Tarski in 1948 using algebraic logic.)

For brevity, we will refer to the above method or procedure as ‘applying Algebraic Logic (AL) to Logic’. This expression might be somewhat misleading since AL itself happens to be a part of logic, and we do not intend to deny this. We will use the expression all the same, and hope, the reader will not misunderstand our intention.

In items (i) and (ii) below we describe two of the main motivations for applying AL to Logic.

(i) This is the more obvious one: When working with a relatively new kind of problem, it often proved to be useful to ‘transform’ the problem into a well understood and streamlined area of mathematics, solve the problem

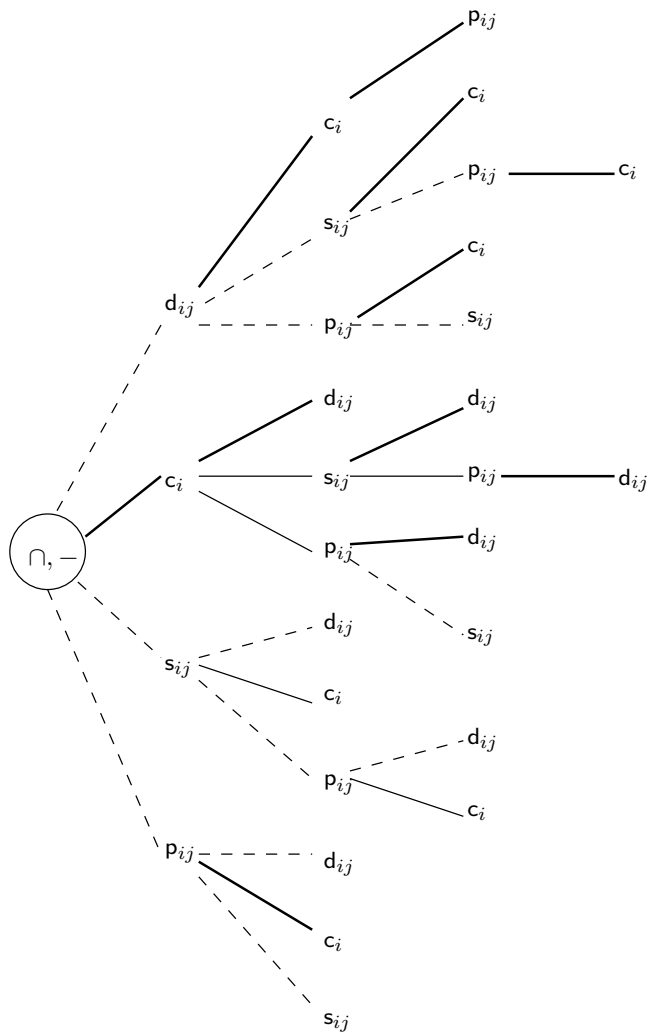


Figure 10.



there and translate the result back. Examples include the method of Laplace Transform in solving differential equations (a central tool in Electrical Engineering).<sup>75</sup>

In the present part we define the algebraic counterpart  $\mathbf{Alg}(\mathcal{L})$  of a logic  $\mathcal{L}$  together with the algebraic counterpart  $\mathbf{Alg}_m(\mathcal{L})$  of the semantical-model theoretical ingredients of  $\mathcal{L}$ . Then we prove equivalence theorems, which to essential logical properties of  $\mathcal{L}$  associate natural and well investigated properties of  $\mathbf{Alg}(\mathcal{L})$  such that if we want to decide whether  $\mathcal{L}$  has a certain property, we will know what to ask from our algebraician colleague about  $\mathbf{Alg}(\mathcal{L})$ . The same devices are suitable for finding out what one has to change in  $\mathcal{L}$  if we want to have a variant of  $\mathcal{L}$  having a desirable property (which  $\mathcal{L}$  lacks). To illustrate these applications we include several examples (which deal with various concrete logics) in Section 7. For all this, first we have to define what we understand by a logic  $\mathcal{L}$  in general (because otherwise it is impossible to define e.g. the function  $\mathbf{Alg}$  associating a class  $\mathbf{Alg}(\mathcal{L})$  of algebras to each logic  $\mathcal{L}$ ).

(ii) With the rapidly growing variety of applications of logic (in diverse areas like computer science, linguistics, AI, law, physics etc.) there is a growing number of new logics to be investigated. In this situation AL offers us a tool for economy and a tool for unification in various ways. One of these is that  $\mathbf{Alg}(\mathcal{L})$  is always a class of algebras, therefore we can apply the same machinery, namely universal algebra, to study all the new logics. In other words, we bring all the various logics to a kind of ‘normal form’ where they can be studied, compared, and even combined by uniform methods. Moreover, for most choices of  $\mathcal{L}$ ,  $\mathbf{Alg}(\mathcal{L})$  tends to appear in the same ‘area’ of universal algebra, hence specialized powerful methods lend themselves to studying  $\mathcal{L}$ . There is a fairly well understood ‘map’ available for the landscape of universal algebra. By using our algebraization process and equivalence theorems, we can project this ‘map’ back to the landscape of possible logics.

In Section 7, we will illustrate the above outlined ‘application of AL to logic’ by using the AL-results of Part I, as follows. In Part I, we studied various distinguished classes of algebras, like e.g.  $\mathbf{RCA}_n$ . Here, after studying the bridge ( $\mathcal{L} \mapsto \mathbf{Alg}(\mathcal{L})$  etc.) between the world of logics and that of algebras, we look up those distinguished logics to which the distinguished algebras of Part I belong. E.g. we will find a certain logic  $\mathbf{L}_n$  for which  $\mathbf{RCA}_n = \mathbf{Alg}(\mathbf{L}_n)$ . Then we will use results in Part I about  $\mathbf{RCA}_n$  to es-

---

<sup>75</sup>At this point we should dispell a misunderstanding: In certain circles of logicians there seems to be a belief that AL applies only to syntactical problems of logic and that semantical and model-theoretic problems are not treated by AL or at least not in their original model theoretic form. Nothing can be as far from the truth as this belief, as e.g. looking into the present part (i.e. Part II) should reveal. A variant of this belief is that the main bulk of AL is about offering a cheap pseudo semantics to Logics as a substitute for intuitive, model theoretic semantics. Again, this is very far from being true.

establish properties of  $\mathbf{L}_n$ . For this, we will use the ‘equivalence theorems’ established in Section 6 (of Part II). Besides  $\mathbf{RCA}_n$  and  $\mathbf{L}_n$ , a similar procedure will be applied to other distinguished classes of algebras (from Part I) and other distinguished logics.

The approach reported here is part of a broader, joint approach with W. J. Blok and D. L. Pigozzi outlined in [Andréka *et al.*, to appear]. The present part contains only a somewhat specialized version of that general approach, in order to suit the special needs of the present work. Besides [Andréka *et al.*, to appear], we refer to [Blok and Pigozzi, 1989; Blok and Pigozzi, 1991; Czelakowski, 1997; Pigozzi, 1991; Font and Jansana, 1994; Font and Jansana, 1997; Palasinska and Pigozzi, 1995; Czelakowski and Pigozzi, 1999], as well as [Andréka *et al.*, 1995]; [Henkin, Monk and Tarski, 1985, sections 5.6, 4.3]; [Andréka *et al.*, 1993; Németi and Andréka, 1994; Madarász, 1998; Hoogland, 1996; Mikulás, 1995] for the more general approach. The semantic aspect of this approach goes back to, e.g., [Andréka and Németi, 1975; Andréka, Gergely and Németi, 1977; Andréka and Sain, 1981].

#### 4 GENERAL FRAMEWORK FOR STUDYING LOGICS

DEFINITION 30 (Logic). By a *logic*  $\mathcal{L}$  we mean an ordered quadruple

$$\mathcal{L} \stackrel{\text{def}}{=} \langle F, \vdash, M, \models \rangle,$$

where (i)–(iv) below hold.

- (i)  $F$  (called the set of *formulas* of  $\mathcal{L}$ ) is a set of finite<sup>76</sup> sequences (called *words*) over some set  $X$  (called the *alphabet* of  $\mathcal{L}$ ).
- (ii)  $\vdash$  (called the *provability relation* of  $\mathcal{L}$ ) is a relation between sets of formulas and formulas, that is,  $\vdash \subseteq \mathcal{P}(F) \times F$ . Following tradition,<sup>77</sup> instead of ‘ $\langle \Sigma, \varphi \rangle \in \vdash$ ’ we write ‘ $\Sigma \vdash \varphi$ ’.
- (iii)  $M$  is a class<sup>78</sup> (called the class of *models* of  $\mathcal{L}$ ).

<sup>76</sup>With this we exclude infinitary languages like  $\mathcal{L}_{\kappa, \lambda}$ ,  $\mathcal{L}_{\infty, \omega}^n$ . This exclusion is not necessary, all the methods go through with some modifications. Actually, occasionally we will look into properties of the finite variable fragment  $\mathcal{L}_{\infty, \omega}^n$  of infinitary logic, because it naturally admits applications of our methods and plays an essential role in finite model theory and in theoretical computer science.

<sup>77</sup>This tradition is used for all binary relations: if  $R$  is a binary relation, then instead of ‘ $\langle a, b \rangle \in R$ ’ we sometimes write ‘ $a R b$ ’.

<sup>78</sup>Although it is not automatically permitted in the ‘most official’ version of set theory (ZF), we *may* assume that for any four classes  $M_1, \dots, M_4$  the tuple ‘ $\langle M_1, \dots, M_4 \rangle$ ’ exists and is again a class. This does not lead to set theoretical paradoxes. What one should avoid is assuming that the collection of *all* classes would (exist and) form a class again

- (iv)  $\models$  (called the *validity relation*) is a relation between  $M$  and  $F$  that is,  $\models \subseteq M \times F$ . Instead of ' $\langle \mathfrak{M}, \varphi \rangle \in \models$ ' we write ' $\mathfrak{M} \models \varphi$ '.

If  $\mathcal{L}$  is a logic, then by  $F_{\mathcal{L}}$ ,  $\vdash_{\mathcal{L}}$ ,  $M_{\mathcal{L}}$ ,  $\models_{\mathcal{L}}$  we denote its corresponding parts.

Intuitively,  $F$  is the collection of 'texts' or 'sentences' or 'formulas' that can be 'said' in the language  $\mathcal{L}$ . For  $\Gamma \subseteq F$  and  $\varphi \in F$ , the intuitive meaning of  $\Gamma \vdash \varphi$  is that  $\varphi$  is provable (or derivable) from  $\Gamma$  with the syntactic inference system (or deductive mechanism) of  $\mathcal{L}$ . In all important cases,  $\vdash$  is subject to certain (well-known) conditions like  $\Gamma \vdash \varphi$  and  $\Gamma \cup \{\varphi\} \vdash \psi$  imply  $\Gamma \vdash \psi$  for any  $\Gamma \subseteq F$  and  $\varphi, \psi \in F$ . The class  $M$  of models is understood in the spirit of model theory: The models  $\mathfrak{M} \in M$  of  $\mathcal{L}$  are thought of as 'possible environments' or 'possible interpretations' or 'possible worlds', cf. [Allén, 1989]. Here a possible world is *not* the same as the technical devices called possible worlds in a Kripke model. The validity relation tells us which texts are 'true' in which possible environments (or worlds or models) under what conditions. Usually  $F$  and  $\vdash$  are defined by what are called grammars in mathematical linguistics.  $\langle F, \vdash \rangle$  together with the grammar defining them is called the *syntactical part* of  $\mathcal{L}$ , while  $\langle M_{\mathcal{L}}, \models \rangle$  is the *semantical part or model theoretical part* of  $\mathcal{L}$ .<sup>79</sup>

As a binary relation between  $M$  and  $F$ ,  $\models$  induces a Galois-connection between  $M$  and  $F$ , and in particular, it defines two closure operators, one on  $M$  and one on  $F$ . Next we collect some of the relevant definitions.

**DEFINITION 31** (Theory of a class of models, models of a set of formulas, semantical consequence, validity).

---

(and variants of this). For more on this cf. [Henkin, Monk and Tarski, 1971, p.34, first 10 lines], [Henkin, Monk and Tarski, 1971, p.25], [Adámek, Herrlich and Strecker, 1990, §2, pp.5-8], or almost any work on abstract model theory.

<sup>79</sup>Cf. Sections 14, 15 of [Gabbay, 1996b] for more intuitive motivation on how and why these parts are highlighted in a logic.

At this point a natural objection suggests itself: Why is  $M_{\mathcal{L}}$  an arbitrary class? Why did we not assume (like in [Barwise and Feferman, 1985]) that  $M_{\mathcal{L}}$  is a class of first order structures or of algebraic systems? The answer is (i)-(iii) below. (i) In institutions theory they do the same what we do and for the same reasons. Cf. the subsection 'Connections with the literature' at the very end of the present Section. (ii) We are developing a general theory, and we do not know in advance what kinds of structures will be the models of our  $\mathcal{L}$ . E.g. they may be classical first order models, they may or may not have a topological structure too, they may be propositional Kripke-models, they may have infinitary relations on them (cf. [Henkin, Monk and Tarski, 1971, §4.3]), they may be models of intensional logic in the sense of Montague etc. Therefore, at the very beginning, we do not want to commit ourselves on some actual kinds of mathematical objects that exactly should be the elements of  $M_{\mathcal{L}}$ . (iii) All the same, during the development of our theory, we will impose *some* structure on  $M_{\mathcal{L}}$  (but only gradually). This structure-imposing process is carried even further in [Andréka *et al.*, to appear], cf. e.g. 'concrete semantical systems' therein.

(i) Let  $K \subseteq M$  and  $\Sigma \subseteq F$ . Then

$$K \models_{\mathcal{L}} \Sigma \quad \text{iff} \quad (\forall \mathfrak{M} \in K)(\forall \varphi \in \Sigma) \mathfrak{M} \models_{\mathcal{L}} \varphi.$$

We will write  $K \models_{\mathcal{L}} \varphi$  in place of  $K \models_{\mathcal{L}} \{\varphi\}$  and  $\mathfrak{M} \models_{\mathcal{L}} \Sigma$  in place of  $\{\mathfrak{M}\} \models_{\mathcal{L}} \Sigma$ .

$\text{Th}_{\mathcal{L}}(K) \stackrel{\text{def}}{=} \{\varphi \in F : K \models_{\mathcal{L}} \varphi\}$ ,  $\text{Th}_{\mathcal{L}}(K)$  is called the *theory* of  $K$ , and  $\text{Mod}_{\mathcal{L}}(\Sigma) \stackrel{\text{def}}{=} \{\mathfrak{M} \in M : \mathfrak{M} \models_{\mathcal{L}} \Sigma\}$ ,  $\text{Mod}_{\mathcal{L}}(\Sigma)$  is called the *class of models* of  $\Sigma$ .

(ii) Semantical consequence, valid formulas: Let  $\Sigma \cup \{\varphi\} \subseteq F$ . Then

$$\Sigma \vDash_{\mathcal{L}} \varphi \quad \text{iff} \quad \text{Mod}_{\mathcal{L}}(\Sigma) \models_{\mathcal{L}} \varphi.$$

We read  $\Sigma \vDash_{\mathcal{L}} \varphi$  as:  $\varphi$  is a *semantical consequence* of  $\Sigma$ . In case of a singleton  $\{\psi\}$  of formulas we write  $\psi \vDash_{\mathcal{L}} \varphi$  in place of  $\{\psi\} \vDash_{\mathcal{L}} \varphi$  for simplicity.<sup>80</sup>

$\vDash_{\mathcal{L}} \varphi$  iff  $M \models_{\mathcal{L}} \varphi$ . In this case we say that  $\varphi$  is a *valid formula* of  $\mathcal{L}$ .

(iii) Axiomatizable classes of models:

$\text{Mod}_{\mathcal{L}} \text{Th}_{\mathcal{L}}(K)$  is called the *axiomatizable hull* of  $K$ .  $K$  is *axiomatizable* iff  $K = \text{Mod}_{\mathcal{L}}(\Sigma)$  for some  $\Sigma \subseteq F$ . In this case we also say that  $\Sigma$  describes or defines  $K$ .

(iv) Provability, or derivability:

$\vdash_{\mathcal{L}} \varphi$  iff  $\emptyset \vdash_{\mathcal{L}} \varphi$ , in this case we say that  $\varphi$  is *provable* or *derivable* in  $\mathcal{L}$ . If  $\Sigma \vdash_{\mathcal{L}} \varphi$ , then we say that  $\varphi$  is *provable from*  $\Sigma$  (in  $\mathcal{L}$ ).

If there is no danger of confusion, we will omit the subscript  $\mathcal{L}$  from  $\vDash_{\mathcal{L}}$ ,  $\text{Th}_{\mathcal{L}}$ ,  $\text{Mod}_{\mathcal{L}}$  etc.

REMARK .  $\text{ThMod}$  and  $\text{ModTh}$  are the two closure operators induced by  $\vDash$ . The semantical consequence relation  $\vDash$  is a binary relation between  $\mathcal{P}(F)$  and  $F$ , just like  $\vdash$  is. To treat  $\vdash$  and  $\vDash$  uniformly, in some places a logical system is defined to be  $\langle F, |\equiv \rangle$  where  $|\equiv \subseteq \mathcal{P}(F) \times F$ . E.g. [Blok and Pigozzi, 1989] uses this definition. In this notion,  $|\equiv$  can mean either the derivability relation or the semantical consequence relation  $\vDash$ . Connections between our conception of a logic and the rest of the literature will be discussed at the end of this section.

---

<sup>80</sup>In the literature of logic, in place of our symbol  $\vDash$ , most often the same symbol is used as the one denoting the validity relation ( $\models$ ). Our choice of using two different symbols for denoting these two (though different but not independent) relations comes from the nature of our Paper (meta-logical considerations).

The definition of a logic in Definition 30 is very broad. Actually, it is too broad for proving interesting theorems about logics. Now we will define a subclass of logics which we will call *algebraizable semantical logics*. The notion of an algebraizable logic is broad enough to cover a very large part of the logics investigated in the literature.<sup>81</sup> On the other hand, the class of algebraizable logics is narrow enough for proving interesting theorems about such logics, that is, we will be able to establish typical logical facts that hold for most logics studied in the literature.

Below, in Definitions 32–39, we collect some common features of logics.

We will discuss the usual extra assumptions one usually makes about a logic  $\mathcal{L}$  in the following order. First we discuss (assumptions on) the distinguished parts of  $\mathcal{L}$  beginning with  $F_{\mathcal{L}}$  and ending with  $\models_{\mathcal{L}}$ . Then in Definition 39, we will discuss (assumptions on) how these parts are put together. Often, what we call ‘extra assumptions’ here will also imply ‘extra structure’.

The set  $F$  of formulas is usually defined by fixing a set  $Cn$  of logical connectives and a set  $P$  of atomic formulas:

DEFINITION 32 ( $\mathcal{L}$  has connectives).

(i) Assume that two sets,  $P$  and  $Cn$  are given, such that every element of  $Cn$  has a finite rank. Then  $F(P, Cn)$  denotes the smallest set  $H$  satisfying (1),(2) below:

- (1)  $P \subseteq H$ , and
- (2) for every  $c \in Cn$  of rank  $k$  and  $\varphi_1, \dots, \varphi_k \in H$ ,  $c(\varphi_1, \dots, \varphi_k) \in H$ .

Note that  $F(P, Cn)$  is the universe of the word-algebra of type  $Cn$  generated by  $P$ .

(ii) We say that  $F_{\mathcal{L}}$  is given by  $\langle P, Cn \rangle$  if  $F_{\mathcal{L}} = F(P, Cn)$ . In this case we say that  $P$  is the set of all *atomic formulas* or *atomic propositions* of  $\mathcal{L}$ , and  $Cn$  is the set of all *logical connectives* of  $\mathcal{L}$ .  $P$  is also called the vocabulary of  $\mathcal{L}$ .<sup>82</sup> The word-algebra generated by  $P$  and using the logical connectives of  $Cn$  as algebraic operations is denoted by  $\mathfrak{F}$ , and is called the *formula algebra* of  $\mathcal{L}$ . Note that  $\mathfrak{F} = \langle F, c^{\mathfrak{F}} \rangle_{c \in Cn}$  where  $c^{\mathfrak{F}}(\varphi_1, \dots, \varphi_k) \stackrel{\text{def}}{=} c(\varphi_1, \dots, \varphi_k) \in F$  for all  $\varphi_1, \dots, \varphi_k \in F$  and  $k$ -ary connective  $c \in Cn$ .

(iii) We say that  $\mathcal{L}$  has connectives if  $F_{\mathcal{L}}$  is given by  $\langle P, Cn \rangle$  for some sets  $P, Cn$ . In this case we usually assume that  $\langle P, Cn \rangle$  is given together with  $\mathcal{L}$ .

---

<sup>81</sup>Moreover, in Remark 40 we will indicate how to extend the methods of the present work from algebraizable logics to a broader class called protoalgebraic semantical logics. The latter is really broad enough to cover most logics in the literature.

<sup>82</sup>Sometimes, informally,  $P$  is also called the set of ‘propositional variables’. Here we emphasize that the connotations of this name are misleading.

Next we turn to *inference systems*  $\vdash_{\mathcal{L}}$ . Inference systems (usually denoted as  $\vdash$ ) are syntactical devices serving to recapture (or at least to approximate) the semantical consequence relation of the logic  $\mathcal{L}$ . The idea is the following. Suppose  $\Sigma \stackrel{\bullet}{\models} \varphi$ . This means that, in the logic  $\mathcal{L}$ , the assumptions collected in  $\Sigma$  semantically imply the conclusion  $\varphi$ . (In any possible world or model  $\mathfrak{M}$  of  $\mathcal{L}$  whenever  $\Sigma$  is valid in  $\mathfrak{M}$ , then also  $\varphi$  is valid in  $\mathfrak{M}$ .) Then we would like to be able to reproduce this relationship between  $\Sigma$  and  $\varphi$  by purely syntactical, ‘finitistic’ means. That is, by applying some formal rules of inference (and some axioms of the logic  $\mathcal{L}$ ) we would like to be able to derive  $\varphi$  from  $\Sigma$  by using ‘paper and pencil’ only. In particular, such a derivation will always be a finite string of symbols. If we can do this, that will be denoted by  $\Sigma \vdash \varphi$ .

Inference systems are usually given by *axioms* and *inference* rules. These axioms and rules use formula-schemes in place of concrete formulas. A formula-scheme is just like a formula, the only difference is that it is built up from formula-variables (i.e. meta-variables ranging over formulas) in place of atomic formulas.

**DEFINITION 33** (Formula-scheme, Hilbert-style inference system).

(i) Assume that  $F_{\mathcal{L}}$  is given by  $\langle P, Cn \rangle$ . We will call  $\phi_i$  ( $i < \omega$ ) *formula-variables*, and  $FV$  will denote the set of all formula-variables, i.e.  $FV = \{\phi_i : i < \omega\}$ . The elements of  $F(FV, Cn)$  are called *formula-schemes*, and  $Fs$  will denote the set of all formula-schemes of  $\mathcal{L}$ . I.e.  $Fs \stackrel{\text{def}}{=} F(FV, Cn)$ . An *instance* of a formula-scheme is obtained by substituting formulas for the formula variables in it. A formula-scheme is called *valid* if all of its instances are valid.

(ii) An *inference-rule* for  $\mathcal{L}$  is a pair  $\langle \langle \Phi_1, \dots, \Phi_k \rangle, \Phi_0 \rangle$ , where every  $\Phi_i$  ( $i \leq k$ ) is a formula scheme of  $\mathcal{L}$ . This inference rule will be denoted by

$$\frac{\Phi_1, \dots, \Phi_k}{\Phi_0}.$$

An *instance* of an *inference rule* is obtained by substituting formulas for the formula variables in the formula schemes occurring in the rule. An inference rule  $\langle \langle \Phi_1, \dots, \Phi_k \rangle, \Phi_0 \rangle$  is called *valid* if  $\{\varphi_1, \dots, \varphi_k\} \stackrel{\bullet}{\models} \varphi_0$  for all instances  $\langle \langle \varphi_1, \dots, \varphi_k \rangle, \varphi_0 \rangle$  of it. Valid inference rules are also called *admissible rules* or strongly sound rules in the literature.

(iii) A *Hilbert-style inference system* (or *calculus*) for  $\mathcal{L}$  is a pair  $\langle Ax, Ru \rangle$  where  $Ax$  is a finite set of formula-schemes, called *axioms*, and  $Ru$  is a finite set of inference rules for  $\mathcal{L}$ .

(iv) A Hilbert-style inference system  $I = \langle Ax, Ru \rangle$  defines a provability or derivability relation  $\vdash$  as follows. Assume  $\Sigma \cup \{\varphi\} \subseteq F$ . We say that  $\varphi$  is *derivable* (or  *$\vdash$ -provable*) from  $\Sigma$  iff there is a finite sequence  $\langle \varphi_1, \dots, \varphi_n \rangle$  of formulas (an  *$\vdash$ -proof of  $\varphi$  from  $\Sigma$* ) such that  $\varphi_n$  is  $\varphi$  and for every  $1 \leq i \leq n$

- $\varphi_i \in \Sigma$  or
- $\varphi_i$  is an instance of an axiom scheme of  $I$  or
- there are  $j_1, \dots, j_k < i$ , and there is an inference rule of  $I$  such that  $\frac{\varphi_{j_1} \dots \varphi_{j_k}}{\varphi_i}$  is an instance of this rule.

We write  $\Sigma \vdash \varphi$  if  $\varphi$  is  $\vdash$ -provable from  $\Sigma$ . Now  $\vdash = \{ \langle \Sigma, \varphi \rangle : \Sigma \vdash \varphi \}$ . We say that  $\vdash$  is given by  $\langle Ax, Ru \rangle$ . Throughout, we identify  $I$  with  $\vdash$ , e.g. we say that  $\varphi$  is an axiom of  $\vdash$ .

Next we turn to the semantical part  $M_{\mathcal{L}}, \models_{\mathcal{L}}$  of  $\mathcal{L}$ . Validity of formulas in models, i.e.  $\models_{\mathcal{L}}$ , is usually defined indirectly by first defining something more basic, namely the *meanings* or denotations of formulas (and of other kinds of syntactic entities belonging to the language) in models. The idea is that the meaning of some syntactic entity (like a noun-phrase, or a sentence) need not always be a truth value. Therefore first we define a so-called meaning function which to each syntactic entity  $\varphi$  and each model  $\mathfrak{M}$  associates some semantic entity  $mng(\varphi, \mathfrak{M})$  called the meaning of  $\varphi$  in  $\mathfrak{M}$ . After knowing what the particular syntactic entities mean in the models, one may be able to derive information concerning which sentences are true or valid in which models.

DEFINITION 34 (Meaning function, compositionality).

(i) Let  $mng$  be any function mapping  $F \times M$  into a class; and let us call  $mng(\varphi, \mathfrak{M})$  the meaning of  $\varphi$  in  $\mathfrak{M}$ . For a fixed  $\mathfrak{M} \in M$ , the function  $mng_{\mathfrak{M}}$  mapping  $F$  to the set of meanings<sup>83</sup> is defined by letting for all  $\varphi \in F$

$$mng_{\mathfrak{M}}(\varphi) \stackrel{\text{def}}{=} mng(\varphi, \mathfrak{M}).$$

We say that  $mng$  is a *meaning-function* for  $\mathcal{L}$  if the validity of a formula depends only on its meaning, i.e., if (\*) below holds:

$$(*) \quad mng_{\mathfrak{M}}(\varphi) = mng_{\mathfrak{M}}(\psi) \implies (\mathfrak{M} \models \varphi \text{ iff } \mathfrak{M} \models \psi).$$

(ii) Assume that  $\mathcal{L}$  has connectives. We say that the meaning-function  $mng$  is *compositional* if the meanings of formulas are built up from the meanings of their subformulas. More precisely,  $mng$  is compositional if,

<sup>83</sup>We use the word ‘meaning’ in the sense Frege used ‘intension’ or ‘sense’. It is important to emphasize that ‘meaning’ is much more general than ‘extension’ or truthvalue (though for some logics the two may coincide). Further denotation (e.g. in [Partee, 1976]) can be identified with what we call ‘meaning’. As an example, let  $\mathcal{L}$  be Richard Montague’s intensional logic. Then  $mng(\varphi, \mathfrak{M})$  is exactly what Montague calls the intension of  $\varphi$  in  $\mathfrak{M}$ . See also [Partee, 1976, pp. 1–18] on meaning(s). It seems that we are using the word ‘meaning’ in the same sense as [Janssen, 1997, pp. 419–470] does. What we call the meaning function  $mng_{\mathfrak{M}}$  is called a meaning assignment in [Janssen, 1997, p. 423].

for each  $\mathfrak{M}$ ,  $mng_{\mathfrak{M}}$  is a homomorphism, or equivalently, if  $ker(mng_{\mathfrak{M}}) = \{(a, b) : mng_{\mathfrak{M}}(a) = mng_{\mathfrak{M}}(b)\}$  is a congruence relation on the formula algebra. Thus  $mng$  is compositional iff the (congruence) condition below is satisfied for all  $k$ -ary connective  $c \in Cn$  and  $\varphi_i, \psi_i \in F$ ,  $1 \leq i \leq k$ :

$$\bigwedge_{i=1}^k mng_{\mathfrak{M}}(\varphi_i) = mng_{\mathfrak{M}}(\psi_i) \implies mng_{\mathfrak{M}}(c(\varphi_1, \dots, \varphi_k)) = mng_{\mathfrak{M}}(c(\psi_1, \dots, \psi_k)).$$

We say that  $\mathcal{L}$  is *compositional* if it has connectives and a compositional meaning-function (w.r.t. the connectives of  $\mathcal{L}$ ). This property is traditionally called *Frege's principle of compositionality*.

(iii) From now on, by a logic  $\mathcal{L}$  we understand a logic with a meaning-function, i.e.  $\mathcal{L} = \langle F, \vdash, M, mng, \models \rangle$ , where  $mng$  is a meaning-function for the rest of  $\mathcal{L}$ .

On the ingredients of a logic: Let  $\mathcal{L} = \langle \dots mng, \models \rangle$  be a logic. Then, using the terminology of Frege, Carnap, Montague as in [van Benthem and ter Meulen, 1997],  $mng$  represents the *intensional* (or denotational) aspects of semantics, while  $\models$  represents the *extensional* (or truth-value oriented) aspect of semantics, in  $\mathcal{L}$ . Therefore those approaches to general logics in which  $mng$  is repressed, seem to be extensionally oriented (cf. e.g. institutions theory e.g. in [Gabbay, 1994, p. 359]), while the ones emphasizing  $mng$  (e.g. [Andréka and Sain, 1981; Henkin, Monk and Tarski, 1985; Epstein, 1990] and [Sain, 1979]) seem to be intensions-oriented (or sense or denotations oriented). The latter intuition is reflected in the fact that [Gabbay, 1994] traces the elements of the class  $\bigcup \{mng_{\mathfrak{M}}(\varphi) : \varphi \in F \text{ and } \mathfrak{M} \in M\}$  as well, and calls them *basic semantical units*.

In many logics (cf. e.g. sentential logics  $\mathbf{L}_S$ ,  $\mathbf{L}_S'$ , modal logic S5 in Section 7) we have a derived connective  $\leftrightarrow$  and a formula denoted as *True* which establish a strong connection between  $mng$  and  $\models$ , namely

- (i)  $mng_{\mathfrak{M}}(\varphi) = mng_{\mathfrak{M}}(\psi)$  iff  $\mathfrak{M} \models \varphi \leftrightarrow \psi$ , and
- (ii)  $\mathfrak{M} \models \varphi$  iff  $mng_{\mathfrak{M}}(\varphi) = mng_{\mathfrak{M}}(True)$ .

In these logics there is a strong connection between  $Th(\mathfrak{M}) \subseteq F$  and the kernel of the meaning function  $ker(mng_{\mathfrak{M}}) \subseteq F \times F$ , namely the kernel of  $mng_{\mathfrak{M}}$  and  $Th(\mathfrak{M})$  are recoverable from each other.<sup>84</sup> We will say that  $\mathcal{L}$  has the *filter-property* iff there are derived connectives that generalize the above situation, as follows.

---

<sup>84</sup>If  $f$  is an arbitrary function, then  $ker(f) \stackrel{\text{def}}{=} \{(a, b) : f(a) = f(b)\}$ .



DEFINITION 35 ( $\mathcal{L}$  has the filter-property).

(i) A  $k$ -ary *derived connective* is a formula-scheme  $\Delta \in Fs$  using the formula variables  $\phi_0, \dots, \phi_{k-1}$  only. If  $\varphi_0, \dots, \varphi_{k-1}$  are formulas, then  $\Delta(\varphi_0, \dots, \varphi_{k-1})$  denotes the instance of  $\Delta$  when we replace  $\phi_0, \dots, \phi_{k-1}$  by  $\varphi_0, \dots, \varphi_{k-1}$  respectively.<sup>85</sup>

(ii) We say that  $\mathcal{L}$  has the *filter-property* iff there exist unary derived connectives  $\varepsilon_0, \dots, \varepsilon_{m-1}$  and  $\delta_0, \dots, \delta_{m-1}$  and binary ones  $\Delta_0, \dots, \Delta_{n-1}$  ( $m, n \in \omega$ ) of  $\mathcal{L}$  such that for all  $\varphi, \psi \in F$  and for all  $\mathfrak{M} \in M$ , properties (1) and (2) below hold.

$$(1) \quad mng_{\mathfrak{M}}(\varphi) = mng_{\mathfrak{M}}(\psi) \iff (\forall i < n)(\mathfrak{M} \models \varphi \Delta_i \psi).$$

$$(2) \quad \mathfrak{M} \models \varphi \iff (\forall j < m)(mng_{\mathfrak{M}}(\varepsilon_j(\varphi)) = mng_{\mathfrak{M}}(\delta_j(\varphi))).$$

In the case of classical logic, we can choose the above derived connectives such that  $\Delta$  is ‘ $\leftrightarrow$ ’,  $\varepsilon(\varphi)$  is  $\varphi$ , and  $\delta(\varphi)$  is ‘True’.

We consider one more basic feature of logics: substitution-invariance properties. Roughly speaking, the idea is that substituting a part (subformula) of a formula with another one which is “equivalent” to the original part should not change the validity or meaning of the formula. In this respect we usually expect a logic to be *substitutional* in the sense of Definition 36 below. If it is not, then we rather treat it as a ‘theory’ of a substitutional logic. For such examples see Section 7 (Examples) or [Andréka *et al.*, 1993].

DEFINITION 36 ( $\mathcal{L}$  has the substitution property).

(i) By a *substitution*  $s$  we understand a function  $s : P \rightarrow F$  (we will ‘substitute’  $p \in P$  with  $s(p) \in F$ ). If  $\varphi \in F$ , then  $\varphi(\overline{p}/s(\overline{p}))$  denotes the formula we obtain from  $\varphi$  after simultaneously substituting  $s(p)$  for every occurrence of  $p$ , for all  $p \in P$  in  $\varphi$ . In other words,  $\varphi(\overline{p}/s(\overline{p})) \stackrel{\text{def}}{=} \hat{s}(\varphi)$ , where  $\hat{s}$  is the (unique) extension of  $s : P \rightarrow F$  to a homomorphism  $\hat{s} : \mathfrak{F} \rightarrow \mathfrak{F}$ .<sup>86</sup>

(ii)  $\mathcal{L}$  has the (*syntactic*) *substitution property* (or  $\mathcal{L}$  is *substitutional*) iff for any formula  $\varphi \in F$  and substitution  $s : P \rightarrow F$

$$\models \varphi \quad \text{implies} \quad \models \varphi(\overline{p}/s(\overline{p})).$$

This means that a formula of  $\mathcal{L}$  is valid iff the corresponding formula scheme of  $\mathcal{L}$  is valid (where we get the corresponding formula scheme by substituting atomic formulas  $p_i \in P$  with formula variables  $\phi_i \in FV$ ).

<sup>85</sup>The algebraic counterpart of ‘derived connective’ is ‘term function’. If  $\Delta$  is binary, then we will write  $\varphi \Delta \psi$  in place of  $\Delta(\varphi, \psi)$ .

<sup>86</sup>Such a unique extension exists because  $\mathfrak{F}$  is the word-algebra generated by  $P$ , i.e., in algebraic terms, it is freely generated by  $P$ .

(iii)  $\mathcal{L}$  has the *semantical substitution property* iff for any model  $\mathfrak{M} \in M$  and substitution  $s : P \rightarrow F$  there is another model  $\mathfrak{N} \in M$  such that

$$mng_{\mathfrak{N}}(p) = mng_{\mathfrak{M}}(s(p)) \quad \text{for all } p \in P.$$

Intuitively, the model  $\mathfrak{N}$  is the substituted version of  $\mathfrak{M}$  along  $s$ .

The semantical substitution property says that the atomic formulas can have the meanings of any other formulas. (This statement will be made precise in Proposition 43.) Examples where we have and do not have this property are given in Section 7.

**PROPOSITION 37.** If a logic  $\mathcal{L}$  has the filter-property and the semantical substitution property, then it has the syntactic substitution property, too.

**Proof.** We give a proof only for that simple case when the filter property is realized by the derived connectives  $\leftrightarrow$  and *True*. Actually we will use only the following one of the two filter-property conditions:

$$(**) \quad \mathfrak{M} \models \varphi \quad \text{iff} \quad mng_{\mathfrak{M}}(\varphi) = mng_{\mathfrak{M}}(\textit{True}) \quad \text{for every } \varphi \text{ and } \mathfrak{M}.$$

Assume that  $\mathcal{L}$  satisfies condition  $(**)$  and has the semantical substitution property. Let  $\mathfrak{M} \in M_{\mathcal{L}}$  and  $s : P \rightarrow F_{\mathcal{L}}$  be arbitrary but fixed. To prove the syntactic substitution property, assume that  $\models \varphi$  for some arbitrary but fixed  $\varphi \in F_{\mathcal{L}}$ . By the semantical substitution property of  $\mathcal{L}$ , there exists a model  $\mathfrak{N} \in M_{\mathcal{L}}$  such that  $mng_{\mathfrak{N}}(p) = mng_{\mathfrak{M}}(s(p))$  for all  $p \in P$ . Using formula induction, it is easy to check that

$$(\dagger) \quad mng_{\mathfrak{N}}(\psi) = mng_{\mathfrak{M}}(\hat{s}(\psi)) \quad \text{for every } \psi \in F_{\mathcal{L}}.$$

Now

$$\begin{aligned} mng_{\mathfrak{M}}(\textit{True}) &= mng_{\mathfrak{M}}(\hat{s}(\textit{True})) && \text{because } \hat{s} \text{ is a homomorphism and} \\ & && \textit{True} \text{ is a constant connective} \\ &= mng_{\mathfrak{N}}(\textit{True}) && \text{by } (\dagger) \\ &= mng_{\mathfrak{N}}(\varphi) && \text{because } \mathfrak{N} \models \varphi \text{ (by } \models \varphi \text{) and } (**) \\ &= mng_{\mathfrak{M}}(\hat{s}(\varphi)) && \text{by } (\dagger). \end{aligned}$$

Then  $\mathfrak{M} \models \hat{s}(\varphi)$  by  $(**)$  again (in the other direction). Thus  $\models \hat{s}(\varphi)$  since  $\mathfrak{M}$  was chosen arbitrarily; which proves the syntactic substitution property, since  $s$  and  $\varphi$  were chosen arbitrarily as well.  $\blacksquare$

Thus, a ‘*fully-fledged*’ logic  $\mathcal{L} = \langle F, \vdash, M, mng, \models \rangle$  sometimes is given as  $\mathcal{L} = \langle \langle P, Cn \rangle, \langle Ax, Ru \rangle, M, mng, \models \rangle$ . Often, not all parts of a logic are given. Sometimes we have only  $\langle F, \vdash \rangle$  and we are searching for a ‘semantics’  $\langle M, mng, \models \rangle$  for it such that e.g.  $\langle F, \vdash, M, mng, \models \rangle$  is complete.<sup>87</sup> Or, even more often, we have  $\langle F, M, mng, \models \rangle$  and we are searching for a provability relation  $\vdash$  such that  $\langle F, \vdash, M, mng, \models \rangle$  would be complete. Sometimes  $\langle F, \vdash \rangle$

<sup>87</sup>Completeness of a logic will be defined in Definition 48 in Section 6.

is called a ‘*deductive logic*’ (or *syntactic one*), while  $\langle F, M, mng, \models \rangle$  is called a ‘*semantic logic*’ (cf. e.g. [Andréka *et al.*, to appear]). (Though, one should keep in mind that  $\langle F, \models \rangle$  is a ‘*deductive logic*’ in this sense.) From now on we will often omit some parts of a logic. Most often we will deal with  $\langle F, M, mng, \models \rangle$  and we will say that  $\mathcal{L} = \langle F, M, mng, \models \rangle$  is a logic or more carefully, a *semantical logic*. Most of the notions are meaningful for it, e.g. that  $\mathcal{L}$  is *compositional*, etc.

DEFINITION 38. (Algebraizable semantical logic, structural logic) Let  $\mathcal{L} = \langle F, M, mng, \models \rangle$  be a logic in the above sense.

- We say that  $\mathcal{L}$  is *structural* if  $\mathcal{L}$  is compositional and has the semantical substitution property.
- We say that  $\mathcal{L}$  is an *algebraizable*<sup>88</sup> semantical logic if  $\mathcal{L}$  is structural and has the filter-property.

In most cases, the set  $P$  of atomic formulas is a parameter in the definition of the logic  $\mathcal{L}$ . Namely,  $P$  is a fixed but *arbitrary* set. So in a sense,  $\mathcal{L}$  is a function of  $P$ , and we could write  $\mathcal{L}^P$  (instead of  $\mathcal{L}$ ) to make this explicit. Most often the choice of  $P$  has only limited influence on the behaviour of  $\mathcal{L}$ . However, we will have to remember that  $P$  is a freely chosen parameter because in certain investigations, the choice of  $P$  does influence the behaviour of  $\mathcal{L}^P$ .

DEFINITION 39. (General logic, algebraizable general logic)

- (i) A *general logic* is a function (or indexed family)

$$\mathbf{L} \stackrel{\text{def}}{=} \langle \mathcal{L}^P : P \text{ is a set} \rangle,$$

where for each set  $P$ ,  $\mathcal{L}^P = \langle F^P, M^P, mng^P, \models^P \rangle$  is a logic in the above sense.

(ii) We say that  $\mathbf{L}$  *has connectives* iff there is a set  $Cn$  of connectives such that for every set  $P$ ,  $Cn$  is the set of connectives of  $\mathcal{L}^P$  in the sense of Definition 32 and  $P$  is the set of atomic formulas<sup>89</sup> of  $\mathcal{L}^P$ , i.e.  $F^P = F(P, Cn)$  for all  $P$ . Sometimes  $P$  is called the *vocabulary* of  $\mathcal{L}^P$ .

(iii)  $\mathbf{L}$  is *compositional* if it has connectives and  $\mathcal{L}^P$  is compositional for all  $P$ .

<sup>88</sup>The definition of algebraizability originates with Blok and Pigozzi, cf. e.g. [Blok and Pigozzi, 1989].

<sup>89</sup>We are making simplifications now. It is not necessary to have  $\mathcal{L}^P$  in  $\mathbf{L}$  for all sets  $P$ . What is needed for our investigations is that for all cardinality  $\kappa$  there be a  $P$  such that the set of atomic formulas of  $\mathcal{L}^P$  has cardinality at least  $\kappa$ .

(iv)  $\mathbf{L}$  has the *filter-property* iff there are derived connectives  $\varepsilon_0, \dots, \varepsilon_{m-1}$ ,  $\delta_0, \dots, \delta_{m-1}$ ,  $\Delta_0, \dots, \Delta_{n-1}$  (common for all possible choices of  $P$ ) such that  $\mathcal{L}^P$  has the filter-property with these, for all  $P$ .

(v)  $\mathbf{L}$  has the *substitution property* iff for all  $P, Q$ ,  $s : P \rightarrow F^Q$ , and  $\varphi \in F^P$ ,

$$\models^P \varphi \quad \text{implies} \quad \models^Q \varphi(\bar{p}/s(\bar{p})).$$

(vi)  $\mathbf{L}$  has the *semantical substitution property* iff for all sets  $P, Q$ ,  $s : P \rightarrow F^Q$  and  $\mathfrak{M} \in M^Q$  there is  $\mathfrak{N} \in M^P$  such that  $mng_{\mathfrak{M}}^Q \circ \hat{s} = mng_{\mathfrak{N}}^P$ .

(vii)  $\mathbf{L}$  is an *algebraizable general logic* iff  $\mathbf{L}$  is compositional, has the filter-property, and has both substitution properties,  $\mathbf{L}$  is *structural* if it is compositional and has the semantical substitution property.

(viii) The notions of a formula-scheme, valid formula-scheme, valid rule, and Hilbert-style inference system for a general logic are the obvious generalizations of their versions given for (non-general) logics  $\mathcal{L}$ , cf. e.g. Definition 33.

(ix) By a *fully fledged general logic* we understand a function

$$\mathbf{L} = \langle \mathcal{L}^P : P \text{ is a set} \rangle$$

such that for each set  $P$ ,  $\mathcal{L}^P = \langle F^P, \vdash^P, M^P, mng^P, \models^P \rangle$  is a fully fledged logic i.e.  $\langle F, \vdash \rangle$  is a deductive logic,  $\langle F, M, mng, \models \rangle$  is a semantical logic, and  $\vdash^P \subseteq \models^P$ . Items (ii)–(vii) above extend to the fully fledged case the natural way.

REMARK 40. We note that  $\mathbf{L}$  is an algebraizable general logic iff  $\mathcal{L}^P$  is an algebraizable semantical logic for all  $P$ , the connectives and the derived connectives for the filter-property are the same for all  $P$ , and the condition below holds for all  $P \subseteq Q$ :

$$\{mng_{\mathfrak{M}}^P : \mathfrak{M} \in M^P\} = \{(mng_{\mathfrak{M}}^Q) \upharpoonright F^P : \mathfrak{M} \in M^Q\}.$$

Intuitively, this condition says that  $\mathcal{L}^P$  is the natural restriction of  $\mathcal{L}^Q$ .

REMARK 41. The theories of semantical logics  $\mathcal{L}_s = \langle F, M, mng, \models \rangle$  and deductive logics  $\mathcal{L}_d = \langle F, \vdash, \rangle$  are best when developed in a parallel fashion, cf. e.g. [Andréka *et al.*, to appear]. Throughout this remark we assume that  $\mathcal{L}_s$  is structural (cf. Definition 38).

We called  $\mathcal{L}_s$  algebraizable iff it has the filter property. An analogous definition for algebraizability of  $\mathcal{L}_d$  was given in the papers of Blok and Pigozzi, cf. [Blok and Pigozzi, 1989]. There are weaker properties of  $\mathcal{L}_d$  studied in the Blok and Pigozzi papers which properties already enable one

to apply (at least part of) the methodology of algebraic logic to the logics in question.

Logics with these properties are called ‘protoalgebraic’, ‘equivalential’, and ‘weakly algebraizable’.<sup>90</sup> (There are other such properties in the literature, but the weakest one facilitating application of our methodology seems to be being protoalgebraic.) As we implied, the properties of being protoalgebraic and equivalential naturally extend to semantical logics.

We call a semantical logic  $\mathcal{L}_s$  *protoalgebraic* iff there is a set  $\Delta(\varphi, \psi) = \{\Delta_i(\varphi, \psi) : i \in I\}$  of derived connectives such that

$$(*) \quad \models \Delta(\varphi, \varphi) \quad \text{and} \quad \mathfrak{M} \models \Delta(\varphi, \psi) \Rightarrow (\mathfrak{M} \models \varphi \Leftrightarrow \mathfrak{M} \models \psi),$$

for all  $\mathfrak{M} \in M$ , and  $\varphi, \psi \in F$ .

$\mathcal{L}_s$  is called *equivalential* iff there is  $\Delta$  as in (\*) above, but such that Condition (1) in Definition 35 holds for this  $\Delta$  and  $\mathcal{L}_s$ .

$\mathcal{L}_s$  is called *weakly algebraizable* iff it is protoalgebraic and there are sets  $\varepsilon(x) = \{\varepsilon_i(x) : i \in I\}$  and  $\delta(x) = \{\delta_i(x) : i \in I\}$  as in Definition 35(2).

Clearly, if  $\mathcal{L}_s$  is both equivalential and weakly algebraizable, then  $\mathcal{L}_s$  is infinitely algebraizable, where ‘infinitely’ means that  $\Delta, \varepsilon, \delta$  may be infinite sets of derived connectives. (If they are finite, then  $\mathcal{L}_s$  is algebraizable.) To keep the rest of this discussion short, we concentrate on protoalgebraic and equivalential (but all what we say can be extended to weakly algebraizable, too).

We note that if  $\mathcal{L}_s$  is protoalgebraic (or equivalential), then so is its deductive counterpart  $\langle F, \overset{\bullet}{\equiv} \rangle$  in the sense of e.g. [Czelakowski and Pigozzi, 1999] Moreover,  $\mathcal{L}_d$  is protoalgebraic/equivalential in the sense of ‘op. cit.’ iff there is a semantical logic  $\mathcal{L}_s$  such that  $\mathcal{L}_d = \langle F, \overset{\bullet}{\equiv}_{\mathcal{L}_s} \rangle$  and  $\mathcal{L}_s$  is protoalgebraic/equivalential, respectively. For the method of proving such equivalence theorems we refer to [Font and Jansana, 1994] and [Andréka *et al.*, to appear]. We also note that

$$\text{protoalgebraic} \supseteq \text{equivalential} \supseteq \text{algebraizable}$$

for semantical logics (the same applies to deductive ones, too). The machinery (algebraization process, equivalence theorems etc) developed in the present work does extend to protoalgebraic and equivalential semantical logics (from algebraizable ones). Cf. e.g. [Hoogland, 1996; Hoogland and Madarász, 1997] for part of this extension. For brevity, in this work we present the above mentioned machinery for the case of algebraizable logics only, at the same time inviting the interested reader to extend this machinery to the protoalgebraic and/or equivalential cases, too.

**Connections with the literature:** What we call a fully fledged logic  $\mathcal{L} =$

<sup>90</sup>Cf. [Czelakowski and Pigozzi, 1999].

$\langle F, \vdash, \dots, \models \rangle$  was called<sup>91</sup> a ‘formalism’ in [Tarski and Givant, 1987, p.16 (section 1.6)], an ‘axiomatic system with semantics  $\langle A, \vdash, \text{Mod}, \models \rangle$ ’ in [Aczel, 1994, p.265] (here  $A$  coincides with our  $F$ ), and a ‘logic  $\langle \text{Sign}, \text{sen}, \text{Mod}, \vdash, \models \rangle$ ’ in [Martì-Oliet and Meseguer, 1994]. More precisely, the latter corresponds to our fully fledged general logics (cf. the subitem below).

What we call a semantic logic was called a ‘semantical system’ in [Andréka *et al.*, to appear], a ‘semantical system  $\langle A, \text{Mod}, \models \rangle$ ’ in [Aczel, 1994, p.265]; and our general logic  $\mathbf{L} = \langle \mathcal{L}^P : \dots \rangle$  corresponds to an institution  $\langle \text{Sign}, \text{sen}, \text{Mod}, \models \rangle$  in [Martì-Oliet and Meseguer, 1994, p.358] (the latter will be elaborated below).

**Connection with institutions:** Institutions theory (e.g. [Martì-Oliet and Meseguer, 1994]) emphasizes the category theoretic aspects of a general logic  $\mathbf{L} = \langle \mathcal{L}^P : \dots \rangle$  (which are downplayed here), and suppresses the intensional aspects represented by *mng*. To see that such a general logic is a category whose objects are the logics  $\mathcal{L}^P$ , let

$$\text{Sign} = \{P : P \text{ occurs as a set of atomic formulas in } \mathbf{L}\}$$

be fixed.<sup>92</sup> (We know that according to our conventions,  $\text{Sign} =$  ‘all sets’, but let us abstract away from this, and just assume that  $\text{Sign}$  is a fixed proper class.) Let  $P, P_1 \in \text{Sign}$ . Next we define what a logic morphism  $h : \mathcal{L}^P \rightarrow \mathcal{L}^{P_1}$  is. A *logic morphism* is a pair  $h = \langle f, \mu \rangle$  such that  $f : \mathfrak{F}^P \rightarrow \mathfrak{F}^{P_1}$  is a homomorphism of the formula-algebras and  $\mu : M^{P_1} \rightarrow M^P$  ‘makes everything commute’, e.g.  $\mathfrak{M} \models f(\varphi) \iff \mu(\mathfrak{M}) \models \varphi$ .<sup>93</sup> Now, if  $\langle f, \mu \rangle$  is such a logic morphism, then  $(f \upharpoonright P)$  is called a *signature morphism*. Let **Sign** be the category of these signature-morphisms (as arrows, and  $\text{Sign}$  itself is the class of objects).

Let **Log** be the category of logics  $\mathcal{L}^P$  and logic morphisms  $h = \langle f, \mu \rangle$  occurring in  $\mathbf{L}$ . Then there is a functor  $\mathbf{F} : \mathbf{Sign} \rightarrow \mathbf{Log}$  sending  $P$  to  $\mathcal{L}^P$ . This  $\mathbf{F}$  is almost the institution we are looking for. There is one ingredient missing, though. Namely,  $M^P$  is not only an arbitrary class, but is a category in all the applications we know of. This category character of  $M^P$  is de-emphasized in the present Paper, but it does show up in the theory later, cf. e.g. the category of concrete semantical systems in [Andréka *et al.*, to appear]. So, let us assume that each  $M^P$  is a category. Then our functor  $\mathbf{F}$  above induces two functors  $fml : \mathbf{Sign} \rightarrow \mathbf{Fmla}$  and  $\text{Mod} : \mathbf{Sign} \rightarrow \mathbf{Cat}^{op}$ , where  $\mathbf{Fmla}$  is the category of formula-algebras (of the form  $\mathfrak{F}^P$ ),

<sup>91</sup>A difference is that the intensional part *mng* is suppressed in most of the quoted works but it is not suppressed in e.g. [Epstein, 1990].

<sup>92</sup>What we call here a set  $P$  of atomic formulas is called a signature in institutions theory.

<sup>93</sup>If meaning functions are also present, then  $\mu$  should induce a function  $\mu^+$  on the meanings such that  $mng_{\mu(\mathfrak{M})}(\varphi) = \mu^+(mng_{\mathfrak{M}}(f(\varphi)))$ , cf. [Andréka *et al.*, to appear].

and for each  $P$ ,

$$\mathbf{F}(P) = \langle fml(P), \text{Mod}(P), \models^P \rangle, \quad \text{while}$$

$$\mathbf{F}(\sigma) = \langle fml(\sigma), \text{Mod}(\sigma) \rangle$$

for the morphisms  $\sigma$  of **Sign**. Now the tuple

$$I(\mathbf{L}) = \langle \mathbf{Sign}, fml, \text{Mod}, \models \rangle$$

is called an *institution*, where

$$\models = \langle \models^P : P \in \text{Sign} \rangle.$$

At this point one can see that to a general logic  $\mathbf{L}$  there is an equivalent institution  $I(\mathbf{L})$ , and conversely to an institution  $\mathbf{I}$  there is a general logic  $L(\mathbf{I})$  such that the two can be recovered from each other. (Here we assumed that in  $L(\mathbf{I})$  the models still form a category.) Therefore, institutions and general logics can be studied interchangeably, depending on the kind of mathematical tools (universal algebra or categories) one wants to use.<sup>94</sup>

In institutions theory, our ‘*fml*’ is denoted by ‘*sen*’, exactly because there *mng* is suppressed and therefore meanings are replaced by truthvalues. So, when the theory is applied to e.g. first-order logic, then attention has to be restricted to sentences (=closed formulas) because meanings of open formulas are more complex objects than just truthvalues.

We do not treat here the different notions of equivalence of logics, morphisms acting between logics, concrete semantical logics, cf. e.g. [Andréka, *et al.*, to appear]. Also, we do not treat here interpretability between logics, and combining logics, [Gabbay, 1996; Gabbay, 1998; Jánossy, Kurucz and Eiben, 1996; Andréka *et al.*, to appear]. These are important and very interesting subjects.

For the rest of this work, one of the most important definitions of Section 4 is that of an algebraizable general logic. It is summarized in Remark 40.

## 5 THE PROCESS OF ALGEBRAIZATION

The algebraic counterpart of classical sentential logic  $\mathcal{L}_S$  is the variety **BA** of Boolean algebras. Why is this so important? The answer lies in the general experience that sometimes it is easier to solve a problem concerning  $\mathcal{L}_S$  by translating it to **BA**, solving the algebraic problem, and then translating the result back to  $\mathcal{L}_S$  (than solving it directly in  $\mathcal{L}_S$ ).

<sup>94</sup>To make this statement hundred percent true, one should include the intensional aspect *mng* into institutions and make  $M^P$  into a category in general logics. We do not see why one would not do these amendments.

In this Section we extend applicability of BA to  $\mathcal{L}_S$  to applicability of algebra in general to logics in general. We will introduce a standard translation method from logic to algebra, which to each logic  $\mathcal{L}$  associates a class  $\mathbf{Alg}(\mathcal{L})$  of algebras. (Of course,  $\mathbf{Alg}(\mathcal{L}_S)$  will be BA.) Further, this translation method will tell us how to find the algebraic question corresponding to a logical question. If the logical question is about  $\mathcal{L}$ , then its algebraic equivalent will be about  $\mathbf{Alg}(\mathcal{L})$ . For example, if we want to decide whether  $\mathcal{L}$  has the property called Craig's interpolation property, then it is sufficient to decide whether  $\mathbf{Alg}(\mathcal{L})$  has the so called amalgamation property (for which there are powerful methods in the literature of algebra). If the logical question concerns connections between several logics, say between  $\mathcal{L}_1$  and  $\mathcal{L}_2$ , then the algebraic question will be about connections between  $\mathbf{Alg}(\mathcal{L}_1)$  and  $\mathbf{Alg}(\mathcal{L}_2)$ . (The latter are quite often simpler, hence easier to investigate.) This 'bridge' also enables us to solve algebraic problems by logical methods (for an example see Section 7).

**DEFINITION 42** (Meaning algebra,  $\mathbf{Alg}_m$ ,  $\mathbf{Alg}$ ). Let  $\mathcal{L} = \langle F, M, mng, |= \rangle$  be a compositional logic with  $F \neq \emptyset$ .

(i) First we turn every model into an algebra. Compositionality of  $mng_{\mathfrak{M}}$  means that we can define an algebra of type  $Cn$  on the set  $\{mng_{\mathfrak{M}}(\varphi) : \varphi \in F\}$  of meanings. This algebra is  $mng_{\mathfrak{M}}(\mathfrak{F})$ , it will be called the *meaning algebra of  $\mathfrak{M}$*  and it will be denoted by  $\mathfrak{Mng}(\mathfrak{M})$ . In more detail, to any logical connective  $c$  of arity  $k$  we can define a  $k$ -ary function  $c^{\mathfrak{M}}$  on the meanings in  $\mathfrak{M}$  by setting for all formulas  $\varphi_1, \dots, \varphi_k$

$$c^{\mathfrak{M}}(mng_{\mathfrak{M}}(\varphi_1), \dots, mng_{\mathfrak{M}}(\varphi_k)) \stackrel{\text{def}}{=} mng_{\mathfrak{M}}(c(\varphi_1, \dots, \varphi_k)).$$

(We could say that  $c^{\mathfrak{M}}$  is the meaning of the logical connective  $c$ .) Then  $\mathfrak{Mng}(\mathfrak{M}) \stackrel{\text{def}}{=} \langle \{mng_{\mathfrak{M}}(\varphi) : \varphi \in F\}, c^{\mathfrak{M}} \rangle_{c \in Cn}$ .

(ii)  $\mathbf{Alg}_m(\mathcal{L})$  denotes the class of all meaning-algebras of  $\mathcal{L}$ , i.e.

$$\mathbf{Alg}_m(\mathcal{L}) \stackrel{\text{def}}{=} \{mng_{\mathfrak{M}}(\mathfrak{F}) : \mathfrak{M} \in M_{\mathcal{L}}\} = \{\mathfrak{Mng}(\mathfrak{M}) : \mathfrak{M} \in M_{\mathcal{L}}\}.$$

(iii) Let  $K \subseteq M_{\mathcal{L}}$ . Then for every  $\varphi, \psi \in F$

$$\varphi \sim_K \psi \stackrel{\text{def}}{\iff} (\forall \mathfrak{M} \in K) mng_{\mathfrak{M}}(\varphi) = mng_{\mathfrak{M}}(\psi).$$

Then  $\sim_K$  is an equivalence relation, which is a congruence on  $\mathfrak{F}$  by compositionality of  $\mathcal{L}$ .  $\mathfrak{F}/\sim_K$  denotes the factor-algebra of  $\mathfrak{F}$ , factorized by  $\sim_K$ . It is called the Lindenbaum–Tarski algebra of  $K$ . Now,

$$\mathbf{Alg}(\mathcal{L}) \stackrel{\text{def}}{=} \mathbf{I} \{\mathfrak{F}/\sim_K : K \subseteq M_{\mathcal{L}}\}.$$



Thus,  $\mathbf{Alg}(\mathcal{L})$  is the class of isomorphic copies of the Lindenbaum–Tarski algebras of  $\mathcal{L}$ .

(iv) Let  $\mathbf{L} = \langle \mathcal{L}^P : P \text{ is a set} \rangle$  be a general logic. Then

$$\mathbf{Alg}_m(\mathbf{L}) \stackrel{\text{def}}{=} \bigcup \{ \mathbf{Alg}_m(\mathcal{L}^P) : P \text{ is a set, } F^P \neq \emptyset \},$$

and

$$\mathbf{Alg}(\mathbf{L}) \stackrel{\text{def}}{=} \bigcup \{ \mathbf{Alg}(\mathcal{L}^P) : P \text{ is a set, } F^P \neq \emptyset \}.$$

REMARK . In the definition of  $\mathbf{Alg}_m(\mathcal{L})$  above, it is important that  $\mathbf{Alg}_m(\mathcal{L})$  is not an abstract class in the sense that it is not closed under isomorphisms. The reason for defining  $\mathbf{Alg}_m(\mathcal{L})$  in such a way is that since  $\mathbf{Alg}_m(\mathcal{L})$  is the class of algebraic counterparts of the *models* of  $\mathcal{L}$ , we need these algebras as concrete algebras and replacing them with their isomorphic copies would lead to loss of information (about semantic-model theoretic matters). See e.g. the algebraic characterization of the *weak Beth definability property*, Theorem 59 in the next Section.

For a logic  $\mathcal{L}$ , let  $Mng_{\mathcal{L}} \stackrel{\text{def}}{=} \{ mng_{\mathfrak{M}} : \mathfrak{M} \in M_{\mathcal{L}} \}$ . That is,  $Mng_{\mathcal{L}}$  is the class of ‘meaning-homomorphisms’ of the logic  $\mathcal{L}$  (or equivalently, the unary meaning-functions induced by the models of  $\mathcal{L}$ ). If  $\mathbf{L}$  is a general logic  $\langle \mathcal{L}^P : P \text{ is a set} \rangle$  then  $Mng^P$  denotes the class of all meaning-homomorphisms of  $\mathcal{L}^P$ . That is,  $Mng^P = Mng_{\mathcal{L}^P}$ . If  $\mathfrak{A}$  is an algebra and  $\mathbf{K}$  is a class of algebras, then  $Hom(\mathfrak{A}, \mathbf{K})$  denotes the class of all homomorphisms  $h : \mathfrak{A} \rightarrow \mathfrak{B}$  where  $\mathfrak{B} \in \mathbf{K}$ .

PROPOSITION 43 (Characterization of structural logics). Let  $\mathcal{L}$  and  $\mathbf{L}$  be a compositional logic and a compositional general logic, respectively. Then (i)–(ii) below hold.

- (i)  $\mathcal{L}$  has the semantic substitution property iff
  - $Mng_{\mathcal{L}} = Hom(\mathfrak{F}, \mathbf{Alg}_m(\mathcal{L}))$ , iff
  - $Mng_{\mathcal{L}} = Hom(\mathfrak{F}, \mathbf{K})$  for some class  $\mathbf{K}$  of algebras.
- (ii)  $\mathbf{L}$  has the semantic substitution property iff
  - $Mng^P = Hom(\mathfrak{F}^P, \mathbf{Alg}_m(\mathbf{L}))$ , for all  $P$ , iff
  - $Mng^P = Hom(\mathfrak{F}^P, \mathbf{K})$ , for all  $P$ , for some  $\mathbf{K}$ .

**Proof.** To prove the first equivalence in (i), assume first that  $\mathcal{L}$  has the semantic substitution property, and let  $h \in Hom(\mathfrak{F}, \mathbf{Alg}_m(\mathcal{L}))$ . Then  $h : \mathfrak{F} \rightarrow mng_{\mathfrak{M}}(\mathfrak{F})$  for some  $\mathfrak{M} \in M_{\mathcal{L}}$ . Let such an  $\mathfrak{M}$  be fixed. We want to show that  $h = mng_{\mathfrak{M}}$  for some  $\mathfrak{M} \in M_{\mathcal{L}}$ . Of course, the kernel  $R (= \{ \langle a, b \rangle \in F : mng_{\mathfrak{M}}(a) = mng_{\mathfrak{M}}(b) \})$  of  $mng_{\mathfrak{M}}$  is an equivalence relation. Let  $F/R$  denote the partition of  $F$  determined by  $R$ . Let  $s : P \rightarrow F/R$  be a choice

function such that  $s(p) \in mng_{\mathfrak{M}}^{-1}(h(p))$ . Thus  $s$  chooses exactly one element from each equivalence class of  $R$  the  $mng_{\mathfrak{M}}$ -image of which is  $h(p)$  for some  $p \in P$ . Such an  $s$  exists (assuming the Axiom of Choice), since  $mng_{\mathfrak{M}}$  is an *onto* function.

Since  $\mathcal{L}$  has the semantic substitution property by hypothesis, to the model  $\mathfrak{M}$  and the substitution  $s$  there exists an  $\mathfrak{N} \in M_{\mathcal{L}}$  such that  $mng_{\mathfrak{N}}(p) = mng_{\mathfrak{M}}(s(p))$  for each  $p \in P$ . Since  $mng_{\mathfrak{M}}(s(p)) = h(p)$  for each  $p \in P$ , by the choice of  $s$ , we have that  $mng_{\mathfrak{N}}(p) = mng_{\mathfrak{M}}(s(p)) = h(p)$  for each  $p \in P$ , thus  $mng_{\mathfrak{N}}$  and  $h$  agree on  $P$ . Since  $\mathfrak{F}$  is freely generated by  $P$  and both  $h$  and  $mng_{\mathfrak{N}}$  are homomorphisms (the latter by compositionality of  $\mathcal{L}$ ), we have that  $h = mng_{\mathfrak{N}}$ . This completes the proof of  $Hom(\mathfrak{F}, \mathbf{Alg}_m(\mathcal{L})) \subseteq Mng_{\mathcal{L}}$ , since  $h$  was chosen arbitrarily.

If  $h \in Mng_{\mathcal{L}}$  then  $h = mng_{\mathfrak{M}}$  for some  $\mathfrak{M} \in M_{\mathcal{L}}$ , and thus  $h \in Hom(\mathfrak{F}, \mathbf{Alg}_m(\mathcal{L}))$  by compositionality of  $\mathcal{L}$ . Therefore  $Mng_{\mathcal{L}} \subseteq Hom(\mathfrak{F}, \mathbf{Alg}_m(\mathcal{L}))$ , which completes the proof of  $Mng_{\mathcal{L}} = Hom(\mathfrak{F}, \mathbf{Alg}_m(\mathcal{L}))$ .

The other direction of the first part of (i) is trivial: Let  $\mathfrak{M} \in M_{\mathcal{L}}$  and  $s : P \rightarrow F$ . We have to show that  $mng_{\mathfrak{M}} \circ \hat{s} \in Mng$ , which is true by  $mng_{\mathfrak{M}} \circ \hat{s} : \mathfrak{F} \rightarrow mng_{\mathfrak{M}}(\mathfrak{F}) \in \mathbf{Alg}_m(\mathcal{L})$ .

To prove the equivalence of the second and third statements in (i), assume that  $Mng_{\mathcal{L}} = Hom(\mathfrak{F}, \mathbf{K})$ . We want to show that  $Mng_{\mathcal{L}} = Hom(\mathfrak{F}, \mathbf{Alg}_m(\mathcal{L}))$ . Notice first that  $Mng_{\mathcal{L}} = Hom(\mathfrak{F}, \mathbf{K})$  implies that  $\mathbf{Alg}_m(\mathcal{L}) \subseteq \mathbf{SK}$ . So let  $h : \mathfrak{F} \rightarrow \mathfrak{A}$ ,  $\mathfrak{A} \in \mathbf{Alg}_m(\mathcal{L})$ . Then  $h : \mathfrak{F} \rightarrow \mathfrak{B}$  for some  $\mathfrak{B} \in \mathbf{K}$ , by  $\mathbf{Alg}_m(\mathcal{L}) \subseteq \mathbf{SK}$ . Thus  $h \in Mng_{\mathcal{L}}$  by  $Mng_{\mathcal{L}} = Hom(\mathfrak{F}, \mathbf{K})$ .

The proof of (ii) is completely analogous, and we omit it. ■

**THEOREM 44** (Connection between  $\mathbf{Alg}_m$  and  $\mathbf{Alg}$ ).

(i) Let  $\mathcal{L}$  be a compositional logic. Then

$$\mathbf{SPAlg}(\mathcal{L}) = \mathbf{SPAlg}_m(\mathcal{L}).$$

(ii) Let  $\mathbf{L}$  be a structural general logic. Then

$$\mathbf{Alg}(\mathbf{L}) = \mathbf{SPAlg}_m(\mathbf{L}).$$

**Proof.** Proof of (i): First we show  $\mathbf{Alg}_m(\mathcal{L}) \subseteq \mathbf{IAlg}(\mathcal{L})$ . Let  $\mathfrak{A} \in \mathbf{Alg}_m(\mathcal{L})$ , say  $\mathfrak{A} = mng_{\mathfrak{M}}(\mathfrak{F})$ . Set  $K = \{\mathfrak{M}\}$ . Then  $\sim_K = \ker(mng_{\mathfrak{M}})$ , so  $\mathfrak{F}/\sim_K$  is isomorphic to  $\mathfrak{A}$ .

To show the other direction, let  $\mathfrak{A} \in \mathbf{Alg}(\mathcal{L})$ , say  $\mathfrak{A} = \mathfrak{F}/\sim_K$  for some  $K \subseteq M_{\mathcal{L}}$ . Then there is a subset  $K' \subseteq K$  such that  $\mathfrak{F}/\sim_K = \mathfrak{F}/\sim_{K'}$  (this holds because  $F$  is a set). Thus we may assume that  $K \subseteq M_{\mathcal{L}}$  is a set. We define for all  $\varphi \in F$

$$h(\varphi/\sim_K) \stackrel{\text{def}}{=} \langle mng_{\mathfrak{M}}(\varphi) : \mathfrak{M} \in K \rangle.$$

This is a sound definition by the definition of  $\sim_K$ . It is not difficult to check that  $h$  is one-to-one and a homomorphism, so  $h : \mathfrak{A} \mapsto \mathbf{P}\langle \text{mng}_{\mathfrak{M}}(\mathfrak{F}) : \mathfrak{M} \in K \rangle$ , showing that  $\mathfrak{A} \in \mathbf{SPAlg}_m(\mathcal{L})$ . Since  $\mathbf{SP}$  is a closure operator, we are done with proving (i).

Proof of (ii): First we note that, by (i),  $\mathbf{Alg}(\mathcal{L}^P) \subseteq \mathbf{SPAlg}_m(\mathcal{L}^P)$  for any set  $P$ , thus  $\mathbf{Alg}(\mathbf{L}) \subseteq \mathbf{SPAlg}_m(\mathbf{L})$  holds.

We are going to prove  $\mathbf{SPAlg}_m(\mathbf{L}) \subseteq \mathbf{Alg}(\mathbf{L})$ . Let  $\mathfrak{A} \in \mathbf{SPAlg}_m(\mathbf{L})$ , say  $\mathfrak{A} \subseteq \mathbf{P}_{i \in I} \mathfrak{A}_i$  for a set  $I$  and algebras  $\mathfrak{A}_i \in \mathbf{Alg}_m(\mathbf{L})$ . Let  $h : \mathfrak{F}^A \rightarrow \mathfrak{A}$  be any onto homomorphism (e.g. we can take for  $h$  the homomorphic extension of the identity mapping  $h' : A \rightarrow A$ ). For each  $i \in I$  let  $\pi_i$  denote the projection function onto  $\mathfrak{A}_i$ , and let  $h_i \stackrel{\text{def}}{=} \pi_i \circ h$ . Then  $h_i : \mathfrak{F}^A \rightarrow \mathfrak{A}_i \in \mathbf{Alg}_m(\mathbf{L})$ . By Proposition 43 (ii) then  $h_i = \text{mng}_{\mathfrak{M}_i}$  for some  $\mathfrak{M}_i \in M^A$ . Let  $K = \{\mathfrak{M}_i : i \in I\}$ . Then it is easy to check that  $h(\varphi) = h(\psi)$  iff  $\varphi \sim_K \psi$  for all  $\varphi, \psi \in F^A$ . Thus  $\mathfrak{A}$  is isomorphic to  $\mathfrak{F}^A / \sim_K \in \mathbf{Alg}(\mathbf{L})$ , and we are done.  $\blacksquare$

We note that we also proved that for structural logics  $\mathcal{L}$ ,

$$\mathbf{Alg}(\mathcal{L}) = \mathbf{SPAlg}_m(\mathcal{L}) \cap \{\mathfrak{A} : |A| \leq |F_{\mathcal{L}}|\}.$$

Now we turn to proving that the equations valid in  $\mathbf{Alg}(\mathcal{L})$  correspond to the valid formula-schemes of  $\mathcal{L}$ , and the quasi-equations valid in  $\mathbf{Alg}(\mathcal{L})$  correspond to the valid rules of  $\mathcal{L}$ . Here we will use the filter-property. If  $\mathcal{L}$  is algebraizable, then the equational and quasi-equational theories of  $\mathbf{Alg}(\mathcal{L})$  recapture the validities and the semantical consequence relation  $\models$  of  $\mathcal{L}$ , respectively. Thus, when a logic  $\mathcal{L}$  is given, it is interesting to investigate the equational and quasi-equational theories of  $\mathbf{Alg}(\mathcal{L})$ . Note that by Theorem 44 above,  $\mathbf{Alg}(\mathcal{L})$  and  $\mathbf{Alg}_m(\mathcal{L})$  have the same equational and quasi-equational theories.

First we note that formulas and formula-schemes are terms in the language of  $\mathbf{Alg}(\mathcal{L})$ . Hence if  $\varphi, \psi \in Fs$ , then  $\varphi = \psi$  is an equation in the language of  $\mathbf{Alg}(\mathcal{L})$  where we consider the formula-variables  $\phi_i$  as algebraic variables (ranging over the elements of the algebras). Similarly, if  $\varphi, \psi \in F$ , then  $\varphi = \psi$  is also an equation in the language of  $\mathbf{Alg}(\mathcal{L})$ , where we consider the elements of  $P$  as algebraic variables.

**THEOREM 45** (Valid rules of  $\mathcal{L}$  and quasi-equations of  $\mathbf{Alg}(\mathcal{L})$ ). *Let  $\mathcal{L}$  be a compositional logic with filter-property. Let  $\varepsilon, \delta, \Delta, m, n$  be as in the definition of filter-property. Then (i)–(ii) below hold.*

(i) A rule  $\langle \langle \varphi_0, \dots, \varphi_{k-1} \rangle, \varphi_k \rangle$  of  $\mathcal{L}$  is valid (or admissible) iff

$$\mathbf{Alg}(\mathcal{L}) \models \bigwedge_{\substack{\ell < k \\ i < m}} \varepsilon_i(\varphi_\ell) = \delta_i(\varphi_\ell) \rightarrow \varepsilon_j(\varphi_k) = \delta_j(\varphi_k) \quad \text{for each } j < m.$$

(ii) A quasi-equation  $\varphi_0 = \psi_0 \wedge \dots \wedge \varphi_{k-1} = \psi_{k-1} \rightarrow \varphi_k = \psi_k$  (with variables from  $FV$ ) is valid in  $\mathbf{Alg}(\mathcal{L})$  iff

the rules

$$\frac{\varphi_0 \Delta_0 \psi_0, \dots, \varphi_0 \Delta_{n-1} \psi_0, \dots, \varphi_{k-1} \Delta_0 \psi_{k-1}, \dots, \varphi_{k-1} \Delta_{n-1} \psi_{k-1}}{\varphi_k \Delta_j \psi_k}$$

are valid in  $\mathcal{L}$  for all  $j < n$ .

**Proof.** Assume that  $\rho$  is a valid rule of  $\mathcal{L}$  of the form  $\langle\langle\varphi_0, \dots, \varphi_{k-1}\rangle, \varphi_k\rangle$ . Let  $req_j$  denote the quasi-equation associated to it in (i). We want to show that  $req_j$  is valid in  $\mathbf{Alg}(\mathcal{L})$ . By Theorem 44 it is enough to prove that it is valid in  $\mathbf{Alg}_m(\mathcal{L})$ . Let  $\mathfrak{A} \in \mathbf{Alg}_m(\mathcal{L})$ , and let  $h : FV \rightarrow A$  be an evaluation of the variables in  $req_j$  such that the hypothesis part of  $req_j$  is true in  $\mathfrak{A}$  under the evaluation  $h$ , i.e. assume that

$$(\star) \quad \mathfrak{A} \models \bigwedge_{\substack{\ell < k \\ i < m}} \varepsilon_i(\varphi_\ell) = \delta_i(\varphi_\ell)[h].$$

We want to show

$$\mathfrak{A} \models \varepsilon_j(\varphi_k) = \delta_j(\varphi_k)[h].$$

By  $\mathfrak{A} \in \mathbf{Alg}_m(\mathcal{L})$ , there is  $\mathfrak{M} \in M_{\mathcal{L}}$  such that  $\mathfrak{A} = mng_{\mathfrak{M}}(\mathfrak{F})$ . For any  $\phi_i \in FV$  take  $\psi_i \in F$  such that  $h(\phi_i) = mng_{\mathfrak{M}}(\psi_i)$  and let  $\langle\langle\varphi'_0, \dots, \varphi'_{k-1}\rangle, \varphi'_k\rangle$  be the instance of our rule  $\rho$  by replacing each  $\phi_i$  with  $\psi_i$ . Then for each  $i < m$  and  $\ell \leq k$  we have that  $\hat{h}(\varepsilon_i(\varphi_\ell)) = mng_{\mathfrak{M}}(\varepsilon_i(\varphi'_\ell))$  and the same for  $\delta$ , i.e.  $\hat{h}(\delta_i(\varphi_\ell)) = mng_{\mathfrak{M}}(\delta_i(\varphi'_\ell))$ . (Here  $\hat{h}$  denotes the homomorphic extension of  $h$  to  $Fs$ .) Then by the filter-property of  $\mathcal{L}$ , and by our assumption  $(\star)$ , we have  $\mathfrak{M} \models_{\mathcal{L}} \varphi'_\ell$  for all  $\ell < k$ . Since  $\langle\langle\varphi_0, \dots, \varphi_{k-1}\rangle, \varphi_k\rangle$  is a valid rule, and  $\langle\langle\varphi'_0, \dots, \varphi'_{k-1}\rangle, \varphi'_k\rangle$  is an instance of it, this implies  $\mathfrak{M} \models_{\mathcal{L}} \varphi'_k$ . Then by the filter-property again,  $mng_{\mathfrak{M}}(\varepsilon_j(\varphi'_k)) = mng_{\mathfrak{M}}(\delta_j(\varphi'_k))$ , i.e.  $\hat{h}(\varepsilon_j(\varphi_k)) = \hat{h}(\delta_j(\varphi_k))$  and we are done.

Conversely, assume that the quasi-equation is valid in  $\mathbf{Alg}(\mathcal{L})$ , and we want to show that the rule is valid. Let  $\langle\langle\varphi'_0, \dots, \varphi'_{k-1}\rangle, \varphi'_k\rangle$  be an instance of the rule that we got by substituting  $\psi_i$  to the formulavariables  $\phi_i$ , for all  $i < \omega$ . Assume  $\mathfrak{M} \in M_{\mathcal{L}}$  and  $\mathfrak{M} \models_{\mathcal{L}} \{\varphi'_0, \dots, \varphi'_{k-1}\}$ . We want to show  $\mathfrak{M} \models_{\mathcal{L}} \varphi'_k$ . By the filter-property we have  $mng_{\mathfrak{M}}(\varepsilon_i(\varphi'_\ell)) = mng_{\mathfrak{M}}(\delta_i(\varphi'_\ell))$ . Let  $h : Fs \rightarrow F$  be a homomorphism such that  $h(\phi_i) = \psi_i$  for all  $i < \omega$ . Then  $h(\varepsilon_i(\varphi_\ell)) = mng_{\mathfrak{M}}(\varepsilon_i(\varphi'_\ell))$  and the same for  $\delta_i$ , thus  $(\star)$  above holds with  $\mathfrak{A} = mng_{\mathfrak{M}}(\mathfrak{F})$ . Thus  $\mathfrak{A} \models \varepsilon_j(\varphi_k) = \delta_j(\varphi_k)[h]$ , since the quasi-equation is valid in  $\mathfrak{A} \in \mathbf{Alg}(\mathcal{L})$ , i.e.  $mng_{\mathfrak{M}}(\varepsilon_j(\varphi'_k)) = mng_{\mathfrak{M}}(\delta_j(\varphi'_k))$ , for all  $j < m$ . By the filter-property then  $\mathfrak{M} \models_{\mathcal{L}} \varphi'_k$  as was to be shown.

We omit the proof of (ii). It is analogous to the above proof of (i).  $\blacksquare$

COROLLARY 46. (Valid formula-schemes, validities, and  $\mathbf{Eq}(\mathbf{Alg}(\mathcal{L}))$ )

(i) Let  $\mathcal{L}$  be a compositional logic with filter-property. Let  $\varepsilon, \delta, \Delta, m, n$  be as in the definition of the filter-property. Then for every formula-scheme  $\varphi$  of  $\mathcal{L}$

$\varphi$  is a valid formula-scheme of  $\mathcal{L}$  iff

$$\mathbf{Alg}(\mathcal{L}) \models \varepsilon_j(\varphi) = \delta_j(\varphi) \quad \text{for all } j < m.$$

(ii) Assume further that  $\mathcal{L}$  is algebraizable. Then for any formulas  $\varphi, \varphi_0, \dots, \varphi_k$  of  $\mathcal{L}$ ,

$\models_{\mathcal{L}} \varphi$  iff

$$\mathbf{Alg}(\mathcal{L}) \models \varepsilon_j(\varphi) = \delta_j(\varphi) \quad \text{for each } j < m.$$

$\{\varphi_0, \dots, \varphi_{k-1}\} \dot{\models}_{\mathcal{L}} \varphi_k$  iff

$$\mathbf{Alg}(\mathcal{L}) \models \bigwedge_{\substack{\ell < k \\ i < m}} \varepsilon_i(\varphi_\ell) = \delta_i(\varphi_\ell) \rightarrow \varepsilon_j(\varphi_k) = \delta_j(\varphi_k) \quad \text{for each } j < m.$$

(iii) The set of valid formula-schemes of  $\mathcal{L}$  is decidable (recursively enumerable) iff  $\mathbf{Eq}(\mathbf{Alg}(\mathcal{L}))$  is decidable (recursively enumerable). The set of valid (admissible) rules of  $\mathcal{L}$  is decidable (recursively enumerable) iff the quasi-equational theory of  $\mathbf{Alg}(\mathcal{L})$  is decidable (recursively enumerable).

(iv) Statements (i) and (ii) above hold for general logics  $\mathbf{L}$  in place of  $\mathcal{L}$ . ■

We say that the *validity problem* of the logic  $\mathcal{L}$  is *decidable* iff the set of valid formulas of  $\mathcal{L}$  is decidable. If  $\mathbf{L} = \langle \mathcal{L}^P : P \text{ is a set} \rangle$  is a general logic, then the validity problem of  $\mathbf{L}$  is decidable iff it is decidable for  $\mathcal{L}^P$ , for all  $P$ .

COROLLARY 47. Let  $\mathcal{L}$  be an algebraizable logic with  $|P| \geq \omega$  or an algebraizable general logic. Then the validity problem of  $\mathcal{L}$  is decidable iff  $\mathbf{Eq}(\mathbf{Alg}(\mathcal{L}))$  is decidable. ■

## 6 EQUIVALENCE THEOREMS (WRITTEN BY J. MADARÁSZ)

In this part we give algebraic characterizations of some logical properties. In the next Section we will apply these theorems to some well known logics. Instead of giving the proofs of the theorems in this Section, we will refer to where they can be found. First we characterize completeness and compactness properties.

DEFINITION 48 (Complete, sound inference systems).

Let  $\mathcal{L} = \langle F, M, mng, \models \rangle$  be a logic, and let  $\vdash \subseteq \mathcal{P}(F) \times F$ .

- $\vdash$  is *weakly complete* for  $\mathcal{L}$  iff each valid formula is derivable, i.e. iff

$$(\forall \varphi \in F) (\models \varphi \implies \vdash \varphi).$$

- $\vdash$  is *finitely complete* for  $\mathcal{L}$  iff consequences of finite sets are derivable, i.e. iff<sup>95</sup>

$$(\forall \Sigma \subseteq_{\omega} F) (\forall \varphi \in F) \left( \Sigma \vDash^{\bullet} \varphi \implies \Sigma \vdash \varphi \right).$$

- $\vdash$  is *strongly complete* for  $\mathcal{L}$  iff any semantical consequence is derivable (not only consequences of finite sets), i.e. iff

$$(\forall \Sigma \subseteq F) (\forall \varphi \in F) \left( \Sigma \vDash^{\bullet} \varphi \implies \Sigma \vdash \varphi \right).$$

- $\vdash$  is *weakly sound* for  $\mathcal{L}$  iff derivable formulas are valid, i.e. iff

$$(\forall \varphi \in F) (\vdash \varphi \implies \models \varphi).$$

- $\vdash$  is *strongly sound* for  $\mathcal{L}$  iff derivable consequences are valid consequences, i.e. iff

$$(\forall \Sigma \subseteq F) (\forall \varphi \in F) \left( \Sigma \vdash \varphi \implies \Sigma \vDash^{\bullet} \varphi \right).$$

- $\vdash$  is *strongly complete and sound* for  $\mathcal{L}$  iff the derivability and semantic consequence relations coincide, i.e. iff

$$(\forall \Sigma \subseteq F) (\forall \varphi \in F) \left( \Sigma \vdash \varphi \text{ iff } \Sigma \vDash^{\bullet} \varphi \right).$$

Let  $\mathcal{I} = \langle Ax, Ru \rangle$  be a Hilbert-style inference system. We say that  $\mathcal{I}$  is (weakly, finitely, strongly) complete for  $\mathcal{L}$  if the derivability relation  $\vdash$  given by  $\langle Ax, Ru \rangle$  is such for  $\mathcal{L}$ . We say that  $\mathcal{I}$  is (weakly, finitely, strongly) complete for a general logic  $\mathbf{L} = \langle \mathcal{L}^P : P \text{ is a set} \rangle$  if  $\mathcal{I}$  is such for all  $\mathcal{L}^P$ . We use an analogous terminology for the soundness properties.

The next theorem is a characterization of existence of strongly complete and sound Hilbert-style inference systems for general logics. It is proved in [Andréka *et al*, 1993, Thm.3.2.21]. A class  $\mathbf{K}$  is a *quasi-variety* iff it can be axiomatized with a set of *quasi-equations*, i.e. equational implications. See a footnote in Section 1.

**THEOREM 49.** *Assume that  $\mathbf{L}$  is an algebraizable general logic. Then there is a strongly complete and sound Hilbert-style inference system for  $\mathbf{L}$  iff  $\mathbf{Alg}(\mathbf{L})$  is a finitely axiomatizable quasi-variety. ■*

<sup>95</sup>  $X \subseteq_{\omega} Y$  denotes that  $X$  is a finite subset of  $Y$ .

In Theorem 50 below we give a sufficient and necessary condition for an algebraizable semantic logic to have a finitely complete Hilbert-style inference system. Its proof can be found in [Andréka *et al.*, 1993, Thm.3.2.3].

**THEOREM 50.** *Assume that  $\mathcal{L}$  is an algebraizable semantic logic and  $Cn(\mathcal{L})$  is finite.<sup>96</sup> Then there is a finitely complete and strongly sound Hilbert-style inference system for  $\mathcal{L}$  iff  $\mathbf{Alg}(\mathcal{L})$  generates a finitely axiomatizable quasi-variety. ■*

The following theorem is a characterization of existence of weakly complete and strongly sound Hilbert-style inference systems. It is [Andréka *et al.*, 1993, Thm. 3.2.4].

**THEOREM 51.** *Assume that  $\mathcal{L}$  is an algebraizable semantic logic and  $Cn(\mathcal{L})$  is finite. Then there is a weakly complete and strongly sound Hilbert-style inference system for  $\mathcal{L}$  iff there is a finitely axiomatizable quasi-variety  $\mathbf{K}$  such that  $\mathbf{Alg}_m(\mathcal{L}) \subseteq \mathbf{K} \subseteq \mathbf{HSPAlg}(\mathcal{L})$ . The same is true for algebraizable general logics. ■*

The following theorem, due to J. Madarász, is a characterization of existence of weakly complete and weakly sound Hilbert-style inference systems. Such an inference-system is sometimes called a Gabbay-style inference system. Its rules are not necessarily valid, but the formula-schemes they derive (from the empty set  $\Sigma = \emptyset$  of premises) should be valid.<sup>97</sup>

**THEOREM 52.** *Assume that  $\mathbf{L}$  is an algebraizable general logic and  $Cn(\mathbf{L})$  is finite. Assume that  $\mathbf{HAlg}(\mathbf{L}) \models \bar{\varepsilon}(x\bar{\Delta}y) = \bar{\delta}(x\bar{\Delta}y) \rightarrow x = y$ . Then there is a weakly complete and weakly sound Hilbert-style inference system for  $\mathbf{L}$  iff there is a finitely axiomatizable quasi-variety  $\mathbf{K}$  such that  $\mathbf{HK} = \mathbf{HAlg}(\mathbf{L})$ .<sup>98</sup> ■*

<sup>96</sup>One can eliminate the assumption of  $Cn(\mathcal{L})$  being finite. Then the finitary character of a Hilbert-style ensured in a more subtle way. Also, ‘finitely axiomatizable quasi-variety’ must be replaced by ‘finite-schema axiomatizable quasi-variety’ in the second clause, cf. e.g. [Monk, 1969], or [Németi, 1991].

<sup>97</sup>Cf. e.g. [Mikulás, 1995; Marx and Venema, 1997; Simon, 1991]. We note that many Gabbay-style rules are even more ‘liberal’ than not being strongly sound in that in addition their form does not satisfy Definition 33 (ii). (It is not known yet which of these two liberties is responsible for their behaviour. Our feeling is that non-strongly soundness is the more essential.) These extremely liberal (Gabbay-style) inference systems correspond to classes  $\mathbf{K} \subseteq \mathbf{Alg}(\mathbf{L})$  such that  $\mathbf{K}$  is finitely axiomatized by  $\forall\exists$ -formulas (of a certain form) and  $\mathbf{HSK} = \mathbf{HAlg}(\mathbf{L})$ . An example for such  $\mathbf{K}$  is the class of rectangularly dense cylindric algebras, the representation theorem of which ([Andréka *et al.*, 1998]) was used in [Mikulás, 1995] for obtaining a weakly-sound completeness theorem for  $\mathbf{L}_n$  (which will be defined in Section 7). Cf. also [Mikulás, 1995, the open problem below Thm.1.3.11, p.27].

<sup>98</sup>Here,  $x\bar{\Delta}y = \{x\Delta_i y : i < n\}$  and for a set  $H$  of formulas,  $\bar{\varepsilon}(H) = \bar{\delta}(H)$  denotes  $\{\varepsilon_i(\varphi) = \delta_i(\varphi) : i < m, \varphi \in H\}$ . We do not know whether the condition ‘ $\mathbf{HAlg}(\mathbf{L}) \models \dots$ ’ is needed for this theorem. It is not needed for direction ‘ $\Rightarrow$ ’. In the other direction, we can always obtain a weakly complete and weakly sound ‘ $\vdash$ ’, but this  $\vdash$  may not be completely Hilbert-style (this  $\vdash$  is more into the ‘Gabbay–Venema–Simon–Mikulás’

DEFINITION 53 (Compactness).

- $\mathcal{L}$  is *satisfiability compact* iff a set  $\Gamma$  has a model whenever all of its finite subsets  $\Sigma$  have models, i.e.

$$(\forall \Sigma \subseteq_{\omega} \Gamma) \text{Mod}(\Sigma) \neq \emptyset \implies \text{Mod}(\Gamma) \neq \emptyset.$$

- $\mathcal{L}$  is *consequence compact* if a consequence of  $\Gamma$  is always a consequence of a finite subset  $\Sigma$  already, i.e. iff

$$\Gamma \vDash \varphi \implies (\exists \Sigma \subseteq_{\omega} \Gamma) \Sigma \vDash \varphi.$$

A general logic  $\mathbf{L} = \langle \mathcal{L}^P : P \text{ is a set} \rangle$  is (satisfiability, consequence) compact if  $\mathcal{L}^P$  is such for all  $P$ .

We note that in general, consequence compactness and satisfiability compactness are independent properties, i.e. neither of them implies the other. However, for algebraizable general logics, consequence compactness implies satisfiability compactness. Also, if we have some kind of negation, then the two notions of compactness coincide.<sup>99</sup> In all our examples in Section 7, the two notions of compactness coincide.

Our next theorem characterizes consequence compactness of algebraizable general logics. For a proof see [Andréka *et al.*, 1993, Thm. 3.2.20], or [Andréka *et al.*, 1995, Cor. 3.10].

**THEOREM 54** (Characterization of compactness). *Assume that  $\mathbf{L}$  is an algebraizable general logic. Then  $\mathbf{L}$  is consequence compact iff  $\mathbf{Alg}(\mathbf{L})$  is closed under taking ultraproducts, i.e. iff  $\mathbf{Alg}(\mathbf{L}) = \mathbf{UpAlg}(\mathbf{L})$ . ■*

Now we turn to characterization of some definability properties. Beth's definability properties of logics were defined e.g. in [Barwise and Feferman, 1985]. Here we give the definitions in the framework of the present Paper.

DEFINITION 55 (Implicit definition, explicit definition).

Let  $\langle \mathcal{L}^P : P \text{ is a set} \rangle$  be a general logic. Let  $P \subsetneq Q$  be sets with  $F^P \neq \emptyset$ , and let  $R \stackrel{\text{def}}{=} Q \setminus P$ .

- A set  $\Sigma \subseteq F^Q$  of formulas defines  $R$  implicitly in  $Q$  iff a  $P$ -model can be extended to a  $Q$ -model of  $\Sigma$  at most one way, i.e. iff

$$(\forall \mathfrak{M}, \mathfrak{N} \in \text{Mod}^Q(\Sigma)) [mng_{\mathfrak{M}}^Q \upharpoonright F^P = mng_{\mathfrak{N}}^Q \upharpoonright F^P \longrightarrow mng_{\mathfrak{M}}^Q = mng_{\mathfrak{N}}^Q].$$

direction).

<sup>99</sup>More on this can be found in [Andréka *et al.*, 1993; Andréka *et al.*, 1995].



- $\Sigma$  defines  $R$  implicitly in  $Q$  in the strong sense iff, in addition, any  $P$ -model that in principle can, indeed can be extended to a  $Q$ -model of  $\Sigma$ , i.e. iff

$\Sigma$  defines  $R$  implicitly in  $Q$  and

$$(\forall \mathfrak{M} \in \text{Mod}^P([\text{Th}^Q \text{Mod}^Q \Sigma] \cap F^P)) (\exists \mathfrak{N} \in \text{Mod}^Q(\Sigma))$$

$$mng_{\mathfrak{M}}^Q \upharpoonright F^P = mng_{\mathfrak{M}}^P.$$

- $\Sigma$  defines  $R$  explicitly in  $Q$  iff any element of  $R$  has an ‘explicit definition’ that works in all models of  $\Sigma$ , i.e. iff

$$(\forall r \in R) (\exists \varphi_r \in F^P) (\forall \mathfrak{M} \in \text{Mod}^Q(\Sigma)) mng_{\mathfrak{M}}^Q(r) = mng_{\mathfrak{M}}^Q(\varphi_r).$$

- $\Sigma$  defines  $R$  local-explicitly in  $Q$  iff the above definition can vary from model to model, i.e. iff

$$(\forall \mathfrak{M} \in \text{Mod}^Q(\Sigma)) (\forall r \in R) (\exists \varphi_r \in F^P) mng_{\mathfrak{M}}^Q(r) = mng_{\mathfrak{M}}^Q(\varphi_r).$$

DEFINITION 56 (Beth definability properties). Let  $\mathbf{L}$  be a general logic.

- $\mathbf{L}$  has the (*strong*) *Beth definability* property iff for all  $P, Q, R$  and  $\Sigma$  as in Definition 55,

$$(\Sigma \text{ defines } R \text{ implicitly in } Q \implies \Sigma \text{ defines } R \text{ explicitly in } Q).$$

- $\mathbf{L}$  has the *local Beth definability* property iff for all  $P, Q, R$  and  $\Sigma$  as in Definition 55,

$$(\Sigma \text{ defines } R \text{ implicitly in } Q \implies \Sigma \text{ defines } R \text{ local-explicitly in } Q).$$

- $\mathbf{L}$  has the *weak Beth definability* property iff for all  $P, Q, R$  and  $\Sigma$  as in Definition 55,

$$(\Sigma \text{ defines } R \text{ implicitly in } Q \text{ in the strong sense} \implies \Sigma \text{ defines } R \text{ explicitly in } Q).$$

DEFINITION 57 (Patchwork property of models). Let  $\mathbf{L}$  be a general logic.  $\mathbf{L}$  has the patchwork property of models iff

for all sets  $P, Q$ , and models  $\mathfrak{M} \in M^P$ ,  $\mathfrak{N} \in M^Q$ ,

$$F^{P \cap Q} \neq \emptyset \text{ and } mng_{\mathfrak{M}}^P \upharpoonright (P \cap Q) = mng_{\mathfrak{N}}^Q \upharpoonright (P \cap Q) \implies$$

$$(\exists \mathfrak{P} \in M^{P \cup Q}) (mng_{\mathfrak{P}}^{P \cup Q} \upharpoonright F^P = mng_{\mathfrak{M}}^P \text{ and } mng_{\mathfrak{P}}^{P \cup Q} \upharpoonright F^Q = mng_{\mathfrak{N}}^Q).$$

Recall that if  $\mathbf{K}$  is a class of algebras, then by a morphism of  $\mathbf{K}$  we understand a triple  $\langle \mathfrak{A}, h, \mathfrak{B} \rangle$ , where  $\mathfrak{A}, \mathfrak{B} \in \mathbf{K}$  and  $h : \mathfrak{A} \rightarrow \mathfrak{B}$  is a homomorphism. A morphism  $\langle \mathfrak{A}, h, \mathfrak{B} \rangle$  is an *epimorphism of  $\mathbf{K}$*  iff for every  $\mathfrak{C} \in \mathbf{K}$  and every pair  $f, k : \mathfrak{B} \rightarrow \mathfrak{C}$  of homomorphisms we have  $f \circ h = k \circ h$  implies  $f = k$ . Typical examples of epimorphisms are the surjections. But for certain choices of  $\mathbf{K}$  there are epimorphisms of  $\mathbf{K}$  which are not surjective. Such is the case, e.g., when  $\mathbf{K}$  is the class of distributive lattices.

Let  $\mathbf{K}_0 \subseteq \mathbf{K}$  be two classes of algebras. Let  $\langle \mathfrak{A}, h, \mathfrak{B} \rangle$  be a morphism of  $\mathbf{K}$ .  $h$  is said to be  *$\mathbf{K}_0$ -extensible* iff for every algebra  $\mathfrak{C} \in \mathbf{K}_0$  and every homomorphism  $f : \mathfrak{A} \rightarrow \mathfrak{C}$  there exists some  $\mathfrak{N} \in \mathbf{K}_0$  and  $g : \mathfrak{B} \rightarrow \mathfrak{N}$  such that  $\mathfrak{C} \subseteq \mathfrak{N}$  and  $g \circ h = f$ . It is important to emphasize here that  $\mathfrak{C}$  is a concrete subalgebra of  $\mathfrak{N}$  and *not* only is embeddable into  $\mathfrak{N}$ .

**THEOREM 58** (Characterization of Beth properties<sup>100</sup>). *Let  $\mathbf{L}$  be an algebraizable general logic which has the patchwork property of models. Then (i)–(iii) below hold.*

- (i)  $\mathbf{L}$  has the Beth definability property iff all the epimorphisms of  $\mathbf{Alg}(\mathbf{L})$  are surjective.
- (ii)  $\mathbf{L}$  has the local Beth definability property iff all the epimorphisms of  $\mathbf{Alg}_m(\mathbf{L})$  are surjective.
- (iii)  $\mathbf{L}$  has the weak Beth definability property iff every  $\mathbf{Alg}_m(\mathbf{L})$ -extensible epimorphism of  $\mathbf{Alg}_m(\mathbf{L})$  is surjective. ■

In the formulation of Theorem 58 (ii),(iii) above, it was important that  $\mathbf{Alg}_m(\mathbf{L})$  is not an abstract class in the sense that it is not closed under isomorphisms, since the definition of  $\mathbf{K}$ -extensibility strongly differentiates isomorphic algebras.

If  $\mathbf{K}$  is a class of algebras, then  $\mathbf{maxK}$  denotes the class of all  $\subseteq$ -maximal elements of  $\mathbf{K}$ :

$$\mathbf{maxK} \stackrel{\text{def}}{=} \{ \mathfrak{A} \in \mathbf{K} : (\forall \mathfrak{B} \in \mathbf{K})(\mathfrak{A} \subseteq \mathfrak{B} \implies \mathfrak{A} = \mathfrak{B}) \}.$$

We note that e.g.  $\mathbf{maxCs}_n$  is the class of all full  $\mathbf{Cs}_n$ s.

We will use the notions of ‘reflective subcategory’, and ‘limits of diagrams of algebras’ as in [Mac Lane, 1971]. We will not recall these. Throughout,

<sup>100</sup>The proof of (i) is in [Németi, 1982] and in [Hoogland, 1996]. A less general version of (i) is proved in [Henkin, Monk and Tarski, 1985, Thm.5.6.10]. Part (ii) is due to J. Madarász. An early version of (iii) is in [Sain, 1990], and the full version is proved in [Hoogland, 1996]. The finite Beth property is obtained from the Beth property by restricting  $R$  to be finite. The emphasis in [Németi, 1982] and [Henkin, Monk and Tarski, 1985] was on the finite Beth property. E. Hoogland and J. Madarász [1997] extended the characterization of Theorem 58(i) to the broader (than algebraizable) class of equivalential logics.

by a reflective subcategory we understand a full and isomorphism closed one.

The weak Beth property was introduced in [Friedman, 1973] (cf. references of [Barwise and Feferman, 1985]) and has been investigated since then, cf. e.g. [Barwise and Feferman, 1985, pp. 73–76, 689–716].

**THEOREM 59** (Characterization of weak Beth property<sup>101</sup>). *Let  $\mathbf{L}$  be an algebraizable general logic which has the patchwork property of models. Assume that every element of  $\mathbf{Alg}_m(\mathbf{L})$  can be extended to a maximal element of  $\mathbf{Alg}_m(\mathbf{L})$ , i.e. that  $\mathbf{Alg}_m(\mathbf{L}) \subseteq \mathbf{SmaxAlg}_m(\mathbf{L})$ . Then conditions (i)–(iii) below are equivalent.*

- (i)  $\mathbf{L}$  has the weak Beth definability property.
- (ii)  $\mathbf{Alg}(\mathbf{L})$  is the smallest full reflective subcategory of  $\mathbf{Alg}(\mathbf{L})$  containing  $\mathbf{maxAlg}_m(\mathbf{L})$ .
- (iii)  $\mathbf{maxAlg}_m(\mathbf{L})$  generates  $\mathbf{Alg}(\mathbf{L})$  by taking limits of diagrams of algebras. I.e. there is no limit-closed proper subclass separating these two classes of algebras. ■

Now we turn to characterizing Craig’s interpolation properties.

**DEFINITION 60** (interpolation properties). Let  $\mathcal{L} = \langle F, M, mng, \models \rangle$  be a logic with connectives. For each formula  $\varphi \in F$  let  $\text{voc}(\varphi)$  denote the set of atomic formulas occurring in  $\varphi$ . Let  $\rightarrow$  be a binary connective of  $\mathcal{L}$ .

- $\mathcal{L}$  has the  $\models$ -interpolation property iff  
for all  $\varphi, \psi \in F$  such that  $\varphi \stackrel{\bullet}{\models} \psi$  there is  $\chi \in F$  such that  $\text{voc}(\chi) \subseteq \text{voc}(\varphi) \cap \text{voc}(\psi)$  and  $\varphi \stackrel{\bullet}{\models} \chi \stackrel{\bullet}{\models} \psi$ .
- $\mathcal{L}$  has the  $\rightarrow$ -interpolation property iff  
for all  $\varphi, \psi \in F$  such that  $\models \varphi \rightarrow \psi$  there is  $\chi \in F$  such that  $\text{voc}(\chi) \subseteq \text{voc}(\varphi) \cap \text{voc}(\psi)$  and  $\models \varphi \rightarrow \chi$  and  $\models \chi \rightarrow \psi$ .

Next we recall from the literature the amalgamation and superamalgamation properties of classes of algebras. Let  $\mathbf{K}$  be a class of algebras. We say that  $\mathbf{K}$  has the *amalgamation property* iff any two algebras in  $\mathbf{K}$  can be jointly embedded into a third element of  $\mathbf{K}$  such that we can require some parts of them to go to the same place. I.e. for any  $\mathfrak{A}, \mathfrak{B}, \mathfrak{C} \in \mathbf{IK}$  with  $\mathfrak{B} \supseteq \mathfrak{A} \subseteq \mathfrak{C}$ , there are  $\mathfrak{N} \in \mathbf{K}$  and injective homomorphisms (embeddings)  $f : \mathfrak{B} \rightarrow \mathfrak{N}$ ,  $h : \mathfrak{C} \rightarrow \mathfrak{N}$  such that  $f \upharpoonright \mathfrak{A} = h \upharpoonright \mathfrak{A}$ .

<sup>101</sup>This is due to I. Sain, J. Madarász, and I. Németi. For the origins of this characterization of weak Beth property see [Sain, 1990, p.223 and on].

By a *partially ordered algebra* we mean a structure  $\langle \mathfrak{A}, \leq \rangle$  where  $\mathfrak{A}$  is an algebra and  $\leq$  is a partial ordering on the universe  $A$  of  $\mathfrak{A}$ . A class  $\mathbf{K}$  of partially ordered algebras has the *superamalgamation property* iff any two algebras as above can be embedded in a third one such that only the necessary coincidences and ordering would hold, i.e. if for any  $\mathfrak{A}_i \in \mathbf{K}$ ,  $i \leq 2$  and for any embeddings  $i_1 : \mathfrak{A}_0 \rightarrow \mathfrak{A}_1$  and  $i_2 : \mathfrak{A}_0 \rightarrow \mathfrak{A}_2$  there exist an  $\mathfrak{A} \in \mathbf{K}$  and embeddings  $m_1 : \mathfrak{A}_1 \rightarrow \mathfrak{A}$  and  $m_2 : \mathfrak{A}_2 \rightarrow \mathfrak{A}$  such that  $m_1 \circ i_1 = m_2 \circ i_2$  and for  $\{j, k\} = \{1, 2\}$ ,  $(\forall x \in A_j)(\forall y \in A_k)[m_j(x) \leq m_k(y) \implies (\exists z \in A_0)(x \leq i_j(z) \text{ and } i_k(z) \leq y)]$ .

DEFINITION 61. Let  $\mathcal{L}$  be a compositional logic. We say that  $\mathcal{L}$  has a *deduction theorem* iff there is a binary derived connective  $\nabla$  such that for all  $\Sigma \subseteq F$ ,  $\varphi, \psi \in F$

$$\Sigma \cup \{\varphi\} \stackrel{\bullet}{\models} \psi \quad \text{iff} \quad \Sigma \stackrel{\bullet}{\models} \varphi \nabla \psi.$$

Such a  $\nabla$  is called a *deduction term*.

THEOREM 62 (Characterization of interpolation properties<sup>102</sup>).  
Let  $\mathcal{L}$  be an algebraizable semantic logic.

- (i) Assume that  $\mathcal{L}$  is consequence compact and usual conjunction  $\wedge$  is in  $Cn(\mathcal{L})$ . Assume that  $\mathcal{L}$  has a deduction theorem. Then  $\mathcal{L}$  has the  $\models$ -interpolation property iff  $\mathbf{Alg}(\mathcal{L})$  has the amalgamation property.
- (ii) Assume that  $\mathbf{Alg}(\mathcal{L})$  consists of normal BAOs. Assume that  $\mathbf{Alg}(\mathcal{L})$  is algebraized via the usual Boolean biconditional  $\leftrightarrow$ , i.e. in the filter-property  $\Delta$  is  $\leftrightarrow$ . Let  $\rightarrow$  denote the usual Boolean implication term. Then  $\mathcal{L}$  has the  $\rightarrow$ -interpolation property iff  $\mathbf{HSPA}(\mathcal{L})$  has the superamalgamation property. ■

The above is only a sample of the equivalence theorems in algebraic logic. Other kinds of investigations are connecting deduction property of a logic  $\mathcal{L}$  with  $\mathbf{Alg}(\mathcal{L})$  having equationally definable principal congruences (EDPC) [Blok and Pigozzi, 1989a; Blok and Pigozzi, 1989c; Blok and Pigozzi, 1997]; theorems connecting e.g. atomicity of the formula-algebra of  $\mathcal{L}$  with Gödel's incompleteness property of  $\mathcal{L}$  ([Németi, 1985; Németi, 1986]), theorems connecting logical meanings to neat-reducts of formula-algebras ([Amer, 1993; Sayed Ahmed, 1997]) etc.

<sup>102</sup>The proof of (i) is in [Czelakowski, 1982, Thm.3]. The proof of (ii) is in [Madarász, 1998], [Madarász, 1997].

## 7 EXAMPLES AND APPLICATIONS

In this Section we give some applications for the previous theorems. Most of the logics we use are well-known, but we recall their definitions for illustrating how they are special cases of the concept defined herein, and also for fixing our notation. More and also different kinds of examples are given in [Andréka *et al.*, 1993], [Németi and Andréka, 1994], [Andréka *et al.*, 1995].

**1. Classical sentential logic  $\mathbf{L}_S = \langle \mathcal{L}_S^P : P \text{ is a set} \rangle$ .**

Below, we often will omit the index  $S$ .

The set of logical connectives is  $Cn = \{\wedge, \neg\}$ ,  $\wedge$  is binary,  $\neg$  is unary. Let  $P$  be any set. Thus the set of formulas of  $\mathcal{L}_S^P$  is  $F_S^P = F(P, Cn)$ .

A model of sentential logic  $\mathcal{L}_S^P$  is a function assigning 0 (false) or 1 (true) to each atomic proposition  $p \in P$ . Thus the class  $M_S^P$  of models of  $\mathcal{L}_S^P$  is  ${}^P 2$ , the set of all functions mapping  $P$  to 2. (Recall that  $2 = \{0, 1\}$ .)

We can extend any model  $\mathfrak{M} : P \rightarrow 2$  to the set  $F_S^P$  of all formulas: for all  $\varphi, \psi \in F_S^P$  we let

$$\mathfrak{M}(\neg\varphi) = \begin{cases} 1 & \text{if } \mathfrak{M}(\varphi) = 0 \\ 0 & \text{if } \mathfrak{M}(\varphi) = 1, \end{cases}$$

$$\mathfrak{M}(\varphi \wedge \psi) = \begin{cases} 1 & \text{if } \mathfrak{M}(\varphi) = \mathfrak{M}(\psi) = 1 \\ 0 & \text{otherwise.} \end{cases}$$

Now, the meaning of  $\varphi$  in  $\mathfrak{M}$  is  $\mathfrak{M}(\varphi)$ , i.e.  $mng_S^P(\varphi, \mathfrak{M}) = \mathfrak{M}(\varphi)$ , and  $\varphi$  is valid in  $\mathfrak{M}$ ,  $\mathfrak{M} \models_S \varphi$ , iff  $\mathfrak{M}(\varphi) = 1$ .

We let  $\mathcal{L}_S^P \stackrel{\text{def}}{=} \langle F_S^P, M_S^P, mng_S^P, \models_S \rangle$  and  $\mathbf{L}_S \stackrel{\text{def}}{=} \langle \mathcal{L}_S^P : P \text{ is a set} \rangle$ .

By this, we have defined  $\mathbf{L}_S$ . We are going to show that  $\mathbf{L}_S$  is an algebraizable general logic with  $\mathbf{Alg}(\mathbf{L}_S) = \mathbf{BA}$ .

That  $\mathbf{L}_S$  is compositional comes immediately from the definition. Let  $\mathbf{2}$  denote the 2-element Boolean algebra with universe 2. Then  $\mathfrak{Mng}(\mathfrak{M}) = \mathbf{2}$  for all  $\mathfrak{M} \in M^P$ ,  $P \neq \emptyset$ . Thus  $\mathbf{Alg}_m(\mathbf{L}_S) = \{\mathbf{2}\}$ , and it can be seen from the definition that any homomorphism  $h : \mathfrak{F}^P \rightarrow \mathbf{2}$  is a meaning-function of some model (namely, that of  $h \upharpoonright P$ ), thus  $Mng^P = \text{Hom}(\mathfrak{F}^P, \{\mathbf{2}\})$ . Thus by Proposition 43,  $\mathbf{L}_S$  has the semantical substitution property, so  $\mathbf{L}_S$  is structural. Then  $\mathbf{Alg}(\mathbf{L}_S) = \mathbf{SP}\{\mathbf{2}\} = \mathbf{BA}$  by Theorem 44 (iii) (and also  $\mathbf{Alg}(\mathcal{L}_S^P) = \text{'BAs of cardinality } \leq \max(\omega, |P|)$ '). It can be seen that  $\mathbf{L}_S$  has the filter-property with  $m = n = 1$ ,  $\Delta_0(\varphi, \psi) \stackrel{\text{def}}{=} (\varphi \leftrightarrow \psi) \stackrel{\text{def}}{=} \neg(\neg\varphi \wedge \psi) \wedge \neg(\varphi \wedge \neg\psi)$ ,  $\varepsilon_0(\varphi) \stackrel{\text{def}}{=} \varphi$  and  $\delta_0(\varphi) \stackrel{\text{def}}{=} \text{TRUE} \stackrel{\text{def}}{=} \neg(\neg p \wedge p)$  for a fixed  $p \in P$ .

Thus  $\mathbf{L}_S$  is an algebraizable general logic with  $\mathbf{Alg}(\mathbf{L}_S) = \mathbf{BA}$ ,  $\mathbf{Alg}_m(\mathbf{L}_S) = \{\mathbf{2}\}$ .

By Theorem 54,  $\mathbf{L}_S$  is compact, because  $\mathbf{BA} = \mathbf{UpBA}$ . By Theorem 49,  $\mathbf{L}_S$  has a strongly complete and sound Hilbert-style inference system, because  $\mathbf{BA}$  is a finitely axiomatizable quasi-variety. Moreover, the proof of Theorem 49 (in [Andréka *et al.*, 1993]) constructs such a Hilbert-style inference system (given any axiomatization of  $\mathbf{BA}$ ). We give here the inference system we get from the proof.

In the next inference system, we will use  $\varphi, \psi, \chi$  and  $\delta$  as formula-variables, and we will use  $\leftrightarrow$  and  $TRUE$  as derived connectives.

The axioms are:

$$\begin{aligned} \varphi \wedge \psi &\leftrightarrow \psi \wedge \varphi, \\ \varphi \wedge (\psi \wedge \chi) &\leftrightarrow (\varphi \wedge \psi) \wedge \chi, \\ \varphi &\leftrightarrow \neg(\neg(\varphi \wedge \psi) \wedge \neg(\varphi \wedge \neg\psi)). \end{aligned}$$

The rules are as follows:

$$\begin{aligned} \frac{}{\varphi \leftrightarrow \varphi}, \quad \frac{\varphi \leftrightarrow \psi}{\psi \leftrightarrow \varphi}, \quad \frac{\varphi \leftrightarrow \psi, \psi \leftrightarrow \chi}{\varphi \leftrightarrow \chi}, \\ \frac{\varphi \leftrightarrow \psi}{\neg\varphi \leftrightarrow \neg\psi}, \quad \frac{\varphi \leftrightarrow \psi, \chi \leftrightarrow \delta}{(\varphi \wedge \chi) \leftrightarrow (\psi \wedge \delta)}, \\ \frac{\varphi \leftrightarrow TRUE}{\varphi}, \quad \frac{\varphi}{\varphi \leftrightarrow TRUE}. \end{aligned}$$

It is easy to check that  $\mathbf{L}_S$  has the patchwork property.  $\mathbf{L}_S$  has the Beth property by Theorem 58 (i), because epimorphisms are surjective in  $\mathbf{BA}$ . By Theorem 58 (iii),  $\mathbf{L}_S$  has the weak Beth property. So by Theorem 59,  $\mathbf{2}$  generates  $\mathbf{BA}$  by limits, since  $\mathbf{maxAlg}_m(\mathbf{L}_S) = \mathbf{Alg}_m(\mathbf{L}_S) = \{\mathbf{2}\}$ .

A deduction term for  $\mathbf{L}_S$  is  $\phi_0 \rightarrow \phi_1 \stackrel{\text{def}}{=} \neg(\phi_0 \wedge \neg\phi_1)$ . Since  $\mathbf{BA}$  has superamalgamation, Theorem 62 implies that  $\mathbf{L}_S$  has the interpolation properties.

The validity problem of  $\mathbf{L}_S$  is decidable, the set of admissible rules of  $\mathbf{L}_S$  is decidable, the set of valid formula-schemes of  $\mathbf{L}_S$  is decidable by Theorem 45 and Corollary 46, because the quasi-equational theory of  $\mathbf{BA}$  is decidable.

## 2. Sentential logic in a slightly different form, $\mathbf{L}_S'$ .

The set of connectives, thus the set of formulas are just like in the previous case. The models are different. Let  $P$  be a set.

$$M'_S \stackrel{\text{def}}{=} \{\langle W, v \rangle : W \text{ is a non-empty set and } v : P \rightarrow \mathcal{P}(W)\}.$$

Thus a model  $\mathfrak{M} = \langle W, v \rangle$  is a non-empty set together with an assignment assigning a subset of  $W$  to each  $p \in P$ . Let  $\mathfrak{M} = \langle W, v \rangle$  be a model. We call  $W$  the set of possible situations (or states, or worlds) of  $\mathfrak{M}$ . For any formula  $\varphi$ , we define  $\mathfrak{M}, w \Vdash \varphi$ , which we read as ‘ $\varphi$  is true in  $\mathfrak{M}$  at  $w$ ’, as follows:

$$\mathfrak{M}, w \Vdash p \quad \text{iff} \quad w \in v(p), \quad \text{for } p \in P.$$

$$\mathfrak{M}, w \Vdash \neg\varphi \quad \text{iff} \quad \mathfrak{M}, w \not\Vdash \varphi,$$

$$\mathfrak{M}, w \Vdash \varphi \wedge \psi \quad \text{iff} \quad (\mathfrak{M}, w \Vdash \varphi \text{ and } \mathfrak{M}, w \Vdash \psi).$$

We say that  $\varphi$  is valid in  $\mathfrak{M}$  if  $\mathfrak{M}, w \Vdash \varphi$  for all  $w \in W$ .

The above amounts to saying that the meaning-function  $mng_{\mathfrak{M}}$  is the homomorphic extension of  $v$  into the algebra  $\langle \mathcal{P}(W), \cap, \setminus \rangle$ , i.e.  $mng_{\mathfrak{M}} : \mathfrak{F}^P \rightarrow \langle \mathcal{P}(W), \cap, \setminus \rangle$ , and

$$\mathfrak{M}, w \Vdash \varphi \quad \text{iff} \quad w \in mng_{\mathfrak{M}}(\varphi),$$

$$\mathfrak{M} \models \varphi \quad \text{iff} \quad W = mng_{\mathfrak{M}}(\varphi).$$

Now,  $\mathbf{L}_S'$  is defined. It is compositional,

$$\begin{aligned} \mathbf{Alg}_m(\mathbf{L}_S') &= \text{setBA} \stackrel{\text{def}}{=} \mathbf{S}\{\langle \mathcal{P}(W), \cap, \setminus \rangle : W \text{ is a non-empty set}\} = \\ &= \text{‘the class of all non-trivial set Boolean algebras’}, \\ \text{and } Mng^P(\mathbf{L}_S') &= \text{Hom}(\mathfrak{F}^P, \text{setBA}). \end{aligned}$$

Thus  $\mathbf{L}_S'$  is algebraizable,  $\mathbf{Alg}(\mathbf{L}_S') = \mathbf{BA}$ ,  $\mathbf{Alg}_m(\mathbf{L}_S') = \text{setBA}$ . By Theorem 45, it has the same semantical consequence relation  $\models_{\bullet}$ , same admissible rules, same valid formula-schemes and same valid formulas as  $\mathbf{L}_S$ .

### 3. Modal logic S5.

The set of connectives is  $\{\wedge, \neg, \diamond\}$ ,  $\wedge$  binary,  $\neg, \diamond$  unary. The class of models is the same as for  $\mathbf{L}_S'$ . The ‘meaning of  $\diamond$ ’ is as follows:

$$\mathfrak{M}, w \Vdash \diamond\varphi \quad \text{iff} \quad (\mathfrak{M}, w' \Vdash \varphi \text{ for some } w' \in W).$$

This is the same as saying that

$$mng_{\mathfrak{M}}(\diamond\varphi) = C_0^W(mng_{\mathfrak{M}}(\varphi)), \quad \text{where for any set } X \subseteq W$$

$$C_0^W(X) = \begin{cases} W & \text{if } X \neq 0 \\ 0 & \text{otherwise.} \end{cases}$$

The rest of the definition of **S5** goes the same way as in the case of  $\mathbf{L}_S'$  above.

It can be checked that **S5** has the filter property with the same terms as sentential logic  $\mathbf{L}_S$ .

Thus **S5** is an algebraizable general logic with  $\mathbf{Alg}_m(\mathbf{S5}) = \mathbf{Cs}_1 = \mathbf{S}\{\langle \mathcal{P}(W), \cap, \setminus, C_0^W \rangle : W \text{ is a nonempty set}\}$ ,  $\mathbf{Alg}(\mathbf{S5}) = \mathbf{SPCs}_1 = \mathbf{RCA}_1$ .

A deduction term for **S5** is  $\neg\Diamond\neg\phi_0 \rightarrow \neg\Diamond\neg\phi_1$ .

Therefore **S5** is decidable, compact, has a strongly complete and sound Hilbert-style inference system, and has the Beth and interpolation properties by Theorems 45, 54, 49, 58, 62, because  $\mathbf{RCA}_1$  is a decidable, finitely axiomatizable variety having the superamalgamation property, see Theorem 17(ii) .

#### 4. Arrow logic $\mathbf{L}_{REL}$ .

The field of Arrow Logics grew out of application areas in Logic, Language and Computation, and plays an important role there, cf. e.g. [van Benthem, 1996; van Benthem, 1991a], and the proceedings of the Arrow Logic day at the conference 'Logic at Work' (Dec.1992, Amsterdam). These arrow logics go back to the investigations in [Tarski and Givant, 1987]. Tarski defined in 1951 basically  $\mathcal{L}_{REL}$  to give the first example of an undecidable propositional logic.

The set of connectives of  $\mathbf{L}_{REL}$  is  $\{\wedge, \neg, \circ, ^{-1}\}$ ,  $\wedge, \circ$  binary,  $\neg, ^{-1}$  unary. The models are as in **S5**, except that we require that the elements of  $W$  be all pairs over some set  $U$ , i.e.

$$M_{REL}^P \stackrel{\text{def}}{=} \{\langle W, v \rangle : W = U \times U \text{ for some } U \text{ and } v : P \rightarrow \mathcal{P}(W)\}.$$

The definition of  $\mathfrak{M}, w \Vdash \varphi$  is as in the previous case, and we define the meanings of  $\circ$  and  $^{-1}$  as

$$\mathfrak{M}, \langle u, z \rangle \Vdash \varphi \circ \psi \quad \text{iff} \quad (\mathfrak{M}, \langle u, x \rangle \Vdash \varphi \text{ and } \mathfrak{M}, \langle x, z \rangle \Vdash \psi \text{ for some } x),$$

$$\mathfrak{M}, \langle u, z \rangle \Vdash \varphi^{-1} \quad \text{iff} \quad \mathfrak{M}, \langle z, u \rangle \Vdash \varphi.$$

This amounts to saying that the meaning of  $\circ$  is relation composition, and the meaning of  $^{-1}$  is relation conversion, i.e.

$$mng_{\mathfrak{M}}(\varphi \circ \psi) = mng_{\mathfrak{M}}(\varphi) \circ mng_{\mathfrak{M}}(\psi),$$



$$mng_{\mathfrak{M}}(\varphi^{-1}) = (mng_{\mathfrak{M}}(\varphi))^{-1}.$$

Otherwise everything is the same as before, e.g.  $\mathfrak{M} \models \varphi$  iff  $mng_{\mathfrak{M}}(\varphi) = W$ .

Now,  $\mathbf{L}_{REL}$  is an algebraizable general logic with  $\mathbf{Alg}(\mathbf{L}_{REL}) = \mathbf{BRA}$ ,  $\mathbf{Alg}_m(\mathbf{L}_{REL}) = \mathbf{setBRA}$ .

Thus, by our equivalence theorems in Section 6 and by our algebraic theorems on  $\mathbf{BRA}$  in Section 1, we obtain that  $\mathbf{L}_{REL}$  is undecidable, compact, has no finitely complete and strongly sound Hilbert-style inference system. Since in  $\mathbf{BRA}$  epimorphisms are not surjective,<sup>103</sup>  $\mathbf{L}_{REL}$  does not have the Beth property.

A deduction term for  $\mathbf{L}_{REL}$  is  $TRUE \circ \phi_0 \circ TRUE \rightarrow TRUE \circ \phi_1 \circ TRUE$ . Since  $\mathbf{BRA}$  does not have the amalgamation property, by Theorem 62  $\mathbf{L}_{REL}$  does not have the interpolation property.

We can add ‘equality’ to  $\mathbf{L}_{REL}$ , obtaining  $\mathbf{L}_{REL}^{\bar{=}}$  as follows. We add  $\text{ld}$  to the set of connectives as a zero-ary connective, and we define its meaning as for any model  $\mathfrak{M} = \langle U \times U, v \rangle$

$$mng_{\mathfrak{M}}(\text{ld}) = \{ \langle u, u \rangle : u \in U \}.$$

Then  $\mathbf{L}_{REL}^{\bar{=}}$  is an algebraizable general logic with  $\mathbf{Alg}(\mathbf{L}_{REL}^{\bar{=}}) = \mathbf{RRA}$ ,  $\mathbf{Alg}_m(\mathbf{L}_{REL}^{\bar{=}}) = \mathbf{setRRA}$ . Just as in the previous cases we get properties of  $\mathbf{L}_{REL}^{\bar{=}}$  by using the theorems about  $\mathbf{RRA}$  stated in Section 1, and the equivalence theorems stated in Section 6.

### 5. First-order logic with $n$ variables, with substituted atomic formulas, $\mathbf{L}_n'$ .

Let  $n \in \omega$ , let  $V_n \stackrel{\text{def}}{=} \{v_i : i < n\}$ , our set of variables. Let  $\mathcal{R}$  be any set (our relation symbols). The set  $P$  of atomic formulas of the logic  $\mathcal{L}'_n \stackrel{\text{def}}{=} \mathcal{L}'_n{}^{\mathcal{R}}$  is

$$P \stackrel{\text{def}}{=} \{R(x_0, \dots, x_{n-1}) : R \in \mathcal{R}, x_0, \dots, x_{n-1} \in V_n\}.$$

The set of connectives is  $\{\wedge, \neg, v_i = v_j, \exists v_i : i, j < n\}$ ,  $\wedge$  binary,  $\neg, \exists v_i$  unary, and  $v_i = v_j$  zero-ary.<sup>104</sup> This defines the set  $F'_n{}^{\mathcal{R}}$  of formulas of  $\mathcal{L}'_n{}^{\mathcal{R}}$ . The class of models is the usual one,

$$M_n^{\mathcal{R}} \stackrel{\text{def}}{=} \{ \langle M, R^{\mathfrak{M}} \rangle_{R \in \mathcal{R}} : M \text{ is a nonempty set and } R^{\mathfrak{M}} \subseteq {}^n M \text{ for all } R \in \mathcal{R} \}.$$

<sup>103</sup>See the methods in [Sain, 1990].

<sup>104</sup>Notice that  $v_i = v_j$  is not an atomic formula but rather a zero-ary logical connective.

Let  $\mathfrak{M} = \langle M, R^{\mathfrak{M}} \rangle_{R \in \mathcal{R}}$  be a model. Then we define  $\models$  as before: Let  $\varphi \in F$  and  $h \in {}^n M$ . We call  $h$  an evaluation of the variables. Because of the tradition, we will write  $\mathfrak{M} \models \varphi[h]$  in place of  $\mathfrak{M}, h \models \varphi$ .

$$\mathfrak{M} \models R(v_{i_0} \dots v_{i_{n-1}})[h] \quad \text{iff} \quad \langle h(i_0), \dots, h(i_{n-1}) \rangle \in R^{\mathfrak{M}},$$

$$\mathfrak{M} \models v_i = v_j[h] \quad \text{iff} \quad h(i) = h(j),$$

$$\mathfrak{M} \models \exists v_i \varphi[h] \quad \text{iff} \quad (\mathfrak{M} \models \varphi[h'] \text{ for some } h' \in {}^n M \text{ such that } h \text{ and } h' \text{ differ at most at } i.)$$

$$\mathfrak{M} \models (\varphi \wedge \psi)[h] \text{ and } \mathfrak{M} \models \neg \varphi[h] \text{ are as before.}$$

$$\mathfrak{M} \models_n \varphi \quad \text{iff} \quad (\mathfrak{M} \models \varphi[h] \text{ for all } h \in {}^n M).$$

We define<sup>105</sup>

$$mng_{\mathfrak{M}}(\varphi) \stackrel{\text{def}}{=} \{h \in {}^n M : \mathfrak{M} \models \varphi[h]\}.$$

We define  $\mathcal{L}'_n{}^{\mathcal{R}} \stackrel{\text{def}}{=} \langle F_n{}^{\mathcal{R}}, M_n{}^{\mathcal{R}}, mng, \models_n \rangle$ , and  $\mathbf{L}'_n \stackrel{\text{def}}{=} \langle \mathcal{L}'_n{}^{\mathcal{R}} : \mathcal{R} \text{ is a set} \rangle$ .

Now,  $\mathbf{L}'_n$  is compositional, and has the filter property. But it is not structural in general (and then it is not semantically structural either) as the following example shows. Let

$$p \stackrel{\text{def}}{=} R(v_0, v_1) \quad \text{and} \quad q \stackrel{\text{def}}{=} R(v_0, v_0).$$

Then it can be checked that

$$\begin{aligned} \models_2 q &\leftrightarrow \exists v_1 (v_1 = v_0 \wedge p), & \text{but} \\ \not\models_2 p &\leftrightarrow \exists v_1 (v_1 = v_0 \wedge p). \end{aligned}$$

Another example is the following. Let

$$p \stackrel{\text{def}}{=} R(v_0, v_1) \quad \text{and} \quad q \stackrel{\text{def}}{=} R(v_1, v_0).$$

Then it can be checked that

$$\begin{aligned} \models_2 \exists v_1 p &\leftrightarrow \exists v_1 (v_1 = v_0 \wedge \exists v_0 q), & \text{but} \\ \not\models_2 \exists v_1 q &\leftrightarrow \exists v_1 (v_1 = v_0 \wedge \exists v_0 q). \end{aligned}$$

The logic  $\mathcal{L}'_n$  is *not structural* because the meanings of the atomic formulas are not independent of each other: as soon as we know the meaning of  $R(v_0 \dots v_{n-1})$ , this will determine the meanings of  $R(x_0 \dots x_{n-1})$  where  $x_0, \dots, x_{n-1}$  are arbitrary variables. It would be natural to treat

<sup>105</sup>In the literature,  $mng_{\mathfrak{M}}(\varphi)$  is called the relation defined by  $\varphi$  in the model  $\mathfrak{M}$ . Thus  $\mathfrak{Mng}(\mathfrak{M})$  is the algebra of  $n$ -variable definable relations in  $\mathfrak{M}$ .

only  $R(v_0 \dots v_{n-1})$  as an atomic formula. Then we would like to obtain the substituted atomic formulas  $R(x_0 \dots x_{n-1})$  as ‘complex’, built-up formulas. We will achieve this in two different ways. In the first case we will use Tarski’s observation that substitution can be expressed with quantifiers and equality<sup>106</sup> and in the second case we will introduce substitutions explicitly as logical connectives.<sup>107</sup>

**6. First-order logic with  $n$  variables, structural version<sup>108</sup>  $\mathbf{L}_n$ , for  $n < \omega$  and for  $n$  any ordinal.**

This is exactly like the previous example, except that we keep as atomic formulas only  $R(v_0, \dots, v_{n-1}), R \in \mathcal{R}$ . Since the order of the variables is fixed in our atomic formulas, we will simply write  $R$  in place of  $R(v_0, \dots, v_{n-1})$ . The set of connectives, the class of models, and the meaning function are exactly as before, the only difference is that now the set of atomic formulas is  $\mathcal{R}$  itself. When  $n$  is an infinite ordinal, everything is analogous (then  $R$  stands for  $R(v_0 \dots v_i \dots)_{i < n}$ ).

Notation:  $\mathbf{L}_n = \langle \mathcal{L}_n^{\mathcal{R}} : \mathcal{R} \text{ is a set} \rangle$ ,  $\mathcal{L}_n^{\mathcal{R}} = \langle F_n^{\mathcal{R}}, \dots \rangle$ .

Then,  $\mathbf{L}_n$  is compositional, and  $\text{Mng}_{\mathcal{L}_n^{\mathcal{R}}} = \text{Hom}(\mathfrak{F}^{\mathcal{R}}, \mathbf{Cs}_n)$ , so  $\mathbf{L}_n$  has the semantic substitution property by Proposition 43.  $\mathbf{L}_n$  has the filter property (with  $\phi_0, TRUE, \phi_0 \leftrightarrow \phi_1$  as  $\varepsilon, \delta, \Delta$ ), so  $\mathbf{L}_n$  is an algebraizable general logic. It is easy to check that  $\mathbf{Alg}_m(\mathcal{L}_n^{\mathcal{R}})$  is the class of  $|\mathcal{R}|$ -generated  $\mathbf{Cs}_n$ s, so  $\mathbf{Alg}_m(\mathbf{L}_n) = \mathbf{Cs}_n$ . Thus  $\mathbf{Alg}(\mathbf{L}_n) = \mathbf{RCA}_n$  by Theorem 44(ii), since  $\mathbf{RCA}_n = \mathbf{SPCs}_n$ . Using  $\mathbf{Alg}_m(\mathbf{L}_n) = \mathbf{Cs}_n$  and  $\mathbf{Alg}(\mathbf{L}) = \mathbf{RCA}_n$ , we begin to apply the theorems in Sections 2,6 to  $\mathbf{L}_n$ .

$\mathbf{L}_n$  is compact for all  $n$  by Theorem 54, because  $\mathbf{RCA}_n = \mathbf{UpRCA}_n$  by Theorem 17. For finite  $n$ , a deduction term for  $\mathbf{L}_n$  is  $\neg \exists v_0 \dots \exists v_{n-1} \neg \phi_0 \rightarrow \neg \exists v_0 \dots \exists v_{n-1} \neg \phi_1$ .

**THEOREM 63.** *Let  $n > 2$ . There is no weakly complete and strongly sound Hilbert-style inference system for  $\mathbf{L}_n$ . As a contrast, there are strongly complete and sound Hilbert-style inference systems for  $\mathbf{L}_2, \mathbf{L}_1, \mathbf{L}_0$ .*

**Proof.** For  $n > 2$ , this follows from Theorem 51, because  $\mathbf{RCA}_n$  is a non-finitely axiomatizable variety (by Theorems 17, 18, 19). For  $n \leq 2$ , this follows from Theorem 49, because  $\mathbf{RCA}_n, n \leq 2$  is a finitely axiomatizable quasi-variety (by Theorem 17). ■

<sup>106</sup>C.f. [Tarski, 1951; Tarski, 1965].

<sup>107</sup>For more detail see [Blok and Pigozzi, 1989, Appendix C], and [Henkin, Monk and Tarski, 1985, §4.3].

<sup>108</sup>This is called a *full restricted* first-order language in [Henkin, Monk and Tarski, 1985, §4.3]. ‘Restricted’ refers to the fact that we keep atomic formulas only with a fixed sequence of variables, and ‘full’ refers to the fact that the arity (or rank) of each relation symbol is  $n$ . This logic is investigated in [Blok and Pigozzi, 1989, Appendix C], too. Most of what we say in this example, can be generalized to the infinitary version  $\mathbf{L}_{\infty\omega}^n$  of  $\mathbf{L}_n$  studied in finite model theory, see Example 11 herein.

Soon we will give a strongly sound and complete inference system  $\models_{2+}$  for  $\mathbf{L}_2$ .

The above negative result can be meaningfully generalized to most known variants<sup>109</sup> of  $\mathbf{L}_n$ ,  $\mathbf{L}_n$  without equality, and the infinitary version  $\mathbf{L}_\infty^\omega$  of  $\mathbf{L}_n$  studied e.g. in finite model theory (e.g. [Ebbinghaus and Flum, 1995; Otto, 1997]). See Example 11 herein.

The proof of Theorem 63 above is a typical example of applying algebraic logic to logic. There are analogous theorems (using the same ‘general methodology’). An example is provided by the positive results giving completeness theorems for relativized versions of  $\mathbf{L}_n$  cf. e.g. [Andréka, van Benthem and Néméti, 1997] or [Néméti, 1996]. Different kinds of positive results relevant to Theorem 63 above are in [Sain, 1995; Sain and Gyuris, 1994].

**OPEN PROBLEM 64.** Is there a weakly complete and weakly sound Hilbert-style inference system for  $\mathbf{L}_n$ ,  $n > 2$ ?

By Theorem 52, Open Problem 64 above is equivalent to Problem 25 (i.e. whether  $\mathbf{RCA}_n = \mathbf{HK}$  for some finitely axiomatizable quasi-variety  $\mathbf{K}$ ), because  $\mathbf{RCA}_n \models (x \leftrightarrow y) = 1 \rightarrow x = y$ , where  $x \leftrightarrow y = -(x \oplus y)$ . Actually, in the present case a positive answer would imply the existence of a strongly complete and weakly sound ‘ $\vdash$ ’ for  $\mathbf{L}_n$ , because  $\mathbf{Alg}(\mathbf{L}_n)$  is a variety.

Next we turn to investigating inference systems suggested by the connections between  $\mathbf{RCA}_n$ ,  $\mathbf{CA}_n$ , and  $\mathbf{SNr}_n \mathbf{CA}_m$ , for  $m \geq n$  (see Theorem 28). We will define two provability relations,  $\vdash_n$  and  $\vdash_{n,m}$  for  $\mathbf{L}_n$ . (Of these,  $\vdash_n$  is given by a Hilbert-style inference system, while  $\vdash_{n,m}$  is not.)

In the following we will heavily use that  $F_n^{\mathcal{R}} \subseteq F_{n+m}^{\mathcal{R}}$  (which is so because the atomic formulas of  $\mathcal{L}_n^{\mathcal{R}}$  and  $\mathcal{L}_{n+m}^{\mathcal{R}}$  are identified).

**DEFINITION 65** (Provability relations  $\vdash_n$  and  $\vdash_{n,m}$  for  $\mathbf{L}_n$ <sup>110</sup>).

(i) First we define  $\vdash_n$  which will be given by the Hilbert-style inference system  $\langle Ax_n, Ru_n \rangle$ . In the formula-schemes below we will use  $\varphi, \psi$  as formula-variables (instead of  $\phi_0, \phi_1$ ), and  $\forall v_i, \rightarrow$  are derived connectives:

$$\begin{aligned} \forall v_i \varphi &\stackrel{\text{def}}{=} \neg \exists v_i \neg \varphi \\ \varphi \rightarrow \psi &\stackrel{\text{def}}{=} \neg(\neg \varphi \wedge \psi). \end{aligned}$$

Recall that  $\exists v_i, v_i = v_j$  are logical connectives for  $i, j < n$ .

$Ax_n$  consists of the following formula-schemes of  $\mathbf{L}_n$ : For all  $i, j, k < n$

<sup>109</sup>E.g. to  $\mathbf{L}_n'$  of Example 5.

<sup>110</sup> $\vdash_{n,m}$ , in a slightly different form, is defined in [Henkin, Monk and Tarski, 1985, p. 157], and also in [Néméti, 1986; Néméti, 1992; Néméti, 1995; Blok and Pigozzi, 1989; Czelakowski and Pigozzi, 1999]. In [Blok and Pigozzi, 1989],  $\vdash_n$  is denoted by  $\vdash_{PR_n}$ . One could get the definition of  $\vdash_n$  by mechanically translating the  $\mathbf{CA}_n$ -axioms, as we did with the  $\mathbf{BA}$ -axioms in Example 1, and then polishing the so obtained axioms and rules.

$\chi$ ,  $\chi$  a propositional tautology,<sup>111</sup> i.e., a valid formula-scheme of  $\mathbf{L}_0$

$$\forall v_i(\varphi \rightarrow \psi) \rightarrow (\forall v_i\varphi \rightarrow \forall v_i\psi)$$

$$\forall v_i\varphi \rightarrow \varphi$$

$$\forall v_i\forall v_j\varphi \rightarrow \forall v_j\forall v_i\varphi$$

$$\forall v_i\varphi \rightarrow \forall v_i\forall v_i\varphi$$

$$\exists v_i\varphi \rightarrow \forall v_i\exists v_i\varphi$$

$$v_i = v_i$$

$$v_i = v_j \rightarrow (v_i = v_k \rightarrow v_j = v_k)$$

$$\exists v_i(v_i = v_j)$$

$$v_i = v_j \rightarrow \forall v_k(v_i = v_j) \quad \text{if } k \neq i, j.$$

$$v_i = v_j \rightarrow (\varphi \rightarrow \forall v_i(v_i = v_j \rightarrow \varphi)), \quad \text{if } i \neq j.$$

$Ru_n$  consists of the rules *Modus Ponens* ( $MP$ ) and *Generalization* ( $G$ ) <sub>$i$</sub>  for  $i < n$ , where ( $MP$ ) and ( $G$ ) <sub>$i$</sub>  are, respectively:

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi} \quad \text{and} \quad \frac{\varphi}{\forall v_i\varphi}.$$

Now,  $\vdash_n$  is the derivability relation given by  $\langle Ax_n, Ru_n \rangle$  (for  $\mathbf{L}_n$ ).

(ii) To define  $\vdash_{n,m}$ , let  $n \leq m$ . Let  $Ax_m^n \subseteq F_m^{\mathcal{R}}$  consist of the following formulas:

$$(R) \quad R \rightarrow \forall v_i R \quad \text{if } n \leq i < m, \quad \text{and } R \in \mathcal{R}.$$

Now the inference system  $\vdash_{n,m} \subseteq \mathcal{P}(F_n^{\mathcal{R}}) \times F_n^{\mathcal{R}}$  is defined to be ' $(Ax_m^n \vdash_m)$  restricted to  $\mathbf{L}_n$ ', i.e.

$$\vdash_{n,m} \stackrel{\text{def}}{=} \{ \langle \Sigma, \varphi \rangle : \Sigma \cup Ax_m^n \vdash_m \varphi, \quad \Sigma \cup \{ \varphi \} \subseteq F_n^{\mathcal{R}} \}.$$

Thus, in an  $\vdash_{n,m}$ -proof, in addition to the instances of  $Ax_m$ , we also can use  $R \rightarrow \forall v_i R$ , for  $n \leq i < m$ ,  $R \in \mathcal{R}$ . If  $\vdash_{n,m} \varphi$ , then we say that ' $\varphi$  is provable with  $m$  variables', or ' $\varphi$  is  $m$ -variable provable'.

It is not hard to check that both  $\vdash_n$  and  $\vdash_{n,m}$  are strongly sound for  $\mathbf{L}_n$ , and  $\vdash_n$  is given by a Hilbert-style inference system (if  $n$  is finite).

<sup>111</sup>To keep  $Ax_n$  finite for  $n < \omega$ , we replace the infinitely many schemes here with  $Ax_0$ , where  $\langle Ax_0, \{(MP)\} \rangle$  is a strongly complete and sound Hilbert-style inference system for  $\mathbf{L}_0$ . Such systems are known, cf. e.g. [Andréka *et al.*, to appear].

Therefore, from Theorem 63 we can conclude that there are infinitely many valid  $\mathbf{L}_n$ -formulas which are not  $\vdash_n$ -provable, i.e.  $\vdash_n$  is incomplete for  $\mathbf{L}_n$  (in a rather strong way). On the other hand, we will see below that  $\vdash_{n,\omega}$  is strongly complete for  $\mathbf{L}_n$ .

We are going to prove that the inference system  $\vdash_n$  is the logical equivalent of the algebraic axiom system defining the variety  $\mathbf{CA}_n$ .

Any formula  $\varphi \in \mathcal{L}_n^{\mathcal{R}}$  can be identified with a term in the algebraic language of  $\mathbf{CA}_n$  such that the elements of  $\mathcal{R}$  are considered as (algebraic) variables, assuming that we identify the operations of  $\mathbf{CA}_n$  with the connectives of  $\mathbf{L}_n$ . Hence  $\varphi = 1$  is an equation in the language of  $\mathbf{CA}_n$  (for  $\varphi \in \mathcal{L}_n^{\mathcal{R}}$ ). We write  $\varphi \in \mathcal{L}_n^{\mathcal{R}}$  for  $\varphi \in F_n^{\mathcal{R}}$ . Also,  $\varphi \in \mathbf{L}_n$  means  $(\exists \text{ set } \mathcal{R}) \varphi \in \mathcal{L}_n^{\mathcal{R}}$ .

**THEOREM 66.** *Let  $\varphi \in \mathcal{L}_n^{\mathcal{R}}$ ,  $n$  any ordinal,  $\mathcal{R}$  any set. Then (i)–(iii) below hold for all  $n \leq m$ .*

- (i)  $\vdash_n \varphi$  iff  $\mathbf{CA}_n \models \varphi = 1$ .
- (ii)  $\vdash_{n,m} \varphi$  iff  $\mathbf{SNr}_n \mathbf{CA}_m \models \varphi = 1$ .
- (iii)  $\models_n \varphi$  iff  $\mathbf{RCA}_n \models \varphi = 1$ .

**Proof.** (i)–(iii) are proved in [Henkin, Monk and Tarski, 1985] as Corollary 4.3.26, Theorem 4.3.25, and Theorem 4.3.17, respectively. See also 4.3.57, 4.3.59 therein. To check that  $\vdash_r$  of [Henkin, Monk and Tarski, 1985] is the same as our  $\vdash_{n,m}$ , it suffices to check that in the proof of 4.3.22, only the axioms of our  $Ax_n$  are used. ■

Now we are ready to state the logical corollaries of Theorem 28(ii).

**COROLLARY 67.**

- (i)  $\vdash_{n,n+\omega}$  is strongly complete for  $\mathbf{L}_n$ , for all  $n$ .
- (ii)  $\vdash_{n,n+m}$  is not even weakly complete for  $\mathbf{L}_n$ , if  $m < \omega$ .
- (iii) Let  $\varphi \in \mathbf{L}_n$ . Then  $\models_n \varphi$  iff  $\vdash_{n,n+m} \varphi$  for some  $m < \omega$ .
- (iv) For all  $n, m < \omega$  there are valid  $n$ -variable formulas which cannot be proved with  $m$  variables. For each valid  $n$ -variable formula  $\varphi$  there is an  $m < \omega$  such that  $\varphi$  is  $m$ -variable provable.

**Proof.** (i): Weak completeness of  $\vdash_{n,n+\omega}$  follows immediately from Theorem 66, Theorem 28 ( $\mathbf{RCA}_n = \mathbf{SNr}_n \mathbf{CA}_{n+\omega}$ ). For  $n < \omega$ , then strong completeness follows, because  $\mathbf{L}_n$  is compact and has a deduction theorem. For  $n \geq \omega$ ,  $\mathbf{L}_n$  is still compact, but it does not have a deduction theorem.

However,  $\vdash_{n,n+\omega}$  is strongly complete for  $\mathbf{L}_n$  by [Henkin, Monk and Tarski, 1985, 4.3.23(ii)].

(ii)–(iv) follow from Theorem 66 and Theorem 28. E.g., assume  $\varphi \in \mathbf{L}_n$ . Then  $\models_n \varphi$  iff  $\mathbf{RCA}_n \models \varphi = 1$  iff (by  $\mathbf{RCA}_n = \mathbf{SNr}_n \mathbf{CA}_{n+\omega}$ )  $\mathbf{SNr}_n \mathbf{CA}_{n+\omega} \models \varphi = 1$  iff  $\models_{n,n+\omega} \varphi$  iff (by the definition of  $\vdash_{n,m}$ )  $\vdash_{n,n+m} \varphi$  for some  $m < \omega$ . ■

We note that Corollary 67 speaks also about usual first-order logic, because an  $n$ -variable formula is valid in  $\mathbf{L}_n$  iff it is valid in usual first-order logic (and every first order formula  $\varphi$  has a normal form  $\varphi'$  which is in  $\mathbf{L}_n$ , for some  $n \in \omega$ ).

To investigate further the provability relations  $\vdash_n, \vdash_{n,m}$ , now we compare their ‘deductive powers’. Results in cylindric algebra theory yield the following.

**THEOREM 68** (The deductive powers of  $\vdash_n, \vdash_{n,m}$ ).

(i) For any  $1 < n < \omega$ ,  $\vdash_n \neq (\vdash_{n+1} \upharpoonright \mathbf{L}_n)$ , i.e. there is an  $n$ -variable formula  $\varphi$  such that

$$\not\vdash_n \varphi \quad \text{but} \quad \vdash_{n+1} \varphi.$$

(ii) If  $n \geq \omega$ , then  $\vdash_{n,m} \neq \vdash_{n,m+1}$  for all  $m < \omega$ , i.e. there is a  $\varphi \in \mathbf{L}_\omega$  such that

$$\not\vdash_{\omega,m} \varphi \quad \text{and} \quad \vdash_{\omega,m+1} \varphi.$$

(iii)  $\vdash_2 \neq \vdash_{2,3} = \vdash_{2,m}$  for all  $m \geq 3$ .

(iv) If  $n \leq 1$  or  $n \geq \omega$ , then for any  $n$ -variable  $\varphi$ ,  $\vdash_n \varphi$  iff  $\vdash_m \varphi$ , for all  $m \geq n$ . If  $n \leq 1$ , then  $\vdash_n = \vdash_{n,m}$  for all  $m \geq n$ .

**Proof.** (i) follows from Theorem 66 and [Henkin, Monk and Tarski, 1971, 2.6.14], as follows. Let  $2 \leq n < \omega$ . Then there is an equation  $e$  in the language of  $\mathbf{CA}_n$  such that  $\overline{\mathbf{CA}}_n \not\models e$  and  $\mathbf{CA}_{n+1} \models e$ , by [Henkin, Monk and Tarski, 1971, 2.6.14]. We may assume that  $e$  is of form  $\varphi = 1$ , and then Theorem 66(i) finishes the proof.

Similarly, (iv) follows from Theorem 66 and [Henkin, Monk and Tarski, 1971, 2.6.8, 2.6.9], and from  $\mathbf{CA}_n = \mathbf{RCA}_n = \mathbf{SNr}_n \mathbf{CA}_{n+m}$  if  $n \leq 1$ .

(ii) follows from (Theorem 66(ii) and)  $\mathbf{SNr}_n \mathbf{CA}_{n+m} \neq \mathbf{SNr}_n \mathbf{CA}_{n+m+1}$  for all  $n \geq \omega$ ,  $m < \omega$  which is an unpublished result of Don Pigozzi.

(iii) follows from  $\mathbf{CA}_2 \neq \mathbf{SNr}_2 \mathbf{CA}_3 = \mathbf{RCA}_2 = \mathbf{SNr}_2 \mathbf{CA}_\omega$ , see [Henkin, Monk and Tarski, 1971, 2.6.42, 3.2.65]. ■

REMARK 69. [Henkin, Monk and Tarski, 1971, Problem 2.12] asks, for any  $3 \leq n \leq m < \omega$ , whether  $\mathbf{SNr}_n\mathbf{CA}_m = \mathbf{SNr}_n\mathbf{CA}_{m+1}$ . A negative answer (for all such  $n, m$ ) is implied by the statement that  $\mathbf{SRaCA}_m \neq \mathbf{SRaCA}_{m+1}$  for  $3 \leq m < \omega$ . The consequence of such an answer for proof theory is that  $\vdash_{3,m} \neq \vdash_{3,m+1}$ . Results in this direction are proved e.g. in [Maddux, 1983; Maddux, 1991a; Andr eka, 1997; Goldblatt, 1999]. Hirsch, Hodkinson and Maddux recently proved:

THEOREM 69.1 ([Hirsch, Hodkinson, and Maddux, to appear])

1. For each  $m$  with  $3 \leq m < \omega$ ,  $\mathbf{SRaCA}_m$  strictly contains  $\mathbf{SRaCA}_{m+1}$ .
2. For  $3 \leq m \leq n < \omega$ ,  $\mathbf{SNr}_n\mathbf{CA}_m$  strictly contains  $\mathbf{SRaCA}_{m+1}$ .
3. For each  $m$  with  $3 \leq m < \omega$  there is a 3-variable first-order sentence  $\varphi$  such that  $\vdash_{3,m+1} \varphi$  but  $\not\vdash_{3,m} \varphi$ .

Furthermore, the relation algebra constructed in [Hirsch, Hodkinson, and Maddux, to appear] that witnesses (1) has the property of being generated by a single element. It follows that  $\varphi$  can be taken to be a sentence in a signature consisting of a single binary relation symbol.

Given that  $\mathbf{SNr}_n\mathbf{CA}_m, \mathbf{SRaCA}_m$  strictly contain  $\mathbf{SNr}_n\mathbf{CA}_{m+1}, \mathbf{SRaCA}_{m+1}$  respectively, for  $3 \leq n \leq m < \omega$ , it is natural to ask if the inclusions can be axiomatized finitely. In other words, is there a finite set  $\sigma_m$  of first-order sentences such that for any  $\mathfrak{A} \in \mathbf{SRaCA}_m$ , we have  $\mathfrak{A} \models \sigma_m$  if and only if  $\mathfrak{A} \in \mathbf{SRaCA}_{m+1}$  (and a similar question about neat cylindric reducts)? For  $m = 3$ , the answer is yes:  $\sigma_3$  can be taken to be single sentence expressing the associative law. For  $m > 3$ , we get a negative answer; again, this has a consequence for proof theory.

THEOREM 69.2 ([Hirsch and Hodkinson, to appear])

1. For  $4 \leq m < \omega$ , the variety  $\mathbf{SRaCA}_{m+1}$  cannot be axiomatized relative to  $\mathbf{SRaCA}_m$ , using only finitely many first-order sentences.
2. For  $3 \leq n < m < \omega$ , the variety  $\mathbf{SNr}_n\mathbf{CA}_{m+1}$  cannot be axiomatized relative to  $\mathbf{SNr}_n\mathbf{CA}_{m+1}$ , using only finitely many first-order sentences.
3. There is no finite set of  $n$ -variable schemata whose  $n$ -variable instances, when added to  $\vdash_{n,m}$  as axioms, yield  $\vdash_{n,m+1}$ .

REMARK .  $\vdash_{n,m}$  is a (structural) derivability relation for  $\mathbf{L}_n$  in the sense of [Blok and Pigozzi, 1989], i.e. for any  $\varphi \in \mathcal{L}_n^{\mathcal{R}}$  and  $s : \mathcal{R} \rightarrow \mathcal{L}_n^{\mathcal{R}}$ , if  $\vdash_{n,m} \varphi$  then  $\vdash_{n,m} \hat{s}(\varphi)$ . This follows from Theorem 66(ii). Theorem 66(ii) also implies that  $\vdash_{n,n+1}$  for  $n < \omega$  can be given by some Hilbert-style inference system  $\langle Ax'_n, Ru'_n \rangle$ ; while  $\vdash_{n,n+m}$  with  $m \geq 2$  cannot be given with such. The latter is so because  $\mathbf{SNr}_n\mathbf{CA}_{n+1}$  is a finitely axiomatizable variety while  $\mathbf{SNr}_n\mathbf{CA}_{n+m}$ ,  $m \geq 2$  is not finitely axiomatizable (see [Andr eka, 1997,



Thm. 2.3]), and then one can use the presently discussed ‘methodology of algebraization’, cf. Theorem 50, to infer the above information.

$\vdash_{n,n+\omega}$  is strongly complete for  $\mathbf{L}_n$ , but the  $\vdash_{n,n+\omega}$ -proofs use formulas that are not in  $\mathbf{L}_n$ . Different kinds of complete inference systems for  $\mathbf{L}_n$ , where the proofs use only  $\mathbf{L}_n$ -formulas, are in [Simon, 1991; Venema, 1991; Marx and Venema, 1997; Mikulás, 1995; Mikulás, 1996]. A common feature of the latter inference systems is that they are not strongly sound. (This is natural to expect because by Theorem 63 there cannot exist strongly sound and complete Hilbert-style inference systems for  $\mathbf{L}_n$  if  $n > 2$ .)

Now we turn to  $\mathbf{L}_2$ . If  $n \leq 1$ , then  $\vdash_n$  is strongly complete for  $\mathbf{L}_n$  by Theorem 66, because then  $\mathbf{CA}_n = \mathbf{RCA}_n$ .  $\vdash_2$  is not complete for  $\mathbf{L}_2$  (but  $\vdash_{2,3}$  is). We will show that if we add a rule or axiom expressing

$$(*) \quad |Dom(R)| \leq 1 \implies R = Dom(R) \times Rng(R),$$

(and the same for  $Rng(R)$ ), then we get a strongly complete Hilbert-style inference system for  $\mathbf{L}_2$ . Namely, consider the following formula-schema and rule for  $i \neq j, i, j < 2$ :

$$(SA) \quad \exists v_0 \varphi \wedge \exists v_1 \varphi \wedge \neg \varphi \rightarrow \exists v_i (\exists v_j (v_0 = v_1 \wedge \exists v_i \varphi) \wedge v_0 \neq v_1)$$

$$(SR) \quad \frac{[\exists v_i \varphi \wedge \exists v_j (v_0 = v_1 \wedge \exists v_i \varphi)] \rightarrow v_0 = v_1}{\varphi \leftrightarrow (\exists v_0 \varphi \wedge \exists v_1 \varphi)}.$$

**THEOREM 70.** *Both  $\langle Ax_2 \cup (SA), Ru_2 \rangle$  and  $\langle Ax_2, Ru_2 \cup (SR) \rangle$  are strongly complete for  $\mathbf{L}_2$ .*

**Proof.** This follows from Theorem 66 and Theorem 17(iii). ■

Now we turn to checking what Theorems 58 and 62 say about definability and interpolation properties of  $\mathbf{L}_n$ .  $\mathbf{L}_n$  has the patchwork property of models.

So  $\mathbf{L}_n$  for  $n \geq 2$  does not have the local Beth definability property by Theorem 58, because epimorphisms are not surjective in  $\mathbf{Cs}_n$  (see [Kiss *et al.*, 1983] and [Madarász, 1999, T.7.4.(i)], for  $2 \leq n < \omega$  see [Andréka, Comer and Németi, 1983], for  $n \geq \omega$  see [Németi, 1988]; [Sain, 1990, Thm.10]).

It is proved in [Kearnes, Sain and Simon, to appear] that  $\mathbf{L}_3$  does not have the weak Beth property, and we conjecture that this extends to  $3 < n < \omega$ , while  $\mathbf{L}_2$  has the weak Beth property. It is proved in [Kearnes, Sain and Simon, to appear] that for  $n < \omega$ ,  $\mathbf{L}_n$  has the weak local Beth property. By Theorem 62,  $\mathbf{L}_n$  does not have the interpolation property for all  $n > 1$ , since  $\mathbf{RCA}_n$  does not have the amalgamation property (see [Kiss *et al.*, 1983], this is a result of [Comer, 1969]).

Further definability and interpolation results for  $\mathbf{L}_n, \vdash_n$  (both  $n < \omega$  and  $n \geq \omega$ ) are in [Madarász, 1997b]. That paper is devoted to solving problems from [Pigozzi, 1972].

Summing up:

$\mathbf{L}_0$  is equivalent to sentential logic  $\mathbf{L}_S'$ .

$\mathbf{L}_1$  is equivalent to **S5**.

$\mathbf{L}_2$ : Our characterization theorems in Section 6 and the corresponding algebraic theorems in Section 2 give the following properties for  $\mathbf{L}_2$ :  $\mathbf{L}_2$  is decidable, it has a strongly complete and sound Hilbert-style inference system, which can be obtained from the equational axiomatization of  $\mathbf{RCA}_2$ .  $\mathbf{L}_2$  has the finite model property. The algebraic version of this is stated in [Henkin, Monk and Tarski, 1985, 3.2.66]. It does not have the Beth (definability) property, and it does not have the interpolation property. We conjecture that  $\mathbf{L}_2$  has the weak Beth property.

$\mathbf{L}_n$  for  $3 \leq n < \omega$ : The characterization theorems and the corresponding algebraic theorems give the following properties of  $\mathbf{L}_n$ :  $\mathbf{L}_n$  is undecidable,  $\mathbf{L}_n$  does not have a strongly sound and complete Hilbert-style inference system. It is open whether it has a weakly sound and weakly complete Hilbert-style inference system, cf. Problem 25 and Theorem 52.  $\mathbf{L}_n$  has neither the Beth property nor the interpolation property.

$\mathbf{L}_\omega$ : This is called ‘Finitary logic of infinitary relations’. Model theoretic results (using AL) are in [Németi, 1990].

### 7. First-order logic with $n$ variables with substitutions, with and without equality, $\mathbf{L}_n^{s=}$ , $\mathbf{L}_n^s$ , ( $n \leq \omega$ ).

First we define  $\mathbf{L}_n^{s=}$ . The set of connectives is  $\{\wedge, \neg, \exists v_i, [v_i/v_j], [v_i, v_j], v_i = v_j : i, j < n\}$ ,  $\wedge$  binary,  $v_i = v_j$  zero-ary, and the rest unary. Everything is as in the previous example, we only have to give the meanings of the logical connectives  $[v_i/v_j], [v_i, v_j]$ . Let  $\mathfrak{M}$  be a model, and recall<sup>112</sup> the operations  $[i/j], [i, j]$  mapping  $n$  to  $n$ . Now

$$\text{mng}_{\mathfrak{M}}([v_i/v_j]\varphi) = \{h \in {}^n M : h \circ [i/j] \in \text{mng}_{\mathfrak{M}}(\varphi)\},$$

$$\text{mng}_{\mathfrak{M}}([v_i, v_j]\varphi) = \{h \in {}^n M : h \circ [i, j] \in \text{mng}_{\mathfrak{M}}(\varphi)\}.$$

By this, we have defined  $\mathbf{L}_n^{s=}$ . It is not hard to check that  $\mathbf{L}_n^{s=}$  is an algebraizable general logic.

The theory of quasi-polyadic algebras **QPAs** is analogous with that of cylindric algebras. Exactly as cylindric algebras are the algebraic counterparts of quantifier logics with equality, **QPAs** are the algebraic counterparts of quantifier logics *without* equality, cf. Section 3 herein.  $\mathbf{RQPA}_n$  and  $\mathbf{QP}_{s_n}$  denote the classes of representable **QPAs** of dimension  $n$  and quasi-polyadic

<sup>112</sup> $[i/j]$  sends  $i$  to  $j$  and leaves everything else fixed, and  $[i, j]$  interchanges  $i$  and  $j$  and leaves everything else fixed.

set algebras (of dimension  $n$ ) respectively as introduced e.g. in [Németi, 1991] and in Section 3 herein. Analogously,  $\mathbf{QPEA}_n$  and  $\mathbf{QPse}_n$  denote the same classes but with equality.

Now,  $\mathbf{Alg}_m(\mathbf{L}_n^{s=}) = \mathbf{QPse}_n$ ,  $\mathbf{Alg}(\mathbf{L}_n^{s=}) = \mathbf{RQPEA}_n$ .

If we omit equality from the set of connectives, then we get the equality-free version  $\mathbf{L}_n^s$  of the logic. This is also an algebraizable general logic with  $\mathbf{Alg}_m(\mathbf{L}_n^s) = \mathbf{QPs}_n$ ,  $\mathbf{Alg}(\mathbf{L}_n^s) = \mathbf{RQPA}_n$ .

Now we turn to showing how to retrieve substituted atomic formulas  $R(v_{i_0} \dots v_{i_{n-1}})$  in  $\mathbf{L}_n$ ,  $\mathbf{L}_n^s$ . Here we assume  $n < \omega$ .

First we treat the case  $\mathbf{L}_n^s$ . Since a finite mapping can always be written as a product of  $[i, j]$ s and  $[i/j]$ s, we obtain that for any sequence  $x_0, \dots, x_{n-1}$  of variables there is a sequence  $[i_1, j_1], \dots, [i_\ell, j_\ell]$  of ‘substitutions’ such that for all models  $\mathfrak{M}$  and relation symbols  $R$ ,

$$mng_{\mathfrak{M}}(R(x_0 \dots x_{n-1})) = mng_{\mathfrak{M}}([v_{i_1}, v_{j_1}] \dots [v_{i_\ell}, v_{j_\ell}]R).$$

(Here the first meaning-function is taken from  $\mathcal{L}'_n$ , while the second one from  $\mathcal{L}^s_n$ .) This shows that in  $\mathbf{L}_n^s$  we do have our substituted atomic formulas back as ‘complex’ formulas. (On the other hand, the expressive power of  $\mathbf{L}_n^s$  is not bigger than that of  $\mathbf{L}'_n$ , because of the following. It can be proved with a simple induction that the meaning of the formula  $[v_i, v_j]\varphi$  is the same as that of the formula we get from  $\varphi$  by interchanging  $v_i$  and  $v_j$  in it everywhere (in the connectives  $\exists v_i$  also), and the meaning of the formula  $[v_i/v_j]\varphi$  coincides with that of the formula we get from  $\varphi$  by replacing  $v_i$  everywhere in it with  $v_j$ . Here  $\varphi$  is a formula of  $\mathcal{L}'_n$ .)

Now we show how to get substituted atomic formulas back in  $\mathbf{L}_n$  by using Tarski’s observation that substitution can be expressed with quantifiers and equality. By the above, it is enough to express the meaning of the formulas  $[v_i/v_j]\varphi$  and  $[v_i, v_j]\varphi$ , for  $i \neq j$ . So let  $\mathfrak{M}$  be a model and  $h$  an evaluation of the variables in  $M$ . Then it can be checked that

$$\mathfrak{M} \models_n ([v_i/v_j]\varphi \leftrightarrow \exists v_i(v_i = v_j \wedge \varphi))[h],$$

and if  $k \neq i, j$ ,  $\mathfrak{M} \models_n \varphi \leftrightarrow \exists v_k \varphi$ , then

$$\mathfrak{M} \models_n ([v_i, v_j]\varphi \leftrightarrow [v_i/v_k][v_k/v_j][v_j/v_i]\varphi)[h].$$

Thus to express  $[v_i, v_j]$  we need one extra free variable. We can get this e.g. by treating  $\mathcal{L}'_n^{\mathcal{R}}$  as the following theory of  $\mathcal{L}^{\mathcal{R}}_{n+1}$ :

$$\{(\exists v_n R) \leftrightarrow R : R \in \mathcal{R}\}$$

and then treat the atomic formula  $R(v_0, \dots, v_{n-1})$  of  $\mathcal{L}'_n$  as the atomic formula  $R$  of  $\mathcal{L}_{n+1}$ . For more on this see [Henkin, Monk and Tarski, 1985; Blok and Pigozzi, 1989, §4.3].

### 8. First-order logic, ranked<sup>113</sup> version, $\mathbf{L}_{FOL}^{ranked}$ .

The set of connectives is  $Cn = \{\wedge, \neg, \exists v_i, v_i = v_j : i, j < \omega\}$ ,  $\wedge$  binary,  $\neg, \exists v_i$  unary, and  $v_i = v_j$  zero-ary. (This is the same as that of  $\mathbf{L}_\omega$ .)

Let  $\mathcal{R}$  be a set (the set of relation-symbols), and let  $\rho : \mathcal{R} \rightarrow \omega$  be a function (the *rank-function*,  $\rho(R)$  is the *rank* of  $R$ ). First we define the logic  $\mathcal{L}_{FOL}^\rho$ .

Our atomic formulas will be  $R(v_0, \dots, v_{\rho(R)-1})$  for  $R \in \mathcal{R}$ . We do not include  $R(v_{i_0}, \dots, v_{i_{\rho(R)-1}})$  into the set of atomic formulas for the same reason as in our previous examples: because they would immediately make our logic unstructural. However, these substituted atomic formulas will be present in our logic as (complex) formulas, because they can be expressed by quantifiers and equality (see our previous remark on this). Since the sequence  $(v_0, \dots, v_{\rho(R)-1})$  of variables is determined by  $\rho$ , we will just write  $R$  in place of  $R(v_0, \dots, v_{\rho(R)-1})$ . (This will be convenient also when we will compare our present logic with  $\mathbf{L}_\omega$ .) Thus the set of atomic formulas is  $\mathcal{R}$ . Then the formula-algebra  $\mathfrak{F}^\rho$  of  $\mathcal{L}_{FOL}^\rho$  has universe  $F(\mathcal{R}, Cn)$ .

The models are  $\mathfrak{M} = \langle M, R^{\mathfrak{M}} \rangle_{R \in \mathcal{R}}$  where  $R^{\mathfrak{M}}$  is a  $\rho(R)$ -ary relation on  $M$  for all  $R \in \mathcal{R}$ . i.e.

$$M^\rho = \{ \langle M, R^{\mathfrak{M}} \rangle_{R \in \mathcal{R}} : R \subseteq {}^{(\rho R)}M \text{ for all } R \in \mathcal{R} \}.$$

Validity and the meaning function are practically the same as those of  $\mathbf{L}_\omega$ , therefore we only give here the concise algebraic definition: Let  $\mathfrak{M}$  be a model.

$$mng_{\mathfrak{M}}(R) = \{ h \in {}^\omega M : h \upharpoonright \rho(R) \in R^{\mathfrak{M}} \}, \text{ and}$$

$$mng_{\mathfrak{M}} : \mathfrak{F}^\rho \rightarrow \langle \mathfrak{P}({}^\omega M), c_i, \text{ld}_{ij} \rangle_{i, j < \omega} \text{ is a homomorphism.}$$

$$\mathfrak{M} \models \varphi \quad \text{iff} \quad mng_{\mathfrak{M}}(\varphi) = {}^\omega M.$$

Now

$$\mathbf{L}_{FOL}^{ranked} = \langle \mathcal{L}_{FOL}^\rho : \rho \text{ is a function into } \omega \rangle.$$

Let  $\mathbf{L} = \mathbf{L}_{FOL}^{ranked}$ . Then  $\mathbf{L}$  is compositional, and has the filter-property. Also we have that

$$\mathbf{Alg}_m(\mathbf{L}) = \text{Csf}_\omega \quad \text{and} \quad \mathbf{Alg}(\mathbf{L}) = \text{Lf}_\omega.$$

The second statement is proved in [Henkin, Monk and Tarski, 1985, 4.3.28(iii)].

<sup>113</sup>These are called ordinary languages in [Henkin, Monk and Tarski, 1985, §4.3].

But  $\mathbf{L}$  is not substitutional, even  $\mathcal{L}_{FOL}^\rho$  is not substitutional if  $\rho \neq \emptyset$ . An example is: Let  $R$  be an  $n$ -ary relation symbol in  $\rho$  and let  $\varphi$  denote the formula  $v_0 = v_1 \wedge \dots \wedge v_0 = v_n$ . Then

$$\models R \rightarrow \forall v_n R \quad \text{while} \quad \not\models \varphi \rightarrow \forall v_n \varphi.$$

It is easy to see that  $\mathbf{L}$  is compact. Since  $\mathbf{Uplf}_\omega \neq \mathbf{Lf}_\omega$ , this logic shows that the condition of structurality in Theorem 54 is necessary.

We can extend our inference system  $\vdash_\omega$  of the non-ranked logic  $\mathbf{L}_\omega$  to get a complete one for  $\mathbf{L}_{FOL}^{ranked}$ , as follows.

For any rank-function  $\rho : \mathcal{R} \rightarrow \omega$ , let  $Ax^\rho$  denote the set of the following formulas:

$$(R^i) \quad R \rightarrow \forall v_i R, \quad \text{if } \rho(R) \leq i < \omega, \quad \text{and } R \in \mathcal{R}.$$

( $Ax^\rho$  is a straightforward modification of  $Ax_n^m$ .) Then  $\vdash^\rho$  is defined as

$$\vdash^\rho \stackrel{\text{def}}{=} \{ \langle \Sigma, \varphi \rangle : \Sigma \cup Ax^\rho \vdash_\omega \varphi, \Sigma \cup \{ \varphi \} \subseteq \mathcal{L}_{FOL}^\rho \}.$$

Now,  $\vdash^\rho$  provides a complete inference system for the ranked version of first-order logic  $\mathbf{L}_{FOL}^{ranked}$ . I.e.:

**THEOREM 71** (Gödel's completeness theorem). *For every formula  $\varphi$  of  $\mathcal{L}_{FOL}^\rho$  we have*

$$\models \varphi \quad \text{iff} \quad \vdash^\rho \varphi.$$

**Proof.** This is a corollary of Theorem 66(ii) and  $\mathbf{Lf}_\omega \subseteq \mathbf{RCA}_\omega = \mathbf{HSPCsf}_\omega$  (Theorems 28(i), 18(iii)), as follows. Let  $\equiv \stackrel{\text{def}}{=} \{ (\varphi, \psi) : \vdash^\rho \varphi \leftrightarrow \psi \}$ . Then  $\mathfrak{F}^\rho / \equiv \in \mathbf{Lf}_\omega$  by Theorem 66(i) and  $\mathbf{Rng}(\rho) \subseteq \omega$ . Assume  $\not\vdash^\rho \varphi$ . Then  $\mathfrak{F}^\rho / \equiv \not\models \varphi = 1$ , hence  $\mathbf{Csf}_\omega \not\models \varphi = 1$  by  $\mathbf{Lf}_\omega \subseteq \mathbf{HSPCsf}_\omega$ , i.e.  $\not\models \varphi$ . ■

We note that  $\vdash^\rho$  is the *usual* inference system of first-order logic (up to unessential variations in picking the axioms and in notation). The following is a purely logical corollary of the non-finitizability of  $\mathbf{RCA}_\omega$ .

**COROLLARY.** The usual inference system  $\vdash^\rho$  of first-order logic ( $\mathcal{L}_{FOL}^\rho$ ) is *not complete* w.r.t. the *formula schemas* of that logic. That is, there are valid formula schemas  $\Phi$  of  $\mathcal{L}_{FOL}^\rho$  such that  $\not\vdash^\rho \Phi$ .

**Outline of proof.** We want to prove that there are valid schemes<sup>114</sup>  $\Phi$  of  $\mathcal{L}_{FOL}^\rho$  such that although  $\models \Phi$ , we have  $\not\vdash^\rho \Phi$ . This is so because there is no

<sup>114</sup>If  $\Phi$  is a formula schema of  $\mathcal{L}$  (cf. Definition 33), then by  $\mathcal{L}$ -derivability  $\vdash_{\mathcal{L}}$  of  $\Phi$  we understand the natural extension of Definition 33 to a mixed language consisting of both  $\mathcal{L}$ -formulas and schemas. I.e., in a derivation  $\langle \Phi_1, \dots, \Phi_n \rangle$  of  $\Phi$ ,  $\Phi_i$  is built up from atomic formulas  $p_j \in P$  of  $\mathcal{L}$  and formula-variables  $\phi_j \in FV$  (using the connectives  $Cn$  of  $\mathcal{L}$ ).

difference between the schema languages of  $\mathcal{L}_{FOL}^\rho$  and  $\mathbf{L}_\omega$ , and also the valid schemes of  $\mathbf{L}_{FOL}^{ranked}$  and  $\mathbf{L}_\omega$  coincide by Corollary 46, because  $\mathbf{Eq}(\mathbf{L}_{FOL}^\rho) = \mathbf{Eq}(\mathbf{RCA}_\omega)$ , and the  $\vdash^\rho$ -provable and  $\vdash_\omega$ -provable schemas coincide.<sup>115</sup>

How is it possible that there is an  $\vdash^\rho$ -unprovable valid formula-scheme  $\Phi$ ? This means that though each instance of  $\Phi$  in  $\mathcal{L}_{FOL}^\rho$  is  $\vdash^\rho$ -provable (because of Theorem 71), these  $\vdash^\rho$ -proofs vary from instance to instance. We cannot give a ‘uniform’  $\vdash^\rho$ -proof for these instances, in spite of there being a uniform ‘cause’  $\Phi$  of their validity. ■

REMARK . The above corollary can be strengthened in the following direction. Let  $\vdash$  be defined by some “reasonable” generalization of finite schema for  $\mathcal{L}_{FOL}^\rho$ . Then  $\vdash$  cannot be sound and complete for all formula schemas of  $\mathcal{L}_{FOL}^\rho$ . Here we do not define what we mean by a reasonable generalization of finite schema, but it can be done by analyzing the usual axiomatizations of  $L_{\omega\omega}$  and analyzing the proof of the above corollary. Of course, what we have in mind admits the usual axiomatizations of  $L_{\omega\omega}$  as special cases.

Theorem 28(i) stating  $\text{Dc}_\omega \subseteq \text{RCA}_\omega$  can be used to overcome schema-incompleteness of  $\vdash^\rho$ . Using this theorem, one can obtain enriched inference systems  $\vdash^{\rho+}$  by adding brand new variables  $w_i$  ( $i < \omega$ ) to the language and new axioms postulating the effects of the fact that  $w_i$  does not occur in the *old* formulas. Roughly, these axioms say that

$$\begin{aligned} R \rightarrow \forall v_i R & \quad \text{if } \rho(R) \leq i < \omega \quad \text{and} \\ \phi_j \rightarrow \forall w_i \phi_j & \quad \text{if } i < \omega, \phi_j \text{ is a formula-variable.} \end{aligned}$$

These inference systems are strongly complete for the formula-schemas of  $\mathbf{L}_{FOL}^{ranked}$ . These completeness theorems (based on the Dc-representation result) are proved in [Andréka, Gergely and Németi, 1977; Henkin, Monk and Tarski, 1985].

The reason for  $\mathbf{L}_{FOL}^{ranked}$  not being substitutional is that the atomic formulas cannot take the meanings of any formula, because an atomic formula has a fixed finite rank, while formulas can have meanings of arbitrarily large finite ranks. This will be repaired in our next example.

---

<sup>115</sup>This is not quite trivial, but can be proved with CA-theoretic methods, e.g. one can use [Henkin, Monk and Tarski, 1971, 2.5.26].

### 9. First-order logic, rank-free<sup>116</sup> (or type-less) version, $\mathbf{L}_{FOL}$ .

The set of connectives is as in the previous case. Let  $\mathcal{R}$  be a set (of relation symbols). Then the set of atomic formulas of  $\mathcal{L}_{FOL}^{\mathcal{R}}$  is  $\mathcal{R}$ , as before.

The models will be different (as the information  $\rho : \mathcal{R} \rightarrow \omega$  is missing): We only know that  $R$  denotes a finitary relation, we do not know what its arity is. The actual arity will be given by the model. I.e., the models are  $\mathfrak{M} = \langle M, R^{\mathfrak{M}} \rangle_{R \in \mathcal{R}}$  where  $R^{\mathfrak{M}}$  is an arbitrary finitary relation on  $M$  for all  $R \in \mathcal{R}$ ,

$$M^{\mathcal{R}} = \{ \langle M, R^{\mathfrak{M}} \rangle_{R \in \mathcal{R}} : (\forall R \in \mathcal{R})(\exists n \in \omega) R \subseteq {}^n M \}.$$

Validity and the meaning function are the same as in the previous case, the only difference is that

$$mng_{\mathfrak{M}}(R) = \{ h \in {}^\omega M : h \upharpoonright n \in R^{\mathfrak{M}} \text{ for some } n \}.$$

Let  $\mathbf{L}_{FOL}$  denote the system of these logics. Now this general logic is structural, since

$$Mng^{\mathcal{R}} = Hom(\mathfrak{F}^{\mathcal{R}}, Csf_{\omega}).$$

Thus  $\mathbf{L}_{FOL}$  is an algebraizable general logic with

$$\mathbf{Alg}_m(\mathbf{L}_{FOL}) = Csf_{\omega} \quad \text{and} \quad \mathbf{Alg}(\mathbf{L}_{FOL}) = \mathbf{SPCsf}_{\omega}.$$

Thus Theorem 54 says that  $\mathbf{L}_{FOL}$  is not compact because, by Theorem 18(iii),  $\mathbf{SPCsf}_{\omega} \neq \mathbf{UpSPCsf}_{\omega}$ . Or vice versa, one can prove the algebraic theorem  $\mathbf{SPCsf}_{\omega} \neq \mathbf{UpSPCsf}_{\omega}$  by showing that  $\mathbf{L}_{FOL}$  is not compact, as follows:  $\mathbf{L}_{FOL}$  is not compact because the set  $\Sigma = \{ \neg(R \leftrightarrow \exists v_i R) : i < \omega \}$  of formulas, where  $R$  is any relation symbol, is not satisfiable while all of its finite subsets are. Thus Theorem 54 says that  $\mathbf{SPCsf}_{\omega} \neq \mathbf{UpSPCsf}_{\omega}$  because  $\mathbf{Alg}(\mathbf{L}_{FOL}) = \mathbf{SPCsf}_{\omega}$ .

Theorem 51 admits a generalization to logics like  $\mathbf{L}_{FOL}$  above. Then we obtain the following corollary of this generalized result and of Theorem 19 (saying that  $\mathbf{RCA}_n$  is not finite schema axiomatizable).

**COROLLARY 72.** *Assume that  $\langle Ax, Ru \rangle$  defines a strongly sound and weakly complete inference system  $\vdash$  for  $\mathbf{L}_{FOL}$ . Then  $\langle Ax, Ru \rangle$  must involve an infinite set of formula-variables. I.e.,  $\mathbf{L}_{FOL}$  is not finite-schema axiomatizable. The same applies for  $\mathbf{L}_{\omega}$  in place of  $\mathbf{L}_{FOL}$ . ■*

Improved versions of this negative result are in [Andréka, 1997] where it is proved that  $\langle Ax, Ru \rangle$  has to be extremely complex, too, besides involving infinitely many formula-variables. Positive results kind of side-stepping Corollary 72 above are in [Sain, 1995; Sain, 2000; Sain and Gyuris, 1994].

<sup>116</sup>Rank-free first-order logic was introduced in [Henkin and Tarski, 1961], and elaborated in more detail in [Andréka, 1973], [Andréka, Gergely and Néméti, 1977, sec. IV]. See also [Henkin, Monk and Tarski, 1985, section 4.3]. A nice proof system for this logic is given in [Simon, 1991].

These present expansions of  $\mathbf{L}_{FOL}$  with further logical connectives, such that the new  $\mathbf{L}_{FOL}^+$  becomes finite schema axiomatizable.

At this point the reader might have the impression that Corollary 72 seems to contradict Gödel's completeness theorem. However, Gödel's theorem holds for the ranked version  $\mathbf{L}_{FOL}^{ranked}$  of first order logic but not for  $\mathbf{L}_{FOL}$ . The essential difference between these two logics is that  $\mathbf{L}_{FOL}^{ranked}$  is *not* structural (substitutional). No structural version of first order logic is known for which Gödel's completeness theorem would hold. More precisely, the only such versions are the logics presented in [Sain and Gyuris, 1994] etc. cited above. The presently discussed issue is highly relevant to the propositional modal versions of first order logic, cf. e.g. [van Benthem, 1997; van Benthem, 1996; van Benthem and ter Meulen, 1997; Venema, 1995a; Marx and Venema, 1997].

Now we briefly compare our three versions of FOL: non-ranked  $\mathbf{L}_\omega$ , the ranked version  $\mathbf{L}_{FOL}^{ranked}$ , and the rank-free one,  $\mathbf{L}_{FOL}$ . The same formula-schemes are valid in them, and they have the same admissible rules by Theorem 45, because the same quasi-equations are true in their algebraized forms by  $\mathbf{SPUpCsf}_\omega = \mathbf{RCA}_\omega = \mathbf{Alg}(\mathbf{L}_\omega)$ . Also, this set of admissible formula-schemes is recursively enumerable, and the validity problem in these logics is not decidable, by Theorems 18, 45 and Corollary 47.

As a contrast, here we will give a logic which has a decidable validity problem and at the same time the set of valid formula-schemes is not even recursively enumerable.

### 10. Equality logic, monadic logic.

First we treat equality logic  $\mathcal{L}^e$ . This is the same as first-order logic with  $\omega$  variables and with no atomic formulas, i.e.  $\mathcal{L}^e \stackrel{\text{def}}{=} \mathcal{L}_\omega^0$ . Therefore, this is not a general logic. Notice that the set of formulas is non-empty because  $v_i = v_j$  is a zero-ary logical connective. A model is just a set  $M$  and the meaning-algebra  $\mathfrak{Mng}(M)$  of this model is the subalgebra of  $\langle \mathfrak{P}(\omega M), C_i, D_{ij} \rangle_{i,j < \omega}$  generated by  $\{D_{ij} : i, j < \omega\}$ . These are called the *minimal cylindric set algebras*, and their class is denoted by  $\text{setMn}_\omega$ , while  $\text{Mn}_\omega \stackrel{\text{def}}{=} \mathbf{IsetMn}_\omega$ .

$\mathcal{L}^e$  is an algebraizable semantic logic with  $\mathbf{Alg}_m(\mathcal{L}^e) = \text{setMn}_\omega$  and  $\mathbf{Alg}(\mathcal{L}^e) = \mathbf{SPMn}_\omega$ . ( $\mathcal{L}^e$  is substitutional because its set of atomic formulas is empty.)

It is well known that the validity problem of  $\mathcal{L}^e$  is decidable, it has the finite model property, and it admits an elimination-of-quantifiers theorem. (See e.g. [Monk, 1964a].)

However, the set of valid formula-schemes of  $\mathcal{L}^e$  is not even recursively enumerable. This is so by Corollary 46, because  $\mathbf{Eq}(\text{Mn}_\omega)$  is not recursively enumerable.<sup>117</sup>

<sup>117</sup>This was proved by M. Rubin, and independently by I. Németi, see [Németi, 1987].



More generally, consider now ranked first-order logics  $\mathcal{L}_\omega^\rho$ . Ranked first-order logic  $\mathcal{L}_n^\rho$  with  $n$  variables,  $n < \omega$  can be defined analogously for  $\rho : \mathcal{R} \rightarrow n$ . Let  $n \leq \omega$ . If every relation symbol is unary, i.e. if  $\text{Rng}\rho \subseteq \{1\}$ , then  $\mathcal{L}_n^\rho$  is called a *monadic* logic. Let  $\mathcal{L}_n^{m\mathcal{R}}$  denote monadic logic with relation symbols  $\mathcal{R}$ , i.e.  $\mathcal{L}_n^{m\mathcal{R}} \stackrel{\text{def}}{=} \mathcal{L}_n^\rho$  where  $\rho = \mathcal{R} \times \{1\}$ .

$\mathcal{L}_n^{m\mathcal{R}}$  is a compositional logic with filter-property. It is not substitutional.

It is known that the validity problem of monadic logics is also decidable, they have the finite model property, and they admit elimination of quantifiers. (See also [Monk, 1964a].)

Let  $n > 2$ . If  $n$  is infinite, then the valid schemes of  $\mathcal{L}_n^{m\mathcal{R}}$  are not recursively enumerable. If  $\rho$  is not monadic, then the valid schemes of  $\mathcal{L}_n^\rho$  are recursively enumerable (and the validities become undecidable). (If  $n \leq 2$ , then the set of valid schemes of  $\mathcal{L}_n^\rho$  is decidable.) These are proved in [Németi, 1987] by showing that the equational theories of the corresponding classes of algebras are not recursively enumerable (and using Theorem 45). The logical implications and the reasons for this behaviour are also explained carefully in [Németi, 1987].

### 11. Infinitary version $\mathbf{L}_{\infty\omega}^n$ of the finite variable fragments $\mathbf{L}_n$ .

Let  $\kappa$  be an infinite cardinal.  $\mathbf{L}_{\kappa\omega}^n$  is obtained from  $\mathbf{L}_n$  by adding  $\kappa$ -ary conjunction to the logical connectives. More formally, let  $F_n$  be the set of formulas of  $\mathcal{L}_n^{\mathcal{R}} = \langle F_n, \dots \rangle$ . Let  $F_\kappa^n$  be the smallest set satisfying (i)–(iii) below.

$$(i) \quad F_\kappa^n \supseteq F_n,$$

$$(ii) \quad F_\kappa^n \text{ is closed under the connectives of } \mathcal{L}_n,$$

$$(iii) \quad |H| < \kappa \Rightarrow (\wedge H) \in F_\kappa^n, \quad \text{for any } H \subseteq F_\kappa^n.$$

The models of  $\mathcal{L}_{\kappa\omega}^{n\mathcal{R}}$  are the same as those of  $\mathcal{L}_n^{\mathcal{R}}$ , and  $mng_\kappa, F_\kappa$  are the obvious generalizations of the definition given for  $\mathcal{L}_n^{\mathcal{R}}$ . Then

$$\mathcal{L}_{\kappa\omega}^{n\mathcal{R}} \stackrel{\text{def}}{=} \langle F_\kappa^n, M_n, mng_\kappa, F_\kappa \rangle \quad \text{and} \quad \mathbf{L}_{\kappa\omega}^n \stackrel{\text{def}}{=} \langle \mathcal{L}_{\kappa\omega}^{n\mathcal{R}} : \mathcal{R} \text{ is a set} \rangle.$$

$\mathbf{L}_{\infty\omega}^n$  is obtained from  $\mathbf{L}_{\kappa\omega}^n$  by removing all conditions of the form ' $\dots < \kappa$ '. That is,  $F_{\infty\omega}^n := \bigcup \{F_{\kappa\omega}^n : \kappa \text{ is a cardinal}\}$ , etc.  $\mathbf{L}_{\infty\omega}^n$  and  $\mathbf{L}_{\kappa\omega}^n$  (with  $\kappa > \omega$ ) are *not* logics in the sense of Definition 30 because they involve infinitely long strings of symbols. All the same,  $\mathbf{L}_{\infty\omega}^n$  is an interesting mathematical structure whose study is motivated by studying logics. Most properties of logics make sense for the 'pseudo-logic'  $\mathbf{L}_{\infty\omega}^n$ , too. Studying mathematical structures like  $\mathbf{L}_{\infty\omega}^n, \mathbf{L}_{\kappa\omega}^n$  seems to be useful for obtaining a better understanding of logics (in the sense of Definition 30).

Most of the results obtained for  $\mathbf{L}_n$  via the methods of algebraic logic can be pushed through for  $\mathbf{L}_{\infty\omega}^n$  by the same kinds of algebraic methods. In particular, by stretching the algebraic methods which lead to Theorem 63, one can obtain the following. The notions of formula schema, inference system, axiom schema, rule schema can be generalized to  $\mathbf{L}_{\kappa\omega}^n$  the natural way. Herein we do not go into the details of this.

**COROLLARY 73.** *Assume  $\vdash$  is a strongly sound and weakly complete provability relation for  $\mathbf{L}_{\infty\omega}^n$  or for  $\mathbf{L}_{\kappa\omega}^n$  ( $\kappa \geq \omega$ ). Then  $\vdash$  is not definable by a Hilbert-style inference system. Moreover, any schema  $\langle Ax, Ru \rangle$  axiomatizing  $\vdash$  must involve infinitely many formula variables (cf. Definition 33 for  $\langle Ax, Ru \rangle$  axiomatizing  $\vdash$ .)*

The next table summarizes the algebraic counterparts of some of the distinguished logics.

Table 1.

Logic $\mathcal{L}$	$\mathbf{Alg}(\mathcal{L})$	$\mathbf{Alg}_m(\mathcal{L})$	substitutional, compact
$\mathbf{L}_s$ sentential logic	BA	$\{2\}$	+ +
$\mathbf{S5}$ modal logic	$\mathbf{RCA}_1$	$\mathbf{Cs}_1$	+ +
$\mathbf{L}_{REL}$ arrow logic	BRA	setBRA	+ +
$\mathbf{L}'_n$ $n$ -var. FOL with substituted atomic fmlas	$\mathbf{Rd}_{ca}$ $\mathbf{RPEA}_n$	$\mathbf{Rd}_{ca}$ $\mathbf{Pse}_n$	- +
$\mathbf{L}_n$ structural $n$ -var. FOL	$\mathbf{RCA}_n$	$\mathbf{Cs}_n$	+ +
$\mathbf{L}_\omega$ finitary FOL of $\omega$ -ary rels	$\mathbf{RCA}_\omega$	$\mathbf{Cs}_\omega$	+ +
$\mathbf{L}_{FOL}^{ranked}$ ranked FOL	$\mathbf{Lf}_\omega$	$\mathbf{Csf}_\omega$	- +
$\mathbf{L}_{FOL}$ rank-free FOL	$\mathbf{SPCsf}_\omega$	$\mathbf{Csf}_\omega$	+ -

## ACKNOWLEDGEMENTS

Research supported by the Hungarian Foundation for Basic Research Grants T16448, T23234.

*Mathematical Institute of Hungarian Academy of Sciences, Budapest.*

## BIBLIOGRAPHY

- [Aczel, 1994] P. Aczel. Schematic consequence. In [Gabbay, 1994, pp. 261–272].
- [Adámek, Herrlich and Strecker, 1990] J. Adámek, H. Herrlich and G. Strecker. *Abstract and concrete categories, or the joy of cats*. John Wiley and Sons, 1990.
- [Allén, 1989] S. Allén, ed. *Possible worlds in humanities, arts and sciences* (Proc. of Nobel Symposium 65) W. de Gruyter, Berlin, 1989.
- [Amer, 1993] M. Amer. Cylindric algebras of sentences (abstract). *Journal of Symbolic Logic*, **58,2**, 743, 1993.
- [Andréka, 1973] H. Andréka. Algebraic investigation of first-order logic. (In Hungarian) *PhD thesis, Eötvös Loránd University, Budapest*, ix+162 pp., 1973.
- [Andréka, 1977] H. Andréka. Universal algebraic logic. (In Hungarian) *Dissertation with the Hung. Academy of Sci., Budapest*, 199 pp., 1977.
- [Andréka, 1994] H. Andréka. Weakly representable but not representable relation algebras. *Algebra Universalis*, **32**, 31–43, 1994.
- [Andréka, 1997] H. Andréka. Complexity of the equations valid in algebras of relations. Part I: Strong non-finitizability; Part II: Finite axiomatizations. *Annals of Pure and Applied Logic*, **89**, 149–209 and 211–229, 1997.
- [Andréka *et al.*, to appear] H. Andréka, W. J. Blok, I. Németi, D. L. Pigozzi and I. A. Sain. Abstract algebraic logic. In preparation.
- [Andréka, van Benthem and Németi, 1996] H. Andréka, J. A. F. K. van Benthem and I. Németi. Finite model property of the bounded fragment of first-order logic and cylindric-relativized set algebras. Manuscript, 1996.
- [Andréka, van Benthem and Németi, 1997] H. Andréka, J. A. F. K. van Benthem and I. Németi. Modal languages and bounded fragments of predicate logic. *Journal of Philosophical Logic*, **27**, 217–274, 1998.
- [Andréka, Comer and Németi, 1983] H. Andréka, S. D. Comer and I. Németi. Epimorphisms in cylindric algebras. Manuscript, 1983.
- [Andréka, Gergely and Németi, 1973] H. Andréka, T. Gergely and I. Németi. Purely algebraical construction of first order logics. *Publications of Central Res. Inst. Physics Budapest*, **KFKI-73-71**, 46pp., 1973.
- [Andréka, Gergely and Németi, 1977] H. Andréka, T. Gergely and I. Németi. On universal algebraic construction of logics. *Studia Logica*, **36**, 9–47, 1977.
- [Andréka *et al.*, 1998] H. Andréka, S. Givant, Sz. Mikulás, I. Németi and A. Simon. Notions of density that imply representability in algebraic logic. *Annals of Pure and Applied Logic*, **91**, 93–190, 1998.
- [Andréka, Givant and Németi, 1994] H. Andréka, S. Givant and I. Németi. The lattice of varieties of representable relation algebras. *Journal of Symbolic Logic*, **59,2**, 631–661, 1994.
- [Andréka, Givant and Németi, 1995] H. Andréka, S. Givant and I. Németi. Perfect extensions and derived algebras. *Journal of Symbolic Logic*, **60,3**, 775–796, 1995.
- [Andréka, Givant and Németi, 1997] H. Andréka, S. Givant and I. Németi. Decision problems for equational theories of relation algebras. *Memoirs of Amer. Math. Soc.*, **126,604**, xiv+126pp., 1997.
- [Andréka, Goldblatt and Németi, 1998] H. Andréka, R. Goldblatt and I. Németi. Relativized quantification: some canonical varieties of sequence-set algebras. *Journal of Symbolic Logic*, **63,1**, 163–184, 1998.

- [Andréka, Hodkinson and Németi, 1999] H. Andréka, I. Hodkinson and I. Németi. Finite algebras of relations are representable on finite sets. *Journal of Symbolic Logic*, **64**,1, 243–267, 1999.
- [Andréka, Jónsson and Németi, 1991] H. Andréka, B. Jónsson and I. Németi. Free algebras in discriminator varieties. *Algebra Universalis*, **28**, 401–447, 1991. Abstracted in [Bergman, Maddux and Pigozzi, 1990], 1–14.
- [Andréka *et al.*, 1993] H. Andréka, Á. Kurucz, I. Németi and I. A. Sain. Methodology of applying algebraic logic to logic. *Math. Inst. Budapest, Preprint, 1993*. Shortened version ('Applying Algebraic Logic to Logic') in: Algebraic Methodology and Software Technology (AMAST'93), eds. M. Nivat, C. Rattray, T. Rus and G. Scollo, in Series "Workshops in Computing", Springer-Verlag, 1994, 7–28. Also: Material for Summer School on Algebraic Logic and the Methodology of Applying it, Budapest, 1994. <http://circle.math-inst.hu/pub/algebraic-logic/meth.dvi>
- [Andréka *et al.*, 1996] H. Andréka, Á. Kurucz, I. Németi, I. A. Sain and A. Simon. Causes and remedies for undecidability in arrow logics and in multi-modal logics. In: Arrow logic and Multi-Modal Logic, M. Marx, L. Pólos, M. Masuch eds, CSLI Publications and FOLLI, Stanford, pp. 63–99, 1996.
- [Andréka, Maddux and Németi, 1991] H. Andréka, R. D. Maddux and I. Németi. Splitting in relation algebras. *Proc. Amer. Math. Soc.*, **111**,4, 1085–1093, 1991.
- [Andréka and Mikulás, 1994] H. Andréka and Sz. Mikulás. Lambek calculus and its relational semantics: completeness and incompleteness. *Journal of Logic, Language and Information (JoLLI)*, **3**, 1–37, 1994.
- [Andréka, Monk and Németi, 1991] H. Andréka, J. D. Monk and I. Németi, eds. *Algebraic Logic. (Proc. Conf. Budapest, 1988)* Colloq. Math. Soc. J. Bolyai **54**, North-Holland, Amsterdam v+746 pp. 1991.
- [Andréka and Németi, 1975] H. Andréka and I. Németi. A simple purely algebraic proof of the completeness of some first order logics. *Algebra Universalis*, Vol **5**, 8–15, 1975.
- [Andréka and Németi, 1990] H. Andréka and I. Németi. Constructing unusual relation set algebras. Manuscript, Mathematical Institute, Budapest. August, 1990.
- [Andréka and Németi, 1993] H. Andréka and I. Németi. Axiomatization of identity-free equations valid in relation algebras. *Algebra Universalis*, **35**, 256–264, 1996.
- [Andréka *et al.*, 1995] H. Andréka, I. Németi, I. A. Sain and Á. Kurucz. General algebraic logic including algebraic model theory: an overview. In: *Logic Colloquium'92 (Proc. 1992 Summer Meeting of Assoc. Symb. Log.)*, eds. L. Csirmaz, D. M. Gabbay and M. de Rijke, pp. 1–60, CSLI Publications, Stanford, 1995.
- [Andréka, Németi and Thompson, 1990] H. Andréka, I. Németi and R. J. Thompson. Weak cylindrical set algebras and weak subdirect indecomposability. *Journal of Symbolic Logic*, Vol **55**, No 2, 577–588, 1990.
- [Andréka and Sain, 1981] H. Andréka and I. A. Sain. Connections between initial algebra semantics of CF languages and algebraic logic. In *Mathematical Logic in Computer Science* (Proc. Conf. Salgótarján Hungary 1978) (Editors: Dömölki, B. and Gergely, T.) Colloq. Math. Soc. J. Bolyai, **26**, North-Holland, Amsterdam, 25–83, 1981.
- [Andréka and Thompson, 1988] H. Andréka and R. J. Thompson. A Stone-type representation theorem for algebras of relations of higher rank. *Trans. Amer. Math. Soc.*, **309**,2, 671–682, 1988.
- [Anellis and Houser, 1991] I. H. Anellis and N. Houser. The nineteenth century roots of universal algebra and algebraic logic: A critical-bibliographical guide for the contemporary logician. In [Andréka, Monk and Németi, 1991] 1–36.
- [Barwise and Feferman, 1985] J. Barwise and S. Feferman, eds. *Model-Theoretic Logics*. Springer-Verlag, Berlin, 1985.
- [Bednarek and Ulam, 1977] A. R. Bednarek and S. M. Ulam. Some remarks on relational composition in computational theory and practice. *Fundamentals of Computation Theory '77*, Lecture Notes in Computer Science Vol **56**, Springer-Verlag, Berlin, 33–38, 1977.
- [Bergman, Maddux and Pigozzi, 1990] C. H. Bergman, R. D. Maddux and D. L. Pigozzi, eds. Algebraic logic and universal algebra in computer science. *Lecture Notes in Computer Science, Springer Verlag*, Vol **425**, xi+292 p. 1990.

- [Biró, 1992] B. Biró. Non-finite-axiomatizability results in algebraic logic. *Journal of Symbolic Logic*, Vol **57**, 3, 832–843, 1992.
- [Biró and Shelah, 1988] B. Biró and S. Shelah. Isomorphic but not lower-base-isomorphic cylindric set algebras. *Journal of Symbolic Logic*, **53**, 3, 846–853, 1988.
- [Blok, 1980] W. J. Blok. The lattice of modal logics: An algebraic investigation. *Journal of Symbolic Logic*, Vol **45**, 221–238, 1980.
- [Blok and Pigozzi, 1986] W. J. Blok and D. L. Pigozzi. Protoalgebraizable logics. *Studia Logica*, **45**, 337–369, 1986.
- [Blok and Pigozzi, 1989] W. J. Blok and D. L. Pigozzi. Algebraizable logics. *Memoirs Amer. Math. Soc.*, Vol **77**, **396**, vi+78 pp. 1989.
- [Blok and Pigozzi, 1989a] W. J. Blok and D. L. Pigozzi. On the structure of varieties with equationally definable principal congruences, Part III. Preprint, 1989.
- [Blok and Pigozzi, 1989b] W. J. Blok and D. L. Pigozzi. On the structure of varieties with equationally definable principal congruences, Part IV. Preprint, 1989.
- [Blok and Pigozzi, 1989c] W. J. Blok and D. L. Pigozzi. The deduction theorem in algebraic logic. Preprint, 1989.
- [Blok and Pigozzi, 1991] W. J. Blok and D. L. Pigozzi. Local Deduction Theorems in Algebraic Logic. In *[Andréka, Monk and Németi, 1991]*, pp. 75–109.
- [Blok and Pigozzi, 1991a] W. J. Blok and D. L. Pigozzi. Introduction to *Studia Logica* Vol 50, No 3/4 (1991).
- [Blok and Pigozzi, 1997] W. J. Blok and D. L. Pigozzi. Abstract Algebraic Logic and the Deduction Theorem. Submitted.
- [Bloom and Brown, 1973] S. L. Bloom and D. J. Brown. Classical abstract logics. *Dissertationes Mathematicae (Rozprawy Matematyczne)*, Vol **CII**, 43–56, 1973.
- [Brown and Suszko, 1973] D. J. Brown and R. Suszko. Abstract logics. *Dissertationes Mathematicae (Rozprawy Matematyczne)*, Vol **CII**, 9–41, 1973.
- [Burris and Sankappanavar, 1981] S. Burris and H. P. Sankappanavar. A course in universal algebra. *Graduate Texts in Mathematics*, Springer-Verlag, New York, 1981.
- [Cohn, 1965] P. M. Cohn. *Universal Algebra*. Harper and Row, New York, 1965.
- [Comer, 1969] S. D. Comer. Classes without the amalgamation property. *Pacific J. Math.*, **28**, 309–318, 1969.
- [Comer, 1972] S. D. Comer. A sheaf-theoretic duality for cylindric algebras. *Trans. Amer. Math. Soc.*, Vol **169**, 75–87, 1972.
- [Craig, 1965] W. Craig. Boolean notions extended to higher dimensions. In: *The theory of models* (eds: J.W.Addison, L.Henkin, A.Tarski), North-Holland, 55–69, 1965.
- [Craig, 1974] W. Craig. *Logic in algebraic form*. North-Holland, Amsterdam, viii+204pp, 1974.
- [Craig, 1974a] W. Craig, W., Unification and abstraction in algebraic logic. *Studies in algebraic logic, Math. Assoc. Amer.*, pp. 6–57, 1974.
- [Crvenkovic and Madarász, 1992] S. Crvenkovic and R. Madarász. Relation Algebras. (In Serbian), Beograd, 1992.
- [Czelakowski, 1982] J. Czelakowski. Logical matrices and the amalgamation property. *Studia Logica*, **51**, 4, 329–342, 1982.
- [Czelakowski, 1993] J. Czelakowski. Logic, algebras, and consequence operators. Preprint, 79pp. 1992.
- [Czelakowski, 1997] J. Czelakowski. Protoalgebraic Logic. Kluwer, 1997 (to appear).
- [Czelakowski and Pigozzi, 1999] J. Czelakowski and D. L. Pigozzi. Amalgamation and interpolation in abstract algebraic logic. In: *Models, Algebras, and Proofs, X*. Caicedo and C. H. Montenegro eds, Lecture Notes in Pure and Applied Mathematics Vol.203, Marcel Decker, Inc., 187–265, 1999.
- [Daigneault, 1969] A. Daigneault. Lawvere's elementary theories and polyadic and cylindric algebras. *Fund. Math.*, Vol **66**, 307–328, 1969.
- [Daigneault and Monk, 1963] A. Daigneault and J. D. Monk. Representation theory for polyadic algebras. *Fund. Math.*, Vol **52**, 151–176, 1963.
- [De Morgan, 1964] A. De Morgan. On the syllogism, no. IV, and on the logic of relations. *Transactions of the Cambridge Philosophical Society*, Vol **10**, 331–358, 1964.
- [Dipert, 1984] R. Dipert. Peirce, Frege, the logic of relations, and Church's theorem. *History and Philosophy of Logic*, Vol **5**, 49–66, 1984.

- [Diskin, 1991] Z. B. Diskin. Polyadic algebras for non-classical logics. Parts I–IV. *Latv. Mat. Ezhegodnik Vols 35, 36*, 1991.
- [Diskin, 1996] Z. B. Diskin. Abstract Universal Algebraic Logic. Part I: A unified framework of structural hyperlogics for integrating the deductive and model-theoretical approaches. Part II: Algebraizable logics and algebraic semantics (Galois connections, compactness, constructivisability) In: Proceedings of the Latvian Academy of Sciences Vol 50, No 1 pp.13–21, 22–30, 1996.
- [Diskin, 1997] Z. B. Diskin. When is semantically defined logic algebraizable? *Acta Universitatis Latviensis*, **595**, 57–82, 1997.
- [Düntsch, 1993] I. Düntsch. A note on cylindric lattices. In: Algebraic Methods in Logic and in Computer Science, ed. C. Rauszer, Banach Center Publications Vol 28, Polish Academy of Sciences, Warszawa, pp.231–238, 1993.
- [Ebbinghaus and Flum, 1995] H.-D. Ebbinghaus and J. Flum. Finite Model Theory. Springer Verlag, Berlin, 1995.
- [Epstein, 1990] R. L. Epstein. The semantic foundations of logic. Vol 1: Propositional logics. Kluwer, Dordrecht, 1990.
- [Ferenczi, 1991] M. Ferenczi. Measures defined on free products of formula algebras and analogies with cylindric homomorphisms. In *[Andréka, Monk and Németi, 1991]* pp. 173–181.
- [Font and Jansana, 1994] J. M. Font and R. Jansana. On the sentential logics associated with strongly nice and semi-nice logics. *Bulletin of the IGPL*, **2**, 55–76, 1994.
- [Font and Jansana, 1997] J. M. Font and R. Jansana. A general algebraic semantics for sentential logics. Lecture Notes in Logic Vol. 7, Springer, Berlin, 1997.
- [Friedman, 1973] H. M. Friedman. Beth’s theorem in cardinality logics. *Israel Journal of Mathematics*, **14**, 205–212, 1973.
- [Gabbay, 1994] D. M. Gabbay, ed. What is a logical system? Oxford Science Publications, Clarendon Press, Oxford, 1994.
- [Gabbay, 1996] D. M. Gabbay. Labelled deductive systems, Part I. Oxford University Press, 1996.
- [Gabbay, 1998] D. M. Gabbay. *Fibring Logics*, Oxford University Press, 1998.
- [Gabbay, 1996b] D. M. Gabbay. Fibring semantics and the weaving of logics. Part I: modal and intuitionistic logics. *Journal of Symbolic Logic*, **61**, **4**, 1057–1120, 1996.
- [Givant, 1991] S. Givant. Tarski’s development of logic and mathematics based on the calculus of relations. In *[Andréka, Monk and Németi, 1991]* pp. 189–215.
- [Givant, 1994] S. Givant. The structure of relation algebras generated by relativizations. Contemporary Mathematics Vol 156, American Mathematical Society, Providence, xv+134pp. 1994.
- [Givant, 1999] S. Givant. Universal classes of simple relation algebras. *Journal of Symbolic Logic*, Vol **64**,**2**, 575–589, 1999.
- [Goldblatt, 1990] R. I. Goldblatt. Varieties of complex algebras. *Annals of Pure and Appl. Logic*, Vol **44**, 173–242, 1990.
- [Goldblatt, 1991] R. I. Goldblatt. On closure under canonical embedding algebras. In *[Andréka, Monk and Németi, 1991]* pp. 217–229.
- [Goldblatt, 1995] R. I. Goldblatt. Elementary generation and canonicity for varieties of Boolean algebras with operators. *Algebra Universalis*, **34**, 551–607, 1995.
- [Goldblatt, 2000] R. I. Goldblatt. Algebraic Polymodal Logic. *Logic Journal of the IGPL*, **8**,**4**, 391–448, 2000.
- [Goldblatt, 1999] L. Gordeev. *Combinatorial principles relevant to finite variable logic*. Preprint, 1999.
- [Grätzer, 1979] G. Grätzer. *Universal Algebra*. Second edition. Springer Verlag, New York, 1979.
- [Halmos, 1954] P. R. Halmos. Polyadic Boolean algebras. *Proc. Nat. Acad. Sci. U.S.A.*, Vol **40**, 296–301, 1954.
- [Halmos, 1960] P. R. Halmos. Polyadic algebras. *Summaries of talks presented at the Summer Institute for Symbolic Logic, Cornell University, (1957)*, second edition, Communications Research Division, Institute for Defense Analyses, pp. 252–255. 1960.

- [Halmos, 1962] P. R. Halmos. Algebraic logic. *Chelsea Publ. Co.*, New York, 271pp, 1962.
- [Henkin, 1955] L. Henkin. The representation theorem for cylindric algebras. In: Mathematical interpretation of formal systems, North-Holland, Amsterdam, pp. 85-97. 1955.
- [Henkin, 1970] L. Henkin. Extending Boolean operations. *Pac. J. Math.*, Vol **22**, 723-752, 1970.
- [Henkin and Monk, 1974] L. Henkin and J. D. Monk. Cylindric algebras and related structures. *Proc. Tarski Symp.*, *Amer. Math. Soc.*, Vol **25**, 105-121, 1974.
- [Henkin, Monk and Tarski, 1971] L. Henkin, J. D. Monk and A. Tarski. Cylindric Algebras. Part I, *North-Holland, Amsterdam* 1971.
- [Henkin, Monk and Tarski, 1985] L. Henkin, J. D. Monk and A. Tarski. Cylindric Algebras. Part II *North-Holland, Amsterdam*, 1985.
- [Henkin *et al.*, 1981] L. Henkin, J. D. Monk, A. Tarski, H. Andréka and I. Németi. Cylindric Set Algebras. *Lecture Notes in Mathematics*, Vol **883**, Springer-Verlag, Berlin, vi+323pp. 1981.
- [Henkin and Tarski, 1961] L. Henkin and A. Tarski. Cylindric algebras. In: Lattice theory, Proceedings of symposia in pure mathematics, vol.2, ed. R. P. Dilworth, American Mathematical Society, Providence, pp.83-113. 1961.
- [Hirsch and Hodkinson, 1997] R. D. Hirsch and I. Hodkinson. Step by step — building representations in algebraic logic. *Journal of Symbolic Logic*, **62**, 225-279, 1997.
- [Hirsch and Hodkinson, 1997a] R. D. Hirsch and I. Hodkinson. Relation algebras with  $n$ -dimensional bases. Preprint, Imperial College and University College, London, 1997.
- [Hirsch and Hodkinson, to appear] R. D. Hirsch and I. Hodkinson. Relation algebras from cylindric algebras, II. Submitted.
- [Hirsch, Hodkinson, and Maddux, to appear] R. D. Hirsch, I. Hodkinson, and R. D. Maddux. Relation algebra results of cylindric algebras nad an application to proof theory. Submitted.
- [Hodkinson, 1997] I. Hodkinson. Atom structures of cylindric algebras and relation algebras. *Annals of Pure and Applied Logic*, Vol **89,2-3**, 117-148, 1997.
- [Hoogland, 1996] E. Hoogland. Algebraic characterizations of two Beth definability properties. Master Thesis, University of Amsterdam, 1996.
- [Hoogland, 1997] E. Hoogland. Beth definability in almost algebraizable logics. In preparation.
- [Hoogland and Madarász, 1997] E. Hoogland and J. Madarász. Definability properties beyond algebraizable logics. In preparation.
- [Janssen, 1997] T. M. V. Janssen. Compositionality. In [van Benthem and ter Meulen, 1997, pp. 417-474].
- [Jánossy, Kurucz and Eiben, 1996] A. Jánossy, Á. Kurucz and Á. E. Eiben. Combining algebraizable logics. *Notre Dame Journal of Formal Logic*, **37,2**, 366-380, 1996.
- [Jipsen, 1992] P. Jipsen. Computer aided investigations of relation algebras. Ph.D. Dissertation, Vanderbilt University, Tennessee, 1992.
- [Johnson, 1969] J. S. Johnson. Nonfinitizability of classes of representable polyadic algebras. *Journal of Symbolic Logic*, Vol **34**, 344-352, 1969.
- [Johnson, 1973] J. S. Johnson. Axiom systems for logic with finitely many variables. *Journal of Symbolic Logic*, Vol **38**, 576-578, 1973.
- [Jónsson, 1982] B. Jónsson. Varieties of relation algebras. *Algebra Universalis*, Vol **15**, 273-298, 1982.
- [Jónsson, 1991] B. Jónsson. The theory of binary relations. In [Andréka, Monk and Németi, 1991], pp. 245-292.
- [Jónsson, 1995] B. Jónsson. The preservation theorem for canonical extensions of Boolean Algebras with Operators. In: Lattice Theory and its Applications, K. A. Baker and R. Wille eds., Heldermann Verlag, pp. 121-130. 1995.
- [Jónsson and Tarski, 1951] B. Jónsson and A. Tarski. Boolean algebras with operators. Parts I-II *Amer. J. Math.*, Vol **73, 74**, 891-939, 127-162, 1951, 1952.
- [Jónsson and Tsinakis, 1991] B. Jónsson and C. Tsinakis. Relation algebras as residuated Boolean algebras. *Algebra Universalis*, **30**, 469-478, 1993.

- [Kearnes, Sain and Simon, to appear] K. Kearnes, I. A. Sain and A. Simon. Beth properties of finite variable fragments of first-order logic. Manuscript.
- [Kiss *et al.*, 1983] E. W. Kiss, L. Márki, P. Pröhle and W. Tholen. Categorical algebraic properties. A compendium on amalgamation, congruence extension, epimorphisms, residual smallness and injectivity. *Studia Sci. Math. Hungar.*, **18**, 79–141, 1983.
- [Kurucz, 1997] Á. Kurucz. Decidability in algebraic logic. Ph.D. Dissertation, Mathematical Institute Budapest, 1997.
- [Kurucz *et al.*, 1995] Á. Kurucz, I. Németi, I. A. Sain and A. Simon. Decidable and undecidable modal logics with a binary modality. *Journal of Logic, Language and Information*, **4**, 191–206, 1995.
- [Kurucz *et al.*, 1993] Á. Kurucz, I. Németi, I. A. Sain and A. Simon. Undecidable varieties of semilattice-ordered semigroups, of Boolean algebras with operators, and logics extending Lambek calculus. *Bulletin of the IGPL*, **1**, 91–98, 1993.
- [Lyndon, 1956] R. C. Lyndon. The representation of relation algebras. II *Ann. of Math.*, series 2, Vol. **63**, 294–307, 1956.
- [Mac Lane, 1971] S. Mac Lane. Categories for the working mathematician. Springer Verlag, 1971.
- [Madarász, 1995] J. X. Madarász. The Craig interpolation theorem in multi-modal logics. *Bulletin of the Section of Logic*, **24,3**, 147–151, 1995.
- [Madarász, 1998] J. X. Madarász. Interpolation in algebraizable logics; semantics for non-normal multi-modal logic. *Journal of Applied Nonclassical Logic*, **8,1-2**, 67–105, 1998.
- [Madarász, 1997] J. X. Madarász. Craig interpolation in algebraizable logics; meaningful generalization of modal logics. *Journal of the IGPL*, to appear. Also Preprint of Math. Inst. Hungar. Acad. Sci.
- [Madarász, 1999] J. X. Madarász. Interpolation and amalgamation, pushing the limits. Part I and Part II. *Studia Logica*, **61,3**, 311–345, 1998; **62,1**, 1–19, 1999.
- [Madarász, 1997b] J. X. Madarász. A sequel to “Amalgamation, congruence extension, and interpolation properties in algebras. In preparation.
- [Madarász, Németi and Sági, 1997] J. X. Madarász, I. Németi and G. Sági. On the finitization problem of relation algebras (completeness problems for the finite variable fragments). Extended abstract. *Bulletin of the Section of Logic*, **26,3**, 139–143, 1997.
- [Maddux, 1978] R. D. Maddux. Topics in relation algebras. Ph.D. Thesis. University of California, Berkeley, 1978.
- [Maddux, 1978b] R. D. Maddux. Sufficient conditions for representability of relation algebras. *Algebra Universalis*, **8**, 162–172, 1978.
- [Maddux, 1980] R. D. Maddux. The equational theory of  $CA_3$  is undecidable. *Journal of Symbolic Logic*, **45**, 311–316, 1980.
- [Maddux, 1983] R. D. Maddux. A sequent calculus for relation algebras. *Annals of Pure and Applied Logic*, **25**, 73–101, 1983.
- [Maddux, 1989] R. D. Maddux. Nonfinite axiomatizability results for cylindric and relation algebras. *Journal of Symbolic Logic*, **54, 3**, 951–974, 1989.
- [Maddux, 1991] R. D. Maddux. The origin of relation algebras in the development and axiomatization of the calculus of relations. *Studia Logica*, **50**, 421–456, 1991.
- [Maddux, 1991a] R. D. Maddux. The neat embedding problem and the number of variables required in proofs. *Proc. Amer. Math. Soc.*, **112**, 195–202, 1991.
- [Makkai, 1987] M. Makkai. Stone duality for first order logic. *Advances in Mathematics (Academic Press, New York and London)* Vol **65**, No 2 1987.
- [Makkai and Paré, 1989] M. Makkai and R. Paré. Accessible categories: The foundation of categorical model theory. Contemporary Mathematics, Amer. Math. Soc., vol.104, Providence, Rhode Island, 1989.
- [Makkai and Reyes, 1977] M. Makkai and G. E. Reyes. First Order Categorical Logic. *Lecture Notes in Mathematics* Vol **611**, Springer-Verlag, Berlin, 1977.
- [Maksimova, 1977] L. L. Maksimova. Craig’s interpolation theorem and amalgamable varieties. *Soviet Math. Dokl., American Mathematical Society 1978*, Vol **18**, No 6, 1977.



- [Martí-Oliet and Meseguer, 1994] N. Martí-Oliet and J. Meseguer. General logics and logical frameworks. In [Gabbay, 1994, pp. 355–392].
- [Marx, 1995] M. Marx. Algebraic relativization and arrow logic. PhD Dissertation, University of Amsterdam, 1995.
- [Marx, Németi and Sain, 1996] M. Marx, I. Németi and I. A. Sain. On conjugated Boolean algebras with operators. Manuscript, 1996.
- [Marx, Pólos and Masuch, 1996] M. Marx, L. Pólos and M. Masuch. *Arrow Logic and Multi-Modal Logic*. Studies in Logic, Language and Information, CSLI Publication, Stanford, 1996.
- [Marx and Venema, 1997] M. Marx and Y. Venema. *Multi-dimensional Modal Logic*. Kluwer Academic Publishers, 1997.
- [Masuch and Pólos, 1992] M. Masuch and L. Pólos, eds. *Logic at Work*. (Proc. Conf. Amsterdam December 1992). See the Arrow Logic Day section.
- [McKenzie, 1966] R. N. McKenzie. The representation of relation algebras. Dissertation, University of Colorado, 1966.
- [McKenzie, McNulty and Taylor, 1987] R. N. McKenzie, G. F. McNulty and W. F. Taylor. Algebras, Lattices, Varieties, Vol. I. *The Wadsworth and Brooks/Cole Mathematics Series*, Monterey, California, 1987.
- [Mikulás, 1995] Sz. Mikulás. Taming Logics. PhD Dissertation, University of Amsterdam, 1995.
- [Mikulás, 1996] Sz. Mikulás. Gabbay-style calculi. In: H. Wansing (ed), *Proof Theory of Modal Logic*, Kluwer, pp. 243–252. 1996.
- [Monk, 1961] J. D. Monk. Studies in cylindric algebras. Ph.D. Thesis, University of California, Berkeley, 1961.
- [Monk, 1964] J. D. Monk. On representable relation algebras. *Michigan Math. J.*, Vol **11**, 1964.
- [Monk, 1964a] J. D. Monk. Singular cylindric and polyadic equality algebras. *Trans. Amer. Math. Soc.*, **112**, 185–205, 1964.
- [Monk, 1969] J. D. Monk. Nonfinitizability of classes of representable cylindric algebras. *Journal of Symbolic Logic*, **34**, 331–343, 1969.
- [Monk, 1970] J. D. Monk. On an algebra of sets of finite sequences. *Journal of Symbolic Logic*, **35**, 19–28, 1970.
- [Monk, 1971] J. D. Monk. Provability with finitely many variables. *Proc. Amer. Math. Soc.*, **27** 353–358, 1971.
- [Monk, 1974] J. D. Monk. Connections between combinatorial theory and algebraic logic. In: Studies in Math., vol. 9. *Math. Assoc. Amer.* pp. 58–91.
- [Monk, 1977] J. D. Monk. Some problems in algebraic logic. *Colloq. Inter. de Logic, CNRS*, **249**, 83–88, 1977.
- [Monk, 1991] J. D. Monk. Structure problems for cylindric algebras. In [Andréka, Monk and Németi, 1991], pp. 413–429.
- [Monk, 1993] J. D. Monk. Lectures on Cylindric Set Algebras. In *Algebraic Methods in Logic and in Computer Science*, ed: C. Rauszer, Banach Center Publications Vol 28, Polish Academy of Sciences, Warszawa, pp. 253–290. 1993.
- [Németi, 1982] I. Németi. Surjectiveness of epimorphisms is equivalent to Beth definability property in general algebraic logic. *Preprint, Mathematical Institute, Budapest. 1982.*
- [Németi, 1985] I. Németi. Logic with three variables has Gödel's incompleteness property — thus free cylindric algebras are not atomic. *Preprint No 49/85, Math. Inst. Hungar. Acad. Sci., Budapest* July, 1985.
- [Németi, 1985a] I. Németi. Exactly which varieties of cylindric algebras are decidable? Preprint No. 34/1985, Math. Institut Budapest, 1985.
- [Németi, 1986] I. Németi. Free algebras and decidability in algebraic logic. (In Hungarian) *Dissertation for D.Sc. with Hung. Academy of Sciences, Budapest*, xviii+169 pp. 1986.
- [Németi, 1987] I. Németi. On varieties of cylindric algebras with applications to logic. *Annals of Pure and Applied Logic*, **36**, 235–277, 1987.

- [Németi, 1988] I. Németi. Epimorphisms and definability in relation-, polyadic-, and related algebras. Invited lecture at the “Algebraic Logic in Computer Science” Conference, Ames, Iowa, USA, June 1988.
- [Németi, 1990] I. Németi. On cylindric algebraic model theory. In [Bergman, Maddux and Pigozzi, 1990], pp. 37–76 1990.
- [Németi, 1991] I. Németi. Algebraizations of Quantifier Logics, an introductory overview, 12th version. Math. Inst. Budapest, Preprint, No 13/1996. Electronic address:  
<http://circle.math-inst.hu/pub/algebraic-logic/survey.dvi>  
 An extended abstract of this appeared in *Studia Logica*, **50**, 485–569, 1991.
- [Németi, 1992] I. Németi. Decidability of weakened versions of first-order logic. In: Proc. Logic at Work, December, 1992, Amsterdam.
- [Németi, 1995] I. Németi. Decidable versions of first order logic and cylindric-relativized set algebras. In: Logic Colloquium’92 (Proc. Veszprém, Hungary 1992), eds: L. Csirmaz, D. M. Gabbay and M. de Rijke, Studies in Logic, Language and Computation, CSLI Publications, pp. 177–241. 1995.
- [Németi, 1996] I. Németi. Fine-structure analysis of first order logic. In: *Arrow Logic and Multi-Modal Logic*, M. Marx, L. Pólos, and M. Masuch eds, CSLI Publications, Stanford, California, 221–247, 1996.
- [Németi and Andréka, 1994] I. Németi and H. Andréka. General algebraic logic: a perspective on “what is logic”. In: *What is a Logical System*, ed: D. M. Gabbay, Clarendon Press, Oxford, pp. 393–444. 1994.
- [Németi and Sági, to appear] I. Németi and G. Sági. On the equational theory of representable polyadic equality algebras. *Journal of Symbolic Logic*, to appear.
- [Németi and Sági, 1997] I. Németi and G. Sági. Geometries and relation algebras. Mathematical Institute Budapest, Preprint, 1997.
- [Németi and Simon, 1997] I. Németi and A. Simon. Relation algebras from cylindric algebras and polyadic algebras. *Logic Journal of the IGPL*, **5,4**, 575–588, 1997.
- [Otto, 1997] M. Otto. Bounded variable logics and counting. (A study in Finite Models.) Springer Lecture Notes in Logic, Vol. 9. 1997.
- [Palasinska and Pigozzi, 1995] K. Palasinska and D. L. Pigozzi. Implication in abstract algebraic logic. Preprint, *Cracow University of Technology and Iowa State University*, December, 1995.
- [Partee, 1976] B. H. Partee. Montague Grammar. Academic Press, New York, 1976.
- [Pigozzi, 1972] D. L. Pigozzi. Amalgamation, Congruence-Extension, and Interpolation Properties in Algebra. *Algebra Universalis*, **1,3**, 269–349, 1972.
- [Pigozzi, 1991] D. L. Pigozzi. Fregean Algebraic Logic. In [Andréka, Monk and Németi, 1991], pp. 473–502.
- [Pigozzi and Salibra, 1993] D. L. Pigozzi and A. Salibra. Polyadic algebras over non-classical logics. In: Algebraic Methods in Logic and in Computer Science, Banach Center Publications, vol.28, Institute of Mathematics, Polish Academy of Sciences, Warszawa, pp. 51–66. 1993.
- [Plotkin, 1994] B. I. Plotkin. *Universal algebra, Algebraic logic, and Databases*. Nauka, Fizmatlit (1991) (In Russian.) English version: Kluwer Academic Publishers, 1994.
- [Pratt, 1990] V. R. Pratt. Dynamic algebras as a well behaved fragment of relation algebras. In [Bergman, Maddux and Pigozzi, 1990], pp. 77–110. 1990.
- [Pratt, 1991] V. R. Pratt. Dynamic Algebras. *Studia Logica*, **50**, 3–4, 571–605, 1991.
- [Pratt, 1992] V. R. Pratt. Origins of the calculus of binary relations. In: Proc. 7th Annual IEEE Symposium on Logic in Computer Science, Santa Cruz, CA, pp. 248–254. 1992.
- [Quine, 1936] W. V. O. Quine. Toward a calculus of concepts. *Journal of Symbolic Logic*, **1**, 2–25, 1936.
- [Resek and Thompson, 1991] D. Resek and R. J. Thompson. An equational characterization of relativized cylindric algebras. In [Andréka, Monk and Németi, 1991] pp. 519–538.
- [de Rijke and Venema, 1995] M. de Rijke and Y. Venema. Sahlqvist’s Theorem for Boolean Algebras with Operators with an application to cylindric algebras. *Studia Logica*, **54,1**, 61–79, 1995.

- [Rybakov, 1997] V. Rybakov. Admissibility of Logical Inference Rules, Elsevier, North-Holland, New-York, Amsterdam (Studies in Logic and the Foundations of Math. vol. 132), 617 pp. 1997.
- [Sain, 1979] I. A. Sain. There are general rules for specifying semantics: Observations on abstract model theory. *Computational Linguistics and Computer Languages (CL&CL)*, **13**, 195–250, 1979.
- [Sain, 1990] I. A. Sain. Beth's and Craig's properties via epimorphisms and amalgamation in algebraic logic. In [Bergman, Maddux and Pigozzi, 1990], 209–226, 1990.
- [Sain, 1995] I. A. Sain. On the problem of finitizing first order logic and its algebraic counterpart. (A survey of results and methods.) In: Logic Colloquium'92 (Proceedings of Logic Colloquium'92, eds.: L. Csirmaz, D. M. Gabbay and M. de Rijke), CSLI Publications, Stanford, pp. 243–292. 1995.
- [Sain, 2000] I. A. Sain. On the Search for a Finitizable Algebraization of First Order Logic. *Logic Journal of the IGPL*, **8,4**, 495–588, 2000.
- [Sain and Gyuris, 1994] I. A. Sain and V. Gyuris. Finite schema axiomatizable algebraization of first order logic. *Logic Journal of the IGPL*, **5,5**, 699–751, 1977.
- [Sain and Thompson, 1991] I. A. Sain and R. J. Thompson. Strictly finite schema axiomatization of quasi-polyadic algebras. In [Andréka, Monk and Németi, 1991], pp. 539–571.
- [Sayed Ahmed, 1997] T. M. Sayed Ahmed. Algebras of sentences of logic. Master Thesis, Cairo University, 1997.
- [Sági, 1995] G. Sági. Non-computability of the consequences of the axioms of the omega-dimensional polyadic algebras. *Preprint, Math. Inst., Budapest*, 1995.
- [Sági and Németi, 1997] G. Sági and I. Németi. Modal logic of substitutions, completeness results. Preprint, Mathematical Institute Budapest, 1997.
- [Serény, 1985] G. Serény. Compact cylindric set algebras. *Bulletin of the Section of Logic*, Warsaw-Lódz, 57–64, 1985.
- [Serény, 1997] G. Serény. Saturatedness in cylindric algebraic model theory. *Logic Journal of the IGPL*, **5,1**, 25–48, 1997.
- [Shelah, 1991] S. Shelah. On a problem in cylindric algebra. In [Andréka, Monk and Németi, 1991] 645–664, 1991.
- [Simon, 1991] A. Simon. Finite Schema Completeness for Typeless Logic and Representable Cylindric Algebras. *Algebraic Logic (Proc. Conf. Budapest (1988))* Colloq. Math. Soc. J. Bolyai Vol **54**, 665–670. North-Holland, Amsterdam, 1991.
- [Simon, 1993] A. Simon. What the Finitization Problem is Not. In *Algebraic Methods in Logic and in Computer Science*, (Proc. Banach Semester Fall (1991)), Banach Center Publications, Vol 28, pp.95–116. Warsaw, 1993.
- [Simon, 1996] A. Simon. A new proof of the representation theorem for Quasi-projective Relation Algebras. Manuscript, Mathematical Institute, Budapest, 1996.
- [Simon, 1997] A. Simon. Non-representable algebras of relations. Ph.D. Dissertation, Mathematical Institute, Budapest, 1997.
- [Tarski, 1951] A. Tarski. Remarks on the formalization of the predicate calculus. (abstract) *Bull. Amer. Math. Soc.*, **57**, 81–82, 1951.
- [Tarski, 1965] A. Tarski. A simplified formalization of predicate logic with identity. *Archiv für Math. Logik u. Grundl.*, **7** 61–79, 1965.
- [Tarski and Givant, 1987] A. Tarski and S. Givant. A formalization of set theory without variables. *AMS Colloquium publications, Providence, Rhode Island* Vol **41**, 1987.
- [van Benthem, 1989] J. A. F. K. van Benthem. Modal logic and relational algebra. Preprint, Institute for Language, Logic and Information, University of Amsterdam, 1989.
- [van Benthem, 1991] J. A. F. K. van Benthem. General dynamics. In: *Theoretical Linguistics*, Ph. A. Luelsdorff, ed. pp. 151–201. 1991.
- [van Benthem, 1991a] J. A. F. K. van Benthem. *Language in Action. (Categories, Lambdas and Dynamic Logic)*, Vol. 130 of *Studies in Logic*, x+350 pp. North-Holland, 1991.
- [van Benthem, 1994] J. A. F. K. van Benthem. A note on dynamic arrow logics. In *Logic and Information Flow*, J. van Eijck and A. Visser eds. pp. 15–29. MIT Press, Cambridge, MA, 1994.

- [van Benthem, 1996] J. A. F. K. van Benthem. Exploring logical dynamics. *Studies in Logic, Language and Information*, CSLI Publications, Stanford, 1996.
- [van Benthem, 1997] J. A. F. K. van Benthem. Modal foundations for predicate logics. *Logic Journal of the IGPL*, **5,2**, 259–286, 1997.
- [van Benthem and ter Meulen, 1997] J. A. F. K. van Benthem and A. ter Meulen, eds. *Handbook of Logic and Language*. North-Holland, Amsterdam, 1997.
- [Venema, 1991] Y. Venema. Many-dimensional modal logic. *Doctoral dissertation, University of Amsterdam*, ITLI (Institute for Language, Logic, and Information), vii+178pp. 1991.
- [Venema, 1995] Y. Venema. Cylindric modal logic. *Journal of Symbolic Logic*, **60,2**, 591–623, 1995.
- [Venema, 1995a] Y. Venema. A modal logic of quantification and substitution. In: Logic Colloquium '92, L. Csirmaz, D. M. Gabbay and M. de Rijke, eds., CSLI Publications and FOLLI, Stanford, pp. 293–309. 1995.
- [Venema, 1996] Y. Venema. Atom structures and Sahlqvist equations. *Algebra Universalis*, **38**, 185–199, 1997.
- [Venema, 1997] Y. Venema. Atom structures. Manuscript, Free University, Amsterdam, 1997.
- [Werner, 1978] H. Werner. Discriminator algebras. *Akademie Verlag, Berlin*, 1978.
- [Zlatoš, 1985] P. Zlatoš. On conceptual completeness of syntactic-semantical systems. *Periodica Math. Hungar.*, **16**, 3, 145–174, 1985.



## BASIC MANY-VALUED LOGIC

Many-valued logic is a vast field with hundreds of published papers and numerous monographs devoted to it. I have attempted to keep this survey to manageable length by focusing on many-valued logic as an independent discipline. This means that such topics as the use of many-valued logics for proving the independence of axioms in propositional logic have been omitted.

I am indebted to Gordon Beavers, Peter O’Hearn, Wolfgang Rautenberg and Andrzej Wroński for comments on the earlier version of this survey, and to Daniele Mundici for his constructive criticism of the revised version.

### 1 EARLY HISTORY AND MOTIVATION

#### 1.1 Introduction

Although anticipations of many-valued logic are to be found in Peirce and Vasiliev, the modern era in the subject must be dated from the early papers of Łukasiewicz and Post. Independently, these authors gave the first published systematic descriptions of many-valued logical systems, the former motivated by philosophical, the latter by mathematical considerations.

#### 1.2 Łukasiewicz and Future Contingency

In his philosophical papers, now conveniently available in English translation [Łukasiewicz, 1970], Łukasiewicz engages in an ongoing battle with determinism and logical coercion. His farewell lecture of 1918 contains the following striking passage:

I have declared a spiritual war upon all coercion that restricts man’s creative activity. There are two kinds of coercion. One of them is *physical* . . . the other . . . is *logical*. We must accept self-evident principles and the theorems resulting therefrom . . . That coercion originated with the rise of Aristotelian logic and Euclidean geometry [Łukasiewicz, 1970, p. 84].

These and similar passages in ‘On Determinism’ [1970, p. 110] show that many-valued logic was not just a mathematical toy for Łukasiewicz, but rather a weapon of the most fundamental importance in his fight against the mental strait-jacket of Aristotelian logic, a weapon that he classed with non-Euclidean geometry as a tool for liberating people from the tyranny of rigid intellectual systems.

In ‘On Determinism’ he argued that if statements about future events are already true or false, then the future is as much determined as the past and differs from the past only in so far as it has not yet come to pass. His way

out of this deterministic impasse is to reject the law of excluded middle, that is, the assumption that every proposition is true or false. A third truth-value is added, to be read as ‘possible’. The resulting system of logic was developed by Łukasiewicz and his collaborators between 1920 and 1930. Their technical results appeared in the famous compendium [Łukasiewicz and Tarski, 1930], the philosophical background in [Łukasiewicz, 1930] to which we now turn.

### 1.3 Łukasiewicz’s 3-valued Matrices and their Motivation

The original 3-valued system of propositional logic is based on two connectives,  $\rightarrow$  and  $\neg$  that are intended to generalize the implication and negation connectives of classical logic. Their truth tables are as follows:

$\rightarrow$	0	$\frac{1}{2}$	1	$\neg$
0	1	1	1	1
$\frac{1}{2}$	$\frac{1}{2}$	1	1	$\frac{1}{2}$
1	0	$\frac{1}{2}$	1	0

Here 0 stands for ‘false’, 1 for ‘true’ and  $\frac{1}{2}$  for ‘possible’. A formula is said to be a three-valued tautology if it always takes the value 1, no matter what values are assigned to its variables. The value 1 is said to be the ‘designated value’ because of its special role in defining tautologies.

How did Łukasiewicz hit on his tables? Unfortunately, he is not very explicit on this crucial point. All that he tells us is that “the desired equations I obtained on the basis of detailed considerations, which were more or less plausible to me” (Łukasiewicz [1930], also [1970, p. 166]). However it is possible to make a guess. Let’s think of Łukasiewicz’s truth-values as *sets* of classical truth-values, that is,  $0 = \{F\}$ ,  $1 = \{T\}$ ,  $\frac{1}{2} = \{T, F\}$ . The intention here is that each set of classical values represents the set of values that a proposition may take in the future. Thus the proposition ‘Ronald Reagan was elected president of the USA’ has the truth value  $\{1\} = T$ , since it is determined now and henceforth to be true, while the proposition ‘A thermonuclear war will have taken place by 2500’ has the value  $\{T, F\} = \frac{1}{2}$ , since according to our current knowledge it is determined as neither true nor false, and may take either value in the future.

Now, given a ‘truth value’ (i.e. a set of classical truth values) for each of  $\varphi$  and  $\psi$ , how do we go about computing the truth value of  $\varphi \rightarrow \psi$ ? It might seem that the following idea should work: take a classical truth value from the set assigned to  $\varphi$ , a classical truth value from the set assigned to  $\psi$ , compute the value of the *classical* conditional ( $\varphi \rightarrow \psi$ )—the set of all values you get in this way is the truth value. So, for example, if  $\psi$  has the

value  $\{T\}$ ,  $\varphi \rightarrow \psi$  must have the value  $\{T\}$  as well; similarly for the case where  $\varphi$  has the value  $\{F\}$ . So far, so good:

$\rightarrow$	$\{F\}$	$\{T, F\}$	$\{T\}$	$\neg$
$\{F\}$	$\{T\}$	$\{T\}$	$\{T\}$	$\{T\}$
$\{T, F\}$	$\{T, F\}$		$\{T\}$	$\{T, F\}$
$\{T\}$	$\{F\}$	$\{T, F\}$	$\{T\}$	$\{F\}$

But how do we fill in the remaining central entry? According to our way of looking at the matter, it should be  $\{T, F\}$ ; but Łukasiewicz’s table has  $\{T\}$ , or rather 1. Why? The reason is not far to seek; Łukasiewicz wants  $\varphi \rightarrow \varphi$  to be a three-valued tautology.

In fact, Łukasiewicz has taken over from classical logic two basic assumptions that he does not critically examine, in spite of the polemical character of his attack on Aristotelian logic. The assumptions are:

1. Logic should be formulated as in *Principia Mathematica* using axioms, substitution and *modus ponens*.
2. The values of complex propositions should be a function of the values of their component parts (generalized extensionality).

Given these assumptions, we can see how the central entry in the truth-table is forced. If the central entry were  $\frac{1}{2}$ , then there would be no three-valued tautologies. But *should* the central entry be  $\frac{1}{2}$  (or  $\{T, F\}$  in the modified notation)? Let’s re-examine the whole question.

If  $\varphi$  and  $\psi$  express the *same proposition*, then  $\varphi \rightarrow \psi$  certainly should have the value 1, whether or not we are unsure about the value of  $\varphi$  (and hence  $\psi$ ). But if this condition *doesn’t* hold, then it doesn’t follow. For example, suppose  $\varphi$  is the statement ‘A global nuclear war has taken place by the year 2500’, and  $\psi$  is the statement ‘The human race is extinct by the year 2600’. Obviously, we are presently unsure about the truth value of  $\varphi$  and  $\psi$ . But what about  $\varphi \rightarrow \psi$ ? If we read it as some kind of counterfactual, then we might think it true or false, or be undecided about it. Or we could reason according to our earlier scheme:  $\varphi$  could be true or false, so could  $\psi$ , the propositions are logically independent, so  $\varphi \rightarrow \psi$  could have either value.

The conclusion here seems inescapable. The logic of the ‘possible’ in Łukasiewicz’s sense is just not truth-functional (an observation first made by Gonseth [1941]). This is no more surprising than the fact that the probability calculus is not truth-functional, and it holds for the same reasons. However, it throws in doubt Łukasiewicz’s claim to provide a serious logical and philosophical alternative to Aristotelian logic.



### 1.4 Other Łukasiewicz Logics

Łukasiewicz generalized his 3-valued logic to  $n$  values and also to an infinite-valued system in 1922. The matrix for the infinite-valued system is defined on the rational numbers in the closed unit interval from 0 to 1. For  $x, y$  in the interval, we have:  $x \rightarrow y = \min(1, 1 - x + y)$ ,  $\neg x = 1 - x$ . If instead of the whole rational interval, a finite subset closed under the above functions is chosen, the result is the  $n$ -valued Łukasiewicz connectives, for some  $n$ .

Łukasiewicz himself [1930] expressed a philosophical preference for the infinite-valued logic. It should be pointed out, though, that the transition to infinitely many values makes no difference to the critique given above.

### 1.5 Post's Many-valued Systems

Emil Post [1921] independently gave a formal development of many-valued logic. His  $m$ -valued systems, defined on the set  $\{0, \dots, m - 1\}$  (we are altering Post's notation slightly) have as primitive operators a generalized disjunction and a generalized negation:

$$\begin{aligned} x \vee y &= \min(x, y) \\ \neg x &= x + 1 \pmod{m}. \end{aligned}$$

Post's remarkable paper includes not only a proof of functional completeness for his system of connectives, but also a general method of constructing a complete axiomatization of the system  $T_m^n$ , where the values  $0, \dots, n$  are designated, for  $0 \leq n < m - 1$ .

### 1.6 Bochvar and the Paradoxes

The work of the Russian logician Bochvar [1939] represents a new philosophical motivation for many-valued logic; its use as a means of avoiding the logical paradoxes. His system introduces the intermediate value  $I$  in addition to the classical values  $T$  and  $F$ . His idea is to avoid logical paradoxes such as Russell's and Grelling's by declaring the crucial sentences involving them to be meaningless (having the value  $I$ ). Bochvar's basic tables for his connectives are as follows. When only the values  $T, F$  are involved, they are exactly like their classical counterparts; any formula having a meaningless component is meaningless. For example, Bochvar's conjunction and negation have the tables:

$\wedge$	$T$	$I$	$F$	$\neg$
$T$	$T$	$I$	$F$	$F$
$I$	$I$	$I$	$I$	$I$
$F$	$F$	$I$	$F$	$T$

If we take  $T$  as the only designated value, then it is clear that there are no tautologies in the system. This feature of the system can be repaired by adding an ‘assertion operator’  $Ap$ , that is intended to represent the ‘external assertion’ of a proposition  $p$ , so that  $Ap$  can be considered as the assertion ‘ $p$  is true’ in a two-valued metalanguage. Thus  $Ap$  is  $T$  if  $p$  is true, otherwise  $Ap$  is false. Using this operator, we can define the ‘external connectives’ that always take the values  $T$  or  $F$ . For example, the external negation  $\sim p$  is defined as  $\neg Ap$ , external conjunction  $p \& q$  as  $Ap \wedge Aq$ :

$\&$	$T$	$I$	$F$	$\sim$
$T$	$T$	$F$	$F$	$F$
$I$	$F$	$F$	$F$	$T$
$F$	$F$	$F$	$F$	$T$

If we confine our logic to internal negation, it would seem that we can avoid Russell’s paradox. Denoting the truth-value of a formula  $\varphi$  by  $\llbracket \varphi \rrbracket$ , we assume the basic comprehension principle:  $\llbracket a \in \{x \mid \dots x \dots\} \rrbracket = \llbracket \dots a \dots \rrbracket$ . Then defining  $R = \{x \mid \neg x \in x\}$ , the equation that results from substitution in the above, namely  $\llbracket R \in R \rrbracket = \neg \llbracket R \in R \rrbracket$  is consistent, since  $\llbracket R \in R \rrbracket$  can be  $I$ . However, as Church pointed out in his review [1939], if we define  $R' = \{x \mid \sim x \in x\}$ , paradox again results. The Russell paradox rules out the presence of the assertion operator or external negation.

### 1.7 Kleene’s System

In 1938, Kleene introduced yet another 3-valued logic [1938], see also [1952]. His connectives are defined as follows:

$q$	$p \wedge q$			$p \vee q$			$p \supset q$			$p \equiv q$			$\neg p$
	$T$	$I$	$F$	$T$	$I$	$F$	$T$	$I$	$F$	$T$	$I$	$F$	
$T$	$T$	$I$	$F$	$T$	$T$	$T$	$T$	$I$	$F$	$T$	$I$	$F$	$F$
$I$	$I$	$I$	$F$	$T$	$I$	$I$	$T$	$I$	$I$	$I$	$I$	$I$	$I$
$F$	$F$	$F$	$F$	$T$	$I$	$F$	$T$	$T$	$T$	$F$	$I$	$T$	$T$

Kleene’s motivation arises from the theory of recursive functions. In that theory, if we think of a machine designed to respond ‘true’ or ‘false’ to certain questions, then for certain inputs the machine may not provide an answer, perhaps by going into an infinite loop, or by exhausting its computing capacity. In that case, we can think of the machine’s response as undefined or ‘ $T$ ’. In this light, the truth-tables can be seen as rules for computing the truth-values of complex predicates. For example, if we are

computing the value of a disjunction, then we give the disjunction the value  $T$  as soon as the machine gives the answer  $T$  for either disjunct. Notice that for Kleene (unlike Bochvar) a compound sentence can have a truth-value even if some of its components lack a truth-value. Kleene also considers a set of tables identical with Bochvar's, which he calls the 'weak connectives'; the tables above are for Kleene's 'strong connectives'.

## 2 GENERAL THEORY OF MANY-VALUED LOGICS

In the preceding section we have surveyed a somewhat heterogeneous collection of logical systems, that have in common the idea of enlarging the set of classical truth-values, with varied interpretations for the added non-classical truth-values, such as 'meaningless', 'undefined' or 'presently undetermined'. In the present section, we abandon philosophical and motivational discussion and attempt a systematization.

### 2.1 *The Matrix Method*

We assume in what follows languages for sentential logic with an infinite supply of sentential variables  $p, q, r, p_1, q_1, r_1, \dots$ . If  $L, L'$  are such languages, we say that  $L$  is a sublanguage of  $L'$  if it is based on the same connectives as  $L'$ , and is a subset of  $L'$ . To simplify notation let us assume that we have one two-place connective  $(\varphi * \psi)$  and a one-place connective  $-\varphi$ . If  $\Gamma$  is a set of formulas, we write  $Var(\Gamma)$  for the set of variables in  $\Gamma$ , and  $Var(\varphi)$  for the set of variables in a formula  $\varphi$ .

A *substitution* for a language  $L$  is a function from the set of variables into  $L$ . If  $\varphi$  is a formula of  $L$  and  $g$  a substitution, then we write  $\varphi^g$  for the result of applying the substitution  $g$  to the variables in  $\varphi$ . If  $\Gamma$  is a set of formulas of  $L$  and  $g$  a substitution,  $\Gamma^g$  is the set of all formulas  $\varphi^g$  for  $\varphi$  in  $\Gamma$ .

A *matrix* for the language  $L$  consists of (1) an abstract algebra  $\mathfrak{A}$  of the appropriate type, i.e. a non-empty set  $A$  with a two-place operation  $x + y$  and a one-place operation  $fx$  defined on  $A$ , (2) a non-empty subset  $D \subseteq A$ —the elements of  $D$  are the *designated elements* of  $A$ . For example, in Łukasiewicz's 3-valued logic,  $A = \{0, \frac{1}{2}, 1\}$ ,  $D = \{1\}$ ,  $x + y = x \rightarrow y$ , and  $fx = \neg x$ .

If  $g$  is an assignment of elements of the matrix to propositional variables in  $L$ , then  $g$  can be extended to all of  $L$ , that is, we can define:  $g(\varphi * \psi) = g(\varphi) + g(\psi)$ , and  $g(-\varphi) = f(g(\varphi))$ .

The basic concepts of universal algebra extend in a straightforward way to matrices. If the elements of  $\mathfrak{A}_1$  are a subset of  $\mathfrak{A}_2$ , and the operations and designated values of  $\mathfrak{A}_1$  are just the restriction of those of  $\mathfrak{A}_2$  to the domain of  $\mathfrak{A}_1$ , then  $\mathfrak{A}_1$  is said to be a *submatrix* of  $\mathfrak{A}_2$ . For example, it is

easy to see that the classical matrices are submatrices of all the 3-valued logics introduced above.

## 2.2 Consequence Relations

To define the logical system determined by a matrix, it is possible to generalize the classical concept of tautology, following Łukasiewicz's lead (see Section 1.3); but the result is insufficient in the sense that some systems contain no tautologies at all. For example, in Bochvar's system, if we take only  $T$  as a designated value, then there are no tautologies. Thus the concept of tautology is inadequate in that by employing it we cannot distinguish between systems that have quite distinct matrix definitions—for example, the Bochvar system with  $\wedge$  alone, and the Bochvar system with  $\vee$  alone. To obtain a concept adequate to the general case, we need the notion of consequence relation.

Let  $\Gamma \cup \{\psi\}$  be a finite set of formulas of  $L$ . We say that  $\psi$  is a *consequence* of  $\Gamma$  with respect to the matrix  $M$ ,  $\Gamma \vDash_M \psi$ , if the following holds: for every assignment  $g$  of elements of  $M$  to variables in  $L$ , if  $g(\varphi) \in D$  for all  $\varphi \in \Gamma$  then  $g(\psi) \in D$ . A formula  $\varphi$  is a *tautology* with respect to  $M$  if  $\emptyset \vDash_M \varphi$ —we abbreviate this as  $\vDash_M \varphi$ . We also abbreviate  $\Gamma \cup \Delta \vDash_M \psi$  as  $\Gamma, \Delta \vDash_M \psi$  and  $\Gamma \cup \{\varphi\} \vDash_M \psi$  as  $\Gamma, \varphi \vDash_M \psi$ . If  $F$  is a family of matrices, then we say that  $\psi$  is a consequence of  $\Gamma$  with respect to  $F$ ,  $\Gamma \vDash_F \psi$ , if  $\Gamma \vDash_M \psi$  for all matrices  $M$  in  $F$ .

If  $F$  is a family of matrices for a language  $L$ , then the consequence relation  $\vDash_F$  satisfies the conditions:

$$\begin{aligned} & \Gamma \vDash_F \psi \text{ if } \psi \in \Gamma \\ \text{If } & \Gamma, \varphi \vDash_F \psi \text{ and } \Delta \vDash_F \varphi, \text{ then } \Gamma, \Delta \vDash_F \psi \quad (\text{Cut}). \end{aligned}$$

A *sequent* we define as a pair  $\langle \Gamma, \varphi \rangle$  consisting of a set  $\Gamma$  of formulas of  $L$ , and a formula  $\varphi$  of  $L$ , written as  $\Gamma \vdash \varphi$ . If  $\vDash$  is a consequence relation, we say that a sequent  $\Gamma \vdash \varphi$  is in  $\vDash$  if  $\Gamma \vDash \varphi$ .

If  $L$  is a language, we define a *consequence relation in  $L$*  as any relation between finite sets of formulas and formulas in  $L$  that obeys the above rules. If  $\Gamma$  is a finite set of formulas of  $L$ , and  $\vDash$  a consequence relation in  $L$ , then we say that  $\Gamma$  is *consistent with respect to  $\vDash$* , or simply *consistent* (where  $\vDash$  is understood) if there is a formula  $\varphi$  of  $L$  so that  $\Gamma \not\vDash \varphi$ .

A matrix  $M$  *validates* a sequent  $\Gamma \vdash \varphi$  if  $\Gamma \vDash_M \varphi$ ;  $M$  *validates* a consequence relation  $\vDash$  if it validates all the sequents in  $\vDash$ . If an abstract consequence relation  $\vdash$  coincides with a consequence relation  $\vDash_M$  determined by a matrix  $M$ , then we say that  $M$  is a characteristic matrix for  $\vdash$ . Similarly, if an abstract consequence relation  $\vdash$  coincides with a consequence relation  $\vDash_F$  determined by a set of matrices  $F$ , then we say that  $F$  is a characteristic set of matrices for  $\vdash$ .

It is natural to ask what the general properties are that hold for consequence relations with characteristic sets of matrices. It is obvious that a consequence relation  $\vDash_F$  must satisfy the rule of uniform substitution:

If  $\Gamma \vDash_F \psi$  and  $g$  is a substitution, then  $\Gamma^g \vDash_F \psi^g$ .

Let us say that a consequence relation is *structural* if it satisfies this rule. This added condition is sufficient to characterize consequence relations with characteristic sets of matrices.

**THEOREM 1.** *If  $\vDash$  is a structural consequence relation then  $\vDash$  has a characteristic set of matrices.*

**Proof.** Let  $\vDash$  be a structural consequence relation, and suppose that  $\Gamma \not\vDash \varphi$ . We wish to show that there is a matrix  $M$  validating  $\vDash$ , so that  $\Gamma \not\vDash_M \varphi$ .

Define the matrix  $M$  as follows: the elements of the matrix are the formulas of  $L$ , where  $\varphi + \psi = (\varphi * \psi)$ ,  $f(\varphi) = -\varphi$ . The designated elements  $D$  of  $M$  are those formulas  $\varphi$  such that  $\Gamma \vDash \varphi$ . If we assign each variable in  $L$  to itself, then under this assignment all the formulas in  $\Gamma$  take a designated value, and  $\varphi$  is not designated, so  $\Gamma \vDash \varphi$  is not validated by  $M$ .

It remains to show that  $M$  validates  $\vDash$ . Assume that  $\Pi \vDash \psi$ , and that  $g$  is an assignment of variables so that  $g(\Pi) \subseteq D$ . Thinking of  $g$  as a substitution, we have  $g(\Pi) = \Pi^g$  and  $g(\psi) = \psi^g$ . By the substitution rule, we have  $\Pi^g \vDash \psi^g$ . For all  $\eta \in \Pi^g$ ,  $\Gamma \vDash \eta$ , so by the Cut rule,  $\Gamma \vDash \psi^g$ , that is,  $g(\psi) = \psi^g \in D$ , completing the proof. ■

Consequence relations that have a single characteristic matrix are clearly structural, but in addition satisfy a further condition:

If  $\Gamma, \Delta \vDash \varphi$ ,  $Var(\Gamma \cup \{\varphi\}) \cap Var(\Delta) = \emptyset$ , and  $\Delta$  is consistent then  $\Gamma \vDash \varphi$ .

We define a consequence relation to be *uniform* if it satisfies this added condition. This condition is sufficient to characterize consequence relations having a characteristic matrix.

As a preliminary to proving this result, we give a lemma on consequence relations. If  $\vDash$  is a consequence relation in a language  $L$ , and  $L'$  is a sublanguage of  $L$ , we define the *natural extension* of  $\vDash$  to  $L'$  to be the consequence relation  $\vDash'$  in  $L'$  defined by:  $\Gamma' \vDash' \varphi'$  if and only if there is a formula  $\varphi$  in  $L$ , a set of formulas  $\Gamma$  of  $L$  and a substitution  $\sigma$  so that  $\varphi^\sigma = \varphi'$ ,  $\Gamma^\sigma \subseteq \Gamma'$ , and  $\Gamma \vDash \varphi$ .

**LEMMA 2.** *If  $\vDash$  is a uniform structural consequence relation in  $L$ ,  $L'$  is a sublanguage of  $L$ , and  $\vDash'$  is the natural extension of  $\vDash$  to  $L'$ , then  $\vDash'$  is also a uniform structural consequence relation.*

**Proof.** Let  $\Gamma \cup \{\varphi\}$  be a finite set of formulas in  $L'$ , and  $\sigma$  a bijective substitution whose domain is  $Var(\Gamma \cup \{\varphi\}) \setminus Var(L)$ , and whose range is a

set of variables in  $L$  disjoint from  $Var(\Gamma \cup \{\varphi\})$ . Then  $\Gamma \vDash' \varphi$  if and only if  $\Gamma^\sigma \vDash \varphi^\sigma$ . Using this fact, it is not hard to show that each of the properties in the lemma is preserved in passing from  $\vDash$  to  $\vDash'$ . ■

**THEOREM 3.** *If  $\vDash'$  is a uniform structural consequence relation then  $\vDash$  has a characteristic matrix.*

**Proof.** Let  $L$  be the language of the consequence relation  $\vDash$ . We enlarge  $L$  to a new language  $L'$  by adding sufficiently many propositional variables so that there is a family of substitutions  $\{\sigma(X) : X \text{ a consistent subset of } L\}$ , where each  $\sigma(X)$  is a bijective substitution defined on  $Var(L)$  whose range is a set of variables in  $L'$ , and for distinct  $X, Y$  the ranges of  $\sigma(X)$  and  $\sigma(Y)$  are distinct. By Lemma 2, the natural extension  $\vDash'$  of  $\vDash$  to  $L'$  is a uniform structural consequence relation.

Define a matrix  $M$  by taking as elements of the matrix the formulas of  $L'$ ; an element  $\varphi$  of  $M$  belongs to the set  $D$  of designated values if there are sets of formulas  $X_1, \dots, X_n$  in  $L$  consistent with respect to  $\vDash$  so that

$$\overline{X_1}, \dots, \overline{X_n} \vDash' \varphi,$$

where  $\overline{X} = X^{\sigma(X)}$ .

We need to verify that the defined matrix  $M$  validates  $\vDash$ . Assume that  $\Gamma \vDash \varphi$ , and let  $g$  be an assignment in  $M$  so that  $g(\Gamma) \subseteq D$ . By the definition of  $D$ , for any  $\psi \in \Gamma$ , there are sets of formulas  $X_1, \dots, X_m \subseteq L$  consistent with respect to  $\vDash$  so that

$$\overline{X_1}, \dots, \overline{X_m} \vDash' \psi^g.$$

By the definition of  $\vDash'$ ,  $\Gamma^g \vDash' \varphi^g$ , so by the Cut rule,

$$\overline{X_1}, \dots, \overline{X_n} \vDash' \varphi^g,$$

that is to say,  $g(\varphi) \in D$ .

For the converse, let us suppose that  $\Gamma \vDash_M \varphi$ , where  $\Gamma \cup \{\varphi\} \subseteq L$ . If  $\Gamma$  is not consistent with respect to  $\vDash$ , then  $\Gamma \vDash \varphi$ . Thus assuming  $\Gamma$  consistent with respect to  $\vDash$ , let  $\sigma = \sigma(\Gamma)$ . Then  $\Gamma^\sigma = \overline{\Gamma} \subseteq D$ , so  $\sigma(\varphi) \in D$ . Hence there are sets of formulas  $Y_1, \dots, Y_m$  in  $L$  consistent with respect to  $\vDash$  so that

$$\overline{\Gamma}, \overline{Y_1}, \dots, \overline{Y_m} \vDash' \varphi^\sigma.$$

Since  $Y_1, \dots, Y_m$  are consistent with respect to  $\vDash$ ,  $\overline{Y_1}, \dots, \overline{Y_m}$  are consistent with respect to  $\vDash'$ . By construction, the ranges of the substitutions  $\sigma(\Gamma), \sigma(Y_1), \dots, \sigma(Y_m)$  are mutually disjoint. It follows by repeated use of the uniformity of  $\vDash'$  that  $\overline{\Gamma} \vDash' \varphi^\sigma$ . Since the substitution  $\sigma$  is bijective, there is a substitution  $\sigma'$  so that  $\Gamma^{\sigma\sigma'} = \overline{\Gamma}$  and  $\varphi^{\sigma\sigma'} = \varphi$ . Because  $\vDash'$  is structural, it follows that  $\Gamma \vDash' \varphi$ , hence  $\Gamma \vDash \varphi$ . ■

This result, a generalization of the classic result of Lindenbaum, is due to Los and Suszko [1958]. The consequence relations considered above are *finitary*, that is to say, the domain of the relations consists solely of finite sets. It is possible to consider a more general concept of consequence relation in which this restriction is dropped. In this case, a version of Theorem 3 also holds. However, the condition of uniformity needs to be replaced by the following stronger condition: If  $\Gamma, \Delta \vDash \varphi$  and  $\Delta$  is a union of consistent sets that have no variables in common with one another or with  $\Gamma$  or  $\varphi$ , then  $\Gamma \vDash \varphi$ . For the details of this result see [Wójcicki, 1970] and [Shoesmith and Smiley, 1971]. The reader is also referred to the book of Wójcicki [1988] for this and many other results in the general theory of consequence relations.

We conclude this subsection by noting that not all logics can be considered as many-valued logics in the sense that the consequence relations corresponding to them fail to satisfy the uniformity condition. For example, in the minimal logic of Johansson (see Chapter 4.3), the sequent  $P, \neg P \vdash \neg Q$  is valid, though the set  $\{P, \neg P\}$  is consistent, so the uniformity condition fails. The same thing is true for the modal logics S1, S2 and S3 of C.I. Lewis (see [Shoesmith and Smiley, 1971, Theorem 6]).

### 2.3 The Bochvar Consequence Relation

Now that we have formulated the general concept of consequence relation, it is possible to address the problem of finding complete sets of rules for the consequence relations of the matrices discussed above.

For Bochvar's matrices, it is easiest to formulate the consequence relation as a special form of the classical consequence relation. Let us write  $\vDash_{\mathbf{B}}$  for the consequence relation determined by Bochvar's matrices, with  $T$  the sole designated value.

**THEOREM 4.**  $\Sigma \vDash_{\mathbf{B}} \varphi$  holds if and only if (1)  $\Sigma \vDash \varphi$  is classically valid; (2) If  $\Sigma$  is classically consistent, then every variable in  $\varphi$  occurs in  $\Sigma$ .

**Proof.** If  $\Sigma \vDash_{\mathbf{B}} \varphi$ , then the first condition holds because the classical two-valued matrices are submatrices of Bochvar's matrices. If the second condition fails, then we can assign the values  $T$  and  $F$  to the variables in  $\Sigma$ , but the value  $I$  to an extra variable in  $\varphi$  so as to satisfy  $\Sigma$  but give  $\varphi$  the value  $I$ .

For the converse, if the two conditions hold, then any valuation in the Bochvar matrices giving all formulas in  $\Sigma$  the value  $T$  must assign values  $T$  or  $F$  to all variables in  $\Sigma$ , showing that  $\varphi$  must also take the value  $T$ . ■

### 2.4 The Kleene Consequence Relation

In the case of Kleene's truth tables, we proceed by adding a number of sequents as axioms to our basic structural rules for the abstract consequence

relation in Section 2.2, together with an added rule of inference. We take the operators  $\wedge, \vee, \neg$  as primitive, since we have the definitions:

$$p \supset q = \neg p \vee q, \quad p \equiv q = (p \wedge q) \vee (\neg p \wedge \neg q).$$

Let us define  $\vdash_{\mathbf{K}}$  as the smallest structural consequence relation that contains the sequents:

$$\begin{array}{ll} p \vdash_{\mathbf{K}} \neg\neg p & p, \neg p \vdash_{\mathbf{K}} q \quad \neg\neg p \vdash_{\mathbf{K}} p \\ p \wedge q \vdash_{\mathbf{K}} p & p \vdash_{\mathbf{K}} p \vee q \\ p \wedge q \vdash_{\mathbf{K}} q & q \vdash_{\mathbf{K}} p \vee q \\ p, q \vdash_{\mathbf{K}} p \wedge q & \neg p, \neg q \vdash_{\mathbf{K}} \neg(p \vee q) \\ \neg p \vdash_{\mathbf{K}} \neg(p \wedge q) & \neg(p \wedge q) \vdash_{\mathbf{K}} \neg p \vee \neg q \\ \neg q \vdash_{\mathbf{K}} \neg(p \wedge q) & \neg(p \vee q) \vdash_{\mathbf{K}} \neg p \wedge \neg q, \end{array}$$

and is closed under the rule of inference:

$$\text{If } \Sigma, \varphi \vdash_{\mathbf{K}} \eta \text{ and } \Sigma, \psi \vdash_{\mathbf{K}} \eta \text{ then } \Sigma, \varphi \vee \psi \vdash_{\mathbf{K}} \eta \quad (\text{Dilemma}).$$

Let  $\vDash_{\mathbf{K}}$  stand for the consequence relation determined by the Kleene matrices. Before proving the completeness theorem for the Kleene consequence relation, we need a lemma that will also be useful in later sections. We define a consequence relation  $\vDash$  to be  $\Sigma$ -prime if  $\Sigma \vDash \varphi \vee \psi$  implies  $\Sigma \vDash \varphi$  or  $\Sigma \vDash \psi$  for any  $\varphi, \psi$ .

**LEMMA 5.** *If  $\vDash$  is a consequence relation satisfying the Dilemma rule, and  $\Sigma \not\vDash \varphi$ , then there is a  $\Sigma$ -prime consequence relation  $\vDash'$  extending  $\vDash$  so that  $\Sigma \not\vDash' \varphi$ .*

**Proof.** Consider the family  $F$  of all consequence relations  $\vDash''$  extending  $\vDash$  that satisfy the Dilemma rule, but  $\Sigma \not\vDash'' \varphi$ . The union of any chain of consequence relations in  $F$  is also in  $F$ , so by Zorn's lemma  $F$  contains a maximal element  $\vDash'$ .

It remains to show that  $\vDash'$  is  $\Sigma$ -prime. Suppose that  $\Sigma \vDash' \psi_1 \vee \psi_2$ , but neither  $\Sigma \vDash' \psi_1$  nor  $\Sigma \vDash' \psi_2$ . Define relations  $\vDash_1$  and  $\vDash_2$  by the definitions:  $\Theta \vDash_1 \eta$  if and only if  $\Theta, \psi_1 \vDash' \eta$  and  $\Theta \vDash_2 \eta$  if and only if  $\Theta, \psi_2 \vDash' \eta$ . Both  $\vDash_1$  and  $\vDash_2$  are consequence relations that satisfy the Dilemma rule and properly extend  $\vDash'$ . It follows by the maximality of  $\vDash'$  that  $\Sigma \vDash_1 \varphi$  and  $\Sigma \vDash_2 \varphi$ , that is to say,  $\Sigma, \psi_1 \vDash' \varphi$  and  $\Sigma, \psi_2 \vDash' \varphi$ . By the Dilemma rule,  $\Sigma, \psi_1 \vee \psi_2 \vDash' \varphi$ . However, since  $\Sigma \vDash' \psi_1 \vee \psi_2$ , it follows by the Cut rule that  $\Sigma \vDash' \varphi$ , a contradiction.  $\blacksquare$

**THEOREM 6.**  $\Sigma \vdash_{\mathbf{K}} \varphi$  if and only if  $\Sigma \vDash_{\mathbf{K}} \varphi$ .

**Proof.** Suppose that  $\Sigma \not\vdash_{\mathbf{K}} \varphi$ . By Lemma 5 there is a  $\Sigma$ -prime consequence relation  $\vdash$  extending  $\vdash_{\mathbf{K}}$  in which  $\Sigma \not\vdash \varphi$ . Now define an assignment of values  $g$  as follows:

$$\begin{array}{l} g(p) = T \text{ iff } \Sigma \vdash p, \quad g(p) = F \text{ if } \Sigma \vdash \neg p, \\ g(p) = I \text{ otherwise.} \end{array}$$



We note that this assignment is consistent, because if  $\Sigma \vdash p$  and  $\Sigma \vdash \neg p$  then  $\Sigma \vdash \varphi$  by  $p, \neg p \vdash q$  and the cut rule.

We can prove by using the basic sequents and the Cut and Contraposition rules that  $\Sigma \vdash_{\mathbf{K}} \varphi \wedge \psi$  holds if and only if  $\Sigma \vdash_{\mathbf{K}} \varphi$  and  $\Sigma \vdash_{\mathbf{K}} \psi$ , that  $\Sigma \vdash_{\mathbf{K}} \neg(\varphi \wedge \psi)$  holds if and only if  $\Sigma \vdash_{\mathbf{K}} \neg\varphi \vee \neg\psi$ , and that  $\Sigma \vdash_{\mathbf{K}} \neg(\varphi \vee \psi)$  holds if and only if  $\Sigma \vdash_{\mathbf{K}} \neg\varphi$  and  $\Sigma \vdash_{\mathbf{K}} \neg\psi$ . Using these equivalences, we argue inductively that for any  $\psi$ ,  $\Sigma \vdash \psi$  if and only if  $g(\psi) = T$ , and  $\Sigma \vdash \neg\psi$  if and only if  $g(\psi) = F$ , completing the proof. ■

Those who are familiar with relevance logic will notice a similarity between the logic just defined and first degree entailments [4.7]. The Kleene system does not contain the paradox of material implication  $p \vdash q \vee \neg q$ ; however it contains  $p, \neg p \vdash q$ , so it is not free of the paradoxes of material implication. The relationship between the two systems can be briefly indicated by noting that while Kleene allows for the possibility ‘neither true nor false’, Anderson and Belnap allow for the possibility ‘both true and false’. We are imagining a computational system (for example) that is attempting to act on the basis of inconsistent information. The reader is referred to Belnap’s interesting paper [Belnap, 1977] for more details of this interpretation.

An application of Kleene’s truth tables to a problem of a somewhat different sort is to be found in the chapter by Blamey on partial logic in Volume 7 of this *Handbook* and in Visser’s chapter on the liar paradox in a later volume.

### 2.5 Łukasiewicz Consequence (Finite Case)

In this section we give an axiomatization of all the finite-valued Łukasiewicz logics. Before giving the axiomatization, we prove a lemma on definability in these logics, that will also prove important in later sections.

First, we make a notational change in the systems, to facilitate the proof. In the  $m + 1$ -valued logic of Łukasiewicz,  $\mathbf{L}_{m+1}$ , we take the truth values to be  $0, 1, \dots, m$ , with implication as modified subtraction; for  $x, y \in \{0, \dots, m\}$ ,  $x \rightarrow y$  is  $y - x$ , where  $y - x$  is  $y - x$  if  $y \geq x$ , 0 otherwise. Negation  $\neg x$  is  $m - x$ . So, for example, the basic truth tables for  $\mathbf{L}_5$  are:

$\rightarrow$	0	1	2	3	4	$\neg$
0	0	1	2	3	4	4
1	0	0	1	2	3	3
2	0	0	0	1	2	2
3	0	0	0	0	1	1
4	0	0	0	0	0	0

In contrast to Łukasiewicz, we are taking the smallest value to be the ‘truest’, the largest to be the ‘falsest’. We note that if we define  $(\varphi \vee \psi)$  as

$(\varphi \rightarrow \psi) \rightarrow \psi$  then  $x \vee y = \min(x, y)$  and defining  $\varphi \wedge \psi$  as  $\neg(\neg\varphi \vee \neg\psi)$ , we have  $x \wedge y = \max(x, y)$ . For  $k \in \{0, \dots, m\}$ , we define the function  $J_k$  by :  $J_k(x) = 0$  if  $x = k$ ,  $J_k(x) = m$  otherwise. A function defined on  $\{0, \dots, m\}$  is said to be  $\mathbf{L}_{m+1}$ -definable if it can be expressed in terms of  $\rightarrow$  and  $\neg$ .

LEMMA 7. *For any  $k \in \{0, \dots, m\}$ ,  $J_k$  is  $\mathbf{L}_{m+1}$ -definable.*

**Proof.** We first show that for any  $k$  where  $0 \leq k \leq m$ , the function

$$I_k(x) = \begin{cases} m & \text{if } x \leq k \\ 0 & \text{if } x > k \end{cases}$$

is definable. We define inductively:

$$(a) \ H_1(x) = \neg x, \quad (b) \ H_{n+1}(x) = x \rightarrow H_n(x).$$

It is easy to see that for any  $n$ ,  $H_n(x) = m - nx$ . Define  $I_0(x) = H_m(x)$ . The function  $I_k$ ,  $0 \leq k \leq m$  is defined by induction on  $k$ . Suppose  $I_q$  defined for  $q \leq k$ . Let  $r$  be the largest integer such that  $H_r(k + 1) > 0$ . Define  $p = H_r(k + 1) - 1$ . By construction,  $p \leq k$ . It follows that we can define:  $I_{k+1}(x) = \neg I_p(H_r(x))$ . Then we can define:  $J_0(x) = \neg I_0(x)$ ,  $J_k(x) = \neg I_k(x) \wedge I_{k-1}(x)$ . ■

With the  $J$  functions defined, it is a simple matter to axiomatize the consequence relation of  $\mathbf{L}_{m+1}$ . Let  $\vdash_{m+1}$  be the smallest structural consequence relation containing the following sequents for all  $x, y, z \in \{0, \dots, m\}$ :

- (A)  $\vdash J_0(p) \vee \dots \vee J_m(p)$
- (B)  $J_x(p), J_y(p) \vdash q$ , for  $x \neq y$
- (C)  $J_x(p), J_y(q) \vdash J_{x \rightarrow y}(p \rightarrow q)$
- (D)  $J_x(p) \vdash J_{\neg x}(\neg p)$
- (E)  $J_0(p) \vdash p$
- (F)  $p \vdash J_0(p)$ ,

and closed under the Dilemma rule.

THEOREM 8.  $\Sigma \vdash_{m+1} \varphi$  iff  $\Sigma \models_{m+1} \varphi$ .

**Proof.** If  $\Sigma \not\vdash_{m+1} \varphi$  then by Lemma 5 there is a  $\Sigma$ -prime consequence relation  $\vdash$  extending  $\vdash_{m+1}$  such that  $\Sigma \not\vdash \varphi$ . Now define an assignment of values in  $\mathbf{L}_{m+1}$  as follows:  $g(p) = k$  iff  $\Sigma \vdash J_k(p)$ . By the sequents (A) and (B), the function  $g$  is well defined. Using (C) and (D), we can show that for any  $\varphi$ ,  $\Sigma \vdash J_k(\varphi)$  iff  $g(\varphi) = k$ . By (F), every formula in  $\Sigma$  takes the designated value 0; by (E),  $\varphi$  cannot take the value 0. Thus  $\Sigma \not\vdash_{m+1} \varphi$ . ■

The above axiomatization of the  $\mathbf{L}_{m+1}$  consequence relation is convenient for the purposes of the completeness proof, though hardly elegant. A more perspicuous axiomatization of  $\mathbf{L}_3$  was provided by Wajsberg [1931], who shows that the tautologies of this logic can be axiomatized by the axioms:

1.  $p \rightarrow (q \rightarrow p)$
2.  $(p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r))$
3.  $(\neg p \rightarrow \neg q) \rightarrow (q \rightarrow p)$
4.  $((p \rightarrow \neg p) \rightarrow p) \rightarrow p,$

using the rules of substitution and *modus ponens*.

The techniques we have just used for  $\mathbf{L}_{m+1}$  apply to any finite-valued logic in which analogues of the  $J_k$  functions are definable, for example Post's many-valued systems. For more general results in Łukasiewicz logics, including the case where a different set of designated values is adopted in  $\mathbf{L}_m$ , the reader is referred to the monograph of Rosser and Turquette [1952] and to the papers of Rose (see the bibliographies of [Rescher, 1969] and [Wolf, 1977]) who has investigated in great depth the many possibilities in this area. Grigolia [1973; 1977] provided axiomatic versions of the logics  $\mathbf{L}_n$ , for  $n > 3$ .

The relations between the finite-valued systems  $\mathbf{L}_m$  is completely settled by the following elegant result of Lindenbaum.

**THEOREM 9.**  $\vDash_m \subseteq \vDash_n$  iff  $n - 1$  divides  $m - 1$ .

**Proof.** We first have to settle the case where  $m < n$ . Define  $\varphi \leftrightarrow \psi$  as  $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$ . It is easy to see that for any assignment of values  $g$  in  $\mathbf{L}_m$ ,  $g(\varphi \leftrightarrow \psi) = |g(\varphi) - g(\psi)|$ , where  $|x|$  is the absolute value of  $x$ . For  $k > 1$ , let the formula  $\delta_k$  be the disjunction of all formulas of the form  $p_i \leftrightarrow p_j$ ,  $0 \leq i < j \leq k$ . For example,  $\delta_2$  is:

$$(p_0 \leftrightarrow p_1) \vee (p_1 \leftrightarrow p_2) \vee (p_0 \leftrightarrow p_2).$$

Now  $\delta_k$  is a tautology of  $\mathbf{L}_m$  if and only if  $m \leq k$ ; this follows from the pigeon-hole principle. If there are  $k$  truth values or fewer, at least two distinct variables in  $\delta_k$  must take the same value, so  $\delta_k$  takes the value 0; if there are more than  $k$  truth values, simply assign a different value to each variable. It follows that if  $m < n$ ,  $\delta_m$  is a tautology of  $\mathbf{L}_m$ , but not of  $\mathbf{L}_n$ , showing that  $\vDash_m$  is not contained in  $\vDash_n$ .

So, assuming that we have  $m \geq n$ , we wish to show that  $\vDash_{m+1} \subseteq \vDash_{n+1}$  if and only if  $n$  divides  $m$ . If  $m = qn$  then the matrix  $\mathbf{L}_{n+1}$  is isomorphic to a submatrix of  $\mathbf{L}_{m+1}$  via the mapping  $f : x \rightarrow qx$ , so  $\vDash_{m+1} \subseteq \vDash_{n+1}$ . To prove the converse, let us assume that  $m$  is not divisible by  $n$ . Consider the formula  $H_m(H_{n-1}(p) \leftrightarrow p)$ ; see the proof of Lemma 7 for the

definition of  $H_m$ . For any assignment  $g$  in  $\mathbf{L}_{m+1}$ ,  $g(H_{n-1}(p) \leftrightarrow p) > 0$ , for if  $g(H_{n-1}(p) \leftrightarrow p) = 0$  for  $g(p) = k$ ,  $H_{n-1}(k) = g(p) = k$ , so that  $m - k(n - 1) = k$ , implying that  $m = kn$ , contrary to assumption. Thus  $H_m(H_{n-1}(p) \leftrightarrow p)$  is a tautology of  $\mathbf{L}_{m+1}$ . But if we give  $p$  the value 1 in  $\mathbf{L}_{n+1}$ , then  $H_m(H_{n-1}(p) \leftrightarrow p)$  takes the value  $n$ , showing that  $\vDash_{m+1}$  is not contained in  $\vDash_{n+1}$ . ■

### 2.6 Infinite-valued Consequence

The intersection of all the finite consequence relations  $\vDash_m$  forms a consequence relation  $\vDash_\omega$ . In fact, this consequence relation has as a characteristic matrix the infinite-valued matrix of Łukasiewicz defined on the rational numbers in the unit interval, with 1 as designated value (we are returning in this section to the original notation of §1.4). Let  $R$  denote this matrix.

**THEOREM 10.**  *$R$  is a characteristic matrix for  $\vDash_\omega$ .*

**Proof.** Let  $\gamma_1, \dots, \gamma_n$  be rational numbers in the unit interval, where  $\gamma_i = a_i/b_i$ , and let  $k$  be the least common multiple of the  $b_i$ 's. Application of the operations  $\rightarrow$  and  $\neg$  in  $R$  produces only rational numbers expressible in the form  $c/k$ , so that the smallest submatrix of  $R$  containing  $\gamma_1, \dots, \gamma_n$  is finite. It follows that  $\Sigma \vDash_R \theta$  holds if and only if  $\Sigma \vDash_M \theta$  for every finite submatrix  $M$  of  $R$ . But these finite submatrices are (up to isomorphism) exactly the matrices of  $\mathbf{L}_m$ ,  $m \geq 2$ . ■

There is no finite characteristic matrix for  $\vDash_\omega$ , as can be seen by consideration of the formulas  $\delta_k$  used in the proof of Theorem 9. If  $N$  is a finite characteristic matrix with  $k$  truth values, then  $\delta_k$  must be valid in  $N$ , because any formula that has  $(p \leftrightarrow p)$  as a disjunct is a theorem of  $\mathbf{L}_\omega$ . But none of the formulas  $\delta_k$  is valid in  $\vDash_\omega$ .

Axiomatization of the infinite-valued logic  $\mathbf{L}_\omega$  was first accomplished by Wajsberg in 1935, but his proof did not appear in print. The difficulty of the proof may be gauged by the fact that the first published proof by Rose and Rosser [1958] runs to over fifty pages containing many intricate combinatorial lemmas; Rose and Rosser show that all the tautologies of  $\mathbf{L}_\omega$  can be derived from the axiom set conjectured to be complete by Łukasiewicz.

1.  $p \rightarrow (q \rightarrow p)$
2.  $(p \rightarrow q) \rightarrow ((q \rightarrow r) \rightarrow (p \rightarrow r))$
3.  $((p \rightarrow q) \rightarrow q) \rightarrow ((q \rightarrow p) \rightarrow p)$
4.  $(\neg p \rightarrow \neg q) \rightarrow (q \rightarrow p)$
5.  $((p \rightarrow q) \rightarrow (q \rightarrow p)) \rightarrow (q \rightarrow p)$

together with the rules of substitution and *modus ponens*. The fifth axiom was shown to be derivable from the remaining four by Meredith [1958], and independently by Chang [1958b].

We omit the completeness proof for  $\mathbb{L}_\omega$ , as no really simple proof seems available. The reader is referred to the very useful survey of Rosser [1960] for an overview of this subject. Other published proofs of the completeness result use such diverse methods as quantifier elimination in the first order theory of divisible totally-ordered Abelian groups [Chang, 1958a; Chang, 1959], the representation of free Abelian lattice-ordered groups [Cignoli, 1993], and algebraic geometry [Panti, 1995]. An elementary self-contained proof due to Cignoli and Mundici appears in [1997].

### 2.7 Finite Axiomatizability

In the case of the Łukasiewicz logics, the consequence relation generated by a given matrix could be axiomatized by adding a finite number of sequents to the basic consequence rules together with the rule of uniform substitution (the added Dilemma rule can be replaced by a finite set of sequents). If a consequence relation can be formulated in this way, we say that it is finitely axiomatizable. It is natural to ask whether for every finite matrix  $M$ , the consequence relation  $\vDash_M$  is finitely axiomatizable. This question was answered in the negative by Wroński [1976], who gave an example of a 6-valued matrix whose consequence relation is not finitely axiomatizable. Urquhart [1977] gave another example of a five-valued matrix, and finally Wroński [1979] improved Urquhart's example to show that the matrix:

·	0	1	2
0	2	0	2
1	2	2	2
2	2	2	2

with 2 the only designated value has a consequence relation that is not finitely axiomatizable. This result is the best possible, for every two-valued consequence relation can be finitely axiomatized [Rautenberg, 1981].

### 2.8 Definable Functions

Beside the problem of axiomatization just discussed, the most thoroughly investigated in many-valued logic is the question of which functions are definable in a given many-valued matrix.

It is a familiar fact that the two-valued classical matrix is functionally complete; every truth function is definable using only  $\wedge$  and  $\neg$ , or in fact by using only one binary connective, which can be  $(\varphi \mid \psi) = \neg(\varphi \wedge \psi)$ , the

Sheffer stroke or  $(\varphi \downarrow \psi) = \neg(\varphi \vee \psi)$ . Let us define a logical matrix to be functionally complete if every function  $f(\vec{a})$  defined on  $M$  is expressible by a formula  $\varphi(\vec{p})$  definable from the sentential variables and the basic connectives in the matrix.

Before discussing functions definable in various logics, we prove a useful lemma that gives a generalization of the disjunctive normal form used in classical logic. Let  $F$  be a set of functions defined on the set  $M = \{0, \dots, m\}$ —we treat constants as 0-place functions. Let us say that a subset  $N$  of  $M$  is  $F$ -closed if it is closed under the application of functions in  $F$ . The  $F$ -closure of a set  $X \subseteq M$  is defined as the smallest  $F$ -closed set containing  $X$ .

LEMMA 11. *Let  $F$  be a set of functions defined on  $M = \{0, \dots, m\}$  containing  $\max$ ,  $\min$  and all  $J_k$  for  $k \in M$ . Then an  $n$ -place function  $f$  defined on  $M$  is  $F$ -definable if and only if  $f(\vec{x})$  is in the  $F$ -closure of  $\{x_1, \dots, x_n\}$  for all  $\vec{x}$  in  $M$ .*

**Proof.** The condition is obviously necessary. For the converse, assume that  $f$  satisfies the condition and that  $\vec{a}$  is an  $n$ -tuple in  $M$ . The formula  $\psi(\vec{a})$ , defined as  $\max(J_{a_1}(x_1), \dots, J_{a_n}(x_n), f(\vec{a}))$  takes the value  $f(\vec{a})$  for  $\vec{x} = \vec{a}$  and  $m$  for  $\vec{x} \neq \vec{a}$ . It follows that the function  $\min(\psi(\vec{a}), \vec{a}$  an  $n$ -tuple in  $M$ ) coincides with  $f$ , so  $f$  is  $F$ -definable. ■

THEOREM 12. *The  $m + 1$ -valued logic of Post is functionally complete.*

**Proof.** Recall that the basic connectives are  $x \vee y = \min(x, y)$  and  $\neg x = x + 1 \pmod{m + 1}$  defined on  $M = \{0, \dots, m\}$ . We first show that any one-place function is Post-definable. The function

$$T(x) = \min(x, x + 1, \dots, x + m)$$

always takes the value 0 (this is a generalization of the law of excluded middle). It follows that for any  $k \in M$  the function

$$T_k(x) = \min[\min(T(x) + 1, x) + m, x + k + 1] + m$$

takes the value  $m$  for  $x \neq 0$  and  $k$  for  $x = 0$ . Now let  $f(x)$  be any  $m + 1$ -valued function in one variable. Then the function

$$\min(T_{f(0)}(x), T_{f(1)}(x + m), \dots, T_{f(m)}(x + 1))$$

coincides with  $f$ . Thus the  $J_k$  functions are definable and  $\max(x, y)$  is defined as  $h(\min(h(x), h(y)))$ , where  $h(x) = m - x$ . We can now apply Lemma 11. Since all one-place functions are Post-definable, the only non-empty Post-closed subset of  $M$  is  $M$  itself. It follows that Post's connectives form a functionally complete set. ■

The two operators of Post's logic that we have just shown to be sufficient can in fact be replaced by a single operator. Using the set  $\{0, \dots, m-1\}$  as the set of truth values, as before, define  $w(x, y) = \min(x, y) + 1$ , where the addition sign stands for addition modulo  $m$ . Then we can define Post's negation  $\neg x$  as  $w(x, x)$  and  $\min(x, y) = \neg^m w(x, y)$ . By Theorem 12,  $w$  generates all  $m$ -valued functions;  $w$  is said to be a Sheffer function for  $m$ -valued logic. Note that the Sheffer stroke of two-valued logic can be written as  $\min(x, y) + 1$ . A great deal of research effort has gone into the problem of characterizing Sheffer functions in  $m$ -valued logic, culminating in the paper of Rosenberg [1970] that gives a complete characterization of these functions.

## 2.9 Definable Functions in $L_m$

Lukasiewicz's logics  $L_{m+1}$  for  $m > 1$  are not functionally complete. This is easily seen from the fact that for any formula  $\varphi(p_1, \dots, p_n)$  if  $p_1, \dots, p_n$  are given only the classical values 0 and  $m$  then  $\varphi(p_1, \dots, p_n)$  can take only the values 0 or  $m$ . This functional incompleteness can be remedied in the case of  $L_3$  by adding a function  $Tp$  taking the constant value 1. Słupecki [1936] added this function to  $L_3$  and axiomatized the resulting logic by adding the axiom  $Tp \leftrightarrow T\neg p$ . In fact, to make  $L_3$  functionally complete it is sufficient to add *any* function not definable in  $L_3$ . A set of connectives that has this property we shall call *precomplete*, that is, a set  $S$  of  $m$ -valued connectives is precomplete if and only if it is not functionally complete, but the addition of a function to  $S$  that is not already definable in  $S$  results in a functionally complete set.

It can be shown that the connectives in  $L_6$  form a precomplete set, while those in  $L_7$  do not. A complete characterization of those  $m$  for which the  $L_m$  connectives form a precomplete set follows from the next theorem. Let  $\gcd(x, y)$  be the greatest common divisor of  $x$  and  $y$ ; for a finite set  $X = \{x_1, \dots, x_n\}$  let  $\gcd(X)$  be the greatest common divisor of  $x_1, \dots, x_n$ . The key lemma in the characterization of  $L_m$ -definable functions is:

**LEMMA 13.** *Let  $F$  be a set of functions on  $\{0, \dots, m\}$  containing all  $L_{m+1}$ -definable functions. If  $X \subseteq \{0, \dots, m\}$  is  $F$ -closed then  $\gcd(x, y) \in X$  for all  $x, y \in X$ .*

**Proof.** For  $x, y \in X$ , if  $x > y$  then  $x - y \in X$ ; hence by repeated subtraction the remainder of  $x$  on division by  $y$  is also in  $X$ . The Euclidean algorithm for finding  $\gcd(x, y)$  works simply by repeatedly computing remainders. Thus  $\gcd(x, y) \in X$ . ■

**THEOREM 14.** *An  $n$ -place function  $f$  defined on  $M = \{0, \dots, m\}$  is  $L_{m+1}$ -definable if and only if for any  $n$ -tuple  $\vec{a}$ ,  $\gcd(\{a_1, \dots, a_n, m\})$  divides  $f(\vec{a})$ .*

**Proof.** The necessity of the condition follows from the fact that if  $k$  divides  $x$  and  $y$  then it divides  $x \dot{-} y$ .

In view of Lemmas 7 and 11, to prove sufficiency we need only characterize the  $L_{m+1}$ -closed sets. Let  $X_k$  be defined as the set of multiples of  $k$  in  $M$ . We now show that the  $L_{m+1}$ -closed sets are exactly the sets  $X_k$  for  $k$  a divisor of  $m$ . It is easy to see that  $X_k$  is closed for  $k$  a divisor of  $m$ . Conversely, if  $X$  is an  $L_{m+1}$ -closed set, let  $k = \text{gcd}(X)$ . By Lemma 13,  $k \in X$ . Since  $m$  is in every  $L_{m+1}$ -closed set,  $k$  is a divisor of  $m$ , say  $m = qk$ . Now let  $y = pk$  be any multiple of  $k$  in  $M$ . Then  $y = m - (q - p)k$ , hence  $y \in X$ , showing that  $X = X_k$ .

The stated characterization of  $L_{m+1}$ -definable functions now follows easily from the description of the  $L_{m+1}$ -closed sets. ■

Theorem 14 actually gives us more information than we have stated, namely a characterization of all functions definable from a set of functions extending the Lukasiewicz set. Let  $X$  be a subset of the divisors of  $m$ . Define  $X$  to be lcm-closed if (1)  $1 \in X$  and (2) if  $x, y \in X$  then  $\text{lcm}(x, y)$ , the least common multiple of  $x$  and  $y$ , is in  $X$ . Further, for  $Y \subseteq \{0, \dots, m\}$ ,  $Y \neq \emptyset$ , let  $F(Y)$  be defined as the set of functions on  $\{0, \dots, m\}$  that satisfy the condition: for any  $k \in Y, a_1, \dots, a_n \in X_k$  implies  $f(\vec{a}) \in X_k$ . It is easy to see that  $F(Y)$  is closed under function composition.

**COROLLARY 15.** *The sets of  $m + 1$ -valued functions that are closed under composition and extend the  $L_{m+1}$ -definable functions are exactly the sets  $F(Y)$ , where  $Y$  is an lcm-closed subset of  $\{0, \dots, m\}$ .*

**Proof.** By theorem 14, the closed sets of such a set of functions must all be of the form  $X_k$  for  $k$  a divisor of  $m$ . Furthermore, the set  $\{k \mid X_k \text{ an } F\text{-closed set}\}$  is lcm-closed. It remains only to check that every lcm-closed subset of  $M$  arises as the  $F$ -closed sets of some set of functions  $F$ . Accordingly, let  $Y$  be an lcm-closed subset of  $M$ . We wish to show that the sets  $X_k$  for  $k \in Y$  are exactly the  $F(Y)$ -closed sets. Thus let  $X$  be an  $F(Y)$ -closed subset of  $M$ ; by Theorem 14,  $X = X_k$  for some divisor  $k$  of  $m$ . Let  $p = \text{lcm}(\{q \mid X \subseteq X_q, q \in Y\})$ . Then  $X \subseteq X_p$ ; to show  $X = X_p$  it is sufficient to show that  $p \in X$ . Define a function  $f$  by:  $f(k) = p, f(a) = 0$  for  $a \neq k$ . If  $k \in X_q$  for  $q \in Y$  then  $X = X_k \subseteq X_q$  hence  $p \in X_q$  because  $X_p$  is the smallest set  $X_r, r \in Y$ , containing  $X$ . It follows that  $f \in F(Y)$ . Thus  $p \in X$ , completing the proof. ■

**COROLLARY 16.** *The connectives of  $L_{m+1}$  form a precomplete set if and only if  $m$  is prime.*

**Proof.** The only lcm-closed subsets of divisors of  $m$  are  $\{1, m\}$  and  $\{1\}$ , provided  $m$  is a prime. If  $m$  is not prime, there are others. ■



The characterization of the  $L_{m+1}$ -definable functions in Theorem 14 is due to McNaughton [1951]. McNaughton's paper contains as its main result a characterization of the real-valued functions definable in  $L_\omega$ . McNaughton's original proof of this result makes use of a non-constructive *reductio ad absurdum* argument, based on a procedure of unbounded search; for a constructive proof, see [Mundici, 1994].

### 3 CHARACTERIZING FUNCTIONAL COMPLETENESS

A question that has been extensively investigated is the problem of determining whether a given set of  $m$ -valued connectives is functionally complete. One of the most useful theorems gives the Słupecki criterion for functional completeness. Let us first define an  $m + 1$ -valued function to be *essential* if it depends on at least two variables and takes on all values from  $M = \{0, \dots, m\}$ .

**THEOREM 17** ([Słupecki, 1939]). *Let  $m \geq 2$ , and  $F$  a set of functions on  $M$  that contains all one-place  $m+1$ -valued functions. Then  $F$  is functionally complete if and only if  $F$  contains an essential function.*

Let us define a set of  $m$ -valued functions to be *maximal* if it is closed under function composition and is precomplete. It is not hard to see (by a use of Zorn's lemma) that any proper subset of the set of all  $m$ -valued functions defined on  $M$  that is closed under function composition is contained in a maximal set. Thus we have:

**THEOREM 18.** *A set  $F$  of  $m+1$ -valued functions is functionally complete if and only if it is not contained in a maximal set of functions.*

Theorem 18 reduces the problem of characterizing functional completeness to the problem of describing the maximal sets of functions in  $m$ -valued logic. A complete characterization of these sets has been provided by Rosenberg [1965; 1970]. To state this result, we need some definitions. Let  $R$  be an  $n$ -place relation on the set  $M = \{0, \dots, m\}$ . A function of  $k$  variables defined on  $M$  is said to *preserve*  $R$  if  $\vec{a}_1, \dots, \vec{a}_n \in R$  implies that  $\langle f\vec{a}_1, \dots, f\vec{a}_n \rangle \in R$ . For example, let  $R$  be the partial order of  $\{0, 1\}$  defined by  $0 \leq 0, 0 \leq 1, 1 \leq 1$ . Then a function preserving  $R$  is a monotone Boolean function: these functions form a maximal set in two-valued logic. The maximal sets in  $m+1$ -valued logic can be described completely in terms of the preservation of certain relations on  $M$ .

Let  $p$  be a prime number. A group  $G = \langle M, \cdot \rangle$  is called  *$p$ -elementary Abelian* if  $G$  is Abelian and  $px = x \cdot x \cdot \dots \cdot x$  ( $p$  times) is the zero of the group for all  $x \in M$ . It is a well-known fact that  $p$ -elementary groups in  $M$  exist if and only if  $m + 1 = p^q$  for some  $q \geq 1$ .

A  $k$ -place relation  $R$  on  $M$  is *central* if  $R \neq M^k$  and there is a non-empty proper subset  $C$  of  $M$  such that

1.  $a_j \in C \Rightarrow \vec{a} \in R$  for any  $k$ -tuple  $\vec{a}$ ;
2.  $\vec{a} \in R$  and  $\vec{b}$  is a permutation of  $\vec{a} \Rightarrow \vec{b} \in R$ ;
3.  $a_i = a_j \Rightarrow \vec{a} \in R$  for any  $i \neq j$ .

Let  $2 < h \leq m+1$  and let  $q \geq 1$ . We say that the family  $T = \{\Theta_1, \dots, \Theta_q\}$  of equivalence relations on  $M$  is *h-regular* if

1. Each  $\Theta_j$  has  $h$  equivalence classes;
2. The intersection  $\bigcap_{j=1}^q \epsilon_j$  of arbitrary equivalence classes  $\epsilon_j$  of  $\Theta_j$  ( $j = 1, \dots, q$ ) is non-empty.

The relation *determined* by  $T$  is the relation  $\lambda_T$  of all  $h$ -tuples  $\vec{a}$  in  $M^h$  having the property that for each  $1 \leq j \leq q$  at least two elements among  $a_1, \dots, a_h$  are equivalent on  $\Theta_j$ .

Let  $\text{Pres}(R)$  stand for the functions preserving a relation defined on  $M = \{0, \dots, m\}$ . We are now ready to state the beautiful and deep characterization theorem of Rosenberg.

**THEOREM 19.** *Every maximal class of functions in  $m + 1$ -valued logic is of the form  $\text{Pres}(R)$  where  $R$  is one of the following types of relation on  $M$ :*

1. A partial order on  $M$  with least and greatest element;
2. A relation  $\{\langle x, sx \rangle \mid x \in M\}$  where  $s$  is a permutation of  $M$  with  $(m + 1)/p$  cycles of the same prime length  $p$ ;
3. A four-place relation of the form  $\{\langle a_1, a_2, a_3, a_4 \rangle \in M^4 \mid a_1 \cdot a_2 = a_3 \cdot a_4\}$  where  $\langle M, \cdot \rangle$  is a  $p$ -elementary Abelian group.
4. A non-trivial equivalence relation  $R$  on  $M$  ( $R \neq M^2$ ,  $R$  is not the identity on  $M$ );
5. A central relation on  $M$ ;
6. A relation  $\lambda_T$  determined by an  $h$ -regular family  $T$  of equivalence relations on  $M$ .

*A set of connectives in  $m + 1$ -valued logic is functionally complete if and only if for every relation  $R$  described under (1)–(6) there is an  $f$  in the set not preserving  $R$ .*

The reader is referred to the excellent survey article [Rosenberg, 1984] for more details and further references on the problem of definability of functions.

### 3.1 Post Algebras

Classical propositional logic can be cast in algebraic form as the theory of Boolean algebras. A similar transformation on Post's many-valued systems produces the theory of Post algebras.

To motivate the algebraic developments that follow, we introduce the notion of  $m$ -valued set. A set in the ordinary sense can be considered as a map from a collection of individuals into the set of classical truth-values; given a fixed universal collection  $U$ , any subset  $X \subseteq U$  is uniquely determined by its characteristic function  $f$  defined as:  $f(a) = T$  if  $a \in X$ ,  $f(a) = F$  if  $a \notin X$ . Generalizing to the  $m$ -valued case, we define an  $m$ -valued set  $X$  defined on a collection  $U$  to be a map from  $U$  into the truth-values  $\{0, \dots, m-1\}$ . The equation  $X(a) = k$  is to be read as: ' $a$  is an element of  $X$ ' has truth-value  $k$ . The classical operations of intersection and union generalize in an obvious way to the family of all  $m$ -valued sets on  $U$ , which we shall write as  $P(U, m)$ . For  $X, Y \in P(U, m)$ , define:

$$\begin{aligned} X \wedge Y(a) &= \min[X(a), Y(a)] \\ X \vee Y(a) &= \max[X(a), Y(a)]. \end{aligned}$$

Obviously, these operations are the algebraic counterpart of conjunction and disjunction.  $P(U, m)$  forms a distributive lattice with respect to  $\wedge$  and  $\vee$ ; if we define  $m$ -valued set containment by:

$$X \leq Y \text{ if and only if } X(a) \leq Y(a) \text{ for all } a \in U,$$

then  $X \wedge Y$  is the greatest lower bound,  $X \vee Y$  the least upper bound of  $X$  and  $Y$ , and  $X \wedge (Y \vee Z) = (X \wedge Y) \vee (X \wedge Z)$ .

With respect to  $m$ -valued set containment,  $P(U, m)$  has a greatest element  $\vee$  and a least element  $\wedge$ ;  $\vee$  is the constant function with value  $m-1$ ,  $\wedge$  the constant function with value 0. More generally, let  $C_k$  be the constant function defined on  $U$  with value  $k$ ; then  $P(U, m)$  contains a chain  $\wedge = C_0 \leq C_1 \leq \dots \leq C_{m-1} = \vee$ . (Note that here  $C_0$  is the 'falsest',  $C_{m-1}$  the 'truest' truth-value, reversing our previous ordering in which 0 is the 'truest' value.)

An important role in  $P(U, m)$  is played by the  $m$ -valued sets that are classical or two-valued in the sense that  $X(a)$  is either 0 or  $m-1$ . Algebraically, the classical sets are exactly the *complemented* elements of  $P(U, m)$ , the elements for which there is a (unique) complement  $\bar{x}$  such that  $x \wedge \bar{x} = \wedge$ ,  $x \vee \bar{x} = \vee$ . The centre of a lattice,  $C(L)$  is defined as the set of complemented elements of  $L$ . Any  $m$ -valued set defined on  $U$  can be expressed as a combination of constant functions and classical sets. Let  $X$  be any element of  $P(U, m)$ . For  $k$  any truth-value, let  $D_k(X)$  be the two-valued set on  $U$  that has the value  $m-1$  at  $a$  if  $X(a) \geq k$ , and the value 0 at  $a$  if  $X(a) < k$ . It is easy to see that:

$$X = (D_1(X) \wedge C_1) \vee \dots \vee (D_{m-1}(X) \wedge C_{m-1}).$$

What we have shown, in fact, is an algebraic version of the disjunctive normal form theorem provided by Lemma 11.

After this preliminary discussion, the following abstract definition of a Post algebra should be easy to understand.

**DEFINITION 20** (Traczyk [1963]). A Post algebra of order  $m + 1$  is a distributive lattice  $L$  with greatest element  $\vee$  and least element  $\wedge$  that satisfies the conditions:

1.  $L$  has a subchain  $\wedge = c_0 \leq c_1 \leq \dots \leq c_m = \vee$  such that every element  $a \in L$  can be written as  $a = (a_1 \wedge c_1) \vee \dots \vee (a_m \wedge c_m)$ , where each  $a_i \in C(L)$ ,
2. If  $a \in C(L)$  and  $(a \wedge c_i) \leq c_{i-1}$  for some  $i, i > 0$ , then  $a = \wedge$ .

It follows immediately from (2) that the  $c_i$ 's are all distinct.

The join representation of an element as postulated in (1) is not unique in general; a unique representation can be obtained by adding an extra condition.

**THEOREM 21** (Epstein [1960]). *If  $L$  is a Post algebra of order  $m + 1$ ,  $m > 0$ , then every element  $a \in L$  has a unique representation*

$$a = (a_1 \wedge c_1) \vee \dots \vee (a_m \wedge c_m) \text{ where } a_i \in C(L), \\ a_1 \geq a_2 \geq \dots \geq a_m.$$

**Proof.** Let  $a = (a_1 \wedge c_1) \vee \dots \vee (a_m \wedge c_m)$ . For  $1 \leq i \leq m$ , define  $b_i = a_i \vee a_{i+1} \vee \dots \vee a_m$ . Then  $a = (b_1 \wedge c_1) \vee \dots \vee (b_m \wedge c_m)$  is a representation of the required sort.

To show uniqueness, suppose  $a = (d_1 \wedge c_1) \vee \dots \vee (d_m \wedge c_m)$ , and  $d_1 \geq d_2 \geq \dots \geq d_m$ . Then for  $1 \leq k \leq m, b_k \wedge c_k \leq a$ , so

$$b_k \wedge \overline{d_k} \wedge c_k \leq \overline{d_k} \wedge c_k \wedge a \\ = \overline{d_k} \wedge c_k \wedge [(d_1 \wedge c_1) \vee \dots \vee (d_m \wedge c_m)] \\ = [(d_1 \wedge c_1 \wedge \overline{d_k} \wedge c_k) \vee \dots \vee (d_m \wedge c_m \wedge \overline{d_k} \wedge c_k)].$$

Since  $\overline{d_1} \leq \overline{d_2} \leq \dots \leq \overline{d_m}$ ,  $(d_j \wedge c_j \wedge \overline{d_k} \wedge c_k) = \wedge$  for  $j \geq k$ . It follows that  $b_k \wedge \overline{d_k} \wedge c_k \leq c_{k-1}$ . But then  $b_k \wedge \overline{d_k} = \wedge$ , by Definition 20, so  $b_k \leq d_k$ . Symmetrically,  $d_k \leq b_k$  so  $b_k = d_k$ . ■

Theorem 21 allows us to introduce operations in any Post algebra that correspond to the unique representation of an element. For any Post algebra  $L$  of order  $m + 1, a \in L$ , let  $D_k(a)$ , for  $1 \leq k \leq m$ , be the unique element  $a_k \in L$  in the representation of the theorem. These new operations can be used undefined in an equational formulation of Post algebras:

DEFINITION 22 (Traczyk [1964]). A Post algebra  $L$  of order  $m+1$  ( $m > 0$ ) can be defined as an algebra with two-place operations  $\wedge, \vee$ , one-place operations  $\neg, D_1, D_2, \dots, D_m$  and constants  $c_0, \dots, c_m$  that satisfies the conditions:

1.  $L$  is a distributive lattice with respect to  $\wedge, \vee$ , with least element  $c_0$  and greatest element  $c_m$ ;
2.  $L$  forms a de Morgan algebra with  $\neg$ , that is,  $\neg(x \wedge y) = \neg x \vee \neg y$ , and  $\neg\neg x = x$ ;
3.  $c_i \wedge c_j = c_i$  for  $i \leq j$ ;
4.  $D_i(x \vee y) = D_i(x) \vee D_i(y)$  and  $D_i(x \wedge y) = D_i(x) \wedge D_i(y)$ ;
5.  $D_i(x) \vee \neg D_i(x) = c_m$ ,  $D_i(x) \wedge \neg D_i(x) = c_0$ ;
6.  $D_i(x) \wedge D_j(x) = D_j(x)$  for  $i \leq j$ ;
7.  $D_i(\neg x) = \neg D_{m-i}(x)$ ;
8.  $D_i(c_j) = c_m$  for  $i \leq j$ ;  $D_i(c_j) = c_0$  for  $j < i$ ;
9.  $x = (D_1(x) \wedge c_1) \vee \dots \vee (D_m(x) \wedge c_m)$ .

To show that this definition is equivalent to Definition 20, we first observe that in any Post algebra we can introduce an operation  $\neg a$  as follows: if  $a = (a_1 \wedge c_1) \vee \dots \vee (a_n \wedge c_n)$ , where  $a_1 \geq \dots \geq a_n$ , define  $\neg a$  to be  $(\overline{a_n} \wedge c_1) \vee \dots \vee (\overline{a_1} \wedge c_n)$ . Then it is easy to check that the above postulates are satisfied.

For the converse, first note that  $c_0 \leq \dots \leq c_m$ , where we define:  $x \leq y$  if and only if  $x \wedge y = x$ . Furthermore, (5) implies that  $D_i(x)$  is in  $C(L)$ . A representation of the type given in Theorem 21 is guaranteed by (6) and (9). It remains only to check that (2) of Definition 20 is satisfied. Accordingly, assume that  $a \wedge c_i \leq c_{i-1}$ , where  $a \in C(L)$ . Then by (4)  $D_i(a) \wedge D_i(c_i) \leq D_i(c_{i-1})$ ; but by (8),  $D_i(c_i) = c_m$  and  $D_i(c_{i-1}) = c_0$ , so  $D_i(a) = c_0$ , hence  $D_m(a) = c_0$ . Now let  $\overline{a}$  be the complement of  $a$  in  $C(L)$ . By (4) and (8),  $D_m(a) \vee D_m(\overline{a}) = D_m(a \vee \overline{a}) = c_m$ . Thus  $D_m(\overline{a}) = c_m$ . But then by (9),  $D_m(\overline{a}) \leq \overline{a}$ , so  $\overline{a} = c_m$ , hence  $a = c_0$ .

It is well known that every Boolean algebra is isomorphic to a field of sets, that is, a collection of subsets of a set  $U$  that is closed with respect to intersection, union and complementation with respect to  $U$ . A similar theorem holds for Post algebras of order  $m+1$ . Let us define a concrete Post algebra of order  $m+1$  to consist of a family of  $m+1$ -valued sets defined on a collection  $U$ , together with operations  $\wedge, \vee, D_k$  and the constant functions  $c_k$  as defined earlier together with the operation  $\neg X$  defined as:  $\neg X(a) = m - X(a)$ .

**THEOREM 23** (Wade [1945]). *Every Post algebra  $L$  of order  $m + 1$  is isomorphic to a concrete Post algebra of  $m + 1$ -valued sets.*

**Proof.** We have seen in Theorem 21 that every element of  $L$  corresponds uniquely to an  $m$ -tuple of elements in  $C(L)$  that satisfy  $a_1 \geq \dots \geq a_m$ . Since  $C(L)$  is a Boolean algebra, it is isomorphic to a field of sets. Let  $\varphi(a)$  for  $a \in C(L)$  be the corresponding subset of  $U$  in the field of sets to which  $C(L)$  is isomorphic. Now for an element  $a$  of  $L$ , define an  $m + 1$ -valued function on  $U$  by:

$$X_a(z) = \max\{k \mid z \in \varphi(D_k(a))\}.$$

We wish to show that the map  $\psi : a \rightarrow X_a$  is an isomorphism with respect to the operations of  $L$ . For any  $a, b \in L$ ,  $a \leq b$  if and only if  $D_k(a) \leq D_k(b)$  for all  $1 \leq k \leq m$ ; this follows readily from Theorem 21. It is easy to see from this that  $a \leq b$  if and only if  $X_a \leq X_b$ . It is a straightforward exercise to check that  $\psi$  is an isomorphism with respect to the operations  $D_k$ , and that  $\psi(c_k)$  is the constant function  $C_k$ . ■

### 3.2 Generalized Post Algebras and Algorithmic Logic

Various generalizations of Post algebras have been considered in the literature of which the most widely investigated have been Post algebras of order  $\omega^+$ .

The simplest example of a Post algebra of order  $\omega^+$  is the matrix  $P_\omega$  defined on the ordinals  $0, 1, 2, \dots, \omega$ , where these are ordered as a chain  $0 < 1 < 2 < \dots < \omega$ ,  $x \vee y = \max(x, y)$ ,  $x \wedge y = \min(x, y)$ ; we have in addition the operations  $D_k$  for  $k$  a positive integer, where  $D_k(x) = \omega$  if  $k \leq x$ ,  $D_k(x) = 0$  otherwise. An implication operation and negation operator  $\neg$  are defined by:  $x \Rightarrow y = \omega$  if  $x \leq y$ ,  $x \Rightarrow y = y$  if  $x > y$ ,  $\neg x = \omega$  if  $x = 0$ ,  $\neg x = 0$  if  $x \neq 0$ .  $P_\omega$  plays the same role in  $\omega^+$ -valued Post algebras as the  $m$ -valued Post matrices play in Post algebras of order  $m$ . Post algebras of order  $\omega^+$  can be axiomatized by a set of equations similar to the identities of Definition 22.

The motivation for considering these generalized Post algebras arises from the theory of programming languages. It is possible to add to a classical propositional language containing predicates and individual terms operators that represent composition, branching and iteration operations on programmes. Using these operations, it is possible to express programmes and expressions representing their properties in a compact language (for details, see [Salwicki, 1970]). The extension of this idea to many-valued logic is convenient in cases where a programme has a branching structure. For example, there may be an instruction in the programme of the form: do one of  $P_0 \dots P_n$  according to whether conditions  $A_0 \dots$  or  $A_n$  are realised. In this situation, we can use the generalized truth-values  $c_0, \dots, c_n$  as devices that

keep track of which of  $A_0, \dots, A_n$  are true, so that a convenient language for such programmes is the language of  $\omega^+$ -valued logic, or Post algebras of order  $\omega^+$ . This application of many-valued logic to computer languages is due to H. Rasiowa. The reader is referred to her survey article [Rasiowa, 1977] for more details and references.

The theory of Post algebras arises from the case of an  $m$ -valued logic that is functionally complete. Other many-valued logics with connectives that are not functionally complete have been given algebraic form, such as the many-valued logics of Łukasiewicz. The  $MV$ -algebras of Chang [1958a] constitute an algebraic version of the infinite-valued logic of Łukasiewicz, while the finite-valued Łukasiewicz logics have been given algebraic form as  $MV_n$ -algebras [Grigolia, 1977]. For background on these algebraic structures, the reader is referred to the monograph [Gottwald, 1989].

### 3.3 Many-valued Predicate Logic

Many-valued logics can be extended to include quantification in a straightforward way. The basic idea here (as in classical logic) is that universal quantification can be treated as extended conjunction, existential quantification as extended disjunction.

To make ideas definite, let us suppose that we are dealing with the  $m$ -valued logic of Łukasiewicz. We extend the language of the propositional calculus  $L_m$  to include individual variables, predicates of any degree and the quantifiers  $\forall$  and  $\exists$ . Let us call the resulting language  $L$ . Given a non-empty set  $I$  of individuals, we add to  $L$  a stock of individual constants  $a, b, c, \dots$ , etc. uniquely correlated with individuals  $a, b, c, \dots$ , in  $I$ ; call the resulting language  $L(I)$ . Then an  $m$ -valued structure for  $L$  over the set of individuals  $I$  is defined to be an assignment of a value in  $\{0, \dots, m-1\}$  to the atomic sentences in  $L(I)$  that contain no free variables. Thus if  $P$  is an  $n$ -place predicate in  $L$ , we assign a value  $\llbracket Pa_1 \dots a_n \rrbracket$  to every sentence  $Pa_1 \dots a_n$  for  $a_1 \dots a_n$  in  $I$ . We extend this assignment of values to sentences in  $L(I)$  that contain connectives by the earlier truth-tables. For quantified sentences in  $L(I)$ , we have the definitions:

$$\begin{aligned}\llbracket \forall x \varphi \rrbracket &= \max\{\llbracket \varphi[a/x] \rrbracket \mid a \in I\}, \\ \llbracket \exists x \varphi \rrbracket &= \min\{\llbracket \varphi[a/x] \rrbracket \mid a \in I\}.\end{aligned}$$

Comparison of these rules for quantifiers with the conjunction and disjunction rules of Section 1.2 shows that if  $I$  is a finite set  $\{a_1, \dots, a_n\}$ , then  $\forall x \varphi$  is equivalent to  $\varphi[a_1/x] \wedge \dots \wedge \varphi[a_n/x]$  and  $\exists x \varphi$  is equivalent to  $\varphi[a_1/x] \vee \dots \vee \varphi[a_n/x]$ .

If we take 0 as the sole designated value, as before, then the consequence relation of the quantified version of  $L$  can be axiomatized. To carry out this axiomatization, we add to the language  $L$  an infinite number of new

constants  $a_0, a_1, a_2, \dots$  that we shall call parameters. The quantified version of  $L_m, L_m \mathbf{QC}$ , is axiomatized by adding to the rules of Section 2.5 the following scheme for sequents:

$$(F) \quad J_k(\varphi[a/x]) \vdash J_k(\forall x\varphi) \vee J_{k+1}(\forall x\varphi) \vee \dots \vee J_{m-1}(\forall x\varphi),$$

where  $a$  is any parameter, together with the rule:

$$\frac{\Gamma, \varphi \rightarrow J_k(\psi[a/x]) \vdash \Delta}{\Gamma, \varphi \rightarrow J_k(\forall x\psi) \vdash \Delta} (G),$$

where  $a$  is a parameter that does not occur in the conclusion of the rule. The existential quantifier we are taking as defined by:

$$\exists x\varphi = \neg\forall x\neg\varphi.$$

Let  $\vdash$  be a consequence relation in a language  $L$  and  $\vdash'$  a consequence relation in a language  $L'$  that is an extension of  $L$ . We say that  $\vdash'$  is a conservative extension of  $\vdash$  if exactly the same sequents of  $L$  are provable in  $\vdash$  and  $\vdash'$ . That is to say, if  $\Gamma \cup \{\varphi\}$  is a set of formulas in  $L$ , then  $\Gamma \vdash \varphi$  if and only if  $\Gamma \vdash' \varphi$ .

Let us denote the consequence relation of  $L_m \mathbf{QC}$  by  $\vdash_{Q_m}$ . Let  $L'$  be the language that results from the language of  $L_m \mathbf{QC}$  by adding an infinite set of new constants  $c_0, c_1, c_2, \dots$ . We now define  $\vdash'_{Q_m}$  to be the smallest consequence relation in the language  $L'$  that contains  $\vdash_{Q_m}$ , together with all sequents of the form:

$$\vdash J_k(\forall x\varphi) \rightarrow J_k(\varphi[c/x]),$$

where  $\varphi$  is a formula of  $L'$ , and  $c$  is a new constant that does not occur in  $\varphi$ .

LEMMA 24.  $\vdash'_{Q_m}$  is a conservative extension of  $\vdash_{Q_m}$ .

**Proof.** Let  $\Gamma \vdash'_{Q_m} \delta$  hold, where  $\Gamma \cup \{\delta\}$  is a set of formulas in the language of  $L_m \mathbf{QC}$ . The proof of  $\Gamma \vdash'_{Q_m} \delta$  can involve only finitely many sequents of the form

$$\vdash J_k(\forall x\varphi) \rightarrow J_k(\varphi[c/x]).$$

Let  $\Theta$  stand for the set of sentences of  $L'$  that occur in these sequents. Now perform the following transformation on the proof of  $\Gamma \vdash'_{Q_m} \delta$ ; (1) replace each sequent  $\Sigma \vdash \psi$  by  $\Sigma, \Theta \vdash \psi$ , (2) replace each new constant  $c$  throughout the resulting set of sequents by a new parameter  $a$  that does not occur anywhere in the proof. We can now prove by induction on the length of the proof that for every sequent  $\Sigma \vdash \psi$  in the old proof, the new



sequent  $\Sigma', \Theta' \vdash \psi'$  that results from the replacement process is provable in  $L_m \mathbf{QC}$ . In particular the sequent

$$\Gamma, \Theta' \vdash \delta$$

is provable in  $L_m \mathbf{QC}$ , where  $\Theta'$  is a set of formulas of the form

$$J_k(\forall x\varphi) \rightarrow J_k(\varphi[a/x]),$$

where  $a$  does not occur in  $\Gamma, \delta$  or  $\varphi$ . Since  $\vdash_{Q_m} J_k(\forall x\varphi) \rightarrow J_k(\forall x\varphi)$  it follows by the rule (G) that  $\Gamma \vdash_{Q_m} \delta$ .  $\blacksquare$

Before proving  $L_m \mathbf{QC}$  complete, we need a precise definition of validity. Let  $\Gamma \vdash \delta$  be a sequent of  $L_m \mathbf{QC}$  with no free variables. Then  $\Gamma \vdash \delta$  is  $m$ -valid,  $\Gamma \vDash_m \delta$ , if and only if for every  $m$ -valued structure over a set  $I$   $\llbracket \varphi \rrbracket = 0$  for all  $\varphi \in \Gamma$  implies that  $\llbracket \delta \rrbracket = 0$ —we are assuming that the definition of structure is extended to include an interpretation in  $I$  for all the parameters in the language of  $L_m \mathbf{QC}$ . Sequents  $\Gamma \vdash \delta$  that contain free variables are defined to be  $m$ -valid if and only if the sequents that result from them by replacing all free variables by parameters are  $m$ -valid.

**THEOREM 25.**  $\Gamma \vdash_{Q_m} \delta$  if and only if  $\Gamma \vDash_m \delta$ .

**Proof.** That all sequents provable in  $L_m \mathbf{QC}$  are  $m$ -valid follows by a straightforward induction on the length of proofs. The only non-trivial step concerns rule (G). Suppose that the sequent

$$\Gamma, \varphi \rightarrow J_k(\forall x\psi) \vdash \delta$$

is not  $m$ -valid. Then we can find an interpretation  $\llbracket \cdot \rrbracket$  in a set of individuals  $I$  so that  $\llbracket \theta \rrbracket = 0$  for  $\theta \in \Gamma$ ,  $\llbracket \varphi \rightarrow J_k(\forall x\psi) \rrbracket = 0$ , but  $\llbracket \psi \rrbracket \neq 0$ . Let  $b$  be an element of  $I$  such that  $\llbracket \forall x\psi \rrbracket = \llbracket \psi[b/x] \rrbracket$ —such a  $b$  must exist by the truth definition for  $\forall x\psi$ . Now define a new interpretation  $\llbracket \cdot \rrbracket'$  that  $\llbracket \forall x\psi \rrbracket' = \llbracket \psi[a/x] \rrbracket'$ . Then  $\llbracket \theta \rrbracket = 0$ , for  $\theta \in \Gamma$ ,  $\llbracket \varphi \rightarrow J_k(\psi[a/x]) \rrbracket = 0$ , and  $\llbracket \delta \rrbracket' = \llbracket \delta \rrbracket$ . Thus the premiss of an application of rule (G) is also not  $m$ -valid.

For completeness, let  $\Gamma \vdash \delta$  be a sequent with no free variables that is not provable in  $L_m \mathbf{QC}$ . Then  $\Gamma \vdash \delta$  is also not in the consequence relation  $\vdash'$ , by Lemma 24. Let  $\vdash''$  be a  $\Gamma$ -prime consequence relation containing  $\vdash'$  that does not contain  $\Gamma \vdash \delta$ —existence of such a consequence relation is guaranteed by Lemma 5. Now take as domain of individuals for an  $m$ -valued structure the set  $C$  of all the new constants  $c_0, c_1, \dots$ , together with the set of parameters. For  $P$  an  $n$ -place predicate, define:

$$\llbracket Pd_1, \dots, d_n \rrbracket = k \Leftrightarrow \Gamma \vdash'' J_k(Pd_1, \dots, d_n),$$

where  $d_1, \dots, d_n$  are constants or parameters. We can now prove inductively that for all sentences  $\varphi$  in the language of  $\vdash'$ ,  $\llbracket \varphi \rrbracket = k$  if and only if  $\Gamma \vdash''$

$J_k(\varphi)$ . The inductive steps for the propositional connectives are exactly as in Theorem 8. For the quantifier, it is sufficient to show that for any  $\varphi, \Gamma \vdash'' J_k(\forall x\varphi)$  if and only if

$$\max\{p \mid \exists c(\Gamma \vdash'' J_p(\varphi[c/x]))\} = k.$$

First, if  $\Gamma \vdash'' J_k(\forall x\varphi)$ , since  $\vdash'' J_k(\forall x\varphi) \rightarrow J_k(\varphi[c/x])$  for some  $c$ , it follows that  $\Gamma \vdash'' J_k(\varphi[c/x])$ . Now if  $\Gamma \vdash'' J_p(\varphi[d/x])$  for some  $d$ , with  $p > k$ , then by (F), we must have  $\Gamma \vdash'' J_q(\forall x\varphi)$  for some  $q \geq p$ , a contradiction. It follows that

$$\max\{p \mid \exists c(\Gamma \vdash'' J_p(\varphi[c/x]))\} = k.$$

Second, if  $\Gamma \not\vdash'' J_k(\forall x\varphi)$ , then  $\Gamma \vdash'' J_q(\forall x\varphi)$  for some  $q \neq k$ , so that

$$\max\{p \mid \exists c(\Gamma \vdash'' J_p(\varphi[c/x]))\} = q,$$

by the argument just given, so the condition is sufficient as well as necessary. We have thus proved that in the interpretation just defined,  $\llbracket \gamma \rrbracket = 0$  for all  $\gamma \in \Gamma$ , and  $\llbracket \delta \rrbracket \neq 0$ , so that  $\Gamma \vdash \delta$  is not  $m$ -valid. ■

Further results on the axiomatization of many-valued predicate logic, including systems with generalized quantifiers, can be found in the monograph of Rosser and Turquette [1952].

The axiomatization of the quantified versions of infinite-valued logics presents more difficult problems. If we take as our set of truth-values the real numbers between 0 and 1, with the only designated value 0, we have a system  $L_R\mathbf{QC}$  of infinite-valued predicate logic with the Łukasiewicz connectives. Bruno Scarpellini [1962] has shown that the valid formulas of  $L_R\mathbf{QC}$  cannot be recursively axiomatized. Louise Hay [1963] axiomatized this logic by making use of an infinitary rule of inference.

### 3.4 Set Theory in Many-valued Logic

The idea of avoiding the paradoxes of set theory by altering the underlying logic rather than the comprehension axiom is an old one, as we saw in Section 1.4. In this section we consider some of the research that has been done since Bochvar's early work on that topic.

The system for which this question has been most fully investigated are the Łukasiewicz systems. Here we immediately notice that the finite-valued systems are non-starters. We consider a version of  $m$ -valued predicate logic that contains only the binary predicate of  $\in$  of membership, together with a predicate  $=$  for identity. We add to the valid formulas of  $L_m\mathbf{QC}$  in this language the comprehension axiom scheme:

$$\text{(COM): } \forall \vec{x} \exists y \forall z (z \in y \leftrightarrow \varphi(z, \vec{x})).$$

We shall show that in the resulting system any formula whatever can be derived. Define:  $(\varphi^0 \rightarrow \psi) = \psi$ ,  $(\varphi^{n+1} \rightarrow \psi) = (\varphi \rightarrow (\varphi^n \rightarrow \psi))$ . Now by substituting  $(z \in z)^m \rightarrow \psi$  for  $\varphi(z, \vec{x})$  in (COM), we have for some  $c$  that  $c \in c \leftrightarrow ((c \in c)^m \rightarrow \psi)$ , hence  $(c \in c)^{m+1} \rightarrow \psi$ . But in  $\mathbf{L}_m\mathbf{QC}$ ,  $(\varphi^{m+1} \rightarrow \psi) \rightarrow (\varphi^m \rightarrow \psi)$  is valid, so that by modus ponens, we have  $(c \in c)^m \rightarrow \psi$ . But then  $c \in c$  follows, hence  $\psi$  by  $m$  applications of modus ponens. (This version of Russell's paradox, using only implication, is due to Curry.)

The crucial step in the preceding proof, the step from  $(\varphi^{m+1} \rightarrow \psi)$  to  $(\varphi^m \rightarrow \psi)$ , is invalid in infinite-valued Lukasiewicz logic, a fact that encourages a hope that the result of adding (COM) to the infinite-valued predicate logic  $\mathbf{L}_m\mathbf{QC}$  is consistent. The earliest result along these lines was that of Skolem [1957], who proved that the system in which (COM) is restricted to formulas  $\varphi(x, \vec{z})$  with no free variables is consistent. This result was improved by Chang [1963] who extended the class of formulas admissible on the right-hand side of the comprehension scheme to those in which every bound variable is restricted to occur only in the second place in atomic formulas of the form  $v \in w$ . Chang also proved consistent the form of (COM) in which the right-hand formula has no restrictions on its bound variables, but contains no free variables except  $z$ . Another result of a similar type was obtained by Fenstad [1964] who proved the consistency of the scheme where the free variable  $z$  is allowed to occur only in the first place in atomic formulas of the form  $v \in w$ . All of these proofs of consistency use the Brouwer fixed point theorem. Richard B. White [1979] finally solved the consistency problem for the full system by a proof-theoretical argument. Starting from the axiom system of Hay mentioned above, he used normalization techniques to show the full comprehension axiom consistent in infinite-valued logic.

Some more negative results are forthcoming if we consider adding other axioms of set theory beside comprehension. For example, let (EXT) be the sentence  $\forall x \forall y (\forall z (z \in x \leftrightarrow z \in y) \rightarrow x = y)$ . Then the system resulting from the addition of (EXT) and (COM) to  $\mathbf{L}_\omega\mathbf{QC}$  is inconsistent, even if we restrict ourselves to the instance of (COM) allowed by the first result of Chang mentioned above.

The extent to which mathematics can be developed in many-valued logic remains a largely open question; the reader is referred to White's paper [1979] for a discussion. There does not seem to be a clear natural interpretation for set theory in many-valued logic that would play a role similar to that of the cumulative type hierarchy in classical set theory.

## 4 DEVELOPMENTS SINCE 1960

After the initial period of development in the 1920s and early 1930s, many-valued logic remained a kind of intellectual backwater. However, the area has seen a recent revival of interest, in large part due to the efforts of engineers and computer scientists. In this part we sketch three recent developments. First, an interpretation of the Łukasiewicz systems that links these logics with other areas of non-classical logic; second, the contemporary activity in the area of ‘fuzzy logic’; lastly, work on the logic of significance.

4.1 *Model Structures on Commutative Monoids*

Although the Łukasiewicz systems of many-valued logic were the first non-classical systems to be investigated in depth, the work in this area has remained somewhat apart from the mainstream of work in non-classical logic. In particular, the semantical methods involving relational model structures, ‘possible worlds’ and the like, that have proved so fruitful in areas like modal logic, seem to have no clear connection with traditional many-valued logic. In this section, we show how these ideas can be brought to bear on Łukasiewicz systems, and thereby uncover a somewhat unexpected connection with relevance logic.

We begin by describing a model theory that initially will seem to have little or no connection with many-valued logic. We are concerned with a propositional language containing  $\wedge, \vee$ , and  $\rightarrow$  as primitive symbols. Let  $\mathfrak{A} = \langle A, +, 0, \leq \rangle$  be a (totally) ordered commutative monoid, i.e. an algebra equipped with an associative, commutative operation with 0 an element satisfying  $x + 0 = x$  for all  $x \in A$ , and  $\leq$  a total ordering of  $A$  that satisfies the condition that  $x \leq y$  implies  $x + z \leq y + z$  for all  $x, y, z$ . Let  $A_0$  be the set  $\{x \in A : x \geq 0\}$ . A model structure over  $\mathfrak{A}$  consists of a subset  $\llbracket P \rrbracket \subseteq A_0$  for each propositional variable  $P$  in the language. The sets  $\llbracket P \rrbracket$  are required to be *increasing* (a subset  $B \subseteq A$  is increasing if  $x \in B, x \leq y$  imply that  $y \in B$ ). Given a model over  $\mathfrak{A}$ , we define truth at a point  $x$  in  $A_0$  by induction:

1.  $x \vDash P \Leftrightarrow x \in \llbracket P \rrbracket$
2.  $x \vDash (\varphi \wedge \psi) \Leftrightarrow (x \vDash \varphi \ \& \ x \vDash \psi)$
3.  $x \vDash (\varphi \vee \psi) \Leftrightarrow (x \vDash \varphi \vee x \vDash \psi)$
4.  $x \vDash (\varphi \rightarrow \psi) \Leftrightarrow \forall y \in A_0 (y \vDash \varphi \Rightarrow x + y \vDash \psi)$ .

Given this definition, we can prove by induction on the complexity of  $\varphi$  that the points in  $A_0$  at which a formula  $\varphi$  is true form an increasing set.

A striking feature of the definition of truth just given is its close resemblance to the truth definition for a model theory for relevance logic due to

Routley and the present author (see the Chapter by Bäuerle and Cresswell in a later volume of this *Handbook* for details). The differences between that model theory and the present are two-fold: first, in the case of the relevance logic **R**, the underlying structure is assumed to be a semilattice, so that  $x + x = x$  is generally valid; second, in relevance logic the subsets assigned to variables do not have to be increasing. Apart from these two differences, the truth definition, for  $\wedge, \vee$  and  $\rightarrow$  is word for word the same.

A formula  $\varphi$  is said to be valid in this model theory if  $0 \vDash \varphi$  in any model structure over an ordered commutative monoid. The reader can easily verify that the following axiom schemes are all valid:

1.  $\varphi \rightarrow (\psi \rightarrow (\varphi \wedge \psi))$
2.  $(\varphi \rightarrow \psi) \rightarrow ((\theta \rightarrow \varphi) \rightarrow (\theta \rightarrow \psi))$
3.  $(\varphi \rightarrow (\theta \rightarrow \psi)) \rightarrow (\theta \rightarrow (\varphi \rightarrow \psi))$
4.  $(\varphi \wedge \psi) \rightarrow \varphi$
5.  $(\varphi \wedge \psi) \rightarrow \psi$
6.  $((\varphi \rightarrow \psi) \wedge (\varphi \rightarrow \theta)) \rightarrow (\varphi \rightarrow (\psi \wedge \theta))$
7.  $\varphi \rightarrow (\varphi \vee \psi)$
8.  $\psi \rightarrow (\varphi \vee \psi)$
9.  $((\varphi^k \rightarrow \psi) \wedge (\theta^k \rightarrow \psi)) \rightarrow ((\varphi \vee \theta)^k \rightarrow \psi)$
10.  $(\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$ .

Let us call the system consisting of these axioms together with the rule of *modus ponens* **C**. There is a close resemblance between **C** and certain systems of relevance logic. The differences lie in (1), which reflects the hereditary condition, in (10) which reflects the total ordering condition (although it is valid in the logic **RM**) and in the absence of Contraction  $(\varphi \rightarrow (\varphi \rightarrow \psi)) \rightarrow (\varphi \rightarrow \psi)$ , which corresponds to our omission of the equation  $x + x = x$ .

We postpone until a later section a discussion of the intuitive interpretation to be attached to this model theory. We conclude the section with a completeness proof for **C**.

LEMMA 26. *The following theorem schemes are provable in C:*

- (a)  $\varphi \rightarrow \varphi$
- (b)  $(\varphi \rightarrow \psi) \rightarrow ((\varphi \vee \psi) \rightarrow \psi)$
- (c)  $(\varphi \rightarrow \sigma \vee \psi) \rightarrow ((\varphi \rightarrow \sigma) \vee (\varphi \rightarrow \psi))$ .

**Proof.** To prove (a), we substitute a theorem for  $\psi$  in Axiom (1), then use Axioms (3) and (4). Schema (b) follows immediately from (a), (1), (3) and (9). Finally, for (c), from (b) we have by Axiom (2):

$$(\sigma \rightarrow \psi) \rightarrow ((\varphi \rightarrow (\sigma \vee \psi)) \rightarrow (\varphi \rightarrow \psi)),$$

hence by Axioms (2) and (8),

$$(\sigma \rightarrow \psi) \rightarrow [(\varphi \rightarrow (\sigma \vee \psi)) \rightarrow ((\varphi \rightarrow \sigma) \vee (\varphi \rightarrow \psi))].$$

Similarly,

$$(\psi \rightarrow \sigma) \rightarrow [(\varphi \rightarrow (\sigma \vee \psi)) \rightarrow ((\varphi \rightarrow \sigma) \vee (\varphi \rightarrow \psi))],$$

so by Axioms (6) and (10),

$$(\varphi \rightarrow \sigma \vee \psi) \rightarrow ((\varphi \rightarrow \sigma) \vee (\varphi \rightarrow \psi)).$$

■

We define a  $\mathbf{C}$  theory to be a set of formulas that contains all theorems of  $\mathbf{C}$  and is closed under *modus ponens*. If  $\Delta$  is a  $\mathbf{C}$  theory,  $\Delta(\varphi)$  is defined to be the smallest  $\mathbf{C}$  theory containing  $\Delta \cup \{\varphi\}$ .

LEMMA 27. *If  $\psi \in \Delta(\varphi)$ , then for some  $k$ ,  $\varphi^k \rightarrow \psi \in \Delta$ .*

**Proof.** By induction on the length of the proof of  $\psi$  in  $\Delta(\varphi)$ . For  $\theta$  an axiom of  $\mathbf{C}$   $\varphi^0 \rightarrow \theta \in \Delta$ ; by Lemma 26,  $\varphi \rightarrow \varphi$  is provable in  $\mathbf{C}$ . For the induction step, assume that  $\varphi^k \rightarrow \theta$  and  $\varphi^m \rightarrow (\theta \rightarrow \psi)$  are in  $\Delta$  and that  $\psi$  is derived from  $\theta \rightarrow \psi$  and  $\theta$  in  $\Delta(\varphi)$ . By repeated use of Axiom (2), we have  $(\theta \rightarrow \psi) \rightarrow ((\varphi^k \rightarrow \theta) \rightarrow (\varphi^k \rightarrow \psi))$  as a theorem of  $\mathbf{C}$  so by Axiom (3), we have  $(\theta \rightarrow \psi) \rightarrow (\varphi^k \rightarrow \psi)$  in  $\Delta$ . By using (2) again, we have  $\varphi^{k+m} \rightarrow \psi$  in  $\Delta$ . ■

THEOREM 28.  $\vdash_{\mathbf{BC}} \varphi$  if and only if  $\varphi$  is valid in all model structures over an ordered commutative monoid.

**Proof.** Let  $\varphi$  be a formula not provable in  $\mathbf{C}$ ; then there is a  $\mathbf{C}$  theory that we shall call  $0$  that is maximal in the family of  $\mathbf{C}$  theories not containing  $\varphi$ . Now let  $A$  be the family of sets of formulas  $\Sigma$  that satisfy the conditions: (a)  $0 \subseteq \Sigma$ ; (b)  $\varphi \rightarrow \psi \in 0, \varphi \in \Sigma \Rightarrow \psi \in \Sigma$ ; (c)  $\Sigma$  is prime, that is  $(\pi \vee \psi) \in \Sigma \Rightarrow \pi \in \Sigma$  or  $\psi \in \Sigma$ ; (d)  $\theta, \pi \in \Sigma \Rightarrow \theta \wedge \pi \in \Sigma$ . We have to show that  $0$  itself is prime. Suppose  $(\pi \vee \psi) \in 0, \pi \notin 0, \psi \notin 0$ . By maximality,  $\varphi \in 0(\pi)$  and  $\varphi \in 0(\psi)$ , so by Lemma 27,  $\pi^k \rightarrow \varphi, \psi^m \rightarrow \varphi \in 0$  for some  $k, m$ . By (1), we can assume  $k = m$ . Now by Axiom (9),  $(\pi \vee \psi)^k \rightarrow \varphi \in 0$  so  $\varphi \in 0$ , a contradiction.

On  $A$ , we define the operation  $+$  by :

$$\Delta + \Sigma = \{\psi \mid \exists \sigma(\sigma \rightarrow \psi \in \Delta \wedge \sigma \in \Sigma)\}.$$

We have to show that  $A$  with the defined  $+$  operation is a commutative monoid. First, though, we have to show that  $A$  is closed under  $+$ . It is straightforward to check that conditions (a), (b) and (d) are satisfied by  $\Delta + \Sigma$ , by using the appropriate theorems of  $\mathbf{C}$ . Finally, we note that by Lemma 26 (c),  $\Delta + \Sigma$  is prime. Commutativity of  $+$  follows from Axiom (3), associativity from Axiom (2). Axiom (10) implies that  $A$  is totally ordered by containment:  $\Delta \subseteq \Sigma$  implies  $\Delta + \Gamma \subseteq \Sigma + \Gamma$  is trivially true. That  $\Delta + 0 = \Delta$  follows by definition and Lemma 26 (a).

We now let  $\llbracket P \rrbracket = \{\Delta \in A \mid P \in \Delta\}$ . We have to verify that in the resulting model,  $\Delta \vDash \varphi$  if and only if  $\varphi \in \Delta$  for any  $\Delta$ . The induction steps for  $\wedge$  and  $\vee$  are straightforward. Assuming the claim for  $\theta, \psi$ , we prove it for  $\theta \rightarrow \psi$ . If  $\theta \rightarrow \psi \in \Delta$ , then if  $\theta \in \Sigma, \psi \in \Delta + \Sigma$  by definition, so  $\Delta \vdash \theta \rightarrow \psi$  holds. Conversely, if  $\Delta \vDash \theta \rightarrow \psi$ , let  $\Sigma = \{\sigma \mid \theta \rightarrow \sigma \in 0\}$ . By (1),  $\Sigma$  satisfies (a), by (2), it satisfies (b), and  $\Sigma$  is prime by Lemma 26 (c). The condition (d) follows from (6). Thus  $\Sigma \in A$ , so  $\Delta + \Sigma \vDash \psi$  since  $\Sigma \vDash \theta$ . Thus  $\psi \in \Delta + \Sigma$  by inductive hypothesis. This means that for some  $\pi, \pi \rightarrow \psi \in \Delta, \pi \in \Sigma$  so  $\theta \rightarrow \pi \in 0$ . But then  $(\pi \rightarrow \psi) \rightarrow (\theta \rightarrow \psi) \in 0$  so that  $\theta \rightarrow \psi \in \Delta$ . This completes the induction. Since  $\varphi \notin 0, 0 \not\vDash \varphi$  in this model, so the proof of completeness is finished. ■

#### 4.2 Model Structures for $L_\omega$

The logic  $\mathbf{C}$  we have just discussed bears a strong resemblance to  $L_\omega$ , but does not coincide with it. For a model theory adequate to  $L_\omega$ , we have to add several features to our earlier model structures.

For  $L_\omega$ , we postulate that the underlying algebra is not simply an ordered commutative monoid, but an ordered Abelian group. That is, we add to the postulates of Section 4.1 the additional requirement that for each element  $x$  in  $A$  there is an element  $-x \in A$  such that  $x + (-x) = 0$ . As usual, we abbreviate  $x + (-y)$  as  $x - y$ ; note that  $x \leq y$  if and only if  $y - x \geq 0$ . It may be noted here that Chang [1959] proved the completeness of the Łukasiewicz axioms for the infinite-valued calculus by interpreting formulas as universal sentences on totally ordered Abelian groups. Mundici [1986] gives further developments of Chang's ideas, featuring the equivalence between MV-algebras and Abelian lattice-ordered groups with strong units.

To accommodate negation, we add to our earlier language a constant  $\perp$  to denote the constant false proposition, defining  $\neg A$  as  $A \rightarrow \perp$ . We now define a model structure over an Abelian group as an assignment of subsets  $\llbracket P \rrbracket \subseteq A$  to propositional variables and also  $\llbracket \perp \rrbracket$ , so that  $\llbracket P \rrbracket$  and  $\llbracket \perp \rrbracket$  are non-empty subsets that have a least element, and  $\llbracket \perp \rrbracket \subseteq \llbracket P \rrbracket$  for any  $P$ .

**THEOREM 29.**  $\models_{\omega} \varphi$  if and only if  $\varphi$  is valid in all model structures over an ordered Abelian group.

**Proof.** We need to verify that all the axioms of  $L_{\omega}$  (see Section 2.6) are valid in any such model structure. These are all easy to check, with the possible exception of

$$((\varphi \rightarrow \psi) \rightarrow \psi) \rightarrow ((\psi \rightarrow \varphi) \rightarrow \varphi).$$

We shall prove this by showing that  $(\varphi \rightarrow \psi) \rightarrow \psi$  and  $\varphi \vee \psi$  take the same value at any point  $x$  in a model structure. If  $x \models \varphi \vee \psi$  then  $x \models (\varphi \rightarrow \psi) \rightarrow \psi$  follows immediately. Assume  $x \models (\varphi \rightarrow \psi) \rightarrow \psi$ , but not  $x \models \varphi$  or  $x \models \psi$ . It follows that the least point  $z$  at which  $\varphi$  is true is smaller than the least point  $y$  at which  $\psi$  is true. Now consider the point  $y - z$ . For any  $u$ , if  $u \models \varphi$  then  $u \geq z$ , so  $(y - z) + u \geq y$  hence  $(y - z) + u \models \psi$ . Thus  $y - z \models \varphi \rightarrow \psi$ . It follows that  $(y - z) + x \models \psi$ . But since  $x \not\models \varphi$ ,  $x < z$ , so  $(y - z) + x < y$ , which is a contradiction.

For completeness, recall that an invalid formula of  $L_{\omega}$  is refutable in a finite matrix for  $L_m$ . Suppose  $\varphi$  is invalid in such a matrix. Consider the ordered Abelian group  $I$  of the integers. For a propositional variable  $P$ , let  $g(P)$  be the value assigned to  $P$  in  $L_{m+1}$ ; set  $\llbracket P \rrbracket = \{x \in I \mid x \geq g(P)\}$ . Furthermore, set  $\llbracket \perp \rrbracket = \{x \in I \mid x \geq m\}$ . We can now prove by induction that for any formula  $\varphi$ , integer  $x$ ,  $0 \leq x \leq m$ ,  $x \models \varphi$  if and only if  $x \geq g(\varphi)$ . The easy inductive proof is left to the reader. Now since  $g(\varphi) \neq 0$ ,  $\varphi$  is invalid in the model structure over the integers. ■

It should be noted that the condition that for each formula there is a least point at which it is true is essential. Consider the ordered Abelian group of the reals, and set  $\llbracket P \rrbracket = (1, \infty)$ ,  $\llbracket q \rrbracket = [2, \infty)$ . Then  $x \models p \rightarrow q$  if  $x \geq 1$ , so that  $1 \models (p \rightarrow q) \rightarrow q$ . But  $1 \not\models p \vee q$ , so that not all theorems of  $L_{\omega}$  are validated.

It is possible to give a direct completeness proof for  $L_{\omega}$  by using the model theory just expounded. The reader is referred to Scott's paper [1974] for details.

### 4.3 Do Errors Add Up?

We have not yet discussed the intuitive meaning of the model theory of the previous section, due independently to Scott [1974] and Urquhart [1973].

Urquhart [1973] attempts a tense logical interpretation of the model theory, which seems plausible because of Łukasiewicz's concern with future contingents. However, if we think of the points in a model as moments of time, and the statements as 'coming to be true' at moments in time, the truth condition for implication (see Section 4.1) just does not seem to make sense.



A more plausible reading is suggested in Scott [1974] who reads the points in a model as representing *degrees of error*. Thus 0 represents no error at all (truth), while the higher points in a model represent greater degrees. With this interpretation, the semantical condition for implication takes on an interesting aspect. It says that when we apply *modus ponens* the degrees of error add up. That is, if  $\varphi \rightarrow \psi$  is true with degree of error  $k$ ,  $\varphi$  with degree of error  $m$ , then  $\psi$  can be inferred with degree of error  $k + m$ . Scott suggests that this idea might have applications in areas where we are thinking of something like approximate equality. Suppose  $P$  is a space of point with a real-valued metric on it where  $|a - b|$  is the distance between  $a$  and  $b$ . We can adopt the convention that a degree of error  $\leq i$  can be tolerated, so that a statement  $\varphi$  is designated if  $i \vDash \varphi$ , where ' $x \vDash \varphi$ ' is to be read as 'the degree of error of  $\varphi$  is  $\leq x$ '. Now given this choice of designated values, the sequent

$$a = b, b = c \vDash a = c$$

is not valid—approximate equality is not transitive. However, the conditional assertion

$$a = b \vDash b = c \rightarrow a = c$$

is valid, in fact is just a restatement of the triangle inequality.

These kinds of examples have a certain persuasiveness, but difficulties remain, as Smiley [1976] points out in his comments on Scott [1976]. The difficulty is the same as the one we pointed out in Section 1.3. We want to say that 'if  $A$ , then  $A$ ' is true with degree of error zero. But we *don't* want to accept 'if  $A$ , then  $B$ ' for any  $B$  that happens to have the same degree of error as  $A$ . Similar problems arise with conjunction and negation. The statement ' $A$  and not  $A$ ' ought to be completely false—but it isn't, if it has an intermediate value of error. In fact, it isn't even clear how negation can make sense in most contexts involving error. Suppose we are making guesses as to the position of (say) an enemy ship on a map. Then we can measure the degree of error of 'not  $A$ ' where  $A$  represents a guess of a position? A little thought is all that is needed to see that it is just not a proposition of the right type for degree of error to be defined.

Although there are difficulties with the "degrees of error" interpretation of Łukasiewicz's logic, Mundici [1992] has suggested a very interesting application of the logic in reasoning under uncertainty. The application has its origins in a question of Stanislaw Ulam:

Someone thinks of a number between one and one million (which is just less than  $2^{20}$ ). Another person is allowed to ask up to twenty questions, to each of which the first person is supposed to answer only yes or no. Obviously the number can be guessed by asking first: Is the number in the first half million? then again reduce the reservoir of numbers in the next question by one-half, and so on. Finally the number is obtained in less than  $\log_2(1,000,000)$ . Now suppose one were allowed to lie once or twice, then how many questions would one need to get the right answer? One clearly needs more than  $n$  questions for guessing one of  $2^n$  objects because one does not know when the lie was told. [1976, p. 281]

Mundici [1992] has shown that Ulam's guessing game with lies can be analysed in a natural way using many-valued logic. In the game with no lies, the second player's current state of knowledge can be represented by a function from the set of numbers  $S$  to the Boolean values  $\{0, 1\}$ ; a number has the value 1 just in case it has not been ruled out by any of the questions so far. Similarly, in the case of Ulam's guessing game where  $k$  lies are permitted, the second player's current state of knowledge is represented by a function  $F : S \rightarrow \{0, 1/(k+1), \dots, k/(k+1), 1\}$ , where, for each  $y \in S$ , the quantity  $1 - F(y)$  is just the number  $n$  of answers currently falsified by  $y$ , divided by  $k+1$ . In this representation, the Łukasiewicz truth-value is the distance, measured in units of  $k+1$ , from the condition of falsifying too many answers in Ulam's game with  $k$  lies (since only  $k$  lies are permitted, all numbers of errors exceeding  $k$  are represented by the value  $1 = k+1/k+1$ ). If we represent states of knowledge in arbitrary Ulam games by the variables  $s, t, \dots$ , then the equation  $s \wedge s = s$  is valid just in the games where  $k = 0$ . Thus failure of contraction in Łukasiewicz logics is given a natural meaning in the context of this interpretation; in contrast to the classical case of no lies ( $k = 0$ ), asking the same question twice provides us with useful information. Mundici [1992; 1993] shows that equations valid in all Ulam games with an arbitrary number of lies coincide with tautologies in the infinite-valued calculus of Łukasiewicz (where we rewrite  $s \rightarrow t$  as  $\neg(s \wedge \neg t)$ ).

#### 4.4 Fuzzy Logic

One of the reasons for the recent revival of interest in many-valued logic is the growth of research in the area of 'fuzzy sets' and 'fuzzy logic'. Since its inception in the mid 1960s this subject has seen an explosive growth and there are now hundreds of papers in the area, numerous volumes of conference proceedings and a journal entirely devoted to fuzzy matters. Here we can only touch on the field as it relates to logical questions.

The growth of the field is largely owing to the enthusiastic advocacy of L. A. Zadeh [1965], who introduced the concept of a 'fuzzy set'. Given a collection  $X$  of elements, a fuzzy set on  $X$  is characterized by a membership function  $f(x)$  that associates with each point in  $X$  a real number in the interval  $[0, 1]$ . The nearer the value of  $f(x)$  to unity, the higher the grade of membership in the set. When the range of  $f$  is restricted to  $\{0, 1\}$ , we have the characteristic function of an ordinary classical set.

The intention behind the definition of a fuzzy set is to represent predicates in ordinary discourse that are vague or lacking in precisely defined criteria for membership. For example, the class of numbers much greater than 1 is certainly rather vague. Zadeh suggests that it can be modelled by a fuzzy set defined on the real line with a characteristic function satisfying  $f(0) = 0, f(1) = 0, f(5) = 0.01, f(100) = 0.95, f(500) = 1$ , and so forth.

Complement, containment, union and intersection Zadeh defines as fol-

lows. If  $f$  is the function associated with a fuzzy set on  $X$ , then the complement of the set has the characteristic function  $1 - f(x)$ . Containment is defined by  $f \leq g$ , i.e.  $f(x) \leq g(x)$  for all  $x \in X$ . Union and intersection correspond to  $\max(f, g)$  and  $\min(f, g)$  respectively.

Zadeh proposes to use these ideas to model vague predicates such as ‘beautiful’, ‘tall’, ‘long’, and so forth. A philosophical application is suggested by Goguen [1969] who attempts a solution of the classical paradox of the bald man (or paradox of the heap). The paradox runs as follows. We are inclined to admit the truth of the two following sentences:

1. A man with 20,000 hairs on his head is not bald,
2. If you remove one hair from a man who is not bald, then he remains not bald.

However by applying (2) 20,000 times along with *modus ponens*, we derive the absurd conclusion that a man with zero hairs on his head is not bald. Goguen suggests that we think of ‘bald’ as a fuzzy predicate. Then if we attach to the implication (2) a truth value slightly less than 1 and adopt Lukasiewicz’s implication, we find that although we may start by assigning the truth value 1 to (1), each successive application of *modus ponens* lowers the truth value of the sentence ‘A man with 20,000 –  $x$  hairs is bald’. In this way, Goguen argues, fuzzy logic avoids the paradox of the heap.

One immediate objection that presents itself to this line of approach is the extremely artificial nature of the attaching of precise numerical values to sentences like ‘73 is a large number’ or ‘Picasso’s *Guernica* is beautiful’. In fact, it seems plausible to say that the nature of vague predicates precludes attaching precise numerical values just as much as it precludes attaching precise classical truth values. The artificiality of Zadeh’s and Goguen’s approach emerges readily from a little reflection on the numerical example given above.

Zadeh is of course aware of the artificial nature of his procedure, and in later publications he introduces the idea of ‘fuzzy truth values’. A fuzzy truth value (such as ‘true’, ‘very true’, ‘not so true’) is a fuzzy subset of the real line. The assignment of truth values to sentences now takes a form such as ‘The compatibility of the numerical truth value 0.8 with the linguistic truth value “true” is 0.7’. It seems, however, that the ‘fuzzification’ of truth values (see Bellman and Zadeh [1977]) has only pushed the original problem one stage back, as we still have numerically precise values for compatibility with the fuzzy truth values.

Another problem that arises with fuzzy logic is a difficulty very similar to the difficulties involved in interpreting Lukasiewicz’s logics. How are we to interpret the operations on fuzzy sets? If we interpret intersection and union as the algebraic equivalent of conjunction and disjunction, and complementation as negation, then things don’t seem to work out right. Suppose we

are dealing with a problem in pattern recognition (a case frequently discussed in the fuzzy literature). Then a given object  $x$  may be a triangle (say) to degree 0.9;  $f_{\Delta}(x) = 0.9$ . If the complement of  $f_{\Delta}$  represents ‘is not a triangle’ and union disjunction, then  $\max(f_{\Delta}, 1 - f_{\Delta})$  should represent ‘is a triangle or isn’t a triangle’ and should be the constant 1 function; but it isn’t. So operations on fuzzy sets don’t correspond to ordinary logical connectives; but it’s difficult to make out what they *are* supposed to represent. It can be seen that the root difficulty here is identical with the difficulties in Łukasiewicz’s systems.

The idea of fuzzy set suggests that the notion of a rule of inference can itself be “fuzzified”; following up the original suggestion of Goguen [1969] (see the remarks above about the paradox of the heap) Jan Pavelka [1979] has investigated this idea in some detail in a series of papers. Pavelka defines a many-valued rule of inference from a fuzzy set of formulas by specifying a rule of inference in the ordinary sense, together with a rule for computing the value of the conclusion from the values of the premisses. Pavelka proves both axiomatizability and non-axiomatizability results for the case where the underlying algebra is a chain with a residuation operation.

Fuzzy logic has become a fairly active sub-discipline of software engineering with its own voluminous literature. For a sample of articles in the area, the reader can consult [Baldwin, 1996]. The topics in such articles do not have much to do with logic in the classical sense of a canon of inference. The collection [Höhle and Klement, 1995] is a useful compendium of work on the algebraic and logical foundations of fuzzy set theory; a selection of Zadeh’s own papers is available in [Zadeh, 1996].

#### 4.5 *The Logic of Significance*

We have seen how Bochvar extended the classical truth values by adding a third truth value to be read as ‘meaningless’.

His ideas were extended by Halldén [1949] who formalized a version of three-valued sentential logic containing a one-place operator  $S\varphi$  in addition to the usual propositional connectives. ‘ $S\varphi$ ’ is to be read as ‘ $\varphi$  is a significant proposition’ and takes the value  $T$  if  $\varphi$  takes the value  $T$  or  $F$ , otherwise the value  $F$ . Taking Bochvar’s tables for the propositional connectives as basic, and  $T$  as the only designated value, we now have valid formulas like:

$$S(\varphi \wedge \psi) \leftrightarrow S\varphi \wedge S\psi.$$

Halldén’s ideas have been greatly extended in the work of Goddard and Routley [1973]. They include not only syntactical and semantical analyses of sentential significance logics, but also formalizations of quantified and higher order significance logics. Their book is the fullest and richest discussion in the literature of the logic of significance, and its role in the history of logic

and philosophy.

## 5 RETROSPECTIVE

### 5.1 *What is Many-valued Logic?*

So far we have avoided the question posed above by taking a more or less historical approach. Many-valued logic from this point of view consists simply in the systems developed by Łukasiewicz, Post, Bochvar and Kleene, or systems closely related to these.

Nevertheless, this characterization leaves something to be desired. It would be better if we could give a more analytical account of what many-valued logic is, and what distinguishes it from classical logic. A definition that is explicitly or implicitly adopted by many authors is that a many-valued logic is involved whenever we assign to formulas in a logical system values in an algebra that is not the two element Boolean algebra (i.e. classical truth tables). By this definition, Boolean-valued models for set theory count as many-valued logic (see for example, the introductory remarks in [Mostowski, 1979, Lecture V]). But this broad usage of the phrase ‘many-valued logic’ does not have a great deal to commend it. How broad it is emerges clearly from the fact that by Theorem 3, any uniform structural consequence relation can be considered as a many-valued logic.

In order to get a more sensible idea of what many-valued logic is, let’s return to the ideas of the pioneers to see if we can extract some common core to their assumptions.

According to the interpretation defended here what is characteristic of a many-valued logic is not so much the formal apparatus of multiple truth values as the relationship of the formalism of multiple truth values to the intuitive interpretation. The key ideas that the systems of Łukasiewicz, Bochvar and Kleene have in common are:

1. To the classical truth values is added one or more extra truth values with meanings like ‘possible’, ‘meaningless’ or ‘undetermined’. These truth values are usually considered as linearly ordered.
2. The rules for assigning values to complex formulas satisfy a generalized rule of truth functionality; the value assigned to a complex formula is a function of the values assigned to its components.

Since we are not dealing with a given intuitive interpretation of the formalism, but several, we shall begin by considering Łukasiewicz’s interpretation and related ideas.

## 5.2 *The Logic of Uncertainty*

Lukasiewicz wished to use his logic to describe situations involving uncertainty and the ‘open future’. As I have emphasised in the discussion of Section 1, this idea seems to be definitely incorrect, provided that the connectives of Lukasiewicz’s logic are to be read as corresponding to the ordinary language connectives of conjunction, disjunction, negation and implication. The logic of uncertainty is simply not truth-functional.

In fact, we can establish something stronger than the previous weak negative claim. Arguments of Ramsey, de Finetti and Savage (for details, see [Jeffrey, 1965]) establish that subjective probability values must obey the rules of the probability calculus. More precisely, if we assume that probability values represent betting quotients, then the rules of the probability calculus for subjective probability emerge automatically as consistency conditions for a rational agent. Now the probability calculus is not truth-functional; the probability value of a conjunction is not a function of the values of its conjuncts, because the conjuncts may or may not be stochastically independent. These simple considerations suffice to throw considerable doubt on Lukasiewicz’s ideas.

Quite similar remarks apply to interpretations that consider the multiple values as representing degrees of error, or degrees of precision or vagueness. It would seem that where one is attempting to formalize concepts of uncertainty, vagueness and so forth, the rules of the probability calculus provide a much more attractive model than the framework of ideas provided by many-valued logic. I do not wish, however, to classify probability theory as many-valued logic because it violates the basic truth-functionality principle (2). It is a curious fact that in spite of his polemics against classical logic Lukasiewicz came to grief by clinging to the classical principles of truth-functionality.

## 5.3 *‘Undefined’ as a Truth Value*

The polemics in the previous section against any interpretation of many-valued logic relating to subjective uncertainty leaves unscathed the applications where an intermediate value stands for ‘undefined’ or ‘meaningless’.

As we have noted, the idea of using a many-valued logic as a foundation for set theory is currently in doubt. However, there remains Kleene’s strong truth tables and their application in recursive function theory and other areas where a formula fails to be assigned a truth value for practical or theoretical reasons.

Here as in the previous case, the crucial questions turn around the general principle of truth-functionality. It is possible to offer a plausible argument that in fact, the logic of ‘undefined’ is not truth-functional. Suppose that the sentence  $\varphi$  fails to have a truth-value; evidently  $\neg\varphi$  must also fail to

have a value. But  $\varphi \vee \neg\varphi$  should be true, not undefined (contrary to Kleene's tables). On the other hand, if  $\varphi$  and  $\psi$  are logically independent propositions with no truth value, then  $\varphi \vee \psi$  may certainly be undefined. We shall describe briefly in the next section an approach alternative to many-valued logic that takes the foregoing ideas as basic.

#### 5.4 *Supervaluations*

The method of supervaluations was introduced in [van Fraassen, 1966] as a means of providing semantics for free logic, that is, a logic containing non-denoting terms. Let  $L$  be a language for classical predicate logic containing predicate letters, individual variables, quantifiers and classical connectives. To give a specific interpretation  $I$  for  $L$ , we specify a non-empty domain of discourse  $D$  and an extension in  $D$  for the predicates in  $L$ . Furthermore, for each individual constant or name  $a$  in  $L$ , we either assign  $a$  a denotation in  $D$ , or leave the denotation of  $a$  undefined. For atomic sentences  $Pa$ ,  $Pa$  is true under the interpretation if  $a$  has a denotation  $d(a)$  in  $D$  and  $d(a)$  is in the extension of  $P$ , false if  $d(a)$  is defined and not in the extension of  $P$ ; otherwise  $Pa$  has no truth value. Thus we are considering an interpretation in which 'truth value gaps' exist. How do we define truth and falsity in this interpretation? One method might be to take 'undefined' as a third truth value, then use Kleene's three-valued logic. Van Fraassen, however, wishes to hew to the classically valid arguments. Accordingly, he considers all possible extensions of the truth value assignments to atomic formula in  $I$ . Let us call any such extension that fills in the truth value gaps in  $I$  arbitrarily a classical extension of  $I$ . In any such classical extension we can assess the truth value of any formula  $\varphi$  by the classical definition, bearing in mind that  $\forall x\psi(x)$  is true in such an extension if and only if  $\psi(d)$  is true for any  $d \in D$ . Then a *supervaluation* over  $I$  is a function that assigns  $T(F)$  exactly to those statements assigned  $T(F)$  by all the classical extensions of  $I$ . Let us say that a formula  $\varphi$  is SL-valid if it is true in all supervaluations.

The definition of supervaluation has the attractive feature that it makes SL-valid exactly the theorems of classical 'free logic'. For details of the completeness proof and further work on supervaluation and free logic, the reader is referred to Bencivenga's contribution to this Handbook (Chapter 4.5). Here our main concern is with supervaluations as an alternative to many-valued logic.

One of the most striking features of the supervaluation approach is that it makes SL-valid the law of excluded middle, but not the law of bivalence. For example, if the name 'Bruce Wayne' fails to denote in an interpretation  $I$ , then 'Bruce Wayne reads Proust' fails to have a truth value in the supervaluation over  $I$ . This at first sight seems counter-intuitive, since we have a true disjunction, with neither disjunct true. Our surprise is lessened if we reflect on what it means to say that a formula containing non-denoting

terms is true under a supervaluation. To say that  $\varphi \vee \neg\varphi$  is true is to say that it *would* be true if we pretend that the non-denoting terms in it have a reference. Such a formula is true only in an ‘as if’ kind of way, not a full-blooded classical way. It is important to note that if we regard ‘undefined’ as a third truth value (a move that Van Fraassen strenuously resists) then a supervaluation is not truth functional. For example, if ‘Dick Grayson’ is another non-denoting term, then ‘Dick Grayson reads Baudelaire’ has no truth value; but it is clear that ‘Bruce Wayne reads Proust or Dick Grayson reads Baudelaire’ has no truth value, in contrast to any instance of the law of excluded middle.

For many purposes, the theory of supervaluations seems superior to the older approaches involving many-valued logic. The supervaluation approach has the very real advantage that it allows us all of classical (free) logic, while admitting the possibility of truth value gaps. The method of supervaluations accomplishes these seemingly incompatible objectives precisely by abandoning the principle of truth functionality, which as we have seen is the basic source of difficulty in interpreting many-valued logic. It does so admittedly at the price of abandoning such classical principles as the idea that a disjunction is true only if one of its disjuncts is true. According to Quine [1953] such an abandonment is a ‘desperate extremity’. Those who like Quine wish to stick to this principle come what may should perhaps find Kleene’s truth table more congenial.

### 5.5 *Summing Up*

In a survey of mathematical logic and logical positivism, Zbigniew Jordan gave the following remarkable assessment of many-valued logic:

Without any doubt it is a discovery of the first order, eclipsing everything done in the field of logical research in Poland ([Jordan, 1945] in [McCall, 1967, p. 389]).

This passage is the more striking if one reflects that among the results said to be eclipsed are: Chwistek’s simple type theory, Kuratowski’s work on the projective hierarchy, Jaśkowski’s work on natural deduction, Lindenbaum’s results and above all, Tarski’s fundamental work on methodology, definability and the theory of truth.

It is difficult to agree with this estimate of Łukasiewicz’s work. While Tarski’s ideas have proved their fruitfulness in virtually every area of modern logic, Łukasiewicz’s many-valued systems have remained somewhat marginal to the mainstream of logical research. It seems likely that this marginal position is related to the difficulties in finding a natural interpretation of the extra truth-values discussed above.



Although many-valued logics have not fulfilled the revolutionary role that Lukasiewicz and others hoped for them, interest in these systems is currently increasing, and they will doubtless continue to inspire interesting work in both mathematics and philosophy.

*Further reading.* [Malinowski, 1993], [Rine, 1984], [Wójcicki and Malinowski, 1977].

*University of Toronto, Canada.*

## BIBLIOGRAPHY

- [Baldwin, 1996] J.F. Baldwin, editor. *Fuzzy Logic*. Wiley, 1996.
- [Bellman and Zadeh, 1977] R. E. Bellman and L. A. Zadeh. Local and fuzzy logics. In *Modern Uses of Multiple-valued Logic*. D. Reidel, Dordrecht and Boston, 1977.
- [Belnap, 1977] N. D. Belnap. A useful four-valued logic. In *Modern Uses of Multiple-valued Logic*. D. Reidel, Dordrecht and Boston, 1977.
- [Bochvar, 1939] D. A. Bochvar. Ob odnom trézhnacnom isčislenii i égo priménénii k analiza paradoksov klassičekogo rässirénnoho funkcional'nogo isčisléniiá. (On a 3-valued logical calculus and its application to the analysis of contradictions.). *Matématiceskij sbornik*, 4:287–308, 1939.
- [Chang, 1958a] C. C. Chang. Algebraic analysis of many valued logics. *Transactions of the American Mathematical Society*, 88:467–490, 1958.
- [Chang, 1958b] C. C. Chang. Proof of an axiom of Lukasiewicz. *Transactions of the American Mathematical Society*, 87:55–56, 1958.
- [Chang, 1959] C. C. Chang. A new proof of the completeness of the Lukasiewicz axioms. *Transactions of the American Mathematical Society*, 93:74–80, 1959.
- [Chang, 1963] C. C. Chang. The axiom of comprehension in infinite valued logic. *Mathematica Scandinavica*, 13:9–30, 1963.
- [Church, 1939] A. Church. Review of Bochvar [Bochvar, 1939]. *Journal of Symbolic Logic*, 4:98–99, 1939.
- [Cignoli and Mundici, 1997] R. Cignoli and D. Mundici. An elementary proof of Chang's completeness theorem for the infinite-valued calculus of Lukasiewicz. *Studia Logica*, 58:79–97, 1997.
- [Cignoli, 1993] R. Cignoli. Free lattice-ordered Abelian groups and varieties of mv-algebras. In *Ninth Latin American symposium on mathematical logic*, pages 113–118. Universidad Nacional del Sur, 1993. Notas de Matemática Vol. 38.
- [Epstein, 1960] G. Epstein. The lattice theory of Post algebras. *Transactions of the American Mathematical Society*, 95:300–317, 1960.
- [Fenstad, 1964] J. E. Fenstad. On the consistency of the axiom of comprehension in the Lukasiewicz infinite-valued logic. *Mathematica Scandinavica*, 14:65–74, 1964.
- [Goddard and Routley, 1973] L. Goddard and R. Routley. *The Logic of Significance and Context*. Scottish Academic Press, Edinburgh and London, 1973.
- [Goguen, 1969] J. A. Goguen. The logic of inexact concepts. *Synthese*, 19:325–373, 1969.
- [Gonseth, 1941] F. Gonseth, editor. *Les entretiens de Zurich sur les fondements et la méthode des sciences mathématiques 6-9 Décembre 1938*. Leemann, Zurich, 1941.
- [Gottwald, 1989] S. Gottwald. *Mehrwertige Logik*. Akademie-Verlag, Berlin, 1989. Expanded English edition: *A Treatise on Many-valued Logics*, RSP, 2000.
- [Grigolia, 1973] R.S. Grigolia. An algebraic analysis of n-valued systems of Lukasiewicz-Tarski (russian). *Proceedings of the University of Tbilisi A*, 6-7:121–132, 1973.
- [Grigolia, 1977] R.S. Grigolia. Algebraic analysis of Lukasiewicz-Tarski's n-valued logical systems. In R. Wójcicki and G. Malinowski, editors, *Selected papers on Lukasiewicz sentential Calculi*, pages 81–92. Ossolineum, 1977.

- [Halldén, 1949] S. Halldén. The logic of nonsense. *Uppsala Universitets årsskrift*, 9:132, 1949.
- [Hay, 1963] L. Hay. Axiomatization of the infinite-valued predicate calculus. *Journal of Symbolic Logic*, 28:77–86, 1963.
- [Höhle and Klement, 1995] U. Höhle and E. P. Klement, editors. *Non-classical logics and their applications to fuzzy subsets*. Kluwer, 1995.
- [Jeffrey, 1965] R. Jeffrey. *The Logic of Decision*. McGraw Hill, New York, 1965.
- [Jordan, 1945] Z. A. Jordan. *The Development of Mathematical Logic and of Logical Positivism in Poland between the Two Wars*. Oxford University Press, 1945. Partially reprinted in [McCall, 1967].
- [Kleene, 1938] S. C. Kleene. On a notation for ordinal numbers. *Journal of Symbolic Logic*, 3:150–155, 1938.
- [Kleene, 1952] S. C. Kleene. *Introduction to Metamathematics*. Van Nostrand, Amsterdam and Princeton, 1952.
- [Los and Suszko, 1958] J. Los and R. Suszko. Remarks on sentential logics. *Indagationes Math*, 20:177–183, 1958.
- [Lukasiewicz and Tarski, 1930] J. Lukasiewicz and A. Tarski. Untersuchungen über den Aussagenkalkül. *Comptes rendus de la Société des Sciences et des Lettres de Varsovie*, 23:1–21, 1930. English translation in [Lukasiewicz, 1970].
- [Lukasiewicz, 1930] J. Lukasiewicz. Philosophische Bemerkungen zu mehrwertigen Systemen des Aussagenkalküls. *Comptes rendus de la Société des Sciences et des Lettres de Varsovie*, 23:51–77, 1930. English translation in [Lukasiewicz, 1970].
- [Lukasiewicz, 1970] J. Lukasiewicz. *Selected Works*. North-Holland, Amsterdam, 1970. Edited by L. Borkowski.
- [Malinowski, 1993] G. Malinowski. *Many-valued Logics*. Oxford, 1993.
- [McCall, 1967] S. McCall, editor. *Polish Logic 1920–1939*. Oxford University Press, 1967.
- [McNaughton, 1951] R. McNaughton. A theorem about infinite-valued sentential logic. *Journal of Symbolic Logic*, 16:1–13, 1951.
- [Meredith, 1958] C. A. Meredith. The dependence of an axiom of Lukasiewicz. *Transactions of the American Mathematical Society*, 87:54, 1958.
- [Mostowski, 1979] A. Mostowski. Models of set theory. In *Foundational Studies, Selected Works*, volume 1. North-Holland, Amsterdam, 1979. Lectures delivered in Varenna, September 1968.
- [Mundici, 1986] D. Mundici. Interpretation of AF  $C^*$ -algebras in Lukasiewicz sentential calculus. *Journal of Functional Analysis*, 65:15–63, 1986.
- [Mundici, 1992] D. Mundici. The logic of Ulam's game with lies. In Christina Bicchieri and Maria Luisa Dalla Chiara, editors, *Knowledge, belief and strategic interaction*. Cambridge University Press, 1992.
- [Mundici, 1993] D. Mundici. Ulam's game, Lukasiewicz logic and AF  $C^*$ -algebras. *Fundamenta Informaticae*, 18:151–161, 1993.
- [Mundici, 1994] D. Mundici. A constructive proof of McNaughton's theorem in infinite-valued logic. *Journal of Symbolic Logic*, 59:596–602, 1994.
- [Panti, 1995] G. Panti. A geometric proof of the completeness of the Lukasiewicz calculus. *Journal of Symbolic Logic*, 60:563–578, 1995.
- [Pavelka, 1979] J. Pavelka. On fuzzy logic, I,II,III. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 25:45–52, 119–134, 447–464, 1979.
- [Post, 1921] E. Post. Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, 43:163–185, 1921.
- [Quine, 1953] W. V. Quine. On a so-called paradox. *Mind*, 62:65–67, 1953. Reprinted in *The Ways of Paradox and Other Essays*, Random House, New York, 1966.
- [Rasiowa, 1977] H. Rasiowa. Many-valued algorithmic logic as a tool to investigate programs. In *Modern Uses of Multiple-valued Logic*. D. Reidel, Dordrecht and Boston, 1977.
- [Rautenberg, 1981] W. Rautenberg. 2-element matrices. *Studia Logica*, 40:315–353, 1981.
- [Rescher, 1969] N. Rescher. *Many-valued Logic*. McGraw-Hill, New York, 1969.

- [Rine, 1984] D.C. Rine, editor. *Computer Science and Multiple-valued Logic*. North-Holland, 1984.
- [Rose and Rosser, 1958] A. Rose and J. B. Rosser. Fragments of many-valued statement calculi. *Transactions of the American Mathematical Society*, 87:1–53, 1958.
- [Rosenberg, 1965] I. Rosenberg. La structure des fonctions de plusieurs variables sur un ensemble fini. *C. R. Acad. Sci. Paris, Ser. A.B.*, 260:3817–3819, 1965.
- [Rosenberg, 1970] I. Rosenberg. Über die funktionale Vollständigkeit in den mehrwertigen Logiken (Struktur der Funktionen von mehreren Veränderlichen auf endlichen Mengen). *Rozprawy Cs. Akademie Ved. Ser. Math. Nat. Sci.*, 80:3–93, 1970.
- [Rosenberg, 1984] I. Rosenberg. Completeness properties of multiple-valued logic algebras. In David C. Rine, editor, *Computer Science and Multiple-valued logic*, pages 150–192. North-Holland, 1984.
- [Rosser and Turquette, 1952] J. B. Rosser and A. R. Turquette. *Many-valued Logics*. North-Holland, Amsterdam, 1952.
- [Rosser, 1960] J. B. Rosser. Axiomatisation of infinite-valued logics. *Logique et Analyse*, 3:137–153, 1960.
- [Salwicki, 1970] A. Salwicki. Formalised algorithmic languages. *Bull. Acad. Pol. Sci., Ser. Math. Astron. Phys.*, 18:227–232, 1970.
- [Scarpellini, 1962] B. Scarpellini. Die Nicht-Axiomatisierbarkeit des unendlichwertigen Prädikaten-kalküls von Lukasiewicz. *Journal of Symbolic Logic*, 27:159–170, 1962.
- [Scott, 1974] D. Scott. Completeness and axiomatizability in many-valued logic. In *Proceedings of the Tarski Symposium*. Proceedings of Symposia in Pure Mathematics, Vol. XXV, American Mathematical Society, Rhode Island, 1974.
- [Scott, 1976] D. Scott. Does many-valued logic have any use? In S. Körner, editor, *Philosophy of Logic*. Blackwell, Oxford, 1976.
- [Shoesmith and Smiley, 1971] D.J. Shoesmith and T.J. Smiley. Deducibility and many-valuedness. *J. of Symbolic Logic*, 36:610–622, 1971.
- [Skolem, 1957] T. Skolem. Bemerkungen zum Komprehensionsaxiom. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 3:1–17, 1957.
- [Słupecki, 1936] J. Słupecki. Der volle dreiwertige Aussagenkalkül. *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie*, 29:9–11, 1936. English translation in [McCall, 1967].
- [Słupecki, 1939] J. Słupecki. Completeness criterion for systems of many-valued propositional calculus (in Polish). *Comptes rendus des Séances de la Société des Sciences et des Lettres de Varsovie*, 32:102–109, 1939. English translation in *Studia Logica*, 30, 153–157, 1972.
- [Smiley, 1976] T. Smiley. Comment on Scott [1976]. In S. Körner, editor, *Philosophy of Logic*. Blackwell, Oxford, 1976.
- [Traczyk, 1963] T. Traczyk. Axioms and some properties of Post algebras. *Colloq. Math*, 10:193–209, 1963.
- [Traczyk, 1964] T. Traczyk. An equational definition of a class of Post algebras. *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astronom. Phys.*, 12:147–150, 1964.
- [Ulam, 1976] S. Ulam. *Adventures of a Mathematician*. Scribner, New York, 1976.
- [Urquhart, 1973] A. Urquhart. An interpretation of many-valued logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 19:111–114, 1973.
- [Urquhart, 1977] A. Urquhart. A finite matrix whose consequence relation is not finitely axiomatizable. *Rep. Math. Logic*, 9:71–71, 1977.
- [van Fraassen, 1966] B. van Fraassen. Singular terms, truth value gaps and free logic. *Journal of Philosophy*, 63:481–495, 1966.
- [Wade, 1945] C. I. Wade. Post algebras and rings. *Duke Math. J.*, 12:389–395, 1945.
- [Wajsberg, 1931] M. Wajsberg. Aksjomatyzacja trójwartściowego rachunku zdań (Axiomatisation of the 3-valued propositional calculus). *Comptes rendus des séances de la Société des Sciences et des Lettres de Varsovie*, 24:125–148, 1931. English translation in [McCall, 1967].
- [White, 1979] R. B. White. The consistency of the axiom of comprehension in the infinite-valued predicate logic of Lukasiewicz. *Journal of Philosophical Logic*, 8:509–534, 1979.

- [Wójcicki and Malinowski, 1977] R. Wójcicki and G. Malinowski, editors. *Selected Papers on Lukasiewicz Sentential Calculi*. Ossolineum, Wrocław, 1977.
- [Wójcicki, 1970] R. Wójcicki. Some remarks on the consequence operation in sentential logics. *Fundamenta Mathematica*, 68:269–279, 1970.
- [Wójcicki, 1988] R. Wójcicki. *Theory of Logical Calculi: Basic Theory of Consequence Operations*. Kluwer, Dordrecht and Boston, 1988.
- [Wolf, 1977] R. G. Wolf. A survey of many-valued logic (1966–1974). In *Modern Uses of Multiple-valued Logic*. D. Reidel, Dordrecht and Boston, 1977.
- [Wroński, 1976] A. Wroński. On finitely based consequence operations. *Studia Logica*, 35:453–458, 1976.
- [Wroński, 1979] A. Wroński. A three element matrix whose consequence operation is not finitely based. *Bulletin of the Section of Logic, Polish Academy of Sciences*, 8:68–71, 1979.
- [Zadeh, 1965] L. A. Zadeh. Fuzzy sets. *Information and Control*, 8:338–353, 1965.
- [Zadeh, 1996] L. A. Zadeh. *Fuzzy sets, Fuzzy logics and Fuzzy systems*. World Scientific, Singapore, 1996.



## ADVANCED MANY-VALUED LOGICS

## 1 INTRODUCTION

Let me begin with a brief discussion of the name of this chapter: the adjective “advanced” in the title can only be understood in the temporal sense; the bulk of Urquhart’s chapter in this Handbook was written for the first edition in the early 1980s and, therefore, does not cover recent results in depth. Perhaps “complementary” would be an altogether more fitting qualification for the present text. It is not required to have read “Basic Many-Valued Logic” in order to use my chapter. It is (I hope) not more difficult to read, either. On the other hand, you will find few overlaps and for sure some quite different points of view. Urquhart’s chapter, for example, covers functional completeness, model theory, or theory of consequence relations very well, and I do not repeat this material.

So what *are* the characteristics of the present chapter? First of all, I included a lot of material on proof theory: not only results on axiomatization, but as well on automated deduction; I tried to stress connections between MVL and mainstream topics of mathematics and computer science (for example, linear optimization, constraint programming, or circuit design) whenever possible; I concentrated on topics that, in my opinion, are currently active and promising areas for research; if suitable, applications and implemented systems are mentioned; at the end of this chapter, a list of resources is included that will be continually updated on the web.

Section 2 serves to set up the notational framework for many-valued logic needed later on. It also includes more or less standard material on abstract algebra and inference systems to make this chapter self-contained. Many-valued logic is sometimes criticized for lack of philosophical or mathematical motivation of the additional truth values. If you should have such doubt, then I try to convince you that it is worthwhile to read on in Section 2.5, immediately after the tools for making precise statements have been assembled.

In the last decade or so extremely rich algebraic and geometric structures in (infinite-valued) Łukasiewicz logic and related systems were exhibited; moreover, convincing semantics of many-valued logics coming from such diverse fields as coding theory, functional analysis, geometry, or quantum physics were given. Along with this, new completeness proofs for Łukasiewicz’s axiom system and new proofs of McNaughton’s fundamental characterization of Łukasiewicz logic were found. In Section 4 some strands of this research are reviewed. This is prepared in the preceding Section 3 on algebraic aspects in general, where Łukasiewicz logic is discussed

as a particular member of a large family of many-valued logics based on residuated lattices.

The field of fuzzy logic was not taken seriously by mainstream logicians until a couple of years ago (it was more or less regarded as an *ad hoc* notation used by engineers). This perception has changed.<sup>1</sup> In Section 5, I discuss the re-evaluated relationship between many-valued and fuzzy logic.

Although there existed implementations of satisfiability checkers for certain MVLs as far back as 1967, serious implementation efforts (and along with it, investigations into computational properties of MVL) were only begun as part of the wave of non-classical deduction (mainly for AI applications) that surged from the mid-1980s onwards. By now many results from classical deduction have been generalized to the many-valued case, often in a generic way. Additional sources of complexity in deduction arising from many-valuedness were identified and tools to cope with this added complexity were developed. In a sense, computational MVL is more mature than, say, computational modal logic. Tools for analyzing many-valued logics and for finding formal proofs in them are available. The main developments are sketched in Section 6.

Design of multimedia databases capable of dealing with heterogeneous, distributed, richly typed data is acknowledged as one of the future challenges of database theory [Gray, 1996]. Deductive databases can form the basis for the more active role expected in this context from a database. Many-valued logic can enhance the expressivity of deductive databases, particularly when dealing with inconsistent and/or incomplete information. Closely related to automated deduction and deductive databases, logic programming forms a link between both in terms of expressivity and procedurality. MVL plays quite different roles here: on the one hand, it can be used to provide semantics to extended classes of logic programs, on the other hand, logic programs themselves may be extended to accommodate many-valued connectives. Many-valued logic from the point of view of deductive databases and logic programming is the topic of Section 7.

Discrete function manipulation has been a traditional tool in circuit design and a long-standing topic of applied MVL research. In the last decade design techniques based on *Binary and Many-Valued Decision Diagrams* became predominant in switch-level design. A less known fact which I also elaborate on in Section 8 are the close connections between decision diagrams and proof theory of many-valued logic.

Questions on the complexity and decidability of problems associated to classical logic can be freshly asked in the many-valued context, along with new ones. Some answers are as expected, but some surprises are in store as well, see Section 9.

---

<sup>1</sup>Both, their former opinions about and their re-evaluation of fuzzy logic were expressed, for example, by logicians S. Gottwald, P. Hájek, and P. Schmitt in personal communication with the author [1996].

Many researchers, for example, Gabbay [1996], claim that applications of AI in logic demand combination of various (and diverse) non-classical logics within one system of reasoning. It is, therefore, important to understand how MVL interacts with other logics and what the options for combining them are. An overview of results achieved so far is in Section 10.

The chapter is closed with Section 11, a list of resources available to those interested in MVL research.

## 2 PRELIMINARIES

In many-valued logic, we work with standard propositional and first-order languages, but since we have different and/or additional connectives and quantifiers it is convenient to regard these not as fixed, but to parameterize many-valued logic languages with sets of connectives and possibly quantifiers.

### 2.1 Propositional Logic

As usual, a countably infinite set of propositional variables  $\Sigma = \{p, q, r, \dots\}$  is a **propositional signature**. Let  $\Theta$  be a finite set of operator names, called **connectives**; the **arity** of  $\theta \in \Theta$  is a non-negative integer given by a function  $\alpha$ . Connectives with arity 0 are called **logical constants**. Given a **propositional language**  $\mathbf{L}^0 = \langle \Theta, \alpha \rangle$  and a signature  $\Sigma$ , the set of  **$\mathbf{L}_\Sigma^0$ -formulas** is defined inductively as usual:

1. Members of  $\Sigma$  and logical constants are  $\mathbf{L}_\Sigma^0$ -formulas.
2. If  $\theta \in \Theta$ ,  $\alpha(\theta) = r > 0$ , and  $\varphi_1, \dots, \varphi_r$  are  $\mathbf{L}_\Sigma^0$ -formulas, then also  $\theta(\varphi_1, \dots, \varphi_r)$  is an  $\mathbf{L}_\Sigma^0$ -formula.

The concrete signature usually is not relevant, and then we omit it from the index.

EXAMPLE 1. Throughout this article,  $\bar{0}$  is a logical constant, all connectives of the form  $\neg_x, \nabla$  are unary, while  $\rightarrow_x, \vee, \wedge, \oplus_x, \odot_x$  are binary ones.

1. The language  $\mathbf{L}_c^0$  of classical propositional logic contains the connectives  $\bar{0}, \wedge, \rightarrow$ .
2. The language  $\mathbf{L}_L^0$  of Łukasiewicz logic contains the connectives  $\bar{0}, \odot_L, \rightarrow_L$ .
3. The language  $\mathbf{L}_G^0$  of Gödel logic contains the connectives  $\bar{0}, \wedge, \rightarrow_G$ .
4. The language  $\mathbf{L}_\Pi^0$  of product logic contains the connectives  $\bar{0}, \odot_\Pi, \rightarrow_\Pi$ .



5. We have the family of languages  $\mathbf{L}_t^0$  of  $t$ -norm logics containing the connectives  $\bar{0}, \odot_t, \rightarrow_t$ .
6. The language  $\mathbf{L}_P^0$  of Post logic contains the connectives  $\neg_P, \wedge$ .
7. The language of the paraconsistent logic  $\mathbf{L}_J^0$  contains the connectives  $\neg, \nabla$ , and  $\vee$ .

We say **negation** to the connectives  $\neg_x$ , **implication** to  $\rightarrow_x$ , **conjunction** to  $\wedge$ , **disjunction** to  $\vee$ , **product** to  $\odot_x$ , and **sum** to  $\oplus_x$ . This can be qualified with a language, for example, when one says “Łukasiewicz implication”, the connective  $\rightarrow_L$  is meant.

Given a signature, in classical logic, a propositional variable can either be true or false. Each such interpretation of the variables determines an interpretation of arbitrary formulas in a fixed way. In many-valued logic, we must be more flexible: instead of true and false, a value from an arbitrary, non-empty set of truth values  $N$  can be assigned to a variable. Moreover, for each member of  $\Theta$ , its behaviour on  $N$  must be fixed. As we will consider many different systems, it is convenient to introduce some terminology:

A **propositional matrix**  $\mathbf{A}^0 = \langle N, (A_\theta)_{\theta \in \Theta} \rangle$  for a propositional language  $\mathbf{L}^0$  consists of a non-empty set of **truth values**  $N$  and a collection  $(A_\theta)_{\theta \in \Theta}$  of operations on  $N$  such that  $A_\theta : N^{\alpha(\theta)} \rightarrow N$  for each  $\theta \in \Theta$ .  $|N|$  denotes the cardinality of  $N$ .

A pair  $\mathcal{L}^0 = \langle \mathbf{L}^0, \mathbf{A}^0 \rangle$ , where  $\mathbf{L}^0$  is a propositional language and  $\mathbf{A}^0$  is a matrix for  $\mathbf{L}^0$ , is called ( $N$ -valued) **propositional logic**.

**EXAMPLE 2.** For each  $n \geq 2$  let  $\mathbf{n} = \{0, \frac{1}{n-1}, \dots, \frac{n-2}{n-1}, 1\}$  be the truth value set of cardinality  $n$  consisting of equidistant rational numbers. With  $[0, 1]$  we denote the closed real unit interval. In the following, let  $N$  be either  $[0, 1]$  or  $\mathbf{n}$  for some  $n$ . In all logics considered below,  $A_{\bar{0}} = 0$ ,  $A_{\wedge} = \min$ , and  $A_{\vee} = \max$ . In each of 1.–5. below, implication is defined as  $A_{\rightarrow_x}(i, j) = \sup\{k \mid A_{\odot_x}(i, k) \leq j\}$  with the help of product. This process is called **residuation**.

1. In **classical propositional logic**  $\mathcal{L}_c^0$ ,  $n = 2$  (hence,  $N = \{0, 1\}$ ). If residuation is based on  $\wedge$ , the usual definitions result.
2. In **Łukasiewicz logic** [Łukasiewicz, 1920; Łukasiewicz and Tarski, 1930]  $\mathcal{L}_L^0$ ,  $A_{\odot_L}(i, j) = \max\{0, i + j - 1\}$ .
3. In **Gödel logic** [Gödel, 1932]  $\mathcal{L}_G^0$ ,  $A_{\wedge} = A_{\odot_G} = \min$ .
4. In **product logic** [Hájek *et al.*, 1996]  $\mathcal{L}_\Pi^0$ ,  $A_{\odot_\Pi} = \cdot$  (multiplication). Product logic is only defined for  $N = [0, 1]$ , because none of the sets  $\mathbf{n}$  is closed under  $A_{\odot_\Pi}$ .

5. For  $N = [0, 1]$ , an operator  $* : [0, 1]^2 \rightarrow [0, 1]$  is a **triangular norm** (*t*-norm, for short) if it is

**commutative:**  $i * j = j * i$  for all  $i, j \in [0, 1]$

**associative:**  $(i * j) * k = i * (j * k)$  for all  $i, j, k \in [0, 1]$

**non-decreasing:**  $i_1 \leq i_2$  implies  $i_1 * j \leq i_2 * j$  for all  $i_1, i_2, j \in [0, 1]$  (hence, by commutativity, also  $j_1 \leq j_2$  implies  $i * j_1 \leq i * j_2$  for all  $i, j_1, j_2 \in [0, 1]$ )

**left-neutral in 1:**  $1 * i = i$  for all  $i \in [0, 1]$  (hence, by commutativity,  $*$  is also right-neutral; in particular,  $0 = 0 * 1 \geq 0 * i$ , thus  $0 * i = i * 0 = 0$  for all  $i \in [0, 1]$ ).

A so-called ***t*-norm logic**  $\mathcal{L}_t^0$  is obtained from any *continuous* (in the usual sense) *t*-norm  $A_{\odot_t}$  by restriction of  $A_{\odot_t}$  to  $N$ , whenever  $N$  is closed under  $\odot_t$ .

6. In **Post logic** [Post, 1921]  $\mathcal{L}_P^0$ ,  $N = \mathbf{n}$  is finite and

$$A_{\neg_P}(i) = \begin{cases} 1 & i = 0 \\ i - \frac{1}{n-1} & i > 0 \end{cases} .$$

7. In **paraconsistent logic**<sup>2</sup>  $\mathcal{L}_J^0$ ,  $N = \mathbf{n}$  is finite and

$$A_{\nabla}(i) = \begin{cases} 2i & 0 \leq i < \frac{1}{n-1} \\ i + \frac{1}{n-1} & \frac{1}{n-1} \leq i < 1 \\ 1 & i = 1 \end{cases} .$$

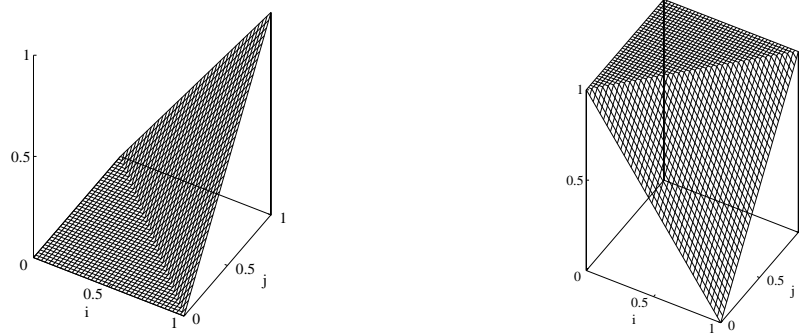
It is straightforward to check that each of the products of Łukasiewicz, Gödel, and product logic is a continuous *t*-norm. The central place of these particular *t*-norms is justified by the fact that any continuous *t*-norm can be represented with Łukasiewicz, Gödel, and product *t*-norms alone (Section 5.2).

Some operators of Łukasiewicz and Gödel logic for  $N = [0, 1]$  are displayed in Figures 1 and 2. Observe, how implication results from residuation of the product.

In *t*-norm theory one often considers  $\bar{0}$  and  $\odot_t$  as primitive operators, in algebra,  $\neg$  and  $\rightarrow$  or  $\oplus$ . The base is not really relevant, but when everything must be expressed with the help of base operators, then we have to translate back and forth to versions used in the respective literature all the time. Therefore, I use a nearly minimal and mostly uniform base for each logic

---

<sup>2</sup>For  $N = \mathbf{3}$  this logic was defined in [D'Ottaviano and da Costa, 1970], beyond  $\mathbf{3}$  several possibilities to define  $\nabla$  exist. The definition used here was suggested by João Marcos. Note that  $A_{\nabla}$  converges towards the identity function as  $n$  increases. For  $N = \mathbf{2}$ , too,  $A_{\nabla}$  is the identity.

Figure 1. Łukasiewicz operators for  $N = [0, 1]$ .

and add further connectives as abbreviations, if needed. For example, in each  $t$ -norm logic, negation and sum connectives are obtained by writing

- (1)  $\neg_t \varphi$  for  $\varphi \rightarrow_t \bar{0}$   
 (2)  $\varphi \oplus_t \psi$  for  $\neg_t(\neg_t \varphi \odot_t \neg_t \psi)$

Further, one can stipulate

- (3)  $\varphi \wedge_t \psi$  for  $\varphi \odot_t (\varphi \rightarrow_t \psi)$   
 (4)  $\varphi \vee_t \psi$  for  $((\varphi \rightarrow_t \psi) \rightarrow_t \psi) \wedge ((\psi \rightarrow_t \varphi) \rightarrow_t \varphi)$

An  $N$ -valued matrix  $\mathbf{A}^0$  is called **functionally complete**, if every  $m$ -ary function  $f : N^m \rightarrow N$  can be defined using operators from  $\mathbf{A}^0$  alone. The matrix of classical logic is functionally complete and the matrix of  $n$ -valued Post logic  $\mathcal{L}_P^0$  is functionally complete for each  $n$ . None of the other matrices is functionally complete. Functional completeness is thoroughly discussed in Chapter Urquhart's chapter in this Volume of this Handbook.

A **propositional interpretation**  $\mathbf{I}$  determines the truth value of each variable in a given signature  $\Sigma$ , hence it is simply a mapping  $\mathbf{I} : \Sigma \rightarrow N$ . For each propositional many-valued logic  $\mathcal{L}^0$  its matrix  $\mathbf{A}^0$  uniquely (simple exercise) determines the extension of any  $\Sigma$ -interpretation to a **propositional valuation** function on  $\mathbf{L}_\Sigma^0$  (for which the same symbol is used) via

$$(5) \quad \mathbf{I}(\theta(\varphi_1, \dots, \varphi_r)) = A_\theta(\mathbf{I}(\varphi_1), \dots, \mathbf{I}(\varphi_r)) .$$

Let  $D$  be any subset of  $N$  and  $\Psi$  a set of  $\mathbf{L}_\Sigma^0$ -formulas. We say that  $\Psi$  is  **$D$ -satisfiable** if there is a  $\Sigma$ -interpretation  $\mathbf{I}$  such that  $\mathbf{I}(\varphi) \in D$  for all

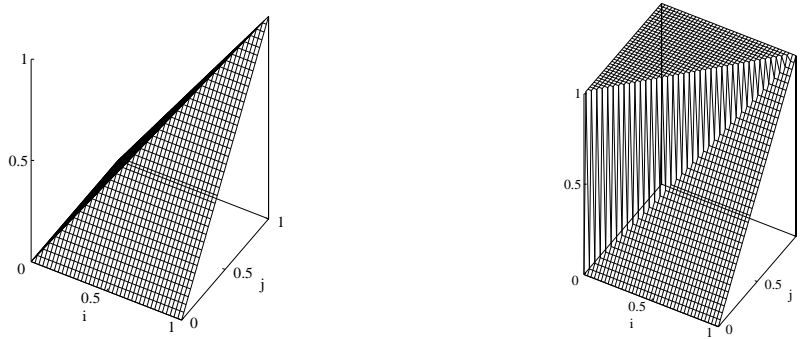


Figure 2. Gödel operators for  $N = [0, 1]$ .

$\varphi \in \Psi$ . Such an  $\mathbf{I}$  is called  $D$ -**model** of  $\Psi$ , in symbolic notation  $\mathbf{I} \models_D \Psi$ . The set  $\Psi$  is  $D$ -**valid** if all  $\Sigma$ -interpretations are  $D$ -models of  $\Psi$ , in symbols  $\models_D \Psi$ . We usually write  $\models_D \varphi$  instead of  $\models_D \{\varphi\}$ . Finally, a formula  $\varphi$  is a  $D$ -**consequence** of  $\Psi$  if every  $D$ -model of  $\Psi$  is as well a  $D$ -model of  $\varphi$ , in symbols  $\Psi \models_D \varphi$ .

The well-known duality between satisfiability and validity known from classical logic, extends as follows:  $\varphi$  is  $D$ -valid iff it is not  $\overline{D}$ -satisfiable.

$D = \emptyset$  is not excluded for technical reasons; obviously, no formula is  $\emptyset$ -satisfiable. The notions of  $D$ -validity and -satisfiability are due to [Kirin, 1966].

In many cases, the set  $D$  is considered to be fixed in a given many-valued logic. It is then called the set of **designated truth values**, and one writes “satisfiable” instead of “ $D$ -satisfiable”, “ $\models$ ” instead of “ $\models_D$ ”, etc.

EXAMPLE 3. For all the logics of Example 2, with the exception of  $\mathcal{L}_J^0$ , the usual choice for the designated truth values is  $D = \{1\}$ . Here are some examples of formulas that are valid in each  $t$ -norm logic:<sup>3</sup>

$$(A1) \quad (\varphi \rightarrow_t \psi) \rightarrow_t ((\psi \rightarrow_t \chi) \rightarrow_t (\varphi \rightarrow_t \chi))$$

$$(A2) \quad (\varphi \odot_t \psi) \rightarrow_t \varphi$$

$$(A3) \quad (\varphi \odot_t \psi) \rightarrow_t (\psi \odot_t \varphi)$$

$$(A4) \quad (\varphi \odot_t (\varphi \rightarrow_t \psi)) \rightarrow_t (\psi \odot_t (\psi \rightarrow_t \varphi))$$

<sup>3</sup>These are the axioms of basic  $t$ -norm logic [Hájek, 1998], see below.

$$(A5a) \quad (\varphi \rightarrow_t (\psi \rightarrow_t \chi)) \rightarrow_t ((\varphi \odot_t \psi) \rightarrow_t \chi)$$

$$(A5b) \quad ((\varphi \odot_t \psi) \rightarrow_t \chi) \rightarrow_t (\varphi \rightarrow_t (\psi \rightarrow_t \chi))$$

$$(A6) \quad ((\varphi \rightarrow_t \psi) \rightarrow_t \chi) \rightarrow_t (((\psi \rightarrow_t \varphi) \rightarrow_t \chi) \rightarrow_t \chi)$$

$$(A7) \quad \bar{0} \rightarrow_t \varphi$$

Moreover, in any  $t$ -norm logic,  $\{\varphi, \varphi \rightarrow_t \psi\} \models \psi$ .

In  $\mathcal{L}_J^0$ , one fixes  $D = N - \{0\}$ , which has the effect that  $\nabla$  characterizes designatedness.

We may define a strong notion of equivalence that can be paraphrased as “ $\{i\}$ -consequence in both directions for all  $i \in N$ ”:  $\mathbf{L}_\Sigma^0$ -formulas  $\varphi, \psi$  are **strongly equivalent**, briefly  $\varphi \equiv \psi$ , when  $\mathbf{I}(\varphi) = \mathbf{I}(\psi)$  for all  $\Sigma$ -interpretations  $\mathbf{I}$ . For example, with the definitions in (3), respectively, in (4) one has  $\varphi \wedge \psi \equiv \varphi \wedge_x \psi$ , respectively,  $\varphi \vee \psi \equiv \varphi \vee_x \psi$  in every  $t$ -norm logic. Therefore, only the symbols  $\vee$  and  $\wedge$  will be used in all these logics.

EXAMPLE 4. It is easy to see that for  $N = \mathbf{3}$  the matrices of the logics  $\mathcal{L}_L^0$  and  $\mathcal{L}_J^0$  define the same functions:  $\nabla\varphi \equiv \neg\varphi \rightarrow_L \varphi$  and, vice versa,  $\varphi \rightarrow_L \psi \equiv (\nabla\neg\varphi \vee \psi) \wedge (\neg\varphi \vee \nabla\psi)$ ,  $\varphi \odot_L \psi \equiv \neg(\neg\varphi \rightarrow_L \psi)$ , and  $\bar{0} \equiv \neg(\neg\nabla\varphi \vee \nabla\varphi)$ .

REMARK 5. A choice other than  $D = \{1\}$  for the set of designated truth values can have interesting consequences: as remarked earlier, the three-valued matrices of  $\mathcal{L}_L^0$  and  $\mathcal{L}_J^0$  are identical up to the base of operators. The choice of  $D = N - \{0\} = \{\frac{1}{2}, 1\}$  for  $\mathcal{L}_J^0$ , however, renders the latter a **paraconsistent logic**. This is a logic equipped with a negation connective  $\neg$  such that a syntactically inconsistent set of formulas containing, say,  $\varphi$  and  $\neg\varphi$  does not have every formula as a logical consequence (as it is the case in classical logic). Indeed, in three-valued  $\mathcal{L}_J^0$  one has, for example,  $\{\varphi, \neg\varphi\} \not\models \psi$  and  $\{\varphi, \neg\varphi \vee \psi\} \not\models \psi$ . One can even characterize the consequence relation syntactically within the logic. Let  $\varphi \supset \psi \equiv \neg\nabla\varphi \vee \psi$ , then  $\Psi \models \varphi$  iff  $\models (\bigwedge_{\psi \in \Psi} \psi) \supset \varphi$  ( $\Psi$  finite).  $\{1\}$ -consequence can be recovered with the connective  $\Delta\varphi \equiv \neg\nabla\neg\varphi$ , which restricts models to  $\{1\}$ -models:  $\models \Delta\varphi$  iff  $\models_{\{1\}} \varphi$ . On the other hand, unsatisfiability can easily be expressed using, for example, the formula  $\Delta\varphi \wedge \neg\Delta\varphi$ . These properties motivated the definition of three-valued  $\mathcal{L}_J^0$  in [D’Ottaviano and da Costa, 1970], where it was called  $J_3$ . For more than three truth values similar relationships hold in  $\mathcal{L}_J^0$ .

Paraconsistent logic is a field of research in its own right and discussed at length elsewhere in this Handbook. Many paraconsistent logics, such as the systems  $P^1$  of [Sette, 1973; Sette and Carnielli, 1995] and  $LFI1$  (whose matrix defines the same functions as  $J_3$ ) of [Carnielli *et al.*, to appear, 2000], can be seen as a particular many-valued logic with more than one designated value.

## 2.2 First-Order Logic

A **first-order signature**  $\Sigma$  is a triple  $\langle \mathbf{P}_\Sigma, \mathbf{F}_\Sigma, \alpha_\Sigma \rangle$ , where  $\mathbf{P}_\Sigma$  is a non-empty family of **predicate symbols**,  $\mathbf{F}_\Sigma$  a possibly empty family of **function symbols** disjoint from  $\mathbf{P}_\Sigma$ , and  $\alpha_\Sigma$  assigns an arity to each member of  $\mathbf{P}_\Sigma \cup \mathbf{F}_\Sigma$ .

Let  $\text{Term}_\Sigma$  be the set of  $\Sigma$ -terms, defined inductively as usual over a countably infinite set of **object (or individual) variables**  $\text{Var} = \{x_0, x_1, \dots\}$ :

1. Members of  $\text{Var}$  and  $c \in \mathbf{F}_\Sigma$  with  $\alpha(c) = 0$  are  $\Sigma$ -terms.
2. If  $f \in \mathbf{F}_\Sigma$ ,  $\alpha(f) = r > 0$ , and  $t_1, \dots, t_r$  are  $\Sigma$ -terms, then  $f(t_1, \dots, t_r)$  is a  $\Sigma$ -term.

$\text{Term}_\Sigma^0$  are the variable-free terms in  $\text{Term}_\Sigma$ , the so-called **ground terms**. The **atoms** are defined as:

$$\text{At}_\Sigma = \{p(t_1, \dots, t_r) \mid p \in \mathbf{P}_\Sigma, \alpha_\Sigma(p) = r, t_i \in \text{Term}_\Sigma\} .$$

A **first-order language** is a triple  $\mathbf{L} = \langle \Theta, \Lambda, \alpha \rangle$ , where  $\langle \Theta, \alpha \rangle$  is propositional language and  $\Lambda$  is a finite set of **first-order quantifiers**.

EXAMPLE 6. All of the propositional languages  $\mathbf{L}_x^0$  defined in Example 1 are extended to first-order languages  $\mathbf{L}_x$  by adding the quantifier set  $\Lambda = \{\forall, \exists\}$ .

The set of  **$\mathbf{L}_\Sigma$ -formulas** of a first-order language  $\mathbf{L}$  over a first-order signature  $\Sigma$  is inductively defined by:

1. The atoms over  $\Sigma$  are  $\mathbf{L}_\Sigma$ -formulas.
2. If  $\theta \in \Theta$ ,  $\alpha(\theta) = r > 0$ , and  $\varphi_1, \dots, \varphi_r$  are  $\mathbf{L}_\Sigma$ -formulas, then also  $\theta(\varphi_1, \dots, \varphi_r)$  is an  $\mathbf{L}_\Sigma$ -formula.
3. If  $\lambda \in \Lambda$ ,  $\varphi \in \mathbf{L}_\Sigma$ , and  $x \in \text{Var}$ , then  $(\lambda x)\varphi$  is an  $\mathbf{L}_\Sigma$ -formula and  $\varphi$  is the **scope** of  $(\lambda x)\varphi$ .

A variable  $x \in \text{Var}$  occurs **bound** in a formula  $\varphi$ , if  $\varphi$  contains a subformula of the form  $(\lambda x)\psi$ . It occurs **free**, if there is an occurrence of  $x$  in  $\varphi$  that is not in the scope of a subformula of the form  $(\lambda x)\psi$ .

A classical first-order formula  $\varphi$  is said to be in **conjunctive normal form** (CNF) iff it has the form  $(\forall x_1) \cdots (\forall x_r) \bigwedge_{k=1}^M \bigvee_{l=1}^{J_k} L_{kl}$ , where the  $L_{kl}$  are literals and  $\{x_1, \dots, x_r\}$  are the free variables in the scope. For any  $1 \leq k \leq M$ , the expression  $(\forall y_1) \cdots (\forall y_m) \bigvee_{l=1}^{J_k} L_{kl}$ , where  $\{y_1, \dots, y_m\}$  are the free variables in the scope, is a **clause** of  $\varphi$ .

If  $\mathbf{L} = \langle \Theta, \Lambda, \alpha \rangle$  is a first-order language, then a **first-order matrix** is a triple  $\mathbf{A} = \langle N, (A_\theta)_{\theta \in \Theta}, (Q_\lambda)_{\lambda \in \Lambda} \rangle$ , where  $\langle N, (A_\theta)_{\theta \in \Theta} \rangle$  is a propositional

matrix for  $\langle \Theta, \alpha \rangle$  and  $Q_\lambda : \mathcal{P}^+(N) \rightarrow N$  for each  $\lambda \in \Lambda$ , where  $\mathcal{P}^+(N)$  are the non-empty subsets of  $N$ .  $Q_\lambda$  is called the **distribution function** of the quantifier  $\lambda$ .

A pair  $\mathcal{L} = \langle \mathbf{L}, \mathbf{A} \rangle$ , where  $\mathbf{L}$  is a first-order language and  $\mathbf{A}$  is a first-order matrix for  $\mathbf{L}$ , is called ( $N$ -valued) **first-order logic**.

EXAMPLE 7. From each first-order language  $\mathbf{L}_x$  of Example 6 one obtains a first-order logic  $\mathcal{L}_x$  over  $N = [0, 1]$  or  $N = \mathbf{n}$  using distribution functions  $Q_\forall = \inf$  and  $Q_\exists = \sup$ .<sup>4</sup> The reader is invited to check that  $\mathcal{L}_c$  is classical first-order logic.

Given a first-order signature  $\Sigma$ , a **first-order structure**  $\mathbf{M} = \langle \mathbf{D}, \mathbf{I} \rangle$  fixes a non-empty set  $\mathbf{D}$ , the **domain** of discourse, and the meaning of function and predicate symbols via an interpretation  $\mathbf{I}$  that maps each function symbol  $f \in \mathbf{F}_\Sigma$  of arity  $r$  into a function  $\mathbf{I}(f) : \mathbf{D}^r \rightarrow \mathbf{D}$ , and each predicate symbol  $p \in \mathbf{P}_\Sigma$  of arity  $r$  into a function  $\mathbf{I}(p) : \mathbf{D}^r \rightarrow N$ . Observe that for  $\mathbf{F}_\Sigma = \emptyset$  and  $\mathbf{P}_\Sigma$  containing only 0-ary predicate symbols  $\mathbf{I}$  reduces to a propositional interpretation.

Like in the propositional case, for each first-order logic  $\mathcal{L}$  its matrix  $\mathbf{A}$  uniquely determines for any first-order structure  $\mathbf{M}$  a valuation function on arbitrary terms and  $\mathbf{L}_\Sigma$ -formulas. In addition, one must specify the meaning of object variables that might occur within formulas. This is done as usual with a **variable assignment**  $\beta : \text{Var} \rightarrow \mathbf{D}$ .

For given  $\mathbf{M}$  and  $\beta$ , the **first-order valuation** function  $v_{\mathbf{M},\beta}$  maps terms into  $\mathbf{D}$  and  $\mathbf{L}_\Sigma$ -formulas into  $N$ . For  $t \in \text{Term}_\Sigma$ , one writes  $t^{\mathbf{M},\beta}$  instead of  $v_{\mathbf{M},\beta}(t)$ . The definition is by induction:

$$(6) \quad x^{\mathbf{M},\beta} = \beta(x), \quad x \in \text{Var}$$

$$(7) \quad f(t_1, \dots, t_r)^{\mathbf{M},\beta} = \mathbf{I}(f)(t_1^{\mathbf{M},\beta}, \dots, t_r^{\mathbf{M},\beta}), \quad f \in \mathbf{F}_\Sigma, \quad \alpha(f) = r$$

$$(8) \quad v_{\mathbf{M},\beta}(p(t_1, \dots, t_r)) = \mathbf{I}(p)(t_1^{\mathbf{M},\beta}, \dots, t_r^{\mathbf{M},\beta}), \quad p \in \mathbf{P}_\Sigma, \quad \alpha(p) = r$$

$$(9) \quad v_{\mathbf{M},\beta}(\theta(\varphi_1, \dots, \varphi_r)) = A_\theta(v_{\mathbf{M},\beta}(\varphi_1), \dots, v_{\mathbf{M},\beta}(\varphi_r)), \quad \theta \in \Theta, \quad \alpha(\theta) = r$$

$$(10) \quad v_{\mathbf{M},\beta}((\lambda x)\varphi) = Q_\lambda(\{v_{\mathbf{M},\beta_x^d}(\varphi) \mid d \in \mathbf{D}\}), \quad \lambda \in \Lambda$$

The expression  $\{v_{\mathbf{M},\beta_x^d}(\varphi) \mid d \in \mathbf{D}\}$  in (10) is the **distribution** of  $\varphi$  at  $x$ . Equation (9) is, of course, the first-order version of (5).

Our definition of first-order structure automatically ensures that it is **safe** in the sense of [Hájek, 1998]:  $v_{\mathbf{M},\beta}(\varphi)$  is a total function in any  $N$ -valued first-order logic for all  $\varphi$  and  $\beta$ .

<sup>4</sup>Here, it would not do to have only rational numbers as truth values, because the rationals are not closed under inf and sup.

Satisfaction is defined analogously to the propositional case, with the exception that the presence of assignments gives rise to one more concept, just as in classical logic.

Let  $D$  be a non-empty subset of  $N$  and  $\Psi$  a set of  $\mathbf{L}_\Sigma$ -formulas. We say that  $\Psi$  is  $D$ -**satisfiable** if there is a first-order structure  $\mathbf{M}$  over  $\Sigma$  and a variable assignment  $\beta$  such that  $v_{\mathbf{M},\beta}(\varphi) \in D$  for all  $\varphi \in \Psi$ , for short write  $\mathbf{M}, \beta \models_D \Psi$ . When  $\mathbf{M}, \beta \models_D \Psi$  for all variable assignments  $\beta$ ,  $\Psi$  is said to be  $D$ -**true** in  $\mathbf{M}$ , for short  $\mathbf{M} \models_D \Psi$ , and  $\mathbf{M}$  is called  $D$ -**model** of  $\Psi$ . The formula set  $\Psi$  is  $D$ -**valid** if it is true in all first-order  $\Sigma$ -structures, in symbols  $\models_D \Psi$ . Logical consequence is defined as before, that is,  $\Psi \models_D \varphi$  iff every  $D$ -model of  $\Psi$  is as well  $D$ -model of  $\varphi$ . The conventions about dropping  $D$  when it is obvious and identifying  $\{\varphi\}$  with  $\varphi$  are as above.

A **substitution** is a mapping  $\sigma : \text{Var} \rightarrow \text{Term}_\Sigma$ . It is extended to terms and formulas as usual:

- $\sigma(c) = c$ , if  $c \in \mathbf{F}_\Sigma$  with  $\alpha(c) = 0$
- $\sigma(f(t_1, \dots, t_r)) = f(\sigma(t_1), \dots, \sigma(t_r))$ , if  $f(t_1, \dots, t_r) \in \text{Term}_\Sigma$ ,  $f \in \mathbf{F}_\Sigma$  with  $\alpha(f) = r > 0$
- $\sigma(p) = p$ , if  $p \in \mathbf{P}_\Sigma$  with  $\alpha(p) = 0$
- $\sigma(p(t_1, \dots, t_r)) = p(\sigma(t_1), \dots, \sigma(t_r))$ , if  $p(t_1, \dots, t_r) \in \text{At}_\Sigma$ ,  $p \in \mathbf{P}_\Sigma$  with  $\alpha(p) = r > 0$
- $\sigma(\theta) = \theta$ , if  $\theta$  is a logical constant
- $\sigma(\theta(\varphi_1, \dots, \varphi_r)) = \theta(\sigma(\varphi_1), \dots, \sigma(\varphi_r))$ , if  $\theta(\varphi_1, \dots, \varphi_r) \in \mathbf{L}_\Sigma$ ,  $\theta \in \Theta$  with  $\alpha(\theta) = r > 0$
- $\sigma((\lambda x)\varphi) = (\lambda x)\sigma'(\varphi)$  for  $(\lambda x)\varphi \in \mathbf{L}_\Sigma$  and  $\lambda \in \Lambda$ , where  $\sigma'(x) = x$  and  $\sigma' = \sigma$  otherwise.

If a substitution is the identity for all but finitely many object variables  $X = \{x_1, \dots, x_r\} \subset \text{Var}$ , then it is a **substitution for  $X$** , and it is written as  $\sigma = \{x_1/t_1, \dots, x_r/t_r\}$ , where  $\sigma(x_i) = t_i$ ,  $1 \leq i \leq r$ . Application of substitutions is usually written postfix (note that  $\varphi(\sigma \circ \rho) = (\varphi\rho)\sigma = \varphi\rho\sigma$ ). When the image on  $X$  of a substitution for  $X$  consists of ground terms one has a **ground substitution** for  $X$ .

We assume without loss of generality that all substitutions  $\sigma$  for  $X$  are **free** with respect to the formulas  $\varphi$  they are applied to, that is, no variable in the image of  $\sigma$  on  $X$  occurs bound in  $\varphi$ . This can easily be achieved by **bound renaming** of variables in  $\varphi$ : for each bound variable  $x$  in  $\varphi$  replace all occurrences of  $x$  with a  $y \in \text{Var}$  not occurring elsewhere. Bound renaming obviously preserves the models of a formula.



EXAMPLE 8. For the logics of Example 7, with the exception of  $\mathcal{L}_J$ , the usual choice for the designated truth values is  $D = \{1\}$ . Here are some examples of formulas that are valid in each first-order  $t$ -norm logic:<sup>5</sup>

$$(A8) \quad (\forall x)\varphi \rightarrow_t \varphi\{x/t\}, \text{ for any } t \in \text{Term}_\Sigma^0$$

$$(A9) \quad \varphi\{x/t\} \rightarrow_t (\exists x)\varphi, \text{ for any } t \in \text{Term}_\Sigma^0$$

$$(A10) \quad (\forall x)(\varphi \rightarrow_t \psi) \rightarrow_t (\varphi \rightarrow_t (\forall x)\psi), \text{ whenever } x \text{ not free in } \varphi$$

$$(A11) \quad (\forall x)(\varphi \rightarrow_t \psi) \rightarrow_t ((\exists x)\varphi \rightarrow_t \psi), \text{ whenever } x \text{ not free in } \psi$$

$$(A12) \quad (\forall x)(\varphi \vee \psi) \rightarrow_t ((\forall x)\varphi \vee \psi), \text{ whenever } x \text{ not free in } \psi$$

Moreover, in any  $t$ -norm logic,  $\{\varphi\} \models (\forall x)\varphi$ .

The semantics of quantifiers in many-valued logic is not straightforward. In logics having disjunction- and conjunction-like connectives  $\vee$  and  $\wedge$ , these can be used to define existential and universal quantifiers  $\exists$  and  $\forall$ , where  $v_{\mathbf{M},\beta} \models (\exists x)\varphi$  iff  $\bigvee_{d \in \mathbf{D}} (v_{\mathbf{M},\beta_x^d} \models \varphi)$  and similar for  $\forall$ .<sup>6</sup> This approach is taken in Section 3.3 of Chapter Urquhart's chapter in this Volume in the present volume. More generally, whenever  $N$  is a complete lattice with operations  $\prod$  and  $\sqcup$ , universal and existential quantifiers  $\Pi$  and  $\Sigma$  can be defined as above, and they are characterized by  $Q_\Pi(S) = \prod_{i \in S} i$ , respectively, by  $Q_\Sigma(S) = \sqcup_{i \in S} i$ , see [Zach, 1993; Baaz and Fermüller, 1995a; Hähnle, 1998]. Here, we take a more general stance.

The idea of considering distributions of values is encountered in two-valued logic as well: generalized two-valued quantifiers  $\xi$  are obtained from  $Q_\xi : 2^{\mathbf{D}} \rightarrow \{\text{true}, \text{false}\}$  based on the distribution  $\{d \in \mathbf{D} \mid v_{\mathbf{M},\beta_x^d}(\varphi) = \text{true}\}$ .

The present notion of a many-valued quantifier is due to [Rosser and Turquette, 1952, Chapter IV], but appears implicitly already in [Mostowski, 1948]. The simplified definition used in (10) is from [Mostowski, 1961]. The phrase **distribution quantifier** for referring to quantifiers of this kind was coined by Carnielli [1987].

Even more general many-valued quantifiers can be had by refining distributions. Consider, for example, the distribution obtained by regarding  $v_{\mathbf{M},\beta_x^d}(\varphi)$  as a function in  $d$ . In that case, the semantics of a quantifier would be a mapping  $(\mathbf{D} \rightarrow N) \rightarrow N$  and thus could distinguish between different domains and domain elements. This generality is greater than one usually cares for, though.

So-called  $(r, m)$ -ary quantifiers  $(\lambda x_1, \dots, x_r)(\varphi_1, \dots, \varphi_m)$  are based on this idea [Gottwald, 2000]. In principle, to each function  $v : \mathbf{D}^r \rightarrow N^m$  defined as

$$v_{\varphi_1, \dots, \varphi_m}(d_1, \dots, d_r) = \langle v_{\mathbf{M}, \beta_{x_1, \dots, x_r}^{d_1, \dots, d_r}}(\varphi_1), \dots, v_{\mathbf{M}, \beta_{x_1, \dots, x_r}^{d_1, \dots, d_r}}(\varphi_m) \rangle$$

<sup>5</sup>These are the axioms of first-order basic  $t$ -norm logic [Hájek, 1998], see below.

<sup>6</sup>In abuse of notation, the symbol  $\vee$  is used to denote meta-level disjunction as well.

a different truth value is assigned, but distinction between domains (and domain elements) is renounced. Therefore, only with each possible *distribution* (that is: image) of  $v$  a truth value is associated. The distribution of  $v$  for non-empty  $\mathbf{D}$  is a non-empty subset of  $N^m$ . More precisely, the semantics of an  $(r, m)$ -ary quantifier  $\lambda$  is given by a function  $Q_\lambda : (2^{N^m} - \emptyset) \rightarrow N$  and the truth value is computed by

$$(11) \quad v_{\mathbf{M},\beta}((\lambda x_1, \dots, x_r)(\varphi_1, \dots, \varphi_m)) = Q_\lambda(\{v_{\varphi_1, \dots, \varphi_m}(d_1, \dots, d_r) \mid d_i \in \mathbf{D}, 1 \leq i \leq r\}) .$$

For  $r = m = 1$ , equation (10) is obtained.

### 2.3 Algebra

In the terminology of universal algebra (see, for example, [Grätzer, 1979; Cohn, 1981; Meinke and Tucker, 1992]),  $\mathbf{L}_\Sigma^0$  are the elements of the **free term algebra** of  $\mathbf{L}^0$  generated by  $\Sigma$ , and  $\mathbf{A}^0$  is simply a (one-sorted) **abstract algebra** over  $N$  similar to  $\mathbf{L}^0$ . An interpretation maps the generators  $\Sigma$  of  $\mathbf{L}_\Sigma^0$  into  $N$ , and can by standard results (for example, [Meinke and Tucker, 1992, Corollary 3.4.9]) be uniquely extended to a homomorphism from  $\mathbf{L}_\Sigma^0$  to  $N$ . This extension is given by condition (5), which states, of course, homomorphy of  $\mathbf{I}$ .<sup>7</sup>

We need several particular abstract algebras. An **Abelian monoid**  $\langle A, 1, * \rangle$  has a binary operation  $*$ , which is associative and commutative, and neutral with respect to the constant 1. An Abelian monoid is extended to an **Abelian group**  $\langle G, 1, ^{-1}, * \rangle$  by adding a unary operation  $^{-1}$ , which for every  $x \in G$  gives an **inverse element**  $x^{-1}$  such that  $x * x^{-1} = 1$ . It is easy to show that  $x^{-1}$  is uniquely defined. The *binary* operation  $x/y$  is an abbreviation for  $x * y^{-1}$ . The operator  $^{-1}$  is an **involution**, if  $(x^{-1})^{-1} = x$  for all  $x \in G$ .

In addition, monoids and groups can be equipped with a partial order  $\leq$ , with respect to which its operator  $*$  must be *non-decreasing*. A lattice-ordered Abelian group is called an  **$\ell$ -group**, a totally ordered Abelian group is an  **$o$ -group** for short.

If  $*$  is a  $t$ -norm on  $[0, 1]$ , then  $\langle [0, 1], 1, *, \leq \rangle$  is a totally ordered Abelian monoid.

Let  $x^r$  stand for  $\overbrace{x * \dots * x}^{r \text{ times}}$ . An element  $x \in G$  is **idempotent** if  $x^2 = x$ , it is **nilpotent** if  $x^m = 0$  for some  $m$ .

Let  $N$  be a set with partial order  $\leq$ . If  $M \subseteq N$  has the property that  $i \in M$  and  $j \leq i$  ( $j \geq i$ ) imply  $j \in M$ , then  $M$  is a **downset** or **order**

---

<sup>7</sup>Some authors prefer to start with an algebraic framework right away; I refrained from this to keep this chapter accessible to a broad audience.

**ideal** (**upset** or **order filter**) of  $N$ . The collection of order ideals of  $N$  is denoted by  $\mathcal{O}(N)$ .

Obviously, for each  $S \subseteq N$ ,  $\uparrow S = \{x \in N \mid x \geq i \text{ for some } i \in S\}$  and  $\downarrow S = \{x \in N \mid x \leq i \text{ for some } i \in S\}$  are upsets, respectively, downsets of  $N$ . If  $N$  is finite, *all* upsets and downsets are of this form. Instead of  $\uparrow\{i\}$  and  $\downarrow\{i\}$  one writes  $\uparrow i$  and  $\downarrow i$ .

A **lattice** is an abstract algebra of the form  $\mathbf{L} = \langle L, \sqcup, \sqcap \rangle$  equipped with a partial order  $\leq$  such that any two elements  $i, j$  of  $N$  have the unique supremum  $i \sqcup j$  (the **join**) and infimum  $i \sqcap j$  (the **meet**).

A lattice can be defined by join and meet alone; then, one explicitly stipulates that these are associative, commutative, idempotent and absorptive operations. With this in mind we write  $\sqcup\{i_1, \dots, i_n\}$  for  $i_1 \sqcup (i_1 \sqcup (\dots (i_{n-1} \sqcup i_n) \dots))$  and similarly with  $\sqcap$ . Given  $\sqcup$  or  $\sqcap$ , the order can be reconstructed via:  $i \leq j$  iff  $i \sqcap j = i$  iff  $i \sqcup j = j$ . Any finite lattice is **bounded**: there is a (unique) minimal element 0 and maximal element 1 in  $L$ . A lattice is distributive iff for all  $i, j, k \in L$ :  $i \sqcap (j \sqcup k) = (i \sqcap j) \sqcup (i \sqcap k)$ .

A special lattice is the **Boolean set lattice** for a set  $N$ ,  $\mathbf{2}^N = \langle 2^N, \emptyset, N, \cap, \cup \rangle$ , where  $\cap$  is set intersection,  $\cup$  is set union,  $\leq$  is set inclusion.

Let  $\mathbf{L}$  be a lattice and  $I, F \subseteq L$ . If  $i, j \in I$  ( $i, j \in F$ ) imply  $i \sqcup j \in I$  ( $i \sqcap j \in F$ ) and  $I$  is a non-empty order ideal ( $F$  is a non-empty order filter) of  $L$ , then  $I$  is called an **ideal** ( $F$  is called a **filter**) of  $\mathbf{L}$ . For each  $i \in L$ ,  $\downarrow i$  is an ideal of  $\mathbf{L}$ . In finite lattices, all filters and ideals are order filters and ideals of the form  $\uparrow i$ , respectively,  $\downarrow i$ . A non-trivial ideal (filter) of  $\mathbf{L}$  is a **prime ideal** (**prime filter**), if for all  $i, j \in L$ ,  $i \sqcap j \in L$  ( $i \sqcup j \in L$ ) implies  $i \in L$  or  $j \in L$ .

A lattice element  $x \in L$  is called **meet-irreducible** (**join-irreducible**), if

1.  $x \neq \top$  ( $x \neq \perp$ ) and
2.  $x = i \sqcap j$  ( $x = i \sqcup j$ ) implies  $x = i$  or  $x = j$  for all  $i, j \in L$ .

The sets of meet-irreducible and join-irreducible elements of  $\mathbf{L}$  are denoted  $\mathcal{M}(\mathbf{L})$  and  $\mathcal{J}(\mathbf{L})$ , respectively.

## 2.4 Inference

An inference system for a logic tries to capture its valid consequences in a purely syntactical way, and so makes it possible to effectively enumerate them. The most traditional type of inference system is the **Hilbert calculus**. It consists of decidable sets of **axioms** and **rule schemata**. An axiom is a **formula schema**, that is, simply a propositional formula or a first-order formula whose atoms are 0-ary predicates, possibly with a proviso. Examples are (A1–A12). An **instance** of a formula schema is any formula obtained by replacing identical atoms in it with identical

formulas while obeying the proviso (if any). For example, the formula  $(\forall x)(p(y) \rightarrow_t q(x)) \rightarrow_t (p(y) \rightarrow_t (\forall x)q(x))$  is an instance of (A8), but the formula  $(\forall x)(p(x) \rightarrow_t q(x)) \rightarrow_t (p(x) \rightarrow_t (\forall x)q(x))$  is not. A rule schema is a pair  $\langle \Psi, \varphi \rangle$ , where  $\Psi$  is a non-empty set of formula schemata, called **premiss**, and  $\varphi$  is a formula schema, called **conclusion**. Whenever the premiss is a finite set, the rule is called **finitary**. Finitary rules are denoted like

$$(12) \quad \frac{\text{premiss}_1 \quad \cdots \quad \text{premiss}_r}{\text{conclusion}} .$$

A rule instance is obtained by instantiating each of its rule schemata. Let us call a set of axioms and rule schemata over a fixed logical language  $\mathbf{L}$  a **Hilbert calculus**. An example is the calculus  $\mathcal{BL}$ , consisting of axioms (A1–A7), together with a rule schema that is the  $t$ -norm version of **modus ponens**:

$$(13) \quad \frac{\varphi \quad \varphi \rightarrow_t \psi}{\psi} .$$

Each Hilbert calculus  $\mathcal{HK}$  induces a **provability relation**  $\vdash_{\mathcal{HK}}$  between sets  $\Psi$  of  $\mathbf{L}$ -formulas and  $\mathbf{L}$ -formulas  $\theta$  as follows:

1. If  $\theta \in \Psi$  or  $\theta$  is an instance of an axiom of  $\mathcal{HK}$ , then  $\Psi \vdash_{\mathcal{HK}} \theta$ .
2. If  $\Psi \vdash_{\mathcal{HK}} \varphi_i$  for  $1 \leq i \leq r$ , and there is an instance of a rule schema in  $\mathcal{HK}$  with premisses  $\varphi_1, \dots, \varphi_r$  and conclusion  $\theta$ , then  $\Psi \vdash_{\mathcal{HK}} \theta$ .

One abbreviates  $\emptyset \vdash \theta$  with  $\vdash \theta$ . A formula  $\theta$  is **provable** from  $\Psi$  (in  $\mathcal{HK}$ ), if  $\Psi \vdash_{\mathcal{HK}} \theta$ . If  $\Psi = \emptyset$ , then  $\theta$  is simply called provable.

Let  $\mathcal{L}$  be a logic and  $\mathcal{HK}$  a calculus over the same language  $\mathbf{L}$ . One says that  $\mathcal{HK}$  is **sound** (for  $\mathcal{L}$ ), whenever  $\Psi \vdash_{\mathcal{HK}} \varphi$  implies  $\Psi \models \varphi$ .  $\mathcal{HK}$  is **complete**, if all valid formulas of  $\mathcal{L}$  are provable in  $\mathcal{HK}$ . It is **strongly complete**, if  $\Psi \models \varphi$  implies  $\Psi \vdash_{\mathcal{HK}} \varphi$  for all (sets of)  $\mathbf{L}$ -formulas.

Example 3 and a straightforward induction shows soundness of  $\mathcal{BL}$  for each  $t$ -norm logic  $\mathcal{L}_t^0$ . Conversely, it can be shown [Cignoli *et al.*, 2000] that  $\mathcal{BL}$  is complete for the intersection of all  $t$ -norm logics: if a  $\mathbf{L}_t^0$ -formula  $\varphi$  is valid in all logics  $\mathcal{L}_t^0$ , then  $\vdash_{\mathcal{BL}} \varphi$ .

A logic, for which one has a sound and complete calculus is said to be **axiomatizable**. If, moreover, the calculus consists of a finite set of (first-order) axioms and finitary (first-order) rule schemata, its logic is **finitely (first-order) axiomatizable**.

Hilbert calculi provide crisp syntactic characterizations of many logics. For example, a sound and complete axiomatization of Gödel logic  $\mathcal{L}_G^0$  is obtained [Dummett, 1959] by adding the single axiom schema

$$(G) \quad \varphi \rightarrow_G (\varphi \odot_G \varphi)$$

to  $\mathcal{BL}$  (and replacing indices  $t$  with  $G$  in the other schemas). Similar results exist for the other famous  $t$ -norm logics (see also Section 4 below). A sound and complete calculus  $\mathcal{CL}$  for classical propositional logic  $\mathcal{L}_c^0$ , by the way, is obtained from  $\mathcal{BL}$  by adding the **non tertium datur** [Hájek, 1998]:

$$(C) \quad \varphi \vee \neg\varphi.$$

In classical logic strong completeness can be reduced to completeness by way of the **deduction theorem**: if  $\Psi$  is any set of  $\mathbf{L}_c^0$ -formulas and  $\varphi, \rho \in \mathbf{L}_c^0$ , then  $\Psi \cup \{\rho\} \vdash_{\mathcal{CL}} \varphi$  iff  $\Psi \vdash_{\mathcal{CL}} \rho \rightarrow \varphi$ . Since in any deduction only a finite number of formulas from the premiss are used this allows to “shuffle” all required premisses to the right-hand side. In many-valued logics the deduction theorem does only hold in modified form, and sometimes not at all. For example,  $\varphi \vdash_{\mathcal{BL}} (\varphi \odot_t \varphi)$ , but  $\not\vdash_{\mathcal{BL}} \varphi \rightarrow_t (\varphi \odot_t \varphi)$ . Let  $\varphi^m$  stand for  $\varphi \odot_t \cdots \odot_t \varphi$  ( $m$  copies of  $\varphi$ ). Then at least the following version of the deduction theorem holds in  $\mathcal{BL}$ :

**THEOREM 9** ([Hájek, 1998]). *If  $\Psi$  is any set of  $\mathbf{L}_t^0$ -formulas and  $\varphi, \rho \in \mathbf{L}_t^0$ , then  $\Psi \cup \{\rho\} \vdash_{\mathcal{BL}} \varphi$  iff  $\Psi \vdash_{\mathcal{BL}} \rho^m \rightarrow \varphi$  for some  $m \in \mathbb{N}$ .*

For the particular  $t$ -norm logic of Łukasiewicz the theorem was proven already in [Pogorzelski, 1964]. This improved by giving a concrete upper bound for the number  $m$ , depending on  $\Psi, \varphi, \rho$  (an exponent in the number of variable occurrences in the formulas) in [Ciabattoni, 2000b; Aguzzoli and Ciabattoni, 2000]. For the  $t$ -norm logic of Gödel  $m = 1$  is sufficient, that is, the classical deduction theorem holds for Gödel logic.

Another version of completeness, so-called **Pavelka style completeness** assumes that a logic  $\mathcal{L}$  is expressive enough to characterize **partial truth** or **graded truth**: for each formula  $\varphi \in \mathbf{L}$  and truth value  $i \in N$  there must be a formula  $\varphi_i \in \mathbf{L}$  such that for all interpretations  $\mathbf{I}$ ,  $\mathbf{I}(\varphi_i) \in D$  iff  $\mathbf{I}(\varphi) \geq i$ . In the case  $N = [0, 1]$  and  $D = \{1\}$ , for instance, it is sufficient, if the unary connective  $\uparrow i$  is definable for all rational  $i = \frac{d}{m} \in N$  with  $d, m \in \mathbb{N}$ :

$$(14) \quad \uparrow i(j) = \begin{cases} 0 & 0 \leq j < \frac{d-1}{m} \\ m \cdot j - d + 1 & \frac{d-1}{m} \leq j < i \\ 1 & i \leq j \leq 1 \end{cases}$$

In Łukasiewicz logic the existence of  $\uparrow i$  follows directly from McNaughton’s Theorem. Explicit constructions of similar formulas can be found in [Mundici and Olivetti, 1998; Hájek, 1998].

Even if  $\Psi \vdash \varphi$  does not hold, it may well be the case that  $\Psi \vdash \varphi_i$  holds for some  $i \in N$ . It is natural to define the **provability degree**

$|\varphi|_{\Psi} = \sup\{i \mid \Psi \vdash \varphi_i\}$  as an upper bound on the level of truth for which  $\varphi$  can be proven from  $\Psi$ .

On the other hand, for any model  $\mathbf{I}$  of  $\Psi$ , the value  $i = \mathbf{I}(\varphi)$  says that  $\varphi$  is an  $\{i\}$ -consequence of  $\Psi$  at most. The **truth degree**  $\|\varphi\|_{\Psi} = \inf\{\mathbf{I}(\varphi) \mid \mathbf{I} \text{ model of } \Psi\}$  is a lower bound on the validity of  $\varphi$  relative to  $\Psi$ . An axiomatization is **Pavelka complete**, if provability degree and truth degree coincide, that is,<sup>8</sup>

$$(15) \quad \|\varphi\|_{\Psi} \leq |\varphi|_{\Psi}$$

for all  $\Psi$  and  $\varphi$ . Pavelka-completeness of Łukasiewicz logic was shown by Pavelka [1979a; 1979b; 1979c] after whom the concept is named, in simplified form in [Hájek, 1998]. The tool used in this investigation, **rational Pavelka logic**  $\mathcal{L}_{\text{RPL}}^0$ , is an extension of Łukasiewicz logic, where rational constants are built into the language. It is discussed in Section 4.3 below.

Pavelka completeness breaks down in general for logics with non-continuous connectives, such as Goguen implication of product logic [Hájek, 1998, 4.1.22]. Also from  $\|\varphi\|_{\Psi} = i$  one can, in general, not deduce  $\Psi \vdash \varphi_i$ : consider  $\Psi = \{\uparrow i(p) \mid i < 1\}$  and  $\varphi = p$ , for which  $\|\varphi\|_{\Psi} = |\varphi|_{\Psi} = 1$ , but  $\Psi \not\vdash \varphi_1$ . For finite  $\Psi$ , there are positive results.

**Internal and External Calculi** Some authors find it convenient to distinguish between external and internal calculi [Hähnle and Escalada-Imaz, 1997; Ciabatonni, 2000b; Baaz *et al.*, to appear, 2000]:

**Internal calculi:** the objects constructed during a proof are from the same logical language as the goal to be proven; a typical example are Hilbert calculi.

**External calculi:** the objects occurring during a formal proof are from an *extended* language that may involve elements from the semantics such as designators for truth values, worlds or even non-logical expressions such as constraints; a typical example are the signed calculi developed in Section 6.3.

Proof theorists often only accept calculi of the first kind and regard the second option as a kind of “cheating”. On the other hand, if a uniform and computationally efficient treatment of deduction is desired, there seems to be no alternative to external calculi: otherwise, highly indeterministic rules are inevitable, even if an internal axiomatization exists.

A somewhat extreme position of gaining a classical logic approach to deduction in non-classical logic would be to formulate the “external” elements in the second approach as a meta theory in classical logic. For a

<sup>8</sup>If  $\vdash$  is sound, then also  $\|\varphi\|_{\Psi} \geq |\varphi|_{\Psi}$ .

wide range of logics this is even possible in first-order logic. The “meta theory” of finite-valued logic in particular can always be captured without having to move to a higher-order stage. From the viewpoint of efficiency, however, this is a disastrous strategy. In fact, this approach was used to create challenge problems for first-order theorem provers [Anantharaman and Bonacina, 1990].

### 2.5 *What is the Meaning of Truth Values?*

A frequently heard objection against MVL is the missing “logical meaning” of truth values; more precisely, I was sometimes asked how, in general, the structure of the truth value set  $N$  is reflected in the provability relation  $\vdash$ . There is, however, no meaningful relation, in general.<sup>9</sup> This is closely related to another objection: that a useful distinction of what is and what is not a many-valued logic is lacking in the first place. Indeed, as is pointed out in Section 2.2 of Urquhart’s chapter in this Volume, just about *any* notion of logical consequence can be represented within the framework of many-valued logic, that is, matrix semantics [Wójcicki, 1988]. On the other hand, any attempt undertaken so far to restrict many-valued logic to certain classes of matrices lead to exclusion of natural examples. Thus we have the situation that not every instance falling into the framework of matrix semantics can be *naturally* considered as a many-valued logic while, at the same time, a useful restriction is elusive.

I must admit that I cannot provide a natural and non-trivial definition of MVL either, but neither do I follow the conjecture expressed in Section 5.5 of Urquhart’s chapter in this Volume that the lack of such a definition implies that “Łukasiewicz’s many-valued systems have remained somewhat marginal to the mainstream of logical research.”

If this were true, I think many other classes of logical systems would have to be considered as marginal as well.

To explicate this point, consider the case of modal logic, see Bull and Segerberg’s chapter in Volume 3 of this *Handbook*. It is trivial to re-interpret any many-valued logic as a *two-valued* modal logic (see Section 10) even if this is, of course, completely unintuitive in most cases. Or one could make also a case against, say, substructural logics, which are defined solely by proof theoretic means, and often have only an awkward model semantics (as an arbitrary example, consider non-commutative linear logic). But what is a substructural logic? If the answer is “anything weaker than classical logic and obtained by imposing restrictions upon the structural rules of sequent calculi”, then one could stipulate absurd things, say, requiring every other contracted formula to be of opposite polarity. So are modal and substructural logic marginal to the mainstream of logical research?

---

<sup>9</sup>Perhaps one should not speak of *truth values* at all, but only of *values*, however, I chose to follow tradition here.

I think not, and because no one cares too much about the boundaries of an area provided that sufficiently many and rich results exist that clearly fall within it. And from this point of view, the scale tilts in favor of MVL in the past decades. I hope to demonstrate this in the remaining part of the present chapter, but I want to seize the opportunity to present some concrete usages of MVL right here:

Underlying the notion of a  $t$ -norm (p. 301) is the **truth ordering** or **certainty ordering** of  $N$ , which identifies 0 with the least true or certain value and 1 with the most true or certain value. Let us write  $F$  for 0 and  $T$  for 1, if this is the intended meaning, see left part of Figure 3. In the vast majority of applications of  $t$ -norm-based logics, the order is assumed to be total, see Section 5.

A fundamentally different interpretation of truth values, found in AI and programming, is the **knowledge ordering** or **information ordering**, where 0 means to know nothing at all and 1 to know everything (which could actually be too much . . . ). Let us write  $\perp$  for 0 and  $\top$  for 1, if this is the intended meaning, see the central part of Figure 3.

The distinction between these two usages of truth values is also stressed in philosophical treatments of non-classical logic [Haack, 1996, p. 113], where our second kind of interpretation is referred to as *epistemological uncertainty*.

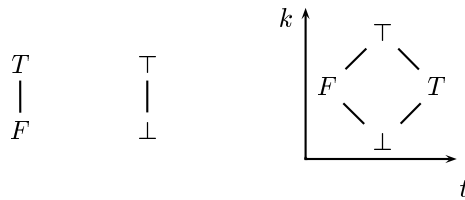


Figure 3. Truth ordering, knowledge ordering, and the bilattice *FOUR*.

It can be useful to have both orderings present in an MVL at the same time. Following [Belnap Jr., 1977; Fitting, 1991a], imagine we have a set of distributed agents working on the same problem. The two-valued answer from each of these agents is from the set  $\{F, T\}$  that is ordered by truth. What, if we have no answer from any agent, or differing answers from two or more agents? In the first case, we know nothing, thus we should assign the truth value  $\perp$  to model this situation; in the latter case, we assign  $\top$  to model total, here even inconsistent, knowledge. The uniquely defined values  $F$  and  $T$  appear in between. The resulting **knowledge diamond lattice** is depicted in the right part of Figure 3. This lattice and the many-valued logics induced by it are investigated in many papers, starting perhaps with [Lukasiewicz, 1957] who thought of the resulting logic as a *modal logic* and



was trying to capture Aristotle's Modal Syllogistic with it. Lukasiewicz considered an implication obtained by residuation of the lattice infimum as well as several unary connectives for adding/deleting/swapping support of truth/falsity. Both, [Lukasiewicz, 1957] and [Belnap Jr., 1977], have a negation connective that swaps  $\perp$  with  $\top$  and  $F$  with  $T$ .

The knowledge diamond lattice is particularly useful to capture paraconsistent reasoning, that is, making useful deductions in the presence of inconsistency [Belnap Jr., 1977; Lu *et al.*, 1991]. Further uses are discussed in Section 7.5.

The knowledge lattice becomes the truth diamond lattice, if rotated by 90 degrees, counterclockwise. It turns out that such interlaced lattices, where both kinds of ordering are present, can be generalized from the four-valued case:

**DEFINITION 10** ([Ginsberg, 1988; Fitting, 1991a]). A **bilattice** is an abstract algebra of the form  $\langle B, \sqcup, \sqcap, \oplus, \otimes \rangle$  equipped with two partial orders  $\leq_t$  and  $\leq_k$  such that:

- $\langle B, \sqcup, \sqcap \rangle$  with  $\leq_t$  and  $\langle B, \oplus, \otimes \rangle$  with  $\leq_k$  are complete lattices,
- $\sqcup, \sqcap$  are monotone with respect to  $\leq_k$  and  $\oplus, \otimes$  are monotone with respect to  $\leq_t$ .

A **bilattice with negation** in addition contains an involution  $^{-1}$  which preserves the knowledge order and reverses the truth order.

Truth values are sometimes motivated by application domains and have technical rather than logical meaning. Here is an example, taken from [Hayes, 1986; Hähnle and Kernig, 1993]: An MOS transistor has different signal strengths at source and drain terminals, due to a physical effect called degradation. We model this using a seven-valued logic with  $N = \{F, T, \tilde{F}, \tilde{T}, \top, \tilde{\top}, \perp\}$  and the ordering  $\leq$  depicted in Figure 4. Values  $F$  and  $T$  represent full strength signals, while  $\tilde{F}$  and  $\tilde{T}$  represent the degraded signals. In a faulty circuit, each of those may clash at a node resulting in undefined values  $\top$  and  $\tilde{\top}$ . The value  $\perp$  represents unconnected nodes (“no signal”). In this setup, a node that connects two signals  $x$  and  $y$  is simply computed by taking the join in the lattice induced by  $\leq$ . Note that we have essentially two knowledge diamond lattices stacked on top of each other. At the same time each degraded signal is below each full strength signal. MOS transistors are modeled by propositional connectives in this logic whose semantics is determined by the technical dimensions of NMOS and PMOS transistors. There is no natural algebraic or proof theoretical characterization of the resulting logic which, at the same time, is obviously rather useful. It is examples like this that justify a generic approach to many-valued logic as defended in the present article.

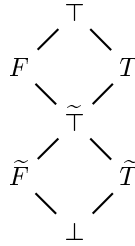


Figure 4. Seven-valued lattice used for modeling MOS transistors.

An example from a quite different domain is found in [Kerber and Kohlhasse, 1996], where a four-valued logic is used to model the phenomenon of presuppositions in natural language semantics.

Even truth value sets without any natural order have important applications, which makes it worth while to study their computational proof theory (Section 6). Consider an undirected finite graph  $G = (V, E)$  with vertices  $V$  and edges  $E \subseteq \{S \subseteq V \mid |S| = 2\}$ . Each vertex  $v \in V$  has a color  $c(v) \in C$ , where  $C$  is a finite set of colors. The well-known **graph colorability problem** asks whether there is a function  $c : V \rightarrow C$  such that  $|\{c(v), c(w)\}| = 2$  for each  $\{v, w\} \in E$ . This can be easily modeled in a  $C$ -valued logic [Manyà, 1996]. The basic idea is to represent vertices  $v$  by propositional variables  $p_v$  and to exploit that  $p_v$  is  $C'$ -satisfiable iff  $v$  can be colored by one of the colors in  $C' \subseteq C$ . Many-valued logic satisfiability procedures as outlined below in Section 6 perform quite favorably on such problems, if compared to traditional approaches [Béjar and Manyà, 1999c]. More generally, satisfiability checking in finite-valued logic looks like an interesting alternative to mixed integer programming or constraint solving when tackling combinatorial optimization problems over finite domains.

Finally, I want to mention two uses of many-valued logic in philosophical arguments: first, the heap paradox or **Sorites paradox** (see, for example, [Cargile, 1969]) and its resolution in Lukasiewicz logic. The paradox, in one form, goes:

- (A) One grain of sand is certainly not a heap;
- (B<sub>*i*</sub>) adding just one grain of sand to  $i$  many grains that are not yet a heap ( $NH_i$ ), does not result in a heap;
- (C) 100,000 grains of sand are a heap.

Taking (A), (B<sub>*i*</sub>)<sub>1 ≤ *i* ≤ 99,999</sub>, and (C) together is classically inconsistent; modifying (B<sub>*i*</sub>) to the extent that going from, say, 27,000 to 27,001 grains of sand produces a heap, seems implausible, just like dropping either (A)

or (C). In Łukasiewicz logic one can resolve this paradox by admitting  $(B_i)$  to be somewhat less than true, like in  $\uparrow_{i_0}(NH_i \rightarrow_L NH_{i+1})$ , using the connective (14) and Łukasiewicz implication. Now one has models provided that  $1/(1-i_0) < 99,999$ . Application of  $(B_i)$ , if repeated often enough, “uses up” the trust put into the conclusion and thus resolves the paradox. This modeling is criticized in Section 4.4 of Urquhart’s chapter in this Volume of this Handbook for the “artificiality” of the truth value  $i_0$ , but at least the relation that  $r$  must have to the number of grains is quite clear, even though not independent from it.

More important is a recent result [Hájek and Paris, 1997; Hájek *et al.*, 2000] on the logical formalization of the **liar paradox** (“the sentence I am just stating is false”) based on  $\mathcal{L}_L$  and using a many-valued truth predicate. It is shown that this idea leads to a consistent definition of the truth predicate in Peano Arithmetic, which is well-known to be impossible in classical logic. Interestingly, arithmetic can be kept classical in this setup, only the definition of truth must be many-valued.

### 3 ALGEBRAIC PERSPECTIVE

We remarked already that each propositional matrix  $\mathbf{A}^0$  for a language  $\mathbf{L}^0$  is an abstract algebra, Example 2 lists several concrete instances. It is well-known that the matrix  $B_2 = \langle \mathbf{2}, 0, \min \rangle$  of classical propositional logic, identical to the **two-element Boolean algebra** up to notation, plays a special role: a formula  $\varphi \in \mathbf{L}_c^0$  is valid in *the particular* algebra  $B_2$  iff it is valid in *all* Boolean algebras. This **algebraic completeness** has, among others, three important advantages:

1. it provides an abstract, algebraic characterization of validity;
2. the problem of checking validity in a class of algebras can be reduced to the problem of checking validity in one so-called **canonical algebra**;
3. it often goes a long way towards proving completeness of a particular axiomatization (this was the original motivation for the definition of MV-algebras in [Chang, 1959], see below).

It is natural to try to extend algebraic results to many-valued logics. Among the first landmarks were Chang’s [1958] MV-algebras, which are complete with respect to the matrix of Łukasiewicz logic, that is, the canonical MV-algebra. This led to a drastically simplified and shortened completeness proof for the axiomatization of Łukasiewicz logic compared to [Rose and Rosser, 1958].

Thus it is fruitful to investigate the algebraic structure of (many-valued) logics, not only for its own sake. In the present section the main results in

algebraic treatments of many-valued logics, an area that has been thriving recently, are given.

As so many important logics turn out to be continuous  $t$ -norms, the latter structures are the most thoroughly investigated. Recall the observation from Section 2.3 that each  $t$ -norm  $*$  induces a lattice-ordered Abelian monoid  $\langle L, 1, *, \leq \rangle$ . Conversely, such a structure falls short of being a  $t$ -norm in two respects (ignoring total order for the moment): first, in general, the upper bound of the lattice induced by  $\leq$  on  $L$  needs not to be 1; second, there is no residuated implication (see Example 2). This motivates the following definition.

**DEFINITION 11.** Let  $\langle L, 1, *, \leq \rangle$  be a lattice-ordered Abelian monoid. The abstract algebra of the form  $\langle L, 0, 1, \sqcap, \sqcup, *, \Rightarrow \rangle$  is a **residuated lattice** (RL) provided that 0, 1,  $\sqcap$ ,  $\sqcup$  are minimal element, maximal element, infimum, and supremum, respectively, of the lattice induced by  $\leq$  on  $L$  and, moreover,

$$(16) \quad i \leq (j \Rightarrow k) \quad \text{iff} \quad i * j \leq k$$

for all  $i, j, k \in L$ . The operations  $*$  and  $\Rightarrow$  are said to form an **adjoint pair**.

One can show that the adjointness condition actually restricts  $t$ -norms  $*$  to those that are left continuous (in the usual sense) in both arguments. In a similar manner, full continuity of  $t$ -norms can be characterized algebraically:

**DEFINITION 12.** A residuated lattice  $\langle L, 0, 1, \sqcap, \sqcup, *, \Rightarrow \rangle$  is **divisible** iff for all  $i, j \in L$  with  $i \leq j$  there exists some  $k \in L$  such that  $i = j * k$ .

An alternative characterization of divisibility in residuated lattices is

$$(17) \quad i \sqcap j = i * (i \Rightarrow j)$$

for all  $i, j \in L$ . In other words, the lattice operators can be recovered from an adjoint pair alone (the join operation is recovered as  $i \sqcup j = ((i \Rightarrow j) \Rightarrow j) \sqcap ((j \Rightarrow i) \Rightarrow i)$ ). This justifies definitions (3) and (4). Divisible residuated lattices are, therefore, usually introduced in the form  $\langle L, 0, 1, *, \Rightarrow \rangle$ .

A residuated lattice determined by a  $t$ -norm  $*$  is divisible iff  $*$  is continuous. Everything that remains to characterize continuous  $t$ -norms is linearity:

**DEFINITION 13.** A residuated lattice  $\langle L, 0, 1, \sqcap, \sqcup, *, \Rightarrow \rangle$  satisfies **prelinearity** iff

$$(18) \quad (i \Rightarrow j) \sqcup (j \Rightarrow i) = 1$$

for all  $i, j \in L$ .

Note that this condition is somewhat weaker than stipulating that  $L$  be totally ordered, which would amount to saying that one of  $i \sqcap j = i$  or

$i \sqcap j = j$  (equivalently, one of  $i \sqcup j = i$  or  $i \sqcup j = j$ ) holds for all  $i, j \in L$ . In a **prelinear residuated lattice**,  $\sqcup$  is definable, and (18) is sometimes called *proof by case distinction*.

In each residuated lattice, one can define a unary negation operator “ $-$ ” by taking  $-i = i \Rightarrow 0$  for all  $i \in L$  (in hindsight, this justifies (1)). Now  $\langle L, 0, *, - \rangle$  is an Abelian group, so an **involutive** residuated lattice structure is one, where the negation operator is an involution (see Section 2.3). One part of the equation characterizing involutions,  $--i \leq i$ , holds in any residuated lattice.

Altogether, there are now four useful restrictions of residuated lattices, which can be combined in various ways: prelinearity (18), divisibility (17), involutive negation, and conflating the  $t$ -norm operator  $*$  with the lattice operator  $\sqcap$ . The latter is succinctly expressed in the equation

$$(19) \quad i^2 = i .$$

Adding involutive negation to residuated lattices, respectively, identification of  $*$  and  $\sqcap$  yields well-known structures: Girard monoids, the standard semantics of linear logic [Girard, 1987], respectively, Heyting algebras, the algebras that characterize intuitionistic logic (see van Dalen’s chapter in Volume 7 of this *Handbook*). Prelinear residuated lattices are investigated in [Esteva and Godo, 1999] under the name of **QBL-algebra** (shorthand for *quasi-BL-algebra*) and divisible, prelinear residuated lattices were baptized **BL-algebra** by [Hájek, 1998] (BL stands for “basic  $t$ -norm logic”).<sup>10</sup>

Do BL-algebras capture the tautologies of continuous  $t$ -norms? Clearly, each continuous  $t$ -norm determines a BL-algebra on  $[0, 1]$ , a so-called  **$t$ -algebra**. But in fact  $t$ -algebras can prove no more tautologies than BL-algebras, in other words, BL-algebras are complete with respect to  $t$ -algebras [Cignoli *et al.*, 2000]. Thus the label “basic  $t$ -norm logic” is fully justified, because BL-algebras capture the intersection of all logics based on continuous  $t$ -norms over  $[0, 1]$ .

Let us discuss some of the axioms of basic  $t$ -norm logic (A1–A7) in this light. For adjoint pairs based on a  $t$ -norm  $A_{\odot_t}$ , residuated implication characterizes the order on  $N$ :

$$(20) \quad i \leq j \text{ iff } A_{\rightarrow_t}(i, j) = 1 \text{ for all } i, j \in N$$

For the truth values  $\{0, 1\}$ ,  $\rightarrow_t$  behaves like classical implication. Together, this accounts for BL axioms (A5a–b). Further, (A1) expresses transitivity of the truth value ordering, (A3) is commutativity of  $A_{\odot_t}$ . Axiom (A2) is a direct consequence of  $A_{\odot_t}$  being non-decreasing.

<sup>10</sup>In [Hájek, 1998] and related publications “basic  $t$ -norm logic” is just called “basic logic”; in accordance with [Gottwald, 2000] I use the clarification “basic  $t$ -norm logic” to avoid confusion with other systems of “basic logic”. Likewise, “prelinear residuated lattice” is used henceforth, rather than “QBL-algebra”.

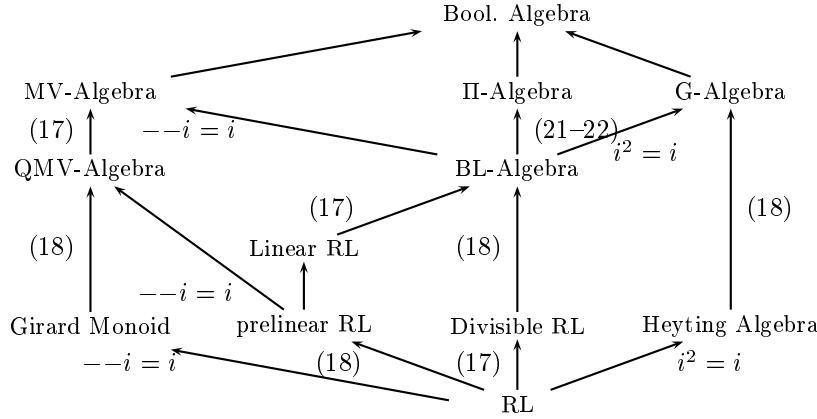


Figure 5. Hierarchy of some algebraic structures related to  $t$ -norms

We remarked already on page 301 that each of Łukasiewicz, Gödel, and product logic is a continuous  $t$ -norm logic. It is not surprising, therefore, that the algebras of these logics can be obtained as extensions of BL-algebra.

Let us start with Łukasiewicz logic. An **MV-algebra** is simply an involutive BL-algebra. Alternatively, one may extend prelinear residuated lattices with an involution to yield a **QMV-algebra** [Esteva and Godo, 1999], and then obtain MV-algebras with divisibility.

Recall that in Gödel logic the  $t$ -norm coincides with the infimum on the totally ordered lattice  $[0, 1]$ . Hence, **G-algebras** are either obtained as linearly ordered Heyting algebras or, alternatively, as those BL-algebras satisfying (19).

Unfortunately,  **$\Pi$ -algebras**, corresponding to product logic, are less intuitive. It turns out that a  $\Pi$ -algebra is a BL-algebra, where

$$(21) \quad i \sqcap -i = 0$$

$$(22) \quad --k \leq (i * k \Rightarrow j * k) \Rightarrow (i \Rightarrow j)$$

hold. Equation (21) ensures that the fact  $i > 0$  is indicated by  $-i = 0$ , hence  $--i = 1$ . Then, equation (22) is sufficient to ensure: if  $k > 0$ , then  $i * k = j * k$  implies  $i = j$ . This divisibility by non-zero elements is just what characterizes product algebras (recall that  $A_{\odot \Pi}$  was multiplication on  $[0, 1]$ ).

The combination of any two of MV-,  $\Pi$ -, or G-algebras yields Boolean algebra. On the other hand, merely adding “**tertium non datur**”  $i \sqcup -i$  to BL gives Boolean algebra as well (see also page 312). The picture so far is summarized in Figure 5.

All classes of algebras mentioned so far have been finitely axiomatized: residuated lattices, that is, the logic of left-continuous  $t$ -norms under the label **monoidal logic** by Höhle [1995]; prelinear residuated lattices and QMV-algebras in [Esteva and Godo, 1999], the axiomatization of BL-algebras is **basic  $t$ -Norm Logic** [Hájek, 1998]; Girard monoids (linear logic) and Heyting algebras (intuitionistic logic) were mentioned already; MV-algebras characterize the axioms of Łukasiewicz logic [Chang, 1959], see also Section 4; G-algebras were shown to be sound and complete for Gödel logic in [Dummett, 1959], while for the more recent  $\Pi$ -algebras and product logic the same was established in [Hájek *et al.*, 1996]. An overview of recent results for  $t$ -norm-based logics is contained in [Gottwald, 2000].

Let us come back now to the question of algebraic completeness raised at the beginning of this section. We mentioned already that the matrix of Łukasiewicz logic is the canonical MV-algebra [Chang, 1959]; analog results hold for the matrix of Gödel logic and G-algebras [Dummett, 1959], product logic and  $\Pi$ -algebras [Hájek *et al.*, 1996]. The weaker structures seem not to have canonical representations, although the result of [Cignoli *et al.*, 2000], that it is sufficient to consider BL-algebras induced by continuous  $t$ -norms, goes some way towards a canonical representation.

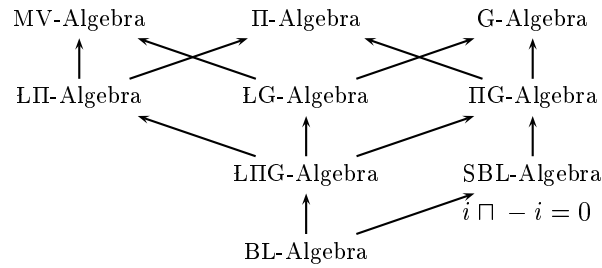


Figure 6. Algebraic structures between BL-algebras and MV-,  $\Pi$ -, and G-algebras

The hierarchy depicted in Figure 5 can be refined in various directions of which we mention two possibilities:

1. In [Cignoli *et al.*, 2000] the intersections between MV-,  $\Pi$ -, or G-algebras were investigated, see Figure 6. All algebras can be obtained from BL-algebra by adding certain equations. It turns out that stipulating  $i \sqcap -i = 0$  for all  $i$  takes one from LPIG-Algebra to PIG-Algebra. In residuated lattices, this determines “ $-$ ” to be Gödel negation. Adding this equation alone gives SBL-algebras, which correspond to  $t$ -norm-based logics with Gödel negation [de Baets *et al.*, 1999; Esteva *et al.*, 2000].

2. In [Hájek, 1998] the language of  $t$ -norm-based logics is extended with certain connectives, and the resulting algebras are investigated.

(a) The unary connective  $\Delta$  has the semantics

$$(23) \quad A_{\Delta}(i) = \begin{cases} 1 & \text{if } i = 1 \\ 0 & \text{otherwise} \end{cases}$$

and, therefore, states that its argument has a designated truth value. It is investigated in [Baaz, 1996; Hájek, 1998]. Resulting algebras are called  $L_{\Delta}$ -algebra, etc. It is noteworthy that with the help of  $\Delta$  one can define Gödel implication and negation.

(b) Logical constants of the form  $\bar{r}$ , where  $r$  is a rational number from  $[0, 1]$  and  $A_{\bar{r}} = r$ . Adding such constants to Łukasiewicz logic, results in **rational Pavelka logic** [Pavelka, 1979a; Pavelka, 1979b; Pavelka, 1979c], which is further discussed in Section 4. Rational Gödel logic, obtained in the same way, is discussed in [de Baets *et al.*, 1999].

Post algebras [Rasiowa, 1974] are parameterized with  $n$  and thus are difficult to compare with monoidal logics. Finite Łukasiewicz logics become functionally complete when logical constants for all truth values are added and are then indistinguishable from Post algebras. Infinite Post algebras [Rasiowa, 1973] are usually of order type  $\omega^+$  (a chain isomorphic to the natural numbers topped by an infinite element  $\omega$ ) and cannot be directly compared either.

It should be mentioned that a different, and more general, notion of algebraization of logics than the traditional approach employed here was developed in [Blok and Pigozzi, 1989; Blok and Pigozzi, to appear, 2000]. This has been applied, among other systems, to various many-valued logics [Ángel J. Gil *et al.*, 1997].

An overview of some algebraic structures related to many-valued logics is [Iturrioz *et al.*, 2000].

## 4 LUKASIEWICZ LOGIC

A wide range of rather deep results has been established for Łukasiewicz logic in the last decade or so, and, without doubt, it is the most intensely researched many-valued logic. Let us start this section with a classical result.

### 4.1 McNaughton's Theorem

Denote with  $\mathbf{I}_{p_1, \dots, p_r}^{i_1, \dots, i_r}$  an interpretation that fixes  $\mathbf{I}(p_j) = i_j$  for  $1 \leq j \leq r$ . Any propositional formula  $\varphi$  over  $r$  variables, say,  $p_1, \dots, p_r$ , determines for



any given logic in a natural way a function  $f_\varphi : N^r \rightarrow N$  via  $f_\varphi(i_1, \dots, i_r) = \mathbf{I}_{p_1, \dots, p_r}^{i_1, \dots, i_r}(\varphi)$  for all  $i_1, \dots, i_r \in N$ . Each  $\mathbf{L}_L^0$ -formula  $\varphi$ , in particular, determines for Łukasiewicz logic a function  $f_\varphi : [0, 1]^r \rightarrow [0, 1]$ . It is easy to prove (or see [Cignoli *et al.*, 1999, 3.1.8]) that for every  $\varphi \in \mathbf{L}_L^0$  over  $r$  variables,  $f_\varphi$  has the following properties:

1.  $f_\varphi$  is continuous;
2. there is a finite set  $\mathcal{P}$  of linear polynomials in  $r$  variables over  $[0, 1]$  with integral coefficients such that for each  $\vec{v} \in [0, 1]^r$ , there is a polynomial  $P \in \mathcal{P}$  with  $f_\varphi(\vec{v}) = P(\vec{v})$ .

McNaughton [1951] showed that, conversely, for every function  $f : [0, 1]^r \rightarrow [0, 1]$  satisfying the above properties, there is a formula  $\varphi$  of Łukasiewicz logic such that  $f = f_\varphi$ . McNaughton originally gave an indirect argument, but as shown in [Mundici, 1994], the formula  $\varphi$  can be effectively constructed from  $f$ . Mundici's proof was simplified in [Aguzzoli, 1999].

In the following I elaborate a little on McNaughton's Theorem, because it is a good place to see how familiar and straightforward classical concepts unfold in a surprisingly complex and rich manner against a many-valued background.

The classical counterpart of McNaughton's Theorem is the well-known fact that for every function  $b : \{0, 1\}^r \rightarrow \{0, 1\}$  there is a  $\mathbf{L}_c^0$ -formula  $\varphi$  such that  $b_\varphi = b$ , in other words, every  $r$ -ary two-valued function can be represented by a suitable propositional formula over  $r$  variables. There are many ways to compute  $\varphi$  for a given  $b$ . Perhaps the most straightforward method is as follows: let  $\text{ON}(b) = \{\vec{v} \mid b(\vec{v}) = 1\}$  be the values for which  $b$  is 1, the **ON-set** of  $b$ ; let  $p_1, \dots, p_r$  be propositional variables corresponding to the arguments of  $b$ . The elements of  $\text{ON}(b)$  are easily characterized: let  $C(\vec{v}) = \bigwedge_{j=1}^r L_j$ , where  $L_j = p_j$ , if the  $j$ -th component of  $\vec{v}$  is 1, and  $L_j = \neg p_j$  otherwise. Then one takes as  $\varphi$  simply the formula  $\bigvee_{\vec{v} \in \text{ON}(b)} C(\vec{v})$ , which enumerates all argument vectors of  $b$ , where  $b$  has value 1.

The formula  $\varphi$  one obtains from this construction is in disjunctive normal form. The literals  $L_j$  correspond to primitive functions by means of which the desired  $b$  is composed.

For finite-valued logics, often the same technique can be used. One needs to consider not merely one ON-set, but  $|D|$  many such sets, one for each designated truth value, see also Section 6. This works at least when a logic is expressive enough to specify that a variable  $p$  must evaluate to a truth value  $i$  for some  $i \in D$ .

In the infinite case, the enumeration strategy would lead to a formula of infinite size, because  $\text{ON}(f)$  is in general an infinite subset of  $[0, 1]^r$ . Even so, the idea of decomposing  $f$  into a normal form, whose primitive functions  $h$  one knows how to represent, can be fruitfully applied here as well.

It is fairly easy to find a  $\mathbf{L}_L^0$ -formula  $\varphi$  such that  $f = f_\varphi$ , if  $f$  is a linear polynomial with integer coefficients [Rose and Rosser, 1958]. The reason is that linear polynomials with integer coefficients can be defined by addition alone, for example,

$$m \cdot i = \overbrace{i + \dots + i}^{m \text{ times}},$$

and addition is, up to truncation to  $[0, 1]$ , present in the form of the connective  $\oplus_L$ . As a consequence, it is easy to represent convex polyhedra, defined by polynomials with integer coefficients.

It is a tempting idea to try to combine a McNaughton function  $f$  disjunctively from the polyhedral cones defined by its non-differentiable parts, but this does not work. Assume, we wanted to find a formula representation of the unary function depicted on the left in Figure 7. The polyhedral cone originating in  $(\frac{1}{3}, \frac{2}{3})$  is larger than the function value below the dashed line; one needs other, more primitive building blocks to compose McNaughton functions.

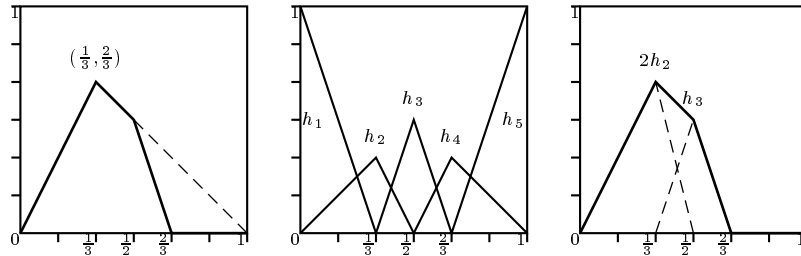


Figure 7. Construction of one-dimensional McNaughton function

An ingenious solution based on Farey sequences of Schauder hats is due to [Mundici, 1994; Mundici and Pasquetto, 1995]. I sketch the one-dimensional case following [Cignoli *et al.*, 1999].

The **Farey partition**  $\text{Farey}_n$  of  $[0, 1]$  is a finite sequence of rational numbers from  $[0, 1]$ , inductively defined by:

1.  $\text{Farey}_0 = \langle 0, 1 \rangle$
2. If  $\text{Farey}_n$  is a sequence of length  $m$ , then  $\text{Farey}_{n+1}$  is a sequence of length  $2m-1$  and is obtained by inserting between any two consecutive numbers  $\frac{a}{b}$  and  $\frac{c}{d}$  in  $\text{Farey}_n$ , the number  $\frac{a+c}{b+d}$ .

$\text{Farey}_1 = \langle 0, \frac{1}{2}, 1 \rangle$ ,  $\text{Farey}_2 = \langle 0, \frac{1}{3}, \frac{1}{2}, \frac{2}{3}, 1 \rangle$ , etc.; observe that the non-simplified denominators correspond to Pascal's triangle.

Trivially,  $|\text{Farey}_n| = 2^n + 1$ , moreover, it was proven by Cauchy that each member of the rational unit interval occurs in  $\text{Farey}_n$  for some  $n$ . With each sequence  $\text{Farey}_n = \langle \frac{a_1}{b_1}, \dots, \frac{a_u}{b_u} \rangle$  one associates a sequence Schauder $_n$  of the same length, whose elements are certain simple univariate functions, so-called **Schauder hats** as follows:  $h_1$  connects  $(0, 1)$  and  $(\frac{a_1}{b_1}, 0)$ ,  $h_u$  connects  $(1, 0)$  and  $(\frac{a_u}{b_u}, 0)$ , whereas for each  $1 < k < u$ ,  $h_k$  connects  $(\frac{a_k}{b_k}, \frac{1}{b_k})$  with  $(\frac{a_{k-1}}{b_{k-1}}, 0)$  on the one hand and  $(\frac{a_{k+1}}{b_{k+1}}, 0)$  on the other (Schauder $_2$  is depicted in Figure 7). Define the **multiplicity** of each Schauder hat  $h_k$  to be  $\mu_k = b_k$ . It is easy to see that  $\sum_{k=1}^u \mu_k h_k = 1$ .

Using Schauder hats, a unary McNaughton function  $f$  is easy to construct: Let  $n$  be the smallest number such that the non-differentiable points of  $f$  are in  $\text{Farey}_n$ . In the example in Figure 7, these are  $\frac{1}{3}$ ,  $\frac{1}{2}$ , and  $\frac{2}{3}$ , thus we use  $\text{Farey}_2$ .

Since  $f$  has integral coefficients, the value of  $f$  at each  $\frac{a_k}{b_k}$  is an integral multiple of  $\frac{1}{b_k} = h_k(\frac{a_k}{b_k})$ . Therefore, a suitable linear combination  $g$  of the hats  $h_k$  coincides with  $f$  on  $\text{Farey}_n$ . And, since both  $f$  and  $h_k$  are linear on each interval  $[\frac{a_k}{b_k}, \frac{a_{k+1}}{b_{k+1}}]$ ,  $f$  and  $g$  coincide on  $[0, 1]$ . In the example (right most picture),  $f = 2h_2 + h_3$ .

The construction is completed by noting that Schauder hats have a formula representation: Schauder $_1$  is given by  $h_0 = p$ ,  $h_1 = \neg_L p$ ; assume we have formulas for Schauder $_n = \langle h_1, \dots, h_u \rangle$ , then Schauder $_{n+1} = \langle k_1, \dots, k_{2u-1} \rangle$  is given by:

- $k_1 = h_1 \ominus (h_1 \wedge h_2)$ ,
- $k_{2j} = h_j \wedge h_{j+1}$ , for  $1 \leq j < u$ ,
- $k_{2j-1} = h_j \ominus (h_{j-1} \oplus_L h_{j+1})$ , for  $1 < j < u$ ,
- $k_{2u-1} = h_u \ominus (h_{u-1} \wedge h_u)$ ,

where  $iA_{\ominus}j = \max\{0, i - j\}$ .

The  $n$ -dimensional case of the constructive proof of McNaughton's theorem proceeds along similar lines, but adds considerable technical complications. For a start, the sequence of rational points in  $[0, 1]$ , where  $f$  is non-differentiable, becomes a set of at most  $(n - 1)$ -dimensional faces. These give rise to a unimodular, simplicial triangulation, the  $n$ -dimensional version of Farey partitions. Details are given in [Mundici, 1994; Cignoli *et al.*, 1999].

## 4.2 Mixed Integer Programming

It is well-known (see, for example, [Hooker, 1988; Jeroslow, 1988]) that propositional classical formulas in conjunctive normal form correspond to

certain 0-1 integer programs. More precisely, given a set  $\Gamma$  of classical disjunctive clauses over the signature  $\Sigma$  one transforms each clause

$$(24) \quad C = p_1 \vee \cdots \vee p_k \vee \neg p_{k+1} \vee \cdots \vee \neg p_{k+m}$$

into a linear inequation

$$(25) \quad \sum_{i=1}^k p_i - \sum_{j=k+1}^m p_j \geq 1 - m$$

Here, the variables from  $\Sigma$  are interpreted as function variables ranging over  $\{0, 1\}$ . It is easy to see that the resulting set of inequations is solvable iff  $\Gamma$  is satisfiable: recall that  $A_{\vee}(i, j) = \max\{i, j\}$  and  $A_{\neg}(i) = 1 - i$ , so clause  $C$  is satisfiable iff  $f_C \geq 1$  iff (25) holds.

McNaughton's theorem suggests that this embedding of logic into integer programming can be generalized to cover Lukasiewicz logic (after all, a McNaughton function is defined by linear polynomials) and, indeed, it turns out to be possible [Hähnle, 1994c; Hähnle, 1997]. To aid the presentation of this result, let us recall some facts and definitions about Mixed Integer Programming (MIP). As background reading, for example, [Schrijver, 1986] is recommended.

With the expression **linear inequation** I mean in the following always a term of the form  $a_1 p_1 + \cdots + a_m p_m \geq c$ , where  $a_1 p_1 + \cdots + a_m p_m$  is a linear polynomial over variables  $\{p_1, \dots, p_m\}$  and integral coefficients  $\{a_1, \dots, a_m, c\}$ . The type of the variables can be any truth value set  $N$  as defined in Example 2. The expression  $a_1 p_1 + \cdots + a_m p_m$  is called **linear term**.

**DEFINITION 14.** Let  $\mathbf{J}$  be a finite set of linear inequations and  $K$  a linear term. Let  $\Sigma$  be the set of variables occurring in  $\mathbf{J}$  and  $K$ . Assume the type  $N$  of each variable is finite. Then  $\langle \mathbf{J}, K \rangle$  is a **bounded integer program** (IP) with **cost function**  $K$ .<sup>11</sup> If the type of variables is either  $\{0, 1\}$  or  $[0, 1]$ , then one has a **bounded 0-1 mixed integer program** (MIP). When all variables in  $\Sigma$  run over infinite  $N$ , the result is a **bounded linear program** (LP).

A variable assignment  $\sigma : \Sigma \rightarrow [0, 1]$  that respects the type of each variable and such that all inequations in  $\mathbf{J}\sigma$  are satisfied is called a **feasible solution** of  $\langle \mathbf{J}, K \rangle$ . A variable assignment  $\sigma$  such that the value of  $K\sigma$  is minimal among all feasible solutions is called an **optimal solution**.  $\langle \mathbf{J}, K \rangle$  is **feasible** iff there are feasible solutions.

Only the feasibility part of (M)IPs/LPs is required, cost functions are not considered in the following.

<sup>11</sup>The adjective *integer* is justified, because the elements of  $N$  can without loss of generality assumed to be of the form  $\{0, 1, \dots, n - 1\}$ .

DEFINITION 15. Let  $M \subseteq [0, 1]^k$ .  $M$  is a **MIP-representable** set if there is an MIP  $\mathbf{J}$  over variables  $\Sigma' = \{x_1, \dots, x_k\}$  with type  $[0, 1]$  and variables  $\Sigma''$  with type  $\{0, 1\}$  such that

$$M = \{\vec{x} \mid \vec{x} \text{ is feasible solution of } \mathbf{J}\sigma \text{ for some } \sigma : \Sigma'' \rightarrow \{0, 1\}\}.$$

A many-valued logic is **MIP-representable** iff for all its connectives  $\theta \in \Theta$  the function  $A_\theta$ , is an MIP-representable subset of  $[0, 1]^{\alpha(\theta)+1}$ . The variable in a relational MIP-representation of a function that holds the function value is called **output variable**, the variables that hold the function arguments are called **argument variables**.

All finite-valued logics are MIP-representable, simply because  $A_\theta$  is a finite subset of  $[0, 1]^{\alpha(\theta)+1}$ . It is more interesting that Łukasiewicz logic is MIP-representable. Specifically, an MIP representation of  $\oplus_L$  is given by

$$\begin{array}{ll} (i) & x + y + z - i \geq 0 \\ (ii) & -x - y + z + i \geq 0 \\ (iii) & x + y - z \geq 0 \\ (iv) & -x - y + z \geq -1 \\ (v) & -z + i \geq 0 \end{array}$$

where  $x$  and  $y$  are argument variables,  $i$  is output variable and  $z$  is an additional variable with type  $\{0, 1\}$ . To see this, first set  $z = 0$ . Then the polynomial  $P_1(x, y) = i = x + y$  is defined by (i) and (ii), inequations (iii) and (v) are trivially satisfied, and (iv) determines the area in which  $P_1$  equals  $\oplus_L$ . The case  $z = 1$  is similar. An MIP-representation of  $\neg_L$  is straightforward:

$$\begin{array}{ll} -x - i \geq -1 \\ x + i \geq 1 \end{array}$$

All other connectives are definable with  $\oplus_L$  and  $\neg_L$ . Now McNaughton's Theorem can be strengthened to provide a direct link between MIP-representable logics and MIP:

THEOREM 16 (Generic MIP version of McNaughton's Theorem).

1. If  $\varphi(\vec{p})$  is a formula of an MIP-representable logic then there is an MIP  $\mathbf{J}_\varphi$  with argument variables  $\vec{p}$  and output variable  $y$  whose feasible solutions restricted to  $(\vec{p}, y)$  are the function  $f_\varphi(\vec{p})$ . Moreover, the size of  $\mathbf{J}_\varphi$  is linear in the size of  $\varphi$ .
2. Let  $\mathbf{J}$  be an MIP over variables  $\Sigma$ . Then there is a  $\Sigma$ -formula  $\varphi_{\mathbf{J}} \in \mathbf{L}_{Luk}$  which is satisfiable iff  $\mathbf{J}$  is feasible.

The first part of the theorem now is non-trivial to prove. In return, it provides a direct way to perform deduction in MIP-representable logics.

The second part of the theorem simply says that there is a “backend” to McNaughton’s result that allows to go from MIP to McNaughton functions.

It is instructive to sketch the proof of Part 1 (see [Hähnle, 1997] for Part 2): assume we have an MIP-representation of  $A_\theta$  with output variable  $y_\psi$  and argument variables  $x_{\varphi_1}, \dots, x_{\varphi_k}$  for each complex subformula  $\psi = \theta(\varphi_1, \dots, \varphi_k)$  of  $\varphi$ . Connect the MIP-representations for each such  $\psi$  by adding equations  $x_\psi = y_\psi$ ; furthermore, add equations  $x_p = p$  for each propositional variable  $p$  of  $\varphi$  to obtain an MIP-representation of  $f_\varphi$  with output variable  $y_\varphi$  and argument variables  $p$ . The size of the MIP-representation of each  $\psi$  is constant and depends only on the connective  $\theta$  and the number of all such MIP-representations is proportional to the number of subformulas in  $\varphi$ , hence it is linear in the length of  $\varphi$ .

A direct realization of this proof idea (written in Eclipse Prolog and taken from [Hähnle, 1994b; Hähnle, 1997]) in the case of Łukasiewicz logic is given in Figure 8. It assumes that  $\neg_L$  is represented by `neg/1`,  $\oplus_L$  by `plus/2`, and atoms  $p$  by `atom(p)`, all in prefix notation. A query of the form “:- `sat(i, phi)`.” builds an MIP representation of `phi` and checks if  $f_{\text{phi}} = i$ , that is,  $\mathbf{I}(\text{phi}) = i$  for some  $\mathbf{I}$ , is possible.

### 4.3 Extensions of Łukasiewicz Logic

Theorem 16, Part 1 does not require integrity of coefficients of the associated functions  $f_\varphi$  in a crucial way, hence it works for logics corresponding to generalized McNaughton functions with possibly non-integral coefficients. To determine the algebraic and logical counterpart of this class of functions is ongoing research. One possibility is to add capability for division by positive integers, for example by supplying infinitely many unary connectives of the form  $\frac{1}{d}$  to Łukasiewicz logic, where  $d$  is a positive natural number, and  $A_{\frac{1}{d}}(i) = \frac{i}{d}$  for  $i \in [0, 1]$ .

Non-continuous connectives, such as Gödel implication and negation, can easily be handled by MIPs with strict inequalities. The connectives of product logic, on the other hand, lead outside MIP and into non-linear programming.

**Rational Pavelka logic** (RPL)  $\mathcal{L}_{\text{RPL}}^0$ , Hajek’s [1998] formalization of Pavelka [1979a; 1979b; 1979c] within the framework of MVL, was mentioned already in Section 2.4 in connection with Pavelka-style completeness.

The language of  $\mathcal{L}_{\text{RPL}}^0$  extends that of  $\mathcal{L}_L^0$  by an infinite number of logical constants of the form  $\bar{i}$  for each rational  $r \in [0, 1]$ , where  $A_{\bar{r}} = r$ . The first-order version  $\mathcal{L}_{\text{RPL}}$  is defined in the same manner. The rational logic constants can be used to express graded truth with  $\bar{r} \rightarrow_L \varphi$ . One has  $rA_{\rightarrow_L} y = \min\{1, 1 - r + y\}$ , which is different from  $\uparrow r(y)$  used in Section 2.4. Moreover,  $f(y) = \min\{1, 1 - r + y\}$  is in general not a McNaughton function. On the other hand,  $rA_{\rightarrow_L} y = \uparrow r(y) = 1$  iff  $r \leq y \leq 1$ .

```

:- lib(r).                % load linear constraint solver

sat(I,plus(Phi,Psi)) :-
    sat(X,Phi),           % Connect X, Phi
    sat(Y,Psi),           % Connect Y, Psi
    truth_var(X),         % X is in [0,1]
    truth_var(Y),         % Y is in [0,1]
    truth_var(I),         % I is in [0,1]
    control_var(Z),       % Z is in {0,1}
    X + Y + Z $>= I,      % (i)
    X + Y - Z $<= I,      % (ii)
    X + Y - Z $>= 0,      % (iii) MIP representation of  $\oplus$ 
    X + Y - Z $<= 1,      % (iv)
    I $>= Z.              % (v)

sat(I,neg(Phi))          :- sat(1-I,Phi).

sat(I,atom(P))           :- I $= P.

control_var(0).
control_var(1).

truth_var(X)             :- 0 $<= X, X $<= 1.

```

Figure 8. A satisfiability checker for infinite-valued Lukasiewicz logic

Provability in  $\mathcal{L}_{\text{RPL}}^0$  can be reduced to  $\mathcal{L}_{\text{L}}^0$ , even though  $\mathcal{L}_{\text{RPL}}^0$ -formulas need not correspond to McNaughton functions [Hájek, 1998]. The idea is to replace each constant  $\bar{r}$  in an RPL-formula  $\varphi$  with a variable  $p_r$  and to add suitable formulas  $\psi_r$ , for which  $\mathbf{I}(\psi_r) = 1$  iff  $\mathbf{I}(p_r) = r$ . Even for  $\mathcal{L}_{\text{RPL}}$  one has that rational Pavelka logic is a conservative extension of first-order Łukasiewicz logic [Hájek *et al.*, to appear, 2000].

#### 4.4 Axiomatization and Completeness

A sound and complete axiomatization [Hájek, 1998] of infinite-valued Łukasiewicz logic is obtained by adding the axiom

$$(I) \quad \neg_{\text{L}} \neg_{\text{L}} \varphi \rightarrow_{\text{L}} \varphi$$

to (A1–A7) of basic  $t$ -Norm logic  $\mathcal{BL}$  (and suitably adapting subscripts of connectives), which expresses that  $\neg_{\text{L}}$  is an involution.

Historically, a much simpler equivalent axiomatization of infinite-valued Łukasiewicz logic was used, consisting of (A1–A2) together with:

$$(A13) \quad ((\varphi \rightarrow_{\text{L}} \psi) \rightarrow_{\text{L}} \psi) \rightarrow_{\text{L}} ((\psi \rightarrow_{\text{L}} \varphi) \rightarrow_{\text{L}} \varphi)$$

$$(A14) \quad (\neg_{\text{L}} \varphi \rightarrow_{\text{L}} \neg_{\text{L}} \psi) \rightarrow_{\text{L}} (\psi \rightarrow_{\text{L}} \varphi)$$

Completeness of Łukasiewicz logic can be established with several strategies: the historically first proof [Rose and Rosser, 1958] has a very syntactic nature. The basic idea is to prove that  $\vdash \varphi \equiv \psi_{\varphi}$  for certain  $\psi_{\varphi}$  with  $f_{\psi_{\varphi}} = f_{\varphi}$ . If  $\varphi$  is a tautology, then  $f_{\varphi} \equiv 1$ , and  $\vdash \psi_{\varphi}$  can be relatively easily established. The proof of  $\vdash \varphi \equiv \psi_{\varphi}$  is by induction on the complexity of  $\varphi$  and not unlike the reduction of  $\mathcal{L}_{\text{L}}^0$ -formulas into MIP described in Section 4.2 (without the polynomial bound). All the reasoning has to take place in terms of Hilbert calculi, so the proofs are very long and tedious to follow.

Completeness of the calculus can be relatively easily obtained from algebraic completeness by showing that the Lindenbaum algebra of the axiomatization is a Wajsberg algebra and, hence, up to notation, an MV-algebra.

To prove algebraic completeness, the key insight of Chang [1959] was to link MV-algebras to  $\ell$ -groups. Since  $\varphi$  holds in all MV-algebras iff it holds in all totally ordered MV-algebras, it is sufficient to consider  $\mathcal{o}$ -groups. From there, [Cignoli and Mundici, 1997a] directly go to the free Abelian group  $\mathbb{Z}^r$  and use linear algebra to embed that into  $\mathbb{R}$ ; on the other hand, [Hájek, 1998; Gottwald, 2000] following [Chang, 1959] take a shortcut by appealing to the Gurevich-Kokorin Theorem on quantifier elimination in  $\mathcal{o}$ -groups. A quite different proof using the de Concini-Procesi theorem on elimination of points of indeterminacy in toric varieties is due to [Panti, 1995]. Further proof techniques are mentioned in [Cignoli *et al.*, 1999].



### 4.5 Further Topics

Deep connections between Łukasiewicz logic and diverse other fields have been established. Examples are functional analysis ( $C^*$ -algebras) [Mundici, 1986] and coding theory (adaptive error-correcting codes and Ulam's game) [Mundici, 1992; Mundici, 1990], these are discussed in Section 4.3 of Urquhart's chapter in this Volume.

Further links to quantum physics [Mundici, 1993], geometry (toric desingularizations) [Mundici, 1996], and advanced algebra [Di Nola and Lettieri, 1994; Cignoli and Mundici, 1997b; Cignoli *et al.*, 1999] are, unfortunately, well beyond the scope of this article. Suffice it to say that MV-algebras are categorically equivalent to  $\ell$ -groups with strong unit [Mundici, 1986].

## 5 FUZZY LOGIC

### 5.1 Many-Valued Logics versus Fuzzy Logic

The main journal of **fuzzy logic** research, *Fuzzy Sets and Systems*, appears bi-weekly with twelve articles on average. This makes about 300 published articles per year alone for this journal. A search in the Library of Congress Catalog yields that at least 150 books with the term *fuzzy logic*, *fuzzy system* or *fuzzy set* in their title are available (of these, I can only mention a few of the many recommendable ones: [Zimmermann, 1991; Gottwald, 1993; Kruse *et al.*, 1994; Hájek, 1998; Turunen, 1999]). This is an indicator of how unwieldy the field has become.

To get a first handle on the meaning of fuzzy logic from an MVL point of view, I cite the inventor of fuzzy logic, Lotfi Zadeh (after [Hájek, 1998]):

"In a narrow sense, fuzzy logic, FLn, is a logical system which aims at a formalization of approximate reasoning. In this sense, FLn is an extension of multivalued logic. However, the agenda of FLn is quite different from that of traditional multivalued logics. In particular, such key concepts in FLn as the concept of a linguistic variable, canonical form, fuzzy if-then rule, fuzzy quantification and defuzzification, predicate modification, truth qualification, the extension principle, the compositional rule of inference and interpolative reasoning, among others, are not addressed in traditional systems. In its wide sense, fuzzy logic, FLw, is fuzzily synonymous with the fuzzy set theory, FST, which is the theory of classes with unsharp boundaries. FST is much broader than FLn and includes the latter as one of its branches."  
—Lotfi Zadeh in [Marks, 1994, Preface]

Hájek comments on this as follows:

"... even if I agree with Zadeh's distinction between many-valued logic and fuzzy logic in the narrow sense, I consider formal calculi of many-valued logic (including non-'traditional' ones, of course) to be the kernel or base of fuzzy logic in the narrow sense and the task of explaining things Zadeh mentions by means of these calculi to be a very promising task (not yet finished)."  
—[Hájek, 1998, p. 2]

... and begins to unfold exactly this scientific program. It must be said that the perception of fuzzy logic in the eyes of logicians has changed quite a bit

in recent years [Hájek, 2000]. Hájek also authors the deepest investigation into FLn from a logical point of view so far [Hájek and Paris, 1997; Hájek, 1998; Hájek, 2000]. His conclusions about the relationship between FLn and MVL seem eminently reasonable:

- FLn is an interesting mathematical pursuit in its own right;
- some of its concepts can be clarified by formalizing them in logic: the nature of fuzzy rules, consistency, proof of general properties; as such, the study of FLn in a many-valued logic framework is of interest to the field of FL at large.

I want to refine the second point with an example: in classical rule-based systems one uses modus ponens  $(\varphi \wedge (\varphi \rightarrow \psi)) \rightarrow \psi$  to derive new facts  $B$  from rules  $r = A' \rightarrow B'$  and facts  $A$ . In fuzzy control (FC) [Zadeh, 1979; Bonissone, 1997] all notions are “fuzzified”:  $A$  is a fuzzy predicate,  $r$  typically is a fuzzy mapping (“for arguments approximately equal to  $A'$  the image is approximately equal to  $B'$ ”) called **Mamdani rule** [Mamdani and Assilian, 1975], etc. As [Hájek, 1998] shows, it is possible to model basic notions of FC adequately within the framework of first-order MVL, specifically, with the logics  $\mathcal{L}_t$  (Example 7). The important thing here is that now notions of FC have a precise, unambiguous semantics and can be analyzed with the tools of formal logic. It becomes clear, for instance, that Mamdani rules are based on logical conjunction (thus  $t$ -norms) rather than implication as is sometimes insinuated in FC [Driankow *et al.*, 1993]. One also can easily compare different interpretations of FC concepts. For example, certain rule schemata turn out to be actually invalid in the sense of many-valued logic [Hájek, 1998, p. 192f], which should raise doubts on their use within FC as well.

The logical analysis of FLn available so far indicates that FLn is deeply rooted in “traditional” MVL: many logics used to analyze FLn (Gödel logic, Łukasiewicz logic) existed well before the term “fuzzy logic” was even invented.

One can, of course, argue which particular many-valued logic should be considered as *the* logical basis of FLn. Hájek [1998; 2000] suggests continuous  $t$ -norms over  $[0, 1]$ , naturally ordered, with residuated implication (see p. 301) and negation (1), in other words, basic  $t$ -norm logic; in the first-order case, the quantifiers are  $\forall, \exists$  (see Examples 6, 7). Other authors advocate weaker systems such as monoidal  $t$ -norm logic [Esteva and Godo, 1999] or stronger systems such as rational Pavelka logic [Novák, 1995].

## 5.2 Some Technical Results of Fuzzy Logic

Here I collect some odd ends and pieces of fuzzy logic that are in some way relevant for the present chapter.

**On  $t$ -norm Theory.** As classes of functions on the real unit interval,  $t$ -norms were investigated in different contexts long before FLn was even around. One of the fundamental results is about the decomposition of continuous  $t$ -norms and it appears already in [Mostert and Shields, 1957].

A continuous  $t$ -norm  $*$  is **Archimedean**, if it has as idempotent elements exactly 0 and 1. This the case for Łukasiewicz product  $\odot_L$  and multiplication  $\odot_{\Pi}$  of product logic, but not for Gödel product  $\odot_G$  which has all elements of  $[0, 1]$  as idempotents. In fact, Gödel product is the only  $t$ -norm with this property.

Recall that a  $t$ -norm  $*$  is nilpotent, if it has other nilpotent elements besides 0; it is called **strict** otherwise. Obviously, Łukasiewicz product  $\odot_L$  is nilpotent and  $\odot_{\Pi}$  of product logic is strict.

It turns out that these examples are actually characteristic in the sense that:

- each nilpotent Archimedean  $t$ -norm is isomorphic to Łukasiewicz product
- and each strict Archimedean continuous  $t$ -norm is isomorphic to the  $t$ -norm of product logic.

Starting from the observation that the idempotents  $\mathcal{I}(*)$  of any continuous  $t$ -norm  $*$  form a closed subset of  $[0, 1]$ , one obtains a well-known representation theorem for continuous  $t$ -norms.

Let  $([a_r, b_r])_{r \in I}$  be a countable family of non-overlapping proper subintervals of  $[0, 1]$ . Now assume a  $t$ -norm  $*_r$  is associated with each interval. One defines the **ordinal sum**  $T : [0, 1]^2 \rightarrow [0, 1]$  of  $([a_r, b_r])_{r \in I}$  and  $(*_r)_{r \in I}$  as

$$(26) \quad T(i, j) = \begin{cases} a_r + (b_r - a_r) \cdot \left( \frac{i - a_r}{b_r - a_r} *_r \frac{j - a_r}{b_r - a_r} \right) & \text{if } i, j \in [a_r, b_r] \\ \min\{i, j\} & \text{otherwise} \end{cases}$$

Observe that an Archimedean  $t$ -norm is an ordinal sum consisting of just one summand and that the empty ordinal sum is the Gödel  $t$ -norm. Non-Archimedean  $t$ -norms  $*$  are decomposed into the ordinal sum of the Archimedean  $t$ -norms defined on the intervals in  $\overline{\mathcal{I}(*)}$ , the closure of  $[0, 1] \setminus \mathcal{I}(*)$ .

**THEOREM 17** ([Mostert and Shields, 1957]). *Each continuous  $t$ -norm is the ordinal sum of a family of continuous Archimedean  $t$ -norms.*

Recall from Section 3 that basic  $t$ -norm logic is the logic of continuous  $t$ -norms [Cignoli *et al.*, 2000]. In the wake of this result an algebraic analog of Theorem 17 is proven in the same paper that allows to decompose totally ordered BL-algebras into G-, MV-, and  $\Pi$ -algebras and ordinal sums thereof.

A different way to lend structure to  $t$ -norms is to give parameterized families of  $t$ -norms that instantiate to well-known members. One of many examples is Frank’s family of  $t$ -norms [Frank, 1979]:

$$(27) \quad i *_r^F j = \log_r \left( 1 + \frac{(r^i - 1)(r^j - 1)}{r - 1} \right)$$

Our standard  $t$ -norms are obtained as limits:  $iA_{\odot_L} j = \lim_{r \rightarrow \infty} i *_r^F j$ ,  $iA_{\odot_G} j = \lim_{r \rightarrow 0} i *_r^F j$ , and  $iA_{\odot_H} j = \lim_{r \rightarrow 1} i *_r^F j$ .

A general and up-to-date reference on  $t$ -norm theory is the book [Klement *et al.*, 2000].

**Fuzzy Mathematics.** There hundreds of papers that claim to fuzzify notions of classical mathematics, but many of these fuzzy notions are defined rather *ad hoc*. Among the more serious attempts, I mention only Thiele’s ongoing program to fuzzify parts of universal algebra, for example, in [Thiele and Schmechel, 1995; Thiele, 1998].

## 6 PROOF THEORY AND COMPUTATIONAL ASPECTS OF MANY-VALUED LOGIC

### 6.1 *Sequent and Tableau Calculi*

Hilbert calculi give a good idea of what the “characteristic truths” of a logic, its axioms, are. They are not well-suited for doing *proof theory*, the analysis and systematic construction of formal proofs. For this purpose, **Gentzen** or **sequent calculi** [Gentzen, 1935] and, more recently, tableaux and resolution calculi are more suitable.

A **sequent** is an expression of the form  $\Gamma \Rightarrow \Delta$ , often read as “if I can prove  $\Gamma$ , then I can prove  $\Delta$ ”. Call  $\Gamma$  the **antecedent**,  $\Delta$  the **succedent** of the sequent. Depending on the purpose (and preferences of the author),  $\Gamma$  and  $\Delta$  are sequences, multisets, or sets of objects. In the present chapter, we use sets.

Certain abbreviations are standard: “ $\Gamma, \varphi$ ” for “ $\Gamma \cup \{\varphi\}$ ”, “ $\Gamma, \Gamma'$ ” for “ $\Gamma \cup \Gamma'$ ”, “ $\Rightarrow \Delta$ ” for “ $\emptyset \Rightarrow \Delta$ ”, etc.

A **sequent calculus** is a set of rule schemata (12), where premisses and conclusion are sequent schemata (the latter are defined analogously to formula schemata). A sequent rule with an empty set of premisses, is called **axiom** in the context of sequent calculi.

Traditionally, sequent rules come in two kinds, **structural rules** and **logical rules**. The latter have the property that some connective occurs only in the succedent (“is introduced”), thus allowing to build up more

complex formulas. A typical example from classical logic (for conjunction) is:

$$\frac{\Gamma \Rightarrow \Delta, \varphi \quad \Gamma \Rightarrow \Delta, \psi}{\Gamma \Rightarrow \Delta, \varphi \wedge \psi}$$

Structural rules involve no connectives, but manipulate the arrangement of objects within sequents. In our setting they will not be needed, because antecedents and succedents are sets and because of the choice of axioms.

A calculus is **analytic** or obeys the **subformula principle** if, for each rule, the premisses contain only subformulas of formulas in the conclusion.

Each sequent calculus  $\mathcal{SK}$  induces a **provability relation** on sequents. A **sequent proof tree** is labeled with sequents, and inductively defined by:

- (1) Each single-node tree labeled with an instance of an axiom of  $\mathcal{SK}$  is a sequent proof.
- (2) If  $(\Pi_i)_{i \in I}$  is a family of sequent proofs and there is an instance of a rule schema in  $\mathcal{SK}$  with conclusion  $\Gamma$ , such that each of its premisses occurs among the root labels of the  $\Pi_i$ , then the tree with root  $\Gamma$  and immediate subtrees  $\Pi_i$  is a sequent proof.

A sequent proof with root label  $\Gamma$  is called a **sequent proof** of  $\Gamma$  (in  $\mathcal{SK}$ ).

Analyticity is a key property, if one is interested in automatic proof search: in a finitary and analytic sequent calculus goal-directed (backwards) search has a finite branching factor. Hilbert style calculi that contain modus ponens are inherently non-analytic.

Sequent calculi permit a variety of structural (for example, through structural rules) or geometric conditions, by which the number and form of derivations can be restricted and, hence, different logics can be characterized. For example, already in Gentzen's [1935] paper, intuitionistic logic was obtained by restriction of succedents to at most singletons.

The presence or absence of this "structural feature" of sequent calculi implies two fundamentally different readings of sequents. The first is the traditional one in proof theory, where antecedents and succedents typically are multisets or sequences of objects. In such calculi, a sequent like  $\langle \gamma_1, \dots, \gamma_n \rangle \Rightarrow \langle \delta_1, \dots, \delta_m \rangle$  typically is interpreted as

$$(28) \quad \gamma_1 \wedge \dots \wedge \gamma_n \models \delta_1 \vee \dots \vee \delta_m \text{ ,}$$

where  $\wedge$  is some sort of conjunction or product operator and  $\vee$  a disjunction or sum.

Several many-valued logics, including Gödel, Lukasiewicz and the paraconsistent logic  $J_3$ , were axiomatized with such sequent calculi [Avron, 1991b; Hösli, 1993; Prijatelj, 1996; Dyckhoff, 1999; Avellone *et al.*, 1999]. In the terminology introduced in Section 2.4 all of them are internal calculi.

If one is interested in analytic calculi, then often some kind of extension of the base mechanism is necessary to deal with many-valued logic. One possibility which, at least in spirit, if not in letter, is internal are **hypersequent calculi** [Pottinger, 1983; Avron, 1987; Avron, 1996]. They result from sequent calculi by admitting finite sets of sequents in the place, where just single sequents stood before. Thus, a **hypersequent** is of the form

$$[\Gamma_1 \Rightarrow \Delta_1 | \cdots | \Gamma_r \Rightarrow \Delta_r],$$

where each  $\Gamma_i \Rightarrow \Delta_i$  is a standard sequent.

In hypersequents, the “|” is commonly interpreted as a disjunction. Several finite- and infinite-valued logics, including Lukasiewicz logic, Gödel logic, and Urquhart’s logic<sup>12</sup> C [Urquhart, 1986] were successfully axiomatized with analytic hypersequent calculi [Avron, 1991b; Avron, 1991a; Avron, 1996; Baaz *et al.*, 1998c; Ciabattoni *et al.*, 1999; Ciabattoni, 2000a; Ciabattoni, 2000b].

In the following I will not discuss these results any further. First of all, because each calculus typically involves a new and specific proof theoretical insight and it is too space-consuming to give the details. Second, general proof theoretical devices like structural modifications, hypersequents and, even more so, display calculi [Belnap, 1982], are by no means specific to many-valued logic and thus belong to a general discussion of proof theory, which is found in other parts of this Handbook.

Let us now look at the second possible interpretation of sequents. In classical logic, by virtue of the deduction theorem, (28) is equivalent to  $\models \bigwedge_{i=1}^n \gamma_i \rightarrow \bigvee_{j=1}^m \delta_j$  or

$$(29) \quad \models \bigvee_{i=1}^n \neg \gamma_i \vee \bigvee_{j=1}^m \delta_j .$$

This interpretation hinges on bi-valuedness of classical logic and gives a hint what one can do in the many-valued case: for  $i \in N$ , let  $\Gamma_i$  be sets of formulas.

The notation

$$(30) \quad \Gamma_1 | \cdots | \Gamma_n ,$$

due to [Rousseau, 1967], represents the assertion “there is an  $i \in N$  and  $\gamma \in \Gamma_i$  such that  $\models_{\{i\}} \gamma$ ”.

This interpretation and form of sequents is usually objected against by proof theorists:

“One should not be able to guess, just from the form of the structures which are used, the intended semantics of a given proof system . . . ” —[Avron, 1996, p. 2]

<sup>12</sup>Note that the definition changed in Urquhart’s chapter in this Volume of the present volume.

But whether one regards, say, hypersequents still as strictly “internal” is a matter of taste, and the explicit incorporation of semantical structures into proof theory can be a great virtue from the point of view of automatic proof search. It also gives us a proof theory that has a specific many-valued flavour. And this is what I intend to detail in the remaining section.

Variants of notation (30) exist (for example, [Takahashi, 1967]), but they are all a little clumsy. [Suchoń, 1974; Surma, 1974; Surma, 1984] suggested to use **signed formulas** of the form  $i:\varphi$ , which stands for  $\models_{\{i\}} \varphi$ . It is only natural to make the final step, and use signed formulas of the form  $S:\varphi$  to represent  $\models_S \varphi$ . This was done by Hähnle [1991a; 1994a] who realized that using “truth value sets as signs” not only greatly simplifies many rules, but also may lead to exponentially shorter derivations. Similar, slightly less general ideas were expressed independently in [Doherty, 1990; Murray and Rosenthal, 1991a].

In the following sections, proof theory for many-valued logic based on signed formulas, that is, with *external calculi* is developed. In particular, the notion of sequent used is that of a **signed sequent**, which is simply a set of signed formulas.

A **tableau rule schema** is a pair  $\langle \varphi, \mathcal{C} \rangle$ , where  $\varphi$  is a formula schema called **premiss**, and  $\mathcal{C}$  a non-empty family of non-empty sets of formula schemata called **conclusion**; the members of  $\mathcal{C}$  are called **extension**. A **closure rule schema** is a set of formula schemata. In the finitary case, tableau rule schemata are written thus:

$$(31) \quad \text{extension} \left\{ \underbrace{\begin{array}{c|c|c} \text{premiss} & & \\ \hline \psi_{11} & \cdots & \psi_{m1} \\ \vdots & & \vdots \\ \psi_{1r_1} & \cdots & \psi_{mr_m} \end{array}}_{\text{conclusion}} \right.$$

In classical logic, one has  $m \leq 2$  and  $r_i \leq 2$ , but for many-valued logic the general case is needed.

EXAMPLE 18. Here is a typical tableau rule schema for classical logic:

$$(32) \quad \frac{\varphi \rightarrow \psi}{\neg\varphi \mid \psi}$$

The usual closure rule schemata in classical logic are  $\{\overline{0}\}$  and  $\{\neg\varphi, \varphi\}$ .

A **tableau calculus** is a set of tableau rule schemata and closure rule schemata. Each tableau calculus  $\mathcal{TK}$  induces a **provability relation** on sets of formulas  $\Psi$ . A **tableau proof tree** (for short, only tableau) for  $\Psi$  is a tree labeled with formulas, inductively defined by:

- (1) The empty tree is a tableau for  $\Psi$ .
- (2) If  $\mathbf{T}$  is a tableau for  $\Psi$ , then so is the tree obtained by appending a node labeled with some  $\varphi \in \Psi$  below any branch of  $\mathbf{T}$ .
- (3) If  $\mathbf{T}$  is a tableau for  $\Psi$ ,  $\varphi$  a label on a branch  $B$  of  $\mathbf{T}$ , and  $\varphi$  is the premiss of a tableau rule instance of  $\mathcal{TK}$  with extensions  $\mathcal{C}$ , then one obtains a tableau for  $\Psi$  by extending  $B$  with  $|\mathcal{C}|$  many linear subtrees, each containing as labels exactly the formulas in an extension of  $\mathcal{C}$ .

A branch  $B$  whose labels are a superset of a closure rule instance of  $\mathcal{TK}$  is called **closed**. A tableau is **closed**, if all its branches are closed. A closed tableau for  $\Psi$  is a **tableau proof** of  $\Psi$  (in  $\mathcal{TK}$ ).

### 6.2 Generic MVL: the Logic of Signed Formulas

DEFINITION 19. A **signed formula** of an  $N$ -valued logic  $\mathcal{L}$  is an expression of the form  $S:\varphi$ , where  $S \subseteq N$ , and  $\varphi \in \mathbf{L}$ . Satisfiability, validity, and consequence of signed formulas are defined with  $D$ -satisfiability, etc., by identifying  $\models_S \varphi$  with  $\models S:\varphi$ . In the case when  $\varphi$  is atomic,  $S:\varphi$  is called a **signed atom**. If  $S$  is a singleton, one speaks of a **monosigned formula**.

It is possible to view a signed formula  $S:\varphi(x_1, \dots, x_m)$  with free variables  $x_1, \dots, x_m$  itself as an atomic expression and to build classical first-order formulas over such atoms, for example,  $S:(p \rightarrow_{\mathbf{L}} q) \vee S':r$ . Here,  $\rightarrow_{\mathbf{L}}$  is Lukasiewicz implication and  $\vee$  is classical disjunction. From this point of view, signs act as a separator between classical and many-valued parts of a formula. This view is stressed in [Murray and Rosenthal, 1994].

A signed formula  $S:\varphi$  implicitly stands for a disjunction over the statements  $\mathbf{I}(\varphi) = i$  for all  $i \in S$ . With this device one can represent some properties of the truth values that can be taken on by formulas more succinctly in signed logic than without signs, as we will see.

On the semantic side, it is often useful to define the meaning of signed formulas more directly using **power algebras** (see [Brink, 1993] for an overview). Let  $\mathbf{A}^0$  be the matrix of an  $N$ -valued propositional logic  $\mathcal{L}^0$ . Define the **power matrix** of  $\mathbf{A}^0$  to be

$$\mathcal{P}(\mathbf{A}^0) = \langle \mathcal{P}^+(N), (A_\theta^+)_{\theta \in \Theta} \rangle ,$$

where  $\mathcal{P}^+(N)$  is the family of non-empty subsets of  $N$  and the **power operation**  $A_\theta^+ : \mathcal{P}^+(N)^{\alpha(\theta)} \rightarrow \mathcal{P}^+(N)$  of  $A_\theta$  is defined as

$$A_\theta^+(S_1, \dots, S_{\alpha(\theta)}) = \{A_\theta(i_1, \dots, i_{\alpha(\theta)}) \mid i_j \in S_j, 1 \leq j \leq \alpha(\theta)\} .$$

One can associate a **power interpretation**  $\mathbf{I}^+ : \Sigma \rightarrow \mathcal{P}^+(N)$  with each power matrix. It is continued on  $\mathbf{L}^0$  exactly like a standard interpretation.



$\mathbf{I}^+ \models_S \varphi$  iff  $\mathbf{I}^+(\varphi) \subseteq S$ . Obviously, if  $\mathbf{I}^+ \models_S \varphi$ , then  $\mathbf{I} \models_S \varphi$  for all interpretations  $\mathbf{I}$  with  $\mathbf{I}(p) \in \mathbf{I}^+(p)$  for all  $p \in \Sigma$ .

The inverse  $(A_\theta^+)^{-1}$  of  $A_\theta^+$  yields a family of  $\alpha(\theta)$ -tuples of subsets of  $N$ . Let  $\mathcal{M}(S) \subseteq \mathcal{P}^+(N)$  be the signs occurring in maximal tuples in  $(A_\theta^+)^{-1}(S)$  (ordered by point-wise set inclusion). We call a family of signs  $\mathcal{S} \subseteq \mathcal{P}^+(N)$  **complete** with respect to  $D \subseteq N$  and  $\mathcal{L}^0$ , if  $D \in \mathcal{S}$  and for all  $\theta \in \Theta$ ,  $S \in \mathcal{S}$ , and  $S' \in \mathcal{M}(S)$ ,  $S'$  is contained in the sublattice of  $2^N$  generated by  $\mathcal{S}$ .

EXAMPLE 20. Consider three-valued Lukasiewicz implication  $\rightarrow_L$ , and let  $\mathcal{S} = \{\{\frac{1}{2}\}, \{0, \frac{1}{2}\}, \{0, 1\}\}$ , which generates  $\mathcal{S} \cup \{\emptyset, \{0\}, N\}$ . Then  $\mathcal{S}$  is not complete with respect to  $\{0, \frac{1}{2}\}$  and  $\mathcal{L}_L^0$ , because  $\mathcal{M}(\{0, \frac{1}{2}\})$  contains  $\{\frac{1}{2}, 1\}$  (because of  $A_{\rightarrow_L}^+(\{\frac{1}{2}, 1\}, \{0\}) = \{0, \frac{1}{2}\}$ ).

A simple, complete system of signs for any  $D$  is

$$(33) \quad \mathcal{S}_{\text{mono}} = \{\{i\} \mid i \in N\},$$

the set of singleton signs. Another trivial example is

$$(34) \quad \mathcal{S}_{\text{full}} = \mathcal{P}^+(N) - \{N\},$$

the set of *all* signs (except  $\emptyset$  and  $N$ ). A non-trivial example is

$$(35) \quad \mathcal{S}_{\text{regular}} = \{\uparrow i \mid i \in N, \uparrow i \neq N\} \cup \{\overline{\uparrow i} \mid i \in N, \uparrow i \neq N\},$$

the set of non-trivial order filters of  $N$  and their complements that are generated by single elements. It is defined for any partially ordered set of truth values. For totally ordered  $N$ , these are exactly the prime ideals/filters and their complements and were called **regular sign** in [Hähnle, 1994a]. The name is kept for the present, more general, definition. Many other sets of signs are possible, for example, a kind of dual of (33):

$$(36) \quad \mathcal{S}_{\text{mono}}^d = \{N - \{i\} \mid i \in N\},$$

All of these systems of signs are trivially complete, because they generate  $2^N$ . The significance of complete families of signs is that they yield simple syntactic characterizations of any finite-valued logic operator:

THEOREM 21. *Let  $\mathcal{S}$  be a complete family of signs for an  $n$ -valued logic  $\mathcal{L}^0$ ,  $S \in \mathcal{S}$ , and  $\varphi = S:\theta(\varphi_1, \dots, \varphi_m)$  (for  $m \geq 1$ ) a signed  $\mathbf{L}_\Sigma^0$ -formula. Let  $\mathbf{I}$  be an arbitrary  $n$ -valued  $\Sigma$ -interpretation.*

*Then there are numbers  $M_1, M_2 \leq n^m$ , index sets  $I_1, \dots, I_{M_1}, J_1, \dots, J_{M_2} \subseteq \{1, \dots, m\}$ , and signs  $S_{rs}, S_{kl} \in \mathcal{S}$  with  $1 \leq r \leq M_1, 1 \leq k \leq M_2$  and  $s \in I_r, l \in J_k$  such that*

$$\varphi \text{ is satisfiable by } \mathbf{I} \\ \text{iff}$$

$$\begin{aligned} \bigvee_{r=1}^{M_1} \bigwedge_{s \in I_r} S_{rs} : \varphi_s \text{ is satisfiable by } \mathbf{I} \\ \text{iff} \\ \bigwedge_{k=1}^{M_2} \bigvee_{l \in J_k} S_{kl} : \varphi_l \text{ is satisfiable by } \mathbf{I}. \end{aligned}$$

The first equivalent expression is called a **signed DNF representation** of  $\varphi$ , the second a **signed CNF representation** of  $\varphi$ .

There are two special cases hidden here: when the image of the function  $A_\theta$  (that is,  $A_\theta^+(N)$ ) is a subset of  $S$ , then  $\varphi$  is valid; likewise, when  $A_\theta^+(N) \cap S = \emptyset$ , then  $\varphi$  is unsatisfiable. Although the theorem holds in those cases as well, it is more convenient to use then  $N:\varphi$ , respectively,  $\emptyset:\varphi$  as a representation of  $S:\varphi$  instead.

For  $\mathcal{S}_{\text{mono}}$ , Theorem 21 was shown in [Rousseau, 1967; Takahashi, 1967], the general case in [Hähnle, 1991a; Hähnle, 1994a] with a slightly different formulation of completeness of signs.

Recall that the semantics of a first-order quantifier  $\lambda$  in many-valued logic is defined via a distribution function  $Q_\lambda : \mathcal{P}^+(N) \rightarrow N$ . Similar as in the propositional case, one may obtain a representation of a signed quantified formula in terms of certain signed instances. Informally, what one needs to do is to characterize the distributions that are mapped to one of the truth values that occur in the sign of a quantified formula.

**THEOREM 22.** *Let  $(Q_\lambda)^{-1}(S) = \{\emptyset \neq I \subseteq N \mid Q_\lambda(I) \in S\}$ ; then a signed quantified formula*

$$\begin{aligned} S:(\lambda x)\varphi(x) \text{ is satisfiable} \\ \text{iff} \\ \bigvee_{I \in (Q_\lambda)^{-1}(S)} \left( \bigwedge_{i \in I} \{i\}:\varphi(c_i) \wedge \bigwedge_{t \in \text{Term}_S^0} I:\varphi(t) \right) \text{ is satisfiable,} \end{aligned}$$

where the  $c_i$  are new Skolem constants.

Each disjunct in this representation says that the distribution of  $\varphi$  at  $x$  is  $I$ : the first conjunction assures that *at least* the elements of  $I$  occur in the distribution, the second conjunction says that *at most* the elements of  $I$  occur.

In the spirit of the remark after Theorem 21, if  $(Q_\lambda)^{-1}(S) = \mathcal{P}^+(N)$ , use  $N:(\lambda x)\varphi(x)$ , and if  $(Q_\lambda)^{-1}(S) = \emptyset$ , use  $\emptyset:(\lambda x)\varphi(x)$ .

For a proof of the theorem, see [Hähnle, 1999], also [Carnielli, 1991; Baaz and Fermüller, 1995b] for a monosigned version. It must be stressed that monosigned first-order rules are *much* more complicated, even for simple quantifiers.

A CNF representation is obtained by duality: compute a DNF representation for  $\overline{S}$  and replace “not  $\bigvee \bigwedge \dots S' \dots$ ” with “ $\bigwedge \bigvee \dots \overline{S'} \dots$ ” using de Morgan’s rules.

### 6.3 Many-Valued Sequent and Tableau Calculi

Recall that in classical logic tableaux and sequent calculi correspond to each other very closely (see, for example, [Fitting, 1996; D'Agostino, 1999]):

The semantics of signed sequents  $\Gamma$  is defined by  $\models \Gamma$  iff  $\models \bigvee_{\gamma \in \Gamma} \gamma$  (compare to the discussion of formulas (29) and (30) as well as Definition 19). Instances of axiomatic sequents are supposed to correspond to valid formulas, and sequent rule schemata preserve validity for all instances: if all premisses are valid, then the conclusion is valid as well. In other words, they are CNF-representations of their premiss. A standard induction argument then shows **soundness**: the existence of a sequent proof with root  $\Gamma$  implies that  $\Gamma$  is valid.

Tableau proofs are completely dual to sequent proofs: let  $\bar{\Gamma} = \{\bar{S}:\varphi \mid S:\varphi \in \Gamma\}$ . A tableau proof shows that the set of formulas  $\bar{\Gamma}$  is unsatisfiable, from which the validity of  $\Gamma$  follows. Thus, tableau rule instances preserve satisfiability: if the premiss is satisfiable, then all formulas in at least one extension are satisfiable; tableau rules are DNF representations of their premiss. Tableau closure indicates unsatisfiability of each tableau branch.

As a consequence, both sequent *and* tableau rules can be derived from Theorems 21 and 22. If  $S_{rs}, S_{kl}$  are as in Theorem 21, then the following sequent, respectively, tableau rules are sound and complete for the connective appearing in the premiss, provided that the set of signs is complete with respect to  $S$ :

$$(37) \quad \frac{\Gamma, \bigcup_{l \in J_1} S_{1l}:\varphi_l \quad \cdots \quad \Gamma, \bigcup_{l \in J_{M_2}} S_{M_2l}:\varphi_l}{\Gamma, \{S:\theta(\varphi_1, \dots, \varphi_m)\}}$$

$$(38) \quad \frac{S:\theta(\varphi_1, \dots, \varphi_m)}{\begin{array}{c|c|c} \vdots & \vdots & \vdots \\ S_{1s}:\varphi_s & \cdots & S_{M_1s}:\varphi_s \\ \vdots & \vdots & \vdots \end{array}}$$

EXAMPLE 23. A tableau rule for sign  $\{0, \frac{1}{2}\}$  and three-valued Łukasiewicz implication is:

$$\frac{\{0, \frac{1}{2}\}:\varphi \rightarrow_L \psi}{\begin{array}{c|c} \{\frac{1}{2}, 1\}:\varphi & \{1\}:\varphi \\ \{0\}:\psi & \{0, \frac{1}{2}\}:\psi \end{array}}$$

For the first-order case, with the notation of Theorem 22, one obtains the tableau rule (the sequent rule is similar and skipped, see [Hähnle, 1999] for details):

$$(39) \quad \frac{S:(\lambda x)\varphi(x)}{\begin{array}{c|c|c} \{i_{11}\}:\varphi(c_1) & \cdots & \{i_{m1}\}:\varphi(c_1) \\ \vdots & & \vdots \\ \{i_{1k_1}\}:\varphi(c_{k_1}) & \cdots & \{i_{mk_m}\}:\varphi(c_{k_m}) \\ I_1:\varphi(t_1) & \cdots & I_m:\varphi(t_m) \end{array}}$$

Here,  $(Q_\lambda)^{-1}(S) = \{I_1, \dots, I_m\}$ ,  $I_j = \{i_{j1}, \dots, i_{jk_j}\}$ , the  $c_1, c_2, \dots$  are new Skolem constants, and the  $t_1, \dots, t_m$  are arbitrary ground terms.

As an immediate simplification, note that, if  $I_j = \{i_{j1}\}$  for some  $j$ , then in the corresponding extension it is sufficient to list merely the signed formula  $I_j:\varphi(t)$ . Moreover, one can always delete signed formulas of the form  $N:\varphi$ , because they are trivially valid.

EXAMPLE 24. For three-valued  $Q_\forall$  and the sign  $\{\frac{1}{2}, 1\}$  (see Example 7) one computes  $(Q_\forall)^{-1}(\{\frac{1}{2}, 1\}) = \{\{\frac{1}{2}\}, \{1\}, \{\frac{1}{2}, 1\}\}$ ; and obtains the rule below on the left, which can be simplified to the rule on the right. Systematic simplification procedures for many-valued quantifier rules are described in [Salzer, 1996b; Hähnle, 1998].

$$\frac{\frac{\{\frac{1}{2}, 1\}:(\forall x)\varphi(x)}{\begin{array}{c|c|c} \{\frac{1}{2}\}:\varphi(c_1) \\ \{1\}:\varphi(c_2) \\ \{\frac{1}{2}, 1\}:\varphi(t_2) \end{array}}{\{\frac{1}{2}\}:\varphi(t_1) \quad \{\frac{1}{2}, 1\}:\varphi(t_2) \quad \{1\}:\varphi(t_3)} \quad \frac{\{\frac{1}{2}, 1\}:(\forall x)\varphi(x)}{\{\frac{1}{2}, 1\}:\varphi(t)}$$

Axiomatic sequents denote elementary valid formulas. They are of the form  $\Gamma, \bigcup_i \{S_i:\varphi\}$  such that  $\bigcup_i S_i = N$ .

Tableau closure rule schemata are completely dual and detect primitive unsatisfiability: their form is  $\bigcup_i \{S_i:\varphi\}$  such that  $\bigcap_i S_i = \emptyset$ . In particular, any branch containing a label  $\emptyset:\varphi$  is closed.

To summarize, for each finite-valued logic one can construct in a generic way sound and complete signed sequent and tableau calculi (see also Theorem 25 below). The reverse question, that is, whether for any signed sequent/tableau calculus with truth value sets as signs there is a finite-logic relative to which it is sound and complete, was answered affirmative in [Baaz *et al.*, 1998b] for families of signs  $\mathcal{S}$  having the property

$$(40) \quad \text{for each } i \in N \text{ there are } S_1, \dots, S_r \in \mathcal{S} \text{ such that } \bigcap_{j=1}^r S_j = \{i\}$$

(in this case  $\mathcal{S}$  is also complete as defined in Section 6.2).

Only finite-valued logics immediately yield finitary signed calculi, but there is a way around: in [Ciabattoni, 2000b; Aguzzoli and Ciabattoni, 2000] an effectively computable function  $f$  from  $\mathbf{L}_L^0$ -formulas into  $\mathbf{N}$  is given such that for each  $\mathbf{L}_L^0$ -formula  $\varphi$ ,  $\varphi$  is valid in  $\infty$ -valued  $\mathcal{L}_L^0$  iff it is valid

in  $f(\varphi)$ -valued  $\mathcal{L}_L^0$ , improving on [Mundici, 1987] and the analysis given in Section 4.2. This result is then used in combination with a notational variant of the family of sequent calculi for finite-valued  $\mathcal{L}_L^0$  based on regular signs (35). A labeling mechanism keeps track of the finite truth value set  $N$ , from which each concrete rule must be selected.

At this point we stop the parallel development of signed sequent and tableau calculi, because their duality is fully unfolded. The remaining material is stated for the tableau case only.

**THEOREM 25 (Completeness).** *Let  $\mathcal{S}$  be a complete family of signs with respect to  $\overline{S} \in \mathcal{S}$  for an  $n$ -valued logic  $\mathcal{L}^0$  and  $\varphi$  a  $\mathbf{L}^0$ -formula.*

*If  $\models_S \varphi$ , then there is a tableau proof for  $\overline{S}:\varphi$  constructed with rules of the form (38) according to Theorem 21.*

**Proof.** If  $S = N$ , the result is trivial; otherwise, assume there were no closed tableau for  $\overline{S}:\varphi$ . There must exist a non-closed branch  $B$  in some tableau for  $\overline{S}:\varphi$ , on which all possible rules were applied.

Identify  $B$  with the set of its labels. For each  $p \in \Sigma$  let  $B(p)$  be the set of signed atoms in  $B$  of the form  $S':p$  for some  $S' \subseteq N$ . We define a power interpretation:

$$\mathbf{I}^+(p) = \begin{cases} \bigcap_{S':p \in B(p)} S' & B(p) \neq \emptyset \\ N & B(p) = \emptyset \end{cases}$$

$\mathbf{I}^+$  is well-defined, because  $B$  is not closed. By structural induction on the depth of formulas one proves that  $\mathbf{I}^+$  satisfies all formulas in  $B$ .

The atomic case follows from the definition of  $\mathbf{I}^+$ . Assume  $\mathbf{I}^+$  satisfies smaller formulas than the complex formula  $S':\psi \in B$ . To this formula, a rule according to Theorem 21 was applied based on a DNF representation  $\bigvee_{r=1}^M C_r$  (the representation is not  $\emptyset:\psi$ , because  $B$  is not closed).

By the definition of a tableau, for some  $C_r$ ,  $1 \leq r \leq M$ , all formulas of  $C_r$  are on  $B$  and, by the induction hypothesis, satisfied by  $\mathbf{I}^+$  (again, the empty sign cannot occur, otherwise  $B$  were closed). By Theorem 21,  $\mathbf{I}^+$  satisfies  $S':\psi$ .

As  $\mathbf{I}^+$  satisfies  $B$ , in particular,  $\mathbf{I}^+ \models \overline{S}:\varphi$ , a contradiction to  $\models_S \varphi$ . ■

The first-order version of this result based on rules of the form (39) holds as well. Completeness is shown by combining Theorem 22 straightforwardly with a standard argument from classical logic [Fitting, 1996]. Some more details are given in [Hähnle, 1994a].

**EXAMPLE 26.** Let us prove that (A8) is  $\{1\}$ -valid in three-valued  $\mathcal{L}_L$ . We need to construct a closed tableau for  $\{0, \frac{1}{2}\}:(\forall x)\varphi(x) \rightarrow_L \varphi(t)$  for any given

$t \in \text{Term}_\Sigma^0$ :

$$\begin{array}{ccc}
 & \{0, \frac{1}{2}\}:(\forall x)\varphi(x) & \rightarrow_L \varphi(t) \\
 & \swarrow & \searrow \\
 \{\frac{1}{2}, 1\}:(\forall x)\varphi(x) & & \{1\}:(\forall x)\varphi(x) \\
 | & & | \\
 \{0\}:\varphi(t) & & \{0, \frac{1}{2}\}:\varphi(t) \\
 | & & | \\
 \{\frac{1}{2}, 1\}:\varphi(t) & & \{1\}:\varphi(t)
 \end{array}$$

The rules of Examples 23, 24 are used in the construction.

The examples made it obvious that it is not trivial to compute minimal tableau and sequent rules for a given first-order matrix. One can polynomially reduce the propositional aspect to minimization of Boolean functions (and vice versa, hence this is an NP-complete problem) [Hähnle, 1994a]. A minimization algorithm for finite-valued distribution quantifiers is given in [Salzer, 1996b]. The system MULTlog<sup>13</sup> [Salzer, 1996a; Vienna Group for Multiple Valued Logics, 1996] is a tool that computes optimal sequent and tableau rules for any given finite-valued first order logic. Its output can be used, for example, to parameterize the system  $\mathcal{I}^{\mathcal{A}P}$  [Beckert *et al.*, 1996], a generic tableau-based theorem prover for many-valued sorted first-order logic and (two-valued) equality. The system Deep Thought [Gerberding, 1996] essentially is a re-implementation of  $\mathcal{I}^{\mathcal{A}P}$  in the language C.

#### 6.4 Many-Valued Analytic Cut

We will need the following definitions: a family  $(S_i)_{i \in I}$  of subsets of a set  $N$  is a **covering** of  $N$ , if  $\bigcup_{i \in I} S_i = N$ . A covering of  $N$  is a **partition** of  $N$  if, moreover,  $S_i \cap S_j = \emptyset$  for all  $i, j \in I, i \neq j$ .

Recall that in classical logic the cut rule in sequent calculus is semantically equivalent to the conjunction of all tautologies of the form  $\varphi \vee \neg\varphi$ , where  $\varphi$  is any formula. If  $\varphi$  is restricted to subformulas of the formulas in the root sequent, then one speaks of an **analytic cut rule**.

Hence, formally the cut rule is a DNF representation of truth; as such it can be conceived as a tableau rule schema with empty (always true) premiss:

$$(41) \quad \frac{}{\varphi \mid \neg\varphi}$$

This rule can be immediately generalized to signed DNF representations:

$$(42) \quad \frac{}{S_1:\varphi \mid \cdots \mid S_m:\varphi}$$

where  $m \geq 2$ , and  $\{S_1, \dots, S_m\}$  is a partition of the set  $N$ . In the left truth table in Figure 9 it is demonstrated (with the rule of Example 23)

<sup>13</sup>[www.logic.at/multlog](http://www.logic.at/multlog)

$\rightarrow_L$	0	$\frac{1}{2}$	1
0	1	1	1
$\frac{1}{2}$	$\frac{1}{2}$	1	1
1	0	$\frac{1}{2}$	1

$\rightarrow_L$	0	$\frac{1}{2}$	1
0	1	1	1
$\frac{1}{2}$	$\frac{1}{2}$	1	1
1	0	$\frac{1}{2}$	1

Figure 9. Different coverings of the sign  $\{0, \frac{1}{2}\}$  in truth table of three-valued Lukasiewicz implication.

how the union of the extensions of a tableau rule correspond to a covering of all those fields of a truth table whose entries occur in the sign (here:  $S = \{0, \frac{1}{2}\}$ ) of the premiss. The covering property is necessary for complete tableau rules, but the covering is not necessarily a partition of the fields containing entries from  $S$ : some fields are possibly covered in more than one extension, in the example, the field with entry 0.

With suitable cut rules one can enforce that the extensions of a rule form a partition of the fields to be covered. In the right table of Figure 9 a covering partition of the truth table fields with entries in  $\{0, \frac{1}{2}\}$  of Lukasiewicz implication is displayed. The tableau rule corresponding to it is as follows:

$$(43) \quad \frac{\{0, \frac{1}{2}\}:\varphi \rightarrow_L \psi}{\begin{array}{l|l} \{\frac{1}{2}, 1\}:\varphi & \{1\}:\varphi \\ \{0\}:\psi & \{\frac{1}{2}\}:\psi \end{array}}$$

**DEFINITION 27.** A signed DNF representation  $\bigvee_r C_r$  of  $S:\theta(\varphi_1, \dots, \varphi_m)$  (for  $m \geq 1$ ) is called a **partitioning DNF representation** iff for any two  $C_i$  and  $C_j$  with  $i \neq j$  the set of signed atoms  $C_i \cup C_j$  contains an instance of a tableau closure rule schema. **Partitioning tableau rules** are those based on partitioning DNF representations.

Just as in classical logic with the analytic cut rule (41) it is possible to *derive* many-valued partitioning rules from arbitrary ones with the help of many-valued analytic cut (42). For instance, with the help of the many-valued cut rule  $\overline{\{0\}:\psi \mid \{\frac{1}{2}, 1\}:\psi}$  one can derive rule (43) from the rule in Example 23.

The classical results of Gentzen [1935] on cut elimination in sequent systems can be recast in the finite-valued logic setting. In [Carnielli, 1991] it is observed that the existence of cut-free sequent proofs follows from the completeness of many-valued tableaux and the duality between sequent and tableau calculi, while [Baaz *et al.*, 1994] give a direct and constructive cut

elimination algorithm for singleton signed calculi.

### 6.5 Exploiting Duality Theory

If a truth value set is equipped with a partial order that defines the connectives and quantifiers of a logic, one can exploit dual representations of ordered structures to improve its calculi. Consider, for example, Birkhoff’s well-known representation theorem for finite distributive lattices saying that each element of a finite distributive lattice either is the top (bottom) element or it can be uniquely represented as a meet (join) of meet-(join-)irreducible elements.

**THEOREM 28 (Birkhoff).** *Let  $\mathbf{L}$  be a finite distributive lattice; then  $\mathbf{L}$  is isomorphic to  $\mathcal{O}(\mathcal{J}(\mathbf{L}))$ , where  $\mathcal{O}$  is ordered by set inclusion.*

*Moreover, if  $i \in L$ , let  $M$  be the minimal elements of  $\mathcal{M}(\mathbf{L}) \cap \uparrow i$  and  $J$  the maximal elements of  $\mathcal{J}(\mathbf{L}) \cap \downarrow i$ . Then  $i = \prod M = \bigsqcup J$  (using the convention  $\prod\{\} = \top$  and  $\bigsqcup\{\} = \perp$ ).*

**EXAMPLE 29.** Consider the distributive lattice  $\mathbf{L}$  on the left of Figure 10.  $\mathcal{J}(\mathbf{L})$  is drawn in the middle, and  $\mathcal{O}(\mathcal{J}(\mathbf{L}))$  on the right.

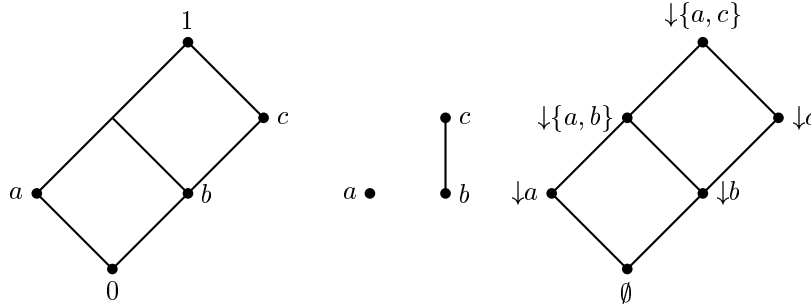


Figure 10. Illustration of Example 29.

Birkhoff’s and other representation theorems [Davey and Priestley, 1990; Goldblatt, 1989; Sofronie-Stokkermans, 2000] allow to replace truth value sets as signs with elements  $j$  of  $\mathcal{J}(\mathbf{L})$ .<sup>14</sup> Formally, using interpretations  $\mathbf{I}^* : \Sigma \rightarrow \mathcal{O}(\mathcal{J}(\mathbf{L}))$ , the following—quasi-classical—semantic conditions hold provided that  $A_\vee$  and  $A_\wedge$  are the lattice operations:

- (44)  $\mathbf{I}^* \models j:p$  iff  $j \in \mathbf{I}^*(p)$
- (45)  $\mathbf{I}^* \models j:(\varphi \vee \psi)$  iff  $\mathbf{I}^* \models j:\varphi$  or  $\mathbf{I}^* \models j:\psi$
- (46)  $\mathbf{I}^* \models j:(\varphi \wedge \psi)$  iff  $\mathbf{I}^* \models j:\varphi$  and  $\mathbf{I}^* \models j:\psi$

<sup>14</sup>In general, the dual space of  $\mathbf{L}$  may be more complicated. I use  $\mathcal{J}(\mathbf{L})$ , which is valid for the case of finite distributive lattices, by way of illustration.



These give directly rise to almost classical tableau and sequent rule schemata. An additional advantage is that  $\mathcal{J}(\mathbf{L})$  is often much smaller than  $\mathbf{L}$ .

A logic solely based on lattice operators is not very interesting, but the representation theorems hold as well, if one adds, for example, monotone operations  $m$  and anti-monotone operations  $a$  on  $\mathcal{J}(\mathbf{L})$ . These induce lattice morphisms  $f_m$  and anti-morphisms  $g_a$  on  $\mathbf{L}$ . The latter can be added to the logical language as additional unary connectives  $F_m$  and  $G_a$ :

$$(47) \quad \mathbf{I}^* \models j:F_m(\varphi) \quad \text{iff} \quad \mathbf{I}^* \models m(j):\varphi$$

$$(48) \quad \mathbf{I}^* \models j:G_a(\varphi) \quad \text{iff} \quad \mathbf{I}^* \not\models a(j):\varphi$$

The last equivalence is realized in a rule like

$$\frac{j:G_a(\varphi)}{a(j):\varphi} .$$

Hence, the signs used in calculi are  $\mathcal{J}(\mathbf{L}) \cup \{\bar{j} \mid j \in \mathcal{J}(\mathbf{L})\}$ . Keeping in mind that  $\mathcal{J}(\mathbf{L})$  is isomorphic to the prime ideals of  $\mathbf{L}$  ordered by set inclusion via  $f : j \mapsto (L - \uparrow j)$  [Davey and Priestley, 1990, Prop. 9.4], one can see that the duals of the family of regular signs (35) are exactly the signs required for  $N = \{0, \frac{1}{n-1}, \dots, 1\}$ . In the general, non-linearly ordered case, not all regular signs are required, because join-irreducible elements alone suffice to construct the dual representation.

To arrive at sound and complete calculi, two more ingredients are required: encoding validity by signed formulas and closure rule schemata. Assume the designated values are  $D = \{1\}$ , then  $\varphi$  is valid iff  $\mathbf{I}^*(\varphi) = N$  for all  $\mathbf{I}^*$  iff never  $\mathbf{I}^*(\varphi) \neq N = \bigsqcup \mathcal{J}(\mathbf{L})$  iff  $\bigvee_{j \in \mathcal{J}(\mathbf{L})} j:\varphi$  has a closed tableau. The single closure rule schema is obviously  $\{j:\varphi, \bar{j}:\varphi\}$ .

The logics based on finite chains and obtained from regular signs with monotone and anti-monotone unary operators were investigated under the label **regular logic** in [Hähnle, 1991b; Hähnle, 1994a]. It was suggested in [Lehmke, 1995; Lehmke, 1996] to take signs from the dual space of a lattice; the totally ordered chain  $[0, 1]$  is discussed in detail.

A systematic treatment of the idea to use dual spaces for lattice-based logics is [Sofronie-Stokkermans, 1997]. In [Sofronie-Stokkermans, 1999b; Sofronie-Stokkermans, to appear, 2000] this is further generalized to ordered relational structures, which can be considered as classes of possible world models, see also Bull and Segerberg's chapter in Volume 3 of this *Handbook* and Section 10 below.

One can express (44)–(48) along with the theory of partially ordered signs in classical propositional (for finite  $N$ ) or in first-order logic [Sofronie-Stokkermans, 1998; Beckert *et al.*, 1999; Sofronie-Stokkermans, to appear, 2000]. For example, (45) becomes

$$(\forall j)(p_{\varphi \vee \psi}(j) \leftrightarrow p_{\varphi}(j) \vee p_{\psi}(j)) .$$

In [Sofronie-Stokkermans, 1999a] translations into classical logic of this kind were used to show decidability of certain non-classical logics that happen to fall in decidable fragments of first-order logic. The idea is pushed one step further in [Ganzinger and Sofronie-Stokkermans, 2000], where suitable translations into of dual space representations classical logic are shown to be a special case of the classical first-order logic with transitive relations that can be efficiently handled with so-called ordered chaining [Bachmair and Ganzinger, 1998]. Thus, refinements based on literal orderings and selection functions that are available for classical logic become directly applicable to many-valued logics. For example, a stronger version of regular negative hyperresolution as discussed below (56), is obtained. In addition, highly sophisticated theorem provers for classical logic become applicable to many-valued logics.

Recall that the many-valued quantifiers  $\forall$  and  $\exists$  are defined via min and max just like  $\wedge$  and  $\vee$  are. Given the above treatment of binary connectives based on lattice operators, it is not surprising that quantifiers  $\Pi$  and  $\Sigma$  defined via  $\sqcap$  and  $\sqcup$  in finite (distributive) lattices possess elegant generic tableau and sequent rules [Hähnle, 1996a; Salzer, 1996b; Hähnle, 1998].

The use of duality theory for automated deduction in many-valued logics has just about started. One could try to base Sofronie-Stokkermans' approach, for example, on Martínez' Priestley-type duality theory, which is available for a wide class of algebraic structures related to many-valued logic, including MV-algebras, linear Heyting algebras, and implicative lattices [Martínez, 1990; Martínez, 1994; Martínez, 1996]. Interestingly, computation of the associated dual spaces can even largely be mechanized [Martínez and Priestley, to appear, 2000].

### 6.6 Normal form computation

Many practically relevant calculi, such as resolution (see Section 6.8), are only defined (or efficient) for formulas in conjunctive normal form.

Up to the right notion of literals, CNF is obviously achieved by repeated application of Theorems 21 and 22, pushing quantifiers in front, if necessary. The details are straightforward and, for the monosigned case, can be found in [Baaz and Fermüller, 1995b].

It is well-known that the naive computation of a CNF of a classical propositional formula is in general exponential in size with respect to the input. This can be avoided using **structure preserving CNF translations** [Tseitin, 1970; Plaisted and Greenbaum, 1986]. The basic idea is to introduce a predicate  $p_\psi(x_1, \dots, x_r)$  for each non-atomic subformula  $\psi$  of  $\varphi$  with free variables  $x_1, \dots, x_r$ , and then replace each occurrence of  $\psi$  with its predicate symbol. Of course, one must state that  $p_\psi$  and  $\psi$  are equivalent.

For each  $\psi = \theta(\rho_1, \dots, \rho_m)$  one obtains the formula

$$(49) \quad (\forall x_1) \cdots (\forall x_r) (p_\psi(x_1, \dots, x_r) \leftrightarrow \theta(p_{\rho_1}, \dots, p_{\rho_m})) \ .$$

Together with the negation of  $\varphi$ , the conjunction of these lines states that  $\varphi$  is unsatisfiable. As each formula (49) has constant depth two, its CNF is of length linear in  $|\varphi|$  (not constant, because of the free variables). There is a linear number of subformulas of  $\varphi$ . Together, one has a quadratic length, satisfiability equivalent<sup>15</sup> CNF of  $\varphi$ . Exactly the same trick is used in the proof of Theorem 16(1).

This idea works for signed formulas just as well, including optimizations based on the polarity of subformula occurrences [Hähnle, 1994d]. The role of classical literals, of course, is taken over by signed atoms: a **signed CNF formula** has the form  $(\forall x_1) \cdots (\forall x_r) \bigwedge_{k=1}^M \bigvee_{l=1}^{J_k} S_{kl}:p_{kl}$ , in which the  $S_j:p_j$  are signed atoms and  $\{x_1, \dots, x_t\}$  are the free variables in the scope. For any  $1 \leq k \leq M$ , the expression

$$(50) \quad (\forall y_1) \cdots (\forall y_m) (S_{k1}:p_{k1} \vee \cdots \vee S_{kJ_k}:p_{kJ_k})$$

where  $\{y_1, \dots, y_m\}$  are the free variables in the scope, is a **signed clause**. Like in classical logic, the quantifier prefix is often not written explicitly and a signed CNF formula is identified with the set of its clauses.

Any signed formula in any finite-valued first-order logic has an at most polynomially larger signed CNF, which is satisfiability equivalent.

Signed clauses are used as if they were sets of signed atoms, which is justified by commutativity, associativity, and idempotency of disjunction. As for sequents, various notations for signed clauses occur in the literature, starting with [Lee and Chang, 1971; Morgan, 1976; Orłowska, 1978].

Like in classical logic, a signed clause is satisfiable if at least one of its literals is satisfiable, a signed CNF formula is satisfiable if all instances of all its clauses are simultaneously satisfiable. The **empty signed clause** has no literals and is denoted with  $\square$ . By definition, it is unsatisfiable.

Exactly like the tableau and sequent rules (see Section 6.5), the translation process can be improved by working with dual representations in the case of truth value lattices [Sofronie-Stokkermans, to appear, 2000].

In [Murray and Rosenthal, 1991b] **signed negation normal form** (NNF) formulas are defined: these are negation-free propositional formulas constructed with  $\wedge$ ,  $\vee$  and with signed atoms as literals. Its relevance is based on the following observation: if, for each connective  $\theta$  occurring in  $D:\varphi$  and all signs  $S$  of a family of signs complete with respect to  $D$ , the number of occurrences of subformulas in a CNF or DNF representation for  $S:\theta(\rho_1, \dots, \rho_m)$  is not greater than in  $S:\theta(\rho_1, \dots, \rho_m)$  itself, then one obtains a linear size signed NNF equivalent of  $D:\varphi$  by repeated substitution

<sup>15</sup>The CNF of  $\varphi$  computed this way has not the same models as  $\varphi$ , because the signature was extended.

of signed subformulas with their CNF or DNF equivalents. This generalizes the well-known fact that classical NNF can be computed in linear time, if no equivalence connectives are present in the input. There are calculi that are specialized for signed NNF, see Section 6.9.

In [Thiele and Lehmke, 1994; Lehmke, 1996] it is observed that every formula of infinite-valued Łukasiewicz logic can be expressed in signed NNF provided that Łukasiewicz sum  $\oplus_L$  and product  $\odot_L$  are used instead of classical disjunction and conjunction, but it can blow up a formula exponentially. They call this **hierarchical normal form**. Again, by using abbreviations for complex subformulas, this can be improved to convert any propositional Łukasiewicz formula into a CNF over so-called bold clauses (see end of Section 6.8).

An internal (in the sense of Section 2.4) DNF was obtained for Łukasiewicz logic in [Mundici, 1996] based on [Mundici, 1991], see also Section 4.1. Clauses are based on  $\odot_L$  instead of  $\wedge$ ; as Łukasiewicz product is not idempotent, a more complex notion of literal is required: formulas  $\varphi$  such that  $f_\varphi$  is a strictly monotone McNaughton function in one variable. Satisfiability of such functions can be characterized with a regular signed atom, that is, a signed atom of the form  $\uparrow i:p$  or  $\downarrow i:p$ . As a consequence, the resolution calculus of [Mundici and Olivetti, 1998] based on this notion of literal does not go beyond signed resolution which is discussed in Section 6.8.

It is non-trivial to prove a McNaughton Theorem for the above notion of literals, that is, to characterize the Łukasiewicz formulas that are literals. This is done in [Aguzzoli, 1998b; Aguzzoli, 1999].

### 6.7 Signed Clause Logic

By virtue of the results of Section 6.6, it is no loss of generality for the purpose of validity checking of formulas in finite-valued first-order logic to work with signed CNF formulas.

Signed CNF formulas are generic or “logic-free” in the sense that their syntax and semantics are fixed and independent of the logic, out of which the translation started. Signed CNF formulas do not contain *any* many-valued connective and are simply a generic and flexible language for denoting many-valued interpretations.

In addition, signed CNF formulas can be motivated from the point of view of **constraint programming** (CP) and **annotated logic programming** (ALP).

If  $N$  is equipped with an ordering, there is a natural notion of a Horn formula. Recall that in classical logic a CNF formula is a **Horn formula** iff each clause contains at most one positive literal. Signed atoms are, in general, neither positive nor negative in any sense, but a natural notion of **polarity** is present in regular signs (35): literals of the form  $\uparrow i:p$  are **positive** while literals of the form  $\overline{\uparrow} i:p$  are **negative**. With this convention,

a **regular Horn formula** is defined exactly as in the classical case. Some calculi based on this formula class are discussed in Section 7.2.

The particular case, when  $N$  is lattice-ordered and  $S$  is an order filter is investigated in annotated logic programming [Kifer and Subrahmanian, 1992] (there,  $S$  is called an *annotation*). Annotated logic programs can be considered as particular signed CNF formulas, see Section 7.3.

On the other hand,  $S:p$  can be read as “ $p$  is constrained to values from  $S$ ” and, hence, as an instance of finite-domain constraint programming [Lu *et al.*, 1997; Castell and Fargier, 1998; Hähnle *et al.*, 2000]. Applications in this spirit are reported for Assumption-Based Reasoning [Haenni and Lehmann, 1998a; Haenni and Lehmann, 1998b].

It is also possible to embed propositional signed CNF formulas into classical monadic first-order logic over finite domains by representing a signed atom  $S:p$ , where  $S = \{i_1, \dots, i_r\}$ , with the following classical formula:

$$(\exists p)(S(p)) \wedge (\forall x)(S(x) \leftrightarrow (S(i_1) \vee \dots \vee S(i_r)))$$

### 6.8 Resolution

A **resolution rule** is a pair  $\langle \mathcal{C}, \varphi \rangle$ , where  $\mathcal{C}$  is a non-empty set of formula schemata, the **premiss** or **parent formulas**, and  $\varphi$  is a formula schema called **resolvent** (of  $\mathcal{C}$ ). A **termination rule** is a non-empty set of formula schemata. A **resolution calculus**  $\mathcal{RK}$  is a finite set of resolution rules and a termination rule.

Each resolution calculus  $\mathcal{RK}$  induces a **provability relation**  $\vdash_{\mathcal{RK}}$  between sets of formulas  $\Psi$  and formulas  $\varphi$ .

- (1) If  $\varphi \in \Psi$ , then  $\Psi \vdash_{\mathcal{RK}} \varphi$ .
- (2) If  $\Psi \vdash_{\mathcal{RK}} \varphi_i$  for  $1 \leq i \leq n$ , and there is an instance of a resolution rule in  $\mathcal{RK}$  with parent formulas  $\varphi_1, \dots, \varphi_n$  and resolvent  $\varphi$ , then  $\Psi \vdash_{\mathcal{RK}} \varphi$ .

If  $\Psi \vdash_{\mathcal{RK}} \varphi$  and  $\varphi$  is an instance of a formula schema in the termination rule of  $\mathcal{RK}$ , then  $\Psi$  is **refuted** (in  $\mathcal{RK}$ ). Obviously, resolution calculi can be considered as Hilbert calculi with an empty set of axioms (up to refutation).

The formula schemata in many-valued resolution calculi are typically restricted to signed clauses of the form (50). In this case, the termination rule contains exactly  $\square$  and it is only possible to refute signed CNF formulas. A resolution calculus is **complete** if every unsatisfiable set of formulas can be refuted and it is **sound** if only unsatisfiable sets of formulas can be refuted.

Only propositional calculi are discussed here, because lifting to first-order logic is done *exactly* like in classical logic and does not provide any insights specific to many-valued logic.

Recall that classical resolution is based on combining clauses that contain unsatisfiable literal sets (closure rule schemata in a tableau framework). Many-valued resolution does exactly the same:

$$(51) \quad \frac{S_1:p \vee C_1 \quad \cdots \quad S_m:p \vee C_m}{C_1 \vee \cdots \vee C_m} \quad \text{if} \quad \bigcap_{1 \leq i \leq m} S_i = \emptyset$$

Compare this to the tableau closure rule defined in Section 6.3. In contrast to classical logic, unsatisfiable literal sets in general are required to admit more than two elements.

EXAMPLE 30. The signed CNF formula  $\Gamma = \{\{0, \frac{1}{2}\}:p, \{0, 1\}:p, \{\frac{1}{2}, 1\}:p\}$  is clearly unsatisfiable, but there is no refutation, if (51) is restricted to  $m \leq 2$ .

There is a complete binary resolution rule for signed logic, though:

$$(52) \quad \frac{S:p \vee C \quad S':p \vee C'}{(S \cap S'):p \vee C \vee C'} \quad \frac{\emptyset:p \vee C}{C}$$

Rule (51) can be simulated by several applications of rules (52). On the other hand, (52) has (intermediate) resolvents which cannot be obtained with (51). Let us call (51) **many-valued hyperresolution** and (52) **many-valued binary resolution**. The literal  $(S \cap S'):p$  in the binary rule is called a **residue**, the rule on the right in (52) is called **reduction rule**.

Similar as for tableau and sequent calculi, historically the monosigned restriction of (52) came first [Lee and Chang, 1971; Lee, 1972; Morgan, 1976; Baaz and Fermüller, 1992]. Orłowska [1978] and Schmitt [1986; 1989] implicitly considered truth value sets as signs in a specialized context.

Rule (51) appeared first in [Hähnle, 1993] and, independently, a close variant in [Haenni and Lehmann, 1998b]; rule (52) is due to [Murray and Rosenthal, 1991b; Murray and Rosenthal, 1993b]. All of the earlier systems included a **many-valued merging** rule

$$(53) \quad \frac{S_1:p \vee \cdots \vee S_m:p \vee C}{(S_1 \cup \cdots \cup S_m):p \vee C}$$

but [Hähnle, 1996b] proved that either of (51) and (52) alone is complete with implicit merging of only identical literals.

It is easy to prove [Murray and Rosenthal, 1994; Hähnle, 1996b] that every CNF formula over signed atoms is logically equivalent to one in which only regular signs (for any given total order) occur. Such a formula is called a **regular formula**. Again using the trick to introduce abbreviations for subformulas (see Section 6.6), one can even show that each signed CNF formula can be reduced to a satisfiability equivalent regular formula, which is only polynomially larger [Beckert *et al.*, 2000b].

In the case of monosigned and regular CNF formulas over a totally ordered truth value set, non-empty residues cannot occur, hence (52) simplifies to

$$(54) \quad \frac{S_1:p \vee C \quad S_2:p \vee C'}{C \vee C'} \quad \text{if } S_1 \cap S_2 = \emptyset$$

which is called **monosigned** or **regular binary resolution**, depending on the form of the signs. Completeness of binary resolution, as well as of ordered resolution and hyperresolution, for monosigned CNF formulas is shown in [Baaz and Fermüller, 1995b]. If  $N$  is totally ordered, one obtains the hyperresolution-like refinements (55) and (56) of regular binary resolution by combining several applications of rule (54) into one [Hähnle, 1994d; Hähnle, 1996b].

$$(55) \quad \frac{\uparrow i_1:p \vee C_1 \quad \cdots \quad \uparrow i_m:p \vee C_m \quad \downarrow j:p \vee C}{C_1 \vee \cdots \vee C_m \vee C} \quad \text{if } \left( \max_{1 \leq k \leq m} i_k \right) > j$$

**regular resolution**

Taking the *maximal*  $i_k$  in the rule above is not strictly necessary: admitting *any*  $i_k > j$  yields a sound and complete calculus, but may lead to longer proofs. For regular formulas, (55) with  $m = 1$  is the same as (54).

EXAMPLE 31. Let  $N = \{0, \frac{1}{2}, 1\}$  and  $\Gamma$  the following regular formula:

$$\{\downarrow 0:p_1 \vee \downarrow \frac{1}{2}:p_2, \uparrow \frac{1}{2}:p_1 \vee \downarrow 0:p_2, \downarrow 0:p_1 \vee \uparrow 1:p_3, \\ \uparrow 1:p_2 \vee \uparrow \frac{1}{2}:p_3, \uparrow 1:p_2 \vee \downarrow 0:p_3\}$$

The last three clauses resolve to  $\downarrow 0:p_1 \vee \uparrow 1:p_2$  by (55), which in turn resolves to  $\downarrow 0:p_1$  with the first clause (by either rule (55) or (54)). From there, one obtains  $\downarrow 0:p_2$  with the second clause. In three more steps the empty clause can be derived.

$$(56) \quad \frac{\downarrow i_1:p \vee C_1 \quad \cdots \quad \downarrow i_m:p \vee C_m \quad \uparrow j_1:p \vee \cdots \vee \uparrow j_m:p \vee C}{C_1 \vee \cdots \vee C_m \vee C}$$

if  $m \geq 1$ ,  $i_l < j_l$  for  $1 \leq l \leq m$ ,  
 $C_1, \dots, C_m, C$  contain only negative literals

**regular negative hyperresolution**

In [Sofronie-Stokkermans, 1998; Sofronie-Stokkermans, to appear, 2000] it is shown that, if  $N$  is a distributive lattice and signs are its prime ideals and their complements, then an analogue of rule (56) is complete, where signs in positive literals are prime ideals and those in negative literals their complements. The  $\bar{U}$ -resolution rule of [Leach *et al.*, 1998] is a special case of Sofronie-Stokkermans' [to appear, 2000] framework.

If  $N$  is a lattice, the following calculus is complete [Beckert *et al.*, 1999]:

$$(57) \quad \frac{\uparrow i:p \vee C \quad \overline{\uparrow j}:p \vee C'}{C \vee C'} \quad \text{if } i \geq j \quad \text{lattice-regular binary resolution} \qquad \frac{\uparrow i:p \vee C \quad \uparrow j:p \vee C'}{\uparrow(i \sqcup j):p \vee C \vee C'} \quad \text{if } i, j \text{ incomparable} \quad \text{lattice-regular reduction}$$

Note that, when  $N$  is totally ordered, the left rule of (57) is the same as (54) for regular formulas.

Regarding techniques to prove completeness of the above-mentioned resolution calculi, one can say that semantic tree arguments [Robinson, 1968] retain much of their clarity. The most straightforward approach is to use  $|N|$ -ary semantic trees [Hähnle, 1994d]. Just as in classical resolution theory, more complex refinements are often better handled by inductive construction of a proof, where the number of atoms or atom occurrences in a formula supplies the induction parameter [Hähnle, 1996b]. Alternatively, via translation of signed logics into classical first-order logic with transitive relations [Ganzinger and Sofronie-Stokkermans, 2000] many completeness results are inherited from classical logic. In fact, where this works, one usually obtains strengthened versions (based on literal orderings and selection functions) of the calculi discussed here.

The restriction of binary resolution to **unit resolution** (the case when one input clause is a unit) is at the heart of the **Davis–Putnam–Loveland procedure** [Davis *et al.*, 1962], together with a case splitting rule (a so-called *pure* rule that discards irrelevant clauses and is not needed for completeness, but for efficiency, is not displayed):

$$(58) \quad \frac{S:p \quad S':p \vee C}{(S \cap S'):p \vee C} \quad \frac{\emptyset:p \vee C}{C} \quad \frac{S_1:p \mid \cdots \mid S_r:p}{(S_j)_{1 \leq j \leq r} \text{ generates } \mathbf{2}^N}$$

This many-valued version was introduced for regular formulas (with total order) in [Hähnle, 1996b]. Regular-DPL was analyzed and improved in [Manyà, 1996; Manyà *et al.*, 1998; Béjar and Manyà, 1999b].

Several resolution-based calculi were also given for logic programs based on signed formulas. They are discussed in Section 7.

Lehmke, in [1995; 1996] describes a resolution system for so-called **weighted bold clauses**. These are signed formulas over  $\mathbf{L}_L^0$ , where the formula part is a **bold clause**: a finite multiset  $M$  of literals plus a “conjunctive correction”  $\delta \in \mathbb{R}$  that determines, whether  $M$  is interpreted as  $\oplus_L$  or  $\odot_L$ :

$$\mathbf{I}((M, \delta)) = \max\{0, \min\{1, \sum_{L \in M} \mathbf{I}(L) - \delta\}\} .$$



Signs are either prime filters of  $[0, 1]$  (that is, open intervals of the form  $(i, 1]$  and closed intervals of the form  $[i, 1]$ ) or again bold clauses. In the first case, the interpretation is as before, see Definition 19. Otherwise,  $\mathbf{I} \models \varphi : \varphi'$  iff  $\mathbf{I}(\varphi') \geq \mathbf{I}(\varphi)$ .

Once again, the trick of introducing abbreviations for complex subformulas (see Section 6.6) is employed to show that any satisfiability problem of Łukasiewicz logic can be expressed with a finite set of weighted bold clauses (the necessity for bold clauses as signs comes from the abbreviations). A sound and complete resolution calculus for weighted bold clauses exists [Lehmke, 1995]. Due to the many parameters of the language (conjunctive correction, two kinds of signs), its rules are quite technical.

The **non-clausal resolution** rule for classical propositional logic takes any two formulas  $\varphi$  and  $\varphi'$  as premisses, in which  $p \in \Sigma$  occurs, and computes the formula  $\varphi\{p/0\} \vee \varphi'\{p/1\}$  as resolvent (the clausal rule is obtained, if  $\varphi$  and  $\varphi'$  are clauses,  $p$  occurs positively in  $\varphi$  and negatively in  $\varphi'$ ). Together with simplification rules of the form  $p \wedge 0 \mapsto 0$  etc., for getting rid of 0 and 1, this is a sound and complete calculus for  $\mathcal{L}_c^0$  [Murray, 1982]. In a finite-valued logic  $\mathcal{L}^0$  one can, in principle, use a straightforward generalization: resolve formulas  $\varphi_i$ ,  $i \in N$ , to  $\bigvee_{i \in N} \varphi_i\{p/i\}$ . In practice, some obstacles must be cleared away first:

- (1) the truth values  $i$  might not be definable in  $\mathcal{L}^0$ ;
- (2)  $\vee$  might not be definable in  $\mathcal{L}^0$ .

Truth values can be replaced by the elements of a suitable subalgebra of the Lindenbaum algebra of  $\mathcal{L}^0$ , so-called **verifiers** [Stachniak, 1988]. One can show that a finite, although in the worst case exponential in  $N$ , number of verifiers is sufficient [Stachniak and O'Hearn, 1990]. The theory of many-valued non-clausal resolution is fully developed in [Stachniak, 1996].

### 6.9 Other Calculi

One deduction method which, like non-clausal resolution, avoids to compute any normal form altogether is the **dissolution rule**. It is available both for classical [Murray and Rosenthal, 1993a] and finite-valued logics [Murray and Rosenthal, 1991b; Murray and Rosenthal, 1994].

Many-valued dissolution operates on formulas in signed NNF (see Section 6.6). The dissolution rule selects in a signed NNF formula an implicitly conjunctively connected pair of signed atoms  $S:p$ ,  $S':p$ , and restructures the formula in such a way that at least one conjoint occurrence of  $S:p$ ,  $S':p$  is replaced with  $(S \cap S'):p$ . Producing  $\emptyset:p$  leads to obvious simplifications like in (52). Every unsatisfiable signed NNF formula is reduced to the empty formula by a finite number of dissolution steps.

A **regular implicate** of a signed NNF formula  $\varphi$  is a clause  $C$  of a regular formula such that  $\varphi \models C$ . A **regular prime implicate** of  $\varphi$  is a regular implicate  $C'$  of  $\varphi$  such that there is no regular implicate  $C \neq C'$  with  $C' \models C$ . The set of prime implicates of a formula is important to know in many applications of classical and many-valued logic, such as diagnosis, circuit minimization, or truth maintenance systems. For some applications it is crucial to avoid computing clause normal form [Ramesh *et al.*, 1997a]. An algorithm for computing regular prime implicates of signed NNF formulas based on many-valued dissolution that avoids normal form is discussed in [Ramesh and Murray, 1994; Ramesh and Murray, 1997].

A conjunctively connected pair  $S:p, S':p$  of signed atoms with  $S \neq S'$  in a signed NNF formula  $\varphi$  can be seen as a generalization of the tableau closure rule and (52). The dual notion, a disjunctively connected pair such that  $S \cap S' \neq \emptyset$ , indicates the presence of certain redundancies of  $\varphi$  in representing  $f_\varphi$  (if  $\mathbf{I}(p) \in (S \cap S')$ ). Such a pair of signed atoms is called an **anti-link** [Ramesh *et al.*, 1997b]. Anti-links can be systematically removed from formulas by a similar rule as dissolution. This can dramatically speed up the computation of irredundant representations of  $f_\varphi$ . The signed NNF case is handled in [Beckert *et al.*, 1998]. An implementation is described in [Geiß, 1997]. The technique of anti-links can improve the methods based on dissolution mentioned above.

In contrast to dissolution, the so-called **TAS method** [Aguilera Venegas *et al.*, 1995] computes a simplified DNF of a given formula in NNF. The input formula is unsatisfiable iff the result is the empty formula. Before each application of the distributive laws towards computing a DNF, unitary models of subformulas are computed and used for simplification. The generalization of the TAS method to signed NNF formulas was done in [Aguilera Venegas *et al.*, 1997] and improved in [Aguilera Venegas *et al.*, 1999].

Both, dissolution and the TAS method can, in principle, be lifted to first-order logic, although the technical difficulties are considerable.

The **connection method** [Bibel, 1987] is another non-clausal deduction system for full first-order logic closely related to tableaux. In [Lee, 1997] a version of the connection method for propositional regular formulas is given.

In [Orłowska, 1991b] a relationship between certain non-classical logics and relation algebras [Tarski, 1941] was established and the concept of **relational proof system** was derived from it. There exist relational proof systems in the spirit of [Rasiowa and Sikorski, 1963] for various many-valued logics including generic ones for arbitrary finite-valued logics [Morgan and Orłowska, 1993; Orłowska, 1991a; Konikowska *et al.*, 1998].

Algorithms based on **local search** outperform deductive decision procedures for checking satisfiability of CNF formulas on some problem classes. In particular, this holds for satisfiable hard random 3-SAT instances, which the fastest implementations of DPL cannot solve within a reasonable time

limit [Selman *et al.*, 1994].

**Regular-GSAT** [Béjar and Manyà, 1999a] is an extension of the classical GSAT procedure [Selman *et al.*, 1992]. It works as follows: first, it tries to find a satisfying interpretation for a regular formula  $\Gamma$  (with a total order on truth values) performing a greedy local search through the space of interpretations. It starts with a randomly generated interpretation  $\mathbf{I}$ . If  $\mathbf{I}$  does not satisfy  $\Gamma$ , then it creates a set  $\mathcal{P} \subseteq \Sigma \times N$ , formed by those variable-value pairs  $\langle p, i \rangle$  that give rise to a maximal decrease (possibly zero or negative) in the total number of unsatisfied clauses of  $\Gamma$  when the truth value of  $\mathbf{I}$  at  $p$  is changed to  $i$ . Next, a propositional variable  $p'$  appearing in  $\mathcal{P}$  and then a truth value  $i'$  from  $\{i \mid \langle p', i \rangle \in \mathcal{P}\}$  are randomly chosen. Finally,  $\mathbf{I}$  is updated to  $i'$  at  $p'$ . Such changes are repeated until either a satisfying interpretation is found or a preset maximum number of changes is reached. The whole process is repeated up to fixed number of times, if no satisfying interpretation is found before.

The superiority of classical local search algorithms over decision procedures such as DPL for certain hard combinatorial problems was found for Regular-GSAT and Regular-DPL [Béjar and Manyà, 1999a] as well. There is experimental evidence that Regular-GSAT outperforms other, including classical, approaches on certain classes of problems [Béjar and Manyà, 1999c].

## 7 LOGIC PROGRAMMING AND DEDUCTIVE DATABASES

This section requires familiarity with the basic notions of Logic Programming as found, for example, in [Lloyd, 1987]. Once again, the discussion is at the propositional level, because lifting to first-order is standard for signed CNF formulas.

### 7.1 Signed Formula Logic Programs

Logic programs are obtained from classical CNF formulas by specifying a preferred direction of evaluation within clauses, which are then called **rule**. Syntactically, rules can be considered as sequents over literals, where the succedent has length one or zero. Traditionally, rules are written in the reverse direction of sequents.

Formally, a **signed formula logic program**<sup>16</sup> (SFLP) [Lu, 1996] is a finite set of **signed rules** of the form

$$(59) \quad S:p \leftarrow S_1:\varphi_1, \dots, S_m:\varphi_m \quad ,$$

where  $S:p$  is a signed atom and the  $S_i:\varphi_i$  are signed formulas of a finite-valued logic. A signed rule is satisfiable iff  $S:p$  is satisfiable or one of  $\overline{S_j}:\varphi_j$ ,

<sup>16</sup>Not be confused with *signed logic programs*, a totally different concept [Turner, 1994].

$1 \leq j \leq m$ , is satisfiable.  $S:p$  is the **head** of the rule,  $S_1:\varphi_1, \dots, S_m:\varphi_m$  its **body**, and  $S_j:\varphi_j$  its **body literals**.

In the light of the results on the existence of a signed CNF of  $S_i:\varphi_i$  (see Section 6.6), one may assume without loss of generality that the  $\varphi_i$  are atomic as well [Lu, 1996]. Similar as in classical logic, if  $S = \emptyset$ , write

$$(60) \quad \leftarrow S_1:\varphi_1, \dots, S_m:\varphi_m \text{ ,}$$

and call such a rule **signed query**. If  $m = 0$ , write

$$(61) \quad S:p$$

and say **signed fact**. If  $m = 0$  and  $S = \emptyset$  one has the empty clause  $\square$ . As usual in logic programming, the goal is to prove  $P \models \bigwedge_{L \in Q} L$  for a logic program  $P$  and literal set  $Q$ , and this is established by showing  $P \cup \{\leftarrow Q\}$  to be unsatisfiable.

The sign  $N$  is not excluded from SFLPs, but it is assumed that any body literal of the form  $N:p$  is automatically deleted; this is justified by the reduction rule in (52).

Unrestricted SFLPs are not easier to solve than general signed CNF formulas (and, hence, as classical CNF formulas): any signed clause (50) can be rewritten into one of the signed rules

$$S_j:p_j \leftarrow \overline{S_1}:p_1, \dots, \overline{S_{j-1}}:p_{j-1}, \overline{S_{j+1}}:p_{j+1}, \dots, \overline{S_m}:p_m$$

for  $1 \leq j \leq m$ . So one cannot expect straight SLD-style (Prolog-like) or unit resolution to be complete for SFLP as it is for classical logic programs. SFLPs encompass the complexities of classical disjunctive logic programming [Lobo *et al.*, 1992] and more.

In this context, two directions for meaningful investigations were suggested:

- (1) define a signed SLD-like resolution calculus for SFLPs, and characterize semantically the valid queries it approximates;
- (2) restrict the language of SFLP so that complete calculi can be defined (analogous to classical logic, where one works with *definite* or *normal* logic programs).

The key tool for the first part are power interpretations, defined in Section 6.2, with the small modification that  $\mathbf{I}^+(p) = \emptyset$  is admitted, hence,  $\mathbf{I}^+ \rightarrow \mathcal{P}(N)$ . Following [Lu, 1996], we speak of an **extended interpretation**. An extended interpretation that satisfies an SFLP  $P$  is called **extended model** of  $P$ . Extended interpretations are naturally ordered according to their “definiteness”:  $\mathbf{I}_1^+ \preceq \mathbf{I}_2^+$  iff  $\mathbf{I}_1^+(p) \supseteq \mathbf{I}_2^+(p)$  for all  $p \in \Sigma$ . Let  $\mathcal{I}^+$  be the family of all extended interpretations (for fixed  $N$  and  $\Sigma$ ),

then  $\langle \mathcal{I}^+, \preceq \rangle$  is a bounded distributive lattice of finite height generated by “definite” extended interpretations  $\mathbf{I}_1^+(p) = \{\mathbf{I}(p)\}$ . Consider the interpretation

$$(62) \quad \mathcal{E}_P(p) = \bigcup_{\mathbf{I}^+ \models P} \mathbf{I}^+(p)$$

defined for each SFLP  $P$ . One can show that it is the **minimal extended model** of  $P$  with respect to  $\preceq$  [Lu, 1996]. Moreover,  $\mathcal{E}_P$  can be effectively computed:

$$(63) \quad \frac{S:p \quad L \leftarrow L_1, \dots, S':p, \dots, L_m}{L \leftarrow L_1, \dots, (\overline{S \cap S'}) : p, \dots, L_m} \quad \text{signed unit resolution} \qquad \frac{S:p \quad S':p}{(S \cap S') : p} \quad \text{signed unit reduction}$$

Both rules are special cases of the left-hand side of (52).

**THEOREM 32.** *For every SFLP  $P$ :  $\mathcal{E}_P \models S:p$  iff  $S:p$  can be derived from  $P$  with rules (63).*

The theorem follows immediately from Theorems 2.18 and 2.19 in [Lu, 1996], where also the top-down version of (63) was shown to be a complete SLD-style resolution calculus for SFLP.<sup>17</sup>

$$(64) \quad \frac{S:p \leftarrow C \quad \leftarrow L_1, \dots, S':p, \dots, L_m}{\leftarrow L_1, \dots, (\overline{S \cap S'}) : p, C, \dots, L_m} \quad \text{signed SLD resolution}$$

**THEOREM 33.** *For every SFLP  $P$  and signed query  $\leftarrow Q$ :  $\mathcal{E}_P \models \bigwedge_{L \in Q} L$  iff  $\square$  can be derived from  $P$  starting with  $\leftarrow Q$  and using rule (64).*

**EXAMPLE 34.** For the SFLP  $P = \{\{1\}:p \leftarrow \{0\}:q, \{1\}:p \leftarrow \{1\}:q, \{1\}:q \leftarrow \{1\}:p\}$ , one has  $\mathcal{E}_P \equiv N = \{0, 1\}$ . Nothing definite can be said about any atom, indeed, no rule in (63) is applicable, and only the empty query succeeds with (64). This is not surprising, since  $P$  corresponds to a non-Horn classical CNF formula.

If we add  $\{0\}:q$  to  $P$ , then  $\mathcal{E}_P(p) = \{1\}$ , but  $\mathcal{E}_P(q) = \emptyset$ . This means  $P$  is not satisfiable by any standard interpretation; indeed, the corresponding classical CNF formula is unsatisfiable.  $\mathcal{E}_P$ , however, gives more information than mere unsatisfiability, for example,  $\mathcal{E}_P \not\models \{0\}:q$ . It is possible to reconstruct paraconsistent logic programming [Kifer and Lozinskii, 1992] within SFLP [Lu, 1996].

<sup>17</sup>Theorems 5.4 and 5.8; there is a mistake in the statement of the latter, which is corrected in Theorem 33 below.

Every SFLP is satisfied by the **trivial extended model** that constantly assigns  $\emptyset$ .

For  $P' = \{\{a, 0\}:p, \{b, 0\}:p\}$  one has  $\mathcal{E}_P(p) = \{0\}$ , which is “definite”.

Note, that SLD resolution is optimized for first-order logic and not very efficient on the propositional level.

### 7.2 Regular Logic Programs

The second direction of research into SFLP starts by restricting the language. The key property of classical definite logic programs is that they are Horn formulas. If  $N$  is totally ordered, regular Horn formulas inherit many properties of their classical counterparts, such as model theoretic characterizations [Hähnle, 1996b].

Complete refinements of regular binary resolution (54) for regular Horn formulas over a *totally* ordered truth value set are **regular unit resolution** [Hähnle, 1994d] (the case  $C = \square$  in rule (54)) and also **regular positive unit resolution** [Manyà, 1996] (where, in addition, the unit input clause must be a positive literal).

A slightly different strand of development, sometimes encountered under the label **fuzzy logic programming**, is (selectively) represented by the papers [Klawonn and Kruse, 1994; Escalada-Imaz and Manyà, 1995; Yasui and Mukaidono, 1996; Vojtáš and Paulik, 1996; Vojtáš, 1998]. Signs are used here in a more restrictive way in that they are attached to rules, not to single literals. On the other hand, instead of the classical  $\vee$ ,  $\wedge$ , and  $\rightarrow$ , connectives  $\oplus_t$ ,  $\odot_t$ , and  $\rightarrow_t$  based on various continuous  $t$ -norms over  $[0, 1]$  are admitted. Most of the cited papers define sound and complete versions of SLD-resolution.

Another possibility to generalize is to consider non-linear orderings. [Beckert *et al.*, 1999] show that the rules below are complete for regular Horn formulas, if  $N$  is a finite upper semi-lattice.

(65)

$$\frac{\uparrow i:p \quad \overline{\uparrow j:p} \vee C}{C} \qquad \frac{\uparrow i:p \quad \uparrow j:p}{\uparrow(i \sqcup j):p}$$

if  $i \geq j$   $i, j$  incomparable

**lattice-regular positive unit resolution**    **lattice-regular unit reduction**

The discussion on duality in Section 6.5 shows that for distributive lattices further improvements are possible, for example, using only prime filters and their complements as signs [Sofronie-Stokkermans, 1998]. The opposite direction, to consider more general orderings than lattices, requires more thought: a core feature of any efficient deduction procedure for Horn formulae is the possibility to represent the conjunction of two unit clauses as a

single unit clause as witnessed by lattice-regular unit reduction (65). This means that the set of signs must be closed under intersection:

$$(66) \quad \text{For all } i, j \in N \text{ there is a } k \in N \text{ such that } \uparrow i \cap \uparrow j = \uparrow k$$

Every non-empty, finite partially ordered set satisfying (66) is, however, already an upper semi-lattice. Therefore, it is inevitable to generalize the language of signs if one wants to go beyond lattices. The most natural choice is to work with upsets of  $N$  that need not be filters, in other words, signs of the form  $\uparrow S$ , where  $S \subseteq N$ . Details are in [Beckert *et al.*, 1999].

We defined regular signs as the non-trivial order filters of  $N$  generated by single elements and their complements (35) and this is the definition most of the work on regular signs is based upon. If  $N$  is totally ordered, then (35) is equivalent to (in fact, this was the original definition in [Hähnle, 1994a]):

$$(67) \quad \mathcal{S}_{\text{regular}}^* = \{\uparrow i \mid i \in N, \uparrow i \neq N\} \cup \{\downarrow i \mid i \in N, \downarrow i \neq N\} .$$

The set of signs  $\mathcal{S}_{\text{regular}}^*$  differs from  $\mathcal{S}_{\text{regular}}$  in the non-linear case; it is less expressive than the latter and not suitable for defining a Horn fragment, because it is not closed under set complement. See also Table 1 below for complexity results for associated satisfiability problems.

### 7.3 Annotated and Paraconsistent Logic Programs

An **annotated logic program** (ALP) [Kifer and Subrahmanian, 1992], sometimes called **paraconsistent logic program**, has the same form as (59), but all occurring signs are positive regular signs and  $N$  is a complete lattice. It should be clear then, that an ALP is simply a regular Horn formula (based on lattice-orders) in rule notation. In this sense, ALP is a special case of SFLP [Lu *et al.*, 1993; Lu *et al.*, 1998].

It is possible to go the other way round: for a truth value set  $N$ , let  $\mathbf{2}^n$  be the inverted power set lattice of  $N$ , that is,  $\emptyset$  is on top. Now each sign  $S \subseteq N$  generates an order filter in  $\mathbf{2}^n$  that is comprised of all subsets of  $S$ , that is, values of extended interpretations satisfying  $S$  [Lu, 1996].

Historically, ALPs preceded regular Horn formulas [Kifer and Lozinskij, 1989; Kifer and Lozinskii, 1992]. They were motivated by Belnap's paraconsistent logic [Belnap Jr., 1977], which is based on the truth sublattice of *FOUR* discussed in Section 2.5. The logic based on this lattice was modeled with Petri Nets [Murata *et al.*, 1991]. Various calculi for ALPs were developed and partly implemented [Blair and Subrahmanian, 1989; Lu *et al.*, 1991; Kifer and Subrahmanian, 1992; Messing and Stackelberg, 1995; Leach and Lu, 1996; Lu, 1996]. Non-monotonic ALP is considered in [Bell *et al.*, 1994], while [Lu and Rosenthal, 1997] is an overview of generalized logic programming.

Remarkably, the general theory of annotated logic programs [Kifer and Subrahmanian, 1992] allows for variables and function symbols to occur in the signs, thus blurring the distinction between truth values and terms. Variables in signs are also admitted in [Hähnle *et al.*, 2000].

#### 7.4 *Deductive Databases & Knowledge Representation*

Relational databases correspond to recursion-free, safe Datalog logic programs (no function symbols) [Ullman, 1988]. Expressiveness of queries, respectively, conciseness of the data can be improved by allowing recursion or admitting function symbols or non-Horn rules. All of these drastically increase computational cost (the first two relaxations even imply undecidability), in contrast to SFLP and annotated logic programming ALP. Therefore, signed logic techniques have been suggested to enhance deductive database technology [Lakshmanan and Sadri, 1994; Subrahmanian, 1994; Ng and Subrahmanian, 1993] and knowledge representation systems [Messing, 1997].

Deductive tasks in databases differ from satisfiability checking: while consistency of a database is important, it is often guaranteed by non-deductive means and rarely performed.

More important tasks are query answering, updates, and query optimization. **Query answering** means to decide whether a conjunction of atoms logically follows from a given logic program. It is important for such algorithms that they can take advantage from the fact that a logic program does not change between subsequent queries, that is, there should be some sort of compilation. This compilation process should be incremental in order to allow **database updates**. As these requirements are fulfilled by standard logic programming techniques, one looks for deductive algorithms in signed logic that closely resemble those for classical logic programming [Leach and Lu, 1996; Messing and Stackelberg, 1995; Messing, 1995]. Another topic derived from classical database theory is the optimization of queries [Lakshmanan and Sadri, 1994].

The expert system architecture Milord II [Puyol-Gruart, 1996] makes essential use of regular Horn formulas. Large knowledge bases are split into modules, each equipped with a local many-valued logic [Agustí-Cullell *et al.*, 1991; Agustí-Cullell *et al.*, 1994]. A mapping determines the global truth degree from the truth degrees computed by each module.

In [Arieli and Avron, 1998] it is argued that the *bilattice FOUR* (Section 2.5) plays a similar role as the two-valued Boolean algebra in the realm of reasoning with incomplete and uncertain information, rather than Belnap's [Belnap Jr., 1977] four-valued truth lattice.



### 7.5 Many-Valued Semantics of Logic Programs

The knowledge lattice  $\langle \{F, T, \perp, \top\}, \leq_k \rangle$  of the bilattice  $\mathcal{FOUR}$  (Definition 10) is a key tool for characterizing the declarative semantics of SLDNF-resolution of general logic programs.

Observe that both sublattices of  $\mathcal{FOUR}$  are compatible in the sense that prime filters/ideals (that is: regular signs) in one lattice have the same status in the other. It is possible, therefore, to handle connectives from both lattices with the same set of regular signs in  $\mathcal{FOUR}$ . In fact, even  $\uparrow_k F$  and  $\uparrow_k T$  suffice.

The operational semantics of general logic programs  $P$  can be modeled with the signed formula  $\uparrow_k T:P$ , if the operators in  $P$  are defined by a suitable  $\mathcal{FOUR}$ -valued matrix. One possibility is: define  $A_{\leftarrow}(i, j) = T$  iff  $j \leq_k i$  and  $A_{\leftarrow}(i, j) = F$  otherwise, that is, rules propagate knowledge and have a two-valued definite result; combination of rules and body literals is done simply by  $\sqcap$  in the truth lattice; to model negation as failure, think of  $\top$  to encompass both  $F$  and  $T$  and  $\perp$  none of them. Now negation just switches support of  $T$  and  $F$ , thus:  $A_{\neg}(F) = T$ ,  $A_{\neg}(T) = F$ , and  $A_{\neg}(i) = i$  otherwise. It is straightforward to transform the signed formula  $\uparrow_k T:P$  into a regular Horn formula or annotated logic program over  $\mathcal{FOUR}$ . At that point, the results stated in Section 7.1 apply.

The model of classical general logic programming sketched above is essentially due to [Stärk, 1991; Plaza, 1996]. The latter shows that SLDNF-resolution is faithfully captured therein. The idea to use three- and four-valued logics to model the semantics of logic programs with negation is due to [Fitting, 1985; Kunen, 1987; Fitting, 1991a].

## 8 MANY-VALUED LOGIC IN CIRCUIT DESIGN

### 8.1 Usage of Many-Valued Logic in Circuit Design

The idea to generalize binary logic technology in hardware designs to many-valued logic is natural and has been intensely explored for at least 30 years. A closer look reveals that many-valued logic can be put to use in quite different ways:

**Many-Valued Hardware.** One can build genuinely many-valued hardware that internally works typically with  $2^n$  voltage levels, where  $n \geq 2$ . The most important advantage is that a higher integration can be achieved, because a  $2^n$ -valued gate or cell is only slightly larger than a two-valued one.

- Many-valued technology is used in various commercially available memory devices [Bauer *et al.*, 1995]. Besides larger storage capacity on the

same area, they offer faster access as well. The need to map binary addresses to many-valued cells affords slightly more complicated periphery, otherwise, existing manufacturing techniques scale up well. A nice overview is [Gulak, 1998].

- The design of complex arithmetic and logic circuits has been proven to be technologically feasible as well. In addition to smaller size, they consume less power and have fewer internal connections [Kawahito *et al.*, 1994]. On the other hand, many-valued design tools of comparable quality to binary ones are not yet available. Together with the higher complexity of many-valued circuits, this leads so far to comparatively higher development cost. As a consequence, complex many-valued circuits did not yet make a commercial impact, which may change as physical barriers are in sight for binary technology.

**Modeling below the Gate Level.** Switch-level models (SLM) [Bryant, 1984; Hayes, 1986] are a well-established formal framework for modeling properties of *binary* circuits on the transistor level in considerable detail. SLMs are used to model phenomena such as propagation and resolution of undefined values, hazard detection, degradation effects, varying capacities, pull-up transistors or depletion mode transistors, see [Eveking, 1991] for a very exhaustive list. It is important to note, however, that all dimensions are symbolic values.

It has been demonstrated that many-valued logic is an appropriate modeling language at the switch-level [Hayes, 1986; Eveking, 1991], which makes it possible to verify designs as well [Bryant and Seger, 1991; Hähnle and Kernig, 1993].

**Representation, Minimization, Synthesis.** Just as in the binary case, it is important to represent, minimize and optimize many-valued functions, that is, to express them in a given logical language with a minimal number of connectives, which is the basis for a realization in hardware. For representation of binary and many-valued functions decision diagrams (Section 8.2) are mostly used in modern design tools.

A large number of many-valued minimization methods can be found in the literature. They fall into two categories:

- *heuristic methods*, for example, ESPRESSO-MV [Rudell and Sangiovanni-Vincentelli, 1987]; [Rudell and Sangiovanni-Vincentelli, 1987]; these are used in practice, because
- *systematic methods* often cannot deal efficiently enough with larger designs [Sasao, 1999], but are successful in certain special contexts, such as designs using field programmable gate arrays (FPGA).

What seems surprising at first is that many-valued minimization can be used for optimizing *binary* designs as well [Sasao, 1981; Sasao, 1993b]. For instance, a single  $2^n$ -valued signed atom can encode  $n$  binary literals. Accordingly, a signed CNF formula (with truth values in  $\{1, \dots, 2^n\}$ ) represents a so-called programmable logic array (PLA) with  $n$ -bit decoder.

It is outside the scope of this article to present technical details of many-valued circuits. Instead, I would like to point out some connections between many-valued switching theory and proof theory of many-valued logics.

## 8.2 Decision Diagrams

Decision diagrams (DD) are a family of data structures originally developed for efficient representation and manipulation of Boolean formulas, but now successfully used for many purposes in computer science, in particular in circuit design tools [Burch *et al.*, 1990]. A standard reference for binary decision diagrams (BDD) is [Bryant, 1986], a survey of BDDs is contained in [Bryant, 1992] and, more recently and exhaustively, in [Sasao and Fujita, 1996; Minato, 1996]. An introduction to BDDs from the point of view of automated theorem proving is [Strother Moore, 1994]. DDs in many-valued logic are discussed in [Kam *et al.*, 1998].

BDDs are a representation of two-valued (Boolean) functions based on the three-place **if-then-else** operator **if**  $i$  **then**  $j$  **else**  $k \equiv (i \wedge j) \vee (\neg i \wedge k)$ , where

$$A_{\text{if-then-else}}(i, j, k) = \begin{cases} j & \text{if } i = 1 \\ k & \text{if } i = 0 \end{cases} .$$

Every Boolean function can be expressed with a formula that contains no connective but **if-then-else**, logical constants 0 and 1, and where variables occur exactly as the first argument of **if-then-else** connectives. For instance,

$$(68) \quad p \rightarrow q \equiv \text{if } p \text{ then (if } q \text{ then 1 else 0) else 1} .$$

Such a representation is called an **if-then-else normal form**, a **BDD**, or a **Shannon tree**. A systematic way to obtain a BDD representation of a formula or logical function  $\varphi$  is provided by the so-called **Shannon expansion**.<sup>18</sup> Assume that  $p$  is a variable occurring in  $\varphi$ , then:

$$(69) \quad \varphi \equiv \text{if } p \text{ then } \varphi\{p/1\} \text{ else } \varphi\{p/0\} .$$

Recursive application of (69) to  $\varphi\{p/i\}$  ( $i = 0, 1$ ) and the variables in it, plus replacement of variable-free formulas with their valuation, obviously

<sup>18</sup>In the BDD and function minimization literature this equation is often attributed to Shannon [1938] or Akers [1978], however, it appears already in [Boole, 1854]. Expansions are sometimes called **decomposition**.

yields a BDD representation of  $\varphi$ , see Figure 11 (top right) for the example in (68) in tree notation.

By definition of **if-then-else**, the right-hand side of the equivalence (69) is valid iff the signed atom  $\{1\}:p$  and  $\varphi\{p/1\}$  are valid or the signed atom  $\{0\}:p$  and  $\varphi\{p/0\}$  are valid.

Usually, BDDs are assumed to be reduced and ordered, abbreviated **ROBDD**. **Reduced** means that the syntax tree of a BDD representation is turned into a directed acyclic graph by identifying isomorphic subtrees and applying the following simplification rule, wherever possible:

$$(\text{if } i \text{ then } j \text{ else } j) \equiv j .$$

**Ordered** means that relative to a given total ordering  $<$  on variables, whenever  $q$  occurs in the body of **if**  $p \dots$  then  $p < q$  must hold. An important property of ROBDDs is that for a given variable ordering the ROBDD representation of any formula is unique, rendering ROBDDs a **canonical normal form** for Boolean functions. This property is crucial for the implementation of BDD manipulation packages, because it allows BDDs to be hashed, which helps to avoid redundant computations.

BDDs are extended to a **many-valued decision diagram** or **MDD** for finite-valued logics in a natural way using an  $(n + 1)$ -ary **switch** connective in  $n$ -valued logic:

$$A_{\text{switch}}(i, j_0, \dots, j_{n-1}) = \begin{cases} j_0 & \text{if } i = 0 \\ j_1 & \text{if } i = \frac{1}{n-1} \\ \dots & \dots \\ j_{n-1} & \text{if } i = 1 \end{cases}$$

Orłowska [1967] gave a proof procedure for propositional Post logic based on an MDD-like structure, using a different (and slightly cryptic) notation. ROMDDs were defined first by [Thayse *et al.*, 1979] and were rediscovered in connection with the growing interest in BDD methods by [Srinivasan *et al.*, 1990], now called **canonical function graph**. Like their binary counterparts  $n$ -valued MDDs are functionally complete, an ROMDD representation is a canonical normal form for any  $n$ -valued function, and they can be computed with the help of a generalized Shannon expansion:

$$(70) \quad \varphi \equiv \begin{cases} \text{switch } i \\ \text{case } 0: & \varphi\{p/0\}; \\ \text{case } \frac{1}{n-1}: & \varphi\{p/\frac{1}{n-1}\}; \\ \dots & \dots \\ \text{case } 1: & \varphi\{p/1\} \end{cases}$$

Obviously, the right-hand side of this equivalence is valid iff for some  $i \in N$  both the signed atom  $\{i\}:p$  and the formula  $\varphi\{p/i\}$  are valid.

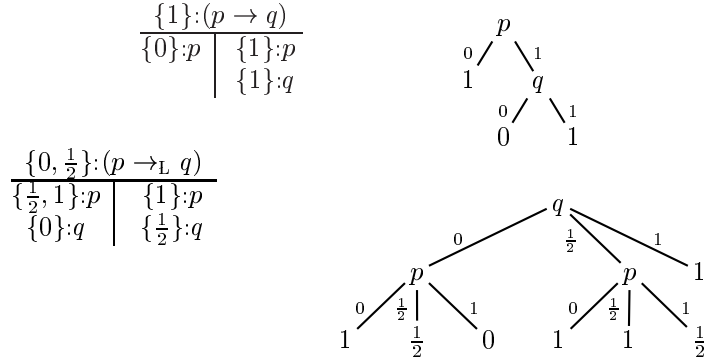


Figure 11. Partitioning tableau rules and DDs.

DDs and partitioning DNF representations (Definition 27) bear a close relationship. Consider the partitioning tableau rule for  $\{1\}:(p \rightarrow q)$  and the ROBDD for  $p \rightarrow q$  depicted in the top row of Figure 11. Edges corresponding to **then** and **else** branches are labeled with 1 and 0, respectively, in the ROBDD. An edge labeled with  $i$  going out of a node  $p$  can be seen as an assertion of the truth value  $i$  to  $p$ , in other words, as the signed formula  $\{i\}:p$ . The following relationship between classical signed tableaux with partitioning rules and (RO)BDDs holds [Posegga, 1993]: if  $B$  is the set of signed atoms corresponding to the edges on a path ending with 1 in a (RO)BDD for  $\varphi$ , then there is an open branch in any tableau for  $\{1\}:\varphi$  with partitioning rules containing exactly  $B$  as signed atoms and vice versa.

This relationship generalizes to many-valued signed tableaux with partitioning rules and (RO)MDDs. Let  $B$  be the set of signed atoms corresponding to edges on a path ending with  $j \in S$  in a (RO)MDD for  $\varphi$ . Then there is an open branch in any signed tableau with partitioning rules for  $S:\varphi$  containing signed atoms  $B'$  such that for each  $\{i\}:p \in B$  there is  $S':p \in B'$  with  $i \in S'$ .

Conversely, if  $B'$  are the signed atoms on an open branch in a signed tableau with partitioning rules for  $S:\varphi$ , then there is  $j \in S$  and a path in a (RO)MDD for  $\varphi$  ending with  $j$  and set  $B$  of signed atoms corresponding to its edges, such that for each  $S':p \in B'$  one has  $\{i\}:p \in B$  for some  $i \in S'$ .

For instance, an ROMDD for  $p \rightarrow_L q$  is displayed in Figure 11 (bottom right—this time, variable  $q$  was expanded first); the paths ending with values from  $\{0, \frac{1}{2}\}$  correspond to the extensions of the partitioning tableau rule for  $\{0, \frac{1}{2}\}:(p \rightarrow_L q)$  depicted on the left.

The relationship between open tableau branches and paths in MDDs is more straightforward, if one admits *set-valued edges* in DDs, which correspond directly to atoms signed with truth value sets. Such diagrams,

however, are not a canonical representation and, hence, less important in practice.

Yet another way to look at DDs is as a non-clausal version of the (signed) DPL procedure (58), whose splitting rule in the singleton sign case is exactly generalized Shannon expansion (70). The main difference is that DPL-proofs are trees and not DAGs.

We stress that DD methods are essentially confined to the ground case, because total ordering of non-ground atoms is not compatible with applying substitutions to them, hence the canonicity property is lost [Posegga and Schmitt, 1995].

Implementations of MDDs are reported in [Srinivasan *et al.*, 1990; Sasao, 1996; Kam *et al.*, 1998].

### 8.3 Polynomial Expressions

The set of paths in a BDD or MDD can be seen as a Boolean polynomial over signed atoms with truth values as coefficients. Let us write a signed atom of the form  $\{i\}:p$  as  $p^i$ , “ $\wedge$ ” as “ $\cdot$ ”, and “ $\vee$ ” as “ $+$ ”. Then, by (69), for example,

$$\begin{aligned}
 p \rightarrow q &\equiv p^1 \cdot (1 \rightarrow q) + p^0 \cdot (0 \rightarrow q) \\
 &\equiv p^1 \cdot (q^1 \cdot (1 \rightarrow 1) + q^0 \cdot (1 \rightarrow 0)) + \\
 &\quad p^0 \cdot (q^1 \cdot (0 \rightarrow 1) + q^0 \cdot (0 \rightarrow 0)) \\
 &\equiv p^1 \cdot q^1 \cdot 1 + p^1 \cdot q^0 \cdot 0 + p^0 \cdot q^1 \cdot 1 + p^0 \cdot q^0 \cdot 1 \\
 (\text{reduction: } &\equiv p^1 \cdot q^1 \cdot 1 + p^0 \cdot 1) \text{ ,}
 \end{aligned}$$

where the last line can also be read off the paths (ending with 1) in the ROBDD in Figure 11. In switching theory [Hachtel and Somenzi, 1996; Sasao, 1999], such a polynomial expression is known as **sum-of-products expression** or **SOP**, the products  $p_1^{i_1} \cdot \dots \cdot p_k^{i_k}$  of the non-reduced form are called **minterm**, and the coefficients 0, 1 **discriminant**. Many readers will have realized that an SOP simply is a DNF.<sup>19</sup>

Replacing disjunction “ $+$ ” with exclusive or “ $\oplus$ ” yields a so-called **exclusive-or-of-products expression** **ESOP**, which is of great practical value [Hachtel and Somenzi, 1996; Sasao, 1999]. We saw that the Shannon expansion yields SOP expressions. Because of  $p \wedge q \equiv 0$  iff  $p \oplus q \equiv p \vee q$ , one obtains ESOP expressions in the same way. There are many kinds of polynomial representations. Most of them, for example, general SOPs and ESOPs are not canonical (that is, more than one (E)SOP has the same ROBDD). On the other hand, certain restrictions of (E)SOPs are canonical: an ESOP, in which only positive signed atoms of the form  $p^1$  occur,

<sup>19</sup>(Signed) NNF formulas (Section 6.6) are known in switching theory as well under the name **factored expression**.

and where each product occurs at most once is called **positive polarity Reed-Muller** (PPRM) expression and is of considerable practical interest in circuit synthesis. The relevance of polynomial expressions comes from the fact that they can be immediately realized with two-level networks. In addition, many kinds of (E)SOP expressions can be read off decision diagrams (based on suitable expansions), which gives directly a method for synthesis of the corresponding circuits. One drawback is that the (E)SOPs produced by DDs are not necessarily of minimal size.

In the many-valued case one may, of course, consider arbitrary signed atoms  $S:p$  in minterms. Written  $p^S$ , they are known in logic design under the name **set literal** or **universal literal** [Sasao, 1981; Dueck and Butler, 1994]. On the other hand, there is no reason to restrict oneself to unary functions as an expansion base for a polynomial representation: recall that (70) can alternatively be written as

$$\varphi \equiv \Sigma_{i \in N} (p^i \cdot \varphi\{p/i\}) ,$$

which can be generalized to

$$(71) \quad \varphi \equiv \Sigma_{i \in I} (\Psi_i \cdot \varphi_i) ,$$

for a certain base of Boolean-valued functions  $(\Psi_i)_{i \in I}$  over the variables of  $\varphi$  and certain many-valued functions  $\varphi_i$  depending on the  $\Psi_i$ . One needs, of course, to impose restrictions on the  $\Psi_i$  to make this work. Examples of generalized expansions are Löwenheim's *orthonormal expansions*, see [Brown, 1990]. Expansions for many-valued logic were investigated in many papers, see [Perkowski, 1992] for an overview and attempt at systematizing the plethora of existing expansions. The complexity of expressions in terms of minimal length is compared in many papers.

Some expansions, for example (70), correspond to partitioning DNF representations (Definition 27), but in general, expansions are a much more general mechanism. With DNF representations and, hence, tableau and sequent rules, the usual point of view taken is that one eliminates a connective by applying it, whereas expansion schemata as used in switching theory often eliminate a variable. Through many-valued cut rules both notions are linked.

## 9 COMPLEXITY AND DECIDABILITY

As it is the case for classical logic and other non-standard logics, a variety of complexity-related questions can be asked in the context of many-valued logic. Some questions, such as the complexity of the sets of satisfiable and valid formulas in various logics, are completely standard; others, such as the maximal size of DNF/CNF representations of many-valued operators, only make sense in a many-valued context. We assume the reader to be

familiar with standard concepts from complexity [Papadimitriou, 1994] and recursion theory [Soare, 1987].

### 9.1 Satisfiability and Validity

Given an  $N$ -valued (propositional or first-order) logic  $\mathcal{L}$ , a truth value set  $D \subseteq N$ , denote with  $D\text{-SAT}_{\mathcal{L}}$  the  $D$ -satisfiable formulas and with  $D\text{-VAL}_{\mathcal{L}}$ , the  $D$ -valid formulas of  $\mathcal{L}$ . Further, let  $\text{SAT}_{\mathcal{L}}=\{1\}\text{-SAT}_{\mathcal{L}}$  and  $\text{VAL}_{\mathcal{L}}=\{1\}\text{-VAL}_{\mathcal{L}}$ .

**THEOREM 35.**  *$\text{SAT}_{\mathcal{L}}$  is NP-complete and  $\text{VAL}_{\mathcal{L}}$  co-NP-complete, whenever  $\mathcal{L} \in \{\mathcal{L}_c^0, \mathcal{L}_G^0, \mathcal{L}_L^0, \mathcal{L}_{\Pi}^0\}$  for any choice of  $N$ .*

**Proof.** (Sketch) For  $\mathcal{L}_c^0$  this is, of course, Cook’s Theorem.

For hardness of the problems associated to a logic  $\mathcal{L}^0 \in \{\mathcal{L}_G^0, \mathcal{L}_L^0, \mathcal{L}_{\Pi}^0\}$ , first observe that by virtue of equations (1), (3) one may assume that  $\mathbf{L}^0$  contains all classical formulas  $\varphi \in \mathbf{L}_c^0$ . Now it is sufficient to define for each  $\varphi \in \mathbf{L}_c^0$  an  $\mathbf{L}^0$ -formula  $\varphi^*$  that restricts  $\mathcal{L}^0$ -interpretations of  $\varphi$  to values in  $\{0, 1\}$ . Then  $\varphi \in \text{SAT}_{\mathcal{L}^0}$  iff  $\varphi \wedge \varphi^* \in \text{SAT}_{\mathcal{L}^0}$ . In the case of  $\mathcal{L}_L^0$ , for example, one can use  $\varphi^* = \bigwedge_{p \text{ occurs in } \varphi} (p \vee \neg p)$ , in the other logics similar formulas work.

For finite-valued logics, NP-, respectively, co-NP-easiness is straightforward, so assume  $N = [0, 1]$  for the remaining cases.

$\text{SAT}_{\mathcal{L}_G^0}=\text{SAT}_{\mathcal{L}_{\Pi}^0}=\text{SAT}_{\mathcal{L}_c^0}$  is shown by giving straightforward polynomial-size embeddings of the latter into the former [Hájek, 1998].

$\text{VAL}_{\mathcal{L}_G^0}$  [Baaz *et al.*, 1998a]: assume there is an interpretation  $\mathbf{I}$  of  $\varphi \in \mathbf{L}_G^0$  over variables  $p_1, \dots, p_m$  such that  $\mathbf{I}(\varphi) < 1$ . There is a trivial order-isomorphism  $o$  mapping  $\mathbf{I}(p_1), \dots, \mathbf{I}(p_m)$  into  $\mathbf{m} + \mathbf{2} = \{0, \frac{1}{m+1}, \dots, \frac{m}{m+1}, 1\}$ . All Gödel operations  $f$  have the property that  $f(i_1, \dots, i_{\alpha(f)}) \in \{i_1, \dots, i_{\alpha(f)}\} \cup \{0, 1\}$ , hence,  $o(\mathbf{I})(\varphi) < 1$ ; now it suffices to guess such an interpretation over  $\mathbf{m} + \mathbf{2}$  and check that  $o(\mathbf{I})(\varphi) < 1$ , which can obviously be done in polynomial time.

$\text{VAL}_{\mathcal{L}_{\Pi}^0}$ : there is a polynomial embedding of  $\mathcal{L}_{\Pi}^0$  into  $\mathcal{L}_L^0$ , see [Baaz *et al.*, 1998a].

$\text{VAL}_{\mathcal{L}_L^0}, \text{SAT}_{\mathcal{L}_L^0}$ : an immediate consequence of Theorem 16(1). ■

**REMARK 36.** Co-NP-completeness of  $\text{VAL}_{\mathcal{L}_L^0}$  was shown by Mundici [1987]. The polynomial embedding of MIP-representable logics described in Section 4.2 also yields NP-easiness of Gödel logic and of the extension of Łukasiewicz logic that is characterized by piecewise linear functions with rational coefficients (discussed at the end of Section 4.3). In [Hájek, 1998] also the complexity of the sets  $N \setminus \{0\}\text{-SAT}_{\mathcal{L}}$  and  $N \setminus \{0\}\text{-VAL}_{\mathcal{L}}$  is considered.

While the decision problems of propositional infinite-valued Łukasiewicz, Gödel and product logic have the same complexity as two-valued logic, the situation changes drastically in the first-order case:



**THEOREM 37.**  $VAL_{\mathcal{L}_G}$  is  $\Sigma_1^-$ -complete and  $VAL_{\mathcal{L}}$  is  $\Pi_2$ -complete, whenever  $\mathcal{L} \in \{\mathcal{L}_L, \mathcal{L}_\Pi\}$ , if  $N = [0, 1]$ .

The proofs are rather technical and must be omitted (they can be found in [Hájek, 1998]).  $\Pi_2$ -completeness of  $\mathcal{L}_L$  was first shown in [Scarpellini, 1962; Ragaz, 1981]. An embedding of  $\mathcal{L}_L$  into  $\mathcal{L}_G$  [Hájek, 1998] can be used to show  $\Pi_2$ -completeness of the latter.  $\Sigma_1$ -easiness follows from the existence of a complete first-order axiomatization [Hájek, 1998]. Surprisingly,  $VAL_{\mathcal{L}_G}$  is *not* recursively enumerable anymore, for example, over the truth value set  $N_* = \{\frac{1}{n+1} \mid n \in \mathbb{N}\} \cup \{0\}$  [Baaz *et al.*, 1996]. So the order type of the truth value set can drastically change the complexity of infinite-valued first-order logic.

More fine-grained investigations were made into the complexity of satisfiability problems associated with signed CNF formulas (50). Let CNF-SAT be the set of satisfiable propositional signed formulas, let 2-CNF-SAT be the restriction of CNF-SAT, where signed clauses contain exactly two signed atoms, and let HORN-SAT be the satisfiable regular Horn formulas (defined in Section 7.2). If the truth value set  $N$ , together with a partial order, is fixed, this is denoted with  $CNF-SAT_N$ ,  $2-CNF-SAT_N$ , and  $HORN-SAT_N$ .

Similar to the classical case, CNF-SAT is NP-complete, but some of its sub-classes are polynomially solvable. NP-hardness of CNF-SAT is trivial, because classical SAT is the same as  $CNF-SAT_{\{0,1\}}$  up to notation; NP-easiness of the problem CNF-SAT for all finite  $N$  is straightforward, see above. Further results are summarized in Table 1 and are collected in [Beckert *et al.*, 2000b].

$2-CNF-SAT_N$  for any  $|N| \geq 3$  and, therefore, 2-CNF-SAT was shown to be NP-hard in [Manyà, 1996; Manyà, 2000] (in contrast to classical 2-CNF-SAT that can be solved in linear time) by embedding the 3-colorability problem of graphs. A direct embedding of classical CNF-SAT into 2-CNF-SAT is given in [Beckert *et al.*, 1999].

Even *regular* 2-CNF-SAT is NP-complete, which can be shown by embedding (general) 2-CNF-SAT into regular 2-CNF-SAT [Beckert *et al.*, 2000a]. Under certain restrictions, however, membership in regular 2-CNF-SAT can be checked in polynomial time: for totally ordered  $N$ , [Manyà, 1996; Manyà, 2000] gives a quadratic-time procedure by a version of signed DPL (58) for regular formulas. A generalization is proven in [Beckert *et al.*, 2000a] for the case when  $N$  is a lattice and all occurring signs are of the form  $\uparrow i$  or  $\downarrow i$ . A linear-time procedure for solving monosigned 2-CNF-SAT is described in [Manyà, 2000].

If  $N$  is totally ordered, the problem of deciding whether a regular Horn formula  $\Gamma$  is satisfiable can be solved in time linear in  $|\Gamma|$  in case  $|N|$  is fixed, and in  $|\Gamma| \log |\Gamma|$  time, otherwise [Hähnle, 1996b]. An algorithm for a particular subclass of regular Horn formulas appeared before in [Escalada-Imaz and Manyà, 1994].

Table 1. Known complexity results for signed SAT problems

	CNF	2-CNF	HORN
<i>classical</i>	NPC	linear [Even <i>et al.</i> , 1976]	linear [Dowling and Gallier, 1984]
<i>monosigned</i>	NPC	linear [Manyà, 2000]	—
<i>regular, N totally ord.</i>	NPC	polynomial [Manyà, 1996]	$ \Gamma  \log  \Gamma $ [Hähnle, 1996b]
<i>regular, N distr. lattice, signs of form <math>\uparrow i</math> and <math>\overline{\uparrow i}</math></i>	NPC	NPC [Beckert <i>et al.</i> , 2000a]	$ \Gamma   N ^2$ [Sofronie-Stokkermans, 1998]
<i>regular, N lattice, signs of form <math>\uparrow i</math> and <math>\overline{\uparrow i}</math></i>	NPC	NPC	polynomial [Beckert <i>et al.</i> , 1999]
<i>regular, N lattice, signs of form <math>\uparrow i</math> and <math>\downarrow i</math></i>	NPC	polynomial [Beckert <i>et al.</i> , 2000a]	—
<i>regular (arbitrary)</i>	NPC	NPC	—
<i>signed (arbitrary)</i>	NPC	NPC [Manyà, 1996]	—

If  $N$  is a finite lattice, regular HORN-SAT is decidable in time linear in the length of the formula and polynomial in the cardinality of  $N$  via a reduction to classical HORN-SAT [Beckert *et al.*, 1999]. For distributive lattices, the more precise bound  $|\Gamma| \cdot |N|^2$  was found independently in [Sofronie-Stokkermans, 1998], which contains also some results on decidable first-order fragments of regular CNF formulas. A closer inspection of the proofs in the cited papers yields immediately that all regular HORN-SAT $_N$  versions have linear complexity.

If the partial order of  $N$  is no lattice, then regular Horn formulas need to be based on signs of the form  $\uparrow S$ , where  $S \subseteq N$  (see end of Section 7.2). This more general notion of regular HORN-SAT is still decidable in time linear in the length of the formula, but exponential in the cardinality of  $N$  [Beckert *et al.*, 1999] provided that  $N$  possesses a maximal element.

If  $N$  is infinite, then regular HORN-SAT is decidable provided that  $N$  is a **locally finite lattice**, that is, every sub-lattice generated by a finite subset is finite [Beckert *et al.*, 1999].

A set of on-line (that is: incremental) algorithms for Horn formulas with numerical uncertainties has been proposed and studied in [Ausiello and Giaccio, 1997]. The resulting complexities are (at most cubic) polynomials, also in the infinite-valued case.

The NP-complete satisfiability problems for regular formulas and Regular-DPL exhibit the phase transition phenomena encountered in many decision

procedures for NP-complete problems [Mitchell *et al.*, 1992]: (i) there is a sharp increase (phase transition) of the percentage of unsatisfiable random CNF-SAT instances around a certain point when the ratio  $\frac{c}{v}$  between the number  $c$  of clauses and the number  $v$  of variables is varied; (ii) there is an easy-hard-easy pattern in the computational difficulty of solving problem instances as  $\frac{c}{v}$  is varied—the hard instances tend to be found near the crossover point [Manyà *et al.*, 1998; Béjar and Manyà, 1999b].

The complexity of deciding logical consequence depends on the availability and form of deduction theorems (such as Theorem 9) for a given logic.

On the CNF level, some authors used many-valued semantics to approximate classical propositional consequence. In some cases this led to polynomial decision procedures and results of this kind were used in knowledge representation [Patel-Schneider, 1990]. An overview of results in this area is in [Cadoli and Schaerf, 1996].

## 9.2 Representations

The size of representations and tableau/sequent rules for many-valued operators (see Section 6) is closely related to the deterministic complexity of decision problems, because the size of rules determines the size of sequent/tableau proofs, and the latter immediately yield upper bounds of the complexity of VAL and CNF-SAT by virtue of the Completeness Theorem 25.

Representations (Theorem 21) and expansions (Section 8.2) do also determine the size of various normal forms for many-valued logic formulas in an essential way.

In [Rousseau, 1967; Rousseau, 1970; Zach, 1993; Hähnle, 1994a; Baaz and Fermüller, 1995b] the following results on the maximal size of signed CNF-/DNF-representations of finite-valued connectives  $\theta$  are stated and proven:

$n =  N , r = \alpha(\theta)$	DNF	CNF
monosigned	$n^r$	$n^{r-1}$
general	$n^{r-1}$	$n^{r-1}$

All bounds are tight. The  $r$ -ary Łukasiewicz sum  $\oplus_L^r$  serves in the proof of all cases; it is defined as  $\oplus_L^r(p_1, \dots, p_r) \equiv p_1 \oplus_L \dots \oplus_L p_r$ .

The method described in Section 6.6 for translating formulas of any finite-valued logic into signed CNF limits the latter's size to be in  $\mathcal{O}(n^R|\varphi|)$ , where  $R = \max\{\alpha(\theta) \mid \theta \in \Theta, \theta \text{ occurs in } \varphi\}$  [Hähnle, 1994d]. For logics defined by distributive lattices, up to exponentially better results are possible by using the dual space representations discussed in Section 6.5. General complexity results for this case and many concrete examples are in [Sofronie-Stokkermans, to appear, 2000].

The branching factor of signed tableau rules for quantifiers (the number of disjuncts in Theorem 22) is at most  $2^n - 2$ . For the dual sequent rules the slightly better bound  $2^{n-1}$  can be obtained [Baaz and Fermüller, 1995b], which is sharp as well.

A slightly different question is to ask how many different signs are needed in general to build a sound and complete signed tableau or sequent calculus for a given  $n$ -valued logic. It is shown in [Baaz *et al.*, 1998b] for families of signs fulfilling condition (40) that this number is logarithmic in  $n$  (and the bound is tight).

One of the results relating to infinite-valued Łukasiewicz logic is that every  $\mathcal{L}_L^0$ -formula over just one variable can be polynomially translated into a regular signed atom with respect to the natural order on  $N = [0, 1]$  [Mundici and Olivetti, 1998]. The complexity of McNaughton functions in one variable is investigated in [Aguzzoli, 1998a; Aguzzoli, 1999]. In [Mundici, 1987] it is shown that a  $\mathcal{L}_L^0$ -formula  $\varphi$  is valid in infinite-valued  $\mathcal{L}_L^0$  iff it is valid in  $(2^{(2^{|\varphi|})^2} + 1)$ -valued  $\mathcal{L}_L^0$ . This bound was later improved to  $(2^{|\varphi|} + 1)$  values [Aguzzoli and Ciabattoni, 2000; Ciabattoni, 2000b]. An analogous result is established for logical consequence.

Space complexity of various kinds of MDDs (Section 8.2) is discussed, for example, in [Sasao, 1993b], where further pointers to the literature can be found. In fact, every known kind of MDD has exponential worst-case space complexity. Increased space complexity is frequently traded in for more efficient computation in practice.

The worst-case, best-case and relative space complexity of various kinds of polynomial representations of finite-valued functions is investigated in many papers, for example, [Sasao and Butler, 1997].

## 10 INTERACTION WITH OTHER NON-CLASSICAL LOGICS

Interaction between many-valued and other non-classical logics can take place in several ways. I start by mentioning general *logical frameworks* such as Gabbay's labeled deductive systems (LDS) [Gabbay, 1996], which can accommodate several non-classical logics, including MVL, at the same time. The development of a general tableau-based deductive framework for LDS was started in [D'Agostino and Gabbay, 1994].

**Fibring** of logics, suggested in [Pfalzgraf, 1991], and developed further by Gabbay [1999], is a general methodology to break up logical systems into simpler components and recombine them. This is exploited in [Beckert and Gabbay, 1998] who provide precise conditions on logics that guarantee modular combination of sound and complete tableau calculi.

The matrix-based semantics defined in Section 2.1 and used throughout this article is a very straightforward and natural generalization of classical semantics. Moreover, it is perfectly adequate for characterizing many-valued

logics. There exist, however, several other kinds of semantics that not only cover certain many-valued logics, but other logics (modal, relevant, etc.) as well. Advantages of taking an alternative approach are that it admits a direct comparative analysis of the logics in question and often gives additional insights.

One semantical framework is the **possible-translation semantics** or **non-deterministic semantics** of Carnielli [2000]. It works with sets of matrices, and formulas are satisfiable, if they are satisfiable in certain subsets of these matrices. There are logics that can be characterized by a finite subset of finite matrices, even though there is no single finite characteristic matrix. Possible-translation semantics can be generalized to offer an alternative to logical fibring [Carnielli and Coniglio, 1999].

The so-called **society semantics** [Carnielli and Lima-Marques, 1999] is an agent-oriented semantics and can be seen as a special case of possible-translation semantics.

The book [Epstein, 1996] develops **set-assignment semantics**, an addition to matrix semantics that allows elegant characterization of a great variety of non-classical logics, including modal logics, intuitionistic logic, paraconsistent logics, and many-valued logics.

In parallel to the semantical frameworks just mentioned, there exist at least two mainly proof theoretical frameworks that allow to compare a variety of non-classical logics, including many-valued ones. These are hypersequent calculi (see Section 6.1) [Pottinger, 1983; Avron, 1987; Avron, 1996] and display sequent calculi [Belnap, 1982]

On the other hand, many-valued logic can also be combined with other logics in a “bottom-up” style, which sometimes allows to prove stronger results than can be obtained within general frameworks. Many-valued modal logics of various sorts were considered in [Fitting, 1991b; Fitting, 1992; Iturrioz, 1993; Fitting, 1995; Baaz and Fermüller, 1996], many-valued temporal logic in [Thiele and Kalenka, 1993; Baaz and Zach, 1994; Baaz *et al.*, 1996], sometimes with surprising results such as *propositional* infinite-valued temporal logic being undecidable [Wagner, 1997]. Sorted, many-valued higher-order logic was used for modeling partial functions [Kerber and Kohlhase, 1994; Kerber and Kohlhase, 1997].

Consider signed formulas  $S:\varphi$  in some many-valued logic  $\mathcal{L}^0$  with truth values  $N$  and family of signs  $\mathcal{S} \subseteq 2^N$ . It is straightforward to define a two-valued possible world semantics, if one considers  $\mathcal{S}$  as the set of possible worlds and by forcing  $\varphi$  in  $i$ , whenever  $S:\varphi$  is satisfiable in  $\mathcal{L}^0$ . The question, of course, is: are there any interesting operators to relate possible worlds? In some cases, this is indeed so, for example, the common possible worlds semantics of Gödel, intuitionistic and certain temporal logics was used to prove undecidability results about the latter [Baaz *et al.*, 1996]. The paper [Orłowska and Iturrioz, 1999] provides a link between Łukasiewicz logic and modal logic.

The converse direction, from modal logic to MVL, is interesting as well, because many-valued logics permit efficient deduction procedures (see Section 6). Any modal logic with set of possible worlds  $N$  and accessibility relation  $R \subseteq N \times N$  can be trivially re-interpreted as a  $2^N$ -valued logic by encoding a possible world interpretation  $\mathbf{I}^* : \Sigma \times N \rightarrow \{0, 1\}$  as a many-valued interpretation  $\mathbf{I} : \Sigma \rightarrow 2^N$ . For modal logics with the finite model property this construction yields finite-valued proof systems characterizing those modal logics [Caferra and Zabel, 1990]. More interesting results are obtained from duality theory of distributive lattices: possible world frames of certain modal logics turn out to be the dual spaces of suitable distributive lattices, not necessarily finite. The technique explained in Section 6.5 due to Sofronie-Stokkermans [1999a; to appear, 2000] leads to proof theoretical characterization of such modal logics in terms of regular logic and, ultimately, first-order classical logic.

Finally, an interesting connection is the usage of many-valued signs to make deduction in intuitionistic and modal logic more efficient [Miglioli *et al.*, 1994; Miglioli *et al.*, 1995; Avellone *et al.*, 1999].

## 11 MVL RESOURCES

REMARK 38. In the present section I collected various resources to draw upon for learning about MVL. For the reader's convenience URLs are supplied, whenever appropriate. Information of this kind is likely to change, so I maintain a web page<sup>20</sup> containing this section in updated and expanded form.

**Books** Classic, but dated, monographs on MVL in general are [Rosser and Turquette, 1952; Rescher, 1969]; the same holds for the collections [Dunn and Epstein, 1977; Rine, 1984].

[Malinowski, 1993] is recommended as a more recent and compact introduction into MVL. Another concise introduction, strong on algebraic aspects, is [Panti, 1998]. The volumes [Bolc and Borowik, 1992; Bolc and Borowik, 2000] are comprehensive and fairly recent, but at least the first part is, unfortunately, seriously flawed by many inaccuracies, see [Hájek and Zach, 1994].

The book [Gottwald, 1989] is comprehensive, but only available in German; a much expanded English edition [Gottwald, 2000] is likely to become the standard monograph on MVL. The work [Hájek, 1998] is broader than its title may suggest—both books were of invaluable help in preparing this chapter. Substantial bibliographies are contained in [Iturrioz, 2000; Iturrioz *et al.*, 2000; Gottwald, 2000]

<sup>20</sup><http://www.cs.chalmer.se/~reiner/mvl-web>

There are, of course, books on more specialized areas within MVL: Among the many books on fuzzy logic I only mention [Novák, 1989; Zimmermann, 1991; Gottwald, 1993; Kruse *et al.*, 1994; Hájek, 1998; Turunen, 1999] and the collections [Marks, 1994; Klir and Yuan, 1996], which contain material on fuzzy logic in the narrow sense. The definitive book on *t*-norm theory is [Klement *et al.*, 2000]. Deductive aspects of MVL are the topic of [Hähnle, 1994a; Stachniak, 1996; Baaz *et al.*, to appear, 2000]. An overview of algebraic structures related to many-valued logics is [Iturrioz *et al.*, 2000]. Algebraic aspects of Łukasiewicz logic are treated in depth in [Cignoli *et al.*, 1999], while philosophical aspects of MVL are discussed in [Zinov'ev, 1963; Haack, 1974; Haack, 1996]. For many-valued switching theory look at [Muzio and Wesselkamper, 1986] and the collections [Sasao, 1993a; Sasao and Fujita, 1996].

**Journals** There is one journal publishing articles on all aspects of many-valued logic: *Multiple-Valued Logic: an International Journal*<sup>21</sup>, published by Gordon & Breach.

Journals with a strong emphasis on one or more aspects of MVL are *Soft Computing: A Fusion of Foundations, Methodologies and Applications*<sup>22</sup>, published by Springer-Verlag; *Mathware & Soft Computing*<sup>23</sup>, published by Universitat Politècnica de Catalunya; *Fuzzy Sets and Systems*<sup>24</sup>, published by Elsevier.

Mainstream logic journals with an editorial interest in many-valued logic include *Studia Logica*<sup>25</sup>, published by Kluwer; *Journal of Applied Non-Classical Logics*<sup>26</sup>, published by Hermès; *Journal of Logic and Computation*<sup>27</sup>, published by Oxford University Press; *Journal of Language, Logic and Computation*<sup>28</sup>, published by Kluwer.

**Organizations and Meetings** *IEEE Computer Society*<sup>29</sup> has a *Technical Committee on MVL*<sup>30</sup> [Kameyama, 1997], which also organizes the only annual conference devoted exclusively to MVL. The meeting is called *International Symposium on Multiple-Valued Logic*, the proceedings are published by the IEEE Computer Society Press. Papers in MVL are, of course, also presented in other *logic-related conferences*<sup>31</sup>. A research network called

<sup>21</sup><http://www.gbhap-us.com/journals/733/733-top.htm>

<sup>22</sup><http://link.springer.de/link/service/journals/00500/index.htm>

<sup>23</sup><http://www.upc.es/ea-smi/mathware/>

<sup>24</sup><http://www.elsevier.nl/inca/publications/store/5/0/5/5/4/5/>

<sup>25</sup><http://kapis.www.wkap.nl/journalhome.htm/0039-3215>

<sup>26</sup><http://www.editions-hermes.fr/periodiques/no.htm>

<sup>27</sup><http://www3.oup.co.uk/logcom/>

<sup>28</sup><http://kapis.www.wkap.nl/journalhome.htm/0925-8531>

<sup>29</sup><http://www.computer.org>

<sup>30</sup><http://wwwj3.comp.eng.himeji-tech.ac.jp/mvl/>

<sup>31</sup><http://cm.bell-labs.com/cm/cs/who/libkin/lics/logic-confs.html>

*Many-Valued Logics for Computer Science Applications*<sup>32</sup> sponsored as a COST Action by the European Community existed between 1995 and 1999 [Tassart *et al.*, 1995]. Its final report [Iturrioz, 2000] contains a bibliography of recent work on MVL with over 600 entries.

#### ACKNOWLEDGEMENTS

For preprints of unpublished books and articles I am indebted to Lluís Godo, Siegfried Gottwald, Daniele Mundici, and Alasdair Urquhart. For constructive criticism and many helpful remarks I am grateful to Walter A. Carnielli (the second reader of this chapter), Agata Ciabattoni, Roberto Cignoli, Siegfried Gottwald, João Marcos, and Daniele Mundici.

The meetings of EC COST Action 15 “Many-Valued Logics for Computer Science Applications” between 1995–1999 were a great source of inspiration. I am deeply grateful to all who contributed to these meetings—without you this chapter simply could not have been written.

*Chalmers University of Technology, Gothenburg, Sweden.*

#### BIBLIOGRAPHY

- [Aguilera Venegas *et al.*, 1995] Aguilera Venegas, G., Perez de Guzmán, I., and Ojeda Aciego, M. Increasing the efficiency of automated theorem proving. *Journal of Applied Non-Classical Logics*, **5**(1), 9–29.
- [Aguilera Venegas *et al.*, 1997] Aguilera Venegas, G., Perez de Guzman, I., and Ojeda Aciego, M. A reduction-based theorem prover for 3-valued logic. *Mathware & Soft Computing*, **IV**(2), 99–127. Special Issue on Deduction in Many-Valued Logic.
- [Aguilera Venegas *et al.*, 1999] Aguilera Venegas, G., Perez de Guzman, I., Ojeda Aciego, M., and Valverde, A. Reducing signed propositional formulas. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **2**(4), 157–166.
- [Agustí-Cullell *et al.*, 1991] Agustí-Cullell, J., Esteva, F., García, P., Godó, L., and Sierra, C. Combining multiple-valued logics in modular expert systems. *Pages 17–25 of: D’Ambrosio, Bruce D., Smets, Philippe, and Bonissone, P. P. (eds.), Proc. 7th Conference on Uncertainty in Artificial Intelligence*. San Mateo/CA: Morgan Kaufmann.
- [Agustí-Cullell *et al.*, 1994] Agustí-Cullell, J., Esteva, F., García, P., Godó, L., López de Mantaras, R., and Sierra, C. Local multi-valued logics in modular expert systems. *Journal of Experimental and Theoretical Artificial Intelligence*, **6**, 303–321.
- [Aguzzoli, 1998a] Aguzzoli, S. The complexity of McNaughton functions of one variable. *Advances in Applied Mathematics*, **21**(1), 58–77.
- [Aguzzoli, 1998b] Aguzzoli, S. A note on the representation of McNaughton lines by basic literals. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **2**(3), 111–115.
- [Aguzzoli, 1999] Aguzzoli, S. Geometric and Proof Theoretic Issues in Lukasiewicz Propositional Logics. Ph.D. thesis, University of Siena, Italy.
- [Aguzzoli and Ciabattoni, 2000] Aguzzoli, S., and Ciabattoni, A. Finiteness in infinite-valued logic. *Journal of Logic, Language and Information*, **9**(1), 5–29.

---

<sup>32</sup><http://www.upmf-grenoble.fr/mvl/>



- [Akers, 1978] Akers, S. B. Binary decision diagrams. *IEEE Transactions on Computers*, **27**(6), 509–516.
- [Anantharaman and Bonacina, 1990] Anantharaman, S., and Bonacina, M. P. An application of the theorem prover SBR3 to many-valued logic. *Pages 156–161 of: Kaplan, S., and Okada, M. (eds.), Proc. 2<sup>nd</sup> International Workshop on Conditional and Typed Term Rewriting Systems, Montreal, Canada*. LNCS, vol. 516. Springer-Verlag.
- [Angel J. Gil *et al.*, 1997] Angel J. Gil, Torrens, A., and Verdú, V. On Gentzen systems associated with the finite linear MV-algebras. *Journal of Logic and Computation*, **7**(4), 473–500.
- [Arieli and Avron, 1998] Arieli, O., and Avron, A. The value of the four values. *Artificial Intelligence*, **102**(1), 97–141.
- [Ausiello and Giaccio, 1997] Ausiello, G., and Giaccio, R. On-line algorithms for satisfiability problems with uncertainty. *Theoretical Computer Science*, **171**(1–2), 3–24.
- [Avellone *et al.*, 1999] Avellone, A., Ferrari, M., and Miglioli, P. Duplication-free tableau calculi and related cut-free sequent calculi for the interpolable propositional intermediate logics. *Logic Journal of the IGPL*, **7**(4), 447–480.
- [Avron, 1987] Avron, A. A constructive analysis of RM. *Journal of Symbolic Logic*, **52**(4), 939–951.
- [Avron, 1991a] Avron, A. Hypersequents, logical consequence and intermediate logics for concurrency. *Annals of Mathematics and Artificial Intelligence*, **4**(3–4), 225–248.
- [Avron, 1991b] Avron, A. Natural 3-valued logics—characterization and proof theory. *Journal of Symbolic Logic*, **56**(1), 276–294.
- [Avron, 1996] Avron, A. The method of hypersequents in the proof theory of propositional non-classical logics. *Pages 1–32 of: Hodges, W., Hyland, M., Steinhorn, C., and Truss, J. (eds.), Logic: from foundations to applications. Proc. Logic Colloquium, Keele, UK, 1993*. New York: Oxford University Press.
- [Baaz, 1996] Baaz, M. Infinite-valued Gödel logics with 0-1-projections and relativizations. *Pages 23–33 of: Hájek, P. (ed.), Proc. GÖDEL'96: Logical Foundations of Mathematics, Computer Science and Physics*. Lecture Notes in Logic, vol. 6. Springer-Verlag.
- [Baaz and Fermüller, 1992] Baaz, M., and Fermüller, C. G. Resolution for many-valued logics. *Pages 107–118 of: Voronkov, A. (ed.), Proc. Logic Programming and Automated Reasoning LPAR, St. Petersburg, Russia*. LNCS, vol. 624. Springer-Verlag.
- [Baaz and Fermüller, 1995a] Baaz, M., and Fermüller, C. G. Nonelementary speedups between different versions of tableaux. *Pages 217–230 of: Baumgartner, P., Hähnle, R., and Posegga, J. (eds.), Proc. 4th Workshop on Deduction with Tableaux and Related Methods, St. Goar, Germany*. LNCS, vol. 918. Springer-Verlag.
- [Baaz and Fermüller, 1995b] Baaz, M., and Fermüller, C. G. Resolution-based theorem proving for many-valued logics. *Journal of Symbolic Computation*, **19**(4), 353–391.
- [Baaz and Fermüller, 1996] Baaz, M., and Fermüller, C. G. Combining many-valued and intuitionistic tableaux. *Pages 65–79 of: Miglioli, P., Moscato, U., Mundici, D., and Ornaghi, M. (eds.), Theorem Proving with Tableaux and Related Methods, 5th International Workshop, TABLEAUX'96, Terrasini, Palermo, Italy*. LNCS, vol. 1071. Springer-Verlag.
- [Baaz and Zach, 1994] Baaz, M., and Zach, R. Approximating propositional calculi by finite-valued logics. *Pages 257–263 of: Proc. 24th International Symposium on Multiple-Valued Logics (ISMVL), Boston/MA, USA*. IEEE CS Press, Los Alamitos.
- [Baaz *et al.*, 1994] Baaz, M., Fermüller, C. G., and Zach, R. Elimination of cuts in first-order many-valued logics. *Journal on Information Processing and Cybernetics*, **29**(6), 333–355.
- [Baaz *et al.*, 1996] Baaz, M., Leitsch, A., and Zach, R. Incompleteness of a first-order Gödel logic and some temporal logics of programs. *Pages 1–15 of: Kleine Büning, H. (ed.), Selected Papers from Computer Science Logic, CSL'95, Paderborn, Germany*. LNCS, vol. 1092. Springer-Verlag.
- [Baaz *et al.*, 1998a] Baaz, M., Hájek, P., Svejda, D., and Krajíček, J. Embedding logics into product logic. *Studia Logica*, **61**(1), 35–47.
- [Baaz *et al.*, 1998b] Baaz, M., Fermüller, C. G., Salzer, G., and Zach, R. Labeled calculi and finite-valued logics. *Studia Logica*, **61**(1), 7–33. Many-valued logics.

- [Baaz *et al.*, 1998c] Baaz, M., Ciabattoni, A., Fermueller, C. G., and Veith, H. Proof theory of fuzzy logics: Urquhart's C and related logics. *Pages 203–212 of: Brim, L., Gruska, J., and Zlatuska, J. (eds.), Proc. 23rd International Symposium Mathematical Foundations of Computer Science, Brno, Czech Republic.* LNCS, vol. 1450.
- [Baaz *et al.*, to appear, 2000] Baaz, M., Fermüller, C. G., and Salzer, G. Automated deduction for many-valued logics. *In: Robinson, A., and Voronkov, A. (eds.), Handbook of Automated Reasoning.* Elsevier Science Publishers.
- [Bachmair and Ganzinger, 1998] Bachmair, L., and Ganzinger, H. Ordered chaining calculi for first-order theories of transitive relations. *Journal of the ACM*, **45**(6), 1007–1049.
- [Bauer *et al.*, 1995] Bauer, M., Alexis, R., Atwood, G., Baltar, B., Fazio, A., Frary, K., Hensel, M., Ishac, M., Javanifard, J., Landgraf, M., Leak, D., Loe, K., Mills, D., Ruby, P., Rozman, R., Sweha, S., Talreja, S., and Wojciechowski, K. A multilevel-cell 32 Mb flash memory. *Pages 132–133, 351 of: 41st Solid-State Circuits Conference, ISSCC. Digest of Technical Papers.* IEEE CS Press.
- [Beckert and Gabbay, 1998] Beckert, B., and Gabbay, D. Fibring semantic tableaux. *Pages 77–92 of: de Swart, H. (ed.), Proc. International Conference on Theorem Proving with Analytic Tableaux and Related Methods, Oisterwijk, The Netherlands.* LNCS, no. 1397. Springer-Verlag.
- [Beckert *et al.*, 1996] Beckert, B., Hähnle, R., Oel, P., and Sulzmann, M. The tableau-based theorem prover  $\exists T^AP$ , version 4.0. *Pages 303–307 of: McRobbie, M., and Slaney, J. (eds.), Proc. 13th Conference on Automated Deduction, New Brunswick/NJ, USA.* LNCS, vol. 1104. Springer-Verlag.
- [Beckert *et al.*, 1998] Beckert, B., Hähnle, R., and Escalada-Imaz, G. Simplification of many-valued logic formulas using anti-links. *Journal of Logic and Computation*, **8**(4), 569–588.
- [Beckert *et al.*, 1999] Beckert, B., Hähnle, R., and Manyà, F. Transformations between signed and classical clause logic. *Pages 248–255 of: Proc. 29th International Symposium on Multiple-Valued Logics, Freiburg, Germany.* IEEE CS Press, Los Alamitos.
- [Beckert *et al.*, 2000a] Beckert, B., Hähnle, R., and Manyà, F. On the regular 2-SAT problem. *Pages 331–336 of: Proc. 30th International Symposium on Multiple-Valued Logics, Portland/OR, USA.* IEEE CS Press, Los Alamitos.
- [Beckert *et al.*, 2000b] Beckert, B., Hähnle, R., and Manyà, F. The SAT problem of signed CNF formulas. *Pages 61–82 of: Basin, D., D'Agostino, M., Gabbay, D., Matthews, S., and Viganò, L. (eds.), Labelled Deduction.* Applied Logic Series, vol. 17. Kluwer, Dordrecht.
- [Béjar and Manyà, 1999a] Béjar, R., and Manyà, F. A comparison of systematic and local search algorithms for regular CNF formulas. *Pages 22–31 of: Hunter, A., and Parsons, S. (eds.), Proc. Fifth European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty, ECSQARU'99, London, UK.* LNCS, vol. 1638. Springer-Verlag.
- [Béjar and Manyà, 1999b] Béjar, R., and Manyà, F. Phase transitions in the regular random 3-SAT problem. *Pages 292–300 of: Raś, Z. W., and Skowron, A. (eds.), Proc. International Symposium on Methodologies for Intelligent Systems, ISMIS'99, Warsaw, Poland.* LNCS, no. 1609. Springer-Verlag.
- [Béjar and Manyà, 1999c] Béjar, R., and Manyà, F. Solving combinatorial problems with regular local search algorithms. *Pages 33–43 of: Ganzinger, H., and McAllester, D. (eds.), Proc. 6th Int. Conference on Logic for Programming and Automated Reasoning, LPAR, Tbilisi, Georgia.* LNCS, vol. 1705. Springer-Verlag.
- [Bell *et al.*, 1994] Bell, C., Nerode, A., Ng, R., and Subrahmanian, V. Mixed integer programming methods for computing nonmonotonic deductive databases. *Journal of the ACM*, **41**(6), 1178–1215.
- [Belnap, 1982] Belnap, Jr., N. D. Display logic. *Journal of Philosophical Logic*, **11**(4), 375–417.
- [Belnap Jr., 1977] Belnap Jr., N. D. A useful four-valued logic. *Pages 8–37 of: Dunn, J. M., and Epstein, G. (eds.), Modern uses of multiple-valued logic.* Reidel, Dordrecht.
- [Bibel, 1987] Bibel, W. *Automated Theorem Proving.* Second revised edn. Vieweg, Braunschweig.

- [Blair and Subrahmanian, 1989] Blair, H. A., and Subrahmanian, V. S. Paraconsistent logic programming. *Theoretical Computer Science*, **68**(2), 135–154.
- [Blok and Pigozzi, 1989] Blok, W. J., and Pigozzi, D. Algebraizable logics. *Memoirs of the American Mathematical Society*, **77**(396), vi+78.
- [Blok and Pigozzi, to appear, 2000] Blok, W. J., and Pigozzi, D. Abstract algebraic logic and the deduction theorem. *Bulletin of Symbolic Logic*.
- [Bolc and Borowik, 1992] Bolc, L., and Borowik, P. *Many-Valued Logics*. Vol. 1: Theoretical Foundations. Springer-Verlag.
- [Bolc and Borowik, 2000] Bolc, L., and Borowik, P. *Many-Valued Logics*. Vol. 2: Automated Reasoning and Practical Applications. Springer-Verlag.
- [Bonissone, 1997] Bonissone, P. P. Soft computing: the convergence of emerging reasoning technologies. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **1**(1), 6–18.
- [Boole, 1854] Boole, G. *An Investigation of the Laws of Thought*. Walton, London. Reprinted by Dover Books, New York, 1954.
- [Brink, 1993] Brink, C. Power structures. *Algebra Universalis*, **30**, 177–216.
- [Brown, 1990] Brown, F. M. *Boolean Reasoning*. Kluwer, Norwell/MA, USA.
- [Bryant, 1984] Bryant, R. E. A switch-level model and simulator for MOS digital systems. *IEEE Transactions on Computers*, **C-33**, 160–169.
- [Bryant, 1986] Bryant, R. E. Graph-based algorithms for Boolean function manipulation. *IEEE Transactions on Computers*, **C-35**, 677–691.
- [Bryant, 1992] Bryant, R. E. Symbolic boolean manipulation with ordered binary decision diagrams. *ACM Computing Surveys*, **24**(3), 293–318.
- [Bryant and Seger, 1991] Bryant, R. E., and Seger, C.-J. H. Formal verification of digital circuits using symbolic ternary system models. *Pages 33–43 of: Clarke, E. M., and Kurshan, R. P. (eds.), Proc. 2nd International Conference on Computer-Aided Verification (CAV'90)*. LNCS, vol. 531. Springer-Verlag.
- [Burch et al., 1990] Burch, J., Clarke, E., McMillan, K., and Dill, D. Sequential circuit verification using symbolic model checking. *Pages 46–51 of: Proc. 27th ACM/IEEE Design Automation Conference (DAC)*. ACM Press.
- [Cadoli and Schaerf, 1996] Cadoli, M., and Schaerf, M. On the complexity of entailment in propositional multivalued logics. *Annals of Mathematics and Artificial Intelligence*, **18**(1), 29–50.
- [Caferra and Zabel, 1990] Caferra, R., and Zabel, N. An application of many-valued logic to decide propositional  $\mathbf{S}_5$  formulae: a strategy designed for a parameterized tableaux-based theorem prover. *Pages 23–32 of: Jorrand, Ph., and Sgurev, V. (eds.), Proc. Artificial Intelligence IV: Methodology, Systems, Applications (AIMSA)*. Elsevier.
- [Cargile, 1969] Cargile, J. The Sorites paradox. *British Journal for the Philosophy of Science*, **20**, 193–202.
- [Carnielli, 1987] Carnielli, W. A. Systematization of finite many-valued logics through the method of tableaux. *Journal of Symbolic Logic*, **52**(2), 473–493.
- [Carnielli, 1991] Carnielli, W. A. On sequents and tableaux for many-valued logics. *Journal of Non-Classical Logic*, **8**(1), 59–76.
- [Carnielli, 2000] Carnielli, W. A. Possible-translations semantics for paraconsistent logics. *Pages 149–163 of: Batens, D., Mortensen, C., Priest, G., and Van Bendegem, J. P. (eds.), Frontiers of Paraconsistent Logic*. Studies in Logic and Computation, vol. 8. Research Studies Press, Baldock, UK.
- [Carnielli and Coniglio, 1999] Carnielli, W. A., and Coniglio, M. E. A categorial approach to the combination of logics. *Manuscrito—Revista Internacional de Filosofia*, **XXII**(2), 69–94.
- [Carnielli and Lima-Marques, 1999] Carnielli, W. A., and Lima-Marques, M. Society semantics and multiple-valued logics. *Pages 33–52 of: Carnielli, W., and D'Ottaviano, I. M. L. (eds.), Advances in Contemporary Logic and Computer Science: Proc. XI Brazilian Logic Conference on Mathematical Logic 1996, Salvador, Brazil*. Contemporary Mathematics, vol. 235. American Mathematical Society, Providence.
- [Carnielli et al., to appear, 2000] Carnielli, W. A., Marcos, J., and de Amo, S. Formal inconsistency and evolutionary databases. *Logic and Logical Philosophy*.

- [Castell and Fargier, 1998] Castell, T., and Fargier, H. Between SAT and CSP: Propositional satisfaction problems and clausal CSPs. *Pages 214–218 of: Prade, H. (ed.), Proc. 13th European Conference on Artificial Intelligence, Brighton*. John Wiley & Sons.
- [Chang, 1958] Chang, C. C. Algebraic analysis of many-valued logics. *Transactions of the American Mathematical Society*, **88**, 467–490.
- [Chang, 1959] Chang, C. C. A new proof of the completeness of the Lukasiewicz axioms. *Transactions of the American Mathematical Society*, **93**, 74–80.
- [Ciabattoni, 2000a] Ciabattoni, A. On Urquhart's C logic. *Pages 113–118 of: Proc. 30th International Symposium on Multiple-Valued Logics, Portland/OR, USA*. IEEE CS Press, Los Alamitos.
- [Ciabattoni, 2000b] Ciabattoni, A. Proof Theoretic Techniques in Many-Valued Logics. Ph.D. thesis, University of Milan, Italy.
- [Ciabattoni *et al.*, 1999] Ciabattoni, A., Gabbay, D. M., and Olivetti, N. Cut-free proof systems for logics of weak excluded middle. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **2**(4), 147–156.
- [Cignoli and Mundici, 1997a] Cignoli, R., and Mundici, D. An elementary proof of Chang's completeness theorem for the infinite-valued calculus of Lukasiewicz. *Studia Logica*, **58**(1), 79–97. Special Issue on Logics with Incomplete Information.
- [Cignoli and Mundici, 1997b] Cignoli, R., and Mundici, D. An invitation to Chang's MV algebras. *In: Droste, M., and Göbel, R. (eds.), Advances in algebra and model theory: selected surveys presented at conferences in Essen, 1994 and Dresden, 1995*. Algebra, logic and applications, no. 9. Gordon & Breach, Amsterdam.
- [Cignoli *et al.*, 2000] Cignoli, R., Esteve, F., Godo, L., and Torrens, A. Basic fuzzy logic is the logic of continuous t-norms and their residua. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **4**(2), 106–112.
- [Cignoli *et al.*, 1999] Cignoli, R. L. O., D'Ottaviano, I. M. L., and Mundici, D. *Algebraic Foundations of Many-Valued Reasoning*. Trends in Logic, vol. 7. Kluwer, Dordrecht.
- [Cohn, 1981] Cohn, P. M. *Universal Algebra*. Second edn. Reidel, Dordrecht.
- [D'Agostino, 1999] D'Agostino, M. Tableaux methods for classical propositional logic. *Pages 45–123 of: D'Agostino, M., Gabbay, D., Hähnle, R., and Posegga, J. (eds.), Handbook of Tableau Methods*. Kluwer, Dordrecht.
- [D'Agostino and Gabbay, 1994] D'Agostino, M., and Gabbay, D. M. A generalization of analytic deduction via labelled deduction systems. part I: Basic substructural logics. *Journal of Automated Reasoning*, **13**(2), 243–281.
- [Davey and Priestley, 1990] Davey, B. A., and Priestley, H. A. *Introduction to Lattices and Order*. Cambridge Mathematical Textbooks. Cambridge University Press, Cambridge.
- [Davis *et al.*, 1962] Davis, M., Logemann, G., and Loveland, D. A machine program for theorem-proving. *Communications of the ACM*, **5**, 394–397.
- [de Baets *et al.*, 1999] de Baets, B., Esteve, F., Fodor, J., and Godo, L. Systems of ordinal fuzzy logic with application to preference modelling. *Pages 47–50 of: Proc. Eusflat-Estjlf Joint Conference, Palma de Mallorca, Spain*. Univ. de las Islas Baleares.
- [Di Nola and Lettieri, 1994] Di Nola, A., and Lettieri, A. Perfect MV-algebras are categorically equivalent to Abelian  $\ell$ -groups. *Studia Logica*, **53**(3), 417–432.
- [Doherty, 1990] Doherty, P. Preliminary report: NM3 — a three-valued non-monotonic formalism. *Pages 498–505 of: Raś, Z., Zemankova, M., and Emrich, M. (eds.), Proc. 5th Int. Symposium on Methodologies for Intelligent Systems, Knoxville/TN, USA*. North-Holland.
- [D'Ottaviano and da Costa, 1970] D'Ottaviano, I. M. L., and da Costa, N. C. A. Sur un problème de Jaśkowski. *Comptes Rendus Hebdomadaires des Séances de l'Académie des Sciences, Série A—Sciences Mathématiques*, **270**(21), 1349–1353.
- [Dowling and Gallier, 1984] Dowling, W., and Gallier, J. Linear-time algorithms for testing the satisfiability of propositional Horn formulæ. *Journal of Logic Programming*, **3**, 267–284.
- [Driankow *et al.*, 1993] Driankow, D., Mellendoorp, H., and Reinfrank, M. *An Introduction to Fuzzy Control*. Springer-Verlag.

- [Dueck and Butler, 1994] Dueck, G. W., and Butler, J. T. Multiple-valued logic operations with universal literals. *Pages 73–79 of: Proc. 24th International Symposium on Multiple-Valued Logic, Boston/MA*. IEEE CS Press, Los Alamitos.
- [Dummett, 1959] Dummett, M. A propositional calculus with denumerable matrix. *Journal of Symbolic Logic*, **24**, 97–106.
- [Dunn and Epstein, 1977] Dunn, J. M., and Epstein, G. (eds.). *Modern Uses of Multiple-Valued Logic*. Reidel, Dordrecht. Invited Papers of 5th ISMVL Symposium 1975 with Bibliography by R. G. Wolf.
- [Dyckhoff, 1999] Dyckhoff, R. A deterministic terminating sequent calculus for Gödel-Dummett logic. *Logic Journal of the IGPL*, **7**(3), 319–326.
- [Epstein, 1996] Epstein, R. L. *The Semantic Foundations of Logic: Propositional Logic*. Second edn. Vol. 1. Oxford University Press.
- [Escalada-Imaz and Manyà, 1994] Escalada-Imaz, G., and Manyà, F. The satisfiability problem for multiple-valued Horn formulae. *Pages 250–256 of: Proc. International Symposium on Multiple-Valued Logics, ISMVL'94, Boston/MA, USA*. IEEE CS Press, Los Alamitos.
- [Escalada-Imaz and Manyà, 1995] Escalada-Imaz, G., and Manyà, F. Efficient interpretation of propositional multi-valued logic programs. *Pages 428–439 of: Bouchon-Meunier, B., Yager, R. R., and Zadeh, L. A. (eds.), Advances in Intelligent Computing, IPMU '94, 5th International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems, Paris, France*. LNCS, vol. 945. Springer-Verlag.
- [Esteva and Godo, 1999] Esteva, F., and Godo, L. QBL: Towards a logic for left-continuous t-norms. *Pages 35–37 of: Proc. Eusflat-Estylf Joint Conference, Palma de Mallorca, Spain*. Univ. de las Islas Baleares.
- [Esteva et al., 2000] Esteva, F., Godo, L., Hájek, P., and Navara, M. Residuated fuzzy logics with an involutive negation. *Archive for Mathematical Logic*, **39**(2), 103–124.
- [Eveking, 1991] Eveking, H. *Verifikation digitaler Systeme: eine Einführung in den Entwurf korrekter digitaler Systeme*. LMI. Teubner, Stuttgart.
- [Even et al., 1976] Even, S., Itai, A., and Shamir, A. On the complexity of timetable and multicommodity flow problems. *SIAM Journal of Computing*, **5**(4), 691–703.
- [Fitting, 1985] Fitting, M. C. A Kripke-Kleene semantics for logic programming. *Journal of Logic Programming*, **4**, 295–312.
- [Fitting, 1991a] Fitting, M. C. Bilattices and the semantics of logic programming. *Journal of Logic Programming*, **11**(2), 91–116.
- [Fitting, 1991b] Fitting, M. C. Many-valued modal logics. *Fundamenta Informaticae*, **XV**, 235–254.
- [Fitting, 1992] Fitting, M. C. Many-valued modal logics II. *Fundamenta Informaticae*, **XVII**, 55–74.
- [Fitting, 1995] Fitting, M. C. Tableaus for many-valued modal logic. *Studia Logica*, **55**(1), 63–68.
- [Fitting, 1996] Fitting, M. C. *First-Order Logic and Automated Theorem Proving*. Second edn. Springer-Verlag, New York.
- [Frank, 1979] Frank, M. J. On the simultaneous associativity of  $F(x, y)$  and  $x + y - F(x, y)$ . *Aequationes Mathematicae*, **19**, 194–226.
- [Gabbay, 1996] Gabbay, D. M. *Labelled Deductive Systems*. Vol. 1—Foundations. Oxford University Press.
- [Gabbay, 1999] Gabbay, D. M. *Fibring Logics*. Oxford Logic Guides, vol. 38. Oxford University Press.
- [Ganzinger and Sofronie-Stokkermans, 2000] Ganzinger, H., and Sofronie-Stokkermans, V. Chaining techniques for automated theorem proving in many-valued logics. *Pages 337–344 of: Proc. 30th International Symposium on Multiple-Valued Logics, Portland/OR, USA*. IEEE CS Press, Los Alamitos.
- [Geiß, 1997] Geiß, K. Vereinfachung großer Formeln in mehrwertiger Logik mit Anti-Links (in German). Master's thesis, Fakultät für Informatik, Universität Karlsruhe.
- [Gentzen, 1935] Gentzen, G. Untersuchungen über das Logische Schliessen. *Mathematische Zeitschrift*, **39**, 176–210, 405–431. English translation [Szabo, 1969].

- [Gerberding, 1996] Gerberding, S. DT—an automated theorem prover for multiple-valued first-order predicate logics. *Pages 284–289 of: Proc. 26th International Symposium on Multiple-Valued Logics, Santiago de Compostela, Spain*. IEEE CS Press, Los Alamitos.
- [Ginsberg, 1988] Ginsberg, M. L. Multi-valued logics. *Computational Intelligence*, **4**(3).
- [Girard, 1987] Girard, J.-Y. Linear logic. *Theoretical Computer Science*, **50**, 1–102.
- [Gödel, 1932] Gödel, K. Zum intuitionistischen Aussagenkalkül. *Anzeiger Akademie der Wissenschaften Wien, mathematisch-naturwiss. Klasse*, **32**, 65–66. Reprinted and translated in [Gödel, 1986].
- [Gödel, 1986] Gödel, K. *Collected Works : Publications 1929–1936*. Vol. 1. Oxford University Press. Edited by Solomon Feferman, John Dawson, and Stephen Kleene.
- [Goldblatt, 1989] Goldblatt, R. Varieties of complex algebras. *Annals of Pure and Applied Logic*, **44**(3), 173–242.
- [Gottwald, 1989] Gottwald, S. *Mehrwertige Logik. Eine Einführung in Theorie und Anwendungen (in German)*. Akademie-Verlag Berlin.
- [Gottwald, 1993] Gottwald, S. *Fuzzy Sets and Fuzzy Logic*. Vieweg, Braunschweig.
- [Gottwald, 2000] Gottwald, S. Axiomatizations of t-norm based logics—a survey. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **4**(2), 63–67.
- [Gottwald, 2000] Gottwald, S. *A Treatise on Many-Valued Logics*. Research Studies Press, 2000.
- [Grätzer, 1979] Grätzer, G. *Universal Algebra*. Second edn. Springer, New York.
- [Gray, 1996] Gray, J. Evolution of data management. *IEEE Computer*, **29**(Oct.), 38–46. Special Issue: 50 Years of Computing.
- [Gulak, 1998] Gulak, P. G. A review of multiple-valued memory technology. *Pages 222–231 of: Proc. International Symposium on Multiple-Valued Logics, ISMVL'98, Fukuoka, Japan*. IEEE CS Press, Los Alamitos.
- [Haack, 1974] Haack, S. *Deviant Logic—Some Philosophical Issues*. Cambridge University Press.
- [Haack, 1996] Haack, S. *Deviant Logic, Fuzzy Logic: Beyond the Formalism*. University of Chicago Press. Revised edition of [Haack, 1974].
- [Hachtel and Somenzi, 1996] Hachtel, G. D., and Somenzi, F. *Logic Synthesis and Verification Algorithms*. Kluwer Academic Publishers, Boston.
- [Haenni and Lehmann, 1998a] Haenni, R., and Lehmann, N. Assumption-based reasoning with finite set constraints. *Pages 1289–1295 of: Proc. Int. Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems IPMU, Paris, France*.
- [Haenni and Lehmann, 1998b] Haenni, R., and Lehmann, N. Reasoning with finite set constraints. *In: Proceedings Workshop 17 Many-valued logic for AI Applications at 13th European Conference in Artificial Intelligence*.
- [Hähnle, 1991a] Hähnle, R. Towards an efficient tableau proof procedure for multiple-valued logics. *Pages 248–260 of: Börger, E., Kleine Büning, H., Richter, M. M., and Schönfeld, W. (eds.), Selected Papers from Computer Science Logic, CSL'90, Heidelberg, Germany*. LNCS, vol. 533. Springer-Verlag.
- [Hähnle, 1991b] Hähnle, R. Uniform notation of tableaux rules for multiple-valued logics. *Pages 238–245 of: Proc. International Symposium on Multiple-Valued Logic, Victoria*. IEEE Press, Los Alamitos.
- [Hähnle, 1993] Hähnle, R. Short normal forms for arbitrary finitely-valued logics. *Pages 49–58 of: Komorowski, J., and Raś, Z. (eds.), Proc. 7th International Symposium on Methodologies for Intelligent Systems (ISMIS), Trondheim, Norway*. LNCS, vol. 689. Springer-Verlag.
- [Hähnle, 1994a] Hähnle, R. *Automated Deduction in Multiple-Valued Logics*. International Series of Monographs on Computer Science, vol. 10. Oxford University Press.
- [Hähnle, 1994b] Hähnle, R. Efficient deduction in many-valued logics. *Pages 240–249 of: Proc. International Symposium on Multiple-Valued Logics, ISMVL'94, Boston/MA, USA*. IEEE CS Press, Los Alamitos.
- [Hähnle, 1994c] Hähnle, R. Many-valued logic and mixed integer programming. *Annals of Mathematics and Artificial Intelligence*, **12**(3,4), 231–264.

- [Hähnle, 1994d] Hähnle, R. Short conjunctive normal forms in finitely-valued logics. *Journal of Logic and Computation*, **4**(6), 905–927.
- [Hähnle, 1996a] Hähnle, R. Commodious axiomatization of quantifiers in multiple-valued logic. *Pages 118–123 of: Proc. 26th International Symposium on Multiple-Valued Logics, Santiago de Compostela, Spain*. IEEE CS Press, Los Alamitos.
- [Hähnle, 1996b] Hähnle, R. Exploiting data dependencies in many-valued logics. *Journal of Applied Non-Classical Logics*, **6**(1), 49–69.
- [Hähnle, 1997] Hähnle, R. Proof theory of many-valued logic—linear optimization—logic design: Connections and interactions. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **1**(3), 107–119.
- [Hähnle, 1998] Hähnle, R. Commodious axiomatization of quantifiers in multiple-valued logic. *Studia Logica*, **61**(1), 101–121. Special Issue on Many-Valued Logics, their Proof Theory and Algebras.
- [Hähnle, 1999] Hähnle, R. Tableaux for many-valued logics. *Pages 529–580 of: D’Agostino, M., Gabbay, D., Hähnle, R., and Posegga, J. (eds.), Handbook of Tableau Methods*. Kluwer, Dordrecht.
- [Hähnle and Escalada-Imaz, 1997] Hähnle, R., and Escalada-Imaz, G. Deduction in many-valued logics: a survey. *Mathware & Soft Computing*, **IV**(2), 69–97.
- [Hähnle and Kernig, 1993] Hähnle, R., and Kernig, W. Verification of switch level designs with many-valued logic. *Pages 158–169 of: Voronkov, A. (ed.), Proc. LPAR’93, St. Petersburg, Russia*. LNCS, vol. 698. Springer-Verlag.
- [Hähnle et al., 2000] Hähnle, R., Hasegawa, R., and Shirai, Y. Model generation theorem proving with finite interval constraints. *Pages 285–399 of: Lloyd, J., Dahl, V., Furbach, U., Kerber, M., Lau, K.-K., Palamidessi, C., Pereira, L. M., Sagiv, Y., and Stuckey, P. J. (eds.), Proc. Computational Logic – CL 2000, First International Conference, London, UK*. LNCS, vol. 1861. Springer-Verlag.
- [Hájek, 1998] Hájek, P. *Metamathematics of Fuzzy Logic*. Trends in Logic: Studia Logica Library, vol. 4. Kluwer Academic Publishers, Dordrecht.
- [Hájek, 2000] Hájek, P. Mathematical fuzzy logic—state of art. *Pages 197–205 of: Buss, S., Hájek, P., and Pudlák, P. (eds.), Proc. Logic Colloquium ’98, Prague, Czech Republic*. Lecture Notes in Logic, vol. 13. Association for Symbolic Logic with A K Peters.
- [Hájek and Paris, 1997] Hájek, P., and Paris, J. A dialogue on fuzzy logic. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **1**(1), 3–5.
- [Hájek and Zach, 1994] Hájek, P., and Zach, R. Review of: Leonard Bolc and Piotr Borowik: Many-Valued Logics 1: Theoretical Foundations. *Journal of Applied Non-Classical Logics*, **4**(2), 215–220.
- [Hájek et al., 1996] Hájek, P., Godo, L., and Esteva, F. A complete many-valued logic with product conjunction. *Archive for Mathematical Logic*, **35**, 191–208.
- [Hájek et al., 2000] Hájek, P., Paris, J., and Sheperdson, J. The liar paradox and fuzzy logic. *Journal of Symbolic Logic*, **65**(1), 339–346.
- [Hájek et al., to appear, 2000] Hájek, P., Paris, J., and Sheperdson, J. Rational Pavelka predicate logic is a conservative extension of Lukasiewicz predicate logic. *Journal of Symbolic Logic*.
- [Hayes, 1986] Hayes, J. P. Pseudo-Boolean logic circuits. *IEEE Transactions on Computers*, **C-35**(7), 602–612.
- [Höhle, 1995] Höhle, U. Commutative, residuated  $\ell$ -monoids. *Pages 53–106 of: Höhle, U., and Klement, E. P. (eds.), Non-Classical Logics and their Applications to Fuzzy Subsets*. Kluwer, Dordrecht.
- [Hooker, 1988] Hooker, J. N. A quantitative approach to logical inference. *Decision Support Systems*, **4**, 45–69.
- [Hösli, 1993] Hösli, B. Robuste Logik. Ph.D. thesis, Eidgenössische Technische Hochschule Zürich.
- [Iturrioz, 1993] Iturrioz, L. *Logics of Approximating Reasoning, Semantically Based on Completely Symmetrical Posets of Cooperating Agents*. Draft, Université Claude Bernard, Lyon 1.

- [Iturrioz, 2000] Iturrioz, L. (ed.). *COST Action 15: Many-valued logics for computer science applications — Final report*. EUR 19204. Office for Official Publications of the European Communities, Luxembourg.
- [Iturrioz *et al.*, 2000] Iturrioz, L., Orłowska, E., and Turunen, E. (eds.). *Atlas of many-valued structures*. Mathematics Report 75. Tampere University of Technology, Department of Information Technology, Tampere, Finland.
- [Jeroslow, 1988] Jeroslow, R. G. *Logic-Based Decision Support. Mixed Integer Model Formulation*. Elsevier, Amsterdam.
- [Kam *et al.*, 1998] Kam, T., Villa, T., Brayton, R. K., and Sangiovanni-Vincentelli, A. Multi-valued decision diagrams: Theory and applications. *Multiple-Valued Logic*, **4**(1-2), 9-62.
- [Kameyama, 1997] Kameyama, M. Technical activities forum: Multiple-valued logic TC stresses innovation. *IEEE Computer*, May, 83-85.
- [Kawahito *et al.*, 1994] Kawahito, S., Ishida, M., Nakamura, T., Kameyama, M., and Higuchi, T. High-speed area-efficient multiplier design using multiple-valued current-mode circuits. *IEEE Transactions on Computers*, **43**(1), 34-42.
- [Kerber and Kohlhase, 1994] Kerber, M., and Kohlhase, M. A mechanization of strong Kleene logic for partial functions. *Pages 371-385 of: Bundy, A. (ed.), Proc. 12th International Conference on Automated Deduction, Nancy/France*. LNCS, vol. 814. Springer-Verlag.
- [Kerber and Kohlhase, 1996] Kerber, M., and Kohlhase, M. A resolution calculus for presuppositions. *Pages 375-379 of: Proc. 12th European Conference on Artificial Intelligence, ECAI-96*. John Wiley & Sons.
- [Kerber and Kohlhase, 1997] Kerber, M., and Kohlhase, M. Mechanising partiality without re-implementation. *Pages 123-134 of: Brewka, G., Habel, C., and Nebel, B. (eds.), Proc. 21st Annual German Conference on Artificial Intelligence (KI-97): Advances in Artificial Intelligence*. LNCS, vol. 1303. Springer-Verlag.
- [Kifer and Lozinskii, 1992] Kifer, M., and Lozinskii, E. L. A logic for reasoning with inconsistency. *Journal of Automated Reasoning*, **9**(2), 179-215.
- [Kifer and Lozinskij, 1989] Kifer, M., and Lozinskij, E. L. RI: A logic for reasoning with inconsistency. *Pages 253-262 of: Proc. Logic in Computer Science LICS*. IEEE Press, Los Alamitos.
- [Kifer and Subrahmanian, 1992] Kifer, M., and Subrahmanian, V. S. Theory of generalized annotated logic programming and its applications. *Journal of Logic Programming*, **12**, 335-367.
- [Kirin, 1966] Kirin, V. G. Gentzen's method of the many-valued propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **12**, 317-332.
- [Klawonn and Kruse, 1994] Klawonn, F., and Kruse, R. A Lukasiewicz logic based Prolog. *Mathware & Soft Computing*, **1**(1), 5-29.
- [Klement *et al.*, 2000] Klement, E. P., Mesiar, R., and Pap, E. *Triangular Norms*. Trends in Logic, vol. 8. Kluwer, Dordrecht.
- [Klir and Yuan, 1996] Klir, G. J., and Yuan, B. (eds.). *Fuzzy Sets, Fuzzy Logic, and Fuzzy Systems: Selected Papers by Lotfi A. Zadeh*. Advances in Fuzzy Systems, vol. 6. World Scientific Publishing, Singapore.
- [Konikowska *et al.*, 1998] Konikowska, B., Morgan, C. G., and Orłowska, E. A relational formalisation of arbitrary finite valued logics. *Logic Journal of the IGPL*, **6**(5), 755-774.
- [Kruse *et al.*, 1994] Kruse, R., Gebhardt, J., and Klawonn, F. *Foundations of Fuzzy System*. Wiley, Chichester.
- [Kunen, 1987] Kunen, K. Negation in logic programming. *Journal of Logic Programming*, **4**, 289-308.
- [Lakshmanan and Sadri, 1994] Lakshmanan, L. V., and Sadri, F. Modeling uncertainty in deductive databases. *Pages 724-733 of: Karagiannis, D. (ed.), Proc. Int. Conf. on Database and Expert Systems Applications, DEXA '94, Athens, Greece*. LNCS, vol. 856.
- [Leach and Lu, 1996] Leach, S. M., and Lu, J. J. Query processing in annotated logic programming: Theory and implementation. *Journal of Intelligent Information Systems*, **6**(1), 33-58.



- [Leach *et al.*, 1998] Leach, S. M., Lu, J. J., Murray, N. V., and Rosenthal, E.  $\mathcal{U}$ -resolution: An inference rule for regular multiple-valued logics. *Pages 154–168 of: Dix, J., Fariñas del Cerro, L., and Furbach, U. (eds.), Proc. 6th European Workshop on Logics in AI (JELIA)*. LNCS, vol. 1489. Springer-Verlag.
- [Lee, 1972] Lee, R. C. T. Fuzzy logic and the resolution principle. *Journal of the ACM*, **19**(1), 109–119.
- [Lee and Chang, 1971] Lee, R. C. T., and Chang, C.-L. Some properties of fuzzy logic. *Information and Control*, **19**(5), 417–431.
- [Lee, 1997] Lee, S. Error tolerance method in multiple-valued logic. *Pages 392–405 of: Gabbay, D. M., Kruse, R., Nonnengart, A., and Ohlbach, H. J. (eds.), Proc. First Int. Joint Conference on Qualitative and Quantitative Practical Reasoning ECSQARU-FAPR, Bad Honnef, Germany*. LNCS, vol. 1244. Springer-Verlag.
- [Lehmke, 1995] Lehmke, S. On resolution-based theorem proving in propositional fuzzy logic with ‘bold’ connectives. Master’s thesis, Universität Dortmund, Fachbereich Informatik.
- [Lehmke, 1996] Lehmke, S. A resolution-based axiomatisation of ‘bold’ propositional fuzzy logic. *Pages 115–119 of: Dubois, D., Klement, E. P., and Prade, H. (eds.), Linz’96: Fuzzy Sets, Logics, and Artificial Intelligence. Abstracts*.
- [Lloyd, 1987] Lloyd, J. W. *Foundations of Logic Programming*. Second edn. Springer, Berlin.
- [Lobo *et al.*, 1992] Lobo, J., Minker, J., and Rajasekar, A. *Foundations of Disjunctive Logic Programming*. MIT Press.
- [Lu, 1996] Lu, J. J. Logic programming with signs and annotations. *Journal of Logic and Computation*, **6**(6), 755–778.
- [Lu and Rosenthal, 1997] Lu, J. J., and Rosenthal, E. Logic-based deductive reasoning in AI systems. *Chap. 29, pages 654–657 of: Tucker, A. B. (ed.), The Computer Science And Engineering Handbook*. CRC Press.
- [Lu *et al.*, 1991] Lu, J. J., Henschen, L. J., Subrahmanian, V. S., and da Costa, N. C. A. Reasoning in paraconsistent logics. *Pages 181–210 of: Boyer, R. (ed.), Automated Reasoning: Essays in Honor of Woody Bledsoe*. Kluwer.
- [Lu *et al.*, 1993] Lu, J. J., Murray, N. V., and Rosenthal, E. Signed formulas and annotated logics. *Pages 48–53 of: Proc. 23rd International Symposium on Multiple-Valued Logics*. IEEE CS Press, Los Alamitos.
- [Lu *et al.*, 1997] Lu, J. J., Calmet, J., and Schü, J. Computing multiple-valued logic programs. *Mathware & Soft Computing*, **IV**(2), 129–153. Special Issue on Deduction in Many-Valued Logic.
- [Lu *et al.*, 1998] Lu, J. J., Murray, N. V., and Rosenthal, E. A framework for automated reasoning in multiple-valued logics. *Journal of Automated Reasoning*, **21**(1), 39–67.
- [Lukasiewicz, 1920] Lukasiewicz, J. O logice trójwartościowej. *Ruch Filozoficzny*, **5**, 169–171. Reprinted and translated in [Lukasiewicz, 1970].
- [Lukasiewicz, 1957] Lukasiewicz, J. *Aristotle’s syllogistic from the standpoint of modern formal logic*. 2nd edn. Clarendon Press, Oxford.
- [Lukasiewicz, 1970] Lukasiewicz, J. *Jan Lukasiewicz, Selected Writings*. North-Holland. Edited by L. Borowski.
- [Lukasiewicz and Tarski, 1930] Lukasiewicz, J., and Tarski, A. Untersuchungen über den Aussagenkalkül. *Comptes Rendus des Séances de la Société des Sciences et des Lettres de Varsovie, Classe III*, **23**, 1–21. Reprinted and translated in [Lukasiewicz, 1970].
- [Malinowski, 1993] Malinowski, G. *Many-Valued Logics*. Oxford Logic Guides, vol. 25. Oxford University Press.
- [Mamdani and Assilian, 1975] Mamdani, E. H., and Assilian, S. An experiment in linguistic synthesis with a fuzzy logic controller. *International Journal of Man-Machine Studies*, **7**(1), 1–13.
- [Manyà, 1996] Manyà, F. Proof Procedures for Multiple-Valued Propositional Logics. Ph.D. thesis, Facultat de Ciències, Universitat Autònoma de Barcelona. Published as [Manyà, 1999].

- [Manyà, 1999] Manyà, F. *Proof Procedures for Multiple-Valued Propositional Logics*. Monografies de l'Institut d'Investigació en Intel·ligència Artificial, vol. 9. IIIA-CSIC, Bellaterra (Barcelona).
- [Manyà, 2000] Manyà, F. The 2-SAT problem in signed CNF formulas. *Multiple-Valued Logic. An International Journal*, **5**.
- [Manyà *et al.*, 1998] Manyà, F., Béjar, R., and Escalada-Imaz, G. The satisfiability problem in regular CNF-formulas. *Soft Computing—A Fusion of Foundations, Methodologies and Applications*, **2**(3), 116–123.
- [Marks, 1994] Marks, R. J. (ed.). *Fuzzy logic technology and applications*. IEEE technology updates. IEEE Press, New York.
- [Martínez, 1990] Martínez, N. G. Priestley duality for Wajsberg algebras. *Studia Logica*, **49**(1), 31–46.
- [Martínez, 1994] Martínez, N. G. A topological duality for lattice-ordered algebraic structures, including  $\ell$ -groups. *Algebra Universalis*, **31**, 516–541.
- [Martínez, 1996] Martínez, N. G. A simplified duality for  $\ell$ -groups and implicative lattices. *Studia Logica*, **56**(1–2), 185–204.
- [Martínez and Priestley, to appear, 2000] Martínez, N. G., and Priestley, H. A. On the Priestley space of lattice-ordered algebraic structures. *Order*.
- [McNaughton, 1951] McNaughton, R. A theorem about infinite-valued sentential logic. *Journal of Symbolic Logic*, **16**(1), 1–13.
- [Meinke and Tucker, 1992] Meinke, K., and Tucker, J. V. Universal algebra. *Pages 189–409 of: Abramsky, S., Gabbay, D. M., and Maibaum, T. S. E. (eds.), Handbook of Logic in Computer Science. Volume 1. Background: Mathematical Structures*. Oxford University Press.
- [Messing, 1995] Messing, B. Knowledge representation in many-valued Horn clauses. *Pages 83–92 of: Proceedings of the 6th Conference of the Spanish Association for Artificial Intelligence, Alicante*. Asociacion Española para la Inteligencia Artificial, AEPIA.
- [Messing, 1997] Messing, B. Combining knowledge with many-valued logics. *Data & Knowledge Engineering*, **23**(3), 297–316. Special Issue on Distributed Expertise.
- [Messing and Stackelberg, 1995] Messing, B., and Stackelberg, P. Regular signed resolution applied to annotated logic programs. Poster abstract. *Page 268 of: Lloyd, J. (ed.), Proceedings of the International Logic Programming Conference, Portland/OR*. MIT Press.
- [Miglioli *et al.*, 1994] Miglioli, P., Moscato, U., and Ornaghi, M. An improved refutation system for intuitionistic predicate logic. *Journal of Automated Reasoning*, **13**(3), 361–374.
- [Miglioli *et al.*, 1995] Miglioli, P., Moscato, U., and Ornaghi, M. Refutation systems for propositional modal logics. *Pages 95–105 of: Baumgartner, P., Hähnle, R., and Posegga, J. (eds.), Proc. 4th Workshop on Deduction with Tableaux and Related Methods, St. Goar, Germany*. LNCS, vol. 918. Springer-Verlag.
- [Minato, 1996] Minato, S. *Binary Decision Diagrams and Applications for VLSI CAD*. Kluwer, Norwell/MA, USA.
- [Mitchell *et al.*, 1992] Mitchell, D., Selman, B., and Levesque, H. Hard and easy distributions of SAT problems. *Pages 459–465 of: Proc. of AAAI-92, San Jose, CA*.
- [Morgan, 1976] Morgan, C. G. A resolution principle for a class of many-valued logics. *Logique et Analyse*, **19**(74–75–76), 311–339.
- [Morgan and Orłowska, 1993] Morgan, C. G., and Orłowska, E. Kripke and relational style semantics and associated tableau proof systems for arbitrary finite valued logics. *In: Proc. Second Workshop on Theorem Proving with Tableau-Based and Related Methods, Marseille*. Tech. Report, MPII Saarbrücken.
- [Mostert and Shields, 1957] Mostert, P. S., and Shields, A. L. On the structure of semigroups on a compact manifold with boundary. *Annals of Mathematics*, **65**, 117–143.
- [Mostowski, 1948] Mostowski, A. Proofs of non-deducibility in intuitionistic functional calculus. *Journal of Symbolic Logic*, **13**(4), 204–207.
- [Mostowski, 1961] Mostowski, A. Axiomatizability of some many valued predicate calculi. *Fundamenta Mathematicæ*, **L**, 165–190.

- [Mundici, 1986] Mundici, D. Interpretation of AF  $C^*$ -algebras in Lukasiewicz sentential calculus. *Journal of Functional Analysis*, **65**, 15–63.
- [Mundici, 1987] Mundici, D. Satisfiability in many-valued sentential logic is NP-complete. *Theoretical Computer Science*, **52**, 145–153.
- [Mundici, 1990] Mundici, D. The complexity of adaptive error-correcting codes. *Pages 300–307 of: Börger, E., Kleine Büning, H., Richter, M. M., and Schönfeld, W. (eds.), Proc. Workshop Computer Science Logic 90, Heidelberg, Germany*. LNCS, vol. 533. Springer-Verlag.
- [Mundici, 1991] Mundici, D. Normal forms in infinite-valued logic: The case of one variable. *Pages 272–277 of: Brger, E., Jäger, G., Kleine Büning, H., and Richter, M. M. (eds.), Proc. Workshop Computer Science Logic 91, Berne, Switzerland*. LNCS, vol. 626. Springer-Verlag.
- [Mundici, 1992] Mundici, D. The logic of Ulam's game with lies. *Pages 275–284 of: Bicchieri, C., and Dalla Chiara, M. L. (eds.), Proc. International Conference Knowledge, Belief and Strategic Interaction, Castiglioncello, Italy, 1989*. Cambridge Studies in Probability, Induction and Decision Theory. Cambridge University Press.
- [Mundici, 1993] Mundici, D. Logic of infinite quantum systems. *International Journal of Theoretical Physics*, **32**(10), 1941–1955.
- [Mundici, 1994] Mundici, D. A constructive proof of McNaughton's Theorem in infinite-valued logic. *Journal of Symbolic Logic*, **59**(2), 596–602.
- [Mundici, 1996] Mundici, D. Lukasiewicz normal forms and toric desingularizations. *Pages 401–423 of: Hodges, W., Hyland, M., Steinhorn, C., and Truss, J. (eds.), Logic: from foundations to applications. Proc. Logic Colloquium 1993, Staffordshire, England*. Oxford University Press, New York.
- [Mundici and Olivetti, 1998] Mundici, D., and Olivetti, N. Resolution and model building in the infinite-valued calculus of Lukasiewicz. *Theoretical Computer Science*, **200**(1–2), 335–366.
- [Mundici and Pasquetto, 1995] Mundici, D., and Pasquetto, M. A proof of the completeness of the infinite-valued calculus of Lukasiewicz with one variable. *Pages 107–123 of: Klement, E. P., and Höhle, U. (eds.), Non-classical Logics and their Applications to Fuzzy Subsets (Selected Papers of Int. Conference on Nonclassical Logics and their Applications 1992, Linz, Austria)*. Kluwer, Dordrecht.
- [Murata *et al.*, 1991] Murata, T., Subrahmanian, V. S., and Wakayama, T. A Petri net model for reasoning in the presence of inconsistency. *IEEE Transactions on Knowledge and Data Engineering*, **3**(3), 281–292.
- [Murray, 1982] Murray, N. V. Completely non-clausal theorem proving. *Artificial Intelligence*, **18**, 67–85.
- [Murray and Rosenthal, 1991a] Murray, N. V., and Rosenthal, E. Improving tableau deductions in multiple-valued logics. *Pages 230–237 of: Proceedings 21st International Symposium on Multiple-Valued Logic, Victoria*. IEEE Press, Los Alamitos.
- [Murray and Rosenthal, 1991b] Murray, N. V., and Rosenthal, E. Resolution and path-dissolution in multiple-valued logics. *Pages 570–579 of: Ras, M., and Zemankova, Z. (eds.), Proc. 6th International Symposium on Methodologies for Intelligent Systems ISMIS, Charlotte/NC, USA*. LNCS, vol. 542. Springer-Verlag.
- [Murray and Rosenthal, 1993a] Murray, N. V., and Rosenthal, E. Dissolution: Making paths vanish. *Journal of the ACM*, **40**(3), 504–535.
- [Murray and Rosenthal, 1993b] Murray, N. V., and Rosenthal, E. Signed formulas: A liftable meta logic for multiple-valued logics. *Pages 275–284 of: Komorowski, J., and Raś, Z. (eds.), Proc. 7th International Symposium on Methodologies for Intelligent Systems (ISMIS), Trondheim, Norway*. LNCS, vol. 689. Springer-Verlag.
- [Murray and Rosenthal, 1994] Murray, N. V., and Rosenthal, E. Adapting classical inference techniques to multiple-valued logics using signed formulas. *Fundamenta Informaticae*, **21**(3), 237–253.
- [Muzio and Wesselkamper, 1986] Muzio, J. C., and Wesselkamper, T. *Multiple-Valued Switching Theory*. Adam Hilger Ltd., Bristol and Boston.
- [Ng and Subrahmanian, 1993] Ng, R., and Subrahmanian, V. S. A semantical framework for supporting subjective and conditional probabilities in deductive databases. *Journal of Automated Reasoning*, **10**(2), 191–235.

- [Novák, 1989] Novák, V. *Fuzzy Sets and their Applications*. Bristol: Adam Hilger.
- [Novák, 1995] Novák, V. A new proof of completeness of fuzzy logic and some conclusions for approximate reasoning. *Pages 1461–1468 of: Proc. FUZZ-IEEE/IFES, Yokohama, Japan*. IEEE CS Press.
- [Orłowska, 1967] Orłowska, E. Mechanical proof procedure for the  $n$ -valued propositional calculus. *Bull. de L'Acad. Pol. des Sci., Série des sci. math., astr. et phys.*, **XV**(8), 537–541.
- [Orłowska, 1978] Orłowska, E. The resolution principle for  $\omega^+$ -valued logic. *Fundamenta Informaticae*, **II**(1), 1–15.
- [Orłowska, 1991a] Orłowska, E. Post relation algebras and their proof system. *Pages 298–307 of: Proc. 21st International Symposium on Multiple-Valued Logic, Victoria/BC, Canada*. IEEE Computer Society Press.
- [Orłowska, 1991b] Orłowska, E. Relational interpretation of modal logics. *Pages 443–471 of: Andréka, H., Nemeti, I., and Monk, D. (eds.), Algebraic Logic*. Colloquia mathematica Societatis János Bolyai, vol. 54. North-Holland, Amsterdam.
- [Orłowska and Iturrioz, 1999] Orłowska, E., and Iturrioz, L. A Kripke-style and relational semantics for logics based on lukasiewicz algebras. *In: Baghrmian, M., and Simons, P. M. (eds.), Lukasiewicz in Dublin: an International Conference on the Work of Jan Lukasiewicz, July 1996*. Oxford University Press.
- [Panti, 1995] Panti, G. A geometric proof of the completeness of the Lukasiewicz calculus. *Journal of Symbolic Logic*, **60**(2), 563–578.
- [Panti, 1998] Panti, G. Multi-valued logics. *Chap. 2, pages 25–74 of: Gabbay, D., and Smets, P. (eds.), Handbook of Defeasible Reasoning and Uncertainty Management Systems*, vol. 1: Quantified Representation of Uncertainty and Imprecision. Kluwer, Dordrecht.
- [Papadimitriou, 1994] Papadimitriou, C. H. *Computational Complexity*. Addison-Wesley, New York.
- [Patel-Schneider, 1990] Patel-Schneider, P. F. A decidable first-order logic for knowledge representation. *Journal of Automated Reasoning*, **6**, 361–388.
- [Pavelka, 1979a] Pavelka, J. On fuzzy logic I: Many-valued rules of inference. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **25**, 45–72.
- [Pavelka, 1979b] Pavelka, J. On fuzzy logic II: Enriched residuated lattices and semantics of propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **25**, 119–134.
- [Pavelka, 1979c] Pavelka, J. On fuzzy logic III: Semantical completeness of some many-valued propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **25**, 447–464.
- [Perkowski, 1992] Perkowski, M. A. The generalized orthonormal expansion of functions with multiple-valued inputs and some of its application. *Pages 442–450 of: Proc. 22nd International Symposium on Multiple-Valued Logic*. IEEE CS Press, Los Alamitos.
- [Pfalzgraf, 1991] Pfalzgraf, J. Logical fiberings and polycontextural systems. *Pages 170–184 of: Jorrand, P., and Kelemen, J. (eds.), Proc. International Workshop on Fundamentals of Artificial Intelligence Research (FAIR), Smolenice, Czechoslovakia*. LNCS, vol. 535. Springer-Verlag.
- [Plaisted and Greenbaum, 1986] Plaisted, D. A., and Greenbaum, S. A structure-preserving clause form translation. *Journal of Symbolic Computation*, **2**, 293–304.
- [Plaza, 1996] Plaza, J. A. On the propositional SLDNF-resolution. *International Journal of Foundations of Computer Science*, **7**(4), 359–406.
- [Pogorzelski, 1964] Pogorzelski, W. A. The deduction theorem for Lukasiewicz many-valued propositional calculi. *Studia Logica*, **15**, 7–23.
- [Posegga, 1993] Posegga, J. Deduktion mit Shannongraphen für Prädikatenlogik erster Stufe (in German). Ph.D. thesis, University of Karlsruhe. Diski 51, infix Verlag.
- [Posegga and Schmitt, 1995] Posegga, J., and Schmitt, P. H. Deduction with first-order Shannon graphs. *Journal of Logic and Computation*, **5**(6), 697–729.
- [Post, 1921] Post, E. L. Introduction to a general theory of elementary propositions. *American Journal of Mathematics*, **43**, 163–185. Reprinted in [van Heijenoort, 1967, pp. 264–283].

- [Pottinger, 1983] Pottinger, G. Uniform, cut-free formulations of T, S4 and S5 (abstract). *Journal of Symbolic Logic*, **48**(3), 900.
- [Prijatelj, 1996] Prijatelj, A. Bounded contraction and Gentzen-style formulation of Lukasiewicz logics. *Studia Logica*, **57**(2-3), 437-456.
- [Puyol-Gruart, 1996] Puyol-Gruart, J. *MILORD II: A Language for Knowledge-Based Systems*. Monografies del IIIA, vol. 1. IIIA-CSIC, Artificial Intelligence Research Institute of the Spanish Scientific Research Council.
- [Ragaz, 1981] Ragaz, M. E. Arithmetische Klassifikation von Formelmengen der unendlichwertigen Logik. Ph.D. thesis, ETH Zürich.
- [Ramesh and Murray, 1994] Ramesh, A., and Murray, N. V. Computing prime implicants/implicates for regular logics. *Pages 115-123 of: Proc. 24th International Symposium on Multiple-Valued Logic, Boston/MA, USA*. IEEE CS Press, Los Alamitos.
- [Ramesh and Murray, 1997] Ramesh, A., and Murray, N. V. Parameterized prime implicant/implicate computations for regular logics. *Mathware & Soft Computing*, **IV**(2), 155-179. Special Issue on Deduction in Many-Valued Logic.
- [Ramesh *et al.*, 1997a] Ramesh, A., Becker, G., and Murray, N. V. CNF and DNF considered harmful for computing prime implicants/implicates. *Journal of Automated Reasoning*, **18**(3), 337-356.
- [Ramesh *et al.*, 1997b] Ramesh, A., Beckert, B., Hähnle, R., and Murray, N. V. Fast subsumption checks using anti-links. *Journal of Automated Reasoning*, **18**(1), 47-84.
- [Rasiowa, 1973] Rasiowa, H. On generalized Post algebras of order  $\omega^+$  and  $\omega^+$ -valued predicate calculi. *Bull. Acad. Polon. Sci., Série Sci. Math. Astr. Phys.*, **XXI**, 209-219.
- [Rasiowa, 1974] Rasiowa, H. *An Algebraic Approach to Non-Classical Logics*. Studies in Logic and the Foundations of Mathematics, vol. 78. North-Holland, Amsterdam.
- [Rasiowa and Sikorski, 1963] Rasiowa, H., and Sikorski, R. *The Mathematics of Metamathematics*. Polish Scientific Publishers, Warsaw.
- [Rescher, 1969] Rescher, N. *Many-Valued Logic*. McGraw-Hill, New York.
- [Rine, 1984] Rine, D. C. (ed.). *Computer Science and Multiple-Valued Logics*. Second edn. North-Holland, Amsterdam. Selected Papers from the International Symposium on Multiple-Valued Logics 1974.
- [Robinson, 1968] Robinson, J. A. The generalized resolution principle. *Pages 77-93 of: Machine Intelligence*, vol. 3. Oliver and Boyd, Edinburgh. Reprinted in [Siekman and Wrightson, 1983].
- [Rose and Rosser, 1958] Rose, A., and Rosser, J. B. Fragments of many-valued statement calculi. *Transactions of the American Mathematical Society*, **87**, 1-53.
- [Rosser and Turquette, 1952] Rosser, J. B., and Turquette, A. R. *Many-Valued Logics*. Amsterdam: North-Holland.
- [Rousseau, 1967] Rousseau, G. Sequents in many valued logic I. *Fundamenta Mathematicæ*, **LX**, 23-33.
- [Rousseau, 1970] Rousseau, G. Sequents in many valued logic II. *Fundamenta Mathematicæ*, **LXVII**, 125-131.
- [Rudell and Sangiovanni-Vincentelli, 1987] Rudell, R., and Sangiovanni-Vincentelli, A. Multiple-valued minimization for PLA optimization. *IEEE Transactions on Computer-Aided Design*, **6**(5), 727-750.
- [Salzer, 1996a] Salzer, G. MULTlog: an expert system for multiple-valued logics. *Pages 50-55 of: Collegium Logicum. Annals of the Kurt-Gödel-Society*, vol. 2. Springer-Verlag, Wien.
- [Salzer, 1996b] Salzer, G. Optimal axiomatizations for multiple-valued operators and quantifiers based on semilattices. *Pages 688-702 of: McRobbie, M., and Slaney, J. (eds.), Proc. 13th Conference on Automated Deduction, New Brunswick/NJ, USA*. LNCS, vol. 1104. Springer-Verlag.
- [Sasao, 1981] Sasao, T. Multiple-valued decomposition of generalized Boolean functions and the complexity of programmable logic arrays. *IEEE Transactions on Computers*, **C-30**(Sept.), 635-643.
- [Sasao, 1993a] Sasao, T. (ed.). *Logic Synthesis and Optimization*. Kluwer, Norwell/MA, USA.

- [Sasao, 1993b] Sasao, T. Logic synthesis with EXOR gates. *Chap. 12, pages 259–286 of: Sasao, T. (ed.), Logic Synthesis and Optimization*. Kluwer, Norwell/MA, USA.
- [Sasao, 1996] Sasao, T. Ternary decision diagrams and their applications. *Chap. 12, pages 269–292 of: Sasao, T., and Fujita, M. (eds.), Representations of Discrete Functions*. Kluwer, Norwell/MA, USA.
- [Sasao, 1999] Sasao, T. *Switching Theory for Logic Synthesis*. Kluwer, Norwell/MA, USA.
- [Sasao and Butler, 1997] Sasao, T., and Butler, J. T. Comparison of the worst and best sum-of-products expressions for multiple-valued functions. *Pages 55–60 of: Proc. 27th International Symposium on Multiple-Valued Logic, Nova Scotia, Canada*. IEEE CS Press, Los Alamitos.
- [Sasao and Fujita, 1996] Sasao, T., and Fujita, M. (eds.). *Representations of Discrete Functions*. Kluwer Academic Publishers, Boston.
- [Scarpellini, 1962] Scarpellini, B. Die Nichtaxiomatisierbarkeit des unendlichwertigen Prädikatenkalküls von Lukasiewicz. *Journal of Symbolic Logic*, **27**(2), 159–170.
- [Schmitt, 1986] Schmitt, P. H. Computational aspects of three-valued logic. *Pages 190–198 of: Siekmann, J. H. (ed.), Proc. 8th International Conference on Automated Deduction*. LNCS, vol. 230. Springer-Verlag.
- [Schmitt, 1989] Schmitt, P. H. Perspectives in multi-valued logic. *Pages 206–220 of: Studer, R. (ed.), Proc. International Scientific Symposium on Natural Language and Logic, Hamburg*. LNCS, vol. 459. Springer-Verlag.
- [Schrijver, 1986] Schrijver, A. *Theory of Linear and Integer Programming*. Wiley-Interscience Series in Discrete Mathematics. John Wiley & Sons.
- [Selman *et al.*, 1992] Selman, B., Levesque, H., and Mitchell, D. A new method for solving hard satisfiability problems. *Pages 440–446 of: Proc. of AAAI-92, San Jose/CA, USA*. AAAI Press.
- [Selman *et al.*, 1994] Selman, B., Kautz, H. A., and Cohen, B. Noise strategies for local search. *Pages 337–343 of: Proc. 12th National Conference on Artificial Intelligence, AAAI'94, Seattle/WA, USA*. AAAI Press.
- [Sette, 1973] Sette, A. M. On the propositional calculus  $p^1$ . *Mathematica Japonicae*, **18**, 173–180.
- [Sette and Carnielli, 1995] Sette, A. M., and Carnielli, W. A. Maximal weakly-intuitionistic logics. *Studia Logica*, **55**(1), 181–203.
- [Shannon, 1938] Shannon, C. E. A symbolic analysis of relay and switching circuits. *AIEE Transactions*, **67**, 713–723.
- [Siekmann and Wrightson, 1983] Siekmann, J., and Wrightson, G. (eds.). *Automation of Reasoning: Classical Papers in Computational Logic 1967–1970*. Vol. 2. Springer-Verlag.
- [Soare, 1987] Soare, R. I. *Recursively Enumerable Sets and Degrees*. Perspectives in Mathematical Logic, Omega Series. Springer-Verlag.
- [Sofronie-Stokkermans, 1997] Sofronie-Stokkermans, V. Fibered Structures and Applications to Automated Theorem Proving in Certain Classes of Finitely-Valued Logics and to Modeling Interacting Systems. Ph.D. thesis, Johannes Kepler Universität Linz, Forschungsinstitut für symbolisches Rechnen.
- [Sofronie-Stokkermans, 1998] Sofronie-Stokkermans, V. On translation of finitely-valued logics to classical first-order logic. *Pages 410–411 of: Prade, H. (ed.), Proc. 13th European Conference on Artificial Intelligence, Brighton*. John Wiley & Sons.
- [Sofronie-Stokkermans, 1999a] Sofronie-Stokkermans, V. On the universal theory of varieties of distributive lattices with operators: Some decidability and complexity results. *Pages 157–171 of: Ganzinger, H. (ed.), Proc. CADE-16, 16th International Conference on Automated Deduction, Trento, Italy*. LNCS, vol. 1632. Springer-Verlag.
- [Sofronie-Stokkermans, 1999b] Sofronie-Stokkermans, V. Representation theorems and automated theorem proving in non-classical logics. *Pages 242–247 of: Proc. 29th International Symposium on Multiple-Valued Logics, Freiburg, Germany*. IEEE CS Press, Los Alamitos.
- [Sofronie-Stokkermans, 2000] Sofronie-Stokkermans, V. Duality and canonical extensions of bounded distributive lattices with operators, and applications to the semantics of non-classical logics I, II. *Studia Logica*, **64**(1–2), 93–132, 151–172.

- [Sofronie-Stokkermans, to appear, 2000] Sofronie-Stokkermans, V. Automated theorem proving by resolution for finitely-valued logics based on distributive lattices with operators. *Multiple-Valued Logic*.
- [Srinivasan *et al.*, 1990] Srinivasan, A., Kam, T., Malik, S., and Brayton, R. E. Algorithms for discrete function manipulation. *Pages 92–95 of: Proc. IEEE International Conference on CAD, Santa Clara/CA, USA*. IEEE CS Press, Los Alamitos.
- [Stachniak, 1988] Stachniak, Z. The resolution rule: An algebraic perspective. *Pages 227–242 of: Bergman, C., Maddux, R., and Pigozzi, D. (eds.), Proc. Conference on Algebraic Logic and Universal Algebra in Computer Science, Ames, USA*. LNCS, vol. 425. Springer-Verlag.
- [Stachniak, 1996] Stachniak, Z. *Resolution Proof Systems: an Algebraic Theory*. Kluwer, Dordrecht.
- [Stachniak and O’Hearn, 1990] Stachniak, Z., and O’Hearn, P. Resolution in the domain of strongly finite logics. *Fundamenta Informaticae*, **XIII**, 333–351.
- [Stärk, 1991] Stärk, R. F. A complete axiomatization of the three-valued completion of logic programs. *Journal of Logic and Computation*, **1**(6), 811–834.
- [Strother Moore, 1994] Strother Moore, J. Introduction to the OBDD algorithm for the ATP community. *Journal of Automated Reasoning*, **12**(1), 33–45.
- [Subrahmanian, 1994] Subrahmanian, V. S. Amalgamating knowledge bases. *ACM Transactions on Database Systems*, **19**(2), 291–331.
- [Suchoń, 1974] Suchoń, W. La méthode de Smullyan de construire le calcul n-valent de Łukasiewicz avec implication et négation. *Reports on Mathematical Logic, Universities of Cracow and Katowice*, **2**, 37–42.
- [Surma, 1974] Surma, S. J. An algorithm for axiomatizing every finite logic. *Reports on Mathematical Logic*, **3**, 57–62.
- [Surma, 1984] Surma, S. J. An algorithm for axiomatizing every finite logic. *Pages 143–149 of: Rine, D. C. (ed.), Computer Science and Multiple-Valued Logics*, second edn. North-Holland, Amsterdam. Selected Papers from the International Symposium on Multiple-Valued Logics 1974.
- [Szabo, 1969] Szabo, M. E. (ed.). *The Collected Papers of Gerhard Gentzen*. North-Holland, Amsterdam.
- [Takahashi, 1967] Takahashi, M. Many-valued logics of extended Gentzen style I. *Science Reports of the Tokyo Kyoiku Daigaku, Section A*, **9**(231), 95–116.
- [Tarski, 1941] Tarski, A. On the calculus of relations. *Journal of Symbolic Logic*, **6**(3), 73–89.
- [Tassart *et al.*, 1995] Tassart, G., Iturrioz, L., Klement, E. P., Mundici, D., Prade, H., Schmitt, P., and Hähnle, R. COST Action 15: Many-valued logics for computer science applications. *Computational Logic*, **2**(2), 32–33.
- [Thayse *et al.*, 1979] Thayse, A., Davio, M., and Deschamps, J.-P. Optimization of multiple-valued decision diagrams. *Pages 171–177 of: Proc. International Symposium on Multiple-Valued Logics, ISMVL’79, Rosemont/IL, USA*. IEEE CS Press, Los Alamitos.
- [Thiele, 1998] Thiele, H. On closure operators in fuzzy deductive systems and fuzzy algebras. *Pages 304–309 of: Proc. 28th International Symposium on Multiple-Valued Logics, Fukuoka, Japan*. IEEE Computer Society, Los Alamitos.
- [Thiele and Kalenka, 1993] Thiele, H., and Kalenka, S. On fuzzy temporal logic. *Pages 1027–1032 of: Proc. 2nd IEEE International Conference on Fuzzy Systems, San Francisco, USA*, vol. II. IEEE Press.
- [Thiele and Lehmke, 1994] Thiele, H., and Lehmke, S. On ‘bold’ resolution theory. *Pages 1945–1950 of: Proc. 3rd IEEE Conference on Fuzzy Systems, Orlando, USA*, vol. III. IEEE Press.
- [Thiele and Schmechel, 1995] Thiele, H., and Schmechel, N. On the mutual definability of fuzzy equivalence relations and fuzzy partitions. *In: Proc. FUZZ-IEEE/IFES, Yokohama, Japan*. IEEE Press.
- [Tseitin, 1970] Tseitin, G. On the complexity of proofs in propositional logics. *Seminars in Mathematics*, **8**. Reprinted in [Siekmann and Wrightson, 1983].
- [Turner, 1994] Turner, H. Signed logic programs. *Pages 61–75 of: Bruynooghe, M. (ed.), Logic Programming: Proc. of the 1994 International Symposium*. The MIT Press.

- [Turunen, 1999] Turunen, E. *Mathematics Behind Fuzzy Logic*. Advances in Soft Computing. Springer-Verlag.
- [Ullman, 1988] Ullman, J. D. *Principles of Database and Knowledge-Bade Systems. Volume I: Classical Database Systems*. Computer Science Press.
- [Urquhart, 1986] Urquhart, A. Many-valued logic. *Chap. 2, pages 71–116 of*: Gabbay, D., and Guenther, F. (eds.), *Handbook of Philosophical Logic, Vol. III: Alternatives in Classical Logic*. Reidel, Dordrecht.
- [van Heijenoort, 1967] van Heijenoort, J. (ed.). *From Frege to Gödel. A Source Book in Mathematical Logic, 1879–1931*. Harvard University Press, Cambridge/MA.
- [Vienna Group for Multiple Valued Logics, 1996] Vienna Group for Multiple Valued Logics. MUltlog 1.0: Towards an expert system for many-valued logics. *Pages 226–230 of*: McRobbie, M., and Slaney, J. (eds.), *Proc. 13th Conference on Automated Deduction, New Brunswick/NJ, USA*. LNCS, vol. 1104. Springer-Verlag.
- [Vojtáš, 1998] Vojtáš, P. Fuzzy reasoning with tunable  $t$ -operators. *Journal of Advanced Computational Intelligence*, **2**, 121–127.
- [Vojtáš and Paulik, 1996] Vojtáš, P., and Paulik, L. Soundness and completeness of non-classical extended SLD-resolution. *Pages 289–301 of*: Dyckhoff, R., Herre, H., and Schroeder-Heister, P. (eds.), *Proc. Extensions of Logic Programming, 5th International Workshop, Leipzig, Germany*. LNCS, vol. 1050.
- [Wagner, 1997] Wagner, H. Nonaxiomatizability and undecidability of an infinite-valued temporal logic. *Multiple-Valued Logic*, **2**(1), 47–58.
- [Wójcicki, 1988] Wójcicki, R. *Theory of Logical Calculi*. Reidel, Dordrecht.
- [Yasui and Mukaidono, 1996] Yasui, H., and Mukaidono, M. A consideration of fuzzy logic programming based on Lukasiewicz's implication. *Japanese Journal of Fuzzy Theory and Systems*, **8**(5), 863–878.
- [Zach, 1993] Zach, R. Proof Theory of Finite-valued Logics. Master's thesis, Institut für Algebra und Diskrete Mathematik, TU Wien. Available as Technical Report TUW-E185.2-Z.1-93.
- [Zadeh, 1979] Zadeh, L. A. A theory of approximate reasoning. *Pages 149–196 of*: Hayes, J. E., and Mikulich, L. I. (eds.), *Machine Intelligence*, vol. 9. Ellis Horwood/John Wiley, New York.
- [Zimmermann, 1991] Zimmermann, H.-J. *Fuzzy Set Theory—And Its Applications*. Second Revised edn. Kluwer, Dordrecht.
- [Zinov'ev, 1963] Zinov'ev, A. A. *Philosophical Problems of Many-Valued Logic*. D. Reidel, Dordrecht.





## INDEX

- |   |  |
|---|--|
| <p> <math>1^{\mathfrak{A}}</math>, 139<br/> <i>FV</i>, 193<br/> <i>Fs</i>, 193<br/> <math>\text{Hom}(\mathfrak{A}, \mathbf{K})</math>, 204<br/> <math>\text{Mng}_{\mathcal{L}}</math>, 204<br/> <math>X \setminus Y</math>, 146<br/> <math>\Delta(x)</math>, 182<br/> <math>\alpha</math>-neat-reduct, 183<br/> <math>c(\Gamma)</math>, 161<br/> <math>\models</math>-interpolation property, 214<br/> <math>\omega</math>, 150, 170<br/> <math>\omega</math>-complete, 58, 65–67, 74<br/> <math>\oplus</math>, 143<br/> <math>\subseteq_{\omega}</math>, 209<br/> <math>\rightarrow</math>-interpolation property, 214<br/> <math>\vdash</math> is given by <math>\langle Ax, Ru \rangle</math>, 194<br/> <math>\vdash_n</math>, 223<br/> <math>\vdash_{n,m}</math>, 223<br/> <math>m</math>-valued set, 270<br/> <math>n</math>-ary identity relation, 159<br/> <math>\mathbf{L}_{\kappa\omega}^n</math>, 236<br/> <math>\mathbf{L}_{\text{FOL}}</math>, 234<br/> <math>\mathbf{Nr}_{\alpha} \text{CA}_{\beta}</math>, 183<br/> <math>\mathbf{Nr}_n(\mathfrak{B})</math>, 170<br/> <math>\mathbf{Rd}_{sc}</math>, 184<br/> <math>\mathcal{L}_{3,2}^=</math>, 153<br/> <math>\mathcal{L}_{3,2}^{\neq}</math>, 149<br/> <math>\mathcal{L}_3^{xy}</math>, 150<br/> <math>\mathcal{P}(U)</math>, 138<br/> <math>\mathfrak{P}(U)</math>, 138<br/> <math>\mathfrak{R}f(U)</math>, 171<br/> <math>\mathfrak{R}(M \times M)</math>, 150<br/> <math>\mathfrak{R}(V)</math>, 141<br/> <math>\mathfrak{Nr}_{\alpha} \mathfrak{A}</math>, 182<br/> <math>\mathfrak{Rel}_{\omega}(U)</math>, 171<br/> <math>\mathfrak{Rel}_n(U)</math>, 160<br/> <math>\text{CA}_{\alpha}</math>, 182 </p> | <p> <math>\text{CA}_{\omega}</math>, 182<br/> <math>\text{Csf}_{\omega}</math>, 171<br/> <math>\text{Dc}_{\alpha}</math>, 182<br/> <math>\text{Dr}(R)</math>, 165<br/> <math>\text{Lf}_{\alpha}</math>, 182<br/> <math>\text{PA}_{\alpha}</math>, 185<br/> <math>\text{QPA}_{\alpha}</math>, 185<br/> <math>\text{RCA}_{\alpha}</math>, 182<br/> <math>\text{RCA}_{\omega}</math>, 171<br/> <math>\text{RPA}_{\alpha}</math>, 185<br/> <math>\text{RQPA}_{\alpha}</math>, 185<br/> <math>\text{RSC}_{\alpha}</math>, 184<br/> <math>\text{Rf}(U)</math>, 171<br/> <math>\text{SC}_{\alpha}</math>, 184<br/> <math>\text{p}_{01}(X)</math>, 185<br/> <math>s_j^i</math>, 165<br/> <math>m</math>-variable provable, 224<br/> <math>{}^{\omega}U</math>, 170<br/> <math>{}^nU</math>, 159<br/> <math>MV</math>-algebras, 274<br/> <math>\mathbf{Alg}(\mathcal{L})</math>, 203, 204<br/> <math>\mathbf{Alg}_m(\mathcal{L})</math>, 203, 204<br/> <math>\mathbf{Eq}(\text{BRA})</math>, 148<br/> <math>\mathbf{Eq}(\mathbf{K})</math>, 148<br/> <math>\mathbf{HK}</math>, 141<br/> <math>\mathbf{IK}</math>, 141<br/> <math>\mathbf{L}_n</math>, 222<br/> <math>\mathbf{L}'_n</math>, 220<br/> <math>\mathbf{PK}</math>, 141<br/> <math>\mathbf{RaCA}_n</math>, 165<br/> <math>\mathbf{Sir}(\mathbf{K})</math>, 142<br/> <math>\mathbf{SK}</math>, 141<br/> <math>\mathbf{UpK}</math>, 141<br/> <math>\text{RBM}</math>, 155<br/> <math>\text{ARA}</math>, 140<br/> <math>\text{BAO}</math>, 139<br/> <math>\text{BA}</math>, 138 </p> |
|---|--|

- BRA, 139
- BSR, 152
- $CA_n$ , 161
- $Cs_2$ , 159
- $Cs_n$ , 160
- Id, 153
- $Id_{ij}$ , 162
- $K_0$ -extensible, 213
- $Mod(E)$ , 140
- QPAs with equality, QPEAs, 185
- QPEAs, 185
- QRA, 158
- RA, 156
- $RCA_n$ , 160
- RRA, 153
- cBRA, 138
- setBRA, 147
  
- Abelian
  - group, 309
  - monoid, 309
- abstract algebra, 309
- absurd, 108
- Adams, E. W., 126
- adjoint pair, 319
- admissible rules, 193
- algebra of binary relations, BRA, 139
- algebra of finitary relations,  $Csf_\omega$ , 172
- algebraic completeness, 318
- algebraizable, 198
- algorithmic logic, 273
- alphabet, 189
- amalgamation property, 214
- analytic
  - calculus, 336
  - cut rule, 345
- Anderson, A. R., 260
- annotated logic program, 362
- annotated logic programming, 351
- antecedent, 335
- anti-link, 357
- Archimedean  $t$ -norm, 334
  
- Aristotelian logic, 249
- arity, 299
- arrow logic  $L_{REL}$ , 219
- assertion operator, 253
- atom, 305
- atomic, 56
- atomic formulas, 192
- atomic propositions, 192
- atomic statements, 122
- atomic substatement, 56
- atomic substatements, 72, 73
- autonomous, 126
- axiom, 310, 335
- axiomatizable, 311
  
- Balwin, J. F., 287
- Barcan formula, 10
- Barnes, R. F., 126
- base set, 160
- basic  $t$ -Norm Logic, 322
- basic pair, 76
- basic truth set, 76
- BDD, 366
- Beavers, G., 249
- Bellman, R. E., 286
- Belnap, N. D., 54, 59, 72, 260
- Bencivenga, E., 290
- Bendall, K., 86, 87, 126
- Bernays, P., 53
- Beth, E. W., 125
- Beth, E. W., 72, 122
- bilattice, 316
  - with negation, 316
- binary probability function, 102, 106, 109, 121
- binary probability functions, 126
- binary probability semantics, 126
- $L$ , 311
- BL-algebra, 320
- Blamey, S., 260
- Bochvar consequence relation, 258
- Bochvar, D. A., 252
- body, 359
- bold clause, 355

- Boolean, 127
  - algebra, 318
  - set lattice, 310
- Boolean algebra with operators, BAO, 139
- Boolean algebra, BA, 138
- Boolean semigroup, BSR, 152
- bound
  - renaming, 307
  - variable, 305
- bounded lattice, 310
- Brouwer fixed point theorem, 278
- Burali–Forti paradox, 121
- calculus, 193
- canonical
  - algebra, 318
  - function graph, 367
  - normal form, 367
- Cantor, 121
- Carnap, R., 77, 86, 106, 107, 124, 126
- certainty ordering, 315
- Chang, C. C., 264
- characteristic matrix, 255
- Church, A., 122, 253
- Cignoli, R., 264
- class of models of  $\Sigma$ ,  $\text{Mod}(\Sigma)$ , 191
- class of models,  $M_{\mathcal{L}}$ , 189
- classical propositional logic, 300
- classical sentential logic, 216
- clause, 305
- closed
  - branch, 339
  - tableau, 339
- closure operation, 139
- closure rule schema, 338
- combining logics, 202
- compact, 211
- complemented closure operation, 139
- complete
  - calculus, 311, 352
  - set of signs, 340
- completeness theorem, 61
- composition for  $n$ -ary relations, 159
- compositional, 195
- compositional meaning-function, 195
- comprehension axiom, 277
- conclusion, 311, 338
- concrete algebra of binary relations, cBRA, 138
- concrete Post algebra, 272
- conditional logic, 55
- conditional probability function, 106
- conjugates, 157
- conjunction, 300
- conjunctive normal form, 305
- connection method, 357
- connectives, 192, 299
- consequence compact, 211
- consequence relations, 255
- consistent, 67
- constraint programming, 351
- converse  $R^{-1}$ , 138
- cost function, 327
- countable model, 64
- countable submodels, 122
- covering, 345
- CPC**, 2, 7
- Curry, H. B., 278
- cut elimination, 36
- cylindric algebra of dimension  $n$ ,  $CA_n$ , 161
- cylindrifications,  $c_i$ , 162
- D*-interpretation, 60
- D*-consequence, 303
- D*-model, 303, 307
- D*-satisfiable, 302, 307
- D*-true, 307
- D*-valid, 303, 307
- Davidson, D., 125
- Davis–Putnam–Loveland procedure, 355
- de Finetti, B., 86, 289
- de Morgan algebra, 272
- decomposition, 366

- deduction term, 215
- deduction theorem, 24, 215, 312
- deductive logic, 198
- deductively closed, 58
- definability properties, 211
- definable functions, 264
- definable functions in  $L_m$ , 266
- degrees of error, 284
- denumerably infinite, 67
- derivable, 190
- derived connective, 196
- designated elements, 254
- designated truth value, 303
- designated value, 250
- determinism, 249
- diagonals, 162
- dilating, 158
- dimension-complemented, 183
- discriminant, 369
- discriminator term, 142
- discriminator variety, 147
- disjunction, 300
- dissolution rule, 356
- distribution, 306
  - function, 306
  - quantifier, 308
- distributive over, 139
- divisible, 319
- domain, 60, 306
- double rewrite, 67
- downset, 309
- dummy representation, 165
- Dunn, J. M., 54, 59, 72
  
- eigen-parameter, 12
- Ellis, B., 86
- empty signed clause, 350
- epimorphism, 213
- Epstein, G., 271
- equality logic, 235
- equational class, 140
- equational implication, 148
- equivalence of logics, 202
- equivalential, 200
  
- ESOP, 369
- essential function, 268
- exclusive-or-of-products expression,
  - see* ESOP
- explicit definition, 211
- extended
  - interpretation, 359
  - model, 359
- extension, 338
- extensional, 106
- extensions of  $L$ , 72
- external calculus, 313
- external connective, 253
  
- factored expression, 369
- Farey partition, 325
- feasible, 327
  - solution, 327
- Fenstad, J. E., 278
- fibring, 375
- Field, H. H., 126
- filter, 310
- filter-property, 196
- finitary, 258
- finitary logic of infinitary relations,
  - $L_\omega$ , 229
- finitary rule, 311
- finite axiomatizability, 264
- finite cardinals, 122
- finite model, 85
- finite-dimensional, 172
- finitely axiomatizable, 311
- finitely axiomatizable approximations, 156
- finitely complete, 209
- first degree entailments, 260
- first-order
  - formula, 305
  - language, 305
  - logic, 55, 306
  - matrix, 305
  - signature, 305
  - structure, 306
  - theory, 71

- valuation, 306
- first-order logic with  $n$  variables, 220
- first-order logic with  $n$  variables with substitutions,  $\mathbf{L}_n^{s=}$ ,  $\mathbf{L}_n^s$ , 229
- first-order logic with  $n$  variables, structural version  $\mathbf{L}_n$ , 222
- first-order logic, rank-free (typeless) version,  $\mathbf{L}_{FOL}$ , 234
- first-order logic, ranked version,  $\mathbf{L}_{FOL}^{\text{ranked}}$ , 231
- Fitch, F. B., 57
- foreign, 57
- formula algebra,  $\mathfrak{F}_{\mathcal{L}}$ , 192
- formula schema, 310
- formula-schemes,  $Fs$ , 193
- formula-variable, 193
- free
  - logic, 290
  - substitution, 307
  - term algebra, 309
  - variable, 305
- Frege, G., 53, 122, 124
- full  $\mathbf{Cs}_\omega$ , 171
- full cylindric set algebra, 159, 160
- full BRA, 147
- fully-fledged, 197
- fully-fledged general logic, 199
- function symbol, 305
- functional completeness, 252
- functionally complete, 264, 302
- functionally dense, 158
- future contingents, 283
- fuzzy logic, 279, 332
  - programming, 361
- fuzzy sets, 285
  
- G-algebra, 321
- Gabbay-style inference system, 210
- Gaifman, H., 86, 87, 127
- Garson, J. W., 126
- general logic, 198
- generalized extensionality, 251
  
- Gentzen calculus, 335
- Gentzen's sequent calculus, 35
- Gentzen, G., 56
- given by  $\langle P, Cn \rangle$ , 192
- Goddard, L., 287
- Gödel logic, 300
- Goguen, J. A., 286
- Goldfarb, W. D., 53, 60
- Gonseth, F., 251
- Gottlieb, D., 126
- Gottwald, S., 274
- graded truth, 312
- graph colorability problem, 317
- Grigolia, R. S., 262
- ground
  - substitution, 307
  - term, 305
- Gumb, R., 126, 127
  
- Höhle, U., 287
- Halldén, S., 287
- Harper, W. L., 126
- Hasenjaeger, G., 58
- Hauptsatz*, see cut elimination
- Hay, L., 277
- HCPC**, 3
- head, 359
- Henkin  $D$ -interpretation, 61
- Henkin equation, 168
- Henkin extension, 58, 59, 66, 68, 74
- Henkin model, 64, 67, 68, 70, 120
- Henkin models, 55, 61, 71, 122
- Henkin, L., 58, 61, 122, 125
- hierarchical normal form, 351
- higher-order logics, 55
- Hilbert calculus, 310, 311
- Hilbert, D., 124
- Hilbert–Frege style systems, 2
- Hilbert-style inference system, 193
- Hintikka set, 86
- Hintikka, J., 72, 79, 80, 85, 86
- Hodges, W., 60
- Horn formula, 351

- Huntington, E. V., 127  
hypersequent, 337
- ideal, 310  
idempotent, 309  
identity, 55  
if-then-else normal form, 366  
implicate  
  regular, 357  
implication, 300  
implicit definition, 211  
inference system, 190  
inference-rule, 193  
infinitary version  $\mathbf{L}_{\infty\omega}^n$  of the finite variable fragments, 236  
infinite-valued consequence, 263  
infinite-valued logic, 252  
infinite-valued matrix, 263  
infinitely extendible, 58, 61, 65–67  
information ordering, 315  
instance, 310  
instance of a formula-scheme, 193  
instance of an inference rule, 193  
institution, 202  
integer program, 327  
intended model, 71, 72, 122  
internal calculus, 313  
interpolation properties, 214  
interpretability between logics, 202  
intuitionistic logic, 55  
inverse element, 309  
involuted semigroup, 140  
involution, 309  
isomorphisms, 75
- Jeffrey, R. C., 86, 289  
Jeffreys, H., 107  
Jennings, R. E., 127  
join, 310  
join-irreducible, 310
- Kearns, J. T., 125
- Keynes, J. M., 107  
Kleene consequence relation, 258  
Kleene, S. C., 253  
Klement, E. P., 287  
knowledge  
  diamond lattice, 315  
  ordering, 315  
Kolmogorov, A. N., 86, 87, 106  
Kripke, S., 125
- $\mathcal{L}$  has a deduction theorem, 215  
 $\mathcal{L}$  has the (syntactic) substitution property, 196  
 $\mathcal{L}$  has the filter-property, 196  
 $\mathcal{L}$  is algebraizable, 198  
 $\mathcal{L}$  is compositional, 195  
 $\mathcal{L}$  is structural, 198  
 $\mathcal{L}$  has connectives, 192  
 $\mathbf{L}$  has connectives, 198  
 $\mathbf{L}$  has the filter-property, 199  
 $\mathbf{L}$  has the semantical substitution property, 199  
 $\mathbf{L}$  has the substitution property, 199  
 $\mathbf{L}$  is an algebraizable general logic, 199  
 $\mathbf{L}$  is compositional, 198  
 $\ell$ -group, 309  
Löwenheim, L., 70, 122  
Löwenheim–Skolem Theorem, 70, 71  
lattice, 310  
lattice-regular  
  binary resolution, 355  
  positive unit resolution, 361  
  reduction, 355  
  unit reduction, 361  
law of excluded middle, 250  
Leblanc, H., 54, 56–58, 61, 62, 67, 85–87, 108, 125–127  
Levin, M. E., 106  
Lewis, C. I., 258  
liar paradox, 318  
Lindenbaum, A., 258

- linear
  - inequation, 327
  - program, 327
  - term, 327
- local Beth definability property, 212
- local search, 357
- locally finite, 183
- locally finite lattice, 373
- logic, 189
- logic morphism, 201
- logic of significance, 279, 287
- logic with three individual variables, 153
- logical
  - constant, 299
  - rule, 335
- logical connectives  $Cn$ , 192
- logical entailment, 62, 69, 102, 120, 122
- logical paradoxes, 252
- logical truth, 62, 69, 102, 120, 122
- logically entailed by  $S$  in the model-set sense, 80
- logically entailed by  $S$  in the probabilistic sense, 88
- logically entailed by  $S$  in the standard sense, 61
- logically entailed by  $S$  in the substitutional sense, 69
- logically entailed by  $S$  in the truth-set sense, 77
- logically entailed by  $S$  in the truth-value sense, 73, 76
- logically equivalent, 107
- logically false, 107
- logically true, 72
- logically true in the model-set sense, 80
- logically true in the probabilistic sense, 88
- logically true in the standard sense, 60, 72
- logically true in the substitutional sense, 69
- logically true in the truth-set sense, 77
- logically true in the truth-value sense, 73, 76
- logically true in the truth-value sense, 73, 76
- Los, J., 258
- Lukasiewicz, J., 249
- Lukasiewicz logic, 300
- Lyndon algebras, 158
- $MV_n$ -algebras, 274
- Mamdani rule, 333
- many-sorted model, 144
- many-valued
  - binary resolution, 353
  - decision diagram, *see* MDD
  - hyperresolution, 353
  - logics, 55
  - merging, 353
  - predicate logic, 274
- Marcus, R., 125
- matrix, 254
- maximally complete, 66
- maximally consistent, 58, 65, 67, 74
- maxK, 213
- McArthur, R. P., 126
- McCarthy, T., 126
- McNaughton, R., 268
- MDD, 367
- meaning algebra of  $\mathfrak{M}$ , 203
- meaning of  $\varphi$  in  $\mathfrak{M}$ , 194
- meaning-function for  $\mathcal{L}$ , 194
- meaning-homomorphism, 204
- measure function, 106
- meet, 310
- meet-irreducible, 310
- Meredith, C. A., 264
- minimal extended model, 360
- minimal logic, 12, 258
- minimal predicate logic with identity, 21
- minterm, 369
- MIP-representable, 328
  - logic, 328



- mixed integer program, 327
- modal logic S5, 218
- modal logics, 55
- model associate, 65–68
- model counterpart, 74
- model set, 120
- model sets, 79, 85, 123, 126
- model structures on commutative monoids, 279
- model-set semantics, 83, 123
- models, 60
- modus ponens, 311
- monadic logic, 235
- monoidal logic, 322
- monosigned
  - binary resolution, 354
  - formula, 339
- Moore, G. E., 124
- Morgan, C. G., 126, 127
- Mostowski, A., 288
- multiplicity, 326
- Mundici, D., 249, 264
- MV-algebra, 321
  
- natural deduction, 15
- neat reduct functor, 171
- neat-reduct, 170
- negation, 300
- negative literal, 351
- nilpotent
  - $t$ -norm, 309
- NIQC**, 15
- non tertium datur, 312
- non-clausal resolution, 356
- non-deterministic semantics, 376
- non-Euclidean geometry, 249
- non-finitizability theorems, 178
- non-standard, 120
- nonrepresentable, 158
- normal Boolean algebra with operators, BAO O, 139
- normalisation theorem, 27
- Nute, D., 127
  
- O’Hearn, P., 249
- $o$ -group, 309
- object variable, 305
- objectual interpretation, 60
- occurrences
  - closed, 16
- occurrences
  - discharged, 16
- ON-set, 324
- optimal solution, 327
- order filter, 310
- order ideal, 310
- ordered Abelian group, 282
- ordinal sum, 334
  
- Panti, G., 264
- paraconsistent logic, 301, 304
- paraconsistent logic program, 362
- paradox of the heap, 286
- paradoxes of set theory, 277
- Parsons, C., 126
- partial functions, 107
- partial logic, 260
- partial truth, 312
- partition, 345
- partitioning
  - DNF representation, 346
  - tableau rule, 346
- patchwork property of models, 212
- pattern recognition, 287
- Pavelka complete, 313
- Pavelka, J., 287
- Peirce, C. S., 249
- $\Pi$ -algebra, 321
- polarity, 351
- polyadic algebra,  $RPA_\alpha, PA_\alpha$ , 185
- Popper, K., 54, 86, 87, 107, 108, 126, 127
- positive integers, 71
- positive literal, 351
- positive polarity Reed-Muller, 370
- possible-translation semantics, 376
- Post algebras, 270
- Post logic, 301

- Post, E., 249, 252
- power
  - interpretation, 339
  - matrix, 339
  - operation, 339
- powerset of  $U$ ,  $\mathcal{P}(U)$ , 138
- precomplete, 266
- predicate symbol, 305
- prelinear residuated lattices, 320
- prelinearity, 319
- premiss, 311, 338, 352
- prime filter, 310
- prime ideal, 310
- prime implicate
  - regular, 357
- probabilistic semantics, 53, 54, 86, 123
- probability associate, 121
- probability associate of  $S$ , 100
- probability function, 86, 106
- probability theory, 102, 105
- product, 300
  - logic, 300
- programming languages, 273
- proof in  $L$ , 57
- propositional
  - formula, 299
  - interpretation, 302
  - language, 299
  - logic, 300
  - matrix, 300
  - signature, 299
  - valuation, 302
- protoalgebraic, 200
- provability
  - degree, 312
  - relation, 311
- provability relation, 189, 336, 338, 352
- provable, 57, 190, 311
- provable from  $\Sigma$ , 191
- provable with  $m$  variables, 224
- pseudo-axiomatisable, 143
- QBL-algebra, 320
- QMV-algebra, 321
- quantifications, 56
- quantifier, 305
- quantifierless, 56
- quantifierless statement, 72
- quasi-equation, 206, 209
- quasi-polyadic algebra,  $\text{QPA}_\alpha$ , 185
- quasi-projective RA,  $\text{QRA}$ , 158
- quasi-variety, 148, 209
- Quine, W. V. O., 57, 77, 125
- Rényi, A., 107
- Ramsey, F. P., 124, 289
- Rasiowa, H., 274
- rational belief, 106
- rational Pavelka logic, 313, 323, 329
- Rautenberg, W., 249, 264
- reduction rule, 353
- refuted, 352
- regular, 172
  - binary resolution, 354
  - formula, 353
  - GSAT, 358
  - Horn formula, 352
  - logic, 348
  - negative hyperresolution, 354
  - positive unit resolution, 361
  - resolution, 354
  - sign, 340
  - unit resolution, 361
- Reichenbach, R., 107
- relation algebra,  $\text{RA}$ , 156
- relation-algebra reduct  $\mathfrak{Ra}\mathfrak{A}$ , 165
- relation-composition, 138
- relational proof system, 357
- relevance logic, 260
- representable Boolean monoid,  $\text{RBM}$ , 155
- representable cylindric algebra of  $n$ -ary relations,  $\text{RCA}_n$ , 160
- representable relation algebra,  $\text{RRA}$ , 153

- representation, 178
- Rescher, N., 262
- residuated lattice, 319
- residuation, 300
- residue, 353
- resolution
  - calculus, 352
  - rule, 352
- resolvent, 352
- ROBDD, 367
- Robinson, A., 72, 125
- Rosenberg, I., 265
- Rosser, J. B., 57, 262
- Routley, R., 280, 287
- rule, 358
  - schema, 310
- Russell's Paradox, 253, 278
- Russell, B., 122, 124
  
- safe, 306
- Salwicki, A., 273
- satisfiability compact, 211
- Scarpellini, B., 277
- Schütte, K., 72, 122, 126
- Schauder hat, 326
- Schotch, P. K., 127
- scope, 305
- Scott, D., 283
- Seager, W., 127
- semantic logic, 198
- semantic tableaux, 44
- semantical consequence, 191
- semantical logic, 198
- semantical part, 190
- sequent, 31, 255, 335
  - calculus, 335
  - proof, 336
  - proof tree, 336
- sequent calculi, 33
- set literal, 370
- set of formulas of  $\mathcal{L}, F_{\mathcal{L}}$ , 189
- set theory in many-valued logic, 277
- set-assignment semantics, 376
  
- Shannon
  - expansion, 366
  - tree, *see* BDD
- Sheffer functions, 265
- Shoenfield, J. R., 72, 125
- Shoemith, D. J., 258
- signature morphism, 201
- signed
  - atom, 339
  - clause, 350
  - CNF
    - formula, 350
    - representation, 341
  - DNF representation, 341
  - fact, 359
  - formula, 338, 339
  - NNF formula, 350
  - query, 359
  - rule, 358
  - sequent, 338
  - SLD resolution, 360
  - unit
    - reduction, 360
    - resolution, 360
- signed formula
  - logic program, 358
- singularly probability functions, 120
- singularly probability functions, 126
- Skolem Paradox, 70
- Skolem, T., 70, 122, 278
- Smiley, T. J., 258
- Smullyan, R. M., 58
- society semantics, 376
- SOP, 369
- Sorites paradox, 317
- sound
  - calculus, 311, 352
- soundness, 342
- soundness theorem, 61
- splitting, 158
- Stalnaker, R., 126, 127
- standard account of logical truth, 66

- standard semantics, 53–55, 60, 62, 121
- state-description, 76, 77
- statements, 55
- strict  $t$ -norm, 334
- strong Beth definability, 212
- Strong Completeness Theorem, 84, 102
- strong connectives, 254
- strong equivalence, 304
- Strong Soundness Theorem, 83, 99, 110
- strongly complete, 209, 311
- strongly complete and sound, 209
- strongly sound, 209
- strongly sound and complete, 118
- strongly sound rules, 193
- structural, 198
- structural consequence relation, 256
- structural rule, 335
- structure preserving CNF translation, 349
- structures, 60
- subdirect product, 142
- subdirectly irreducible, 142
- subformula principle, 336
- subjunctive probability, 289
- submatrix, 254
- substatements, 56
- substitution, 196, 307
- substitution instance, 56, 73, 80
- substitution instances, 61
- substitution property, 196
- Substitution Theorem, 61, 65
- substitution-cylindric algebra,  $SCA_\alpha$ , 184
- substitutional, 120, 196
- substitutional semantics, 53, 61, 64, 69, 71, 122, 124, 125
- substitutionally true
  - trues, 69
- succedent, 335
- sum, 300
- sum-of-products expression, *see* SOP
- superamalgamation property, 215
- supervaluations, 290
- Suszko, R., 258
- switching term, 143
- symmetric difference,  $\oplus$ , 143
- syntactic logic, 198
- syntactical part, 190
- Ślupecki, J., 266, 268
  
- $t$ -algebra, 320
- $t$ -norm logic, 301
- tableau
  - calculus, 338
  - proof, 339
  - proof tree, 338
  - rule schema, 338
- Tarski, A., 58, 60, 125, 250
- TAS method, 357
- tense logics, 55
- term, 305
- term extension, 54, 59, 62, 68, 72, 73
- term rewrites, 54
- termination rule, 352
- termless, 56
- tertium non datur*, 24
- tertium non datur, 321
- The Deduction Theorem, 4
- theory, 58, 122
- theory of  $K$ , 191
- thinning, 39
- Traczyk, T., 271
- triangle rule, 157
- triangular norm, 301
- truth
  - degree, 313
  - ordering, 315
  - value, 300
- truth set, 78, 120
- truth value assignment, 73
- truth value gaps, 290
- truth-set semantics, 76, 123
- truth-value, 72

- truth-value assignment, 72, 73, 76, 77, 79, 80, 102, 120, 123
- truth-value associate, 74, 82, 120
- truth-value counterpart, 74
- truth-value function, 76, 78, 79, 86, 87, 102, 105, 106, 120
- truth-value semantics, 53, 55, 72, 75, 76, 102, 123, 125
- truth-value theory, 102, 105
- truth-values, 123
- Turquette, A. R., 262, 277
- two-valued, 103
- two-valued probability functions, 106
  
- Ulam's guessing game with lies, 285
- Ulam, S., 284
- uncountable, 122
- uniform, 256
- unit, 139
- unit resolution, 355
- universal literal, 370
- universal quantification, 61, 80
- upset, 310
- Urquhart, A., 264
  
- vague predicates, 286
- valid, 193
- valid formulas, 191
- validity problem, 208
- validity relation, 190
- Van Fraassen, B., 127
- van Fraassen, B., 126
- van Fraassen, B. C., 106, 290
- variable assignment, 306
- variety, 140
- Visser, A., 260
- vocabulary, 192
- Von Wright, G. R., 107
  
- Wójcicki, R., 258
- Wade, C. I., 273
- Wajsberg, M., 262
  
- weak Beth definability property, 212
- weak connectives, 254
- weakly algebraizable, 200
- weakly complete, 209
- weakly sound, 209
- Wellington, 108
- White, R. B., 278
- Whitehead, A. N., 122, 124
- Wisdom, W. A., 61, 62, 85, 86
- Wittgenstein, L., 53, 124
- Wolf, R. G., 262
- Wroński, A., 249, 264
  
- Zadeh, L. A., 285
- Zermelo–Fraenkel, 121

# Handbook of Philosophical Logic

2nd Edition

Volume 3

edited by Dov M. Gabbay and F. Guentner



## CONTENTS

Editorial Preface	vii
<b>Dov M. Gabbay</b>	
Basic Modal Logic	1
<b>R. A. Bull and K. Segerberg</b>	
Advanced Modal Logic	83
<b>M. Zakharyashev, F. Wolter and A. Chagrov</b>	
Quantification in Modal Logic	267
<b>J. Garson</b>	
Correspondence Theory	325
<b>J. van Benthem</b>	
Index	409





## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good.!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London

Logic	IT			
	Natural language processing	Program control specification, verification, concurrency	Artificial intelligence	Logic programming
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs

<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical frag- ments</b>	Basic back- ground lan- guage	Program syn- thesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely use- ful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calcu- lus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syn- tax	Modules. Combining languages	Logics of space and time	Combining fea- tures
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game seman- tics gaining ground		
<b>Object level/ metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action- revision mod- els</b>			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model





## BASIC MODAL LOGIC

### Historical Part

#### 1 HISTORICAL OVERVIEW

It is popular practice to borrow metaphors between different fields of thought. When it comes to evaluating modal logic it is tempting to borrow from the anthropologists who seem to agree that our civilisation has lived through two great waves of change in the past, the Agricultural Revolution and the Industrial Revolution. Where we stand today, where the world is going, is difficult to say. If there is a deeper pattern fitting all that is happening today, then many of us do not see it. All we know, really, is that history is pushing on.

The history of modal logic can be written in similar terms, if on a less global scale. Already from the beginning—corresponding to the stage of hunter-gatherer cultures in anthropology—insights into the logic of modality has been gathered, by Aristotle, the Megarians, the Stoics, the medievals, and others. But systematic work only began when pioneers found or forged tools that enabled the to plough and cultivate where their predecessors had had to be content to forage. This was the First Wave, and as with agriculture it started in several places, more or less independently: C. I. Lewis, Jan Łukasiewicz, Rudolf Carnap. These cultures grew slowly, from early this century till the end of the sixth decade, a period of more than 50 years. Then something happened that can well be described as a Second Wave. What brought it out spectacularly was the achievements of the teenage genius of Saul Kripke, but he was not alone, more strictly speaking the first of his kind: the names of Arthur Prior, Stig Kanger, and Jaakko Hintikka must also be mentioned, perhaps also those of J. C. C. McKinsey and Alfred Tarski. Now modal logic became an industry. In the quarter of a century that has passed since, this industry has seen steady growth and handsome returns on invested capital.

Where we stand today is difficult to say. Is the picture beginning to break up, or is it just the contemporary observer's perennial problem of putting his own time into perspective? For a long while one attraction of modal logic was that it was, comparatively speaking, so easy to do—now it is becoming as difficult as the more mature branches of logic. And the sheer bulk of published material is making it difficult to survey. But there is also the increasing differentiation of interests and the subsequent tendency

towards fragmentation. In addition to more traditional pursuits we are now seeing phenomena as diverse as the application of modal predicate logic to philosophical problems at a new level of sophistication (Fine [1977; 1977a; 1980]), the analysis of conditionals started by Stalnaker [1968], Lewis [1973], the generalisation of model theory with modal notions (Mortimer [1974], Bowen [1978]), in-depth studies of the so-called provability interpretation (see Boolos [1979]; see also Craig Smoryński's Chapter in this *Handbook*), the advent of dynamic logic (see Pratt [1980] and David Harel's Chapter in this *Handbook*) and Montague grammar (see Montague [1974]).

This is not the place to go deeply into the history of modal logic, even though we will say something about it in the next few sections. A reader who would like to know more about the beginnings of the discipline is referred to Prior [1962], Kneale and Kneale [1962], and Lemmon [1977]. For the discipline itself, as distinct from its history, the reader may consult a number of textbooks or monographs, from E. J. Lemmon's and Dana Scott's fragment Lemmon [1977], and Hughes and Cresswell [1996]. Schütte [1968], Makinson [1971], Segerberg [1971], Snyder [1971], Zeman [1973], and Gabbay [1976] to the recent and very readable Rautenberg [1979] and Chellas [1980]. Notable journal collections of papers on modal logic include 'Proceedings of a colloquium on modal and many-valued logics' (*Acta Philosophica Fennica*, **16**, 1963), 'In memory of Arthur Prior' (*Theoria*, **36**, 1970), and 'Trends in modal logic' (*Studia Logica*, **39**, 1980). Good bibliographies of early work are found in Feys [1965], Hughes and Cresswell [1996] and Zeman [1973]. Among survey papers from the last few years we recommend Montague [1968], Belnap [1981], Bull [1982; 1983], and Føllesdal [1989].

All writing of history is to some extent arbitrary. The historian, in his quest for order, imposes structure. A favourite stratagem is the imposition of  $n$ -chotomies. As long as the arbitrary element is recognised, the procedure seems perfectly legitimate. This admitted we should like to impose a trichotomy on early modal logic: modern modal logic derives from three fountain-heads which may be classified according to their relation to semantics. The *syntactic* tradition is the oldest and is characterised by the lack of explicit semantics. Then we have the *algebraic* tradition with a semantics of sorts in algebraic terms. Finally there is the *model theoretic* tradition, the youngest one, whose semantics is in terms of models. Possible worlds semantics is the dominating kind of model theoretic semantics, perhaps even, if we take advantage of the vagueness of this term and stretch it a little, the only kind. In the next few sections we propose to give a brief account of each of the three traditions.

## 2 THE SYNTACTIC TRADITION

Modern modal logic began in 1912 when C. I. Lewis filed a complaint in *Mind* to the effect that classical logic fails to provide a satisfactory analysis of implication, ‘the ordinary “implies” of ordinary valid inference’, [Lewis, 1912]. Roughly it is the paradoxes of material implication that Lewis worries about, but his subtle argument goes beyond the vulgar objections, implication is not the only connective that worries him. In fact, his very first analysis concerns disjunction. Consider, he says the following two propositions:

1. Either Caesar died, or the moon is made of green cheese.
2. Either Matilda does not love me, or I am beloved.

If we disregard the complication that there is also an exclusive reading of ‘or’, classical logic will consider that both these propositions are of the form

- (i)  $A \vee B$ .

Yet, Lewis argues, there are more important differences between the two. For example, we know that (1) is true since we know that, as it happens, Caesar is dead, but we know that (2) is true without knowing which of the disjuncts is true. Thus (2) exhibits a ‘purely logical or formal character’ and an ‘independence of facts’ that is lacking in (1).

This much all can agree. But disagreement arises over how to account for the difference between (1) and (2). One possibility would be to hold that while both (1) and (2) are of the same form, viz. (i) they differ in that only (2) satisfies the further condition

- (ii)  $\vdash A \vee B$ ,

where the turnstile  $\vdash$  stands for assertability or provability in some suitable system. But Lewis embraces another possibility. The difference between (1) and (2), he feels, is a difference in meaning. More specifically, he feels that there is a connection between the disjuncts of (2) which is part of the meaning of (2). On this view, the ‘or’ of (1) and the ‘or’ of (2) are different kinds of disjunction, and Lewis proposes to call the former *extensional* and the latter *intensional*. While extensional disjunction is rendered by the traditional, truth-value functional operator  $\vee$ , a novel sort of operator is needed to render intensional disjunction. Lewis himself never introduced a symbol for it, but E. M. Curley, in a recent historical study, uses the symbol  $\boxdot$  [Curley, 1975]. Thus, while (1) is of the form (i), we may say that, according to Lewis, (2) is of the form

- (iii)  $A \boxdot B$ .

The same problem also concerns other connectives. In the case of implication there is, according to Lewis, an extensional kind which is adequately rendered by the ‘arrow’,  $\rightarrow$ , the material implication of ordinary truth-value functional logic. But there is also an intensional kind of implication, called *strict implication* by Lewis, and for this he introduces a new symbol, the ‘fish-hook’,  $\rightarrow$ . The latter is not found, nor definable, in classical logic, and so Lewis proposes to develop a calculus of strict implication.

Thus there is a triad corresponding to (1)–(iii), viz.,

$$(i') \quad A \rightarrow B,$$

$$(ii') \quad \vdash A \rightarrow B,$$

$$(iii') \quad A \rightarrow B.$$

(The condition  $A \vdash B$  is logically equivalent to (ii'); Lewis would also have regarded the condition  $\vdash A \rightarrow B$  as equivalent to (ii').) The reader should notice the difference in theoretical status between  $\rightarrow$  and  $\rightarrow$  on the one hand, and  $\vdash$  on the other. In both cases the first two are, or name, operators belonging to the object language, while the turnstile is part of the metalanguage, standing for provability or deducibility. (Provability may of course be seen as a special case of deducibility, viz. deducibility from the empty set of premises.)

Evidently the crucial question is whether the logical difference between (1) and (2) should be expressed in the object language or not—is it a feature *about* logic or *in* logic? Gerhard Gentzen is often regarded as having opted for the former alternative (although see [Shoemith and Smiley, 1978, p. 33f] concerning the historicity of this view). It is hard to say whether Lewis was aware that there was a choice. However, looking back on his work we must represent him as having favoured (iii) over (ii) and (iii') over (ii') as the logical form of certain propositions. he has been much criticised for this. It has been maintained that his whole enterprise rests on a violation of the use/mention distinction and is hopelessly confused. this is not the place to go into that discussion, all we can do is to refer the reader to [Scott, 1971] which contains what is probably the deepest discussion of this matter and certainly the most constructive one.

The method chosen by Lewis in his search for a calculus of strict implication was the axiomatic one. Lewis' intuitive understanding of logical necessity, logical possibility and related notions was of course (at least) as good as any man's, but he never tried to give it direct systematic expression; what there is, is what is implicit in the axiom systems, plus scattered informal remarks. In other words, there is no formal semantics in Lewis' work; semantics is left at an informal level. In mathematics, there is an important and time-honoured way to proceed, ultimately going back on Euclid. In the case of logic the method may be described as follows. A formal language

is defined. Formulas from this language are understood to be meaningful. A number of them are somehow selected for testing against one's intuition. Some are accepted as valid, some are rejected as nonvalid, some may be difficult to decide. The valid ones one tries to axiomatise so as to give a finite description of an infinite scene. In Lewis' case, the first effort was presented in [Lewis, 1918], a calculus which has since become known as the Survey System. However, if your semantics is only intuitive, as Lewis' was, and consequently vague, then you have a completeness problem: even if you are satisfied that the theses of your system are acceptable, how do you know that your axiom system captures as theses all the formulas that you would find acceptable? The answer is that you do not, and it did not take long for other systems to emerge with, apparently, as good a claim as the Survey System to the title conferred upon it in [Lewis, 1918] as *the* System of Strict Implication. In [Lewis and Langford, 1959] several more were defined and others hinted at. Here Lewis himself defined five systems called **S1**, **S2**, **S3**, **S4**, and **S5**, the survey system coinciding with **S3**. Later **S6** was introduced by Miss Alban and **S7** by Halldén, but in effect there were contemplated already by Lewis [Alban, 1943; Halldén, 1949]. The series of **S**-systems has been extended even further, but those mentioned are the principal ones.

Of modal logicians working in the same vein as Lewis, Oskar Becker is remembered for his early treatise [Becker, 1930], but perhaps it is G. H. Von Wright who should be named the second most important author in the syntactic tradition. In his influential monograph [von Wright, 1951] he remarks that, strictly speaking, modal logic is the logic of the modes of being. In this work and the related paper [von Wright, 1951a], Von Wright sets out to explore modal logic in a wider sense, the logic of the modes of knowledge, belief, norms and similar concepts; this wider sense of the term has since gained currency. These two works marked the beginning of much work in epistemic, doxastic, and deontic logic. Some studies of the same kind had already been published, such as [Mally, 1926] and [Hofstadter and McKinsey, 1955] (see [Follesdal and Hilpinen, 1971] or Von Wright [1968; 1981] for more of the prehistory of deontic logic), but Von Wright's work becomes seminal, especially in deontic logic. (For epistemic and doxastic logic the real trigger was a book written some ten years later by Von Wright's one time student Jaakko Hintikka, but this work [Hintikka, 1962] was written in what we call the model theoretic tradition and so does not belong in this section.)

There are two other subtraditions that should be mentioned under the present heading. One is the development of entailment and relevance logic associated with the names of Alan Ross Anderson and Nuel D. Belnap. This movement concentrated on C. I. Lewis' concern to develop a logic of strict implication, that is, to give a syntactic characterisation of 'the ordinary "implies" of ordinary valid inference'. Early contributions in the axiomatic style were given by [Church, 1951a] and [Ackerman, 1956], but it was only

with Anderson and Belnap and their many students that the project got off the ground. Algebraic and model theoretic semantics came later to this kind of logic than to modal logic, and it is perhaps fair to say that the efforts towards finding an explicit semantics have led to results that are less natural than in modal logic. This may have to do with the fact that while model logicians aim at *improving* classical logic, entailment/relevance logicians wish to *replace* it. Students interested in this subtradition will find the powerful tome [Anderson and Belnap, 1975] a rich source of information. (Cf. also Dunn, in a later volume of this *Handbook*.)

The other subtradition that should be mentioned is that of proof theory. Gentzen methods have never really flourished in modal logic, but some work has been done, mostly on sequent formulations. Early references are [Curry, 1950; Ridder, 1955; Kanger, 1957; Ohnishi and Matsumoto, 1957/59]. A monograph in this tradition is [Zeman, 1973]. In the field of natural deduction [Fitch, 1952] would seem to be the pioneer with [Prawitz, 1965] the classical reference. The recent interest in the provability interpretation of modal logic has spurred renewed interest in the proof theory of particular systems (for example [Boolos, 1979; Leivant, 1981]). In Section 9 we return to this topic.

Finally, let it be remarked that the syntactic tradition in Lewis' spirit is by no means dead. For a recent declaration of allegiance to it by a distinguished logician, see [Grzegorzczak, 1981].

### 3 THE ALGEBRAIC TRADITION

That classical logic is truth-functional is enormously impressive! As shown by the existence of intuitionistic and other dissenting logics, it is by no means self-evident that it should be possible to understand the usual propositional operators in terms of simple truth-conditions (the familiar truth-tables). But given the success of classical logic it is natural to ask if the same treatment can be extended to other operators of interest, for example, modal ones. It is immediately clear that such an extension is not straight-forward, if it exists at all. There are four unary truth-functions (identity, negation, tautology, and contradiction), so if necessity or possibility is to be truth-functional, it would have to be one of them, which is absurd.

But if one insists, nevertheless, that it must be possible to give a truth-functional analysis of 'necessary' and 'possible'? Bright idea: perhaps there are more truth-values than the ordinary two—three, say. This idea occurred to Jan Łukasiewicz around 1918. His first effort was to supplement the ordinary truth-values 1 (truth) and 0 (falsity) with a third truth-value  $\frac{1}{2}$  (possibility (of some kind)). his new truth-tables were as follows:

$\wedge$	1	$\frac{1}{2}$	0	$\vee$	1	<i>half</i>	0	$\rightarrow$	1	$\frac{1}{2}$	0
1	1	$\frac{1}{2}$	0	1	1	1	1	1	1	$\frac{1}{2}$	0
$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$	1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1	1	$\frac{1}{2}$
0	0	0	0	0	1	$\frac{1}{2}$	0	0	1	1	1
$\neg$				$\square$				$\diamond$			
1	0				1	1				1	1
$\frac{1}{2}$	$\frac{1}{2}$				$\frac{1}{2}$	0				$\frac{1}{2}$	1
0	1				0	0				0	0

With 1 singled out as the sole designated truth value, the concept of validity is clear: a formula is valid if and only if it takes the value 1 under all (three-valued) truth-value assignments to its propositional letters. Let the resulting logic be called  $L_3$ . It is an immediate corollary that  $L_3$  is a subsystem of the classical propositional calculus; for if everything to do with the new truth-value  $\frac{1}{2}$  is deleted from the truth-tables, then we get the old, classical ones back.

Exactly what sort of possibility would  $\frac{1}{2}$  represent? the inspiration for his new logic Lukasiewicz had got from Aristotle's discussion of the theoretical status of propositions concerning the future. It is an interesting suggestion that a new truth-value is needed to analyse propositions of type 'there will be a sea-battle tomorrow'; for it might be held that there are points in time when such propositions are meaningful, yet neither true nor false. In other words, if one is not a determinist—and Lukasiewicz definitely was not one—then one will agree that there spare propositions  $P$  such that, today,  $P$  is possible and also  $\neg P$  is possible; that is, that both  $\diamond P$  and  $\diamond \neg P$  are true. This is in agreement with Łukasiewicz' matrix, for if  $P$  has value  $\frac{1}{2}$ , then  $\diamond P$  and  $\diamond \neg P$  take the value 1. So far, so good, but here a difficulty lurks. For under the matrix  $\diamond(P \wedge \neg P)$  gets the value 1 which is absurd intuitively: whatever the future may bring, it will not be both a sea-battle and not a sea-battle tomorrow. The counter-example is flagrant, and it is interesting that Łukasiewicz was not moved by it. What is at issue is evidently whether one can accept a modal logic which validates all instances of the type

$$\diamond A \wedge \diamond B \rightarrow \diamond(A \wedge B).$$

Our counter-example would appear to settle this question in the negative—cf. [Lewis and Langford, 1959, p. 167]—but Łukasiewicz was not impressed. In a paper published only a few years before his death he states that he cannot find any example that refutes the schema in question: 'on the contrary, all seem to support its correctness' [Łukasiewicz, 1953]. He goes on to intimate that when people disagree over questions of this sort, they have different concepts of necessity and possibility in mind.



Once invented, this game admits of endless variation. Even among three-valued logics,  $L_3$  is not the only possibility, and there is literally no end to how many truth-values you may introduce. Łukasiewicz himself extended his ideas first to  $n$ -valued logic, for any finite  $n$ , and then to infinitely-valued logic, where infinite could mean either denumerably infinite or even non-denumerably infinite. In this way the notion of matrix was developed. ([Malinowski, 1977] is a compact and informative reference on Łukasiewicz and his work. For Łukasiewicz's own papers non-Polish speaking readers are referred to the collections [Łukasiewicz, 1970] and [McCall, 1967].)

A *matrix* is given if you have (i) a set of objects, called *truth-values*, (ii) a subset of these, called the *designated* truth-values, and (iii) for every  $n$ -ary propositional operator  $\star$  in your object language, a truth-table for  $\star$  (essentially, an  $n$ -place function from truth-values to truth-values). In tuple talk, if  $\star_0, \dots, \star_{k-1}$  are all your propositional operators, the matrix can be thought of as a  $(k+2)$ -tuple  $\langle A, D, \mathfrak{M}(\star_0), \dots, \mathfrak{M}(\star_{k-1}) \rangle$ , where  $A$  is a non-empty set,  $D$  a non-empty subset of  $A$ , and, for each  $i < k$ ,  $\mathfrak{M}(\star_i)$  is a function from the Cartesian product  $A^{n_i}$  to  $A$ , where  $n_i$  is the arity of  $\star_i$ . It is easy to see how this can be generalised to any number of operators.

Opinions may be divided over what philosophical importance to attach to the logics that Łukasiewicz introduce. However, there can be no doubt that he started or tied in with a line of development which is of great mathematical importance. The matrices that he invented became generalised in two steps. The first one seems like a mere change of terminology: the introduction of the concept of an *algebra* as a tuple  $\langle A, f_0, \dots, f_{k-1} \rangle$ , where  $A$  is a non-empty set and  $f_0, \dots, f_{k-1}$  are operations on  $A$ ; that is, for each  $i < k$  there is a non-negative number  $n_i$  such that  $f_i$  is a function from  $A^{n_i}$  to  $A$ . As before, the generalisation to infinitely many functions is obvious. The connection with the concept of matrix is patent. Roughly speaking, it is only the set of designated elements that has been omitted; and as far as logic is concerned, that concept is needed for the definition of validity, not for the assignment of values of  $A$  to formulas. The most important thing about the new definition of algebra is perhaps that it encourages the study of these structures independently of their connection with logic.

The second step of generalisation was to consider *classes of algebras* rather than one matrix or algebra at the time. Thus, whereas at first algebraic structures (matrices) were introduced in order to study logic, later on logic was used to study algebra. The person who more than anyone deserves credit for this whole development is Alfred Tarski, a student and collaborator of Łukasiewicz. Some papers by Tarski written jointly with J. C. C. McKinsey or Bjarni Jónsson rank with the most important in the history of modal logic.

Among early results stemming from the algebraic tradition are that Lewis' five systems are distinct [Parry, 1934]; the analysis of **S2** and **S4** along with a proof that they are decidable [McKinsey, 1941]; that no logic between **S1**

and **S5**, inclusively, can be viewed as an  $n$ -valued logic, for any finite  $n$  [Dugundj, 1940]; that even though **S5** is not a finitely-valued logic, all its proper extensions are [Scroggs, 1951].

It does not seem as if anyone had ever worked out exactly what the relation is between abstract algebras and the intended applications. But the idea must have been something like this. We are told to think of the elements of a matrix as *truth-values*, but in the case of an algebra one should perhaps rather think of the elements as *propositions* (identifying propositions that are logically equivalent). The class of all propositions, if it exists, would presumably form one gigantic, complicated, universal algebra. But in a given context only a subclass of propositions are at issue, and they will form a simpler, more manageable algebra.

A particularly interesting paper with implications for modal logic is [Jónsson and Tarski, 1951]. If it had been widely read when it was published, the history of modal logic might have looked different. the scope of the paper is quite broad, but we should like to mention one or two results of particular relevance to modern modal logic. First, according to M. H. Stone's famous representation theorem, every Boolean algebra is isomorphic to a set of algebra. In other words, if  $\mathfrak{A} = \langle A, 0, 1, -, \cap, \cup \rangle$  is any Boolean algebra, then there exists a certain set  $U$  and a set  $B$  of subsets of  $U$ , closed under the Boolean operations, such that  $\mathfrak{A}$  is isomorphic to the Boolean algebra  $\mathfrak{B} = \langle B, \emptyset, U, -, \cap, \cup \rangle$ . (See [Rasiowa and Sikorski, 1963] for a good presentation of this and related results.) Jónsson and Tarski extend this result to Boolean algebras with operations (that is, functions from  $A^n$  to  $A$ , for any  $n$ ). If this does not sound too exciting, wait.

Suppose that  $U$  is any non-empty set, and let  $F$  be a family of subsets of  $U$  closed under the Boolean operations. Let  $\mathbf{l}, \mathbf{m} : F \rightarrow F$  be functions satisfying the following conditions:

$$\begin{array}{ll} \text{(l1)} & \mathbf{l}U = U, & \text{(m1)} & \mathbf{m}\emptyset = \emptyset, \\ \text{(l2)} & \mathbf{l}(X \cap Y) = \mathbf{l}X \cap \mathbf{l}Y, & \text{(m2)} & \mathbf{m}(X \cup Y) = \mathbf{m}X \cup \mathbf{m}Y, \\ \text{(lm)} & \mathbf{m}X = U - \mathbf{l}(U - X), & \text{(ml)} & \mathbf{l}X = U - \mathbf{m}(U - X). \end{array}$$

Then, according to Jónsson and Tarski, there exists a uniquely defined binary relation  $R$  on  $U$ —that is  $R \subseteq U \times U$ —such that

$$\begin{array}{ll} \text{(lR)} & \mathbf{l}X = \{x \in U : \forall y(xRy \Rightarrow y \in X)\}, \\ \text{(mR)} & \mathbf{m}X = \{x \in U : \exists y(xRy \& y \in X)\}; \end{array}$$

moreover, of the following conditions, (i1), (i2), and (i3) are mutually equivalent, for  $i = r, s, t$ :

$$\begin{array}{ll} \text{(r1)} & (\forall X \in F)(\mathbf{l}X \subseteq X), \\ \text{(r2)} & (\forall X \in F)(X \subseteq \mathbf{m}X), \end{array}$$

- (r3)  $R$  is reflexive with field  $U$ ;
- (s1)  $(\forall X, Y \in F)(Y \cup \mathbf{l}X = U \text{ iff } X \cup \mathbf{l}Y = U)$ ,
- (s2)  $(\forall X, Y \in F)(Y \cap \mathbf{m}X = \emptyset \text{ iff } X \cap \mathbf{m}Y = \emptyset)$ ,
- (s3)  $R$  is symmetric;
- (t1)  $(\forall X \in F)(\mathbf{l}X \subseteq \mathbf{l}\mathbf{l}X)$ ,
- (t2)  $(\forall X \in F)(\mathbf{m}\mathbf{m}X \subseteq \mathbf{m}X)$ ,
- (t3)  $R$  is transitive.

Conversely, if  $R$  is any binary relation on  $U$ , then  $(\mathbf{l}R)$  and  $(\mathbf{m}R)$  define functions  $\mathbf{l}, \mathbf{m} : F \rightarrow F$  such that again (i1), (i2), and (i3) are mutually equivalent, for  $i = r, s, t$ .

Putting all this together we arrive at the following picture. If we are analysing a class of propositions satisfying certain conditions, then we may try to cast them as an algebra  $\mathfrak{B} = \langle B, 0, 1, -, \cap, \cup, \mathbf{l}, \mathbf{m} \rangle$  where  $\langle B, 0, 1, -, \cap, \cup \rangle$  is a Boolean algebra and  $\mathbf{l}$  and  $\mathbf{m}$  are two additional unary operations. (If an element  $a \in B$  is taken to represent a proposition, then  $\mathbf{l}a$  and  $\mathbf{m}a$  would represent the propositions ‘ $a$  is necessary’ and ‘ $a$  is possible’, respectively.) By the representation theorem, there exists a set  $U$  such that  $\mathfrak{B}$  is isomorphic to an algebra  $\mathfrak{A} = \langle A, \emptyset, U, -, \cap, \cup, \mathbf{l}, \mathbf{m} \rangle$ , where  $A$  is a set of subsets of  $U$  and  $-, \cap, \cup$  are the usual set theoretical operations. Note that it is not claimed that every subset of  $U$  corresponds to a proposition, but that the converse claim is made: to every proposition  $a \in B$  a subset  $\|a\| \subseteq U$  corresponds. Under the intended interpretation it seems reasonable that  $\mathbf{l}$  and  $\mathbf{m}$  should satisfy conditions (l1), (l2), (lm) and (m1), (m2), (ml) above. Consequently Jónsson’s and Tarski’s result applies, and so  $\mathbf{l}$  and  $\mathbf{m}$  are completely determined by a certain binary relation  $R$ . Thus  $\mathfrak{A}$  is completely determined by  $U, R$ , and  $P$ , where  $P$  is the set of elements  $\|P\|$  such that  $P$  is an atomic proposition. In this sense,  $\mathfrak{A}$  is equivalent to the triple  $\langle U, R, P \rangle$ . Moreover, in the special case that the closure of  $P$  under  $\mathbf{l}$  and  $\mathbf{m}$  equals  $\mathfrak{B}u$ ,  $\mathfrak{A}$  is in the same sense equivalent to the pair  $\langle U, R \rangle$ . In view of later developments this is a striking result.

The reader is asked to keep the following observations in mind when readings Sections 4 and 10 below: for all  $a, b \in B$  and  $x \in U$ ,

$$\begin{aligned}
 x \in \|-a\| & \text{ iff } x \notin \|a\|, \\
 x \in \|a \cap b\| & \text{ iff } x \in \|a\| \text{ and } x \in \|b\|, \\
 x \in \|a \cup b\| & \text{ iff } x \in \|a\| \text{ or } x \in \|b\|, \\
 x \in \|\mathbf{l}a\| & \text{ iff } \forall y \in U (xRy \rightarrow y \in \|a\|), \\
 x \in \|\mathbf{m}a\| & \text{ iff } \exists y \in U (xRy \& y \in \|a\|).
 \end{aligned}$$

## 4 THE MODEL THEORETIC TRADITION

If algebraic semantics is discounted, then Rudolf Carnap was the first to provide a semantics for modal logic. Three of the all time greats came together in him. From Frege he got his interest in semantics and, more specifically, learnt to distinguish between intension and extension; and he attributes to Leibniz the notion that necessity is to be analysed as truth in all possible worlds. Moreover, he credits Wittgenstein with some ideas that formed the starting point for part of his own work (Carnap [1942; 1947]).

By a *state-description* let us understand a set of atomic propositions (propositional letters). If  $S$  is a state-description, then we may say what it means that a formula  $A$  holds in  $S$ , which in symbols we write  $\vDash_S A$ :

$$\begin{aligned} \vDash_S P &\text{ iff } P \in S, \text{ if } P \text{ is an atomic proposition,} \\ \vDash_S \neg A &\text{ iff not } \vDash_S A, \\ \vDash_S A \wedge B &\text{ iff } \vDash_S A \text{ and } \vDash_S B, \\ \vDash_S A \vee B &\text{ iff } \vDash_S A \text{ or } \vDash_S B, \\ \vDash_S A \rightarrow B &\text{ iff if } \vDash_S A \text{ then } \vDash_S B. \end{aligned}$$

If one is considering a definite collection  $C$  of state-descriptions, then also the following conditions become meaningful:

$$\begin{aligned} \vDash_S \Box A &\text{ iff, for all } T \in C, \vDash_T A, \\ \vDash_S \Diamond A &\text{ iff, for some } T \in C, \vDash_T A. \end{aligned}$$

Let us say that a formula is *valid in  $C$*  if it holds in every state description in  $C$ , and simply *valid* if it is valid in every collection of state-descriptions. this definition singles out a well-defined subset from the set of all formulas. Interestingly enough, this subset is the same as the set of theses of Lewis' system **S5**. Is this a coincidence? On the surface of it, Carnap's characterisation of **S5** looks very different from the original one due to Lewis.

This still does not look like modern modal logic: possible worlds are missing. According to Hintikka [1975], 'Carnap came extremely close to the basic ideas of possible-worlds semantics, and yet apparently did not formulate them, not even to himself'. this is drawing a very fine line, at least on the level of propositional logic. Carnap does talk about possible worlds. He is quite clear that he wants to latch on to Leibniz' suggestion that a necessary truth is one that holds in all possible worlds. Moreover, he says that his state-descriptions 'represent' possible worlds, which would seem to indicate that the former are (partial) descriptions of the latter. Thus from a formal point of view—Hintikka agrees with this—instead of the collections of state-descriptions that appear in the preceding paragraph, we could just as well have collections of possible worlds, provided only that we find a way of dealing with the first clause in the definition of 'holds in'. One virtue of state-descriptions, not shared by possible worlds, is that it is at once

clear what it means that a given atomic proposition hold in a given state-description. What we need, it seems, is a new primitive to perform this service. This leads us to re-cast Carnap's semantics in the following terms. We call  $\langle U, V \rangle$  a *Carnap-model* if  $U$  is any set (of *possible worlds*) and  $V$  (the *valuation*) is a function assigning to each atomic proposition  $P$  and possible world  $x$  a truth-value  $V(P, x)$  which is either T (truth) or F (falsity). In the definition of 'holds at' the first clause is replaced by this condition:

$$\vDash_x P \text{ iff } V(P, x) = \text{T, if } P \text{ is an atomic proposition.}$$

The other conditions are changed accordingly. In particular, those concerning the modal formulas become

$$\begin{aligned} \vDash_x \Box A &\text{ iff } \forall y \in U \vDash_y A, \\ \vDash_x \Diamond A &\text{ iff } \exists y \in U \vDash_y A. \end{aligned}$$

All this is no improvement on Carnap, but it brings us into line with modern terminology. It should be added that the picture of Carnap given here is a pale one since so much of importance in his work is found at the level of predicate logic, which is not considered in this article.

The next step of importance within the semantic tradition was taken by Arthur Prior. both Lewis and Carnap had been concerned with the analysis of modal concepts in the strict sense, but, as remarked in Section 2, some authors have also tried to model concepts which are called modal in the wide sense (imperative, deontic, etc.). The efforts of the latter had been syntactic, but Prior, whose interests lay in temporal notions, gave an algebraic flavoured analysis which in effect was a model theoretic one. In his book, Prior [1957], he models time as the set  $\omega$  of natural numbers. Thus instead of Carnap models we now meet with structures  $\langle \omega, V \rangle$  which we might call *Prior models* and in which the unspecified collection  $U$  of possible worlds of a Carnap model  $\langle U, V \rangle$  is replaced by the special set  $\omega$  representing a set of points of time.

With the help of Prior models many new operators are definable. In [Prior, 1957] attention is focused on the operators defined by the conditions

$$\begin{aligned} \vDash_t \Box A &\text{ iff } \forall u \geq t \vDash_u A, \\ \vDash_t \Diamond A &\text{ iff } \exists u \geq t \vDash_u A. \end{aligned}$$

Later Prior was to consider also the related operators defined by the conditions

$$\begin{aligned} \vDash_t \Box A &\text{ iff } \forall u > t \vDash_u A, \\ \vDash_t \Diamond A &\text{ iff } \exists u > t \vDash_u A. \end{aligned}$$

There is almost no end to the number of new operators thus definable. Already in [Prior, 1957] one finds conditions like

$$\begin{aligned} \vDash_t \Box A &\text{ iff } \vDash_t A \text{ and } \vDash_{t+1} A, \\ \vDash_t \Diamond A &\text{ iff } \vDash_t A \text{ or } \vDash_{t+1} A; \end{aligned}$$

and later developments have seen a host of others.

Once Prior had shown how to do tense logic, much activity followed. For example, it is natural to study Prior models in which the set  $\omega$  of natural numbers is replaced by the set  $\xi$  of all integers, or the set  $\eta$  of rational numbers, or the set  $\lambda$  of real numbers. Much attention was also devoted to studying the interaction of several temporal and other operators in multi-modal systems. (One among many good references in tense logic is [Rescher and Urquhart, 1971].) Prior's work paved the way for Kamp [1968] where for the first time exact definitions of the notion of tense were offered. For example, according to Kamp, an  $n$ -place tense in discrete time is a function  $f$  from  $(\mathfrak{B}\xi)^n$  to  $\mathfrak{B}\xi$ ; and an  $n$ -ary operators  $\star$  will express this tense if, for all  $t \in \xi$ ,

$$\models_t \star(A_0, \dots, A_{n-1}) \text{ iff } t \in f(\{u : \models_u A_0\}, \dots, \{u : \models_u A_{n-1}\}).$$

With Kamp [1968] tense logic achieved a new level of sophistication. However, much of the early interest concerned more basic problems, for example, that of characterising the operators defined by the first of the three definitions given above. This logic, the so-called Diodorean logic, is not as strong as **S5**, yet stronger than **S4**, as pointed out by Hintikka, Dummett and others. Its true identity was finally settled by S. A. Kripke and R. A. Bull, independently [Bull, 1965]. For an entertaining account of this, see [Prior, 1967, Chapter 2].

All of this is sorted out in the chapter on tense logic (see the chapter by Burgess in a later volume of this *Handbook*. What is important here is that Prior replaces Carnap's *unordered* set of possible worlds (actually, state-descriptions) by an *ordered* set of possible worlds (actually, points of time). In order to stress this difference we should perhaps have introduced the Prior models as triples  $\langle \omega, \leq, V \rangle$ , where  $\leq$  is the ordinary less-than-or-equal-to ordering of the natural numbers. Thus in retrospect it seems that Carnap and Prior between them supplied all the necessary ingredients for modal logic as we know it at present. Already Jónsson and Tarski had explored the mathematics that is needed, and in Carnap and Prior there was sufficient philosophical underpinning to get modern modal logic going. The modern notion of a model is a triple  $\langle U, R, V \rangle$ , where  $U$  is a set (of *possible worlds*, or, more neutrally, *indices*, or even just *points*),  $R$  a binary relation on  $U$  (the *accessibility* relation (Geach) or the *alternativeness* relation (Hintikka)), and  $V$  a valuation. As we say the elements  $U$  and  $V$  were contributed by Carnap, and the relation  $R$  is obtained by generalising ever so slightly over Prior: instead of working with his special cases, we keep as the one general requirement that  $R$  is a binary relation, not necessarily an ordering.

But this is not the way history is usually written. So-called possible worlds semantics or Kripke semantics is commonly attributed to S. A.

Kripke, who laid down the foundations of modern propositional and predicate modal logic in several influential papers (Kripke [1959; 1963; 1963a; 1965]). Relatively less influential were the papers by Jaakko Hintikka and Stig Kanger (Hintikka [1957; 1961; 1963]; Kanger [1957; 1957a; 1957b; 1957c]). Actually the three seem to have been independent of one another; but Kanger published first. Kanger's writings are difficult to decipher, and this fact, paired with the unassuming mode of their publication, may have been what has deprived him of some of the recognition due to him (cf. Hintikka's generous review, [Hintikka, 1969a]). Hintikka has had more impact, especially on the philosophers.

The reason his work has been less important for the formal development of modal logic than that of Kripke is perhaps his style of presentation which tones down mathematical aspects and skips proofs.

## 5 OTHER TRADITIONS

In the preceding sections we have described what seems to us to be the main developments in early modal logic. no history is ever complete, and starts not recorded here have been made without their developing into what we regard as a major tradition. In this section we will briefly mention five or six such starts.

First there is the so-called provability interpretation(s) of modal logic, the embryo of which is found in [Gödel, 1933]. In view of recent development one may perhaps say that this is expanding into a new tradition right now. Via Montague [1963], Friedman [1975] and Solovay [1976] it has begun to generate a literature of its own. For more information on this, see [Boolos, 1979] and Smoryński's chapter in a later volume of this *Handbook*.

Another start, more suggestive than seminal, was made by J. C. C. McKinsey who described what is now known as McKinsey's syntactic interpretation of modal logic [McKinsey, 1945]; McKinsey's idea was perhaps foreshadowed in Fitch [1937; 1939], it is taken up again in [Morgan, 1979]. A third start was made by Alonzo Church in a series of papers ([1946; 1951; 1973–74]); recent contributions to this area are Parsons [1982] and C. A. Anderson [1980]. (Cf. also his chapter in volume 4 of this *Handbook*.) A fourth start worth mentioning was made with the appearance of Arthur Prior's three-valued modal logic **Q**. many-valued modal logic is not a vast field and in any case mainly falls under what we have called the algebraic tradition, but **Q**, first defined in [Prior, 1957], seems to be of particular philosophical interest; see, for example, [Fine, 1977].

Finally there ought to be a tradition called intuitionistic modal logic, but it is debatable whether today even a subtradition can be found under that heading. Perhaps Ditch [1948], Curry [1950] and Prawitz [1965] can be regarded as starts, but they are not very illuminating as analyses of

modality; and work on semantics has, to date, been in the classical spirit (Bull [1965a], Fischer Servi [1977; 1981]). Why intuitionistically minded logicians have not been attracted to this area is not clear, and surely it would be interesting to see an intuitionistic-logical analysis of knowledge (including extra-mathematical knowledge), obligation, imperative, perception, and other notions which are modal in the wide sense.

## Systematic Part

### 6 LOGICS AND DEDUCIBILITY RELATIONS

In the preceding sections our primary concern has been historical. It is now time to being a more systematic exposition. In this section we will give a number of concepts which are useful when it comes to classifying modal logics. First we give a family of (more or less) traditional definitions, and then we develop similar definitions of a slightly more general nature.

Modal logics are often defined as sets of formulas of a certain kind. One might begin by defining a *logic* as a set  $\mathbf{L}$  of formulas satisfying the following conditions:

- (tf)  $A \in \mathbf{L}$ , whenever  $A$  is a tautology in the sense of classical propositional logic;
- (mp) if  $A \rightarrow B \in \mathbf{L}$  and  $A \in \mathbf{L}$ , then  $B \in \mathbf{L}$ ;
- (sb) if  $A \in \mathbf{L}$ , then  $sA \in \mathbf{L}$ , if  $sA$  is the result of uniform substitution of formulas for propositional letters in  $A$ .

Then one might perhaps go on to say that a logic  $\mathbf{L}$  is *classical modal* if it contains the formulas

- K.**  $\Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q)$ ,
- \***.  $\Box T$ ,

(where  $P, Q$  are two propositional letters and  $T$  is either primitive or some chosen tautology) and in addition is closed under replacement of tautological equivalents:

- (rte) If  $A$  and  $B$  are tautologically equivalent and  $C$  and  $C^*$  are identical except that one occurrence of  $A$  in  $C$  has been replaced by an occurrence of  $B$  to give  $C^*$ , then  $C \in \mathbf{L}$  iff  $C^* \in \mathbf{L}$ .

This is a very weak conception of classical modal logic (incidentally, differing from that in [Seegerberg, 1971]), and usually one would require much more, for example, closure under congruence (cgr), monotonicity (mon), or necessitation (nec):



(cgr) if  $A \leftrightarrow B \in \mathbf{L}$ , then  $\Box A \leftrightarrow \Box B \in \mathbf{L}$ ;

(mon) if  $A \rightarrow B \in \mathbf{L}$ , then  $\Box A \rightarrow \Box B \in \mathbf{L}$ ;

(nec) if  $A \in \mathbf{L}$ , then  $\Box A \in \mathbf{L}$ .

A modal logic satisfying (cgr) ((mon), (nec)) would be called *congruential* (*regular*, *normal*). Moreover, a modal logic would be *quasi-congruential* (*quasi-regular*, *quasi-normal*) if it contained some congruential (regular, normal) modal logic. (A logic containing a classical modal logic is of course itself classical modal.) Notice that normality implies regularity implies congruentiality. If  $\Box$  is the only non-Boolean operator, then congruentiality implies replacement of tautological equivalents. (Our terminology is not completely standard, but at least the definitions of ‘logic’, ‘regular’, ‘normal’, and ‘quasi-normal’ appear to be.)

So far tradition. however, there is also a more roundabout way to arriving at similar definitions which begins with deducibility relations instead of with logics. It may be instructive to offer these slightly more general definitions as well. In this paper—and here we offer less than full generality—a *deducibility relation*  $R$  is a set of ordered pairs  $\langle \Gamma, A \rangle$ , where  $\Gamma$  is a set of formulas and  $A$  is a formula. If  $\langle \Gamma, A \rangle \in R$  we say that  $\Gamma$  *yields*  $A$  and write  $\Gamma \vdash_R A$ , or even  $\Gamma \vdash A$  when suppression of the subscript does not lead to confusion. If  $\Gamma \vdash A$  and  $\Gamma = \emptyset$  we write  $\vdash A$  and say that  $A$  is a *thesis* of  $R$ . The set of theses of  $R$  is denoted by  $\text{Th } R$ . We usually write  $A_0, \dots, A_{n-1} \vdash B$  instead of  $\{A_0, \dots, A_{n-1}\} \vdash B$ ; also  $A_0, \dots, A_{n-1}, \Gamma \vdash B$  instead of  $\{A_0, \dots, A_{n-1}\}, \Gamma \vdash B$ . If  $A \vdash B$  and  $B \vdash A$  we write  $A \dashv\vdash B$ .

Common conditions on deducibility relations re reflexivity (RX), (left) monotonicity (LM), cut (CUT), and substitutivity (SB):

(RX)  $A \vdash A$ ;

(LM) if  $\Gamma \vdash A$  and  $\Gamma \subseteq \Delta$ , then  $\Delta \vdash A$ ;

(CUT) if  $\Gamma \vdash C$  and  $C, \Gamma \vdash A$ , then  $\Gamma \vdash A$ ;

(SB) if  $\Gamma \vdash A$ , then  $s\Gamma \vdash sA$ , if  $s\Gamma$  and  $sA$  are the result of uniform substitution in  $\Gamma$  and  $A$ , respectively, of formulas for propositional letters.

A deducibility relation is *Boolean* if it also satisfies the conditions in Table 1 (we assume a truth-value functionally complete set of Boolean operators). A deducibility relation is *compact* if, whenever  $\Gamma \vdash B$ , there are some  $A_0, \dots, A_{n-1} \in \Gamma$ , for some  $n \geq 0$ , such that  $A_0, \dots, A_{n-1} \vdash B$ . Notice that two compact Boolean deducibility relations coincide if they agree on their theses:  $\text{Th}R = \text{Th}R'$  implies that  $R = R'$ .

The concepts defined above for logics may now be given analogous definitions in the context of deducibility relations. first, let us say that a deducibility relation is *n-modal* if

(*n*-M) if  $\Gamma$  tautologically implies  $A$ , then  $\Box^n \Gamma \vdash \Box^n \Gamma \Box^n A$ , provided that  $\Gamma \neq \emptyset$ .

Table 1.

( $\wedge$ E)	If $\Gamma \vdash A \wedge B$ , then $\Gamma \vdash A$ and $\Gamma \vdash B$ .
( $\wedge$ I)	If $\Gamma \vdash A$ and $\Gamma \vdash B$ , then $\Gamma \vdash A \wedge B$ .
( $\vee$ E)	If $\Gamma \vdash A \vee B$ and $A, \Gamma \vdash C$ and $B, \Gamma \vdash C$ , then $\Gamma \vdash C$ .
( $\vee$ I)	If $\Gamma \vdash A$ or $\Gamma \vdash B$ , then $\Gamma \vdash A \vee B$ .
( $\rightarrow$ E)	If $\Gamma \vdash A \rightarrow B$ and $\Gamma \vdash A$ , then $\Gamma \vdash B$ .
( $\rightarrow$ I)	If $A, \Gamma \vdash B$ , then $\Gamma \vdash A \rightarrow B$ .
( $\neg$ E)	If $\Gamma \vdash \neg A$ and $\Gamma \vdash A$ , then $\Gamma \vdash B$ .
( $\neg$ I)	If $A, \Gamma \vdash \neg A$ , then $\Gamma \vdash \neg A$ .
(RAA)	If $\neg A, \Gamma \vdash A$ , then $\Gamma \vdash A$ .

(Here  $\Box^n A$  is the formula consisting of the formula  $A$  preceded by a string of  $n$  occurrences of  $\Box$ , while  $\Box^n \Gamma = \{\Box^n B : B \in \Gamma\}$ . Let us say that a Boolean deducibility relation is *modal* if it is 1-modal, and *strongly modal* if it is  $n$ -modal for all  $n$ .)

Next, let us say that a deducibility relation is *classical* if it is closed under the following condition of replacement under tautological equivalents:

(RTE) If  $A$  and  $B$  are tautologically equivalent, and  $C$  and  $C^*$  are identical except that one occurrence of  $A$  in  $C$  has been replaced by an occurrence of  $B$  to give  $C^*$ , then  $C \dashv\vdash C^*$ .

Finally, let us say that a deducibility relation is *congruential* (*regular*, *normal*) if it satisfies (CGR)((SC1), (SC2)):

(CGR) If  $A \dashv\vdash B$ , then  $\Box A \dashv\vdash \Box B$ ;

(SC1) If  $\Gamma \vdash A$ , then  $\Box \Gamma \vdash \Box A$ , provided that  $\Gamma \neq \emptyset$ ;

(SC2) If  $\Gamma \vdash A$ , then  $\Box \Gamma \vdash \Box A$ .

(Conditions (SC1) and (SC2) are due to Dana Scott, whence the notation.)

Let us now review the situation. It is readily seen that every Boolean deducibility relation  $R$  determines a unique logic, viz. Th  $\mathbf{R}$ . Conversely, every logic  $\mathbf{L}$  determines a compact Boolean deducibility relation Rel  $\mathbf{L}$  in a natural manner:  $\Gamma \vdash B$  iff there are  $A_0, \dots, A_{n-1} \in \Gamma$ , for some  $n \geq 0$ , such that  $((A_0 \wedge \dots \wedge A_{n-1}) \rightarrow B) \in \mathbf{L}$ . Note that

$\mathbf{L} = \text{Th Rel } \mathbf{L}$ , for every logic  $\mathbf{L}$ ,

$\mathbf{R} = \text{Rel Th } \mathbf{R}$ , for every compact, Boolean deducibility relation  $\mathbf{R}$ .

Moreover, note that if  $\mathbf{L}$  is classical modal (and also congruential, regular, or normal, respectively), in the sense of logics, then so is Rel  $\mathbf{L}$ , in the sense

of deducibility relations; and if a compact Boolean deducibility relation is classical modal (and also congruential, regular, or normal, respectively), in the sense of deducibility relations, then so is  $\text{Th } \mathbf{R}$ , in the sense of logics. In view of a preceding remark we know that  $\text{Rel } \mathbf{L}$  is the only compact deducibility relation with  $\mathbf{L}$  as its set of theses. Therefore, evidently, if, as in this paper, one is only interested in compact deducibility relations, it is harmless to restrict oneself to the study of logics; which is what one has usually done traditionally.

For some recent works in which deducibility is seen as primary, rather than thesishood, see [Scott, 1971; Kuhn, 1977; Shoesmith and Smiley, 1978; Gabbay, 1981; Segerberg, 1982]. Ultimately this approach seems to derive from two quite different sources, Gentzen and Tarski.

## 7 A CATALOGUE OF MODAL LOGICS

Almost all recent work in modal logic has been concerned with normal logics. At least from a technical point of view, non-normal, regular or quasi-regular logics—a class which includes **S2**, **S3**, **S6** and **S7**—seem to offer little of interest beyond what normal logics offer, and for that reason we will not treat them here but refer the reader to [Kripke, 1965] and [Lemmon, 1957; Lemmon, 1966]. Among logics that are not even quasi-regular, the congruential merit some attention, and in Section 21 below some are implicit. But with this exception the purview of this paper is normal modal logics.

Over the years an almost astronomical number of modal logics have been put forward. Under such circumstances, naming or identifying logics becomes a problem. The best nomenclature is perhaps the one proposed by E. J. Lemmon in [Lemmon, 1977], and here we will usually employ a variant of it. The smallest normal logic we designate by '**K**' (in honour of Kripke who, curiously enough, seems never to have dealt with this particular logic). If ' $\mathbf{X}_0$ ',  $\dots$ , ' $\mathbf{X}_{m-1}$ ' name any formulas, then ' $\mathbf{KX}_0, \dots, \mathbf{X}_{m-1}$ ' is the *Lemmon code* for the smallest normal logic that contains  $\mathbf{X}_0, \dots, \mathbf{X}_{m-1}$ . Note that, by definition, this logic is closed under substitution.

Lemmon's convention presupposes that formulas have names. Here is a list of formulas with names that either are more or less standard, or else in the opinion of the authors deserves to be:

<b>D.</b>	$\Box P \rightarrow \Diamond P,$
<b>T.</b>	$\Box P \rightarrow P,$
<b>4.</b>	$\Box P \rightarrow \Box \Box P,$
<b>E.</b>	$\Diamond P \rightarrow \Box \Diamond P,$
<b>B.</b>	$P \rightarrow \Box \Diamond P,$
<b>Tr.</b>	$\Box P \leftrightarrow P,$
<b>V.</b>	$\Box P,$
<b>M.</b>	$\Box \Diamond P \rightarrow \Diamond \Box P,$
<b>G.</b>	$\Diamond \Box P \rightarrow \Box \Diamond P,$
<b>H.</b>	$(\Diamond P \wedge \Diamond Q) \rightarrow (\Diamond(P \wedge Q) \vee \Diamond(P \wedge \Diamond Q) \vee \Diamond(Q \wedge \Diamond P)),$
<b>Grz.</b>	$\Box(\Box(P \rightarrow \Box P) \rightarrow P) \rightarrow P,$
<b>Dum.</b>	$\Box(\Box(P \rightarrow \Box P) \rightarrow P) \rightarrow (\Diamond \Box P \rightarrow P),$
<b>W.</b>	$\Box(\Box P \rightarrow P) \rightarrow \Box P.$

the following remarks will make it easier to remember these names. ‘**D**’ stands for deontic, ‘**T**’ comes from ‘**t**’, a name invented by Feys, **4** is the characteristic axiom of Lewis’ **S4**, ‘**E**’ stands for Euclidean, ‘**B**’ for Brouwer, ‘**Tr**’ for trivial, ‘**V**’ for *verum*, ‘**M**’ for McKinsey, ‘**F**’ for Geach, ‘**H**’ for Hintikka, ‘**Grz**’ for Grzegorzczuk, ‘**Dum**’ for Dummett, and ‘**W**’ for (anti-)well-ordered. The strangest of these names is perhaps ‘**B**’ for Brouwer, as the father of mathematical intuitionism was never known to harbour much sympathy for logic, let alone modal logic. The name hails back to Oskar Becker who saw a similarity between the logic **KTb** and intuitionistic logic [Becker, 1930].

Of the many logics that can be defined in terms of the above formulas we list the following:

<b>KT</b> = <b>T</b> = the Gödel/Feys/Von Wright system,
<b>KT4</b> = <b>S4</b>
<b>KT4B</b> = <b>KT4E</b> = <b>S5</b>
<b>KD</b> = deontic <b>T</b> ,
<b>KD4</b> = deontic <b>S4</b> ,
<b>KD4E</b> = deontic <b>S5</b> ,
<b>KTb</b> = the Brouwer system (‘the em Brouwersche system’),
<b>KT4M</b> = <b>S4.1</b> ,
<b>KT4G</b> = <b>S4.2</b> ,
<b>KT4H</b> = <b>S4.3</b> ,
<b>KT4Dum</b> = <b>D</b> = Prior’s Diodorean logic,
<b>KT4Grz</b> = <b>KGrz</b> = Grzegorzczuk’s system,
<b>K4W</b> = <b>KW</b> = Löb’s system,
<b>KTr</b> = <b>KT4BM</b> = the trivial system,
<b>KV</b> = the <i>verum</i> system.

There is no upper bound to the number of normal modal logics, and many—perhaps too many—have found their way into the literature. But the given catalogue includes many of the most studied systems.

If the inconsistent logic, the set of all formulas, is accepted as a normal modal logic—and under the definition given here it must be—then the set of all normal modal logics forms a distributive lattice under the operations g.l.b.  $(\mathbf{L}, \mathbf{L}') =$  the greatest normal logic to be contained in both  $\mathbf{L}$  and  $\mathbf{L}'$  (which is the same as  $\mathbf{L} \cap \mathbf{L}'$ ) and l.u.b.  $(\mathbf{L}, \mathbf{L}') =$  the smallest normal logic to extend both  $\mathbf{L}$  and  $\mathbf{L}'$  (which is *not* the same as  $\mathbf{L} \cup \mathbf{L}'$ ). Much effort has gone into exploring the nature of this enormously complicated lattice. Early contributions were made by Scroggs who mapped out all the extensions of  $\mathbf{S5}$  [Scroggs, 1951]; by Bull who did the same for the extensions of  $\mathbf{S4.3}$  [Bull, 1966]; by Makinson who showed that the trivial system and the *verum* system are the two dual atoms of this lattice [Makinson, 1971]; and by McKinsey and Tarski who showed that there are non-normal extensions of  $\mathbf{S4}$  [McKinsey and Tarski, 1948]. Kit Fine and Wim Block have done more than anyone else to complete the picture, and some of their work is described below. Schumm [1981] sums up some of the things that are known about the elements of the big lattice. Readers interested in the geography of modal logic are also referred to Hansson and Gärdenfors [1973].

## 8 SEMANTIC TABLEAUX AND HINTIKKA SYSTEMS

The deductive systems given in the preceding sections are of so-called Hilbert type, strict on rules and soft on axioms. Most of the deductive systems in the modal logic literature are of this type. From a metamathematical point of view such systems have much to offer. But if one's interest lies in proving theorems *in* a system rather than *about* it, then they are not terribly accommodating. Yet in modal logic they have had relatively little competition from other kinds of deductive systems. The most common system of a different kind is no doubt the procedure due to Hintikka and Kripke (similar ideas in a less developed form are found in [Guillaume, 1958]). Hintikka's work on model system [1957; 1961; 1962; 1963] and Kripke's on semantic tableaux [1963; 1963a] were independent, and even though the two methods are equivalent they are not identical. It would take us too far here to discuss both, and here we will follow Hintikka. For classical logic the general references are the classic works [Beth, 1959] and [Hintikka, 1955] as well as the later monograph [Smullyan, 1968]. an elementary and particularly readable account is given in [Jeffrey, 1990].

We define a set  $\Sigma$  of formulas as *downward saturated* if it satisfies the following conditions:

- (C $\neg$ ) If  $\neg A \in \Sigma$ , then  $A \notin Sigma$ .
- (C $\wedge$ ) If  $A \wedge B \in \Sigma$ , then  $A \in \Sigma$  and  $B \in Sigma$ .
- (C $\vee$ ) If  $A \vee B \in \Sigma$ , then  $A \in \Sigma$  or  $B \in \Sigma$ ,
- (C $\rightarrow$ ) If  $A \rightarrow B \in \Sigma$ , then  $A \in \Sigma$  only if  $B \in \Sigma$ .
- (C $\neg\neg$ ) If  $\neg\neg A \in \Sigma$ , then  $A \in \Sigma$ .
- (C $\neg\wedge$ ) If  $\neg(A \wedge B) \in \Sigma$ , then  $\neg A \in \Sigma$  or  $\neg B \in \Sigma$ .
- (C $\neg\vee$ ) If  $\neg(A \vee B) \in \Sigma$ , then  $\neg A \in \Sigma$  and  $\neg B \in \Sigma$ .
- (C $\neg\rightarrow$ ) If  $\neg(A \rightarrow B) \in \Sigma$ , then  $A \in \Sigma$  and  $\neg B \in \Sigma$ .

The seven last conditions define an effective procedure: given any finite set  $\Sigma$  it is possible to add a finite number of new formulas to it to obtain a set  $\Sigma^*$  which satisfies all the conditions except perhaps (C $\neg$ ); this would be to *embed*  $\Sigma$  in  $\Sigma^*$ . Notice that  $\Sigma^*$  is downwards saturated only if also (C $\neg$ ) holds. The latter condition is evidently of a different character from the others: they *prescribe* membership under *some* conditions, whereas (C $\neg$ ) *proscribes* it under *all*. That is to say, (C $\neg$ ) is a consistency condition.

We are now able to define a deducibility relation as follows:  $\Gamma \vdash B$  if and only if the set  $\Gamma \cup \{\neg B\}$  cannot be embedded in a downwards saturated set. Specifically, if  $\Gamma$  is finite,

- (\*)  $A_0, \dots, A_{n-1} \vdash B$  iff, for every downwards saturated set  $\Sigma$ ,  
if  $A_0, \dots, A_{n-1} \in \Sigma$ , then  $\neg B \notin \Sigma$ .

The reason this deducibility relation is of interest is that it coincides with classical logic:  $\Gamma \vdash A$  iff  $\Gamma$  tautologically implies  $A$ . Furthermore, by the compactness theorem of classical propositional logic,  $\Gamma \vdash B$  only if for some  $n \geq 0$  and some  $A_0, \dots, A_{n-1} \in \Gamma$  we have  $A_0, \dots, A_{n-1} \vdash B$ .

The question arises, how to extend this analysis to modal logic. From a syntactic point of view, all that would be needed is two additional rules, (C $\Box$ ) and (C $\neg\Box$ ) of a similar kind. By ‘similar’ is meant that the rules would have to be such that the Augmented set of rules would again define a (not necessarily effective) procedure. It turns out that in order to do this we have to widen the perspective. What both Hintikka and Kripke did was to consider not just downward saturated sets (respectively, semantic tableaux) but systems of such sets (respectively, tableaux). Let us call a triple  $\langle \Sigma_0, U, R \rangle$  a *Hintikka system* if the following is true. First,  $U$  is a set of downward saturated sets of which  $\Sigma_0$  is one; and  $R$  is a binary relation over  $U$  (called the *alternativeness relation* by Hintikka) which generates  $U$  from  $\Sigma_0$  in the sense that, for each  $\Sigma \in U$ , there are some sets  $\Sigma_1, \Sigma_2, \dots, \Sigma_k \in U$ , for some  $k \geq 0$ , such that  $\Sigma_i R \Sigma_{i+1}$ , for all  $k < k$ , and  $\Sigma_k = \Sigma$ . Second, for every  $\Sigma \in U$  the following conditions are satisfied:

- (C $\Box$ ) If  $\Box A \in \Sigma$ , then  $A \in \Sigma'$ , for all  $\Sigma' \in U$  such that  $\Sigma R \Sigma'$ .
- (C $\neg\Box$ ) If  $\neg\Box A \in \Sigma$ , then  $\neg A \in \Sigma'$ , for some  $\Sigma' \in U$  such that  $\Sigma R \Sigma'$ .

We are now able to define a deducibility relation for modal logic:  $\Gamma \vdash A$  iff the set  $\Gamma \cup \{\neg A\}$  cannot be embedded in a Hintikka system (in the obvious sense: there is no Hintikka system  $\langle \Sigma_0, U, R \rangle$  such that  $\Gamma \cup \{\neg A\} \subseteq \Sigma_0$ ). As Hintikka and Kripke proved (and, in effect, Kanger had proved before them), the deducibility relation thus introduced will coincide with the famous modal logics **T**, **S4**, and **S5**, respectively, if special conditions are placed on the alternativeness relation, viz. reflexivity; reflexivity and transitivity; reflexivity, transitivity, and symmetry; respectively. These are no doubt the most celebrated of all results in modal logic, and much of the success of the new semantics is probably due to the fact that the three most important systems of modal logic can be given such a simple characterisation in these new terms. Other conditions than those mentioned can also be considered, and it turns out that for practically all systems in the literature that have been proposed for their philosophical virtues, a similar model theoretic characterisation is possible.

What we have so far is just a procedure. Primarily it is a disproof procedure (successful if an appropriate Hintikka system is found). Secondly it is also (the beginning of) a proof procedure (successful if it can be shown that no appropriate Hintikka system can be found). In general neither procedure need be effective, though, for the new rule ( $C\neg\Box$ ) may introduce new formula sets, and the implicit procedure may therefore not terminate. In other words, given some conditions on the alternativeness relation and formulas  $A_0, \dots, A_{n-1}, B$ , there is no guarantee that one will ever be able to settle the question whether  $A_0, \dots, A_{n-1} \vdash B$  (even though, as it turns out, in many cases such a guarantee can be given).

From a philosophical point of view it should be noted that what we have above is not yet a semantics in any but a combinatorial sense of the word. As in the case of Carnap—there is of course a close connection between state-descriptions and a downward saturated set—a real semantics is obtained if possible worlds are postulated and downward saturated sets are identified as partial descriptions of them.

We shall append two observations which are of some interest. Let us say that a set of formulas is *upward saturated* if the converses of the above  $C$ -conditions for the classical operators are satisfied, and *maximal consistent* if it is saturated both upward and downward. The first observation is a familiar one: we again get classical propositional logic by stipulating that  $\Gamma \vdash B$  iff  $\Gamma \cup \{\neg B\}$  cannot be embedded in a maximal consistent set. Specifically, if  $\Gamma$  is finite,

$$(\S) \quad A_0, \dots, A_{n-1} \vdash B \text{ iff, for every maximal consistent set } \Sigma, \text{ if } A_0, \dots, A_{n-1} \in \Sigma, \text{ then } B \in \Sigma.$$

This statement, which is nothing but the famous Lindenbaum's Lemma, should be compared to (\*) above.

Suppose now that we call a set  $\langle \Sigma_0, U, R \rangle$  of maximal consistent sets a *Henkin system* if  $U$  is a set of maximal consistent sets of which  $\Sigma_0$  is one, and  $R$  is a binary relation on  $U$  such that  $(C\Box)$  and  $(C\neg\Box)$  as well as their converses are satisfied by every  $\Sigma \in U$ . Then once again we get a deducibility relation by stipulating that  $\Gamma \vdash A$  iff  $\Gamma \cup \{\neg A\}$  cannot be embedded in a Henkin system (in the obvious sense: there is no Henkin system  $\langle \Sigma_0, U, R \rangle$  such that  $\Gamma \cup \{\neg A\} \subseteq \Sigma_0$ ). This suggests the second observation, viz. that the relation between downward saturated sets and maximal consistent sets in classical logic is, in some sense, the same as that between Hintikka systems and Henkin systems in modal logic. In fact, Henkin systems have been more used than Hintikka systems in the study of modern modal logic. They were introduced independently by Makinson [1966], Cresswell [1967], Schütte [1968] and perhaps others. Dana Scott had similar ideas a little earlier and exerted a powerful influence even though he did not publish; cf. Kaplan [1966] and Lemmon [1966; 1977]. Another early reference in this context is [Bayart, 1959].

## 9 NATURAL DEDUCTION IN MODAL LOGIC

Seen in a grand perspective, the Hintikka/Kripke deductive technique is an extension to modal logic of ideas introduced into the study of classical logic by P. Hertz and G. Gentzen. However, some have proposed a more straightforward extension of those ideas. In this section we will consider to what extent such an effort is likely to succeed.

Perhaps the most important work in the latter tradition is Prawitz [1965]. We will begin by giving a standard system of natural deduction for classical propositional logic which is similar to one found there. First there are the *inference rules* listed in Table 2. here ‘E’ and ‘I’ stand for ‘elimination’ and ‘introduction’ respectively, while ‘RAA’ is short for ‘*reductio ad absurdum*’.

Next we should give the *deduction rules*, that is, rules which legislate how inference rules may be used to produce deductions. But deduction rules are cumbersome to state in full detail. Therefore we will make a short-cut. (Readers who are led astray by this short-cut should consult [Prawitz, 1965].)

As usual,  $\Gamma \vdash A$  is defined to mean that there is a deduction where  $A$  is the conclusion (‘the bottom formula’) and where  $\Gamma$  contains all premises (‘undischarged top formulas’). It is immediate that the deducibility relation  $\vdash$  will satisfy the common conditions (RX), (LM), (CUT), and (SB) defined in Section 6. Now we declare—this is the short-cut—that the deduction rules are exactly what it takes to make certain that the conditions of Table 1 of the same section to be satisfied; thus  $\vdash$  is a Boolean deducibility relation. Notice that there is a one-to-one correspondence between the conditions of Table 1 and the inference rules of Table 2. In order to stress the connection we have used the same name for both condition and inference rule: in effect



Table 2.

( $\wedge$ E)	$\frac{A \wedge B}{A}$	$\frac{A \wedge B}{B}$	( $\wedge$ I)	$\frac{A \quad B}{A \wedge B}$
( $\vee$ E)	$\frac{A \vee B}{A \vee B}$	$\frac{\begin{array}{c} B \\ (A) \quad (B) \\ C \quad C \end{array}}{C}$	( $\vee$ I)	$\frac{A \quad B}{A \vee B} \quad \frac{B}{A \vee B}$
( $\rightarrow$ E)	$\frac{A \rightarrow B \quad A}{B}$	$\frac{c}{c}$	( $\rightarrow$ I)	$\frac{\begin{array}{c} (A) \\ B \\ A \rightarrow B \\ (A) \end{array}}{A \rightarrow B}$
( $\neg$ E)	$\frac{\neg A \quad A}{B}$	$\frac{\neg A}{\neg A}$	( $\neg$ I)	$\frac{\begin{array}{c} (A) \\ \neg A \end{array}}{\neg A}$
			( $\neg$ RAA)	$\frac{\begin{array}{c} (\neg A) \\ A \end{array}}{A}$

the condition explains how the inference rule is to be applied. This is needed, especially in the case of the so-called improper inference rules, that is, those containing parentheses: ( $\vee$ E) ( $\rightarrow$ I), ( $\neg$ I), (RAA). What is at issue here is on exactly what premises a conclusion *depends*, and this can be gathered from the observations.

The interest in the system thus presented is that the deducibility relation it defines coincides with that of classical logic:  $\Gamma \vdash A$  iff  $\Gamma$  tautologically implies  $A$ . In order to generalise it to modal logic, the most direct course is to try and devise rules for  $\Box$  of the same kind as those governing the classical operators; in other words, to force the classical pattern on the modal operator. Thus one elimination and one introduction rule are called for, and their form is obvious:

$$(\Box E) \quad \frac{\Box A}{A} \quad (\Box I) \quad \frac{A}{\Box A}$$

This is what Prawitz does. He considers ( $\Box$ E) a proper rule, which means that

$$(\Box E) \quad \text{If } \Gamma \vdash \Box A, \text{ then } \Gamma \vdash A.$$

By contrast, ( $\Box$ I) is very much improper: taking it as a proper rule would literally trivialise modal logic. That is, if one accepts

$$(\Box I) \quad \text{If } \Gamma \vdash A, \text{ then } \Gamma \vdash \Box A,$$

then the resulting deducibility relation coincides with the trivial system defined in Section 7. Thus in all interesting cases the deduction rule for  $(\Box I)$  will have to contain some proviso if the trivial system is to be avoided. Prawitz discusses two possibilities. In one case every premise must be of the form  $\Box A$ , in the other of the form either  $\Box A$  or  $\neg\Box A$ . If we adopt the convention according to which  $\star^n\Sigma = \{\star^n A : A \in \Sigma\}$ , where  $\star$  is any unary propositional operator, then we can give Prawitz's rules the following formulation:

- $(\Box I)_{S4}$  If  $\Gamma \vdash A$ , then  $\Gamma \vdash \Box A$ , provided that, for some set  $\Delta$ ,  $\Gamma = \Box\Delta$ .
- $(\Box I)_{S5}$  If  $\Gamma \vdash A$ , then  $\Gamma \vdash \Box A$ , provided that, for some sets  $\Delta_0$  and  $\Delta_1$ ,  $\Gamma = \Box\Delta_0 \cup \neg\Box\Delta_1$ .

The indexing of the rules is not fortuitous: Prawitz's two systems really coincide with Lewis' **S4** and **S5**. However, it has proved difficult to extend this sort of analysis to the great multitude of other systems of modal logic. It seems fair to say that a deductive treatment congenial to modal logic is yet to be found, for Hilbert systems are not suited for the purpose of actual deduction, and in Hintikka/Kripke systems the alternativeness relation introduces an alien element which, moreover, can become quite unmanageable in special cases.

The situation has given rise to various suggestions. One is that the Gentzen format, which works so well for truth-functional operators, should not be expected to work for intensional operators, which are far from truth-functional. (But then Gentzen works well for intuitionistic logic which is not truth-functional either.) Another suggestion is that the great proliferation of modal logics is an epidemic from which modal logic ought to be cured: Gentzen methods work for the important systems, and the other should be abolished. 'No wonder natural deduction does not work for unnatural systems!' We will now present a deductive system which explores a third alternative: trying to achieve generality at the expense of modifying the Gentzen format (there will be no special E- or I-rules for  $\Box$ ). As far as we know, this system is new; there is a forerunner for some special cases in Segerberg [Segerberg, 1989].

Let us begin by trying to learn from the success of the Hintikka/Kripke venture. This success can perhaps be attributed to a certain division of labour: in Hintikka systems of downward saturated sets the classical conditions govern the relationship between the sets. How can this feature be imitated in the setting of natural deduction? The crux of the matter seems to be that any classically valid argument should remain valid *in any modal context*; the difficulty is to explicate the italicised phrase. The solution seems to be to require that whenever  $\Gamma$  tautologically implies  $A$ , then also  $\Box^n\Gamma \vdash \Box^n A$ . This condition we recognise from Section 6 where it was introduced as the condition that the deducibility relation be strongly modal.

The condition of strong modality may of course be adopted as a new rule in a sequent formulation of our logic. But as a proof-theoretic analysis such a move would not go very far: sequent theories, it would appear, are most naturally understood as meta-logics (theories about deductive systems). However that may be, here is the promised system. First there are the *inference rules* list in Table 3. For each rule in the old system there are now infinitely many rules. It is almost as if each power of  $\Box$  would be an independent operator. As before, we do not state the *deduction rules* but are content to make a number of observations from which they can be reconstructed. We introduce the convention

$${}^n\sqrt{\Gamma} = \{A : \Box^n A \in \Gamma\}.$$

Table 3.

$(\wedge E)^n$	$\frac{\Box^n(A \wedge B)}{\Box^n A}$	$\frac{\Box^n(A \wedge B)}{\Box^n B}$	$(\wedge I)^n$	$\frac{\Box^n A \Box^n B}{\Box^n(A \wedge B)}$
$(\vee E)^n$	$\frac{\Box^n(A \vee B)}{\Box^n B}$	$\frac{(a)_n \quad (b)_n}{C \quad C}$	$(\vee I)^n$	$\frac{\Box^n A}{\Box^n(A \vee B)}$ $\frac{\Box^n B}{\Box^n(A \vee B)}$
$(\rightarrow E)^n$	$\frac{\Box^n(A \rightarrow B) \Box^n A}{\Box^n B}$		$(\rightarrow I)^n$	$\frac{(A)_n}{B}$ $\frac{}{\Box^n(A \rightarrow B)}$
$(\neg E)^n$	$\frac{\Box^n(\neg A) \Box^n A}{\Box^n B}$		$(\neg I)^n$	$\frac{(A)_n}{\neg A}$ $\frac{}{\Box^n \neg A}$
			$(RAA)^n$	$\frac{(\neg A)_n}{A}$ $\frac{}{\Box^n A}$

Notice that the new rules (Table 3) have ‘ $( )_n$ ’, where the old (Table 2) have ‘ $( )$ ’. this new notation also is explained by the observations listed in Table 4. It is easy to check that the deducibility relation defined by this system is classical if  $\Box$  is the only non-Boolean operator. Nor is it difficult to prove that it also satisfies Scott’s Rule (SC2): if  $\Gamma \vdash A$ , then  $\Box \Gamma \vdash \Box A$ . In fact, the system coincides with the minimal normal system **K**.

The given system looks more complicated than the Hilbert type formulation of **K** in Section 6. But for deductive purposes it may be an alternative. If one would like to general modal logic within this framework, different logics would have to be characterised by special axioms. This means giving up the idea of finding characteristic rules for those systems. This is perhaps

Table 4.

$(\wedge E)^n$	If $\Gamma \vdash \Box^n(A \wedge B)$ , then $\Gamma \vdash \Box^n A$ and $\Gamma \vdash \Box^n B$ .
$(\wedge I)^n$	If $\Gamma \vdash \Box^n A$ and $\Gamma \vdash \Box^n B$ , then $\Gamma \vdash \Box^n(A \wedge B)$ .
$(\vee E)^n$	If $\Gamma \vdash \Box^n(A \vee B)$ and ${}^n\sqrt{\Gamma}, A \vdash C$ and ${}^n\sqrt{\Gamma}, B \vdash C$ , then $\Gamma \vdash \Box^n C$ .
$(\vee I)^n$	If $\Gamma \vdash \Box^n A$ or $\Gamma \vdash \Box^n B$ , then $\Gamma \vdash \Box^n(A \vee B)$ .
$(\rightarrow E)^n$	If $\Gamma \vdash \Box^n(A \rightarrow B)$ and $\Gamma \vdash \Box^n A$ , then $\Gamma \vdash \Box^n B$ .
$(\rightarrow I)^n$	If ${}^n\sqrt{\Gamma}, A \vdash B$ , then $\Gamma \vdash \Box^n(A \rightarrow B)$ .
$(\neg E)^n$	If $\Gamma \vdash \Box^n(\neg A)$ and $\Gamma \vdash \Box^n A$ , then $\Gamma \vdash \Box^n B$ .
$(\neg I)^n$	If ${}^n\sqrt{\Gamma}, A \vdash \neg A$ , then $\Gamma \vdash \Box^n \neg A$ .
$(RAA)^n$	If ${}^n\sqrt{\Gamma}, \neg A \vdash A$ , then $\Gamma \vdash \Box^n A$ .

a price worth paying, for—as remarked before—only exceptional systems would seem to be characterisable in terms of reasonably simple rules.

The same point can perhaps be put in the following way. When we go to systems of traditional modal logic stronger than  $\mathbf{K}$ , we should like to preserve classicalness, usually also Scott’s Rule. The best way to do this appears to be to add more in the way of axioms rather than rules. In this manner, modal propositional logics become a bit like theories of ordinary predicate logic. Let  $\Sigma$  be any set of modal formulas closed under substitution (that is,  $A^* \in \Sigma$  whenever  $A^*$  is a substitution instance of some  $A \in \Sigma$ ). Then we define  $\mathbf{L}(\Sigma)$  as the logic got by adopting  $\Sigma$  as a set of new axioms:  $\Gamma \vdash A$  in  $\mathbf{L}(\Sigma)$  iff  $\Gamma \cup \Sigma \vdash A$  in the basic system. It is obvious that  $\mathbf{L}(\Sigma)$  will always be classical. Moreover, if  $\Sigma$  is closed also under necessitation (that is, if  $\Box \Sigma \subseteq \Sigma$ ), then  $\mathbf{L}(\Sigma)$  is a normal logic. In this fashion we preserve more of the Gentzen/Prawitz flavour than the Hintikka/Kripke procedure does, while retaining full generality.

## 10 MODAL ALGEBRAS, FRAMES, GENERAL FRAMES

The sections which follow survey the mainstream of technical modal logic. It is felt that the major results have been fairly represented. However, the selection of secondary results has been decidedly subjective, and another writer might well have chosen different topics. The best unified and detailed presentation in the area is [Goldblatt, 1976], which extends his PhD thesis of 1974 to account for the work of other logicians of that period. A good picture of an earlier stage is given in [Seegerberg, 1971]. The startling difference of content between these two ‘monographs’ reflects the great increase of mathematical sophistication in technical modal logic at that time. This trend was led by Kit Fine, S. K. Thomason and R. I. Goldblatt. A more recent exploitation of algebra in the work of W. J. Blok will not be discussed in detail in this survey.

A *modal algebra*  $\mathfrak{A} = \langle A, 0, 1, -, \cap, \cup, \mathbf{l}, \mathbf{m} \rangle$  consists of a set  $A$  including 0 and 1, with functions  $-, \cap, \cup, \mathbf{l}, \mathbf{m}$  on it which satisfies the conditions that  $\langle A, -, 1, -, \cap, \cup \rangle$  is a Boolean algebra and

$$\mathbf{l}1 = 1, \mathbf{l}(a \cap b) = \mathbf{l}a \cap \mathbf{l}b, \mathbf{m}a = -\mathbf{l} - a,$$

or, equivalently, that

$$\mathbf{m}0 = 0, \mathbf{m}(a \cup b) = \mathbf{m}a \cup \mathbf{m}b, \mathbf{l}a = -\mathbf{m} - a.$$

A valuation  $v$  on  $\mathfrak{A}$  is a function from the propositional formulas to the elements of the algebra which satisfies the conditions

$$\begin{aligned} v(\neg A) &= -v(A), \\ v(A \wedge B) &= v(A) \cap v(B), \\ v(A \vee B) &= v(A) \cup v(B), \\ v(\Box A) &= \mathbf{l}v(A), \\ v(\Diamond A) &= \mathbf{m}v(A). \end{aligned}$$

An algebraic ‘model’  $\langle \mathfrak{A}, v \rangle$  is a modal algebra with a valuation on it, and  $A$  is true or verified in this ‘model’ iff  $v(A) = 1$ . A formula is true in a modal algebra iff it is true in all ‘models’ on that algebra (cf. Section 3).

A *frame*  $\mathfrak{F} = \langle W, R \rangle$  consists of a set  $W$  and a binary relation  $R$  on  $W$ . A valuation  $V$  on  $\mathfrak{F}$  is a function such that  $V(A, x) \in \{T, F\}$  for each propositional formula  $A$  and  $x \in W$ , which satisfies the conditions

$$\begin{aligned} V(\neg A, x) = T &\text{ iff } V(A, x) = F, \\ V(A \wedge B, x) = T &\text{ iff } V(A, x) = T \text{ and } V(B, x) = T, \\ V(A \vee B, x) = T &\text{ iff } V(A, x) = T \text{ or } V(B, x) = T, \\ V(\Box A, x) = T &\text{ iff } \forall y(xRy \rightarrow V(A, y) = T), \\ V(\Diamond A, x) = T &\text{ iff } \exists y(xRy \wedge V(A, y) = T). \end{aligned}$$

A model  $\langle \mathfrak{F}, V \rangle$  is a frame with a valuation on it, and  $A$  is satisfied in it iff

$$V(A, x) = T \text{ for some } x \in W,$$

and is true or verified in it iff

$$V(A, x) = T \text{ for each } x \in @.$$

A formula is true or verified in a frame iff it is true in all models on that frame. (Cf. Section 4.)

A modal logic is *normal* iff it includes all tautologies and the axiom

$$\vdash \Box(P \rightarrow Q) \rightarrow (\Box P \rightarrow \Box Q),$$

and is closed under the rules of substitution for variables, modus ponens, and necessitation,

$$\text{if } \vdash A \text{ then } \vdash \Box A.$$

An alternative to this axiom and necessitation is to take

$$\begin{aligned} &\vdash \Box(P \rightarrow P) \\ &\vdash (\Box P \wedge \Box Q) \rightarrow \Box(P \wedge Q) \end{aligned}$$

and the rule

$$\text{if } \vdash A \rightarrow B \text{ then } \vdash \Box A \rightarrow \Box B,$$

from which

$$\vdash \Box(P \wedge Q) \rightarrow (\Box P \wedge \Box Q)$$

is derivable. (Cf. Section 6.) The minimal normal modal logic is called **K**, and its formulas are true in every modal logic and frame. Well-known formulas which are true in every modal algebra satisfying a corresponding equation, and every frame satisfying a corresponding first-order condition on its relation, are shown in Table 5. Here  $a \leq b$  is an abbreviation for  $a \cap b = a$  or  $a \cup b = b$ . It is convenient to label the extension of **K** with certain axioms by concatenating **K** with their labels, so that the extension of **K** with **T** and **4** is **KT4**, except that **KT** has usually been replaced by **S**. (Cf. Section 7.) Note that the modal algebras verifying **S4** satisfy  $\mathbf{1}a \leq$  and  $\mathbf{1}a = \mathbf{1}a$ , being the closure algebras or interior algebras of McKinsey and Tarski [1944].

When added to **K4**, the formulas in Table 4 are true in every transitive frame satisfying the corresponding condition on its relation. (Here the condition for **·3** is known as connectedness, and the condition for **M** asserts that after each point  $x$  there is a ‘second last’ point  $y$ .) (Of these formulas, **M** was introduced in [McKinsey, 1945], **·3** in [Dummett and Lemmon, 1959], and **Grz** in [Sobinciński, 1964], where it is shown that **T** and **M** are derivable in **K4Grz**. In fact **4** is derivable in **KGrz** by [van Benthem and Blok, 1978].)

A frame  $\mathfrak{F} = \langle W, R \rangle$  determines a modal algebra  $\mathfrak{F}^+$  with carrier  $\mathfrak{B}(W)$ , where  $0 = \emptyset$  and  $1 = W$ ,  $-, \cap, \cup$  are the usual set-theoretic operations,  $\mathfrak{B}(W)$  is the set of subsets of  $W$ , and

$$\begin{aligned} \mathbf{1}_R a &= \{x : \forall y (xRy \rightarrow y \in a)\}, \\ \mathbf{m}_R a &= \{x : \exists y (xRy \wedge y \in a)\}. \end{aligned}$$

Writing  $v(A)$  for  $\{x : V(A, x) = T\}$ , each valuation  $V$  on  $\mathfrak{F}$  determines a subset  $\{v(A) : A \text{ a formula}\}$  of  $\mathfrak{B}(W)$ . This subset is in fact the carrier of a subalgebra of  $\mathfrak{F}^+$ . For many purposes this is the most important point of a valuation, so that it is often preferable to consider *general frames*  $\langle W, R, P \rangle$ , where  $P$  is the carrier of a subalgebra of  $\langle W, R \rangle^+$ . A formula is true or verified in a general frame  $\langle W, R, P \rangle$  iff it is true in each model  $\langle W, R, V \rangle$  for which  $v$  is a function into  $P$ . (General frames were introduced in [Thomason, 1972], though they are foreshadowed in [Makinson, 1970] and in the secondary models of [Bull, 1969; Fine, 1970] and [Kaplan, 1970] for modal

logics with propositional quantifiers.) The construction  $+$  can be extended to general frames  $\mathfrak{F} = \langle W, R, P \rangle$  by taking the carrier of  $\mathfrak{F}^+$  to be  $P$  instead of  $\mathfrak{B}(W)$ .

Table 5.

Label	Formula	Equation	Condition on $R$
<b>T</b>	$\Box P \rightarrow P$	$\mathbf{1}a \leq a$	$\forall x(xRx)$
<b>B</b>	$\Diamond \Box P \rightarrow P$	$\mathbf{m}1a \leq a$	$\forall x \forall y(xRy \rightarrow yRx)$
<b>4</b>	$\Box P \rightarrow \Box \Box P$	$\mathbf{1}a \leq \mathbf{1}1a$	$\forall x \forall y \forall z((xRy \wedge yRz) \rightarrow xRz)$

Table 6.

Label	Formula	Condition on $R$
<b>3</b>	$\Box(\Box P \rightarrow \Box Q) \vee \Box(\Box Q \rightarrow \Box P)$	$\forall x \forall y \forall z((xRy \wedge xRz) \rightarrow (yRz \vee zRy))$
<b>M</b>	$\Box \Diamond P \rightarrow \Diamond \Box P$	$\forall x \exists y(xRy \wedge \forall z \forall w((yRz \wedge yRw) \rightarrow z = -w))$
<b>Grz</b>	$\Box(\Box(P \rightarrow \Box P) \rightarrow P) \rightarrow P$	There is no infinite chain $x_0, x_1, x_2, \dots$ with $x_i R x_{i+1}$ and $x_i \neq x_{i+1}$ , for all $i$ .

A modal algebra  $\mathfrak{A}$  determines a general frame  $\mathfrak{A}_+ = \langle W_{\mathfrak{A}}, R_{\mathfrak{A}}, P_{\mathfrak{A}} \rangle$ , where  $W_{\mathfrak{A}}$  is the set of ultrafilters of  $\mathfrak{A}$ ,

$$xR_{\mathfrak{A}}y \text{ iff } \forall a(a \in y \rightarrow \mathbf{m}a \in x)$$

or, equivalently,

$$\begin{aligned} xR_{\mathfrak{A}}y \text{ iff } \forall a(\mathbf{1}a \in x \rightarrow a \in y), \\ P_{\mathfrak{A}} = \{\{x : a \in x\} : a \in A\}, \end{aligned}$$

i.e. for each element of the modal algebra we take the set of ultrafilters  $x$  containing it. (The filters of  $\mathfrak{A}$  are the subsets  $F$  of  $A$  which satisfy the conditions

$$\begin{aligned} 1 \in F \text{ and not } 0 \in F, \\ \text{if } a, b \in F \text{ then } a \cap b \in F, \\ \text{if } a \in F \text{ and } a \leq b \text{ then } b \in F, \end{aligned}$$

and the ultrafilters  $F$  also satisfy

$$\text{for each } a \in A, \text{ either } a \in F \text{ or } -a \in F$$

note that also not both  $a \in F$  and  $-a \in F$ .) Here we write  $\mathfrak{A}_{\sharp}$  for the underlying frame  $\langle W_{\mathfrak{A}}, R_{\mathfrak{A}} \rangle$ . Note that if  $\mathfrak{A}$  is finite then  $P_{\mathfrak{A}}$  is  $\mathfrak{B}(W_{\mathfrak{A}})$ , and  $\mathfrak{A}_+$  and  $\mathfrak{A}_{\sharp}$  coincide.

Clearly a formula is true in a model  $\langle \mathfrak{F}, V \rangle$  iff it is true in the algebraic ‘model’  $\langle \mathfrak{F}^+, v \rangle$  and hence true in  $\mathfrak{F}$  iff it is true in  $\mathfrak{F}^+$ , since they have the same valuations. It can also be shown that a formula is true in an algebraic ‘model’  $\langle \mathfrak{A}, v \rangle$  iff it is true in  $\langle \mathfrak{A}_\#^+, V \rangle$ , where

$$V(A, x) = T \text{ iff } v(A) \in x.$$

(These constructions and results are due to Lemmon [1966], though they would also have been easy consequences of [Jónsson and Tarski, 1951].) In fact, each modal algebra  $\mathfrak{A}$  is isomorphic to  $(\mathfrak{A}_+)^+$  by similar arguments. Let us consider the properties of  $\mathfrak{A}_+$ . A set  $X \subseteq A$  has the f.i.p. (finite intersection property) iff

$$a_1 \cap \dots \cap a_n \neq 0, \text{ for each } a_1, \dots, a_n \in X.$$

Each set  $X$  with the f.i.p. can easily be extended to a filter, which can in turn be extended to a maximal filter by Zorn’s Lemma. Conversely each subset of a filter has the f.i.p. As a lemma, if  $X$  has the f.i.p. but  $X \cup \{-a\}$  does not, then  $a \in F$ , for each filter  $F$  with  $X \subseteq F$ . It follows immediately that each maximal filter is an ultrafilter. As a second lemma following from the first,  $b \in F$ , for each ultrafilter  $F$  with  $X \subseteq F$ , iff

$$a_1 \cap \dots \cap a_n \leq b, \text{ for some } a_1, \dots, a_n \in X.$$

In both the results above we are concerned with the function  $\phi : A \rightarrow P_{\mathfrak{A}}$  with

$$\phi(a) = \{F : F \text{ an ultrafilter on } \mathfrak{A} \text{ with } a \in F\}.$$

The crucial point is to show that

$$\exists G (FR_{\mathfrak{A}}G \wedge G \in \phi(a)) \text{ iff } F \in \phi(\mathbf{m}a),$$

in order to establish the properties of  $V(\Diamond A, x)$  on  $\mathfrak{A}_+$ , and the properties of  $\mathbf{m}_{R_{\mathfrak{A}}}$  in  $(\mathfrak{A}_+)^+$ . This is immediate from left to right, using the definition

$$FR_{\mathfrak{A}}G \text{ iff } \forall b (b \in G \rightarrow \mathbf{m}b \in F).$$

Going from right to left, suppose that the left-hand side is false, so that

$$\forall G (FR_{\mathfrak{A}}G \rightarrow -a \in G),$$

for the ultrafilter  $F$ . Using the alternative definition

$$FR_{\mathfrak{A}}G \text{ if } \forall b (\mathbf{l}b \in F \rightarrow b \in G)$$

and taking  $X = \{b : \mathbf{l}b \in F\}$ , each ultrafilter  $G$  with  $X \subseteq G$  has  $-a \in G$ . Applying the second lemma above to  $X$  it is easy to show that  $\mathbf{l}(-a) \in F$ , and hence not  $F \in \phi(\mathbf{m}a)$ , as required.



However,  $(\mathfrak{F}^+)_+$  is not in general ‘isomorphic’ to  $\mathfrak{F}$ , for a general frame  $\mathfrak{F}$ . Therefore we need a subclass of the general frames which will include all the general frames  $\mathfrak{A}_+$  and be closed under this pair of operations. In the terminology of [Goldblatt, 1976], given a general frame  $\langle W, R, P \rangle$  write

$$\begin{aligned} Px &= \{S \in P : x \in S\}, \\ MPx &= \{\mathbf{m}_R S : x \in S \wedge S \in P\}. \end{aligned}$$

Then Thomason [1972] defines the conditions

$$\begin{aligned} \text{if } Px = Py \text{ then } x = y & \quad (1\text{-refinement}), \\ \text{if } MPy \subseteq Px \text{ then } xRy & \quad (2\text{-refinement}), \end{aligned}$$

and calls a general frame *refined* when it satisfies both of them. In effect a general frame  $\langle W, R, P \rangle$  has enough propositions in  $P$  to determine  $W$  when it is 1-refined, and enough propositions in  $P$  to determine  $R$  when it is 2-refined. (Kit Fine independently introduced analogous conditions differentiated, tight, and natural for models.) Clearly each general frame  $\mathfrak{A}_+$  determined by a modal algebra  $\mathfrak{A}$  is refined.

As Thomason [1972] shows, for each general frame  $\langle W, R, P \rangle$  there is a refined general frame for which precisely the same formulas are true. One first replaces  $R$  by  $R'$  with

$$xR'y \text{ iff } (\forall S \in P)(y \in S \rightarrow \mathbf{m}_R S \in x),$$

so that  $\langle W, R', P \rangle^+$  is the same as  $\langle W, R, P \rangle^+$  but 2-refinement is satisfied. Then an equivalence relation  $\simeq$  is defined on  $W$  by taking

$$x \simeq y \text{ iff } (\forall S \in P)(x \in S \equiv y \in S).$$

This is a congruence on  $\langle W, R', P \rangle$  in the sense that

$$\text{if } x_1 \simeq x_2 \text{ and } y_1 \simeq y_2 \text{ then } x_1 R' y_1 = x_2 R' y_2.$$

Now the quotient general frame  $\langle W/\simeq, R'/\simeq, P/\simeq \rangle$  with

$$\begin{aligned} W/\simeq &= \{[x] : x \in W\}, \\ [x]R'/\simeq [y] &\text{ iff } xR'y, \\ P/\simeq &= \{\{[x] : x \in S\} : S \in P\}, \end{aligned}$$

is refined, and  $\langle W/\simeq, R'/\simeq, P/\simeq \rangle^+$  is isomorphic to  $\langle W, R', P \rangle^+$ . Thus these two steps yield a refined general frame with an associated modal algebra which is isomorphic to that for the given general frame.

Fine [1975] introduces saturation or compactness conditions on models analogous to  $\cap F \neq \emptyset$ , for each ultrafilter  $F$  of  $\langle W, R, P \rangle^+$ , and

$$\cap\{\mathbf{m}_R S : S \in F\} \subseteq \mathbf{m}_R(\cap F) \quad (2\text{-saturation}).$$

Since each  $x \in W$  generates an ultrafilter  $Px$ , this first condition is equivalent to

$$F = Px, \text{ for some } x \in W \text{ (1-saturation)}$$

for each ultrafilter  $F$  of  $\langle W, R, P \rangle^+$ . Note that applying 2-saturation to the ultrafilter  $Px$  yields

$$\text{if } MPy \subseteq Px \text{ then } \exists z(xRz \wedge Pz = Py) \text{ (2'-saturation).}$$

In Goldblatt [1976] it is shown that 2'-saturation is equivalent to 2-saturation in the presence of 1-saturation, and equivalent to 2-refinement in the presence of 1-refinement. Goldblatt [1976] then introduces the *descriptive* general frames as the refined general frames which also satisfy 1-saturation and, hence, 2-saturation. For each modal algebra  $\mathfrak{A}$  the general frame  $\mathfrak{A}_+$  is descriptive. To see that 1-saturation is satisfied we must consider each ultrafilter  $F$  of  $\langle W_{\mathfrak{A}}, R_{\mathfrak{A}}, P_{\mathfrak{A}} \rangle^+$ , i.e. of  $P_{\mathfrak{A}}$  with members

$$\phi(a) = \{F : F \text{ an ultrafilter of } \mathfrak{A} \text{ with } a \in F\},$$

for each  $a \in A$ . The required  $x \in W_{\mathfrak{A}}$  with  $F = P_{\mathfrak{A}}x$  is  $\{a : \phi(a) \in F\}$ .

It can also be shown that each descriptive general frame  $\mathfrak{F}$  is 'isomorphic' to  $(\mathfrak{F}^+)_+$ , so that the descriptive frames are the required 'duals' of the modal algebras. In Goldblatt [1976] this duality is expressed in terms of category theory, which involves the appropriate morphisms between structures as well as the structures themselves. The appropriate frame morphisms are a slight extension of the pseudo-epimorphisms of Segerberg [1968], which have to be onto. Given frames  $\mathfrak{F} = \langle W, R \rangle$  and  $\mathfrak{F}' = \langle W', R' \rangle$ ,  $\phi : W \rightarrow W'$  is a *frame morphism* iff

$$\begin{aligned} &\text{if } xRy \text{ then } \phi(x)R'\phi(y), \\ &\text{if } \phi(x)R'z \text{ then } \exists y(xRy \wedge \phi(y) = z). \end{aligned}$$

Frame morphisms are extended to models  $\langle W, R, V \rangle$  and general frames  $\langle W, R, P \rangle$  by taking

$$v(P) = \phi^{-1}[v'(P)] = \{x \in W : \phi(x) \in v'(P)\},$$

for each propositional variable  $P$ ,

$$\text{if } S \in P' \text{ then } \phi^{-1}[S] = \{x \in W : \phi(x) \in S\} \in P.$$

As in Segerberg [1968],

$$V(A, x) = T \text{ iff } V'(A', \phi(x)) = T,$$

by an easy induction on the construction of  $A$ . The induction basis uses the condition above on  $V'$ . For the step on  $\Box$ , the first condition on frame

morphisms shows that if  $V(\Box B, x) = F$ , then  $V'(\Box B, \phi(x)) = F$ , and the second condition shows that if  $V'(\Box B, \phi(x)) = F$  then  $V(\Box B, x) = F$ .

Now the descriptive frames  $\mathfrak{F}$  and  $(\mathfrak{F}^+)_+$  can be shown to be frame isomorphic. For each descriptive frame  $\mathfrak{F} = \langle W, R, P \rangle$ , the function  $\theta : W \rightarrow W^{\mathfrak{F}^+}$  with

$$\theta(x) = Px, \text{ for each } x \in W,$$

is a one-one frame morphism from  $\mathfrak{F}$  onto  $(\mathfrak{F}^+)_+$ . To see this,  $\theta$  is one-one because  $\mathfrak{F}$  is 1-refined, and not because  $\mathfrak{F}$  is 1-saturated. Also, by the definition of  $l_R$  and 2-refinement,  $xRy$  iff  $(\forall S \in P)((S \in Py \rightarrow \mathbf{m}_R S \in Px)$  iff  $PxR_{\mathfrak{F}^+}Py$  iff  $\theta(x)R_{\mathfrak{F}^+}\theta(y)$ . To complete the proof that  $\mathfrak{F}$  and  $(\mathfrak{F}^+)_+$  are frame isomorphic, i.e. that  $\theta$  and  $\theta^{-1}$  are general frame morphisms, it can be shown that  $S \in P$  iff  $\theta[S] \in P_{\mathfrak{F}^+}$ .

To establish the category-theoretic contravariant duality, correspondences must be established between homomorphisms of modal algebras and general frame morphisms of descriptive general frames, with the functions applied in opposite directions. Given general frames  $\mathfrak{F} = \langle U, R, P \rangle$ ,  $\mathfrak{G} = \langle V, S, Q \rangle$  and a general frame morphism  $\phi : \mathfrak{F} \rightarrow \mathfrak{G}$ , define  $\phi^+ : \mathfrak{G}^+ \rightarrow \mathfrak{F}^+$  by

$$\phi^+(S) = \phi^{-1}[S], \text{ for each } S \in Q,$$

where  $\phi^{-1}[S] \in P$  by the third condition. It is easy to show that  $\phi^+$  is a homomorphism. Given modal algebras  $\mathfrak{A}, \mathfrak{B}$  and a homomorphism  $\psi : \mathfrak{A} \rightarrow \mathfrak{B}$ , define  $\psi_+ : \mathfrak{B}_+ \rightarrow \mathfrak{A}_+$  by

$$\psi_+(x) = \{a \in A : \psi(a) \in x\}, \text{ for each } x \in W_{\mathfrak{B}}.$$

This set is an ultrafilter in  $W_{\mathfrak{A}}$ , and  $\psi_+$  satisfies the conditions on general frame morphisms. For the first condition, if  $xR_{\mathfrak{B}}y$  and  $\mathbf{1}_a \in \psi_+(x)$  then  $a \in \psi_+(y)$ . For the second condition, if  $\psi_+(x)R_{\mathfrak{A}}z$  then  $\{a : Bla \in x\} \cup \{\psi(b) : b \in z\}$  can be shown to have the f.i.p. Therefore it can be extended to an ultrafilter  $y$ , which satisfies  $xR_{\mathfrak{B}}y$  and  $\psi_+(y) = z$ . For the third condition, if

$$S = \{F : F \text{ an ultrafilter of } \mathfrak{A} \text{ with } a \in F\}$$

in  $P_{\mathfrak{A}}$ , then

$$\psi_+^{-1}[S] = \{G : G \text{ an ultrafilter of } \mathfrak{B} \text{ with } \psi(a) \in G\}$$

in  $P_{\mathfrak{B}}$ .

The category of modal algebras is a variety, and varieties are characterised by being closed under homomorphic images, subalgebras and direct products. So what are the corresponding constructions in the contravariantly dual category of descriptive frames? Frame-morphic images correspond to sub-algebras.

Subframes correspond to homomorphic images, where  $\langle W', R', P' \rangle$  is a *subframe* of  $\langle W, R, P \rangle$  iff  $W'$  is a subset of  $W$  satisfying the condition

$$\text{if } x \in W' \text{ and } xRy \text{ then } y \in W',$$

$R'$  is the restriction of  $R$  to  $W'$ , and  $P'$  is  $\{S \cap W' : S \in P\}$ . The generated submodels  $\langle W_x, R_x, V_x \rangle$  of Segerberg [1970] are a special case of subframes. Here, for  $x \in W$ ,

$$W_x = \{y_n : X R y_1 \wedge \dots \wedge y_{n-1} R y_n, \text{ for some } y_1, \dots, y_{n-1}\},$$

and  $R_x, V_x$  are the restrictions of  $R, V$  to  $RW_x$ . (In the context of Segerberg [1970]  $R$  is transitive, so that it suffices to take  $W_x = \{y : xRy\}$ .) Clearly a formula is true in  $\langle W, R, V \rangle$  iff it is true in all the generated submodels  $\langle W_x, R_x, V_x \rangle$ , a surprisingly important fact as we shall see. Note that if  $\langle W, R, P \rangle$  is refined or descriptive, then so is each  $\langle W_x, R_x, P_x \rangle$ . For 1-saturation use the fact that the ultrafilters of  $\langle W_x, R_x, P_x \rangle^+$  are the restrictions of the ultrafilters of  $\langle W, R, P \rangle^+$  to subsets of  $W_x$ .

Disjoint unions correspond to direct products, in which we consider a set of general frames  $\langle W_i, R_i, P_i \rangle$ , for  $i \in I$ , for which each  $W_i$  and  $W_j$  are disjoint. (This can always be achieved by attaching indices.) The disjoint union  $\langle W, R, P \rangle$  then has  $W = \cup_{i \in I} W_i, R = \cup_{i \in I} R_i$ , and

$$S \in P \text{ iff } S \cap W_i \in P_i, \text{ for each } i \in I.$$

It is easy to show that if each  $\langle W_i, R_i, P_i \rangle$  is refined, then so is their disjoint union. Goldblatt [1976, Section 9] shows that the disjoint union preserves 1-saturation if  $I$  is finite, but not if it is infinite. The attempt to characterise the class of descriptive frames in terms dual to the usual characterisation of varieties fails in view of this point. (Category-theoretic duality is not always as good as it might sound!)

Section 12 of [Goldblatt, 1976] solves this problem by using another characterisation of varieties, as being closed under homomorphic images, subalgebras, finite direct products, and unions of chains. Onto inverse limits correspond to unions of chains, where the inverse limit of a directed set of descriptive frames is a complex construction set out in Section 11 of [Goldblatt, 1976].

Another important construction in varieties is Birkhoff's subdirect product,  $\mathfrak{A}$  being a subdirect product of the modal algebras  $\mathfrak{A}_i$  with  $i \in I$  iff it is isomorphic to a subalgebra of their direct product which has the following property. Since  $\mathfrak{A}$  is a subalgebra of  $\prod_{i \in I} \mathfrak{A}_i$ , there is a one-one homomorphism  $\iota$  from  $\mathfrak{A}$  into  $\prod_{i \in I} \mathfrak{A}_i$ . For each  $i \in I$  there is a projection  $\pi_i$  from  $\prod_{i \in I} \mathfrak{A}_i$  onto  $\mathfrak{A}_i$ . The condition on the subdirect product is that the homomorphisms  $\pi_i \circ \iota$  from  $\mathfrak{A}$  into each  $\mathfrak{A}_i$  be onto, so that each  $\mathfrak{A}_i$  is a homomorphic image of  $\mathfrak{A}$ . Using this condition it is easy to show that a

formula is true in  $\mathfrak{A}$  iff it is true in each  $\mathfrak{A}_i$ . Each homomorphic image of a modal algebra  $\mathfrak{A}$  is isomorphic to a quotient  $\mathfrak{A}/F$ , where  $F$  is an open filter of  $\mathfrak{A}$ , i.e. a filter satisfying the condition

$$\text{if } a \in F \text{ then } \mathbf{1}a \in F.$$

The quotient is defined by taking the equivalence relation

$$a \simeq b \text{ iff } (-a) \cup (a \cup (-b)) \in F$$

and then taking  $\mathfrak{A}/F$  to be  $\{[a] : a \in A\}$  with  $\mathbf{1}[a] = [\mathbf{1}a]$ , etc. In view of this we can restrict attention to  $\mathfrak{A}_i$ 's of the form  $\mathfrak{A}/F_i$  for  $F_i$  an open filter of  $\mathfrak{A}$ .

Birkhoff defined a modal algebra  $\mathfrak{A}$  to be *subdirectly reducible* iff it is a subdirect product of quotients  $\mathfrak{A}/F_i$  with  $F_i$  nontrivial, and showed that every modal algebra is subdirectly reducible to subdirectly irreducible algebras. If some nonunit element  $a$  of  $\mathfrak{A}$  is in every nontrivial open filter  $F$  then  $[a] = [1]$  in each  $\mathfrak{A}/F_i$ , so that  $\mathfrak{A}$  cannot be a subalgebra of  $\prod_{i \in I} \mathfrak{A}/F_i$ . Thus  $v$  is subdirectly irreducible already. Otherwise each non-unit member  $a$  of  $\mathfrak{A}$  lies outside some nontrivial open filter, and applying Zorn's Lemma yields a (nontrivial) maximal open filter  $F_a$  among those not containing  $a$ . Now  $\mathfrak{A}$  is subdirectly reducible to the  $\mathfrak{A}/F_a$ 's, noting that if  $b \neq c$  and  $a = ((-b) \cup c) \cap (b \cup (-c)) \neq 1$  then  $[b] \neq [c]$  in  $\mathfrak{A}/F_a$ . Here each  $\mathfrak{A}/F_a$  is subdirectly irreducible, since  $[a] \in F$  for each nontrivial filter  $F$  of  $\mathfrak{A}/F_a$  by the maximality of  $F_a$  among the open filters of  $\mathfrak{A}$  not containing  $a$ .

In view of Birkhoff's theorem, we can restrict attention to modal algebras with some nonunit element in every nontrivial open filter, when verifying formulas in a modal logic. (The importance of this result in modal logic lies in its use in the recent work of W. J. Blok.) In a closure or interior algebra, an open filter is determined by its open elements, so that a closure or interior algebra is subdirectly irreducible iff it has a maximum nonunit open element, or equivalently, a minimum nonzero closed element. In such an algebra,

$$\text{if } \mathbf{1}a \cup \mathbf{1}b = 1 \text{ then } \mathbf{1}a = 1 \text{ or } \mathbf{1}b = 1,$$

a condition we shall use later. It is easy to see that a modal algebra  $\langle W, R \rangle^+$  is subdirectly reducible to the algebras  $\langle W_x, R_x \rangle^+$  for  $x \in W$ , which are subdirectly irreducible.

In view of the contravariant duality between modal algebras and descriptive general frames, what theorem for the latter corresponds to Birkhoff's Theorem? Note that the lack of a disjoint union of infinitely many descriptive frames will block a dualisation of Birkhoff's proof. Let us say that a general frame  $\mathfrak{F}$  is the subdirect sum of general frames  $\mathfrak{F}_i$  with  $i \in I$  iff it is a frame-morphic image of their disjoint union  $\Sigma_{i \in I} \mathfrak{F}_i$  which has the following property. Since  $\mathfrak{F}$  is a frame-morphic image of  $\Sigma_{i \in I} \mathfrak{F}_i$  there is a frame

morphism  $\phi$  from  $\Sigma_{i \in I} \mathfrak{F}_i$  onto  $\mathfrak{F}$ . For each  $i \in I$  there is embedding frame morphism  $\iota_i$  from  $\mathfrak{F}_i$  into  $\Sigma_{i \in I} \mathfrak{F}_i$ . The condition on the subdirect sums is that the frame morphisms  $\phi \circ \iota_i$  from each  $\mathfrak{F}_i$  into  $\mathfrak{F}$  be embedding, so that each  $\mathfrak{F}_i$  is isomorphic to a subframe of  $\mathfrak{F}$ . In view of this we can restrict attention to  $\mathfrak{F}_i$ 's which are subframes of  $\mathfrak{F}$ . Again it is easy to show that a formula is true in  $\mathfrak{F}$  iff it is true in each  $\mathfrak{F}_i$ . Say that a general frame is subdirectly reducible iff it is a subdirect sum of its proper subframes. Then it is clear that a general frame is subdirectly reducible to its generated subframes, and that these are subdirectly irreducible. So although the disjoint union of descriptive frames is not usually descriptive, Birkhoff's deep result for modal algebras is analogous to the easy, known result that a formula is true in a descriptive general frame iff it is true in its generated subframes, which are again descriptive!

## 11 CANONICAL STRUCTURES

So far we have not constructed any modal algebras or frames. given a normal modal logic  $\mathbf{L}$ , define an equivalence relation  $\simeq_{\mathbf{L}}$  on formulas by taking

$$B \simeq_{\mathbf{L}} C \text{ iff } \vdash_{\mathbf{L}} B \equiv C.$$

Then the *canonical modal algebra*  $\mathfrak{A}_{\mathbf{L}}$  is constructed by taking

$$\begin{aligned} A_{\mathbf{L}} &= \{[B]_{\mathbf{L}} : B \text{ a formula}\}, \\ 0 &= [P \wedge \neg P]_{\mathbf{L}} \text{ and } 1 = [(\neg P) \vee P]_{\mathbf{L}}, \\ \neg[B]_{\mathbf{L}} &= [\neg B]_{\mathbf{L}}, \\ [B]_{\mathbf{L}} \cap [C]_{\mathbf{L}} &= [B \wedge C]_{\mathbf{L}}, \\ [B]_{\mathbf{L}} \cup [C]_{\mathbf{L}} &= [B \vee C]_{\mathbf{L}}, \\ l[B]_{\mathbf{L}} &= [\Box B]_{\mathbf{L}}, \\ \mathbf{m}[B]_{\mathbf{L}} &= [\Diamond B]_{\mathbf{L}}. \end{aligned}$$

That  $\mathfrak{A}_{\mathbf{L}}$  is indeed a modal algebra is easily shown using the defining axioms and rules of normal modal logics. Defining a valuation  $v_{\mathbf{L}}$  by

$$v_{\mathbf{L}}(B) = [B]_{\mathbf{L}}, \text{ for each formula } B,$$

we have

$$v_{\mathbf{L}}(B) = 1 \text{ iff } B \in \mathbf{L},$$

so that the canonical algebraic 'model'  $\langle \mathfrak{A}_{\mathbf{L}}, v_{\mathbf{L}} \rangle$  characterises the normal modal logic  $\mathbf{L}$ . Further, for each valuation  $v$  on  $\mathfrak{A}_{\mathbf{L}}$ ,  $v(B)$  is  $[C]_{\mathbf{L}}$  for some substitution instance  $C$  of  $B$ , so that  $B$  is true in  $\mathfrak{A}_{\mathbf{L}}$  iff it is in  $\mathbf{L}$ .

Given a normal modal logic  $\mathbf{L}$ , a set  $X$  of formulas is inconsistent iff  $\vdash_{\mathbf{L}} \neg(A_1 \wedge \dots \wedge A_n)$ , for some  $A_1, \dots, A_n \in X$ , and is consistent otherwise. (Note the analogy between consistency and the f.i.p. The existence of maximal

consistent sets is proved with Zorn's Lemma, just as for that of maximal filters. However, if  $\mathbf{L}$  has only countably many propositional variables, then a more elementary construction due to Henkin can be used.) Define the *canonical frame*  $\langle W_{\mathbf{L}}, R_{\mathbf{L}} \rangle$  by taking  $W_{\mathbf{L}}$  to be the set of maximal consistent set of formulas, and taking

$$FR_{\mathbf{L}}G \text{ iff } \forall A(A \in G \rightarrow \diamond A \in F)$$

or, equivalently,

$$FR_{\mathbf{L}}G \text{ iff } \forall A(\Box A \in F \rightarrow A \in G).$$

Note the analogy with the construction of the frame  $\mathfrak{A}_{\mathfrak{A}}$  from a modal algebra  $\mathfrak{A}$ . Define a valuation  $V_{\mathbf{L}}$  by taking

$$V_{\mathbf{L}}(B, F) = T \text{ iff } B \in F, \text{ for each formula } B,$$

a definition which is shown to be sound by an induction on the construction of  $B$ . For the induction step on  $B = \diamond C$  it must be shown that

$$\exists G(FR_{\mathbf{L}}G \wedge G \in v_{\mathbf{L}}(C)) \text{ iff } F \in v_{\mathbf{L}}(\diamond C).$$

This proof is exactly analogous to the one used when showing that  $(\mathfrak{A}_+)^+$  is isomorphic to  $\mathfrak{A}$ , using the defining axioms and rules of normal modal logics. Now

$$v_{\mathbf{L}}(B, F) = T, \text{ for each } F \in W_{\mathbf{L}}, \text{ iff } B \in \mathbf{L},$$

since each consistent set of formulas can be extended to a member of  $W_{\mathbf{L}}$ , so that the canonical model  $\langle W_{\mathbf{L}}, R_{\mathbf{L}}, V_{\mathbf{L}} \rangle$  characterises the normal modal logic  $\mathbf{L}$ . Taking

$$P_{\mathbf{L}} = \{v_{\mathbf{L}}(B) : B \text{ a formula}\}$$

gives the canonical general frame  $\langle W_{\mathbf{L}}, R_{\mathbf{L}}, P_{\mathbf{L}} \rangle$ . For each valuation  $V$  on this frame,  $v(B)$  is  $v_{\mathbf{L}}(C)$ , for some substitution instance  $C$  of  $B$ , so that  $B$  is true in  $\langle W_{\mathbf{L}}, R_{\mathbf{L}}, P_{\mathbf{L}} \rangle$  iff it is in  $\mathbf{L}$ . In fact  $\langle W_{\mathbf{L}}, R_{\mathbf{L}}, P_{\mathbf{L}} \rangle$  is  $\mathfrak{A}_{\mathbf{L}^+}$ , so that it has a descriptive general frame characterising  $\mathbf{L}$ .

It does not follow that the canonical frame  $\langle W_{\mathbf{L}}, R_{\mathbf{L}} \rangle$  itself characterises the normal modal logic  $\mathbf{L}$ . Nonetheless, in a number of cases it can be shown that  $R_{\mathbf{L}}$  satisfies some condition for frames to verify  $\mathbf{L}$ , so that  $\langle W_{\mathbf{L}}, R_{\mathbf{L}} \rangle$  does characterise  $\mathbf{L}$ . In particular, the canonical frames for **KT**, **KB**, **K4**, and the logics obtained by combining these axioms, satisfy the first-order conditions on  $R$  given in Section 10. (These completeness proofs were given independently in [Lemmon, 1977], written in 1966, and in [Makinson, 1966].) These partial results suggest a number of important problems which have provided the main motivation for modal logic in the 1970s. Under what conditions is a formula true on the underlying frame  $\langle W, R \rangle$  when it is true on a model  $\langle W, R, V \rangle$  or a general frame  $\langle W, R, P \rangle$ ? Are there logics which

are not characterised by the ordinary frames which verify them? What is the relationship between modal axioms and first-order conditions on  $R$  in the frames  $\langle W, R \rangle$ ? Are there formulas not characterised by the class of frames satisfying some first-order condition? Generalising the problem of completeness, often a problem can be easily solved for descriptive general frames by their duality with the variety of modal algebras, and the difficulty lies in transferring the problem to the underlying frames. We shall return to answers to these questions after studying various particular logics which have attracted attention.

## 12 THE F. M. P. AND FILTRATIONS

A logic  $\mathbf{L}$  is said to have the *f.m.p.* (*finite model property*) iff, for each formula  $A$ ,  $\vdash_{\mathbf{L}} A$  iff  $A$  is true in each finite modal algebra or frame which verifies the formulas of  $\mathbf{L}$ . Thus in showing that  $\mathbf{L}$  has the f.m.p. we must find, for each nonthesis  $A$ , a finite modal algebra or frame which verifies  $\mathbf{L}$  but does not verify  $A$ . Note that modal algebras and frames are interchangeable here. For if  $\mathfrak{F}$  is a finite frame, then of course  $\mathfrak{F}^+$  is a finite modal algebra, and if  $\mathfrak{A}$  is a finite modal algebra, then  $\mathfrak{A}_\# = \mathfrak{A}_+$  is a finite frame. The f.m.p. is important, among other reasons, for giving decidability to a finitely axiomatised normal modal logic. For as Harrop pointed out, we can construct the countably many finite models in some order, checking each one for verifying the finitely many axioms and the given formula  $A$ . Again a problem of independence is raised, which will be considered in a later section: are there logics which are characterised by frames, but not by the finite frames which verify them? (The position of the logics characterised by one finite model in the lattice of modal logics is investigated in detail in [Blok, 1980]. The normal modal logics immediately below these, which also have the f.m.p., are the subject of [Block, 1980a].)

We now consider a pair of methods for constructing finite modal algebras and frames from given structures, both known as filtration. Consider an algebraic ‘model’  $\langle \mathfrak{A}, v \rangle$  and a formula  $A$  with  $v(A) \neq 1$ . Let  $\{A_1, \dots, A_n\}$  be a finite set of formulas including  $A$  and closed under subformulas, and let  $\langle B, 0, 1, -, \cap, \cup \rangle$  be the subalgebra of  $\langle A, 0, 1, -, \cap, \cup \rangle$  generated by  $\{v(A_1), \dots, v(A_n)\}$ , noting that it is non-trivial and finite. (Usually  $A_1, \dots, A_n$  are  $A$  and its subformulas, but sometimes some larger set is preferable.) This Boolean algebra is extended to a finite modal algebra  $\mathfrak{B} = \langle B, 0, 1, -, \cap, \mathbf{l}', \mathbf{m}' \rangle$  by taking

$$\begin{aligned} \mathbf{l}'b &= \cup \{ \mathbf{l}a \in B : a \in B \wedge a \leq b \}, \\ \mathbf{m}'b &= \cap \{ \mathbf{m}c \in B : c \in B \wedge b \leq c \}, \end{aligned}$$

(In the case of a closure or interior algebra  $\mathfrak{A}$ ,  $\mathbf{m}$  is determined by the closed elements of  $\mathfrak{A}$  and  $\mathbf{l}$  by the open elements. Therefore it suffices to take  $\mathbf{l}'b$



to be the union of the open elements of  $B$  contained by  $b$ , and take  $\mathbf{m}'b$  to be the intersection of the closed elements of  $B$  containing  $b$ .) In particular,

$$\begin{aligned} \text{if } \mathbf{l}b \in B \text{ then } \mathbf{l}'b &= \mathbf{l}b, \\ \text{if } \mathbf{m}b \in B \text{ then } \mathbf{m}'b &= \mathbf{m}b, \end{aligned}$$

for each  $b \in B$ . Now  $\mathfrak{B}$  is indeed a modal algebra, satisfying

$$\begin{aligned} \mathbf{l}'1 &= 1 \text{ and } \mathbf{l}'(a \cap b) = \mathbf{l}'a \cap \mathbf{l}'b, \\ \mathbf{m}'0 &= 0 \text{ and } \mathbf{m}'(a \cup b) = \mathbf{m}'a \cup \mathbf{m}'b, \end{aligned}$$

using distributivity and the fact that  $\mathfrak{A}$  satisfies these conditions. Construct a valuation  $w$  on  $\mathfrak{B}$  by taking  $w(P) = v(P) \cap B$ , for each propositional variable  $P$  in  $A_1, \dots, A_n$ , and applying the defining conditions for valuations. We now have

$$a(A_i) = v(A_i) \text{ for } i = 1, \dots, n,$$

so that  $w(A) \neq 1$  in the *filtered* algebraic ‘model’  $\langle \mathfrak{B}, w \rangle$ .

It is not in general true that  $\langle \mathfrak{B}, w \rangle$ , let alone  $\mathfrak{B}$ , verifies a logic  $\mathbf{L}$  verified by  $\mathfrak{A}$ . Nonetheless, in a number of cases it can be shown that each filtration  $\mathfrak{B}$  of  $\mathfrak{A}$  satisfies some condition for modal algebras to verify  $\mathbf{L}$ . In particular, filtrations of algebraic ‘models’ verifying  $\mathbf{KT}$ ,  $\mathbf{KB}$ ,  $\mathbf{Kr}$ , and the logics obtained by combining these axioms, again satisfy the equations given in Section 10. It follows that these logics have the f.m.p. and are decidable, being characterised by the filtrations of their canonical modal algebras. (This technique was introduced in [McKinsey, 1941], and extended in [Lemmon, 1966], to establish many decidability results.)

Now consider a model  $\langle W, R, V \rangle$  and a formula  $A$  with  $v(A) \neq W$ . Again let  $\{A_1, \dots, A_n\}$  be a finite set of formulas including  $A$  and closed under subformulas. Define an equivalence relation  $\simeq$  on  $W$  by taking

$$x \simeq y \text{ iff } V(A_i, x) = V(A_i, y), \text{ for } i = 1, \dots, n,$$

so that  $W$  is partitioned into a finite set  $W'$  of equivalence classes  $[x]$  under  $\simeq$ . Consider finite frames  $\langle W', R' \rangle$  satisfying the conditions

$$\begin{aligned} \text{if } xRy \text{ then } [x]R'[y], \\ \text{if } [x]R; [y] \text{ then [if } V(A_i, x) = T, \text{ for } A_i = \Box A_j, \\ \text{then } V(A_j, y) = T], \text{ for } i = 1, \dots, n. \end{aligned}$$

(A suitable condition in terms of  $\Diamond$  could equally well be used.) There are a number of relations  $R'$  on  $W'$  which satisfy these conditions, e.g.  $\bar{R}$  with

$$\begin{aligned} [x]\bar{R}[y] \text{ iff [if } V(A_i, x) = T, \text{ for } A_i = \Box A_j, \\ \text{then } V(A_j, y) = T], \text{ for } i = 1, \dots, n. \end{aligned}$$

This relation satisfies the first conditions, since if  $xRy$  then the right-hand side of the defining condition holds for all formulas  $B = \Box C$ . This is in fact

the largest such relation  $R'$ . The smallest is the intersection  $\underline{R}$  of all such relations, which again satisfies the two conditions. Construct a valuation  $V'$  on  $\langle W', R' \rangle$  by taking  $V; (P, [x]) = V(P, x)$  for each propositional variable  $P$  in  $A_1, \dots, A_n$ , and applying the defining conditions for valuations. It can now be shown that

$$V'(A_i, [x]) = V(A_i, x), \text{ for } i = 1, \dots, n,$$

by induction on the construction of formulas, so that  $v'(A) \neq W'$  in the *filtered* model  $\langle W', R', V' \rangle$ . for the induction step on  $\Box$ , consider  $A_i = \Box A_j$ . If  $V(\Box A_j, x) = T$  and  $[x]R'[y]$  then  $V(A_j, y) = T$  by the second condition on  $R'$ , and  $V'(A_j[y]) = T$  by the induction hypothesis. Applying this to each  $[y]$  we have  $V'(\Box A_j, [x]) = T$ . If  $V'(\Box A_j, [x]) = T$  and  $xy$ , then  $[x]R'[y]$  by the first condition on  $R$ , so that  $V'(A_j, [y]) = T$  and  $V(A_j, y) = T$  by the induction hypothesis. Applying this to each  $y$  we have  $V(\Box A_j, x) = T$ .

Again it is not in general true that  $\langle W', R', V' \rangle$ , let alone  $\langle W', R' \rangle$ , verifies a logic  $\mathbf{L}$  verified by  $\langle W, R, V \rangle$ . Nonetheless, in a number of cases it can be shown that  $R'$  satisfies some condition for frames to verify  $\mathbf{L}$ . In particular filtrations  $\langle W', \bar{R}, \bar{V} \rangle$  of models verifying  $\mathbf{KT}$ ,  $\mathbf{KB}$ ,  $\mathbf{K4}$ , and the logics obtained by combining these axioms, again satisfy the first-order conditions on  $R$  given in Section 10. This gives alternative proofs of the decidability of these logics. (The construction  $\langle W', \bar{R}, \bar{V} \rangle$  was introduced in [Lemmon, 1977] and was generalised to  $\langle W', R', V' \rangle$  in [Seegerberg, 1968].) In many more cases a further step after filtration, or a variation on the construction  $\langle W', \bar{R}, \bar{V} \rangle$  to suit the axioms involved, will yield a finite frame  $\langle W', R' \rangle$  verifying the logic concerned. We shall see some of these techniques in the following sections.

### 13 UNRAVELLING AND BULLDOZING

(The technique of unravelling was introduced in [Dummett and Lemmon, 1959] and used extensively in [Sahlqvist, 1975], apparently without knowledge of the earlier paper.) Consider a frame  $\langle W, R \rangle$  which is generated by  $w_0 \in W$ , so that  $w_0 R w_1, \dots, w_{n-1} R w_n$ , for some  $w_1, \dots, w_{n-1}$ , for each other  $w_n \in W$ . Construct a new frame  $\langle W^*, R^* \rangle$  by taking

$$\begin{aligned} \langle w_0, \dots, w_n \rangle \in W^* &\text{ iff} \\ &w_1, \dots, w_n \in W \text{ and } w_0 R w_1, \dots, w_{n-1} R w_n, \\ \langle w_0, \dots, w_m \rangle R^* \langle w_0, \dots, w_n \rangle &\text{ iff} \\ &\langle w_0, \dots, w_n = \langle w_0, \dots, w_m \rangle * \langle w_n \rangle. \end{aligned}$$

Thus  $R$  has been unravelled in the sense that if  $u_{n-1} R w_n$  and  $v_{n-1} R w_n$  then  $w_n$  is replaced by  $\langle w_0, \dots, u_{n-1}, w_n \rangle$  and  $\langle w_0, \dots, v_{n-1}, w_n \rangle$  with  $\langle w_0, \dots, u_{n-1} \rangle R^* \langle w_0, \dots, u_{n-1}, w_n \rangle$  and  $\langle w_0, \dots, v_{n-1} \rangle R^* \langle w_0, \dots, v_{n-1}, w_n \rangle$ .

Unravelling is extended to models  $\langle W, R, V \rangle$  by taking  $V^*(P, \langle w_0, \dots, w_n \rangle) = V(P, w_n)$  for each propositional variable  $P$ , and applying the defining conditions for valuations. It is easy to show that

$$V^*(A, \langle w_0, \dots, w_n \rangle) = V(A, w_n), \text{ for each formula } A,$$

by induction on the construction of  $A$ . Since  $\mathbf{K}$  is characterised by the finite frames using filtrations, it is now characterised by the unravelled frames. Note that these unravelled frames are irreflexive, asymmetrical, and intransitive. Therefore none of these conditions characterise a proper extension of  $\mathbf{K}$ .

A frame  $\langle W, R \rangle$  could be defined to be a tree iff there is  $w_0 \in W$  and a relation  $S$  on  $W$  satisfying the conditions, for each  $w_n \in W$  other than  $w_0$ , only one  $w_{n-1} \in W$  with  $w_{n-1}Sw_n$ , for some  $w_1, \dots, w_{n-1} \in W$ ; there is only one  $w_{n-1} \in W$  with  $w_{n-1}Sw_n$  and  $w_mRw_n$  if  $w_mSw_{m+1}, \dots, w_{n-1}Sw_n$ , for some  $Rw_{m+1}, \dots, w_{n-1} \in W$ . A tree could be reflexive or irreflexive. Then trees could be obtained by taking the transitive closures of unravelled frames, with or without the reflexive closure as required. (Sahlqvist [1975] uses a more general notion of tree, and proves a number of results concerning them.)

The *clusters* of a transitive frame  $\langle W, R \rangle$  are defined in [Segeberberg, 1971] to be the equivalence classes of  $W$  under the equivalence relation

$$x \simeq y \text{ iff } (xRy \wedge yRx) \vee x = y.$$

Clusters are divided into three kinds: proper, with at least two elements, all reflexive; simple, with one reflexive element; and degenerate with one irreflexive element. Note that when a nondegenerate cluster is unravelled, it will give rise to many branches of  $\langle W^*, R^* \rangle$  in which the members of the cluster are repeated. Thus unravelling imposes asymmetry on frames, sometimes without losing the property of characterising a given logic.

Another technique for removing nondegenerate clusters and so imposing asymmetry is the bulldozing of Segerberg [1970]. Let us suppose that the logic concerned is an extension of  $\mathbf{K4}$  which has countably many propositional variables  $P_0, P_1, P_2, \dots$  and consider a generated transitive frame  $\langle W, R \rangle$ . Construct a new frame  $\langle W^0, R^0 \rangle$  by first replacing each nondegenerate cluster  $C$  of  $W$  by

$$C^0 = \{ \langle x, i \rangle : x \in C \wedge i = 0, 1, 2, \dots \},$$

and replacing each degenerate cluster  $C = \{x\}$  of  $W$  by  $\{ \langle x, 0 \rangle \}$ , to obtain  $W^0$ . Define  $R^0$  on  $W^0$  by taking

$$\begin{aligned} \langle x, i \rangle R^0 \langle y, j \rangle \text{ iff either not } x \simeq y \text{ and } xRy \\ \text{or } x \simeq y \text{ and } i < j \text{ or } x \simeq y \text{ and } i = j \text{ and } xR_C y, \end{aligned}$$

where  $r_C$  is an arbitrary strict ordering of the proper cluster  $C$  with  $x, y \in C$ . Thus each nondegenerate cluster  $C$  of  $W$  is ‘bulldozed’ into an infinite set  $C^0$  on which  $R^0$  is a strict linear ordering. In  $\langle C^0, R^0 \rangle$  a copy  $\langle y, j \rangle$  of  $y$  occurs after each copy  $\langle x, i \rangle$  of  $x$ , for each  $x, y \in C$ . If  $\langle W, R \rangle$  is reflexive, so that there are no degenerate clusters, modify the construction as follows to make  $\langle W^0, R^0 \rangle$  reflexive as well. Form  $C^0$  as above only for proper clusters  $C$ , and replace simple clusters  $C = \{x\}$  by  $C^0 = (\langle x, 0 \rangle)$ ; and add the clause ‘or  $x = y$ ’ to the right- hand side of the definition of  $R_0^0$ . In this case each proper cluster  $C$  is ‘bulldozed’ into an infinite set  $C_0$  on which  $R_0$  is a linear ordering.

Bulldozing is extended to models  $\langle W, R, V \rangle$  by taking

$$V^0(p_j, \langle x, i \rangle) = V(p_j, x), \text{ for } j = 0, 1, 2, \dots,$$

and applying the defining conditions for valuations. Now

$$V^0(A, \langle x, i \rangle) = V(A, x), \text{ for each formula } A$$

by induction on the construction of  $A$ . (For the induction step on  $\Box$ ,  $V^0(\Box B, \langle x, i \rangle) = F$  iff  $V^0(B, \langle y, j \rangle) = F$ , for some  $\langle y, j \rangle \in W^0$  with  $\langle x, i \rangle R^0 \langle y, j \rangle$ , iff  $V(B, y) = F$ , for some  $y \in W$  with  $\langle x, i \rangle R^0 \langle y, j \rangle$ , (by the induction hypothesis) iff  $V(B, y) = F$ , for some  $y \in W$  with  $xRy$ , (by the definition of  $R^0$  if not  $x \simeq y$ , and by a remark above if  $x \simeq y$ ) iff  $V(\Box B, x) = F$ .)

Now consider any normal modal logic  $\mathbf{L}$  containing **S4.3**. First we shall use

$$\vdash_{\mathbf{L}} \Box(\Box A \rightarrow \Box B) \vee \Box(\Box B \rightarrow \Box A)$$

to show that the canonical frame  $\langle W_{\mathbf{L}}, R_{\mathbf{L}} \rangle$  is connected with

$$\forall x \forall y \forall z ((xR_{\mathbf{L}}y \wedge xR_{\mathbf{L}}z) \rightarrow (yR_{\mathbf{L}}z \vee zR_{\mathbf{L}}y)).$$

Let us suppose that we have maximal consistent sets  $F, G, H$  of  $\mathbf{L}$  with  $FR_{\mathbf{L}}G$ ,  $FR_{\mathbf{L}}H$  but not  $GR_{\mathbf{L}}H$  and not  $HR_{\mathbf{L}}G$ , and obtain a contradiction. Since not  $GR_{\mathbf{L}}H$  there is some  $\Box A \in G$  with not  $A \in H$ , and since not  $HR_{\mathbf{L}}G$  there is some  $\Box B \in H$  with not  $B \in G$ . Just as maximal filters are ultrafilters, it can be shown that a maximal consistent set  $F$  satisfies

$$A \in F \text{ or } \neg A \in F, \text{ for each formula } A.$$

It is easy to deduce that

$$\text{if } A \vee B \in F \text{ then } A \in F \text{ or } B \in F, \text{ for all formulas } A, B.$$

Therefore

$$\begin{aligned} &\vdash_{\mathbf{L}} \Box(\Box A \rightarrow \Box B) \vee \Box(\Box B \rightarrow \Box A) \\ &\text{implies } \Box(\Box A \rightarrow \Box B) \in F \text{ or } \Box(\Box B \rightarrow \Box A) \in F \\ &\text{implies } \Box A, \Box A \rightarrow \Box B \in G \text{ or } \textit{square}B, \Box B \rightarrow \Box A \in H \\ &\text{implies } \Box B \in G \text{ or } \Box A \in H \\ &\text{implies } B \in G \text{ or } A \in H \end{aligned}$$

(since  $\vdash_{\mathbf{L}} \Box P \rightarrow P$ )—the required contradiction.

The canonical frame for  $\mathbf{L}$  is also reflexive and transitive. Clearly its generated subframes  $\langle W_{\mathbf{L}}, R_{\mathbf{L}x} \rangle$  satisfy  $\forall y \forall z (yRz \vee zRy)$ , and bulldozing adds

$$\forall y \forall z (y \neq z \rightarrow \neg(yRz \wedge zRy))$$

to these conditions in  $\langle W_{\mathbf{L}z}^0, R_{\mathbf{L}x}^0 \rangle$ , so that  $R_{\mathbf{L}x}^0$  is a linear ordering in the full sense. Often such frames still verify  $\mathbf{L}$ , so that they characterise it, in particular when  $\mathbf{L}$  is **S4.3** itself. (Segerberg [1970] proves the analogous result for extensions  $\mathbf{L}$  of **K4.3**, using filtrations of the canonical frame which are connected although the canonical frame itself is not. Many other results along these lines are obtained in [Segerberg, 1970; Segerberg, 1971] and [Sahlqvist, 1975].)

#### 14 S4.1 AND S4GRZ

(**K4.1** = **K4M** and **S4.1** = **KT4M** were shown to be characterised by frames satisfying the appropriate conditions in [Lemmon, 1977], written in 1966, and **S4.1** was shown to be characterised by the appropriate finite frames in [Segerberg, 1968]. Independently Bull [1967] gave an algebraic proof of the f.m.p. for **S4.1**, and described a characteristic frame for it. The extension **S4 Grz** of **S4.1** was shown to be characterised by the appropriate finite frames in [Segerberg, 1971].)

Bull [1967] begins by showing that **S4.1** can also be axiomatised by extending **S4** with either of the rules

$$\begin{aligned} &\text{if } \vdash \Diamond A, \vdash \Diamond B \text{ then } \vdash \Diamond(A \wedge B), \\ &\text{if } \vdash \Diamond A, \text{ then } \vdash \Diamond \Box A. \end{aligned}$$

Although a filtration  $\mathfrak{B}$  of the canonical modal algebra for **S4.1** may not verify these rules, an extension  $\mathfrak{B}^+$  of  $\mathfrak{B}$  can be constructed which does. (Thinking in terms of  $\langle W, R \rangle^+$ , where  $R$  satisfies the conditions in Section 10 for verifying **S4.1**, we need to isolate the  $R$ -last points of  $W$ . This is achieved by the following trick.) Taking  $a_B = \cup\{\mathbf{m}b - b : b \in B\}$ , where the join and  $\mathbf{m}$  are that of  $\mathfrak{A}_{\mathbf{S4.1}}$ , we shall consider separately what happens in  $a_B$  and what happens in  $\neg a_B$  (the set of  $R$ -last points, in effect). Let  $\langle B^+, 0, 1, -, \cap, \cup, \mathbf{I}', \mathbf{m}' \rangle$  be the filtration of  $\mathfrak{A}_{\mathbf{S4.1}}$  generated by  $B \cup \{a_B\}$ , and define

$$\begin{aligned} \mathbf{I}^+b &= (\mathbf{I}'b \cap a_B) \cup (b - a_B), \\ \mathbf{m}^+b &= (\mathbf{m}'b \cap a_B) \cup (b - a_B), \end{aligned}$$

for each  $b \in \mathfrak{B}^+$ . The required modal algebra  $\mathfrak{B}^+$  is  $\langle B^+, 0, 1, -, \cap, \cup, \mathbf{I}^+, \mathbf{m}^+ \rangle$ . The canonical modal algebra  $\mathfrak{A}_{\mathbf{S4.1}}$  and the filtrations of it are closure or interior algebras, and it can be shown that  $\mathfrak{B}^+$  is as well. Using the fact

that  $\mathfrak{A}_{\mathbf{S4.1}}$  verifies the first rule above, it can be shown that  $\mathbf{I}^+ a_B = 0$ . From this it follows that

$$\text{if } \mathbf{I}^+ b = 0 \text{ then } \mathbf{I}^+ \mathbf{m}^+ b = 0,$$

so that the second rule above is indeed verified by  $\mathfrak{B}^+$ . Finally it can be shown that  $\mathbf{I}^+ b = \mathbf{I} b$  and  $\mathbf{m}^+ b = \mathbf{m} b$  if these are in  $B$ , so that  $\mathfrak{B}^+$  rejects the given formula  $A$  rejected by  $\mathfrak{B}$ . Thus **S4.1** is characterised by these finite closure or interior algebras  $\mathfrak{B}^+$ .

For the reflexive and transitive frames which verify **S4**, the condition given in Section 10 for  $\langle W, R \rangle$  to verify **M** becomes

$$\forall x \exists y (x R y \wedge \forall z (y R z \rightarrow y = z)),$$

i.e. that each point  $x$  has an  $R$ -last point  $y$  after it. For finite frames it suffices that each final cluster be simple. It is well-known that in **S4** the only non-equivalent formulas obtained by applying  $\neg, \Box, \Diamond$  to  $P$  are  $P$  itself,  $\Box P, \Box \Diamond \Box P, \Box \Diamond P$  and  $\Diamond \Box P, \Diamond \Box \Diamond P, \Diamond P$ , and the negations of these. Thus in **S4.1** there are only 10 of these ‘modalities’. In forming a filtration  $\langle W', \bar{R}, \bar{V} \rangle$  let us take  $\{A_1, \dots, A_n\}$  to be the finite closure of  $A$  and its subformulas under these modalities of **S4.1**. Now these filtrations of the canonical model  $\langle W_{\mathbf{S4.1}}, R_{\mathbf{S4.1}}, V_{\mathbf{S4.1}} \rangle$  have all their final clusters simple, and so characterise **S4.1**. For consider  $[F], [G] \in W'_{\mathbf{S4.1}}$  in a final cluster of such a frame  $\langle W'_{\mathbf{S4.1}}, \bar{R}_{\mathbf{S4.1}} \rangle$ , with  $A_i \in F$ . Since  $[F]$  is in a final cluster, for each  $[H]$  with  $[F] \bar{R}_{\mathbf{S4.1}} [H]$  we have  $[H] R_{\mathbf{S4.1}} [F]$ , and so  $\Diamond A_i \in H$ . Therefore  $\Box \Diamond A_i \in F$ , as well as  $\Box \Diamond A_i \rightarrow \Diamond \Box A_i \in F$ , so that  $\Diamond \Box A_i \in F$ . Now there must be an  $H$  with  $[F] \bar{R}_{\mathbf{S4.1}} [H]$  and  $\Box A_i \in H$ . But since  $R$  is transitive and this is a final cluster,  $[H] \bar{R}_{\mathbf{S4.1}} [G]$  and so  $A_i \in G$ . We have shown that if  $A_i \in F$  then  $A_i \in G$ , so that extending the argument yields

$$A_i \in F \text{ iff } A_i \in G, \text{ for } i = 1, \dots, n,$$

i.e.  $[F] = [G]$ , as required.

For finite reflexive and transitive frames, to satisfy the condition given in Section 10 for  $\langle W, R \rangle$  to satisfy **Grz**, it suffices that each cluster be simple. Unfortunately filtrations  $\langle W', \bar{R}, \bar{V} \rangle$  of the canonical model for **S4 Grz** may not have this property, and it is necessary to replace  $\bar{R}$  by a suitable asymmetric  $R'$ . Given a cluster  $C$  of reflexive, transitive  $\langle W'_{\mathbf{Grz}}, \bar{R}_{\mathbf{Grz}}, \bar{V}_{\mathbf{Grz}} \rangle$ , say that  $x \in C$  is ‘virtually last’ in  $C$  iff there is some  $F_x \in x$  with

$$\forall G ((F_x R_{\mathbf{Grz}} G \wedge [G] \in C) \rightarrow x = [G]).$$

It is clear that the member of a simple cluster of this frame is virtually last. In [Seegerberg, 1971, Chapter II, Section 3], it is shown by a difficult argument that each proper cluster has a virtually last element as well.

Assuming this result, define  $R'_{\mathbf{Grz}}$  on  $W'_{\mathbf{Grz}}$  by taking  $x R'_{\mathbf{Grz}} y$  iff either not  $x \simeq y$  and  $x \bar{R}_{\mathbf{Grz}} y$  or  $x \simeq y$  and  $x r_C y$ , where  $r_C$  is an arbitrary ordering

of  $C$  in which the  $r_C$ -last member of finite  $C$  is virtually last in  $C$ . Now  $R'_{\mathbf{Grz}} \subseteq \bar{R}_{\mathbf{Grz}}$ , and  $\langle W'_{\mathbf{Grz}}, R'_{\mathbf{Grz}} \rangle$  has only simple clusters and so verifies **S4Grz**. Define  $V'_{\mathbf{Grz}}$  on  $\langle W'_{\mathbf{Grz}}, R'_{\mathbf{Grz}} \rangle$  by taking  $V'_{\mathbf{Grz}}(P, [F]) = V_{\mathbf{Grz}}(P, F)$  for each propositional variable  $P$  in  $\{A_1, \dots, A_n\}$ , and applying the defining conditions for valuations. It can be shown that

$$V'_{\mathbf{Grz}}(A, [F]) = \bar{V}_{\mathbf{Grz}}(A_i, [F]), \text{ for } i = 1, \dots, n,$$

by induction on their construction, so that  $\langle W'_{\mathbf{Grz}}, R'_{\mathbf{Grz}}, V'_{\mathbf{Grz}} \rangle$  rejects the given formula as well. For the induction step on  $\Box$ , consider  $A_i = \Box A_j$ , one direction being easy with  $R'_{\mathbf{Grz}} \subseteq \bar{R}_{\mathbf{Grz}}$ . For the difficult direction take  $x$  to be a cluster  $C$  with  $y$  virtually last in  $C$ , and then

$$\begin{aligned} & \bar{V}_{\mathbf{Grz}}(\Box A_j, x) = F \\ & \text{implies } \bar{V}_{\mathbf{Grz}}(\Box A_j, y) = F \\ & \text{implies } V_{\mathbf{Grz}}(\Box A_j, F_y) = F \text{ and} \\ & \quad \forall G((F_y R_{\mathbf{Grz}} G \wedge [G] \in C) \rightarrow y = [G]) \\ & \text{implies } V_{\mathbf{Grz}}(A_j, G) = F, \text{ for some } G \text{ with either} \\ & \quad F_y R_{\mathbf{Grz}} G \text{ and not } [G] \in C \text{ or } y = [G] \in C, \\ & \text{implies } V'_{\mathbf{Grz}}(A_j, [G]) = F \text{ and either not } y \simeq [G] \\ & \quad \text{and } y \bar{R}_{\mathbf{Grz}} [G] \text{ or } y \simeq [G] \text{ and } y r_C [G] \\ & \text{implies } V'_{\mathbf{Grz}}(A_j, [G]) = F \text{ and } x R'_{\mathbf{Grz}} y \text{ and } y R'_{\mathbf{Grz}} [G] \\ & \text{implies } V'_{\mathbf{Grz}}(\Box A_j, x) = F. \end{aligned}$$

With what natural axiom can **S4.1** be extended to **S4Grz**? Clearly we need a formula  $A$  such that **S4A** is characterised by the finite reflexive-and-transitive frames in which all but the final clusters are simple. Segerberg [1971, Chapter II, Section 3] shows that

$$\mathbf{Dum}.\diamond\Box P \rightarrow (\Box(\Box(P \rightarrow \Box P) \rightarrow P) \rightarrow P)$$

(i.e.  $\diamond\Box P \rightarrow \mathbf{Grz}$ ) has this property, so that **S4Grz** is **S4.1Dum**.

## 15 THE TRANSITIVE LOGICS OF FINITE DEPTH

Given a frame  $\langle W, R \rangle$ , say that  $x_1, \dots, x_r \in W$  form a *chain* iff  $x_i R x_{i+1}$  and  $x_i \neq x_{i+1}$  and not  $x_{i+1} R x_i$ , for  $i = 1, \dots, r-1$ . (Thus  $x_1, \dots, x_r$  come from a chain of distinct clusters. We include  $\langle x_1 \rangle$  as a chain.) Say that  $x_1$  has a rank  $r$  in  $\langle W, R \rangle$  iff there is a chain  $\langle x_1, \dots, x_r \rangle$  but no chain  $\langle x_1, \dots, x_r, x_{r+1} \rangle$ . And say that  $\langle W, R \rangle$  itself has rank  $r$  iff each element in it has a rank which is  $\leq r$ , and some element in it has rank  $r$ . In this section (which is derived from work in [Segerberg, 1971]) we study normal extensions of **K4** with characteristic frames of finite depth in this sense.

Define formulas  $B_n$ , for  $n = 1, 2, 3, \dots$  by taking

$$\begin{aligned} B_1 &= B = \diamond\Box P_1 \rightarrow P_1, \\ B_{n+1} &= \diamond(\Box P_{n+1} \wedge \neg B_n) \rightarrow P_{n+1}. \end{aligned}$$

Then transitive  $\langle W, R \rangle$  verifies  $B_n$  iff it has rank  $\leq n$ . For it is easy to show that  $\langle W, R, V \rangle$  rejects  $B_n$  at  $x_0 \in W$  iff there exists  $x_1, \dots, x_n \in W$  with  $x_i R x_{i+1}$  and

$$\begin{aligned} V(P_{n-i}, x_i) &= F, \\ v(B_{n-i}, x_i) &= F, \\ v(\Box P_{n-i}, x_{i+1}) &= T, \end{aligned}$$

for  $i = 0, \dots, n-1$ , by induction from  $n-1$  to 0. And it can be checked that these conditions can hold iff  $x_0, \dots, x_n$  satisfy the conditions for being a chain.

We shall see that any normal logic  $\mathbf{L}$  which contains  $\mathbf{K4B}_n$  has the f.m.p. Consider a formula  $A$  with propositional variables from  $P_1, \dots, P_m$ , and take  $r$  to be maximum of  $m$  and  $n$ . Taking  $\mathbf{L}_r$  to be the restriction of  $\mathbf{L}$  to  $P_1, \dots, P_r$ , it is clear that  $\vdash_{\mathbf{L}} A$  iff  $\vdash_{\mathbf{L}_r} A$ . Suppose that  $A$  is a nonthesis of both logics. The canonical general frame  $\langle W_{\mathbf{L}_r}, R_{\mathbf{L}_r}, P_{\mathbf{L}_r} \rangle$  verifies  $\mathbf{L}$  and rejects  $A$ , and we shall see that it is finite.

Firstly  $\langle W_{\mathbf{L}_r}, R_{\mathbf{L}_r} \rangle$  has rank  $\leq n$ . For if it has a chain  $F_0, \dots, F_n$  then there must be formulas  $A_1, \dots, A_n$  with

$$\Box A_{n-1} \in F_{i+1} \text{ and not } A_{n-1} \in F_i, \text{ for } i = 0, \dots, n-1.$$

Then it is easy to show that the formula  $B'_n$  obtained from  $B_n$  by substituting  $A_i$  for  $P_i, i = 1, \dots, n$ , has not  $B'_n \in F_0$ , in contradiction to the properties of  $W_{\mathbf{L}_r}$ .

Now  $W_{\mathbf{L}_r}$  has finitely many maximal consistent sets of rank  $i$ , by induction from  $i = 1$  to  $i = n$ . Say that a formula is modally atomic iff it is a propositional variable or of the form  $\Box C$  or  $\Diamond C$ . Since a maximal consistent set  $F$ , like an ultrafilter, satisfies the conditions

$$\begin{aligned} \neg A \in F &\text{ iff not } A \in F, \\ A \wedge B \in F &\text{ iff } A \in F \text{ and } B \in F, \\ A \vee B \in F &\text{ iff } A \in F \text{ or } B \in F, \end{aligned}$$

it is determined by its modally atomic formulas. Note that if  $F$  is a maximal consistent set in  $W_{\mathbf{L}_r}$  of rank  $i$  then  $\Box C \in F$  iff

$$C \in \cap \{G : F \simeq G \vee (FR_{\mathbf{L}_r} G \wedge G \text{ has rank } < i)\}$$

and  $\Diamond C \in F$  iff

$$C \in \cup \{G : F \simeq G \vee (FR_{\mathbf{L}_r} G \wedge G \text{ has rank } < i)\}.$$

By the induction hypothesis there are finitely many sets of maximal consistent sets  $G$  with  $(FR_{\mathbf{L}_r} G \wedge G \text{ has rank } < i)$ . There are finitely many ways of allocating  $P_1, \dots, P_r$  to the maximal consistent sets  $G$  with  $F \simeq G$ . Once these items are fixed, the members of each maximal consistent set in the



cluster including  $F$  are determined (by an easy induction on the construction of formulas). In particular the number of maximal consistent sets in the cluster is at most the number of ways of allocating  $P_1, \dots, P_r$  to those sets. It follows that there are finitely many possible sets of modally atomic formulas for  $F$ , and hence finitely many maximal consistent sets  $F$  of rank  $i$  in  $\langle W_{\mathbf{L}_r}, R_{\mathbf{L}_r} \rangle$ .

## 16 THE NORMAL EXTENSIONS OF **S4.3**

(Bull [1966] gives an algebraic proof that every normal extension of **4.3** has the f.m.p. Fine [1971] gives a frame-theoretic proof, together with a description of the lattice of these logics. Both proofs are rather elegant.)

Let  $\mathbf{L}$  be any normal modal logic containing **S4.3**. by what we have seen in Section 10,  $\mathbf{l}$  is characterised by the subdirectly irreducible closure or interior algebras which verify it. Let  $\mathfrak{A}$  be such an algebra. Since  $\mathfrak{A}$  verifies  $\Box(\Box P \rightarrow \Box Q) \vee \Box(\Box Q \rightarrow \Box P)$  and satisfies the condition

$$\text{if } \mathbf{la} \cup \mathbf{lb} = 1 \text{ then } \mathbf{la} = 1 \text{ or } \mathbf{lb} = 1,$$

it is well-connected in the sense that

$$\mathbf{la} \leq \mathbf{lb} \text{ or } \mathbf{lb} \leq \mathbf{la}.$$

It also satisfies the condition

$$\text{if } \mathbf{la} < \mathbf{lb} \text{ then } \mathbf{l}(a \cup (-\mathbf{lb})) = \mathbf{la},$$

where  $\mathbf{la} < \mathbf{lb}$  is  $(\mathbf{la} \leq \mathbf{lb}) \wedge \mathbf{la} \neq \mathbf{lb}$ . This is shown by first applying the same argument to  $\Box(\Box P \rightarrow \Box Q) \vee \Box(\Box(\Box P \rightarrow \Box Q) \rightarrow \Box Q)$ , which can be shown to be a thesis of **S4.3**, so that  $\mathbf{lb} \leq \mathbf{la}$  or  $\mathbf{l}((-\mathbf{lb}) \cup \mathbf{la}) \leq \mathbf{la}$ . But if  $\mathbf{la} < \mathbf{lb}$  then not  $\mathbf{lb} \leq \mathbf{la}$ , and in any interior algebra it can be shown that  $\mathbf{la} \leq \mathbf{l}((-\mathbf{lb}) \cup \mathbf{la}) = \mathbf{l}((-\mathbf{lb}) \cup a)$ . dualising these results, we have

$$\mathbf{ma} \leq \mathbf{mb} \text{ or } \mathbf{mb} \leq \mathbf{ma}, \text{ if } \mathbf{mb} < \mathbf{ma} \text{ then } \mathbf{m}(a - \mathbf{mb}) = \mathbf{ma},$$

for each  $a, b \in A$ .

Given a nonthesis  $A$  of  $\mathbf{l}$  and an algebraic ‘model’  $\langle \mathfrak{A}, v \rangle$  which rejects it, let  $A_1, \dots, A_m$  be  $A$  and its subformulas and let  $\mathfrak{B} = \langle B, 0, 1, -, \cap, \cup \rangle$  be the finite subalgebra of  $\langle A, 0, 1, -, \cap, \cup \rangle$  generated by  $\{v(A_1), \dots, v(A_m)\}$ . Take  $W$  to be the set  $\{b_1, \dots, b_n\}$  of atoms of the atomic Boolean algebra  $\mathfrak{B}$  and define  $R$  on  $W$  by taking

$$b_i R b_j \text{ iff } b_i \leq \mathbf{m}b_j.$$

Now  $\langle W, R \rangle^+$  is a finite closure or interior algebra, such that there is an isomorphism  $\phi$  from  $\mathfrak{B}$  onto the underlying Boolean algebra of  $\langle W, R \rangle^+$  on

$\mathfrak{B}(W)$ . (Note that  $\langle W, R \rangle^+$  is not a filtration of  $\mathfrak{A}$  in the usual sense.) Define a valuation  $V'$  on  $\langle W, R \rangle$  by taking

$$v_i(P) = \phi v(P), \text{ for each propositional variable } P \text{ in } A,$$

and applying the conditions on valuations. We have

$$v'(A_i) = \phi v(A_i), \text{ for } i = 1, \dots, n,$$

because  $\phi$  is a Boolean isomorphism and

$$x \in \phi(\mathbf{m}b) \text{ iff } \exists y(xRy \wedge y \in \phi(b)),$$

for each  $b \in \mathfrak{B}$ . For taking  $b = x_1 \cup \dots \cup x_r$  for atoms  $x_1, \dots, x_r$  of  $\mathfrak{B}$ , we have

$$\begin{aligned} x \in \phi(\mathbf{m}b) & \\ \text{iff } x &\leq \mathbf{m}(x_1 \cup \dots \cup x_r) \\ \text{iff } x &\leq \mathbf{m}x_1 \cup \dots \cup \mathbf{m}x_r \\ \text{iff } x &\leq \mathbf{m}(x_1 \text{ or } \dots \text{ or } x_r) \\ \text{iff } xRx_1 &\text{ or } \dots \text{ or } xRx_r \\ \text{iff } \exists y(xRy \wedge y &\in \phi(b)). \end{aligned}$$

In particular  $\langle W, R \rangle^+$  rejects  $A$ .

To show that  $\langle W, R \rangle^+$  verifies **L**, it is sufficient to construct an embedding homomorphism  $\theta$  from  $\langle W, r \rangle^+$  into  $\mathfrak{A}$ . Suppose that  $b_1, \dots, b_n$  are indexed so that, in their indexed order,  $\mathbf{m}b_{k(1)} = \dots = \mathbf{m}b_{k(2)-1} < \dots < \mathbf{m}b_{k(s)} = \dots = \mathbf{m}b_{k(s+1)-1}$  in  $\mathfrak{A}$ , where  $1 = k(1) < \dots < k(s+1) = n+1$ . Set  $b_{k(0)} = 0$  and note that  $\mathbf{m}b_{k(1)} - \mathbf{m}b_{k(0)}, \dots, \mathbf{m}b_{k(s)} - \mathbf{m}b_{k(s-1)}$  is a disjoint cover of 1. Define  $\theta$  by taking

$$\theta(\phi) = 0;$$

for  $i = 1, \dots, s$ ,

$$\theta(\{b_{k(i)}\}) = \mathbf{m}b_{k(i)} = b_{k(i)+1} \cup \dots \cup b_{k(i+1)-1} - \mathbf{m}b_{k(i-1)};$$

for  $i = 1, \dots, s$  and  $k(i) + j = k(i) + 1, \dots, k(i+1) - 1$ ,

$$\begin{aligned} \theta(\{b_{k(i)+j}\}) &= b_{k(i)+j} - \mathbf{m}b_{k(i+1)}; \\ \theta(\{b_{i(1)}, \dots, b_{i(r)}\}) &= \theta(\{b_{i(1)}\}) \cup \dots \cup \theta(\{b_{i(r)}\}). \end{aligned}$$

It is clear that  $\theta$  is an embedding homomorphism of the underlying Boolean algebras. It can also be shown that

$$\begin{aligned} \mathbf{m}\theta(\{b_{k(i)}\}) &= \mathbf{m}(b_{k(i)} - \mathbf{m}b_{k(i-1)}), \\ \mathbf{m}\theta(\{b_{k(i)+j}\}) &= \theta(\{b_1, \dots, b_{k(i+1)-1}\}), \end{aligned}$$

for  $i = 1, \dots, s$  and  $k(i) + j = k(i), \dots, k(i+1) - 1$ . (The second result uses the first and the lemma of the first paragraph.) But  $\{b_1, \dots, b_{k(i+1)-1}\}$

is the closure of  $\{b_{k(i)+j}\}$  in  $\langle W, R \rangle^+$ , so that  $\theta$  is now easily seen to be a homomorphism w.r.t.  $\mathbf{m}$  as well.

Alternatively,  $\mathbf{L}$  is characterised by the generated submodels  $\langle W_{\mathbf{L}x}, R_{\mathbf{L}x}, V_{\mathbf{L}x} \rangle$  of its canonical model. We know from Section 13 that these satisfy the condition

$$\forall y \forall z (y R_{\mathbf{L}x} z \vee z R_{\mathbf{L}x} y).$$

So, given a nonthesis  $A$  of  $\mathbf{L}$ , let  $\langle W, R, V \rangle$  be a model which satisfies this condition and rejects  $A$ . Let  $\{A_1, \dots, A_n\}$  be  $\diamond A$  and its subformulas, and consider the filtration  $\langle W; \bar{R}, \bar{V} \rangle$  determined by this set of formulas.

Let us first try to prove that finite  $\langle W', \bar{R} \rangle$  verifies each formula verified by  $\langle W, R, V \rangle$  and, hence,  $\mathbf{L}$ , which would establish the f.m.p. for  $\mathbf{L}$ . We must first reduce any model  $\langle W', \bar{R}, V'' \rangle$  to  $\langle W', \bar{R}, \bar{V} \rangle$ . Say a subset of  $W'$  is definable in  $\langle W, \bar{R}, \bar{V} \rangle$  iff it is  $\bar{v}(B)$ , for some formula  $B$ ; that  $\langle W', \bar{P}R, v'' \rangle$  is a definable variant of  $\langle W', \bar{R}, \bar{V} \rangle$  iff  $\bar{v}''(P)$  is definable in  $\langle W', \bar{R}, \bar{V} \rangle$ , for each propositional variable  $P$ ; and that  $\langle W', \bar{R}, \bar{V} \rangle$  is differentiated iff  $\{[w]\}$  is definable, for each  $[w] \in W'$  (cf. 1- refinement). It is easy to show that finite  $\langle W', \bar{R}, \bar{V} \rangle$  is differentiated; that therefore each  $\langle W', \bar{R}, V'' \rangle$  is a definable variant of it' and that therefore if  $\langle W', \bar{R}, \bar{V} \rangle$  verifies  $\mathbf{L}$  then so does each  $W', \bar{R}, V''$ . To show that  $\langle W', \bar{R}, \bar{V} \rangle$  verifies  $\mathbf{L}$ , it would clearly suffice to show that

$$\begin{aligned} &\text{if } xRy \text{ then } [x]\bar{R}[y], \\ &\text{if } [x]\bar{R}[y] \text{ then } \exists z(xRz \wedge z \in [y]). \end{aligned}$$

The first condition is of course true, but unfortunately it is quite possible that the second could fail.

In view of this set-back, let us try to eliminate elements  $\beta$  for which the second condition fails. given  $\alpha, \beta \in W'$ , define  $\beta$  sub  $\alpha$  to hold iff

$$\exists x(x \in \alpha \wedge \forall y(y \in \beta \rightarrow \neg xRy)).$$

Note that if this holds then  $yRx$ , since  $\langle W, R \rangle$  is connected and so  $\beta \bar{R} \alpha$ . Say that  $\beta$  is *eliminable*' iff there is some  $\alpha$  with  $\alpha \bar{R} \beta$  and  $\beta$  sub  $\alpha$ . (Note the similarity of the conditions 'virtually last' and 'eliminable' on the members of a cluster in a filtration.) Take  $U$  to be the set of noneliminable elements of  $V$ , and form  $\langle U, \bar{R}, \bar{V} \rangle$  by restricting  $\bar{R}, \bar{V}$  to  $U$ . It is easy to show that

$$\bar{V}(A_i, [x]) = T \text{ iff } A_i \in x, \text{ for } i = 1, \dots, n \text{ and each } [x] \in U,$$

once the lemma of the following paragraph is proved. It follows that  $\langle U, \bar{R}, \bar{V} \rangle$  rejects the given formula  $A$  and is differentiated.

The lemma is that, for each formula  $\diamond B$  in  $\{A_1, \dots, A_n\}$ , if  $\diamond B \in x$  then there is some  $y$  with  $B \in y$  such that  $[x]\bar{R}[y]$  and  $y$  is not eliminable. This is done by constructing a sequence  $\alpha_0, \alpha_1, \alpha_2, \dots$  in  $W'$  by taking  $\alpha_0 = [x]$ ,

and for each  $i = 1, 2, 3, \dots$ ,

$$\begin{aligned} \alpha_{2i-1} & \text{ is some } [z] \text{ with } B \in z \text{ and not } [z] \text{ sub } \alpha_{2i-2}, \\ \alpha_{2i} & \text{ is some } [z] \text{ with } [z]\bar{R}\alpha_{2i-1} \text{ and } \alpha_{2i-1} \text{ sub } [z]. \end{aligned}$$

It is easy to see that  $B \in \alpha_{2i-1}$  and  $\diamond B \in \alpha_{2i}$ , for  $i = 1, 2, 3, \dots$ . It can be shown that this sequence must terminate, but that it cannot terminate at any  $\alpha_{2i}$ . the required  $y$  is the  $z$  with  $B \in z$  such that the sequence terminates at  $\alpha_{2j-1} = [z]$ .

To complete the argument it will suffice to set up a frame morphism  $\phi$  from some definable variant of  $\langle W, R, V \rangle$  onto  $\langle U, \bar{R}, \bar{V} \rangle$ . Fro then  $\langle U, \bar{R}, \bar{V} \rangle$  will verify  $L$ , as shown in Section 10, and so will each variant of differentiated  $\langle U, \bar{R}, \bar{V} \rangle$ , as in the original ‘proof’. Define  $\phi : W \rightarrow U$  by taking

$$\begin{aligned} \phi(x) & = [x], \text{ if } [x] \in U, \\ & = \text{the first element in some arbitrary ordering of } U \text{ which is} \\ & \quad \bar{R}\text{-first in } \{\alpha : [x]\bar{R}\alpha\}, \text{ otherwise} \end{aligned}$$

—noting that  $\phi$  is onto  $U$ . If  $xRy$  then  $[x]\bar{R}[y]$  and  $[y]\bar{R}\phi(y)$  yield  $\phi(x)\bar{R}\phi(y)$ . If  $\phi(x)\bar{R}\phi(y)$  then we must have some  $z \in \phi(y)$  with  $xRz$ , otherwise  $\phi(y)$  sub  $\phi(x)$  and  $\phi(y)$  would be eliminable. Now  $\phi(y) = \phi(z)$  and  $xRz$  as required. Thus  $\phi$  is an onto frame morphism. Define a valuation  $V'$  on  $\langle W, R \rangle$  by taking  $V'(P, x) = \bar{V}(P, \phi(x))$ , for each propositional variable  $P$ , and applying the conditions on valuations. Then it is easy to show that  $\langle W, R, V' \rangle$  is a definable variant of  $\langle W, R, V \rangle$  and to extend  $\phi$  to a morphism of models.

(What is the relationship between these two proofs? Take  $\langle W, R, V \rangle$  to be a generated sub model of the canonical model of  $\mathbf{L}$ , and take  $\langle \mathfrak{A}, v \rangle$  to be  $\langle \langle W, R \rangle^+, v \rangle$ , for the same valuation. Thus  $\mathfrak{A}$  is indeed a subdirectly irreducible closure or interior algebra verifying  $\mathbf{L}$ . Relabelling the finite frame  $\langle W, R \rangle$  of the first proof as  $\langle W', R' \rangle$ ,  $W'$  is the usual set obtained from  $\{v(A_1), \dots, v(A_n)\}$  in a filtration, but from

$$\alpha R' \vee \text{ iff } \forall x(x \in \alpha \rightarrow \exists y(xRy \wedge y \in \beta)).$$

Since a one–one homomorphism  $\theta$  from  $\langle W', R' \rangle^+$  into  $\langle W, R \rangle^+$  is the dual of a frame morphism  $\phi$  from  $\langle W, R \rangle$  onto  $\langle W', R' \rangle$ , we would expect that all the elements in  $W'$  are noneliminable. To see that this is indeed true, suppose that  $\alpha R' \beta$  and  $\beta$  sub  $\alpha$  and try to obtain a contradiction. In this case there is some  $x \in \alpha$  with  $\forall y(y \in \beta \rightarrow \neg xRy)$  by the definition of  $\beta$  sub  $\alpha$ . then the definition of  $\alpha R' \beta$  give us some  $y \in \beta$  with  $xRy$ —the required contradiction. Unfortunately, the other condition on frame morphisms, that if  $xRy$  then  $[x]R'[y]$ , is not satisfied by this construction. and indeed the frame morphism  $\phi$  of which  $\theta$  i the dual, is not  $\phi(x) = [x]$ , for each  $x \in W$ , but a more complicated function which can be constructed from the definition of  $\theta$  above.)

Say that a nonempty sequence of positive integers is a list. A finite frame  $\langle W, R \rangle$  which verifies **S4.3** must consist of a finite chain of finite clusters, so that it is described by the list of numbers of elements in successive clusters. Say that a list  $t$  contains a list  $s = \langle a_1, \dots, a_m \rangle$  when there is a subsequence  $\langle b_{i_1}, \dots, b_{i_m} \rangle$  of  $t$  with  $a_1 \leq b_{i_1}, \dots, a_m \leq b_{i_m}$ . And that  $t = \langle b_1, \dots, b_n \rangle$  covers  $s$  iff  $t$  contains  $s$  and  $a_m \leq b_n$ . Given finite frames  $\langle W, R \rangle$  and  $\langle U, S \rangle$  which verify **S4.3**, described by lists  $t$  and  $s$ , it is easy to show that if  $t$  covers than in each infinite sequence  $t_1, t_2, t_3, \dots$  of lists there is an infinite subsequence  $t_{i_1}, t_{i_2}, t_{i_3}, \dots$ , such that if  $h < k$  then  $t_{i_h}$  is covered by  $t_{i_k}$ . From this it is easy to deduce that there is no infinite increasing sequence  $\mathbf{L}_1 \subset \mathbf{L}_2 \subset \mathbf{L}_3 \subseteq \dots$  of normal modal logics containing **S4.3**. For take  $A_i$  to be a formula in  $\mathbf{L}_{i+1}$  but not in  $\mathbf{L}_i$ , and take  $t_i$  to be the list describing a suitable finite frame which rejects  $A_i$ . Then the result yields a  $t_j$  with  $i < j$  which covers  $t_i$ , and now  $A_i$  is also not in  $\mathbf{L}_j$  with  $i + 1 \leq j$ , a contradiction.

## 17 THE PRETABULAR EXTENSIONS OF S4

(A normal modal logic is said to be tabular iff it is characterised by a single finite structure, and to be pretabular iff all its proper extensions are tabular. Thus the well-known [Scroggs, 1951] shows that **S5** = **S4B** is a pretabular logic. Maksimova [1975] and [Esaia and Meskhi, 1977] independently prove the very pretty result that there are precisely five pretabular extensions of **S4**. The work of the last four sections provides the background needed for [Esaia and Meskhi, 1977]. The pretabular extensions of **K4** are a much more difficult topic, dealt with by [Block, 1980a]. This paper takes as its starting point the very strong results of [Jónsson, 1967] on the subdirectly irreducible algebras in a variety.)

Consider the finite, generated, reflexive-and-transitive frames  $\langle W, R \rangle$ . Which parameters of these frames can be left unrestricted by the formulas that they verify? It turns out that there are precisely five of them.

1. *The maximum number of points in any final cluster.*
2. *the maximum number of points in any non-final cluster.*

A cluster  $[z]$  is a successor of  $[x]$  iff  $xRz$  but  $[x] \neq [z]$ , and an immediate success for iff, further, there is no cluster  $[y]$  such that  $[z]$  is a successor of  $[y]$  and  $[y]$  is a successor of  $[x]$ . Say that the external branching of a cluster is the number of final clusters which are immediate successors of it. And that the internal branching of a cluster is the number of non-final clusters which are immediate successors of it.

3. *The maximum of the external branching of the clusters.*
4. *The maximum of the internal branchings of the clusters.*

5. *The maximum number of clusters in any chain of cluster*, i.e. the rank of  $\langle W, R \rangle$  in the sense of Section 15.

It is clear that once all five parameters are bounded, the class of reflexive- and- transitive frames satisfying those bounds is finite. Thus if  $\mathbf{L}$  is determined by such a class of frames then it is determined by a single finite frame, namely the finite disjoint union of these finite frames.

For each of the five parameters, given a finite frame  $\langle W, R \rangle$  of the kind being considered, a frame  $\langle W_i, R_i \rangle$  of a certain kind can be constructed, which has the same value of that parameter. The constructions needed are subframes and frame-morphic images. We saw in Section 10 that a class of frames verifying a normal modal logic  $\mathbf{L}$  is closed under them. The five kinds of simple frames and their constructions are as follows.

1.  $\langle W_1, R_1 \rangle$  *has one cluster*. Take the largest final cluster of  $\langle W, R \rangle$ , which is a subframe and has the required properties.
2.  $\langle W_2, R_2 \rangle$  *has two clusters, of which the final one is simple*. Take the largest nonfinal cluster  $[x]$  of  $\langle W, R \rangle$  and form  $\langle W_x, R_x \rangle$ . Take  $W_2 = [x] \cup \{\omega\}$  and define  $R_2$  on it by taking  $xR_2y$  iff  $x \simeq y \vee y = \omega$ . Define a frame morphism  $\phi_2$  from  $W_x$  onto  $W_2$  by taking  $\phi_2(y) = y$  if  $x \simeq y$ ,  $\phi_2(y) = \omega$  otherwise.
3.  $\langle W_3, R_3 \rangle$  *has  $W_3 = \{0, 1, \dots, n\}$  with  $xR_3y$  iff  $x = y \vee x = 0$* . Take  $[x]$  to have the maximal external branching in  $\langle W, R \rangle$  with final clusters  $[y_1], \dots, [y_n]$  immediately succeeding it. Form  $\langle W_x, R_x \rangle$  and define a frame morphism  $\phi_3$  from  $W_x$  onto  $W_3$  by taking  $\phi_3(y) = 0$  if  $y \in [x]$ ,  $\phi_3(y) = i$  if  $y \in [y_i]$ , for  $i = 1, \dots, n$ ,  $\phi_3(y) = 1$  otherwise.
4.  $\langle W_4, R_4 \rangle$  *has  $W_4 = \{0, 1, \dots, n, \omega\}$  with  $xR_4y$  iff  $x = y \vee x = 0 \vee y = \omega$* . Take  $[x]$  to have the maximal internal branching in  $\langle W, R \rangle$ , with nonfinal clusters  $[y_1], \dots, [y_n]$  immediately succeeding it. Form  $\langle W_x, R_x \rangle$  and define a frame morphism  $\phi_4$  from  $W_x$  onto  $W_4$  by taking  $\phi_4(y) = 0$  if  $y \in [x]$ ,  $\phi_4(y) = i$  if  $y \in [y_i]$ , for  $i = 1, \dots, n$ ,  $\phi_4(y) = \omega$  otherwise.
5.  $\langle W_5, R_5 \rangle$  *has  $W_5 = \{2, \dots, n\}$  with  $iR_5j$  if  $i \leq j$* . Suppose that  $\langle W, R \rangle$  has rank  $n$ , with a maximal chain  $\langle x_1, \dots, x_n \rangle$ . Define a frame morphism  $\phi_5$  from  $W$  onto  $W_5$  by taking  $\phi_5(y) = i$  if  $x_i \simeq y$ , for  $i = 1, \dots, n-1$ ,  $\phi_5(y) = n$  otherwise.

Each of these five sets of simple frames characterises a normal modal logic, as follows:

1. **S4B**, known as **S5**.
2. **S4.3B<sub>2</sub>M**
3. **S4GrzB<sub>2</sub>**.

4. **S4GrzB<sub>3</sub>** plus  $\diamond\Box\diamond P \rightarrow \Box\diamond P$ .

5. **S4 · 3Grz**.

For each of these extensions of **S4B<sub>n</sub>** or **S4 · 3** has the f.m.p. by Sections 15 and 16, and it is easy to check the class of finite generated frames which verifies each logic. Any pretabular extension **L** of **S4** must be one of these logics. For pretabular **L** must have the f.m.p. with a class of finite frames in which one of the five parameters is not bounded, as we saw above. Its class of finite frames must therefore include one of the five sets of simple frames. Therefore **L** must be contained in one of the five corresponding logics. But every proper extension of pretabular **L** must be tabular, so that **L** has to be identical with one of these logics.

Finally it can be shown that any nontabular logic is contained in a pretabular logic, and hence in one of these five. But these five logics are pairwise incomparable, so that they must all be pretabular logics.

## 18 THE TRANSITIVE LOGICS OF FINITE WIDTH

(The work of this section is taken from Fine [1974a; 1974b], which extend the ideas of [Fine, 1971] to a wider set of logics.)

Given a frame  $\langle W, R \rangle$  say that points  $x, y \in W$  are incomparable iff  $x \neq y$  and not  $xRy$  and not  $yRx$ . The frame  $\langle W, R \rangle$  is of width  $n$  if it has  $n$  pairwise incomparable points but does not have  $n + 1$  incomparable points. (In particular, for transitive frames,  $\langle W, R \rangle$  is connected iff it is of width 1.) For  $i = 1, \dots, n$  take  $\mathbf{I}_n$  to be the formula

$$\bigwedge_{i=0}^n P \rightarrow \bigvee_{0 \leq i \neq j \leq n} \diamond(P_i \wedge (P_j \vee \diamond P_j)).$$

It is easy to see that a generated frame verifies  $\mathbf{I}_n$  iff it is of width  $\leq n$ .

Various of the nice properties of the connected frames break down at greater widths. As an example of this, there is an infinite increasing chain of normal extensions of **S4I<sub>2</sub>**. Indeed there are continuum many distinct normal extensions of **S4I<sub>2</sub>**. This is shown by defining certain frames  $\mathfrak{F}_1, \mathfrak{F}_2, \mathfrak{F}_3, \dots$  of width 2, and proving that distinct subsets of this set of frames characterise distinct logics. Each frame  $\mathfrak{F}_n = \langle W_n, R_n \rangle$  is defined by taking  $W_n = \{0, \dots, 2n + 4\}$  and taking  $R_n$  to be the restriction to  $W_n$  of  $R$  with

$$\begin{aligned} iRj \text{ iff either } & i = 0 \\ & \text{or } i \text{ is odd, } j \text{ is odd, and } i > j \\ & \text{or } i \text{ is odd, } j \text{ is even, and } i > j + 2 \\ & \text{or } i \text{ is odd, } j \text{ is odd, and } i > j + 4 \end{aligned}$$

For example,  $\mathfrak{F}_2$  is depicted in Figure 1.

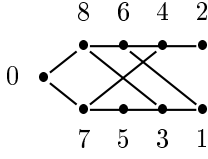


Figure 1.

The result will follow if it can be shown that each  $\mathfrak{F}_n$  rejects a formula  $\neg A_n$  which is verified by every other  $\mathfrak{F}_m$ . In each case  $A_n$  is taken to be the frame formula for  $\mathfrak{F}_n$ , in the following sense.

The frame formula  $A_{\mathfrak{F}}$  for any finite reflexive-and-transitive frame  $\mathfrak{F} = \langle \{0, \dots, r\}, R \rangle$  generated by 0 is the conjunction of the formulas  $P_0$  and

$$\begin{aligned} & \Box(P_0 \vee \dots \vee P_r), \\ & \Box(P_i \rightarrow \neg P_j), \text{ for each } i \neq j, \\ & \Box(P_i \rightarrow \Diamond P_j), \text{ whenever } iRj, \\ & \Box(P_i \rightarrow \neg \Diamond P_j), \text{ whenever not } iRj. \end{aligned}$$

In general, frame formulas have the property that  $A_{\mathfrak{F}}$  can be satisfied in a frame  $\mathfrak{S} = \langle U, S \rangle$  iff, for some  $u \in U$ , there is a frame morphism  $\phi$  from  $\mathfrak{S}_u$  onto  $\mathfrak{F}$ . We know from Section 10 that if this condition holds then each formula satisfied in  $\mathfrak{F}$  can be satisfied at  $u$  in  $\mathfrak{S}$ . but  $A_{\mathfrak{F}}$  is satisfied in  $\mathfrak{F}$  when  $V$  is defined on  $\{0, \dots, r\}$  by taking

$$V(P_{i,j}) = T \text{ iff } i = j, \text{ for each } i = 0, \dots, r,$$

which yields  $V(A_{\mathfrak{F}}, 0) = T$ . For the converse, suppose that there is a  $u \in U$  and a valuation  $V'$  on  $\mathfrak{S}$  with  $V'(A_{\mathfrak{F}}, u) = T$ . Then define a function  $\phi$  from  $U_u$  into  $\{0, \dots, r\}$  by taking

$$\phi(x) = i \text{ iff } V'(P_i, x) = T,$$

for each  $x$  with  $uSx$  and  $i = 0, \dots, r$ . It is straightforward to show, using the construction of  $A_{\mathfrak{F}}$ , that  $\phi$  is an onto frame morphism.

Therefore, to show that  $\neg A_n$  is verified by  $\mathfrak{F}_m$ , i.e.  $A_n$  is not satisfied by  $\mathfrak{F}_m$ , it suffices to show that there is no frame morphism from  $\mathfrak{F}_{m,k}$  onto  $\mathfrak{F}_n$  unless  $m = n$  and  $k = 0$ . Clearly, if  $m < n$  or  $1 \leq k \leq 2n + 6$  then  $\mathfrak{F}_{m,k}$  does not have enough points for there to be a frame morphism from it onto  $\mathfrak{F}_n$ . (Compare  $\mathfrak{F}_{2,k}$  with  $\mathfrak{F}_0$ .) So suppose that  $m > n$  and  $k = 0$  or  $k \geq 2n + 7$ , and that  $\phi$  is a frame morphism from  $\mathfrak{F}_{m,k}$  onto  $\mathfrak{F}_n$ , and try to obtain a contradiction. Firstly it can be shown that  $\phi(1)$  and  $\phi(2)$  are distinct final points of  $\mathfrak{F}_n$ , say  $\phi(1) = 1$  and  $\phi(2) = 2$ . Then it can be shown



that  $\phi(i) = i$ , for  $i \geq 1$ , in  $\mathfrak{F}_{m,k}$ , by induction on odd or even  $i = 1, 2, \dots$ . Now  $i = 2n + 5$  or  $i = 2n + 6$  is in  $\mathfrak{F}_{m,k}$  but not in  $\mathfrak{F}_n$ , so that  $\mathfrak{F}_n$  does not have enough points for  $\phi$  to map  $\mathfrak{F}_{m,k}$  but not in  $\mathfrak{F}_n$ , so that  $\mathfrak{F}_n$  does not have enough points for  $\phi$  to map  $\mathfrak{F}_{m,k}$  into it. (Compare  $\mathfrak{F}_{1,0}, \mathfrak{F}_{2,7}, \mathfrak{F}_{2,8}$  with  $\mathfrak{F}_0$ .) (Check why this argument cannot be used on a connected frame!)

Nonetheless, each normal extension of  $\mathbf{K4I}_n$  is characterised by the transitive frames of width  $\leq n$  which verify it. The proof of this major result is difficult, and all that will be given here is a brief glance at the ideas involved. Let  $\mathbf{L}$  be any normal extension of  $\mathbf{K4I}_n$ . The big difference from the second half of Section 16 is that we are working with infinite  $\langle W_{\mathbf{L}_r}, R_{\mathbf{L}_r}, F_{\mathbf{L}_r} \rangle$  instead of with a finite filtration of  $\langle W_{\mathbf{L}}, R_{\mathbf{L}}, V_{\mathbf{L}} \rangle$ . (Here  $\mathbf{L}_r$  is the restriction of  $\mathbf{L}$  to the propositional variables  $P_1, \dots, P_r$ .) Therefore the problem comes at a different point. It is now immediate that  $\langle W_{\mathbf{L}_r}, R_{\mathbf{L}_r}, V_{\mathbf{L}_r} \rangle$  verifies  $\mathbf{L}$ , but since this differentiated model is not finite, it is no longer true that each variant of it is definable. (Note that just as the canonical general frame is refined, the canonical model is not only differentiated but natural. That is, it satisfies the condition that if  $V(\Box A, x) = T \rightarrow V(A, y) = T$ , for each formula  $A$ , then  $xRy$ .)

As before it is necessary to eliminate certain points from the given frame. Say  $x \in W_{\mathbf{L}_r}$  is eliminable iff, for each formula  $A$ ,

$$\text{if } V(a, x) = T \text{ then } \exists y(xR_{\mathbf{L}_r}y \wedge \neg yR_{\mathbf{L}_r}x \wedge V_{\mathbf{L}_r}(A, y) = T).$$

A reduced canonical model is not formed on the noneliminable points. It must be shown that there are enough noneliminable points, i.e. that if  $V_{\mathbf{L}_r}(A, x) = T$  then there is some noneliminable  $y$  with  $xR_{\mathbf{L}_r}y$  and  $V_{\mathbf{L}_r}(A, y) = T$ , and that they are definable. The proof that the reduced canonical frame verifies  $\mathbf{L}$ , because the definable variants of the reduced canonical model do, uses the facts that  $\langle W_{\mathbf{L}_r}, R_{\mathbf{L}_r}, V_{\mathbf{L}_r} \rangle$  is natural and that  $\langle W_{\mathbf{L}_r}, R_{\mathbf{L}_r} \rangle$  has no infinite ascending  $R$ -chains. (So does the proof of the definability of the noneliminable points.) So a crucial step in the argument is the lengthy proof that a differentiated model which is transitive and of finite width has no such chains.

## 19 THE VEILED RECESSION FRAME

The *recession frame*  $\langle \omega, R \rangle$  is defined on  $\omega = \{0, 1, 2, \dots\}$  by taking

$$mRn \text{ iff } m \leq n + 1 \text{ for each } m, n \in \omega.$$

Thus  $R$  is reflexive, and transitive for increasing numbers, but is not transitive for decreasing numbers, when only  $mRn$  iff  $m = n + 1$ . For any valuation  $V$  on  $\langle \omega, R \rangle$ ,

$$v(\Box A) = [m, \infty) = \{n : m \leq n\} \text{ and } v(\Box \Box A) = [m + 1, \infty),$$

where  $[m-1, \infty)$  is the ‘largest unbroken interval in  $v(A)$ ’. It is easy to verify that the recession frame verifies **KT** · **3**. The *veiled* recession frame  $\langle \omega, R, P \rangle$  is the general frame defined on the recession frame by taking  $P$  to consist of the finite and cofinite subsets of  $\omega$ . (Cofinite subsets are the complements of the finite ones.) In fact Blok has shown that it characterises **KT** · **3M** plus  $\Box(P \rightarrow \Box P) \rightarrow (\Diamond P \rightarrow P)$  and two further axioms, all of which correspond to certain first-order conditions on frames; see [van Benthem, 1978]. The recession frame was introduced in [Makinson, 1969] to show that a certain logic does not have the f.m.p. the veiled recession frame was introduced in [Thomason, 1974] to show that a certain logic is not characterised by frames. Two similar but sharper examples were produced in [van Benthem, 1978]. These four results are discussed in this section. Thomason [1972a] uses the finite fragments of the recession frame with one point added. It shows that a certain formula (10) is verified by any frame verifying a certain infinite set of axioms, of which each finite subset is verified by a frame rejecting (10). It follows that whatever finitary rules are used, a logic with these axioms is not characterised by the frames which verify it. Finally Blok [1980] uses variations on the veiled recession frame to show that there is a continuum of distinct extensions of **KT** which are all verified by the same class of frames! This paper takes as its starting point the very strong results of [Jónsson, 1967] on the subdirectly irreducible algebras in a variety.

These results are usually described as incompleteness theorems, but they are better thought of as showing the independence of various notions of consequence. In each case we have a logic **L** and a formula  $F$ . Firstly there is modal *logical consequence*  $\mathbf{L} \vdash F$ , using the rules of normal modal logics. then for each class  $S$  of structures there is a corresponding notion of *semantic consequence*, with  $\mathbf{L} \vDash F$  iff  $F$  is verified by each structure in  $S$  which verifies **L**. We know from Sections 10, 11, 12 that finite semantic consequence is as strong as (frame) semantic consequence, which is as strong as general (frame) semantic consequence, which is equivalent to algebraic ‘semantic’ consequence and modal logical consequence. The problem is to show that these relative strengths are strict. The method is to show by example that some formula  $F$  is a consequence of **L** in the first sense but not in the second sense.

In order to show that finite semantic consequence is strictly stronger than semantic consequence, take **L** to be **KT** plus

$$(\Box P \wedge \neg \Box^2 P) \rightarrow \Diamond(\Box^2 P \wedge \neg \Box^3 P),$$

and take  $F$  to be 4. If the recession frame verifies this formula, it will show that 4 is not a semantic consequence of this **L**. It is clear that if a valuation  $V$  on  $\langle \omega, R \rangle$  rejects this formula  $m$  then

$$\begin{aligned} V(\Box P, m) &= T, V(\Box^2 P, m) = F, \\ V(\Box^2 P, m+1) &= F \text{ or } V(\Box^3 P, m+1) = T. \end{aligned}$$

In the second case,  $(m+1)Rm$  yields  $V(\Box P, m) = T$  and a contradiction. The first case requires some  $n$  with  $m \leq n$  such that  $V(\Box P, n) = F$ , and some  $k$  with  $n < k+1$  such that  $V(P, k) = F$ . Now  $m \leq k+1$  and so  $V(\Box P, m) = F$ , another contradiction.

To show that 4 is a finite semantic consequence of this  $\mathbf{L}$ , it is sufficient to show that if a model  $\langle W, R, V \rangle$  verifying  $\mathbf{L}$  rejects 4 then  $W$  is infinite. But in a model which rejects 4 we have  $v(\Box^2 P) \subset v(\Box P)$ , which serves as the induction basis for an inductive proof that  $v(\Box^{n+1} P) \subset v(\Box^n P)$ , for  $n \geq 1$ . The induction step uses the fact that

$$\begin{aligned} &\text{if } v(\Box^k P) - v(\Box^{k+1} P) \neq 0, \\ &\text{then } v(\Box^{k+1} P) - v(\Box^{k+2} P) \neq 0, \end{aligned}$$

from the verification of  $(\Box P \wedge \neg \Box^2 P), \Diamond(\Box^2 P \wedge \neg \Box^3 P)$ . The argument can be sharpened to prove the existence of an infinite ascending  $R$ -chain if

$$(P \wedge \Diamond^2 Q) \rightarrow (\Diamond Q \vee \Diamond^2(Q \wedge \Diamond P))$$

is added to  $\mathbf{L}$ . For suppose that  $\langle W, R \rangle$  verifies this formula and rejects 4, having  $x, y, z \in W$  such that  $xRy$  and  $yRz$  but not  $xRz$ . Then taking  $v(P) = \{x\}$  and  $v(Q) = \{z\}$  we have  $V(P, x) = T, V(\Diamond^2 Q, x) = T, V(\Diamond Q, x) = F$ , so that  $V(\Diamond^2(Q \wedge \Diamond P), x) = T$ . It follows that  $V(\Diamond P, z) = T$ , which can only hold if  $zRx$ . This fact, that if  $xRy$  and  $yRz$  but not  $xRz$  then  $zRx$ , can be used to construct an infinite ascending  $R$ -chain from the decreasing sequence  $v(\Box P), v(\Box^2 P), v(\Box P), \dots$  of subsets of  $W$ .

Note that this additional formula is also verified by the recession frame. For if  $V(P, m) = T, V(\Diamond^2 Q, m) = T, V(\Diamond Q, m) = F$  then  $V(Q, m-2) = T, V(Q \wedge \Diamond P, m-2) = T$ , and  $V(\Diamond^2(Q \wedge \Diamond P), m) = T$ . To show that semantic consequence is strictly stronger than general semantic consequence, it only remains to find a formula  $A$  which is verified by the veiled recession frame but is rejected by any frame with an infinite ascending  $R$ -chain. Thomason [1974] does give a complicated formula  $A$  with this property. Now, for each frame verifying the extension of  $\mathbf{KT}$  with the two formulas of recent paragraphs, rejection of 4 implies the rejection of  $A$ , so that verification of  $A$  requires the verification of 4. Taking  $\mathbf{L}$  to be the extension of  $\mathbf{KT}$  with the two stated formulas and  $A$ , 4 is a semantic consequence of  $\mathbf{L}$  but not a general semantic consequence of it.

Another proof that semantic consequence is strictly stronger than general semantic consequence goes as follows. Take  $\mathbf{L}$  to be  $\mathbf{KT} \cdot \mathbf{3M}$  plus

$$\Box(P \rightarrow \Box P) \rightarrow (\Diamond P \rightarrow P),$$

and take  $F$  to be  $P \rightarrow \Box P$ . This formula reduces the modal operators to triviality, with the corresponding condition on  $R$  that if  $xRy$  then  $x = y$ .

Define  $xR^n y$  on a frame  $\langle W, R \rangle$ , for  $n \geq 0$ , taking

$$\begin{aligned} xR^0 y &\text{ iff } x = y, \\ xR^1 y &\text{ iff } xRy, \\ xR^{n+1} y &\text{ iff } xRz_1, \dots, z_n Ry, \text{ for some } z_1, \dots, z_n \in W. \end{aligned}$$

Given a frame  $\langle W, R \rangle$  and  $x, y \in W$  such that  $xRy$  but not  $yR^n x$ , for  $n \geq 0$ , define  $V$  on  $\langle W, R \rangle$  by taking  $V(P, z) = T$  iff  $yR^n z$ , for some  $n \geq 0$ . It is easy to show that  $V(\Box(P \rightarrow \Box P), x) = T$ ,  $V(\Diamond P \rightarrow P, x) = F$ . Therefore in any frame  $\langle W, R \rangle$  which verifies  $\Box(P \rightarrow \Box P) \rightarrow (\Diamond P \rightarrow P)$  we have

(\*) if  $xRy$  then  $yR^n x$ , for some  $n \geq 0$ .

It can be shown that any reflexive frame  $\langle W, R \rangle$  which verifies  $\cdot 3$  satisfies the condition

$$\forall x \forall y \forall z ((xRy \wedge xRz) \rightarrow (\forall u (yRu \rightarrow zRu) \vee \forall v (zRv \rightarrow yRv))).$$

Call this condition strong connectedness, noting that connectedness is the special case with  $u = y$  and  $v = z$ , and that this condition can be derived from the ordinary one and transitivity. It can be shown that if a reflexive, strongly connected frame  $\langle W, R \rangle$  satisfies condition (\*), then it verifies  $\Box(P \rightarrow \Box P) \rightarrow (\Diamond P \rightarrow P)$ . As an application of this result, the recession frame verifies this formula. Thus the veiled recession frame verifies **L** but not  $P \rightarrow \Box P$ .

Suppose that  $\langle W, R \rangle$  is a reflexive, strongly connected frame which satisfies condition (\*). It can be shown that if  $\langle W, R \rangle$  also verifies  $M$  then  $xRy$  implies  $x = y$ , so that any frame which verifies **L** also verifies  $P \rightarrow \Box P$ . For given any  $x \in W$ , define

$$S_n = \{y : yR^n x \wedge \neg \exists m (m < n \wedge yR^m x)\},$$

for  $n \geq 0$ , and define  $V$  on  $\langle W, R \rangle$  by taking

$$V(P, y) = T \text{ iff } \exists m (y \in S_{2m}), \text{ for each } y \in W.$$

Now it can be shown that  $V(\Box \Diamond P, x) = T$ , so that  $V(\Diamond \Box P, x) = T$  by the verification of  $M$ . From this it can be deduced that  $V(\Box P, x) = T$ . Finally we suppose that  $xRy$  and  $x \neq y$ , and obtain a contradiction. For in this case we have  $V(P, y) = T$ , so that  $y \in S_{2m}$ , for some  $m \leq 1$ , and there are some  $z_1, \dots, z_{2m-1} \in W$  with  $yRz_1, \dots, z_{2m-1}Rx$  and not  $z_1Rx$ . Thus  $xRy, xRx, yRz_1$  but not  $xRz_1, xRx$  but not  $yRx$ —which contradicts strong connectedness when we put  $x$  for  $z, z_1$  for  $u$ , and  $x$  for  $v$ .

A third proof that semantic consequence is strictly stronger than general semantic consequence takes **L** to be **KT** plus

$$\Box(\Box(P \rightarrow \Box P) \rightarrow \Box^3 P) \rightarrow P$$

and takes  $F$  to be 4 again. for it can be shown that the veiled recession frame verifies this axiom of  $\mathbf{L}$ , but that each frame which verifies it is transitive. The interest of this example lies in the fact that the extension of  $\mathbf{S4}$  with this axiom is precisely  $\mathbf{S4Grz}$ .

Given a frame  $\langle W, R \rangle$ , consider the evaluation of any formula  $A$  in any model on  $\langle W, R \rangle$ . Our definition of valuations determines  $V(A, x)$  in terms of first-order logic applied to propositions of the form  $yRz$  and  $V(P, y) = T$  for propositional variables  $P$ . Replace each  $yRz$  by an atomic proposition  $R(y, z)$ , and each  $V(P, y) = T$  by an atomic proposition  $P(y)$ . Now the truth of  $A$  in  $\langle W, R \rangle$  can be expressed by a formula in second-order predicate logic with unary predicate parameters  $P, Q$ , etc. and one binary parameter  $R$ . This formula is known as the standard translation  $ST(A)$  of  $A$ . As we have seen,  $ST(A)$  is often equivalent to a first-order predicate formula in  $R$  alone, but this is not always the case. If we take some axiom system for second-order predicate logic then we can introduce yet another notion of consequence. Say that  $F$  is a second-order logical consequence of  $L$  iff  $ST(F)$  is derivable from the standard translations of the formulas of  $\mathbf{L}$ . In fact whenever we have shown that  $F$  is a semantic consequence of  $\mathbf{L}$ , we have used an argument in some unspecified, informal second-order logic to show that  $F$  is a second-order logical consequence of  $\mathbf{L}$ . Clearly semantic consequence is as strong as second-order logical consequence, which is as strong as modal logical consequence.

Van Benthem [1978; 1979a] discuss whether second-order logical consequence is strictly stronger than modal logical consequence. History added point to this question, in that transitivity was derived from  $ST(\mathbf{Grz})$  before 4 was derived in  $\mathbf{KG4z}$ . Of course the answer will depend on the axiomatisation used for second-order predicate logic. For example, close inspection of the informal argument for  $P \rightarrow \Box P$  being a second-order logical consequence of  $\mathbf{KT} \cdot \mathbf{3M}$  plus  $\Box(P \rightarrow \Box P) \rightarrow (\Diamond P \rightarrow P)$ , shows that it involves an Axiom of Choice. It turns out that if this is dropped, then a second-order derivation is no longer possible. Consider the axiomatic second-order logic with just the weak second-order substitution axiom

$$\forall P A \rightarrow S_P^B(A), \text{ for first-order formulas } B.$$

(Here  $S_P^B(A)$  is obtained from  $A$  by substituting  $S_x^t(B)$  for  $P(t)$  throughout, under suitable conditions.) the proof that  $P \rightarrow \Box P$  is not a general semantic consequence of this modal logic used the veiled recession frame, for which the possible values of formulas are the finite and cofinite subsets of  $\omega$ . It can be shown that these are precisely the subsets of  $\omega$  definable by first-order formulas with  $=$  and  $R$  as their only predicate parameters. Since these are the subsets of  $\omega$  to which the weak second-order substitution axiom applies, the same argument shows that  $P \rightarrow \Box P$  is not a second-order logical consequence of this modal logic.

The normal modal logic  $\mathbf{k}$  plus  $\Box(\Box P \rightarrow P) \rightarrow P$  is easily shown to be inconsistent. Define a general frame  $\langle \omega \cup \{\infty\}, R, P \rangle$  by taking

$$xRy \text{ iff } x > y \vee x = \infty,$$

and taking  $P$  to consist of the finite subsets of  $\omega$  and their complements in  $\omega \cup \{\infty\}$ . Then it is easy to show that  $\Box(\Box P \rightarrow P) \rightarrow P$  is satisfied at  $\infty$  by each valuation on  $\langle \omega \cup \{\infty\}, R, P \rangle$  (but not of course verified). So consider the non-normal logic  $\mathbf{K}$  plus  $\Box(\Box P \rightarrow P)$ , from which the rule of necessitation has been dropped. Now  $P \wedge \neg P$  is a second-order logical consequence of this logic, but not a modal logical consequence of it. Van Benthem [1979a] shows how to adopt this argument to give a normal modal logic  $\mathbf{L}$  and a formula  $F$ , such that  $F$  is a second-order logical consequence of  $\mathbf{L}$  but not a modal logical consequence of it.

## 20 INDEPENDENCE RESULTS ABOVE $\mathbf{S4}$

None of the logics used in the previous section is an extension of  $\mathbf{S4}$  (though  $\mathbf{KT} \cdot \mathbf{3M}$  plus  $\Box(P \rightarrow \Box P) \rightarrow (\Diamond P \rightarrow P)$  is a very strong logic in a sense, with no frames between it and triviality). Further, the methods of that section cannot be applied to extensions of  $\mathbf{S4}$ , since transitivity reduces the recession frame to a frame verifying  $\mathbf{S5}$ . For independence results above  $\mathbf{S4}$  we turn to a brief description of the complicated constructions of [Fine, 1972; Fine, 1974a].

In showing that finite semantic consequence is strictly stronger than semantic consequence,  $\mathbf{L}$  is taken to be  $\mathbf{S4}$  plus a certain axiom  $Y \rightarrow Z$ , and  $F$  is taken to be  $\neg Y$ . The frame used to show that  $\neg Y$  is not a consequence of  $\mathbf{S4}$  plus  $Y \rightarrow Z$  consists of three chains of points  $a_i, b_i, c_i$ , for  $i \geq 0$ , with  $R$  a lattice on them, and a final related pair of points  $d, e$ . This frame is illustrated in Figure 2 with  $R$  going from left to right. The points in these chains are described by corresponding formulas  $A_i, B_i, C_i$ , for  $i \geq 0$ , with  $A_0 = P, B_0 = Q, C_0 = R$ . Each  $A_{i+1}$  is

$$\Diamond A_i \wedge \Diamond B_i \wedge \neg \Diamond C_i,$$

expressing the fact that

$$a_{i+1}Ra_i \wedge a_{i+1}Rb_i \wedge \neg a_{i+1}Rc_i$$

and similarly for  $B_{i+1}, C_{i+1}$ . (Remember the frame formulas of the first half of Section 18.) Because of this construction there are theses of  $\mathbf{S4}$  describing the relations between the points. For example, not  $a_iRb_i$  and not  $a_iRc_i$ , and  $\vdash_{\mathbf{S4}} \Box(A_i \rightarrow (\neg \Diamond B_i \wedge \neg \Diamond C_i))$ .

The formula  $Y$  is simply a description of  $a_0, b_0, c_0, d$  in these terms, so that if  $V$  is defined on this frame by taking  $V(P) = \{a_0\}, V(Q) = \{b_0\}$ ,

$V(R) = \{c_0\}, V(S) = \{d\}$ , then  $V(Y, d) = T$ . Thus  $V(\neg Y, d) = F$  and  $\neg Y$  is rejected on this frame as required. but it is also true that if  $V$  is a valuation on this frame with  $V(Y, x) = T$  then  $x$  is  $d$  or  $e$  and  $V(P), V(Q), V(R)$  are a permutation of  $\{a_i\}, \{b_i\}, \{c_i\}$ , for some  $i \geq 0$ . The formula  $Z$  describes a property of four such points, so that again  $V(Z, x) = T$ . Thus  $V(Y \rightarrow Z, x) = T$ , for each  $x \in W$  and each valuation  $V$ , so that this frame verifies  $Y \rightarrow Z$  as required.

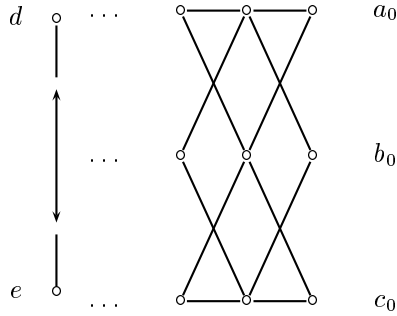


Figure 2.

These formulas also have the property that any frame  $\langle W, R \rangle$  which verifies  $Y \rightarrow Z$  and has a valuation  $V$  which satisfies  $Y$  must be infinite. First it can be shown that if  $V(Y, x) = T$  then  $V(\Diamond A_i, x) = T$ , for  $i \geq 0$ , by an induction on  $i$ . The induction basis with  $i = 0$  uses  $V(Y, x) = T$ , the induction step from  $i = 1$  uses  $V(Y \rightarrow Z, x) = T$ , and the other induction steps use theses of **S4** as above and  $V(Y' \rightarrow Z', x) = T$ , for substitution instances  $Y', Z'$  of  $Y, Z$ . Then it can be shown that  $\vdash_{\mathbf{S4}} A_i \rightarrow A_{i-j}$ , for each  $0 < j < i$ , by an induction using theses of **S4** above. It follows that there must be points  $a_i$  with  $xRa_i$  and  $V(A_i, a_i) = T$ , for  $i \geq 0$ , and with  $a_i \neq a_j$ , for  $i \neq j$ . Thus any finite frame which verifies **S4** plus  $Y \rightarrow Z$  must reject  $\neg Y$ , for otherwise it would satisfy  $Y$  and be infinite.

A similar strategy is used to show that semantic consequence is strictly stronger than general semantic consequence. At first sight Fine [1974] is not about general semantic consequence at all. Instead  $\langle W, R, V \rangle$  strongly verifies  $A$  iff all substitution instances of  $A$  are true in  $\langle W, R, V \rangle$ . But this is clearly equivalent to  $A$  being true on  $\langle W, R, P \rangle$ , where  $P = \{v(B) : B \text{ a formula}\}$ . Unfortunately there are a number of omissions and other typographical slips in this paper. See Bull [1982; 1983]. Again **L** is **S4** plus certain axioms  $E \rightarrow F$  and  $H$ , and the other formula is  $\neg E$ . The underlying frame used in showing that  $\neg E$  is not a general semantic consequence of this logic has two descending  $R$ -chains of points  $b_m, c_m$ , for  $m \geq 0$ , with  $R$  a

lattice on them. It also has a sequence of unrelated points  $a_m$  linked to an ascending  $R$ -chain of points  $d_m$ , for  $m \geq 0$ . (Note that because of the unrelated  $a_m$ 's, this frame is not of finite width.) This frame is illustrated in Figure 3 with  $R$  going from left to right. (As the page is finite, the ascending and descending parts have been overlapped. Each  $d_n$  should be linked to its  $a_n$  from the left, so that  $d_m R a_n$  for each  $m \leq n$ .) The points in the first three sequences are described by corresponding formulas  $A_m, B_m, C_m$ , for  $m \geq 0$ , with  $B_0 = Q_0, B_1 = Q_1, C_0 = R_0, C_1 = R_1$ . Each  $A_m$  is

$$\Diamond B_{m+1} \wedge \Diamond C_{m+1} \wedge \neg \Diamond B) m + 2 \wedge \neg \Diamond C_{m+2},$$

expressing the fact that

$$a_m R b_{m+1} \wedge a_m R c_{m+1} \wedge \neg a_m R b_{m+2} \wedge \neg a_m R c_{m+2},$$

and so on. Because of this construction there are theses of **S4** describing the relations between the points. For example,  $b_{i+1} R b_i$  but not  $b_{i+1} R c_i$ , and

$$\vdash_{\mathbf{S4}} \Box (B_{i+1} \rightarrow (\Diamond B_i \wedge \neg \Diamond C_i)).$$

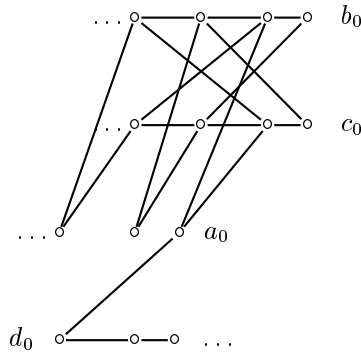


Figure 3.

The formula  $E$  is a description, from the viewpoint of  $d_0$ , of the frame given in Figure 4, together with the fact that there is an  $R$ -chain after it. Thus  $E$  is rejected at  $d_0$  on this frame by taking

$$\begin{aligned} v(P_0) &= \{d_{2m} : m \geq 0\}, & V(P_1) &= \{d_{2m+1} : m \geq 0\}, \\ V(Q_0) &= \{b_0\}, & V(Q_1) &= \{b_1\}, \\ V(R_0) &= \{c_0\}, & V(R_1) &= \{c_1\}. \end{aligned}$$

But it is also true that if  $V$  is a valuation on this frame with  $V(E, x) = T$  then  $V$  must give the propositional variables values which are points in this



configuration. Thus  $x$  must be some  $d_n$ . The formula  $F$  describes the  $R$ -chain beginning at  $d_1$  from the viewpoint of  $d_0$ , so that again  $V(F, x) = T$ . Thus  $V(E \rightarrow F, x) = T$ , for each  $x \in W$  and each valuation  $V$ , so that this

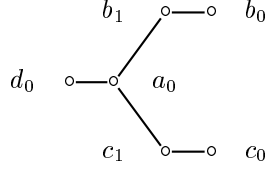


Figure 4.

frame verifies  $E \rightarrow F$ .

These formulas also have the property that any frame  $\langle W, R \rangle$  which verifies  $E \rightarrow F$  and has a valuation  $V$  which satisfies  $E$  at  $x \in W$  must have an infinite ascending  $R$ -chain after  $x$ . To see this, write  $E_n, F_n$  for the formulas obtained from  $E, F$  by replacing  $A_0, A_1$  with  $A_n, A_{n+1}$ , and so on. It can be shown by an induction on  $n$  that there is an  $R$ -chain  $\langle x = y_0, \dots, y_n \rangle$  such that  $V(E_n, y_n) = T$ , for  $n \geq 0$ . (Think of  $y_0, \dots, y_n$  as  $d_m, \dots, d_{m+n}$ .) The induction step uses  $V(E_n \rightarrow F_n, y_n) = T$  and these of **S4** as above. The crucial point is that

$$F_n = \diamond((P_0 \vee P_1) \wedge \neg \diamond A_n \wedge \diamond A_{n+1})$$

sends us from  $y_n$  with  $V(F_n, y_n) = T$  to some  $y_{n+1}$  with  $y_n R y_{n+1}$  and  $V(\diamond A_{n+1}, y_{n+1}) = T$ . Using this infinite ascending  $R$ -chain after  $x$ , it is easy to reject

$$H = S \wedge \Box(S \rightarrow \diamond((\neg S \wedge T) \wedge \diamond((\neg S \wedge \neg T) \wedge \diamond S)))$$

at  $x$  with a suitable valuation. Thus any frame which verifies **S4** plus  $E \rightarrow F$  and  $H$  must reject  $\neg E$ .

Finally, consider again the frame illustrated in Figure 4 above, and the valuation  $V$  on it used to satisfy  $E$  at  $d_0$ . This valuation determines a general frame on it, in which  $P$  is the set of values  $v(B)$  of all formulas  $B$ . We already know that  $E \rightarrow F$  is verified by this general frame and that  $\neg E$  is rejected by it, so it only remains to show that it verifies  $H$ . Suppose then that  $V(\neg H', x) = T$ , for some  $x \in W$ , and some substitution instance  $H'$  of  $H$ , and try to obtain a contradiction. It is clear that  $x$  must be  $d_m$ , for some  $m \geq 0$ , for  $H'$  can only be rejected on a proper cluster or an infinite ascending  $R$ -chain. Note that  $H'$  is constructed from three incompatible propositions  $a, \neg A \wedge B, \neg A \wedge \neg B$ . Further,  $A$  and  $B$  are constructed from propositional variables and formulas  $\Box C_1, \dots, \Box C_k$  with  $\neg, \wedge, \vee$ . Note that

after some  $d_n$ , the formulas  $\Box C_1, \dots, \Box C_k$  must have fixed truth values  $V(\Box C_i, d_j)$ . Consider  $j_1, j_2, j_3$  with  $n \leq j_1, j_2, j_3$  and

$$V(A, d_{j_1}) = V(\neg A \wedge B, d_{j_2}) = V(\neg A \wedge \neg B, d_{j_3}) = T.$$

At least one pair of these  $j$ 's must have an even difference, e.g.  $j_2$  and  $j_3$ . In this case

$$V(\neg A \wedge \neg B, d_{j_2}) = V(\neg A \wedge B, d_{j_2}) = T,$$

using the construction of each  $V(P_i, d_j)$  and the fact about each  $V(\Box C_i, d_j)$ . But this contradicts the mutual incompatibility of these three formulas.

## 21 NEIGHBOURHOOD FRAMES

A *neighbourhood frame*  $\langle U, N \rangle$  consists of a set  $U$  and a function  $N : U \rightarrow \mathfrak{B}(\mathfrak{B}(U))$ . Thus each value  $N(x)$  of  $N$  is a subset of  $\mathfrak{B}(U)$ , the subsets of  $U$  in  $N(x)$  being known as the neighbourhoods of  $x$ . Valuations  $V$  and models on  $\langle U, N \rangle$  are defined as for ordinary frames except that

$$V(\Box A, x) = T \text{ iff } V(A) \in N(x).$$

The canonical neighbourhood model  $\langle U_L, N_L, V_L \rangle$  for a logic  $\mathbf{L}$  is defined as for ordinary frames except that

$$S \in N(F) \text{ iff } \exists A(\Box A \in F \wedge S = \{G : A \in G\}).$$

Satisfaction, verification, and neighbourhood semantic consequence are defined as for ordinary frames. The minimal normal modal logic  $\mathbf{K}$  is characterised by the class of neighbourhood frames  $\langle U, N \rangle$  in which each  $N(x)$  is a filter on  $U$ . Such a neighbourhood frame is said to be normal, and determines a modal algebra on  $\mathfrak{B}(U)$ . Each ordinary frame  $\langle W, R \rangle$  determines a normal neighbourhood frame  $\langle W, N \rangle$  by taking

$$N(x) = \{S : \{y : xRy\} \subseteq S\}, \text{ for each } x \in W.$$

Here  $\langle W, N \rangle$  verifies the same formula as  $\langle W, R \rangle$ . Also each normal neighbourhood frame  $\langle U, N \rangle$  determines an ordinary frame  $\langle U, R \rangle$  by taking

$$xRy \text{ iff } y \in \cap N(x), \text{ for each } x, y \in U.$$

But here  $\langle U, R \rangle$  may not be equivalent to  $\langle U, N \rangle$ , so that we must ask whether semantic consequence is strictly stronger than normal neighbourhood semantic consequence.

(Neighbourhood frames seem to have been created independently by Dana Scott and Montague. See [Seegerberg, 1971] for a full discussion of

them. Gerson [1975] established that normal neighbourhood semantic consequence was strictly stronger than general semantic consequence, while Gerson [1976; 1975a] established that ordinary semantic consequence was strictly stronger than it.)

In showing that normal neighbourhood semantic consequence is strictly stronger than general semantic consequence, the arguments of Thomason [1974] and Fine [1974] can be taken over with only slight alterations. These come when showing that each normal neighbourhood frame which verifies the logic concerned also verifies the other formula 4 or  $E$ . For the first case, if  $\langle U, N \rangle$  verifies S. K. Thomason's axiom

$$(P \wedge \diamond^2 Q) \rightarrow (\diamond Q \vee \diamond^2(Q \wedge \diamond P)),$$

and there are  $R, S, T \subseteq U$  with  $R \subseteq \mathbf{m}S$  and  $S \subseteq \mathbf{m}T$  but not  $R \subseteq \mathbf{m}T$ , then  $T \cap \mathbf{m}R$  is nonempty. Now the proof that, if  $\langle W, R \rangle$  verifies Makinson's axiom

$$(\Box P \wedge \neg \Box^2 P) \rightarrow \diamond(\Box^2 P \wedge \neg \Box^3 P)$$

but rejects 4 then it must be infinite, can be sharpened as follows. If  $\langle U, N \rangle$  verifies both these axioms but rejects 4 then  $U$  contains an infinite sequence of distinct subsets  $W_1, W_2, W_3, \dots$  with  $W_i \subseteq \mathbf{m}W_j$  if  $i < j$ . S. K. Thomason's second axiom  $A$  can be rejected on any  $\langle U, N \rangle$  with this property, so that if a normal neighbourhood frame verifies the logic of [Thomason, 1974] then it verifies 4.

For the second case, suppose that  $\langle U, N \rangle$  verifies  $E \rightarrow F$  and has valuation  $V$  which satisfies  $RE$  at  $u \in U$ . Then it can be shown that  $U$  contains an infinite sequence of distinct subsets  $W_1, W_2, W_3, \dots$  with  $u \in W_i$ , for  $i \geq 0$ , and  $W_i \subseteq \mathbf{m}W_j$  if  $i < j$ , taking  $W_i = v(E_i)$ , for  $i \geq 0$ . Using this finite sequence of sets it is easy to reject  $\neg H$  with  $V$  at  $u$ , so that if a normal neighbourhood frame verifies **S4** plus  $E \rightarrow F$  and  $H$  then it verifies  $\neg E$ .

Gerson [1976] uses a minor variation on the logic  $\mathbf{L}$  of the 'noncompactness' proof in [Thomason, 1972a]. A very complicated argument shows that this logic is verified by a certain normal neighbourhood frame, which is largely determined by an ordinary frame consisting of all finite fragments of the recession frame, with one point added. A further three points are then added and their neighbourhoods specified. Otherwise the argument is like that of [Thomason, 1972a]. Gerson [1975a] uses a version  $\mathbf{L}'$  of the logic  $\mathbf{L}$  of [Fine, 1974], with  $E \rightarrow F_n$  for  $n \geq 1$ . That any ordinary frame which verifies  $\mathbf{L}'$  also verifies  $\neg E$  goes as before. A complicated argument shows that  $\mathbf{L}'$  is verified by a certain normal neighbourhood frame, which is largely determined by an ordinary frame similar to that of [Fine, 1974] illustrated above. The difference is that the infinite ascending  $R$ -chain of  $d_m$ 's has been replaced by an infinity of finite ascending  $R$ -chains  $\langle d_{m,1}, \dots, d_{m,m} \rangle$  for  $m \geq 1$ . A further two points are then added and their neighbourhood specified. Otherwise the argument is fairly similar to that of [Fine, 1974].

22 ELEMENTARY EQUIVALENCE AND  $D$ -PERSISTENCE

Consider the first-order predicate logic with binary predicate constants  $=$  and  $R$ . Write  $\mathfrak{F} \models A$  iff the formula  $A$  of predicate logic is true of the frame  $\mathfrak{F}$ , and similarly for  $\mathfrak{F} \models \Gamma$ , where  $\Gamma$  is a set of predicate formulas. A class  $X$  of frames is *elementary* iff

$$X = \{\mathfrak{F} : \mathfrak{F} \models A\}, \text{ for some formula } A \text{ of predicate logic,}$$

$\Delta$ -*elementary* iff it is an intersection of elementary classes,  $\Sigma$ -*elementary* iff it is a union of elementary classes, and  $\Sigma\Delta$ -*elementary* iff it is an intersection of  $\Sigma$ -elementary classes. Note that  $X$  is  $\Delta$ -elementary iff it is axiomatic, with

$$X = \{\mathfrak{F} : \mathfrak{F} \models \Gamma\}, \text{ for some set } \Gamma \text{ of formulas of predicate logic.}$$

And  $X$  is  $\Sigma\Delta$ -elementary iff it is closed under elementary equivalence, where  $\mathfrak{F}$  and  $\mathfrak{G}$  are *elementarily equivalent* iff

$$\mathfrak{F} \models A \text{ iff } \mathfrak{G} \models A, \text{ for each formula } A \text{ of predicate logic.}$$

The importance of elementarily equivalent frames for modal logic lies in the following lemma. Given a general frame  $\mathfrak{F} = \langle W, R, P \rangle$ , there is a general frame  $\mathfrak{F}' = \langle W', R', P' \rangle$  such that  $\mathfrak{F}'$  is 1- and 2'-saturated (see Section 10),  $\mathfrak{F}^+$  and  $\mathfrak{F}'^+$  are isomorphic,  $\langle W, R \rangle$  and  $\langle W', R' \rangle$  are elementarily equivalent, and there is a frame morphism from  $\langle W', R' \rangle$  onto  $(\mathfrak{F}^+)_\sharp$ .

Alternatively, consider modal logic as usual, again writing  $\langle \mathfrak{F}, V \rangle \models A$  iff the formula  $A$  of modal logic is true in the model  $\langle \mathfrak{F}, V \rangle$ , and so on. A class  $X$  of frames is *modal elementary* iff

$$X = \{\mathfrak{F} : \mathfrak{F} \models A\}, \text{ for some formula } A \text{ of modal logic,}$$

and is *modal axiomatic* iff

$$X = \{\mathfrak{F} : \mathfrak{F} \models \Gamma\}, \text{ for some set } \Gamma \text{ of formulas of modal logic.}$$

Again modal axiomatic is equivalent to modal  $\Delta$ -elementary. A set  $\Gamma$  of formulas of modal logic is *c-persistent* iff  $\langle W_{K\Gamma}, R_{K\Gamma} \rangle \models \Gamma$  (the canonical frame for the normal modal logic  $K$  plus  $\Gamma$ ), *d-persistent* iff if  $\langle \mathfrak{F}, P \rangle \models \Gamma$  then  $\mathfrak{F} \models \Gamma$ , for each descriptive general frame  $\langle v, P \rangle$ , and *r-persistent* iff if  $\langle \mathfrak{F}, P \rangle \models \Gamma$  then  $\mathfrak{F} \models \Gamma$ , for each refined general frame  $\langle \mathfrak{F}, P \rangle$ . Note that *r-persistent* implies *d-persistent*, implies *c-persistent*, implies characterised by frames. Many proofs that a logic is characterised by frames involve *c-persistence*. However  $\mathbf{K}$  plus  $\Box(\Box P \rightarrow P) \rightarrow \Box P$  is characterised by frames but is not *c-persistent* (see [Seegerberg, 1971; van Benthem, 1979]).

A class of frames verifies a *d-persistent* set of formulas iff it is closed under subframes, frame-morphic images, disjoint unions, and both it and

its complement are closed under the construction  $(\mathfrak{F}^+)_{\sharp}$ . We know from Section 10 that the class of frames verifying a set of formulas is closed under subframes and frame-morphic images, and that any frame  $\mathfrak{F}$  is frame-isomorphic to a subframe of  $(\mathfrak{F}^+)_{\sharp}$ . The latter point can be extracted from the proof that a descriptive frame  $\mathfrak{F}$  is isomorphic to  $(\mathfrak{F}^+)_{+}$ , and shows that the complement of a class of frames verifying a set of formulas is closed under the construction  $(\mathfrak{F}^+)_{\sharp}$ . It is easy to show that the class of frames verifying a set of formulas is closed under disjoint unions. If a frame  $\mathfrak{F}$  verifies a set  $\Gamma$  of formulas then so do the modal algebra  $\mathfrak{F}^+$  and the descriptive general frame  $(\mathfrak{F}^+)_{+}$ , by Section 10. If  $\Gamma$  is a  $d$ -persistent set of formulas then  $(\mathfrak{F}^+)_{\sharp}$  also verifies  $\Gamma$ , so that the class of frames verifying a  $d$ -persistent set of formulas is closed under the construction  $(\mathfrak{F}^+)_{\sharp}$ .

Conversely, suppose that a class  $X$  of frames satisfies these closure conditions. Consider the class

$$X^+ = \{\mathfrak{F}^+ : \mathfrak{F} \in X\}$$

of modal algebras and the set

$$\Gamma = \{A : \mathfrak{F}^+ \models A, \text{ for each } \mathfrak{F}^+ \in X^+\}$$

of formulas. If  $\mathfrak{F} \in X$  then  $\mathfrak{F}^+ \models \Gamma$  and so  $\mathfrak{F} \models \Gamma$  by Section 10. For the other direction, suppose that  $\mathfrak{F} \models \Gamma$  and so  $\mathfrak{F}^+ \models \Gamma$ . The set  $\Gamma$  of formulas is closely analogous to the set of equations in modal algebra verified by  $X^+$ , so that the set of all modal algebras verifying  $\Gamma$  is the variety generated by  $X^+$ . Using a theorem of Birkhoff's on varieties, a modal algebra  $\mathfrak{F}^+$  verifies  $\Gamma$  iff it is a homomorphic image of a subalgebra of a direct product of modal algebras  $\{\mathfrak{F}_i^+ : i \in I\}$  in  $X^+$ . Checking the definition of the disjoint union  $\Sigma_{i \in I} \mathfrak{F}_i \in X$ , the direct product  $\Pi_{i \in I} \mathfrak{F}_i^+$  is isomorphic to  $(\Sigma_{i \in I} \mathfrak{F}_i)^+$ . Taking the carrier of the subalgebra to be  $P$ , this subalgebra is  $\langle \Sigma_{i \in I} \mathfrak{F}_i, P \rangle^+$ . Thus there is a homomorphism from  $\langle \Sigma_{i \in I} \mathfrak{F}_i \rangle^+$  onto  $\mathfrak{F}^+$ . As in Section 10, we can dualise from the category of modal algebras to the category of descriptive frames, with homomorphic images going to subframes and subalgebras going to frame-morphic images. Thus  $(\mathfrak{F}^+)_{+}$  is frame-isomorphic to a subframe of  $((\Sigma_{i \in I} \mathfrak{F}_i, P)^+)_{+}$ , and  $((\Sigma_{i \in I} \mathfrak{F}_i, P)^+)_{+}$  is a frame-morphic image of  $((\Sigma_{i \in I} \mathfrak{F}_i)^+)_{+}$ , and  $((\Sigma_{i \in I} \mathfrak{F}_i, P)^+)_{+}$  is a frame-morphic image of  $((\Sigma_{i \in I} \mathfrak{F}_i)^+)_{+}$ . Going to the underlying frames,  $(\mathfrak{F}^+)_{\sharp}$  is frame-isomorphic to a subframe of a frame-morphic image of  $((\Sigma_{i \in I} \mathfrak{F}_i)_{\sharp}^+)_{\sharp}$ . Since  $\Sigma_{i \in I} \mathfrak{F}_i \in X$  and  $X$  is closed under subframes, frame-morphic images, and the construction  $(\mathfrak{F}^+)_{\sharp}$ , we have  $(\mathfrak{F}^+)_{\sharp} \in X$ . Since the complement of  $X$  is also closed under the construction  $(\mathfrak{F}^+)_{\sharp}$ , we have  $\mathfrak{F} \in X$ . Thus  $\mathfrak{F} \models \Gamma$  iff  $\mathfrak{F} \in X$ , so that  $X$  is the class of frames verifying  $\Gamma$ .

It remains to show that  $\Gamma$  is  $d$ -persistent. Supposing that a descriptive frame  $\langle \mathfrak{F}, P \rangle$  verifies  $\Gamma$ , and repeating the previous argument with  $\langle \mathfrak{F}, P \rangle^+$  in place of  $\mathfrak{F}^+$ , will show that  $((\mathfrak{F}, P)^+)_{\sharp} \in X$ . But the descriptive frame

$\langle \mathfrak{F}, P \rangle$  is frame-isomorphic to  $(\langle \mathfrak{F}, P \rangle^+)_+$  by Section 10, so that going to the underlying frames yields that  $\mathfrak{F}$  is frame-isomorphic to  $(\langle \mathfrak{F}, P \rangle^+)_\#$ . Thus  $\mathfrak{F} \in X$  and, hence,  $\mathfrak{F}$  verifies  $\Gamma$ , so that  $\Gamma$  is  $d$ -persistent.

Consider a set of formulas  $\Gamma$  characterised by the class  $X$  of frames which verify it. Then  $\Gamma$  is  $d$ -persistent iff  $X$  is closed under the construction  $(\mathfrak{F}^+)_\#$ . If  $\Gamma$  is  $d$ -persistent then one direction of the result applies, and yields  $X$  closed under the construction  $(\mathfrak{F}^+)_\#$ . If  $X$  is the class of frames verifying  $\Gamma$  and is closed under the construction  $(\mathfrak{F}^+)_\#$ , then the other direction of the result applies. In this case it yields that  $X$  is the class of frames verifying some  $d$ -persistent set of formulas. Inspection of the proof shows that this is the set of semantic consequences of  $\Gamma$ . But since  $\Gamma$  is characterised by frames, it equals its set of semantic consequences so that  $\Gamma$  is a  $d$ -persistent set of formulas.

Combining our lemmas elementary equivalence and  $d$ -persistence yields two important theorems. Firstly, if a set  $\Gamma$  of formulas is characterised by the class  $X$  of frames which verify it and  $X$  is closed under elementary equivalence, then  $\Gamma$  is  $d$ -persistent. For then  $X$  is closed under the construction  $(\mathfrak{F}^+)_\#$  by the first lemma, and so  $\Gamma$  is  $d$ -persistent by the second lemma. Secondly, given a class  $X$  of frames closed under elementary equivalence,  $X$  is modal axiomatic iff it is closed under subframes, frame -morphic images, disjoint unions, and its complement is closed under the construction  $(\mathfrak{F}^+)_\#$ . We have already seen that a modal axiomatic class of frames has these closure properties. If  $X$  is closed under elementary equivalence and these conditions then it satisfies all the closure properties of the theorem on  $d$ -persistent sets, using the first lemma. Thus  $X$  is modal axiomatic; indeed it is determined by a  $d$ -persistent set of formulas.

The presentation here has followed the elegant van Benthem [1979]. The first paper in this area was the important [Fine, 1975]. It defined notions of modal saturation and persistence, and introduced the lemma on classes of frames closed under elementary equivalence. (It worked in terms of models rather than of general frames, but the analogy is close.) It proved the slightly weaker result, that if a set  $\Gamma$  of formulas is characterised by the class  $X$  of frames which verify it and  $X$  is closed under elementary equivalence, then  $\Gamma$  is  $c$ -persistent. The theorem giving the closure conditions for a class  $X$  of frames, which is closed under elementary equivalence, to be axiomatic, is Goldblatt's contribution to Goldblatt and Thomason [1975]. The proof was roughly similar to the one here but more complicated. It too started with the duality between varieties of modal algebras and classes of descriptive frames, and used Fine's lemma and the properties of  $(\mathfrak{F}^+)_\#$  to bridge the gap between the frames and descriptive frames. Fine [1975] used classical modal theory to show that if a set  $\Gamma$  of formulas is  $r$ -persistent then the class  $X$  of frames which verify  $\Gamma$  is  $\Delta$ -elementary (and of course characterises the normal modal logic  $\mathbf{K}$  plus  $\Gamma$ ). It also gives counter-examples to the converse of both its theorems. In the second case the counter-example is

**S4 · 3M.** We know that it is characterised by the elementary class of frames determined by certain conditions. and it is verified by the refined general frame  $\langle \omega, \leq, P \rangle$ , where  $P$  is the set of finite and cofinite subsets of  $\omega$ , but  $\langle \omega, \leq \rangle$  rejects  $M$ .

### 23 MODAL ELEMENTARY AND AXIOMATIC CLASSES

The main construction for this topic is the ultraproduct of frames. Consider frames  $\mathfrak{F}_i \langle W_i, R_i \rangle$  for  $i \in I$ , and an ultrafilter  $G$  on  $I$ . Remember that the members  $f$  of the direct product  $\prod_{i \in I} W_i$  are the functions  $f : I \rightarrow \cup_{i \in I} W_i$  such that  $f(i) \in W_i$ , for each  $i \in I$ . Define an equivalence relation  $\simeq$  on  $\prod_{i \in I} W_i$  by taking

$$f \simeq g \text{ iff } \{i : f(i) = g(i)\} \in G$$

and consider the equivalence classes  $[f]$  under  $\simeq$ . The *ultraproduct*  $\mathfrak{F}_G = \prod_{i \in I} \mathfrak{F}_i / G = \langle W_G, R_G \rangle$  is defined by taking

$$W_G = \prod_{i \in I} W_i / G = \{[f] : f \in \prod_{i \in I} W_i\},$$

$$[f] R_G [g] \text{ iff } \{i : f(i) R_i g(i)\} \in G.$$

To extend this definition to general frames  $\langle F_i, P_i \rangle$ , for  $i \in I$ , it can first be shown that

$$\text{if } f \simeq g \text{ then } \{i : f(i) \in S(i)\} \in G \equiv \{i : g(i) \in S(i)\} \in G,$$

$$S \simeq T \text{ iff } \forall f (\{i : f(i) \in S(i)\} \in G \equiv \{i : f(i) \in T(i)\} \in G),$$

for  $f, g \in \prod_{i \in I} W_i$  and  $S, T \in \prod_{i \in I} P_i$ . This justifies defining

$$[S] = \{[f] : \{i : f(i) \in S(i)\} \in G\},$$

for each  $S \in \prod_{i \in I} P_i$ , and taking

$$P_G = \left\{ [S] : S \in \prod_{i \in I} P_i \right\}.$$

Here the definition of a general frame requires that  $P_G$  be a subalgebra of  $(\prod_{i \in I} \mathfrak{F}_i / G)^+$ . for the case  $\mathbf{m}_{R_G}$  we need

$$\mathbf{m}_{R_G} [S] = [\mathbf{m}S],$$

where

$$(\mathbf{m}S)(i) = \mathbf{m}_{R_i}(S(i)), \text{ for each } i \in I,$$

for each  $S \in \prod_{i \in I} P_i$ . We have

$$\begin{aligned}
 & [f] \in \mathbf{m}_{R_G} [S] \\
 & \text{iff } [f]R_G[g], \text{ for some } [g] \in [S], \\
 & \text{iff } \{i : f(i)R_i g(i)\} \in G \text{ and } \{i : g(i) \in S(i)\} \in G, \\
 & \quad \text{for some } g \in \prod_{i \in I} W_i, \\
 & \text{iff } \{i : f(i)R_u g(u) \wedge g(i) \in S(i)\} \in G, \text{ for some } g \in \prod_{i \in I} W_i, \\
 & \text{iff } \{i : f(i) \in \mathbf{m}_{R_i}(S(i))\} \in G \\
 & \text{iff } \{i : f(i) \in (\mathbf{m}S)(i)\} \in G \\
 & \text{iff } [f] \in [\mathbf{m}S].
 \end{aligned}$$

Given a valuation  $V_i$  on each general frame  $\mathfrak{F}_i = \langle W_i, R_i, P_i \rangle$ , for each  $i \in I$ , define a valuation  $V_G$  on  $\mathfrak{F}_G = \langle W_G, R_G, P_G \rangle$  by taking

$$V_G(P, [f]) = T \text{ iff } [f] \in [V_G(P)] \text{ iff } \{i : V_i(P, f(i)) = T\} \in G,$$

for each propositional variable  $P$ , and apply the defining conditions for valuations. Then the argument like that of the previous paragraph shows that

$$V_G(A, [f]) = T \text{ iff } \{i : V_i(A, f(i)) = T\} \in G,$$

for each formula  $A$ . It is now easy to show that

$$\mathfrak{F}_G \models A \text{ iff } \{i : \mathfrak{F}_i \models A\} \in G.$$

Going from left to right, note that if not  $\{i : \mathfrak{F}_i \models A\} \in G$  then  $\{i : \text{not } \mathfrak{F}_i \models A\} \in G$ , since  $G$  is an ultrafilter. Now use valuations  $V_i$  and points  $f(i)$  with  $V_i(A, f(i)) = F$ , for each  $i$  in the member of  $G$ . Note that taking  $P_i = \mathfrak{B}(W_i)$ , for each  $i \in I$ , does not yield  $P_G = \mathfrak{B}(\prod_{i \in I} W_i/G)$ , so that  $\prod_{i \in I} \langle \mathfrak{F}_i, \mathfrak{B}(W_i) \rangle / G$  is not the same as  $\prod_{i \in I} \mathfrak{F}_i / G$ . Therefore this result for ultraproducts of general frames yields only

$$\text{if } \mathfrak{F}_G \models A \text{ then } \{i : \mathfrak{F}_i \models A\} \in G,$$

for ultraproducts of ordinary frames. (As we shall note later,  $M$  is a counterexample to the converse.) It follows that if  $X$  is a modal elementary class of frames, then its complement is closed under ultraproducts. Similarly, if  $X$  is a modal axiomatic class of frames then its complement is closed under ultrapowers. Here an ultrapower  $\mathfrak{F}^I/G$  is the ultraproduct  $\prod_{i \in I} \mathfrak{F}_i / G$  for which  $\mathfrak{F}_i = \mathfrak{F}$ , for each  $i \in I$ .

Classical model theory proves the following characterisations of the various kinds of elementary classes. A class  $X$  of frames is elementary iff  $X$  and  $\neg X$  are closed under frame isomorphism and ultraproducts. Class  $X$  is  $\Delta$ -elementary iff  $X$  is closed under frame isomorphism and ultraproducts, and  $\neg X$  is closed under ultrapowers. Class  $X$  is  $\Sigma$ -elementary iff  $X$  is closed under ultrapowers, and  $\neg X$  is closed under frame isomorphism and



ultraproducts. Class  $X$  is  $\Sigma\Delta$ -elementary iff  $X$  and  $\neg X$  are closed under isomorphism and ultrapowers. Combining the results so far, it is easy to show that a modal elementary class of frames is elementary if it is closed under ultraproducts. And a modal axiomatic class is  $\Delta$ -elementary iff it is closed under ultraproducts.

Further, a class  $X$  of frames closed under frame isomorphism, subframes, disjoint unions, and ultrapowers is also closed under ultraproducts. For, given  $\mathfrak{F}_i \in X$ , for  $i \in I$ , it is easy to show that  $\Pi_{i \in I} \mathfrak{F}_i / G$  is isomorphic to a subframe of  $(\Sigma_{i \in I} \mathfrak{F}_i)^I / G$ . Now it is easy to show that for a modal elementary class  $X$  of frames, all the following conditions are equivalent:  $X$  is elementary,  $X$  is  $\Sigma$ -elementary,  $X$  is  $\Delta$ -elementary,  $X$  is  $\Sigma\Delta$ -elementary,  $X$  is closed under ultrapowers,  $X$  is closed under ultraproducts. For a modal axiomatic class  $X$  of frames, the conditions elementary and  $\Sigma$ -elementary are equivalent, and the following conditions are equivalent:  $X$  is  $\Delta$ -elementary,  $X$  is  $\Sigma\Delta$ -elementary,  $X$  is closed under ultrapowers,  $X$  is closed under ultraproducts.

Ultraproducts of frames were introduced in [Goldblatt, 1975], and are described in detail in [Goldblatt, 1976]. Goldblatt [1975] obtained some of the results above, and gave a complicated example of frames which verify  $M$  but have an ultraproduct which does not. It follows that the class of frames verifying  $M$  is not (first-order) axiomatic, although [Fine, 1975] shows that **KM** is characterised by the class of frames verifying it. (Therefore this class of frames is characterised by some formula of second-order predicate logic, as in the last part of Section 19.) This result was also proved independently in [van Benthem, 1975], by a direct method. Van Benthem [1976] proved more of the results above, the published version using Goldblatt's ultraproducts. The picture was completed in [Goldblatt, 1976], where there is also a more detailed explanation of the ultraproduct of frames which verify  $M$ .

## 24 TWO FURTHER RESULTS

We have found closure conditions for a modal axiomatic class of frames, provided that it is closed under elementary equivalence and, hence, includes enough saturated frames. Can closure conditions for axiomatic classes of frames still be found when this condition is dropped? A rather complicated answer is provided in [Goldblatt and Thomason, 1975] (originally part of [Thomason, 1975]). Given a frame  $\langle W, R \rangle$ , choosing a general frame  $\langle W, R, P \rangle$  represents a choice of which 'propositions' are to be considered. In then forming  $\langle U, S \rangle = (\langle W, R, P \rangle^+)_\#$ , the members of  $U$  are the ultrafilters on  $P$ , representing 'states-of-affairs', i.e. maximal consistent sets of 'propositions'. The natural definition of  $S$  on these 'states-of-affairs' is, as usual,

$$uSv \text{ iff } (\forall X \in \mathfrak{m}_R)(X \in v \rightarrow \mathfrak{m}_R X \in u).$$

Under what conditions will  $\langle U, s \rangle$  again verify the formulas verified by  $\langle W, R \rangle$ ? Firstly, there must be no ‘new propositions’ in  $\langle U, S \rangle$ , i.e.

$$(\forall Y \subseteq U)(\exists X \in P)(Y = \phi(X)),$$

where  $\phi(X) = \{u \in U : X \in u\}$ , or

$$(\forall Y \subseteq U)(\exists X \in P)(u \in Y \rightarrow X \in u).$$

Secondly, to carry out the necessary induction step on the value of  $\Diamond A$ , we must have

$$(\forall u \in U)(\forall X \in P)(\mathbf{m}_R X \in u \rightarrow (\exists v \in u)(uSv \wedge X \in v)).$$

If  $\langle U, S \rangle$  satisfies these conditions for the carrier  $P$  of some subalgebra of  $\langle W, R \rangle^+$ , then we say that  $\langle U, S \rangle$  is SA-based on  $\langle W, R \rangle$ .

It can be shown, by a fairly difficult proof, that  $\langle U, S \rangle$  is frame-isomorphic to a frame SA-based on  $\langle W, R \rangle$  iff  $\langle U, S \rangle^+$  is a homomorphic image of a subalgebra of  $\langle W, R \rangle^+$ . Now a class of frames is modal axiomatic if it is closed under frame isomorphism, nontrivial disjoint unions, and the construction of  $\langle U, S \rangle$  SA-based on  $\langle W, R \rangle$ . It is easy to show that a modal axiomatic class is closed under these conditions. For the converse, suppose that a class  $X$  of frames is closed under these conditions. As in the theorem in Section 23 on the closure conditions for the class of frames verifying a  $d$ -persistent set of formulas, we take

$$\begin{aligned} X^+ &= \{\mathfrak{F}^+ : \mathfrak{F} \in X\}, \\ \Gamma &= \{A : \mathfrak{F}^+ \vDash A \wedge \mathfrak{F}^+ \in X^+\}, \end{aligned}$$

and show that  $X$  is the class of frames verifying  $\Gamma$ . Again  $\mathfrak{F}^+$  verifies  $\Gamma$  iff it is a homomorphic image of a subalgebra for a direct product of modal algebras  $\{\mathfrak{F}_i^+ : i \in I\}$  in  $X^+$ , where the direct product is isomorphic to  $(\Sigma_{i \in I} \mathfrak{F}_i)^+$  for  $\Sigma_{i \in I} \mathfrak{F}_i \in X$ . By the lemma stated above  $\mathfrak{F}$  must be SA-based on  $\Sigma_{i \in I} \mathfrak{F}_i$ , and so  $\mathfrak{F} \in X$ . Thus if  $\mathfrak{F} \vDash \Gamma$  then  $\mathfrak{F} \in X$ , and the converse is clear.

We are familiar with the duality between modal algebras and descriptive frames, and with the fact that we must shift from frames to descriptive frames before a duality can be established. Can we, as an alternative, shift to some other kind of algebra and then establish a duality with frames proper? This is done in [Thomason, 1975]. The appropriate algebras are the complete atomic modal algebras, i.e. modal algebras based on complete atomic Boolean algebras with

$$\begin{aligned} \mathbf{1} \cap \{b_i : i \in I\} &= \cap \{\mathbf{1}b_i : i \in I\}, \\ \mathbf{m} \cup \{b_i : i \in I\} &= \cup \{\mathbf{m}b_i : i \in I\}. \end{aligned}$$

An atom of a Boolean algebra  $\mathfrak{B} = \langle B, 0, 1, -, \cap, \cup \rangle$  is an element  $a \in B$  with

$$a \leq b \vee a \cap b = 0, \text{ for each } b \in B.$$

Then  $\mathfrak{B}$  is atomic iff

$$\forall b \exists a (a \text{ an atom} \wedge a \leq b),$$

and is complete iff it is closed under the operations  $\cap$  and  $\cup$  for arbitrary subsets  $\{b_i : i \in I\}$  of  $B$ . In a complete atomic Boolean algebra, each element  $b$  is determined by the set of atoms  $a$  with  $a \leq b$ . the appropriate morphisms for the category of complete atomic modal algebras are the complete homomorphisms, i.e. the homomorphisms  $\phi$  with

$$\phi(\cup\{b_i : i \in I\}) = \cup\{\phi(b_i) : i \in I\}.$$

this category is dual to the category of frames and frame morphisms. As far as the structures go, for each frame  $\mathfrak{F}$  the usual modal algebra  $\mathfrak{F}^+$  on  $\mathfrak{B}(W)$  is complete and atomic. For each complete atomic modal algebra  $\mathfrak{A}$  with set of atoms  $\text{At}(\mathfrak{A})$ , we take the frame  $\mathfrak{A}_+ = \langle \text{At}(\mathfrak{A}), R \rangle$  with

$$xRy \text{ iff } x \leq \mathbf{m}y, \text{ for each } x, y \in \text{At}(\mathfrak{A}).$$

For the morphisms, given frames  $\mathfrak{F} = \langle W, R \rangle, \mathfrak{F}' = \langle W', R' \rangle$  and a frame morphism  $\psi : \mathfrak{F} \rightarrow \mathfrak{F}'$ , define  $\psi^+ : \mathfrak{F}'^+ \rightarrow \mathfrak{F}^+$  by taking

$$\psi^+(S) = \psi^{-1}[S], \text{ for each } S \in \mathfrak{B}(W')$$

as before. In the other direction a new definition is needed. given complete atomic modal algebras  $\mathfrak{A}, \mathfrak{B}$  and a complete homomorphism  $\phi : \mathfrak{A} \rightarrow \mathfrak{B}$ , define  $\phi_+ : \mathfrak{B}_+ \rightarrow \mathfrak{A}_+$  by taking

$$\phi_+(y) = x \text{ iff } y \leq \phi(x), \text{ for each } x \in \text{At}(\mathfrak{A}), y \in \text{At}(\mathfrak{B}).$$

To see that this definition is valid, note that  $\{\phi(x) : x \in \text{At}(\mathfrak{A})\}$  is a disjoint cover of  $B$ , since  $\text{At}(\mathfrak{A})$  is a disjoint cover of  $A$  and  $\phi$  is a complete homomorphism. It can be checked that each frame  $\mathfrak{F}$  is 'isomorphic' to  $(\mathfrak{F}^+)_+$ , and that each complete atomic modal algebra  $\mathfrak{A}$  is isomorphic to  $(\mathfrak{A}_+)^+$ , so that these categories are contravariantly dual to each other.

## ACKNOWLEDGEMENTS

This chapter is the result of collaboration on the following terms. Segerberg wrote Section 1–9, Bull Sections 10–24. Although the authors met and together planned the paper, each wrote his part independently of the other will little *ex post script* discussion.

Seeger wishes to thank S. K. Thomason (who conveniently spent part of his sabbatical 1982 at the University of Auckland) for a number of very useful critical comments.

Robert Bull  
*University of Canterbury, New Zealand*

Krister Segerberg  
*University of Uppsala, Sweden*

#### BIBLIOGRAPHY

- [Ackerman, 1956] W. Ackerman. Begründung einer strengen Implikation. *Journal of Symbolic Logic*, **21**, 113–128, 1956.
- [Alban, 1943] M. J. Alban. Independence of the primitive symbols of Lewis' calculi of propositions. *Journal of Symbolic Logic*, **8**, 24–26, 1943.
- [Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap. *Entailment: The Logic of Relevance and Necessity*, Vol. 1, Princeton University Press, Princeton, 1975.
- [Anderson, 1980] C. A. Anderson. Some axioms for the logic of sense and denotation: alternative (0). *Nous*, **14**, 217–234, 1980.
- [Bayart, 1959] A. Bayart. Quasi-adéquation de la logique modale du second ordre S5 et adéquation de la logique du premier ordre S5. *Logique et analyse*, **2**, 99–121, 1959.
- [Becker, 1930] O. Becker. Zur Logik der Modalitäten. *Jarbuch für Philosophie und phänomenologische Forschung*, **11**, 496–548, 1930.
- [Belnap, 1981] N. D. Belnap. Modal and relevance logics: 1977. In *Modern Logic—A Survey*, E. Agazzi, ed. pp. 131–151. Reidel, Dordrecht, 1981.
- [Beth, 1959] E. W. Beth. *The Foundations of Mathematics: A Study in the Philosophy of Science*. North-Holland, Amsterdam, 1959.
- [Blok, 1980] W. J. Blok. The lattice of modal algebras: an algebraic investigation. *Journal of Symbolic Logic*, **45**, 221–236, 1980.
- [Block, 1980a] W. J. Blok. Pretabular varieties of modal algebras. *Studia Logica*, **39**, 101–124, 1980.
- [Boolos, 1979] G. Boolos. *The Unprovability of Consistency: An Essay in Modal Logic*. Cambridge University Press, Cambridge, 1979.
- [Bowen, 1978] K. A. Bowen. *Model Theory for Modal Logic*. Reidel, Dordrecht, 1978.
- [Bull, 1965] R. A. Bull. An algebraic study of Diodorean modal systems. *Journal of Symbolic Logic*, **30**, 58–64, 1965.
- [Bull, 1965a] R. A. Bull. A modal extension of intuitionistic logic. *Notre Dame Journal of Formal Logic*, **6**, 142–146, 1965.
- [Bull, 1966] R. A. Bull. That all normal extensions of **S4.3** have the finite model property. *Zeit Math. Logik Grund.*, **12**, 341–344, 1966.
- [Bull, 1966a] R. A. Bull. MIPC as the formalisation of an intuitionist concept of modality. *Journal of Symbolic Logic*, **31**, 609–616, 1966.
- [Bull, 1967] R. A. Bull. On the extension of **S4** with *CLMpMLp*. *Notre Dame Journal of Formal Logic*, **8**, 325–329, 1967.
- [Bull, 1969] R. A. Bull. On modal logic with propositional quantifiers. *Journal of Symbolic Logic*, **34**, 257–263, 1969.
- [Bull, 1982] R. A. Bull. Review. *Journal of Symbolic Logic*, **47**, 440–445, 1982.
- [Bull, 1983] R. A. Bull. Review. *Journal of Symbolic Logic*, **48**, 488–495, 1983.
- [Carnap, 1942] R. Carnap. *Introduction to Semantics*. Harvard University Press, Cambridge, MA, 1942.
- [Carnap, 1947] R. Carnap. *Meaning and Necessity: A Study in Semantics and Modal Logic*. The University of Chicago Press, Chicago, 1947.

- [Chellas, 1980] B. F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, Cambridge, 1980.
- [Church, 1946] A. Church. A formulation of the logic of sense and denotation. *Abstract. Journal of Symbolic Logic*, **11**, 31, 1946.
- [Church, 1951] A. Church. A formulation of the logic of sense and denotation. In *Structure, Method and Meaning. Essays in Honor of Henry M. Scheffer*, P. Henle *et al.*, eds. pp. 3–24. the Liberal Arts Press, NY, 1951.
- [Church, 1951a] A. Church. The weak theory of implication. In *Kontrolliertes Denken: Untersuchungen zum Logikkalkül unter der Einzelwissenschaften*, Menne *et al.*, eds. pp. 22–37. Kommissions-Verlag Karl Alber, Munich, 1951.
- [Church, 1973–74] A. Church. Outline of a revised formulation of the logic of sense and denotation. *Nous*, **7**, 24–33; **8**, 135–156, 1973–74.
- [Cresswell, 1967] M. Cresswell. A Henkin completeness theorem for *T*. *Notre Dame Journal of Formal Logic*, **8**, 186–190, 1967.
- [Curley, 1975] E. M. Curley. The development of Lewis' theory of strict implication. *Notre Dame Journal of Formal Logic*, **16**, 517–527, 1975.
- [Curry, 1950] H. B. Curry. *A Theory of Formal Deducibility*. University of Notre Dame Press, Notre Dame, IN, 1950.
- [Dugundj, 1940] J. Dugundj. Note on a property of matrices for Lewis and Langford's calculi of propositions. *Journal of Symbolic Logic*, **5**, 150–151, 1940.
- [Dummett and Lemmon, 1959] M. A. E. Dummett and E. J. Lemmon. Modal logics between **S4** and **S5**. *Zeit. Math. Logik. Grund.*, **3**, 250–264, 1959.
- [Esaia and Meskhi, 1977] L. Esakia and V. Meskhi. Five critical modal systems. *Theoria*, **43**, 52–60, 1977.
- [Feys, 1965] R. Feys. *Modal Logics*. Edited with some complements by Joseph Dopp, E. Nauwelaerts, Louvainand Gauthier-Vallars, Paris, 1965.
- [Fine, 1970] K. Fine. Propositional quantifiers in modal logic. *Theoria*, **36**, 336–346, 1970.
- [Fine, 1971] K. Fine. The logics containing **S4.3**. *Zeit. Math. Logik. Grund.*, **17**, 371–376, 1971.
- [Fine, 1972] K. Fine. Logics containing **S4** without the finite model property. In *Conference in Mathematical Logic*, London 1970, W. Hodges, ed. pp. 88–102. Vol. 255 *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 1972.
- [Fine, 1974] K. Fine. An incomplete logic containing **S4**. *Theoria*, **40**, 23–29, 1974.
- [Fine, 1974a] K. Fine. An ascending chain of **S4** logics. *Theoria*, **40**, 110–116, 1974.
- [Fine, 1974b] K. Fine. Logics containing **K4**, Part I. *Journal of Symbolic Logic*, **39**, 31–42, 1974.
- [Fine, 1975] K. Fine. Some connections between elementary and modal logic. In *Proceedings of the Third Scandinavian Logic Symposium*, S. Kanger, ed. pp. 15–31. North-Holland, Amsterdam, 1975.
- [Fine, 1975a] K. Fine. Normal formas in modal logic. *Notre Dame Journal of Formal Logic*, **16**, 229–234, 1975.
- [Fine, 1977] K. Fine. Prior on the construction of possible worlds and instants. In *Worlds, Times and Selves*, A. N. Prior and K. Fine, eds. pp. 116–161. Duckworth, London, 1977.
- [Fine, 1977a] K. Fine. Properties, propositions and sets. *Journal of Philosophical Logic*, **6**, 135–191, 1977.
- [Fine, 1978] K. Fine. Model theory for modal logic. *Journal of Philosophical Logic*, **7**, 125–156, 1978/81.
- [Fine, 1980] K. Fine. First order modal theories, I: Sets. *Nous*, **15**, 177–205; II: Propositions. *Studia Logica*, **34**, 159–202; III: Facts. *Synthese*, **53**, 43–122, 1980, 1981, 1982.
- [Fischer-Servi, 1977] G. Fischer-Servi. On modal logic with an intuitionist base. *Studia Logica*, **36**, 141–149, 1977.
- [Fischer-Servi, 1981] G. Fischer-Servi. Semantics for a class of intuitionist modal calculi. In *Italian Studies in the Philosophy of Science*, M. L. Dalla Chiara, ed. pp. 59–72. Reidel, Dordrecht, 1981.
- [Fitch, 1937] F. B. Fitch. Modal functions in two-valued logic. *Journal of Symbolic Logic*, **2**, 125–128, 1937.

- [Fitch, 1939] F. B. Fitch. Note on modal functions. *Journal of Symbolic Logic*, **4**, 115–116, 1939.
- [Fitch, 1948] F. B. Fitch. Intuitionistic modal logic with quantifiers. *Portugaliae Mathematica*, **7**, 113–118, 1948.
- [Fitch, 1952] F. B. Fitch. *Symbolic Logic: An Introduction*. Ronald Press, NY, 1952.
- [Follesdal, 1989] D. Føllesdal. Von Wright's modal logic. In *The Philosophy of Georg Henrik Von Wright*, P. A. Schilpp, ed. 1989.
- [Follesdal and Hilpinen, 1971] D. Føllesdal and R. Hilpinen. Denontic logic: An introduction. In *Deontic Logic: Introductory and Systematic Readings*, R. Hilpinen, ed. pp. 1–35. Reidel, Dordrecht, 1971.
- [Friedman, 1975] H. Friedman. One hundred and two problems in mathematical logic. *Journal of Symbolic Logic*, **40**, 113–129, 1975.
- [Gabbay, 1976] D. M. Gabbay. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. Reidel, Dordrecht, 1976.
- [Gabbay, 1981] D. M. Gabbay. *Semantical Investigations in Heyting's Intuitionistic Logic*. Reidel, Dordrecht, 1981.
- [Gerson, 1975] M. Gerson. The inadequacy of the neighbourhood semantics for modal logic. *Journal of Symbolic Logic*, **40**, 141–148, 1975.
- [Gerson, 1975a] M. Gerson. An extension of **S4** complete for the neighbourhood semantics but incomplete for the relational semantics. *Studia Logica*, **34**, 333–342, 1975.
- [Gerson, 1976] M. Gerson. A neighbourhood frame for *T* with no equivalent relational frame. *Zeit. Math. Logik. Grund.*, **22**, 29–34, 1976.
- [Gödel, 1933] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematisches Kolloquiums*, **4**, 39–40, 1933.
- [Goldblatt, 1975] R. I. Goldblatt. First-order definability in modal logic. *Journal of Symbolic Logic*, **40**, 35–40, 1975.
- [Goldblatt, 1976] R. I. Goldblatt. Methamathematics of modal logic. *Reports on Mathematical Logic*, **6**, 41–78; **7**, 21–52, 1976.
- [Goldblatt and Thomason, 1975] R. I. Goldblatt and S. K. Thomason. Axiomatic classes in propositional modal logic. In *Algebra and Logic*, J. N. Crossley, ed. pp. 163–173. Vo. 450 of *Lecture Notes in Mathematics*, Springer-Verlag, Berlin, 1975.
- [Grzegoczkyk, 1981] A. Grzegoczkyk. Individualistic formal approach to deontic logic. *Studia Logica*, **40**, 99–102, 1981.
- [Guillaume, 1958] M. Guillaume. Rapports entre calculs propositionnels modaux et topologie impliqués par certaines extensions de la méthode e tableaux sémantiques. *Comptes rendus hebdomadaires des séances de l'Académie des Sciences*, **246**, 1140–1142, 2207–2210; **247**, 1281–1283, Gauthiers- Villars, Paris, 1958.
- [Halldén, 1949] S. Halldén. Results concerning the decision problem of Lewis's calculi **S3** and **S4**. *Journal of Symbolic Logic*, **15**, 230–236, 1949.
- [Hansson and Gärdenfors, 1973] B. Hansson and P. Gärdenfors. A guide to intensional semantics. In *Modality, Morality and other Problems of Sense and Nonsense. Essays Dedicated to Sören Halldén*, pp. 151–167. Gleerup, Lund, 1973.
- [Hilpinen, 1971] R. Hilpinen. *Deontic Logic: Introductory and Systematic Readings*. Reidel, Dordrecht, 1971.
- [Hintikka, 1955] J. Hintikka. Form and content in quantification theory. *Acta Philosophica Fennica*, **8**, 11–55, 1955.
- [Hintikka, 1957] J. Hintikka. *Quantifiers in Deontic Logic*. Societas Scientarum Fennica, Commentationes humanarum litterarum **23:4**, Helsingfors, 1957.
- [Hintikka, 1961] J. Hintikka. Modality and quantification. *Theoria*, **27**, 119–128, 1961. Revised veesion reprinted in Hinktikka [1969].
- [Hintikka, 1962] J. Hintikka. *Knowledge and Belief: An Introduction to the Logic of the Two Notions*. Cornell University Press, Ithaca, NY, 1962.
- [Hintikka, 1963] J. Hintikka. The modes of modality. *Acta Philosophica Fennica*, **16**, 65–82, 1963. Reprinted in Hintikka [1969].
- [Hintikka, 1969] J. Hinktikka. *Models for Modalities: Selected Essays*. Reidel, Dordrecht, 1969.
- [Hintikka, 1969a] J. Hinktikka. Review. *Journal of Symbolic Logic*, **34**, 305–306, 1969.

- [Hintikka, 1975] J. Hintikka. Carnap's heritage in logical semantics. In *Rudolf Carnap, Logical Empiricist: Materials and Perspectives*, J. Hintikka, ed. pp. 217–242. Reidel, Dordrecht, 1975.
- [Hofstadter and McKinsey, 1955] A. Hofstadter and J. C. C. McKinsey. On the logic of imperatives. *Philosophy of Sciences*, **6**, 446–457, 1939.
- [Hughes and Cresswell, 1968] G. E. Hughes and M. J. Cresswell. *A New Introduction to Modal Logic*. Routledge, 1968.
- [Jeffrey, 1990] R. C. Jeffrey. *Formal Logic: Its Scope and Limits*, 3rd Edition. McGraw-Hill, NY, (1st edition 1967), 1990.
- [Jónsson, 1967] B. Jónsson. Algebras whose congruence lattices are distributive. *Mathematica Scandinavica*, **21**, 110–121, 1967.
- [Jónsson and Tarski, 1951] E. Jónsson and A. Tarski. Boolean algebras with operators. Part I. *Am. J. Math.*, **74**, 891–939, 1951.
- [Kamp, 1968] J. A. W. Kamp. *On Tense Logic and the Theory of Order*. PhD Dissertation, UCLA, 1968.
- [Kanger, 1957] S. Kanger. *Provability in Logic*. Dissertation, Stockholm, 1957.
- [Kanger, 1957a] S. Kanger. *New Foundations for Ethical Theory*, Stockholm, 1957. Reprinted in Hilpinen [1971].
- [Kanger, 1957b] S. Kanger. The Morning Star Paradox. *Theoria*, **23**, 1–11, 1957.
- [Kanger, 1957c] S. Kanger. A note on quantification and modalities. *Theoria*, **23**, 131–134, 1957.
- [Kaplan, 1966] D. Kaplan. Review. *Journal of Symbolic Logic*, **31**, 120–122, 1966.
- [Kaplan, 1970] D. Kaplan.  $S_5$  with quantifiable propositional variables, Abstract. *Journal of Symbolic Logic*, **35**, 355, 1970.
- [Kneale and Kneale, 1962] W. Kneale and M. Kneale. *The Development of Logic*. Clarendon Press, Oxford, 1962.
- [Kripke, 1959] S. A. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, **24**, 1–14, 1959.
- [Kripke, 1963] S. A. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, **16**, 83–94, 1963.
- [Kripke, 1963a] S. A. Kripke. Semantical analysis of modal logic I: Normal propositional calculi. *Zeit. Math. Logik. Grund.*, **9**, 67–96, 1963.
- [Kripke, 1965] S. A. Kripke. Semantical analysis of modal logic II: Non-normal modal propositional calculi. In *The Theory of Models*, J. W. Adison *et al.*, eds. pp. 206–220. North-Holland, Amsterdam, 1965.
- [Kuhn, 1977] S. T. Kuhn. *Many-sorted Modal Logics*. Philosophical studies published by the Philosophical Society and the Department of Philosophy, University of Uppsala, Vol. 35, Uppsala, 1977.
- [Leivant, 1981] D. Leivant. On the proof theory of the modal logic for arithmetic provability. *Journal of Symbolic Logic*, **46**, 531–538, 1981.
- [Lemmon, 1957] E. J. Lemmon. New foundations for Lewis modal systems. *Journal of Symbolic Logic*, **22**, 176–186, 1957.
- [Lemmon, 1966] E. J. Lemmon. Algebraic semantics for modal logics. *Journal of Symbolic Logic*, **31**, 46–65, 191–218, 1966.
- [Lemmon, 1977] E. J. Lemmon. *An Introduction to Modal Logic*. In collaboration with D. Scott, Blackwell, Oxford, 1977.
- [Lewis, 1912] C. I. Lewis. Implication and the algebra of logic. *Mind*, **21**, 522–531, 1912.
- [Lewis, 1918] C. I. Lewis. *A Survey of Symbolic Logic*. University of California Press, Berkeley, 1918.
- [Lewis and Langford, 1959] C. I. Lewis and C. H. Langford. *Symbolic Logic*. The Century Co, NY, 1932. Second edn, Dover, NY, 1959.
- [Lewis, 1973] D. Lewis. *Counterfactuals*. Harvard University Press, Cambridge, MA, 1973.
- [Lukasiewicz, 1953] J. Lukasiewicz. A system of modal logic. *Journal of Computing Systems*, **1**, 111–149, 1953.
- [Lukasiewicz, 1970] J. Lukasiewicz. *Selected Works*, L. Borkowski, ed. North Holland, Amsterdam, 1970.
- [McCall, 1967] S. McCall. *Polish Logic 1920–1939*, Clarendon Press, Oxford, 1967.

- [McKinsey, 1941] J. C. C. McKinsey. A solution of the decision problem for the Lewis systems **S2** and **S4** with an application to topology. *Journal of Symbolic Logic*, **6**, 117–134, 1941.
- [McKinsey, 1945] J. C. C. McKinsey. On the syntactical construction of modal logic. *Journal of Symbolic Logic*, **10**, 83–96, 1945.
- [McKinsey and Tarski, 1944] J. C. C. McKinsey and A. Tarski. the algebra of topology. *Annals of Mathematics*, **45**, 141–191, 1944.
- [McKinsey and Tarski, 1948] J. C. C. McKinsey and A. Tarski. Some theormes about the sentential calculi of Lewis and Heyting. *Journal of Symbolic Logic*, **13**, 1–15, 1948.
- [Makinson, 1966] D. Makinson. On some completeness theorems in modal logic. *Zeit. Math. Logik. Grund.*, **12**, 379–384, 1966.
- [Makinson, 1969] D. Makinson. A normal modal calculus between **T** and **S4** without the finite modal property. *Journal of Symbolic Logic*, **34**, 35–38, 1969.
- [Makinson, 1970] D. Makinson. A generalisation of the concept of a relational model for modal logic. *Theoria*, **36**, 331–335, 1970.
- [Makinson, 1971] D. Makinson. *Aspectos de la logica mdoal*, Instituto e matematica. Universidad Nacional del Sur, Bahia Blanca, 1971.
- [Makinson, 1971a] D. Makinson. Some embedding theorems for modal logic. *Notre Dame Journal of Formal Logic*, **12**, 252–254, 1971.
- [Maksimova, 1975] L. L. Maksimova. Pretabular extensions of Lewis' **S4**. *Algebra i logika*, **14**, 28–55, 1975. (In Russian)
- [Malinowski, 1977] G. Malinowski. Historical note. In *selected Papers on Lukasiewicz Sentential Calculi*, R. Wójcicki, ed. pp. 177–187. Polish Academy of Sciences, Wrocław, 1977.
- [Mally, 1926] E. Mally. *Grundgesetze des Sollens: Elemente der Logik des Willens*. Lenscher and Lugensky, Graz, 1926.
- [Montague, 1963] R. Montague. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatisability. *Acta Philosophica Fennica*, **16**, 153–167, 1963. Reprinted in Montague [1974].
- [Montague, 1968] R. Montague. Pragmatics. In *Contemporary Philosophy: A Survey*, Vol. 1. R. Klībasky, ed. pp. 102–122. La Nuova Editrice, Florence, 1968. Reprinted in Montague [1974].
- [Montague, 1974] R. Montague. *Formal Philosophy: Selected Papers of Richard Montague*. Edited, with an introduction by Richmond H. Thomason. Yale University Press, New Haven, 1974.
- [Morgan, 1979] C. Morgan. Modality, analogy, and ideal experiments according to C. S. Pierce. *Synthese*, **41**, 65–83, 1979.
- [Mortimer, 1974] M. Mortimer. Some results in modal model theory. *Journal of Symbolic Logic*, **39**, 496–508, 1974.
- [Ohnishi and Matsumoto, 1957/59] M. Ohnishi and K. Matsumoto. Gentzen method in modal calculi. *Osaka Mathematical Journal*, **9**, 113–130; **11**, 115–120, 1957/1959.
- [Parry, 1934] W. T. Parry. The postulates for 'strict implication'. *Mind*, **43**, 78–80, 1934.
- [Parsons, 1982] C. Parsons. Intensional logic in extensional language. *Journal of Symbolic Logic*, **47**, 289–328, 1982.
- [Pratt, 1980] V. R. Pratt. Application of modal logic to programming. *Studia Logica*, **34**, 257–274, 1980.
- [Prawitz, 1965] D. Prawitz. *Natural Deduction: A Proof-theoretic study*, Stockholm Studies in Philocopy 3, Almqvist and Wiskell, Stockholm, 1965.
- [Prior, 1962] A. N. Prior. *Formal Logic*. Clarendon Press, Oxford, 1955. Second Edition, 1962.
- [Prior, 1957] A. N. Prior. *Time and Modality*. Clarendon Press, Oxford, 1957.
- [Prior, 1967] A. N. Prior. *Past, Present and Future*. Clarendon Press, Oxford, 1967.
- [Rasiowa and Sikorski, 1963] H. Rasiowa and R. Sikorski. *The Mathematics of Meta-mathematics*, Państwowe Wydawnictwo Naukowe, 1963.
- [Rautenberg, 1979] W. Rautenberg. *klassische und nichtklassische Aussagenlogik*, Bieweg, Braunschweig, Wiesbaden, 1979.
- [Rescher and Urquhart, 1971] N. Rescher and A. Urquhart. *Temporal Logic*. Springer-Verlag, NY, 1971.



- [Ridder, 1955] J. Ridder. Die Grntzensschen Schlussverfahren in modalen Aussagenlogiken I. *Indagationes mathematicae*, **17**, 163–276, 1955.
- [Sahlqvist, 1975] H. Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logic. In *Proceedings of the Third Scandinavian Logic Symposium*, S. Kanger, ed. pp. 110–143. North-Holland, Amsterdam, 1975.
- [Schumm, 1981] G. F. Schumm. Bounded properties in modal logic. *Zeit. Math. Logik. Grund.*, **27**, 197–200, 1981.
- [Schütte, 1968] K. Schütte. *Vollständige Systeme modaler und intuitionistischer Logik*. Springer-Verlag, Berlin, 1968.
- [Scott, 1971] D. Scott. On engendering an illusion of understanding. *Journal of Philosophy*, **68**, 787–807, 1971.
- [Scroggs, 1951] S. J. Scroggs. Extensions of the Lewis system **S5**. *Journal of Symbolic Logic*, **16**, 112–120, 1951.
- [Segeberg, 1968] K. Segerberg. Decidability of **S4.2**. *Theoria*, **34**, 7–20, 1968.
- [Segeberg, 1970] K. Segerberg. Modal logics with linear alternative relations. *Theoria*, **36**, 301–322, 1970.
- [Segeberg, 1971] K. Segerberg. *An Essay in Classical Modal Logic*. Philosophical studies published by the Philosophical society and the Department of Philosophy, University of Uppsala, Vol. 13, Uppsala, 1971.
- [Segeberg, 1982] K. Segerberg. *Classical Propositional Operators: An Exercise in the Foundations of Logic*, Clarendon Press, Oxford, 1982.
- [Segeberg, 1989] K. Segerberg. Von Wright's tense-logic. In *The Philosophy of Georg Henrik von Wright*, P. A. Schlipp, ed. 1989.
- [Shoesmith and Smiley, 1978] D. J. Shoesmith and T. J. Smiley. *Multiple-conclusion Logic*. Cambridge University Press, Cambridge, 1978.
- [Smullyan, 1968] R. M. Smullyan. *First-order Logic*. Springer-Verlag, NY, 1968.
- [Snyder, 1971] D. P. Snyder. *Modal Logic and its Applications*. Van Nostrand Reinhold, NY, 1971.
- [Sobinciński, 1964] B. Sobinciński. Family *K* of the non-Lewis modal systems. *Notre Dame Journal of Formal Logic*, **5**, 313–318, 1964.
- [Solovay, 1976] R. S. M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, **25**, 287–304, 1976.
- [Stalnaker, 1968] R. Stalnaker. A theory of conditionals. In *Studies in Logical Theory*, N. Rescher, ed. p. 98–112. Blackwell, Oxford, 1968.
- [Thomason, 1972] S. K. Thomason. Semantic analysis of tense logics. *Journal of Symbolic Logic*, **37**, 150–158, 1972.
- [Thomason, 1972a] S. K. Thomason. Noncompactness in propositional modal logic. *Journal of Symbolic Logic*, **37**, 716–720, 1972.
- [Thomason, 1974] S. K. Thomason. An incompleteness theorem in modal logic. *Theoria*, **40**, 30–34, 1974.
- [Thomason, 1975] S. K. Thomason. Categories of frames for modal logic. *Journal of Symbolic Logic*, **40**, 439–442, 1975.
- [van Benthem, 1975] J. F. A. K. van Benthem. A note on modal formulae and relational properties. *Journal of Symbolic Logic*, **40**, 55–58, 1975.
- [van Benthem, 1976] J. F. A. K. van Benthem. Modal formulas are either elementary or not  $\Sigma\Delta$ -elementary. *Journal of Symbolic Logic*, **41**, 436–438, 1976.
- [van Benthem, 1978] J. F. A. K. van Benthem. Two simple incomplete modal logics. *Theoria*, **44**, 25–37, 1978.
- [van Benthem, 1979] J. F. A. K. van Benthem. Canonical modal logics and ultrafilter extensions. *Journal of Symbolic Logic*, **44**, 1–8, 1979.
- [van Benthem, 1979a] J. F. A. K. van Benthem. Syntactic aspects of modal incompleteness theorems. *Theoria*, **45**, 67–81, 1979.
- [van Benthem and Blok, 1978] J. F. A. K. van Benthem and W. Blok. Transitivity follows from Dummett's axiom. *Theoria*, **44**, 117–118, 1978.
- [von Wright, 1951] G. H. von Wright. *An Essay in Modal Logic*. North Holland, Amsterdam, 1951.
- [von Wright, 1951a] G. H. von Wright. Deontic logic. *Mind*, **60**, 1–15, 1951.

- [von Wright, 1968] G. H. von Wright. An essay in deontic logic and general theory of action with a bibliography of deontic and imperative logic. *Acta Philosophical Fennica*, **21**, 1968.
- [von Wright, 1981] G. H. von Wright. Problems and prospects of deontic logic. A Survey. In *Modern Logic—A Survey*, ed. Evandro Agazzi, ed. pp. 199–423. Reidel, Dordrecht, 1981.
- [Zeman, 1973] J. J. Zeman. *Modal Logic: The Lewis-Modal Systems*. Clarendon Press, Oxford, 1973.



## ADVANCED MODAL LOGIC

This chapter is a continuation of the preceding one, and we begin it at the place where the authors of *Basic Modal Logic* left us about fifteen years ago. Concluding his historical overview, Krister Segerberg wrote: “Where we stand today is difficult to say. Is the picture beginning to break up, or is it just the contemporary observer’s perennial problem of putting his own time into perspective?” So, where did modal logic of the 1970s stand? Where does it stand now? Modal logicians working in philosophy, computer science, artificial intelligence, linguistics or some other fields would probably give different answers to these questions. Our interpretation of the history of modal logic and view on its future is based upon understanding it as part of mathematical logic.

Modal logicians of the First Wave constructed and studied modal systems trying to formalize a few kinds of necessity-like and possibility-like operators. The industrialization of the Second Wave began with the discovery of a deep connection between modal logics on the one hand and relational and algebraic structures on the other, which opened the door for creating many new systems of both artificial and natural origin. Other disciplines—the foundations of mathematics, computer science, artificial intelligence, etc.—brought (or rediscovered<sup>1</sup>) more. “This framework has had enormous influence, not only just on the logic of necessity and possibility, but in other areas as well. In particular, the ideas in this approach have been applied to develop formalisms for describing many other kinds of structures and processes in computer science, giving the subject applications that would have probably surprised the subject’s founders and early detractors alike” [Barwise and Moss 1996]. Even two or three mathematical objects may lead to useful generalizations. It is no wonder then that this huge family of logics gave rise to an abstract notion (or rather notions) of a modal logic, which in turn put forward the problem of developing a general theory for it.

Big classes of modal systems were considered already in the 1950s, say extensions of **S5** [Scroggs 1951] or **S4** [Dummett and Lemmon 1959]. Completeness theorems of Lemmon and Scott [1977],<sup>2</sup> Bull [1966b] and Segerberg [1971] demonstrated that many logics, formerly investigated “piecewise”, have in fact very much in common and can be treated by the same methods. A need for a uniting theory became obvious. “There are two main lacunae in recent work on modal logic: a lack of general results and a lack of negative results. This or that logic is shown to have such and such a property, but very little is known about the scope or bounds of the property.

---

<sup>1</sup>One of the celebrities in modal logic—the Gödel–Löb provability logic **GL**—was first introduced by Segerberg [1971] as an “artificial” system under the name **K4W**.

<sup>2</sup>This book was written in 1966.

Thus there are numerous results on completeness, decidability, finite model property, compactness, etc., but very few general or negative results”, wrote Fine [1974c]. The creation of duality theory between relational and algebraic semantics ([Lemmon 1966a,b], [Goldblatt 1976a,b]), originated actually by Jónsson and Tarski [1951], the establishment of the connection between modal logics and varieties of modal algebras ([Kuznetsov 1971], Maksimova and Rybakov [1974], [Blok 1976]), and between modal and first and higher order languages ([Fine 1975b], [van Benthem 1983]) added those mathematical ingredients that were necessary to distinguish modal logic as a separate branch of mathematical logic.

On the other hand, various particular systems became subjects of more special disciplines, like provability logic, deontic logic, tense logic, etc., which has found reflection in the corresponding chapters of this Handbook.

In the 1980s and 1990s modal logic was developing both “in width” and “in depth”, which made it more difficult for us to select material for this chapter. The expansion “in width” has brought in sight new interesting types of modal operators, thus demonstrating again the great expressive power of propositional modal languages. They include, for instance, polyadic operators, graded modalities, the fixed point and difference operators. We hope the corresponding systems will be considered in detail elsewhere in the *Handbook*; in this chapter they are briefly discussed in the appendix, where the reader can find enough references.

Instead of trying to cover the whole variety of existing types of modal operators, we decided to restrict attention mainly to the classes of normal (and quasi-normal) uni- and polymodal logics and follow “in depth” the way taken by Bull and Segerberg in *Basic Modal Logic*, the more so that this corresponds to our own scientific interests.

Having gone over from considering individual modal systems to big classes of them, we are certainly interested in developing general methods suitable for handling modal logics *en masse*. This somewhat changes the standard set of tools for dealing with logics and gives rise to new directions of research. First, we are almost completely deprived of proof-theoretic methods like Gentzen-style systems or natural deduction. Although proof theory has been developed for a number of important modal logics, it can hardly be extended to reasonably representative families. (Proof theory is discussed in the chapter *Sequent systems for modal logics* in a later volume of this *Handbook*; some references to recent results can be found in the appendix.)

In fact, modern modal logic is primarily based upon the frame-theoretic and algebraic approaches. The link connecting syntactical representations of logics and their semantics is general completeness theory which stems from the pioneering results of Bull [1966b], Fine [1974c], Sahlqvist [1975], Goldblatt and Thomason [1974]. Completeness theorems are usually the first step in understanding various properties of logics, especially those that have semantic or algebraic equivalents. A classical example is Maksimova’s

[1979] investigation of the interpolation property of normal modal logics containing **S4**, or decidability results based on completeness with respect to “good” classes of frames. Completeness theory provides means for axiomatizing logics determined by given frame classes and characterizes those of them that are modal axiomatic.

Standard families of modal logics are endowed with the lattice structure induced by the set-theoretic inclusion. This gives rise to another line of studies in modal logic, addressing questions like “what are co-atoms in the lattice?” (i.e., what are maximal consistent logics in the family?), “are there infinite ascending chains?” (i.e., are all logics in the family finitely axiomatizable?), etc. From the algebraic standpoint a lattice of logics corresponds to a lattice of subvarieties of some fixed variety of modal algebras, which opens a way for a fruitful interface with a well-developed field in universal algebra.

A striking connection between “geometrical” properties of modal formulas, completeness, axiomatizability and  $\bigwedge$ -prime elements in the lattice of modal logics was discovered by Jankov [1963, 1969], Blok [1978, 1980b] and Rautenberg [1979]. These observations gave an impetus to a project of constructing frame-theoretic languages which are able to characterize the “geometry” and “topology” of frames for modal logics ([Zakharyashev 1984, 1992], [Wolter 1996c]) and thereby provide new tools for proving their properties and clarifying the structure of their lattices.

One more interesting direction of studies, arising only when we deal with big classes of logics, concerns the algorithmic problem of recognizing properties of (finitely axiomatizable) logics. Having undecidable finitely axiomatizable logics in a given class [Thomason 1975a; Shehtman 1978c], it is tempting to conjecture that non-trivial properties of logics in this class are undecidable. However, unlike Rice’s Theorem in recursion theory, some important properties turn out to be decidable, witness the decidability of interpolation above **S4** [Maksimova 1979]. The machinery for proving the undecidability of various properties (e.g. Kripke completeness and decidability) was developed in [Thomason 1982] and [Chagrova 1990b,c].

Thomason [1982] proved the undecidability of Kripke completeness first in the class of polymodal logics and then transferred it to that of unimodal ones. In fact, Thomason’s embedding turns out to be an isomorphism from the lattice of logics with  $n$  necessity operators onto an interval in the lattice of unimodal logics, preserving many standard properties [Kracht and Wolter 1999]. Such embeddings are interesting not only from the theoretical point of view but can also serve as a vehicle for reducing the study of one class of logics to another. Perhaps the best known example of such a reduction is the Gödel translation of intuitionistic logic and its extensions into normal modal logics above **S4** [Maksimova and Rybakov 1974; Blok 1976; Esakia 1979a,b]. We will take advantage of this translation to give a brief survey of results in the field of superintuitionistic logics which actually were always

studied in parallel with modal logics (see also Section 5 of *Intuitionistic Logic* in volume 7 of this *Handbook*).

Listed above are the most important general directions in mathematical modal logic we are going to concentrate on in this chapter. They, of course, do not cover the whole discipline. Other topics, for instance, modal systems with quantifiers, the relationship between the propositional modal language and the first (or higher) order classical language, or proof theory are considered in other chapters of this *Handbook*.

It should be emphasized once again that the reader will find no discussions of particular modal systems in this chapter. Modal logic is presented here as a mathematical theory analyzing big families of logics and thereby providing us with powerful methods for handling concrete ones. (In some cases we illustrate technically complex methods by considering concrete logics; for instance Rybakov's [1994] technique of proving the decidability of the admissibility problem for inference rules is explained only for **GL**.)

## 1 UNIMODAL LOGICS

We begin by considering normal modal logics with one necessity operator, which were introduced in Section 6 of *Basic Modal Logic*. Recall that each such logic is a set of modal formulas (in the language with the primitive connectives  $\wedge, \vee, \rightarrow, \perp, \Box$ ) containing all classical tautologies, the modal axiom

$$\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q),$$

and closed under substitution, modus ponens and necessitation  $\varphi/\Box\varphi$ .

### 1.1 The lattice $\mathbf{NExtK}$

First let us have a look at the class of normal modal logics from a purely syntactic point of view. Given a normal modal logic  $L_0$ , we denote by  $\mathbf{NExt}L_0$  the family of its **n**ormal **e**xtensions.  $\mathbf{NExtK}$  is thus the class of all normal modal logics. Each logic  $L$  in  $\mathbf{NExt}L_0$  can be obtained by adding to  $L_0$  a set of modal formulas  $\Gamma$  and taking the closure under the inference rules mentioned above; in symbols this is denoted by

$$L = L_0 \oplus \Gamma.$$

Formulas in  $\Gamma$  are called *additional* (or *extra*) *axioms of  $L$  over  $L_0$* . Formulas  $\varphi$  and  $\psi$  are said to be *deductively equivalent* in  $\mathbf{NExt}L_0$  if  $L_0 \oplus \varphi = L_0 \oplus \psi$ . For instance,  $\Box p \rightarrow p$  and  $p \rightarrow \Diamond p$  are deductively equivalent in  $\mathbf{NExtK}$ , both axiomatizing **T**, however  $(\Box p \rightarrow p) \leftrightarrow (p \rightarrow \Diamond p) \notin \mathbf{K}$ . (For more information on the relation between these formulas see [Chellas and Segerberg 1994] and [Williamson 1994].)

We distinguish between two kinds of derivations from assumptions in a logic  $L \in \text{NExt}\mathbf{K}$ . For a formula  $\varphi$  and a set of formulas  $\Gamma$ , we write  $\Gamma \vdash_L \varphi$  if there is a derivation of  $\varphi$  from formulas in  $L$  and  $\Gamma$  with the help of only modus ponens. In this case the standard deduction theorem— $\Gamma, \psi \vdash_L \varphi$  iff  $\Gamma \vdash_L \psi \rightarrow \varphi$ —holds. The fact of derivability of  $\varphi$  from  $\Gamma$  in  $L$  using both modus ponens and necessitation is denoted by  $\Gamma \vdash_L^* \varphi$ ; in such a case we say that  $\varphi$  is *globally derivable*<sup>3</sup> from  $\Gamma$  in  $L$ . For this kind of derivation we have the following variant of the deduction theorem which is proved by induction on the length of derivations in the same manner as for classical logic.

**THEOREM 1 (Deduction).** *For every logic  $L \in \text{NExt}\mathbf{K}$ , all formulas  $\varphi$  and  $\psi$ , and all sets of formulas  $\Gamma$ ,*

$$\Gamma, \psi \vdash_L^* \varphi \text{ iff } \exists m \geq 0 \Gamma \vdash_L^* \Box^{\leq m} \psi \rightarrow \varphi,$$

where  $\Box^{\leq m} \psi = \Box^0 \psi \wedge \dots \wedge \Box^m \psi$  and  $\Box^n \psi$  is  $\psi$  prefixed by  $n$  boxes.

It is to be noted that in general no upper bound for  $m$  can be computed even for a decidable  $L$  (see Theorem 194). However, if the formula

$$\mathbf{tra}_n = \Box^{\leq n} p \rightarrow \Box^{n+1} p$$

is in  $L$ —such an  $L$  is called  *$n$ -transitive*—then we can clearly take  $m = n$ . In particular, for every  $L \in \text{NExt}\mathbf{K4}$ ,  $\Gamma, \psi \vdash_L^* \varphi$  iff  $\Gamma \vdash_L^* \Box^+ \psi \rightarrow \varphi$ , where  $\Box^+ \psi = \psi \wedge \Box \psi$ . Moreover, a sort of conversion of this observation holds.

**THEOREM 2.** *The following conditions are equivalent for every logic  $L$  in  $\text{NExt}\mathbf{K}$ :*

- (i)  $L$  is  $n$ -transitive, for some  $n < \omega$ ;
- (ii) there exists a formula  $\chi(p, q)$  such that, for any  $\varphi, \psi$  and  $\Gamma$ ,

$$\Gamma, \psi \vdash_L^* \varphi \text{ iff } \Gamma \vdash_L^* \chi(\psi, \varphi).$$

**Proof.** The implication (i)  $\Rightarrow$  (ii) is clear. To prove the converse, observe first that  $\chi(p, q) \vdash_L^* \chi(p, q)$  and so  $\chi(p, q), p \vdash_L^* q$ . By Theorem 1, we then have  $\chi(p, q) \vdash_L^* \Box^{\leq n} p \rightarrow q$ , for some  $n$ . Let  $q = \Box^{n+1} p$ . Then

$$\chi(p, \Box^{n+1} p) \vdash_L^* \Box^{\leq n} p \rightarrow \Box^{n+1} p.$$

And since  $p \vdash_L^* \Box^{n+1} p$ ,  $\chi(p, \Box^{n+1} p) \in L$ . Consequently,  $\mathbf{tra}_n \in L$ . ■

**REMARK.** Note also that (i) is equivalent to the algebraic condition: the variety of modal algebras for  $L$  has equationally definable principal congruences. For more information on this and close results consult [Blok and Pigozzi 1982].

---

<sup>3</sup>This name is motivated by the semantical characterization of  $\vdash_L^*$  to be given in Theorem 19.



The *sum*  $L_1 \oplus L_2$  and *intersection*  $L_1 \cap L_2$  of logics  $L_1, L_2 \in \text{NExt}L_0$  are clearly logics in  $\text{NExt}L_0$  as well. The former can be axiomatized simply by joining the axioms of  $L_1$  and  $L_2$ . To axiomatize the latter we require the following definition. Given two formulas  $\varphi(p_1, \dots, p_n)$  and  $\psi(p_1, \dots, p_m)$  (whose variables are in the lists  $p_1, \dots, p_n$  and  $p_1, \dots, p_m$ , respectively), denote by  $\varphi \underline{\vee} \psi$  the formula  $\varphi(p_1, \dots, p_n) \vee \psi(p_{n+1}, \dots, p_{n+m})$ .

**THEOREM 3.** *Let  $L_1 = L_0 \oplus \{\varphi_i : i \in I\}$  and  $L_2 = L_0 \oplus \{\psi_j : j \in J\}$ . Then*

$$L_1 \cap L_2 = L_0 \oplus \{\Box^m \varphi_i \underline{\vee} \Box^n \psi_j : i \in I, j \in J, m, n \geq 0\}.$$

**Proof.** Denote by  $L$  the logic in the right-hand side of the equality to be established and suppose that  $\chi \in L_1 \cap L_2$ . Then for some  $m, n \geq 0$  and some finite  $I'$  and  $J'$  such that all  $\varphi'_i$  and  $\psi'_j$ , for  $i \in I', j \in J'$ , are substitution instances of some  $\varphi_{i'}$  and  $\psi_{j'}$ , for  $i' \in I, j' \in J$ , we have

$$\Box^{\leq m} \bigwedge_{i \in I'} \varphi'_i \rightarrow \chi \in L_0, \quad \Box^{\leq n} \bigwedge_{j \in J'} \psi'_j \rightarrow \chi \in L_0,$$

from which

$$\bigwedge_{\substack{i \in I', j \in J' \\ 0 \leq k, l \leq m+n}} (\Box^k \varphi'_i \vee \Box^l \psi'_j) \rightarrow \chi \in L_0$$

and so  $\chi \in L$  because  $\Box^k \varphi'_i \vee \Box^l \psi'_j$  is a substitution instance of  $\Box^k \varphi_{i'} \underline{\vee} \Box^l \psi_{j'}$ . Thus,  $L_1 \cap L_2 \subseteq L$ . The converse inclusion is obvious.  $\blacksquare$

Although the sum of logics differs in general from their union, these two operations have a few common important properties.

**THEOREM 4.** *The operation  $\oplus$  is idempotent, commutative, associative and distributes over  $\cap$ ; the operation  $\cap$  distributes over (infinite) sums, i.e.,*

$$L \cap \bigoplus_{i \in I} L_i = \bigoplus_{i \in I} (L \cap L_i).$$

It follows that  $\langle \text{NExt}L_0, \oplus, \cap \rangle$  is a complete distributive lattice, with  $L_0$  and the inconsistent logic, i.e., the set **For** of all modal formulas, being its zero and unit elements, respectively, and the set-theoretic  $\subseteq$  its corresponding lattice order. Note, however, that  $\oplus$  does not in general distribute over infinite intersections of logics. For otherwise we would have

$$(\mathbf{K} \oplus \neg \Box \perp) \oplus \bigcap_{1 \leq n < \omega} (\mathbf{K} \oplus \Box^n \perp) = \bigcap_{1 \leq n < \omega} (\mathbf{K} \oplus \neg \Box \perp \oplus \Box^n \perp),$$

which is a contradiction, since the logic in the left-hand side is consistent (**D**, to be more precise), while that in the right-hand side is not.

If we are interested in finding a simple (in one sense or another) syntactic representation of a logic  $L \in \text{NExt}L_0$ , we can distinguish *finite*, *recursive* and *independent axiomatizations of  $L$  over  $L_0$* . The former two notions mean that  $L = L_0 \oplus \Gamma$ , for some finite or, respectively, recursive  $\Gamma$ , and a set of axioms  $\Gamma$  is independent over  $L_0$  if  $L \neq L_0 \oplus \Delta$  for any proper subset  $\Delta$  of  $\Gamma$ . In the case when  $L_0$  is  $\mathbf{K}$  or any other finitely axiomatizable over  $\mathbf{K}$  logic, we may omit mentioning  $L_0$  and say simply that  $L$  is finitely (recursively, independently) axiomatizable.

It is fairly easy to see that  $L$  is not finitely axiomatizable over  $L_0$  iff there is an infinite sequence of logics  $L_1 \subset L_2 \subset \dots$  in  $\text{NExt}L_0$  such that  $L = \bigoplus_{i>0} L_i$ . This observation is known as *Tarski's criterion*. (It is worth noting that finite axiomatizability is not preserved under  $\cap$ . For example, using Tarski's criterion, one can show that  $\mathbf{D} \cap (\mathbf{K} \oplus \Box p \vee \Box \neg p)$  is not finitely axiomatizable.) The recursive axiomatizability of a logic  $L$ , as was observed by Craig [1953], is equivalent to the recursive enumerability of  $L$ . As for independent axiomatizability, an interesting necessary condition can be derived from [Kleyman 1984]. Suppose a normal modal logic  $L_1$  has an independent axiomatization. Then, for every finitely axiomatizable normal modal logic  $L_2 \subset L_1$ , the interval of logics

$$[L_2, L_1] = \{L \in \text{NExt}\mathbf{K} : L_2 \subseteq L \subseteq L_1\}$$

contains an immediate predecessor of  $L_1$ . Using this condition Chagrov and Zakharyashev [1995a] constructed various logics in  $\text{NExt}\mathbf{K4}$ ,  $\text{NExt}\mathbf{S4}$  and  $\text{NExt}\mathbf{Grz}$  without independent axiomatizations.

To understand the structure of the lattice  $\text{NExt}L_0$  it may be useful to look for a set  $\Gamma$  of formulas which is *complete* in the sense that its formulas are able to axiomatize all logics in the class, and *independent* in the sense that it contains no complete proper subsets. Such a set (if it exists) may be called an *axiomatic basis* of  $\text{NExt}L_0$ . The existence of an axiomatic basis depends on whether every logic in the class can be represented as the sum of “indecomposable” logics. A logic  $L \in \text{NExt}L_0$  is said to be  $\bigoplus$ -irreducible in  $\text{NExt}L_0$  if for any family  $\{L_i : i \in I\}$  of logics in  $\text{NExt}L_0$ ,  $L = \bigoplus_{i \in I} L_i$  implies  $L = L_i$  for some  $i \in I$ .  $L$  is  $\bigoplus$ -prime if for any family  $\{L_i : i \in I\}$ ,  $L \subseteq \bigoplus_{i \in I} L_i$  only if there is  $i \in I$  such that  $L \subseteq L_i$ . It is not hard to see (using Theorem 4) that a logic is  $\bigoplus$ -irreducible iff it is  $\bigoplus$ -prime. This does not hold, however, for the dual notions of  $\bigcap$ -irreducible and  $\bigcap$ -prime logics. We have only one implication in general: if  $L$  is  $\bigcap$ -prime (i.e.,  $\bigcap_{i \in I} L_i \subseteq L$  only if  $L_i \subseteq L$ , for some  $i \in I$ ) then it is  $\bigcap$ -irreducible (i.e.,  $L = \bigcap_{i \in I} L_i$  only if  $L = L_i$ , for some  $i \in I$ ). A formula  $\varphi$  is said to be *prime* in  $\text{NExt}L_0$  if  $L_0 \oplus \varphi$  is  $\bigoplus$ -prime in  $\text{NExt}L_0$ .

**PROPOSITION 5.** *Suppose a set of formulas  $\Gamma$  is complete for  $\text{NExt}L_0$  and contains no distinct deductively equivalent in  $\text{NExt}L_0$  formulas. Then  $\Gamma$  is an axiomatic basis for  $\text{NExt}L_0$  iff every formula in  $\Gamma$  is prime.*

Although the definitions above seem to be quite simple, in practice it is not so easy to understand whether a given logic is  $\bigoplus$ - or  $\bigcap$ -prime, at least at the syntactical level. However, these notions turn out to be closely related to the following lattice-theoretic concept of splitting for which in the next section we shall provide a semantic characterization.

A pair  $(L_1, L_2)$  of logics in  $\text{NExt}L_0$  is called a *splitting pair* in  $\text{NExt}L_0$  if it divides the lattice  $\text{NExt}L_0$  into two disjoint parts: the filter  $\text{NExt}L_2$  and the ideal  $[L_0, L_1]$ . In this case we also say that  $L_1$  *splits* and  $L_2$  *cosplits*  $\text{NExt}L_0$ .

**THEOREM 6.** *A logic  $L_1$  splits  $\text{NExt}L_0$  iff it is  $\bigcap$ -prime in  $\text{NExt}L_0$ , and  $L_2$  cosplits  $\text{NExt}L_0$  iff it is  $\bigoplus$ -prime in  $\text{NExt}L_0$ . Moreover, the following conditions are equivalent:*

- (i)  $(L_1, L_2)$  is a splitting pair in  $\text{NExt}L_0$ ;
- (ii)  $L_1$  is  $\bigcap$ -prime in  $\text{NExt}L_0$  and  $L_2 = \bigcap\{L \in \text{NExt}L_0 : L \not\subseteq L_1\}$ ;
- (iii)  $L_2$  is  $\bigoplus$ -prime in  $\text{NExt}L_0$  and  $L_1 = \bigoplus\{L \in \text{NExt}L_0 : L \not\supseteq L_2\}$ .

Splittings were first introduced in lattice theory by Whitman [1943] and McKenzie [1972] (see also [Day 1977], [Jipsen and Rose 1993]). Jankov [1963, 1968b, 1969], Blok [1976] and Rautenberg [1977] started using splittings in non-classical logic.

A few standard normal modal logics are listed in Table 1. Note that our notations are somewhat different from those used in *Basic Modal Logic*. ( $\mathbf{A}^*$  was introduced by Artemov; see [Shavrukov 1991]. The formulas  $B_n$  bounding depth of frames are defined in Section 15 of *Basic Modal Logic*.)

## 1.2 Semantics

The algebraic counterpart of a logic  $L \in \text{NExt}\mathbf{K}$  is the variety of modal algebras validating  $L$  (for definitions consult Section 10 of *Basic Modal Logic*). Conversely, each variety (equationally definable class)  $\mathcal{V}$  of modal algebras determines the normal modal logic  $\text{Log}\mathcal{V} = \{\varphi : \forall \mathfrak{A} \in \mathcal{V} \mathfrak{A} \models \varphi\}$ . Thus we arrive at a dual isomorphism between the lattice  $\text{NExt}\mathbf{K}$  and the lattice of varieties of modal algebras, which makes it possible to exploit the apparatus of universal algebra for studying modal logics.

It is often more convenient, however, to deal not with modal algebras directly but with their relational representations discovered by Jónsson and Tarski [1951] and now known as general frames. Each *general frame*  $\mathfrak{F} = \langle W, R, P \rangle$  is a hybrid of the usual Kripke frame  $\langle W, R \rangle$  and the modal algebra

$$\mathfrak{F}^+ = \langle P, \emptyset, W, -, \cap, \cup, \Box, \Diamond \rangle$$

in which the operations  $\Box$  and  $\Diamond$  are uniquely determined by the accessibility relation  $R$ : for every  $X \in P \subseteq 2^W$ ,

$$\Box X = \{x \in W : \forall y (xRy \rightarrow y \in X)\}, \quad \Diamond X = -\Box - X.$$

Table 1. A list of standard normal modal logics.

---

<b>D</b>	<b>=</b>	<b>K</b> $\oplus$ $\Box p \rightarrow \Diamond p$
<b>T</b>	<b>=</b>	<b>K</b> $\oplus$ $\Box p \rightarrow p$
<b>KB</b>	<b>=</b>	<b>K</b> $\oplus$ $p \rightarrow \Box \Diamond p$
<b>K4</b>	<b>=</b>	<b>K</b> $\oplus$ $\Box p \rightarrow \Box \Box p$
<b>K5</b>	<b>=</b>	<b>K</b> $\oplus$ $\Diamond \Box p \rightarrow \Box p$
<b>Alt<sub>n</sub></b>	<b>=</b>	<b>K</b> $\oplus$ $\Box p_1 \vee \Box(p_1 \rightarrow p_2) \vee \dots \vee \Box(p_1 \wedge \dots \wedge p_n \rightarrow p_{n+1})$
<b>D4</b>	<b>=</b>	<b>K4</b> $\oplus$ $\Diamond \top$
<b>S4</b>	<b>=</b>	<b>K4</b> $\oplus$ $\Box p \rightarrow p$
<b>GL</b>	<b>=</b>	<b>K4</b> $\oplus$ $\Box(\Box p \rightarrow p) \rightarrow \Box p$
<b>Grz</b>	<b>=</b>	<b>K</b> $\oplus$ $\Box(\Box(p \rightarrow \Box p) \rightarrow p) \rightarrow p$
<b>K4.1</b>	<b>=</b>	<b>K4</b> $\oplus$ $\Box \Diamond p \rightarrow \Diamond \Box p$
<b>K4.2</b>	<b>=</b>	<b>K4</b> $\oplus$ $\Diamond(p \wedge \Box q) \rightarrow \Box(p \vee \Diamond q)$
<b>K4.3</b>	<b>=</b>	<b>K4</b> $\oplus$ $\Box(\Box^+ p \rightarrow q) \vee \Box(\Box^+ q \rightarrow p)$
<b>S4.1</b>	<b>=</b>	<b>S4</b> $\oplus$ $\Box \Diamond p \rightarrow \Diamond \Box p$
<b>S4.2</b>	<b>=</b>	<b>S4</b> $\oplus$ $\Diamond \Box p \rightarrow \Box \Diamond p$
<b>S4.3</b>	<b>=</b>	<b>S4</b> $\oplus$ $\Box(\Box p \rightarrow q) \vee \Box(\Box q \rightarrow p)$
<b>Triv</b>	<b>=</b>	<b>K4</b> $\oplus$ $\Box p \leftrightarrow p$
<b>Verum</b>	<b>=</b>	<b>K4</b> $\oplus$ $\Box p$
<b>S5</b>	<b>=</b>	<b>S4</b> $\oplus$ $p \rightarrow \Box \Diamond p$
<b>K4B</b>	<b>=</b>	<b>K4</b> $\oplus$ $p \rightarrow \Box \Diamond p$
<b>A*</b>	<b>=</b>	<b>GL</b> $\oplus$ $\Box \Box p \rightarrow \Box(\Box^+ p \rightarrow q) \vee \Box(\Box^+ q \rightarrow p)$
<b>Dum</b>	<b>=</b>	<b>S4</b> $\oplus$ $\Box(\Box(p \rightarrow \Box p) \rightarrow p) \rightarrow (\Diamond \Box p \rightarrow p)$
<b>K4BW<sub>n</sub></b>	<b>=</b>	<b>K4</b> $\oplus$ $\bigwedge_{i=0}^n \Diamond p_i \rightarrow \bigvee_{0 \leq i \neq j \leq n} \Diamond(p_i \wedge (p_j \vee \Diamond p_j))$
<b>K4BD<sub>n</sub></b>	<b>=</b>	<b>K4</b> $\oplus$ $B_n$
<b>K4<sub>n,m</sub></b>	<b>=</b>	<b>K4</b> $\oplus$ $\Box^n p \rightarrow \Box^m p$ , for $1 \leq m < n$

---

So, using general frames we can take advantage of both relational and algebraic semantics. To simplify notation, we denote general frames of the form  $\mathfrak{F} = \langle W, R, 2^W \rangle$  by  $\mathfrak{F} = \langle W, R \rangle$ . Such frames will be called *Kripke frames*. Given a class of frames  $\mathcal{C}$ , we write  $\text{Log}\mathcal{C}$  to denote the logic determined by  $\mathcal{C}$ , i.e., the set of formulas that are valid in all frames in  $\mathcal{C}$ ; it is called the *logic of*  $\mathcal{C}$ . If  $\mathcal{C}$  consists of a single frame  $\mathfrak{F}$ , we write simply  $\text{Log}\mathfrak{F}$ .

Basic facts about duality between frames and algebras can be found in the chapters *Basic Modal Logic* and *Correspondence Theory* in this volume. Here we remind the reader of the definitions that will be important in what follows.

A frame  $\mathfrak{G} = \langle V, S, Q \rangle$  is said to be a *generated subframe* of a frame  $\mathfrak{F} = \langle W, R, P \rangle$  if  $V \subseteq W$  is *upward closed* in  $\mathfrak{F}$ , i.e.,  $x \in V$  and  $xRy$  imply  $y \in V$ ,  $S = R \upharpoonright V$  and  $Q = \{X \cap V : X \in P\}$ . The smallest generated subframe  $\mathfrak{G}$  of  $\mathfrak{F}$  containing a set  $X \subseteq W$  is called the *subframe generated by*  $X$ . A frame  $\mathfrak{F}$  is *rooted* if there is  $x \in W$ —a *root* of  $\mathfrak{F}$ —such that the subframe of  $\mathfrak{F}$  generated by  $\{x\}$  is  $\mathfrak{F}$  itself.

A map  $f$  from  $W$  onto  $V$  is a *reduction* (or *p-morphism*) of a frame  $\mathfrak{F} = \langle W, R, P \rangle$  to  $\mathfrak{G} = \langle V, S, Q \rangle$  if the following three conditions are satisfied for all  $x, y \in W$  and  $X \in Q$

- (R1)  $xRy$  implies  $f(x)Sf(y)$ ;
- (R2)  $f(x)Sf(y)$  implies  $\exists z \in W (xRz \wedge f(z) = f(y))$ ;
- (R3)  $f^{-1}(X) \in P$ .

The operations of reduction and generating subframes are relational counterparts of the algebraic operations of forming subalgebras and homomorphic images, respectively, and so preserve validity.

A frame  $\mathfrak{F} = \langle W, R, P \rangle$  is *differentiated* if, for any  $x, y \in W$ ,

$$x = y \text{ iff } \forall X \in P (x \in X \leftrightarrow y \in X).$$

$\mathfrak{F}$  is *tight* if

$$xRy \text{ iff } \forall X \in P (x \in \Box X \rightarrow y \in X).$$

Those frames that are both differentiated and tight are called *refined*. A frame  $\mathfrak{F}$  is said to be *compact* if every subset  $\mathcal{X}$  of  $P$  with the finite intersection property (i.e., with  $\bigcap \mathcal{X}' \neq \emptyset$  for any finite subset  $\mathcal{X}'$  of  $\mathcal{X}$ ) has non-empty intersection. Finally, refined and compact frames are called *descriptive*. A characteristic property of a descriptive  $\mathfrak{F}$  is that it is isomorphic to its bidual  $(\mathfrak{F}^+)_+$ . The classes of all differentiated, tight, refined and descriptive frames will be denoted by  $\mathcal{DF}$ ,  $\mathcal{T}$ ,  $\mathcal{R}$  and  $\mathcal{D}$ , respectively.

When representing frames in the form of diagrams, we denote by  $\bullet$  ir-reflexive points, by  $\circ$  reflexive ones, and by  $\circ\circ$  two-point clusters. An arrow from  $x$  to  $y$  means that  $y$  is accessible from  $x$ . If the accessibility relation is transitive, we draw arrows only to the immediate successors of  $x$ .

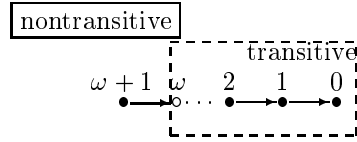


Figure 1.

EXAMPLE 7. (Van Benthem 1979) Let  $\mathfrak{F} = \langle W, R, P \rangle$  be the frame whose underlying Kripke frame is shown in Fig. 1 ( $\omega + 1$  sees only  $\omega$  and the subframe generated by  $\omega$  is transitive) and  $X \subseteq W$  is in  $P$  iff either  $X$  is finite and  $\omega \notin X$  or  $X$  is cofinite in  $W$  and  $\omega \in X$ . It is easy to see that  $P$  is closed under  $\cap$ ,  $-$  and  $\diamond$ . Clearly,  $\mathfrak{F}$  is refined. Suppose  $\mathcal{X}$  is a subset of  $P$  with the finite intersection property. If  $\mathcal{X}$  contains a finite set then obviously  $\bigcap \mathcal{X} \neq \emptyset$ . And if  $\mathcal{X}$  consists of only infinite sets then  $\omega \in \bigcap \mathcal{X}$ . Thus,  $\mathfrak{F}$  is descriptive.

A frame  $\mathfrak{F}$  is said to be  $\varkappa$ -generated,  $\varkappa$  a cardinal, if its dual  $\mathfrak{F}^+$  is a  $\varkappa$ -generated algebra.<sup>4</sup> Each modal logic  $L$  is determined by the free finitely generated algebras in the corresponding variety, i.e., by the Tarski–Lindenbaum (or canonical) algebras  $\mathfrak{A}_L(n)$  for  $L$  in the language with  $n < \omega$  variables. Their duals are denoted by  $\mathfrak{F}_L(n) = \langle W_L(n), R_L(n), P_L(n) \rangle$  and called the *universal frames of rank  $n$  for  $L$* . Analogous notation and terminology will be used for the free algebras  $\mathfrak{A}_L(\varkappa)$  with  $\varkappa$  generators. Note that  $\langle W_L(\varkappa), R_L(\varkappa) \rangle$  is (isomorphic to) the canonical Kripke frame for  $L$  with  $\varkappa$  variables (defined in Section 11 of *Basic Modal Logic*) and  $P_L(\varkappa)$  is the collection of the truth-sets of formulas in the corresponding canonical model. Unless otherwise stated, we will assume in what follows that the language of the logics under consideration contains  $\omega$  variables.

An important property of the universal frame of rank  $\varkappa$  for  $L$  is that every descriptive  $\varkappa'$ -generated frame for  $L$ ,  $\varkappa' \leq \varkappa$ , is a generated subframe of  $\mathfrak{F}_L(\varkappa)$ . Thus, the more information about universal frames for  $L$  we have, the deeper our knowledge about the structure of arbitrary frames for  $L$  and thereby about  $L$  itself.

Although in general universal frames for modal logics are very complicated, considerable progress was made in clarifying the structure of the upper part (points of finite depth) of the universal frames of finite rank for logics in NExt**K4**. The studies in this direction were started actually by Segerberg [1971]. Shehtman [1978a] presented a general method of constructing the universal frames of finite rank for logics in NExt**S4** with the finite model property. Later similar results were obtained by other authors; see e.g. [Bellissima 1985]. The structure of free finitely generated algebras

<sup>4</sup>An algebra is said to be  $\varkappa$ -generated if it contains a set  $X$  of cardinality  $\leq \varkappa$  such that the closure of  $X$  under the algebra's operations coincides with its universe.

for **S4** was investigated by Blok [1976].

Let us try to understand first the constitution of an arbitrary transitive refined frame  $\mathfrak{F} = \langle W, R, P \rangle$  with  $n$  generators  $G_1, \dots, G_n \in P$ . Define  $\mathfrak{V}$  to be the valuation of the set of variables  $\Sigma = \{p_1, \dots, p_n\}$  in  $\mathfrak{F}$  such that  $x \models p_i$  iff  $x \in G_i$ . Say that points  $x$  and  $y$  are  $\Sigma$ -equivalent,  $x \sim_\Sigma y$  in symbols, if the same variables in  $\Sigma$  are true at them; for  $X, Y \subseteq W$  we write  $X \sim_\Sigma Y$  if every point in  $X$  is  $\Sigma$ -equivalent to some point in  $Y$  and vice versa. Let  $d(\mathfrak{F})$  denote the *depth*<sup>5</sup> of  $\mathfrak{F}$ ; if  $\mathfrak{F}$  is of infinite depth, we write  $d(\mathfrak{F}) = \infty$ . For  $d < d(\mathfrak{F})$ ,  $W^{=d}$  and  $W^{>d}$  are the sets of all points in  $\mathfrak{F}$  of depth  $d$  and  $> d$ , respectively;  $W^{<d}$ ,  $W^{\leq d}$ , etc. are defined analogously.  $\mathfrak{F}^{\leq d}$  is the subframe of  $\mathfrak{F}$  generated by  $W^{\leq d}$ . The set of all successors (predecessors) of points in a set  $X \subseteq W$  is denoted by  $X\uparrow$  (respectively,  $X\downarrow$ ); in the transitive case  $X\uparrow = X\uparrow \cup X$  and  $X\downarrow = X\downarrow \cup X$  are then the upward and downward closure operations. A set  $X$  is said to be a *cover* for a set  $Y$  in  $\mathfrak{F}$  if  $Y \subseteq X\downarrow$ . A point  $x$  is called an *atom* in  $\mathfrak{F}$  if  $\{x\} \in P$ .

**THEOREM 8.** *Suppose  $\mathfrak{F} = \langle W, R, P \rangle$  is a transitive refined  $n$ -generated frame, for some  $n < \omega$ . Then*

- (i) *each cluster in  $\mathfrak{F}$  contains  $\leq 2^n$  points;*
- (ii) *for every finite  $d \leq d(\mathfrak{F})$ ,  $W^{=d}$  is a cover for  $W^{\geq d}$  and contains at most  $c_n(d)$  distinct clusters, where*

$$c_n(1) = 2^n + 2^{2^n} - 1, \quad c_n(m+1) = c_n(1) \cdot 2^{c_n(1) + \dots + c_n(m)};$$

- (iii) *every point of finite depth in  $\mathfrak{F}$  is an atom.*

**Proof.** (i) follows from the differentiatedness of  $\mathfrak{F}$  and the obvious fact that precisely the same formulas (in  $p_1, \dots, p_n$ ) are true under  $\mathfrak{V}$  at  $\Sigma$ -equivalent points in the same cluster.

The proof of (ii) proceeds by induction on  $d$ . Let  $x \in W^{>d}$ . Since  $\mathfrak{F}$  is transitive and  $W^{\leq d}$  is finite (by the induction hypothesis), there exists a non-empty upward closed in  $W^{>d}$  set  $X$  (i.e.,  $X = X\uparrow \cap W^{>d}$ ) such that  $x \in X\downarrow$ , points in  $X$  see exactly the same points of depth  $\leq d$  and either

$$(1) \quad \forall u, v \in X \exists w \in u\uparrow \cap X \ w \sim_\Sigma v$$

or

$$(2) \quad \forall u, v \in X \ (u \sim_\Sigma v \wedge \neg uRv).$$

Such a set  $X$  is called *d-cyclic*; it is *nondegenerate* if (1) holds and *degenerate* otherwise. One can readily show that the same formulas are true at  $\Sigma$ -equivalent points in  $X$ . Since  $\mathfrak{F}$  is refined,  $X$  is then a cluster of depth  $d+1$ . Thus,  $W^{>d} \subseteq W^{=d+1}\downarrow$ . The upper bound for the number of distinct

<sup>5</sup>In Section 15 of *Basic Modal Logic*  $d(\mathfrak{F})$  was called the rank of  $\mathfrak{F}$ .

clusters of depth  $d + 1$  follows from the differentiatedness of  $\mathfrak{F}$  and the definition of  $d$ -cyclic sets.

To establish (iii), for every point  $x$  of depth  $d + 1$  one can construct by induction on  $d$  a formula (expressing the definition of the  $d$ -cyclic set containing  $x$ ) which is true in  $\mathfrak{F}$  under  $\mathfrak{V}$  only at  $x$ . For details consult [Chagrov and Zakharyashev 1997]. ■

It is fairly easy now to construct the (generated) subframe  $\mathfrak{F}_{\mathbf{K4}}^{\leq \infty}(n)$  of the universal frame of rank  $n$  for  $\mathbf{K4}$  consisting of finite depth points. Indeed,  $\mathfrak{F}_{\mathbf{K4}}(n)$  is  $n$ -generated, refined and so has the form as described in Theorem 8. On the other hand, it is universal and contains any  $n$ -generated descriptive frame as a generated subframe, which means roughly that it contains all possible points of finite depth that can exist in  $n$ -generated refined frames.

More precisely, assuming that each point is assigned the set of variables in  $\Sigma$  that are true at it, we begin constructing a frame  $\mathfrak{G}_{\mathbf{K4}}(n)$  by putting at depth 1 in it  $2^n$  non- $\Sigma$ -equivalent degenerate clusters and  $2^{2^n} - 1$  non- $\Sigma$ -equivalent non-degenerate clusters with  $\leq 2^n$  non- $\Sigma$ -equivalent points. Suppose that  $\mathfrak{G}_{\mathbf{K4}}^{\leq d}(n)$  is already constructed. Then for every antichain  $\mathfrak{a}$  of clusters in  $\mathfrak{G}_{\mathbf{K4}}^{\leq d}(n)$  containing at least one cluster of depth  $d$  and different from a singleton with a non-degenerate cluster, we add to  $\mathfrak{G}_{\mathbf{K4}}^{\leq d}(n)$  copies of all  $2^n + 2^{2^n} - 1$  clusters of depth 1 so that they would be inaccessible from each other and could see only the clusters in  $\mathfrak{a}$  and their successors. And for every singleton  $\mathfrak{a} = \{C\}$  with a non-degenerate cluster  $C$ , we add to  $\mathfrak{G}_{\mathbf{K4}}^{\leq d}(n)$  copies of those clusters of depth 1 which are not  $\Sigma$ -equivalent to any subset of  $C$  (otherwise the frame will not be refined) so that again they would be mutually inaccessible and could see only  $C$  and its successors in  $\mathfrak{G}_{\mathbf{K4}}^{\leq d}(n)$ .

Let  $\mathfrak{M}_{\mathbf{K4}}(n) = \langle \mathfrak{G}_{\mathbf{K4}}(n), \mathfrak{U}_{\mathbf{K4}}(n) \rangle$  be the resulting model (the relational component of  $\mathfrak{G}_{\mathbf{K4}}(n)$  is completely determined by the construction and its set of possible values is the collection of the truth-sets of formulas in  $\mathfrak{G}_{\mathbf{K4}}(n)$  under  $\mathfrak{U}_{\mathbf{K4}}(n)$ ). It is not hard to show that  $\mathfrak{G}_{\mathbf{K4}}(n)$  is atomic. Moreover, for every point  $x$  in this frame one can construct a formula  $\varphi(p_1, \dots, p_n)$  such that  $x \not\models \varphi$  and, for any frame  $\mathfrak{F}$ ,  $\mathfrak{F} \models \varphi$  iff there is a generated subframe of  $\mathfrak{F}$  reducible to the subframe of  $\mathfrak{G}_{\mathbf{K4}}(n)$  generated by  $x$ . It follows in particular that  $\mathfrak{G}_{\mathbf{K4}}(n)$  is refined. Thus, every  $\mathfrak{G}_{\mathbf{K4}}^{\leq d}(n)$  is a generated subframe of  $\mathfrak{F}_{\mathbf{K4}}(n)$ . On the other hand, by Theorem 8,  $\mathfrak{F}_{\mathbf{K4}}(n)$  contains no clusters of depth  $\leq d$  different from those in  $\mathfrak{G}_{\mathbf{K4}}^{\leq d}(n)$  and so  $\mathfrak{F}_{\mathbf{K4}}^{\leq \infty}(n)$  is isomorphic to  $\mathfrak{G}_{\mathbf{K4}}(n)$ . It worth noting also that, since  $\mathbf{K4}$  has the finite model property, it is characterized by  $\mathfrak{F}_{\mathbf{K4}}^{\leq \infty}(n)$ , and so  $\mathfrak{F}_{\mathbf{K4}}(n)$  is isomorphic to the bidual of  $\mathfrak{F}_{\mathbf{K4}}^{\leq \infty}(n)$ .

The universal frame  $\mathfrak{F}_L(n)$  for an arbitrary consistent logic  $L$  in  $\text{NExt}\mathbf{K4}$  is a generated subframe of  $\mathfrak{F}_{\mathbf{K4}}(n)$ . It can be constructed by removing from  $\mathfrak{F}_{\mathbf{K4}}(n)$  those points at which some formulas in  $L$  are refuted (under



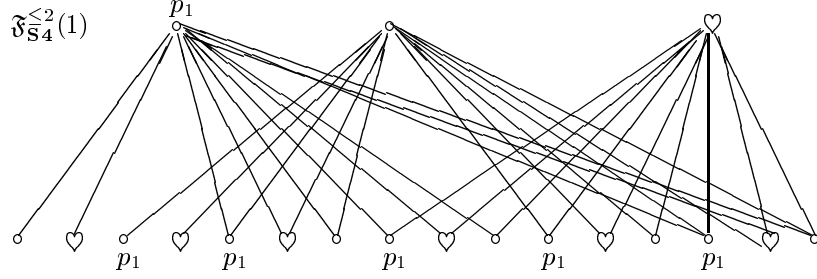


Figure 2.

$\mathfrak{V}_{\mathbf{K}_4}(n)$ . For example,  $\mathfrak{F}_{S_4}^{<\infty}(n)$  is obtained by removing from  $\mathfrak{F}_{\mathbf{K}_4}^{<\infty}(n)$  all irreflexive points and their predecessors. In other words,  $\mathfrak{F}_{S_4}^{<\infty}(n)$  can be constructed in the same way as  $\mathfrak{F}_{\mathbf{K}_4}^{<\infty}(n)$  but using only non-degenerate clusters.  $\mathfrak{F}_{S_4}^{<2}(1)$  (the corresponding model, to be more exact) is shown in Fig. 2, where  $\heartsuit$  denotes the cluster with two points at one of which  $p_1$  is true. To construct  $\mathfrak{F}_{Grz}^{<\infty}(n)$  and  $\mathfrak{F}_{GL}^{<\infty}(n)$ , we take only simple clusters and degenerate clusters, respectively.

In general, this method of constructing universal frames does not work for logics with nontransitive frames. However, using the fact that  $\mathbf{K}$  is characterized by the class of finite intransitive irreflexive trees (see Section 13 of *Basic Modal Logic*), in the same manner as above one can construct an intransitive irreflexive model characterizing  $\mathbf{K}$  and such that  $\mathfrak{F}_{\mathbf{K}}(n)$  is isomorphic to the bidual of the frame associated with this model.

Let us consider now the semantical meaning of splittings. In view of the following observation we focus attention only on splittings by the logics of finite rooted frames.

**THEOREM 9.** *If  $L_1$  splits  $\text{NExt}L_0$  and  $L_0$  has the finite model property then  $L_1 = \text{Log}\mathfrak{F}$ , for some finite rooted frame  $\mathfrak{F}$  validating  $L_0$ .*

**Proof.** Since  $L_2$  in the splitting pair  $(L_1, L_2)$  is a proper extension of  $L_0$ , there is a finite frame  $\mathfrak{G}$  such that  $\mathfrak{G} \models L_0$  and  $\mathfrak{G} \not\models L_2$ . It follows that  $\text{Log}\mathfrak{G} \subseteq L_1$ . As we shall see later (Corollary 86), every extension of a tabular logic is also tabular. So  $L_1 = \text{Log}\mathfrak{F}$  for some finite  $\mathfrak{F} \models L_0$ . And since  $L_1$  is  $\bigcap$ -prime,  $\mathfrak{F}$  must be rooted.  $\blacksquare$

We say that a frame  $\mathfrak{F}$  *splits*  $\text{NExt}L_0$  if  $\text{Log}\mathfrak{F}$  splits  $\text{NExt}L_0$ . The logic  $L_2$  of the splitting pair  $(\text{Log}\mathfrak{F}, L_2)$  is denoted by  $L_0/\mathfrak{F}$  and called the *splitting* of  $\text{NExt}L_0$  by  $\mathfrak{F}$ . This notation reflects the fact that  $L_2$  is the smallest logic in  $\text{NExt}L_0$  which is not validated by  $\mathfrak{F}$ .

**EXAMPLE 10.** We show that  $\mathbf{D} = \mathbf{K}/\bullet$ . Recall that  $\mathbf{D} = \mathbf{K} \oplus \diamond\top$  is characterized by the class of serial frames (in which every point has a suc-

cessor). So if  $\bullet \models L$  then  $L \subseteq \text{Log}\bullet$ ; otherwise no frame for  $L$  has a dead end, which means that  $\diamond\top \in L$  and  $\mathbf{D} \subseteq L$ . The inconsistent logic  $\mathbf{For}$  can be represented as  $\mathbf{D}/\circ$ .

To illustrate some applications of splittings we require a few definitions. Given  $L \in \text{NExt}L_0$ , we say that the *axiomatization problem* for  $L$  above  $L_0$  is decidable if the set  $\{\varphi : L_0 \oplus \varphi = L\}$  is recursive.  $L$  is *strictly Kripke complete* above  $L_0$  if no other logic in  $\text{NExt}L_0$  has exactly the same Kripke frames as  $L$ . If all frames in a set  $\mathcal{F}$  split  $\text{NExt}L_0$ , we call the logic  $\bigoplus\{L_0/\mathfrak{F} : \mathfrak{F} \in \mathcal{F}\}$  the *union-splitting* of  $\text{NExt}L_0$  and denote it by  $L_0/\mathcal{F}$ .

EXAMPLE 11.  $\mathbf{Grz}$  is not a splitting of  $\text{NExt}\mathbf{S4}$ . However, it is a union-

splitting:  $\mathbf{Grz} = \mathbf{S4}/\{\overset{\circ}{\circ\circ}, \overset{\circ}{\circ\circ}\}$ .  $\mathbf{S4.1} = \mathbf{S4}/\{\overset{\circ}{\circ\circ}\}$ . A frame may split the lattice  $\text{NExt}L_0/\mathcal{F}$  but not  $\text{NExt}L_0$ : e.g.  $\circ$  splits  $\text{NExt}\mathbf{K}/\bullet$  but does not split  $\text{NExt}\mathbf{K}$ .

THEOREM 12. Suppose  $L \in \text{NExt}L_0$  and  $L = (\dots(L_0/\mathcal{F}_1)/\dots)/\mathcal{F}_n$ , for a sequence  $\mathcal{F}_1, \dots, \mathcal{F}_n$  of sets of finite rooted frames.

(i) If  $\mathcal{F} = \bigcup_{i=1}^n \mathcal{F}_i$  is finite and  $L$  is decidable then the axiomatization problem for  $L$  above  $L_0$  is decidable. More precisely,

$$\{\varphi : L_0 \oplus \varphi = L\} = \{\varphi \in L : \forall \mathfrak{F} \in \mathcal{F} \mathfrak{F} \not\models \varphi\}.$$

(ii) If  $L$  is Kripke complete then  $L$  is strictly Kripke complete above  $L_0$ .

(iii) The immediate predecessors of  $L$  in  $\text{NExt}L_0$  are precisely the logics  $L \cap \text{Log}\mathfrak{F}$ , for  $\mathfrak{F} \in \mathcal{F}$  such that  $\mathfrak{F}$  is not a reduct of a generated subframe of another frame in  $\mathcal{F}$ .

**Proof.** (i) is left to the reader as an easy exercise.

(ii) Let  $L'$  be a logic in  $\text{NExt}L_0$  with the same Kripke frames as  $L$ . Then obviously  $L' \subseteq L$ . On the other hand, the frames in  $\mathcal{F}$  do not validate  $L'$  and so  $L \subseteq L'$ .

(iii) If  $L'$  is an immediate predecessor of  $L$  in  $\text{NExt}L_0$  then  $\mathfrak{F} \models L'$ , for some  $\mathfrak{F} \in \mathcal{F}$ . Therefore,  $L' \subseteq L \cap \text{Log}\mathfrak{F} \subset L$  and so  $L' = L \cap \text{Log}\mathfrak{F}$ . Suppose now that  $\mathfrak{F}$  is not a reduct of a generated subframe of another frame in  $\mathcal{F}$  and  $L \cap \text{Log}\mathfrak{F} \subseteq L' \subset L$ . Then  $L' \subseteq \text{Log}\mathfrak{F}'$  for some  $\mathfrak{F}' \in \mathcal{F}$ , and hence  $\mathfrak{F}' = \mathfrak{F}$ ,  $L' = L \cap \text{Log}\mathfrak{F}$ . ■

As follows from Theorem 12 and Example 10,  $\mathbf{For}$  has exactly two immediate predecessors in  $\text{NExt}\mathbf{K}$ :  $\mathbf{Verum} = \text{Log}\bullet$  and  $\mathbf{Triv} = \text{Log}\circ$  (and each consistent normal modal logic is contained in one of them). This result is known as Makinson's [1971] Theorem. Moreover, the axiomatization problem for  $\mathbf{For}$  is decidable, i.e., there is an algorithm which decides, given a formula  $\varphi$  whether  $\mathbf{K} \oplus \varphi$  is consistent. Likewise, since  $\mathbf{D} = \mathbf{K} \oplus \diamond\top$  is decidable, there is an algorithm recognizing, given  $\varphi$ , whether  $\mathbf{D} = \mathbf{K} \oplus \varphi$ .

We shall see later in Section 4.4 that in fact not so many properties of logics are decidable (e.g. the axiomatization problem for  $\mathbf{K} \oplus \neg\Diamond\top$  is undecidable; see Theorem 207) and that Theorem 12 (i) provides the main method for proving decidability results of this type.

To determine whether a finite rooted frame  $\mathfrak{F} = \langle W, R \rangle$  splits  $\text{NExt}L_0$ , we need the formulas defined below:

$$\begin{aligned} \Delta_{\mathfrak{F}} &= \{p_x \rightarrow \Diamond p_y : x, y \in W, xRy\} \cup \\ &\quad \{p_x \rightarrow \neg\Diamond p_y : x, y \in W, \neg xRy\} \cup \\ &\quad \{p_x \rightarrow \neg p_y : x, y \in W, x \neq y\}, \\ \sigma_{\mathfrak{F}} &= \bigwedge \Delta_{\mathfrak{F}}, \quad \delta_{\mathfrak{F}} = \sigma_{\mathfrak{F}} \wedge \bigvee \{p_x : x \in W\}. \end{aligned}$$

The meaning of  $\delta_{\mathfrak{F}}$  is explained by the following lemma, in which

$$\Box^{<\omega}\varphi = \{\Box^n\varphi : n < \omega\}.$$

**LEMMA 13.** *For any finite  $\mathfrak{F}$  with root  $r$ , the set of formulas  $\{p_r\} \cup \Box^{<\omega}\delta_{\mathfrak{F}}$  is satisfiable in a frame  $\mathfrak{G}$  iff there is a generated subframe  $\mathfrak{H}$  of  $\mathfrak{G}$  reducible to  $\mathfrak{F}$ . Moreover, if  $\mathfrak{F}$  is cycle free (i.e., contains no path from a point to itself) then  $\omega$  can be replaced by  $n = d(\mathfrak{F}) + 1$ .*

**Proof.** ( $\Rightarrow$ ) Suppose  $\{p_r\} \cup \Box^{<\omega}\delta_{\mathfrak{F}}$  is satisfied at a point  $u$  in  $\mathfrak{G}$ . It is not hard to check that the map  $f$  defined by  $f(v) = x$  iff  $v \models p_x$  is a reduction of the subframe  $\mathfrak{H}$  of  $\mathfrak{G}$  generated by  $u$  to  $\mathfrak{F}$ . If  $\mathfrak{F}$  is cycle free and  $\{p_r\} \cup \Box^{<\omega}\delta_{\mathfrak{F}}$  is satisfied at  $u$  then  $d(\mathfrak{H}) = d(\mathfrak{F})$ . For otherwise an ascending chain of  $n+1$  points starts from  $u$  and so  $\mathfrak{F}$  must contain a cycle.

( $\Leftarrow$ ) Let  $f$  be a reduction of  $\mathfrak{H}$  to  $\mathfrak{F}$ . Define a valuation in  $\mathfrak{G}$  so that  $v \models p_x$  iff  $v \in f^{-1}(x)$ . The reader can readily verify that under this valuation  $\{p_r\} \cup \Box^{<\omega}\delta_{\mathfrak{F}}$  is true at any point in  $f^{-1}(r)$ .  $\blacksquare$

**LEMMA 14.** *For every logic  $L \in \text{NExt}\mathbf{K}$  and every finite rooted frame  $\mathfrak{F}$ ,  $\mathfrak{F} \models L$  iff  $\forall n < \omega \Box^{\leq n}\delta_{\mathfrak{F}} \rightarrow \neg p_r \notin L$ .*

**Proof.** The implication ( $\Rightarrow$ ) follows from Lemma 13. Suppose now that  $\Box^{\leq n}\delta_{\mathfrak{F}} \rightarrow \neg p_r \notin L$ , for every  $n < \omega$ . Then the set  $\{p_r\} \cup \Box^{<\omega}\delta_{\mathfrak{F}}$  is  $L$ -consistent and so it is satisfied in a frame  $\mathfrak{G}$  for  $L$ . By Lemma 13, a generated subframe of  $\mathfrak{G}$  is reducible to  $\mathfrak{F}$ , and hence  $\mathfrak{F} \models L$ .  $\blacksquare$

We are now in a position to characterize finite frames that split  $\text{NExt}L_0$  and to axiomatize splittings.

**THEOREM 15.** *Suppose  $\mathfrak{F}$  is a finite frame with root  $r$  and  $L_0 \in \text{NExt}\mathbf{K}$ . Then  $\mathfrak{F}$  splits  $\text{NExt}L_0$  iff there is  $n < \omega$  such that, for every frame  $\mathfrak{G} \models L_0$ ,  $\Box^{\leq n}\delta_{\mathfrak{F}} \wedge p_r$  is satisfiable in  $\mathfrak{G}$  only if  $\Box^{\leq m}\delta_{\mathfrak{F}} \wedge p_r$  is satisfiable in  $\mathfrak{G}$  for every  $m > n$ . In this case  $L_0/\mathfrak{F} = L_0 \oplus \Box^{\leq n}\delta_{\mathfrak{F}} \rightarrow \neg p_r$ .*

**Proof.** ( $\Rightarrow$ ) Suppose otherwise and consider a sequence  $\{\mathfrak{G}_n : n < \omega\}$  of frames for  $L_0$  such that  $\Box^{\leq n} \delta_{\mathfrak{F}} \wedge p_r$  is satisfiable in  $\mathfrak{G}_n$  but  $\Box^{\leq m} \delta_{\mathfrak{F}} \wedge p_r$  is not satisfied, for some  $m > n$ . By Lemma 14, the former condition implies  $\bigcap_{n < \omega} \text{Log} \mathfrak{G}_n \subseteq \text{Log} \mathfrak{F}$ , while the latter means that  $\mathfrak{F} \not\models \text{Log} \mathfrak{G}_n$ , for every  $n < \omega$ , contrary to  $\text{Log} \mathfrak{F}$  being  $\bigcap$ -prime.

( $\Leftarrow$ ) We show that  $L_0/\mathfrak{F} = L_0 \oplus \Box^{\leq n} \delta_{\mathfrak{F}} \rightarrow \neg p_r$ . Suppose  $L \not\subseteq \text{Log} \mathfrak{F}$ . Then, by Lemma 14, there is  $m < \omega$  such that  $\Box^{\leq m} \delta_{\mathfrak{F}} \rightarrow \neg p_r \in L$ . It follows that  $\Box^{\leq n} \delta_{\mathfrak{F}} \rightarrow \neg p_r \in L$  and so  $L_0 \oplus \Box^{\leq n} \delta_{\mathfrak{F}} \rightarrow \neg p_r \subseteq L$ . ■

For more general versions of this criterion consult [Kracht 1990] and [Wolter 1993].

**COROLLARY 16** (Rautenberg 1980). *Suppose that  $L_0 \in \text{NExt}(\mathbf{K} \oplus \text{tra}_n)$ , for some  $n < \omega$ . Then every finite rooted frame  $\mathfrak{F}$  for  $L_0$  splits  $\text{NExt} L_0$  and  $L_0/\mathfrak{F} = L_0 \oplus \Box^{\leq n} \delta_{\mathfrak{F}} \rightarrow \neg p_r$ .*

In particular, every transitive finite rooted frame splits  $\text{NExt} \mathbf{K4}$ . This result may also be obtained using the fact that all finite subdirectly irreducible algebras split the lattice of subvarieties of a variety with equationally definable principal congruences (see [Blok and Pigozzi 1982]). However, not every frame splits  $\text{NExt} \mathbf{K}$ .

**THEOREM 17** (Blok 1978). *A finite rooted frame  $\mathfrak{F}$  splits  $\text{NExt} \mathbf{K}$  iff it is cycle free. In this case  $\mathbf{K}/\mathfrak{F} = \mathbf{K} \oplus \Box^{\leq n} \delta_{\mathfrak{F}} \rightarrow \neg p_r$ , where  $n = d(\mathfrak{F})$ .*

**Proof.** That frames with cycles do not split  $\text{NExt} \mathbf{K}$  follows from the fact that  $\mathbf{K}$  is characterized by cycle free finite rooted frames. And the converse is an immediate consequence of Lemma 13 and Theorem 15. ■

An element  $x \neq 0$  of a complete lattice  $\mathfrak{L}$  is called an *atom* in  $\mathfrak{L}$  if the zero element  $0$  in  $\mathfrak{L}$  is the immediate predecessor of  $x$ , i.e., there is no  $y$  such that  $0 < y < x$ . Splittings turn out to be closely related to the existence of atoms in finitely generated free algebras; see [Blok 1976], [Bellissima 1984, 1991] and [Wolter 1997c]. We demonstrate the use of splittings by the following

**THEOREM 18** (Blok 1980a). *The lattice  $\text{NExt} \mathbf{K}$  has no atoms.*

**Proof.** If a logic  $L$  is an atom in  $\text{NExt} \mathbf{K}$ , it is  $\bigoplus$ -prime. It follows that  $L$  cosplits  $\text{NExt} \mathbf{K}$  and the logic  $L' = \text{Log} \mathfrak{F}$  in the splitting pair  $(L', L)$  has no proper predecessor that splits  $\text{NExt} \mathbf{K}$ . Add a new irreflexive root to  $\mathfrak{F}$ . By Theorem 17, the resulting frame  $\mathfrak{G}$  splits  $\text{NExt} \mathbf{K}$ , and clearly  $\text{Log} \mathfrak{G} \subset \text{Log} \mathfrak{F}$ , which is a contradiction. ■

A logic is linked with its semantics via completeness theorems. The most general completeness theorem states that every consistent normal modal logic is characterized by the class of (descriptive) frames validating it. Or, if we want to characterize the consequence relations  $\vdash_L$  and  $\vdash_L^*$ , we can use the following

THEOREM 19. (i) For  $L \in \text{NExt}\mathbf{K}$ ,  $\Gamma \vdash_L \varphi$  iff for any model  $\mathfrak{M}$  based on a frame for  $L$  and any point  $x$  in  $\mathfrak{M}$ ,  $x \models \Gamma$  implies  $x \models \varphi$ .

(ii) For  $L \in \text{NExt}\mathbf{K}$ ,  $\Gamma \vdash_L^* \varphi$  iff for any model  $\mathfrak{M}$  based on a frame for  $L$ ,  $\mathfrak{M} \models \Gamma$  implies  $\mathfrak{M} \models \varphi$ .

However, usually more specific completeness results are required. What is the “geometry” of frames for a given logic? Are Kripke or even finite frames enough to characterize it? Questions of this sort will be addressed in the next several sections.

### 1.3 Persistence

The structure of Kripke frames for many standard modal logics can be described by rather simple conditions on the accessibility relation which are expressed in the first order language with equality and a binary (accessibility) predicate  $R$ . (This observation was actually the starting point of investigations in *Correspondence Theory* studying the relation between modal and first (or higher) order languages; see Chapter 4 of this volume.) Moreover, in many cases it turns out that the universal frame  $\mathfrak{F}_L(\omega)$  for such a logic  $L$  also satisfies the corresponding first order condition  $\phi$ . Since  $\phi$  says nothing about sets of possible values in  $P_L(\omega)$ , it follows immediately that the canonical (Kripke) frame  $\kappa\mathfrak{F}_L(\omega)$  also satisfies  $\phi$  and so characterizes  $L$ . Thus we obtain a completeness theorem of the form:

$$\varphi \in L \text{ iff } \mathfrak{F} \models \varphi \text{ for every Kripke frame } \mathfrak{F} \text{ satisfying } \phi.$$

This method of establishing Kripke completeness, known as the *method of canonical models*, is based essentially upon two facts: first, that  $L$  is characterized by its universal frame  $\mathfrak{F}_L(\omega)$  and second, that  $L$  is “persistent” under the transition from  $\mathfrak{F}_L(\omega)$  to its underlying Kripke frame. Of course, instead of  $\mathfrak{F}_L(\omega)$  we can take any other class of frames  $\mathcal{C}$  with respect to which  $L$  is complete and try to show that  $L$  is  $\mathcal{C}$ -persistent in the sense that, for every  $\mathfrak{F} = \langle W, R, P \rangle$  in  $\mathcal{C}$ , if  $\mathfrak{F} \models L$  then  $\kappa\mathfrak{F} = \langle W, R \rangle$  validates  $L$  as well.

PROPOSITION 20. *If a logic is both  $\mathcal{C}$ -complete and  $\mathcal{C}$ -persistent, then it is complete with respect to the class  $\{\kappa\mathfrak{F} : \mathfrak{F} \in \mathcal{C}\}$  of Kripke frames.*

It follows in particular that  $L$  is Kripke complete whenever it is  $\mathcal{DF}$ -, or  $\mathcal{R}$ -, or  $\mathcal{D}$ -persistent. Since every descriptive frame for  $L$  is a generated subframe of a suitable universal frame for  $L$ ,  $L$  is  $\mathcal{D}$ -persistent iff it is persistent with respect to the class of its universal frames. It is an open problem, however, whether *canonicity*, i.e.,  $\mathfrak{F}_L(\omega)$ -persistence, implies  $\mathcal{D}$ -persistence. Here are two simple examples.

THEOREM 21 (van Benthem 1983). *A logic is persistent with respect to the class of all general frames iff it is axiomatizable by a set of variable free formulas.*

It is easily checked that a Kripke frame validates  $\mathbf{Alt}_n$  iff no point in it has more than  $n$  distinct successors (see [Seegerberg 1971]).

**THEOREM 22** (Bellissima 1988). *Every  $L \in \mathbf{NExtAlt}_n$  is  $\mathcal{DF}$ -persistent, for any  $n < \omega$ .*

**Proof.** The proof is based on the fact that, for any differentiated frame  $\mathfrak{F} = \langle W, R, P \rangle$ , any finite  $X \subseteq W$ , and any  $y \in X$ , there is  $Y \in P$  such that  $X \cap Y = \{y\}$ . It follows that at most  $n$  distinct points are accessible from every point in a differentiated frame for  $L$ ; in particular,  $\mathbf{Alt}_n$  is  $\mathcal{DF}$ -persistent. Suppose now that a formula  $\varphi \in L$  is refuted at a point  $x$  under a valuation  $\mathfrak{V}$  in  $\kappa\mathfrak{F}$ ,  $\mathfrak{F}$  a differentiated frame for  $L$ . Let  $X$  be the set of points accessible from  $x$  in  $\leq md(\varphi)$  steps.<sup>6</sup> Since  $X$  is finite, there is a valuation  $\mathfrak{U}$  in  $\mathfrak{F}$  such that  $\mathfrak{U}(p) \cap X = \mathfrak{V}(p)$ , for every variable  $p$ . Consequently,  $\varphi$  is false in  $\mathfrak{F}$  at  $x$  under  $\mathfrak{U}$ , which is a contradiction. ■

The proof of Fine's [1974c] Theorem that all logics of finite width, i.e., logics in  $\mathbf{NExtK4BW}_n$ , for  $n < \omega$ , are Kripke complete (a sketch can be found in Section 18 of *Basic Modal Logic*) may also be regarded as a proof of persistence. Recall that a point  $x$  in a transitive frame  $\mathfrak{F} = \langle W, R, P \rangle$  is called *non-eliminable* (relative to  $R$ ) if there is  $X \in P$  such that  $x \in X$  but no proper successor of  $x$  is in  $X$  (in other words,  $x$  is *maximal* in  $X$ ); in this case we write  $x \in \max_R X$ . Denote by  $W_r$  the set of all non-eliminable points in  $\mathfrak{F}$  and put  $\mathfrak{F}_r = \langle W_r, R_r, P_r \rangle$ , where  $R_r = R \upharpoonright W_r$ ,  $P_r = \{X \cap W_r : X \in P\}$ . (Fine called the frame  $\mathfrak{F}_r$  *reduced*.)

**THEOREM 23** (Fine 1985). *Let  $\mathfrak{F} = \langle W, R, P \rangle$  be a transitive descriptive frame and  $x \in X \in P$ . Then (i) there exists a point  $y \in \max_R X \cap x \uparrow$  and (ii)  $\mathfrak{F}_r$  is a refined frame whose dual  $\mathfrak{F}_r^+$  is isomorphic to  $\mathfrak{F}^+$ .*

**Proof.** (i) Suppose otherwise, i.e., there is no maximal point in  $X \cap x \uparrow$ . Let  $Y$  be a maximal chain of points in  $X \cap x \uparrow$  (that it exists follows from Zorn's Lemma) and  $\mathcal{X} = \{Z \in P : \exists y \in Y \ y \uparrow \cap Y \subseteq Z\}$ . Clearly,  $\mathcal{X}$  is non-empty and has the finite intersection property (because  $X \cap x \uparrow$  has no maximal point). By compactness, we then have a point  $z$  in  $\bigcap \mathcal{X}$  which, by tightness, is maximal in  $Y$ , contrary to  $X \cap x \uparrow$  having no maximal point. (ii) is a consequence of (i). ■

It follows that to establish the Kripke completeness of a logic  $L \in \mathbf{NExtK4}$  it is enough to show that it is persistent with respect to the class

$$\mathcal{RE} = \{\mathfrak{F}_r : \mathfrak{F} \text{ a finitely generated descriptive frame}\}.$$

That is what Fine [1974c] actually did for logics of finite width.

<sup>6</sup>Here  $md(\varphi)$ , the *modal degree* of  $\varphi$ , is the length of the longest chain of nested modal operators in  $\varphi$ .

**THEOREM 24** (Fine 1974c). *All logics of finite width are  $\mathcal{RE}$ -persistent and so Kripke complete.*

Let us return, however, to the method of canonical models. Having tried it for a number of standard systems, Lemmon and Scott [1977] found a rather general sufficient condition for its applicability and put forward a conjecture concerning a further extension (which was proved by Goldblatt [1976b]). This direction of completeness (and correspondence) theory culminated in the theorem of Sahlqvist [1975] who proved an optimal (in a sense) generalization of the condition of [Lemmon and Scott 1977]. To formulate it we require the following definition. Say that a formula is *positive* (*negative*) if it is constructed from variables (negated variables) and the constants  $\top$ ,  $\perp$  using  $\wedge$ ,  $\vee$ ,  $\diamond$  and  $\square$ .

**THEOREM 25** (Sahlqvist 1975). *Suppose  $\varphi$  is a formula which is equivalent in  $\mathbf{K}$  to a formula of the form  $\square^k(\psi \rightarrow \chi)$ , where  $k \geq 0$ ,  $\chi$  is positive and  $\psi$  is constructed from variables and their negations,  $\perp$  and  $\top$  with the help of  $\wedge$ ,  $\vee$ ,  $\square$  and  $\diamond$  in such a way that no  $\psi$ 's subformula of the form  $\psi_1 \vee \psi_2$  or  $\diamond\psi_1$ , containing an occurrence of a variable without  $\neg$ , is in the scope of some  $\square$ . Then one can effectively construct a first order formula  $\phi(x)$  in  $R$  and  $=$  having  $x$  as its only free variable and such that, for every descriptive or Kripke frame  $\mathfrak{F}$  and every point  $a$  in  $\mathfrak{F}$ ,*

$$(\mathfrak{F}, a) \models \varphi \text{ iff } \mathfrak{F} \models \phi(x)[a].$$

(Here  $(\mathfrak{F}, a) \models \varphi$  means that  $\varphi$  is true at  $a$  in  $\mathfrak{F}$  under any valuation.)

**Proof.** We present a sketch of the proof found by Sambin and Vaccaro [1989]. Given a formula  $\varphi(p_1, \dots, p_n)$ , a frame  $\mathfrak{F} = \langle W, R, P \rangle$  and sets  $X_1, \dots, X_n \in P$ , denote by  $\varphi(X_1, \dots, X_n)$  the set of points in  $\mathfrak{F}$  at which  $\varphi$  is true under the valuation  $\mathfrak{V}$  defined by  $\mathfrak{V}(p_i) = X_i$ , i.e.,  $\varphi(X_1, \dots, X_n) = \mathfrak{V}(\varphi)$ . Using this notation, we can say that

$$(\mathfrak{F}, x) \models \varphi(p_1, \dots, p_n) \text{ iff } \forall X_1, \dots, X_n \in P \ x \in \varphi(X_1, \dots, X_n).$$

**EXAMPLE 26.** Let us consider the formula  $\square p \rightarrow p$  and try to extract a first order equivalent for it in the class of tight frames directly from the equivalence above and the condition of tightness. For every tight frame  $\mathfrak{F} = \langle W, R, P \rangle$  we have:

$$\begin{aligned} (\mathfrak{F}, x) \models \square p \rightarrow p & \text{ iff } \forall X \in P \ x \in (\square X \rightarrow X) \\ & \text{ iff } \forall X \in P \ (x \in \square X \rightarrow x \in X) \\ & \text{ iff } \forall X \in P \ (x \uparrow \subseteq X \rightarrow x \in X). \end{aligned}$$

To eliminate the variable  $X$  ranging over  $P$ , we can use two simple observations. The first one is purely set-theoretic:

$$(3) \quad \forall X \in P (Y \subseteq X \rightarrow x \in X) \text{ iff } x \in \bigcap \{X \in P : Y \subseteq X\}.$$

And the second one is just a reformulation of the characteristic property of tight frames:

$$(4) \quad \bigcap \{X \in P : x\uparrow \subseteq X\} = x\uparrow.$$

With the help of (3) and (4) we can continue the chain of equivalences above with two more lines:

$$\begin{aligned} (\mathfrak{F}, x) \models \Box p \rightarrow p & \text{ iff } \dots \\ & \text{ iff } x \in \bigcap \{X \in P : x\uparrow \subseteq X\} \\ & \text{ iff } x \in x\uparrow. \end{aligned}$$

Thus,  $\mathfrak{F} \models \Box p \rightarrow p$  iff  $\forall x \ x \in x\uparrow$  iff  $\forall x \ xRx$ .

The proof of Sahlqvist's Theorem is a (by no means trivial) generalization of this argument. Define by induction  $x\uparrow^0 = \{x\}$ ,  $x\uparrow^{n+1} = (x\uparrow^n)\uparrow$ , and notice that in (4) we can replace  $x\uparrow$  by any term of the form  $x_1\uparrow^{n_1} \cup \dots \cup x_k\uparrow^{n_k}$ , thus obtaining the equality

$$(5) \quad \bigcap \{X \in P : x_1\uparrow^{n_1} \cup \dots \cup x_k\uparrow^{n_k} \subseteq X\} = x_1\uparrow^{n_1} \cup \dots \cup x_k\uparrow^{n_k}$$

which holds for every descriptive frame  $\mathfrak{F} = \langle W, R, P \rangle$ , all  $x_1, \dots, x_k \in W$  and all  $n_1, \dots, n_k \geq 0$ .

A frame-theoretic term  $x_1\uparrow^{n_1} \cup \dots \cup x_k\uparrow^{n_k}$  with (not necessarily distinct) world variables  $x_1, \dots, x_k$  will be called an *R-term*. It is not hard to see that for any *R-term*  $T$ , the relation  $x \in T$  on  $\mathfrak{F} = \langle W, R, P \rangle$  is first order expressible in  $R$  and  $=$ . Consequently, we obtain

**LEMMA 27.** *Suppose  $\varphi(p_1, \dots, p_n)$  is a modal formula and  $T_1, \dots, T_n$  are *R-terms*. Then the relation  $x \in \varphi(T_1, \dots, T_n)$  is expressible by a first order formula (in  $R$  and  $=$ ) having  $x$  as its only free variable.*

Syntactically, *R-terms* with a single world variable correspond to modal formulas of the form  $\Box^{m_1} p_1 \wedge \dots \wedge \Box^{m_k} p_k$  with not necessarily distinct propositional variables  $p_1, \dots, p_k$ . Such formulas are called *strongly positive*. By induction on the construction of  $\varphi$ , one can prove the following

**LEMMA 28.** *Suppose  $\varphi(p_1, \dots, p_n)$  is a strongly positive formula containing all the variables  $p_1, \dots, p_n$  and  $\mathfrak{F} = \langle W, R, P \rangle$  is a frame. Then one can effectively construct *R-terms*  $T_1, \dots, T_n$  (with one variable  $x$ ) such that for any  $x \in W$  and any  $X_1, \dots, X_n \in P$ ,*

$$x \in \varphi(X_1, \dots, X_n) \text{ iff } T_1 \subseteq X_1 \wedge \dots \wedge T_n \subseteq X_n.$$

Now, trying to extend the method of Example 26 to a wider class of formulas, we see that it still works if we replace the antecedent  $\Box p$  in  $\Box p \rightarrow p$



with an arbitrary strongly positive formula  $\psi$ . As to generalizations of the consequent, let us take first an arbitrary formula  $\chi$  instead of  $p$  and see what properties it should satisfy to be handled by our method.

Thus, for a modal formula  $(\psi \rightarrow \chi)(p_1, \dots, p_n)$  with strongly positive  $\psi$  and a descriptive frame  $\mathfrak{F} = \langle W, R, P \rangle$ , we have:

$$\begin{aligned} (\mathfrak{F}, x) \models \psi \rightarrow \chi &\text{ iff } \forall X_1, \dots, X_n \in P \ (x \in \psi(X_1, \dots, X_n) \rightarrow \\ &\quad x \in \chi(X_1, \dots, X_n)) \\ &\text{ iff } \forall X_1, \dots, X_n \in P \ (T_1 \subseteq X_1 \wedge \dots \wedge T_n \subseteq X_n \rightarrow \\ &\quad x \in \chi(X_1, \dots, X_n)) \\ &\text{ iff } \forall X_1, \dots, X_{n-1} \in P \ (T_1 \subseteq X_1 \wedge \dots \wedge T_{n-1} \subseteq X_{n-1} \rightarrow \\ &\quad \forall X_n \in P \ (T_n \subseteq X_n \rightarrow x \in \chi(X_1, \dots, X_n))). \end{aligned}$$

(3) does not help us here, but we can readily generalize it to

$$\begin{aligned} \forall X \in P \ (Y \subseteq X \rightarrow x \in \chi(\dots, X, \dots)) &\text{ iff} \\ (6) \quad x \in \bigcap \{ \chi(\dots, X, \dots) : Y \subseteq X \in P \}. \end{aligned}$$

So

$$\begin{aligned} (\mathfrak{F}, x) \models \psi \rightarrow \chi &\text{ iff } \forall X_1, \dots, X_{n-1} \in P \ (T_1 \subseteq X_1 \wedge \dots \wedge T_{n-1} \subseteq X_{n-1} \rightarrow \\ &\quad x \in \bigcap \{ \chi(X_1, \dots, X_n) : T_n \subseteq X_n \in P \}). \end{aligned}$$

But now (4) and (5) are useless. In fact, what we need is the equality

$$\begin{aligned} \bigcap \{ \chi(\dots, X, \dots) : T \subseteq X \in P \} = \\ (7) \quad \chi(\dots, \bigcap \{ X \in P : T \subseteq X \}, \dots) \end{aligned}$$

which, with the help of (5), would give us

$$(8) \quad \bigcap \{ \chi(\dots, X, \dots) : T \subseteq X \in P \} = \chi(\dots, T, \dots).$$

Of course, (7) is too good to hold for an arbitrary  $\chi$ , but suppose for a moment that our  $\chi$  satisfies it. Then we can eliminate step by step all the variables  $X_1, \dots, X_n$  like this:

$$\begin{aligned} (\mathfrak{F}, x) \models \psi \rightarrow \chi &\text{ iff } \forall X_1, \dots, X_{n-1} \in P \ (T_1 \subseteq X_1 \wedge \dots \wedge T_{n-1} \subseteq X_{n-1} \rightarrow \\ &\quad x \in \chi(X_1, \dots, X_{n-1}, T_n)) \\ &\text{ iff } \dots \text{ (by the same argument)} \\ &\text{ iff } x \in \chi(T_1, \dots, T_n). \end{aligned}$$

And the last relation can be effectively rewritten in the form of a first order formula  $\phi(x)$  in  $R$  and  $=$  having  $x$  as its only free variable. So, finally we shall have  $\mathfrak{F} \models \psi \rightarrow \chi$  iff  $\forall x \phi(x)$ .

Now, to satisfy (7),  $\chi$  should have the property that all its operators distribute over intersections. Clearly,  $\rightarrow$  and  $\neg$  are not suitable for this goal. But all the other operators turn out to be good enough at least in descriptive and Kripke frames. So we can take as  $\chi$  any positive modal formula. The main property of a positive formula  $\varphi(\dots, p, \dots)$  is its *monotonicity* in every variable  $p$  which means that, for all sets  $X, Y$  of worlds in a frame,  $X \subseteq Y$  implies  $\varphi(\dots, X, \dots) \subseteq \varphi(\dots, Y, \dots)$ .

To prove that all positive formulas satisfy (7) in Kripke frames and descriptive frames, recall that  $\Box$  distributes over arbitrary intersections in any frame. As to  $\Diamond$ , we have the following lemma in which a family  $\mathcal{X}$  of non-empty subsets of some space  $W$  is called *downward directed* if for all  $X, Y \in \mathcal{X}$  there is  $Z \in \mathcal{X}$  such that  $Z \subseteq X \cap Y$ .

LEMMA 29 (Esakia 1974). *Suppose  $\mathfrak{F} = \langle W, R, P \rangle$  is a descriptive frame. Then for every downward directed family  $\mathcal{X} \subseteq P$ ,*

$$\Diamond \bigcap_{X \in \mathcal{X}} X = \bigcap_{X \in \mathcal{X}} \Diamond X.$$

Using Esakia's Lemma, by induction on the construction of  $\varphi$  one can prove

LEMMA 30. *Suppose that  $\mathfrak{F} = \langle W, R, P \rangle$  is a Kripke or descriptive frame and  $\varphi(p, \dots, q, \dots, r)$  is a positive formula. Then for every  $Y \subseteq W$  and all  $U, \dots, V \in P$ ,*

$$(9) \quad \bigcap \{ \varphi(U, \dots, X, \dots, V) : Y \subseteq X \in P \} = \varphi(U, \dots, \bigcap \{ X \in P : Y \subseteq X \}, \dots, V).$$

It follows from this lemma and considerations above that Sahlqvist's Theorem holds for formulas  $\varphi = \psi \rightarrow \chi$  with strongly positive  $\psi$  and positive  $\chi$ . The remaining part of the proof is purely syntactic manipulations with modal and first order formulas.

Notice that using the monotonicity of positive formulas, equivalence (6) can be generalized to the following one: for every  $\mathfrak{F} = \langle W, R, P \rangle$ , every positive  $\chi_i(\dots, p, \dots)$  and every  $x_i \in W$ ,

$$(10) \quad \forall X \in P (Y \subseteq X \rightarrow \bigvee_{i \leq n} x_i \in \chi_i(\dots, X, \dots)) \text{ iff } \bigvee_{i \leq n} x_i \in \bigcap \{ \chi_i(\dots, X, \dots) : Y \subseteq X \in P \}.$$

Say that a modal formula  $\psi$  is *untied* if it can be constructed from negative formulas and strongly positive ones using only  $\wedge$  and  $\Diamond$ . If  $\nu(p_1, \dots, p_n)$  is negative then  $\neg\nu(p_1, \dots, p_n)$  is clearly equivalent in  $\mathbf{K}$  to a positive formula; we denote it by  $\nu^*(\neg p_1, \dots, \neg p_n)$ .

LEMMA 31. *Let  $\psi(p_1, \dots, p_n)$  be an untied formula and  $\mathfrak{F} = \langle W, R, P \rangle$  a frame. Then for every  $x \in W$  and all  $X_1, \dots, X_n \in P$ ,*

$$x \in \psi(X_1, \dots, X_n) \text{ iff } \exists y_1, \dots, y_l (\vartheta \wedge \bigwedge_{i \leq n} T_i \subseteq X_i \wedge \bigwedge_{j \leq m} z_j \in \nu_j(X_1, \dots, X_n))$$

where the formula in the right-hand side, effectively constructed from  $\psi$ , has only one free individual variable  $x$ ,  $\vartheta$  is a conjunction of formulas of the form  $uRv$ ,  $T_i$  are suitable  $R$ -terms and  $\nu_j(p_1, \dots, p_n)$  are negative formulas.

We are ready now to prove Sahlqvist's Theorem. To construct a first order equivalent for  $\Box^k(\psi \rightarrow \chi)$  supplied by the formulation of our theorem, we observe first that one can equivalently reduce  $\psi$  to a disjunction  $\psi_1 \vee \dots \vee \psi_m$  of untied formulas, and hence  $\Box^k(\psi \rightarrow \chi)$  is equivalent in  $\mathbf{K}$  to the formula

$$\Box^k(\psi_1 \rightarrow \chi) \wedge \dots \wedge \Box^k(\psi_m \rightarrow \chi).$$

So all we need is to find a first order equivalent for an arbitrary formula  $\Box^k(\psi \rightarrow \chi)$  with untied  $\psi$  and positive  $\chi$ . Let  $p_1, \dots, p_n$  be all the variables in  $\psi$  and  $\chi$  and  $\mathfrak{F} = \langle W, R, P \rangle$  a descriptive or Kripke frame. Then, for any  $x \in W$ , we have:

$$\begin{aligned} (\mathfrak{F}, x) \models \Box^k(\psi \rightarrow \chi) &\text{ iff } \forall X_1, \dots, X_n \in P \ x \in \Box^k(\psi \rightarrow \chi)(X_1, \dots, X_n) \\ &\text{ (by Lemma 31) iff } \forall X_1, \dots, X_n \in P \ \forall y \ (xR^k y \rightarrow (\exists y_1, \dots, y_l \ (\vartheta \wedge \\ &\quad \bigwedge_{i \leq n} T_i \subseteq X_i \wedge \bigwedge_{j \leq m} z_j \in \nu_j(X_1, \dots, X_n)) \rightarrow \\ &\quad y \in \chi(X_1, \dots, X_n))) \\ &\text{ iff } \forall X_1, \dots, X_n \in P \ \forall y, y_1, \dots, y_l \ (\vartheta' \wedge \bigwedge_{i \leq n} T_i \subseteq X_i \wedge \\ &\quad \bigwedge_{j \leq m} z_j \in \nu_j(X_1, \dots, X_n) \rightarrow y \in \chi(X_1, \dots, X_n)) \end{aligned}$$

where  $\vartheta' = xR^k y \wedge \vartheta$ . Let  $\pi_j(p_1, \dots, p_n) = \nu_j^*(\neg p_1, \dots, \neg p_n)$ . We continue this chain of equivalences as follows:

$$\begin{aligned} \text{iff } \forall y, y_1, \dots, y_l \ (\vartheta' \rightarrow \forall X_1, \dots, X_n \in P \ (\bigwedge_{i \leq n} T_i \subseteq X_i \rightarrow \\ \bigvee_{j \leq m+1} z_j \in \pi_j(X_1, \dots, X_n))) \end{aligned}$$

(where  $\pi_{m+1}(p_1, \dots, p_n) = \chi(p_1, \dots, p_n)$  and  $z_{m+1} = y$ )

$$\text{iff } \forall y, y_1, \dots, y_l \ (\vartheta' \rightarrow \bigvee_{j \leq m+1} z_j \in \pi_j(T_1, \dots, T_n)),$$

as follows from (10), Lemma 30 and equality (5). It remains to use Lemma 27. ■

The formulas  $\varphi$  defined in the formulation of Theorem 25 are called *Sahlqvist formulas*. It follows from this theorem that if  $L$  is a  $\mathcal{D}$ -persistent logic and  $\Gamma$  a set of Sahlqvist formulas then  $L \oplus \Gamma$  is also  $\mathcal{D}$ -persistent. Moreover,  $L \oplus \Gamma$  is *elementary* (in the sense that the class of Kripke frames for it coincides with the class of all models for some set of first order formulas in  $R$  and  $=$ ) whenever  $L$  is so.

Other proofs of Sahlqvist's Theorem were found by Kracht [1993] and Jónsson [1994] (the latter is based upon the algebraic technique developed in [Jónsson and Tarski 1951]). Venema [1991] extended Sahlqvist's Theorem to logics with non-standard inference rules, like Gabbay's [1981a] irreflexivity rule. In [Chagrova and Zakharyashev 1995b] it is shown that there is a continuum of Sahlqvist logics above **S4** and that not all of them have the finite model property (above **T** such a logic was constructed by Hughes and Cresswell [1984]). As we shall see later in this chapter, there are even undecidable finitely axiomatizable Sahlqvist logics in **NExtK**. It would be of interest to find out whether such logics exist above **K4** or **S4**.

Kracht [1993] described syntactically the set of first order equivalents of Sahlqvist formulas. To formulate his criterion we require the fragment  $\mathcal{S}$  of first order logic defined inductively as follows. Formulas of the form  $xR^m y$  are in  $\mathcal{S}$  for all variables  $x, y$  and every  $m < \omega$ ; besides, if  $\phi, \phi'$  are in  $\mathcal{S}$  then the formulas

$$\forall x \in y \uparrow^m \phi, \exists x \in y \uparrow^m \phi, \phi \wedge \phi', \text{ and } \phi \vee \phi'$$

are also in  $\mathcal{S}$ . For simplicity we assume that all occurrences of quantifiers in a formula bind pairwise distinct variables. Call a variable  $y$  in a formula  $\phi \in \mathcal{S}$  *inherently universal* if either all occurrences of  $y$  are free in  $\phi$  or  $\phi$  contains a subformula  $\forall y \in x \uparrow^m \phi'$  which is not in the scope of  $\exists$ .

**THEOREM 32** (Kracht 1993). *For every first order formula  $\phi(x)$  (in  $R$  and  $=$ ) with one free variable  $x$ , the following conditions are equivalent:*

- (i)  $\phi(x)$  is classically equivalent to a formula  $\phi'(x) \in \mathcal{S}$  such that any subformula of the form  $yR^m z$  of  $\phi'(x)$  contains at least one inherently universal variable;
- (ii)  $\phi(x)$  corresponds to a Sahlqvist formula in the sense of Theorem 25.

Condition (i) is satisfied, for example, by the formula

$$\forall u \in x \uparrow \forall v \in x \uparrow \exists z \in u \uparrow vRz$$

which corresponds to  $\diamond \Box p \rightarrow \Box \diamond p$ . On the other hand,

$$\phi(x) = \exists y \in x \uparrow \forall z \in y \uparrow zR^0 y$$

does not satisfy (i). In fact, even relative to **S4** the condition expressed by  $\phi(x)$  does not correspond to any Sahlqvist formula. Notice, however, that

$\mathbf{S4} \oplus \Box\Diamond p \rightarrow \Diamond\Box p$  is a  $\mathcal{D}$ -persistent logic whose frames are precisely the transitive and reflexive frames validating  $\forall x\phi(x)$ .

We conclude this section by mentioning two more important results connecting persistence and elementarity (the idea of the proof was discussed in Section 22 of *Basic Modal Logic*.)

**THEOREM 33.**

- (i) (Fine 1975b, van Benthem 1980) *If a logic  $L$  is characterized by a first order definable class of Kripke frames then  $L$  is  $\mathcal{D}$ -persistent.*
- (ii) (Fine 1975b) *If  $L$  is  $\mathcal{R}$ -persistent then the class of Kripke frames for  $L$  is first order definable.*

It is an open problem whether every  $\mathcal{D}$ -persistent logic is determined by a first order definable class of Kripke frames; for more information about this and related problems consult [Goldblatt 1995].

#### 1.4 The degree of Kripke incompleteness

All known logics in  $\mathbf{NExtK}$  of “natural origin” are complete with respect to Kripke semantics. On the other hand, there are many examples of “artificial” logics that cannot be characterized by any class of Kripke frames (see Sections 19, 20 of *Basic Modal Logic* or the examples below). To understand the phenomenon of Kripke incompleteness Fine [1974b] proposed to investigate how many logics may share the same Kripke frames with a given logic  $L$ . The number of them is called the *degree of Kripke incompleteness* of  $L$ . Of course, this number depends on the lattice of logics under consideration. The degree of Kripke incompleteness of logics in  $\mathbf{NExtK}$  was comprehensively studied by Blok [1978]. In this section we present the main results of that paper following [Chagrov and Zakharyashev 1997].

By Theorem 12, all Kripke complete union-splittings of  $\mathbf{NExtK}$  have degree of incompleteness 1. And it turns out that no other union-splitting exists.

**THEOREM 34** (Blok 1978). *Every union-splitting of  $\mathbf{NExtK}$  has the finite model property.*

**Proof.** Let  $\mathcal{F}$  be a class of finite rooted cycle free frames. We prove that  $L = \mathbf{K}/\mathcal{F}$  has the finite model property using a variant of filtration, which is applied to an  $n$ -generated refined frame  $\mathfrak{F} = \langle W, R, P \rangle$  for  $L$  refuting a formula  $\varphi(p_1, \dots, p_n)$  under a valuation  $\mathfrak{V}$ .

Since  $\mathfrak{F}$  is differentiated, for every  $m \geq 1$  there are only finitely many points  $x$  in  $\mathfrak{F}$  such that  $x \models \Box^m \perp \wedge \neg \Box^{m-1} \perp$ ; we shall call them *points of type  $m$* . Given  $\Delta \subseteq \mathbf{Sub}\varphi$ ,  $\mathbf{Sub}\varphi$  the set of all subformulas in  $\varphi$ , we put  $m_\Delta = m$  if  $m$  is the minimal number such that a point in  $\mathfrak{F}$  is of type  $\leq m$

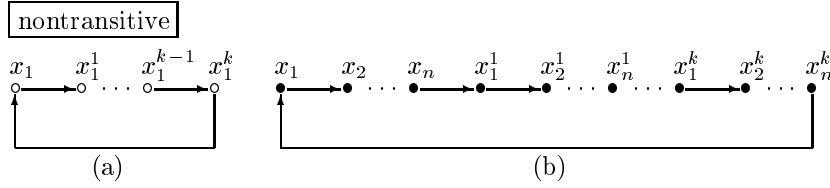


Figure 3.

whenever  $x \models \Delta$  and the formulas in  $\mathbf{Sub}\varphi - \Delta$  are false at  $x$  (under  $\mathfrak{V}$ ); if no such  $m$  exists, we put  $m_\Delta = 0$ . Let

$$k = \max\{m_\Delta : \Delta \subseteq \mathbf{Sub}\varphi\}, \quad \Gamma = \mathbf{Sub}(\varphi \wedge \Box^k \perp).$$

Now we divide  $\mathfrak{F}$  into two parts:  $W_1$  consisting of points of type  $\leq k$  and  $W_2 = W - W_1$ . For  $x, y \in W$ , put  $x \sim y$  if either  $x, y \in W_1$  and  $x = y$  or  $x, y \in W_2$  and exactly the same formulas in  $\Gamma$  are true at  $x$  and  $y$ . Let  $\mathfrak{N} = \langle \mathfrak{G}, \mathfrak{U} \rangle$  be the smallest filtration (see Section 12 of *Basic Modal Logic*) of  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  through  $\Gamma$  with respect to  $\sim$ . Since  $W_1$  is finite,  $\mathfrak{G}$  is also finite and, by the Filtration Theorem,  $(\mathfrak{M}, x) \models \psi$  iff  $(\mathfrak{N}, [x]) \models \psi$ , for every  $\psi \in \Gamma$ . So it remains to show that  $\mathfrak{G} \models L$ . Notice that  $[x]$  in  $\mathfrak{G}$  is of type  $m \leq k$  iff  $x$  has type  $m$  in  $\mathfrak{F}$ . Moreover, there is no  $[x]$  of type  $l > k$ . For otherwise  $x \not\models \Box^k \perp$  and  $m_\Delta = 0$  for  $\Delta = \{\psi \in \mathbf{Sub}\varphi : x \models \psi\}$ , which means that arbitrary long chains (of not necessarily distinct points) start from  $[x]$ , contrary to  $[x]$  being of type  $l$ . Thus  $\mathfrak{G}$  consists of two parts: points of type  $\leq k$ , which form the generated subframe  $\langle W_1, R \upharpoonright W_1 \rangle$  of  $\mathfrak{F}$ , and points involved in cycles. Since  $\mathfrak{F} \models L$  and frames in  $\mathcal{F}$  are cycle free, it follows from Lemma 13 and Theorem 17 that  $\mathfrak{G} \models L$ . ■

**THEOREM 35** (Blok 1978). *If a logic  $L$  is inconsistent or a union-splitting of  $\mathbf{NExtK}$ , then  $L$  is strictly Kripke complete. Otherwise  $L$  has degree of Kripke incompleteness  $2^{\aleph_0}$  in  $\mathbf{NExtK}$ .*

**Proof.** That **For** is strictly complete follows from Example 10 and Theorem 12. Suppose now that a consistent  $L$  is not a union-splitting and  $L'$  is the greatest union-splitting contained in  $L$ . Since  $L'$  has the finite model property, there is a finite rooted frame  $\mathfrak{F} = \langle W, R \rangle$  for  $L'$  refuting some  $\varphi \in L$  and such that every proper generated subframe of  $\mathfrak{F}$  validates  $L$ . Clearly,  $\mathfrak{F}$  is not cycle free. Let  $x_1 R x_2 R \dots R x_n R x_1$  be the shortest cycle in  $\mathfrak{F}$  and  $k = md(\varphi) + 1$ . We construct a new frame  $\mathfrak{F}'$  by extending the cycle  $x_1, \dots, x_n, x_1$  as is shown in Fig. 3 ((a) for  $n = 1$  and (b) for  $n > 1$ ). More precisely, we add to  $\mathfrak{F}$  copies  $x_1^1, \dots, x_n^k$  of  $x_i$  for each  $i \in \{1, \dots, n\}$ , organize them into the nontransitive cycle shown in Fig. 3 and draw an arrow from  $x_i^j$  to  $y \in W - \{x_1, \dots, x_n\}$  iff  $x_i R y$ . Denote the resulting frame

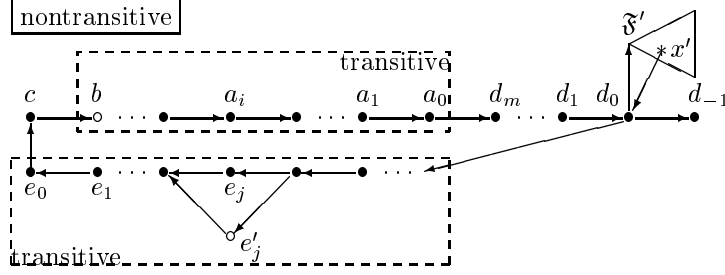


Figure 4.

by  $\mathfrak{F}' = \langle W', R' \rangle$  and let  $x' = x_n^k$ . By the construction,  $\mathfrak{F}$  is a reduct of  $\mathfrak{F}'$ . Therefore, for every models  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  and  $\mathfrak{M}' = \langle \mathfrak{F}', \mathfrak{V}' \rangle$  such that

$$\mathfrak{V}'(p) = \mathfrak{V}(p) \cup \{x_i^j : x_i \in \mathfrak{V}(p), j < k\}$$

and for every  $x \in W$ ,  $\psi \in \mathbf{Sub}\varphi$ , we have  $(\mathfrak{M}, x) \models \psi$  iff  $(\mathfrak{M}', x) \models \psi$ . So we can hook some other model on  $x'$ , and points in  $W$  will not feel its presence by means of  $\varphi$ 's subformulas. The frame to be hooked on  $x'$  depends on whether  $\bullet \models L$  or  $\circ \models L$ . We consider only the former alternative.

Fix some  $m > |W'|$ . For each  $I \subseteq \omega - \{0\}$ , let  $\mathfrak{F}_I = \langle W_I, R_I, P_I \rangle$  be the frame whose diagram is shown in Fig. 4 ( $d_0$  sees the root of  $\mathfrak{F}'$ , all points  $e_i$  and  $e'_j$  and is seen from  $x'$ ; the subframes in dashed boxes are transitive,  $e'_i \in W_I$  iff  $i \in I$ , and  $P_I$  consists of sets of the form  $X \cup Y$  such that  $X$  is a finite or cofinite subset of  $W_I - \{b, a_i : i < \omega\}$  and  $Y$  is either a finite subset of  $\{a_i : i < \omega\}$  or is of the form  $\{b\} \cup Y'$ , where  $Y'$  is a cofinite subset of  $\{a_i : i < \omega\}$ . It is not hard to see that the points  $a_i$ ,  $c$ ,  $e_i$  and  $e'_i$  are characterized by the variable free formulas

$$\alpha_0 = \diamond(\delta_m \wedge \diamond(\delta_{m-1} \wedge \dots \wedge \diamond\delta_0) \dots) \wedge \neg\delta^2(\delta_m \wedge \diamond(\delta_{m-1} \wedge \dots \wedge \diamond\delta_0) \dots),$$

$$\alpha_{i+1} = \diamond\alpha_i \wedge \neg\delta^2\alpha_i, \quad \gamma = \diamond^2\alpha_0 \wedge \neg\delta\alpha_0,$$

$$\epsilon_0 = \diamond\gamma, \quad \epsilon_{i+1} = \diamond\epsilon_i \wedge \neg\delta^2\epsilon_i, \quad \epsilon'_{i+1} = \diamond\epsilon_i \wedge \neg\delta^+\epsilon_{i+1},$$

(in the sense that  $x \models \alpha_i$  iff  $x = a_i$ , etc.), where

$$\delta_0 = \diamond\Box\perp, \quad \delta_1 = \diamond\delta_0 \wedge \neg\delta_0, \quad \delta_2 = \diamond\delta_1 \wedge \neg\delta_1 \wedge \neg\delta^+\delta_0,$$

$$\delta_{k+1} = \diamond\delta_k \wedge \neg\delta_k \wedge \neg\delta^+\delta_{k-1} \wedge \dots \wedge \neg\delta^+\delta_0.$$

Define  $L_I$  to be the logic determined by the class of frames for  $L$  and  $\mathfrak{F}_I$ , i.e.,  $L_I = L \cap \text{Log}\mathfrak{F}_I$ . Since  $\neg(\epsilon'_i \wedge \diamond^{m+6}\neg\varphi) \in L_J - L_I$  for  $i \in I - J$  ( $\varphi$  is refuted at the root of  $\mathfrak{F}'$ ),  $|\{L_I : I \subseteq \omega - \{0\}\}| = 2^{\aleph_0}$ .

Let us show now that  $L_I$  has the same Kripke frames as  $L$ . Since  $L_I \subseteq L$ , we must prove that every Kripke frame for  $L_I$  validates  $L$ . Suppose there is a rooted Kripke frame  $\mathfrak{G}$  such that  $\mathfrak{G} \models L_I$  but  $\mathfrak{G} \not\models \psi$ , for some  $\psi \in L$ . Since  $\psi$  is in  $L$ , it is valid in all frames for  $L$ , in particular,  $\bullet \models \psi$ . And since  $\psi \notin L_I$ ,  $\psi$  is refuted in  $\mathfrak{F}_I$ . Moreover, by the construction of  $\mathfrak{F}_I$ , it is refuted at a point from which the root of  $\mathfrak{F}'$  can be reached by a finite number of steps. Therefore, the following formulas are valid in  $\mathfrak{F}_I$  and so belong to  $L_I$  and are valid in  $\mathfrak{G}$ :

$$(11) \quad \neg\psi \rightarrow \bigvee_{i=0}^l \diamond^i \gamma,$$

$$(12) \quad \neg\psi \rightarrow \bigwedge_{i=0}^l \square^i (\gamma \rightarrow \square(\square_0(\square_0 p \rightarrow p) \rightarrow p)),$$

where  $p$  does not occur in  $\psi$  and  $l$  is a sufficiently big number so that any point in  $\mathfrak{F}_I$  is accessible by  $\leq l$  steps from every point in the selected cycle and every point at which  $\psi$  may be false, and  $\square_0 \chi = \square(\diamond \alpha_0 \rightarrow \chi)$ . According to (11),  $\mathfrak{G}$  contains a point at which  $\gamma$  is true. By the construction of  $\gamma$ , this point has a successor  $y$  at which, by (12),  $\square_0(\square_0 p \rightarrow p) \rightarrow p$  is true *under any valuation* in  $\mathfrak{G}$  and  $y \models \diamond \alpha_0$ . Define a valuation  $\mathfrak{U}$  in  $\mathfrak{G}$  by taking  $\mathfrak{U}(p) = y \uparrow$ . Then  $y \models \square_0(\square_0 p \rightarrow p)$ , from which  $y \models p$  and so  $y \in y \uparrow$ . Now define another valuation  $\mathfrak{U}'$  so that  $\mathfrak{U}'(p) = y \uparrow - \{y\}$ . Since  $y$  is reflexive, we again have  $y \models \square_0(\square_0 p \rightarrow p)$ , whence  $y \models p$ , which is a contradiction. ■

This construction can be used to obtain one more important result.

**THEOREM 36** (Blok 1978). *Every union-splitting  $\mathbf{K}/\mathcal{F}$  has  $\varkappa \leq \aleph_0$  immediate predecessors in  $\text{NExt}\mathbf{K}$ , where  $\varkappa$  is the number of frames in  $\mathcal{F}$  which are not reducts of generated subframes of other frames in  $\mathcal{F}$ . Every consistent logic different from union-splittings has  $2^{\aleph_0}$  immediate predecessors in  $\text{NExt}\mathbf{K}$ . (**For** has 2 immediate predecessors in  $\text{NExt}\mathbf{K}$ .)*

**Proof.** The former claim follows from Theorem 12. To establish the latter, we continue the proof of Theorem 35. One can show that  $L$  is finitely axiomatizable over  $L_I$  (the proof is rather technical, and we omit it here). Then, by Zorn's Lemma,  $\text{NExt}L_I$  contains an immediate predecessor  $L'_I$  of  $L$ . Besides,  $L_I \oplus L_J = L$  whenever  $I \neq J$ . Indeed,

$$L_I \oplus L_J = (L \cap \text{Log}\mathfrak{F}_I) \oplus (L \cap \text{Log}\mathfrak{F}_J) = L \cap (\text{Log}\mathfrak{F}_I \oplus \text{Log}\mathfrak{F}_J)$$

and if  $i \in I - J$  then, for every  $\chi \in L$  and a sufficiently big  $l$ ,

$$\neg \bigvee_{k=0}^l \diamond^k \epsilon'_i \rightarrow \chi \in \text{Log}\mathfrak{F}_I, \quad \neg \epsilon'_i \in \text{Log}\mathfrak{F}_J,$$



from which  $\chi \in \text{Log}\mathfrak{F}_I \oplus \text{Log}\mathfrak{F}_J$  and so  $L \subseteq \text{Log}\mathfrak{F}_I \oplus \text{Log}\mathfrak{F}_J$ . It follows that  $L'_I \neq L'_J$  whenever  $I \neq J$ . ■

It is worth noting that tabular logics, proper extensions of **D** and extensions of **K4** are not union-splittings in  $\text{NExt}\mathbf{K}$ . Similar results hold for the lattices  $\text{NExt}\mathbf{D}$  and  $\text{NExt}\mathbf{T}$ , where every consistent logic has degree of incompleteness  $2^{\aleph_0}$  (see [Blok 1978, 1980b]). It would be of interest to describe the behavior of this function in  $\text{NExt}\mathbf{K4}$ ,  $\text{NExt}\mathbf{GL}$ ,  $\text{NExt}\mathbf{S4}$ ,  $\text{NExt}\mathbf{Grz}$  (where Theorem 34 does not hold and where every tabular logic has finitely many immediate predecessors) and other lattices of logics to be considered later in this chapter.

### 1.5 Stronger forms of Kripke completeness

In the two preceding sections we were considering the problem of characterizing logics  $L \in \text{NExt}\mathbf{K}$  by classes of Kripke frames. The same problem arises in connection with the two consequence relations  $\vdash_L$  and  $\vdash_L^*$  as well. Theorem 19 shows a way of introducing the corresponding concepts of completeness.

With each Kripke frame  $\mathfrak{F}$  let us associate a consequence relation  $\models_{\mathfrak{F}}$  by putting, for any formula  $\varphi$  and any set  $\Gamma$  of formulas,  $\Gamma \models_{\mathfrak{F}} \varphi$  iff  $(\mathfrak{M}, x) \models \Gamma$  implies  $(\mathfrak{M}, x) \models \varphi$  for every model  $\mathfrak{M}$  based on  $\mathfrak{F}$  and every point  $x$  in  $\mathfrak{F}$ . Clearly, a modal logic  $L$  is Kripke complete iff, for any *finite* set of formulas  $\Gamma$  and any formula  $\varphi$ ,  $\Gamma \not\vdash_L \varphi$  only if there is a Kripke frame  $\mathfrak{F}$  for  $L$  such that  $\Gamma \not\models_{\mathfrak{F}} \varphi$ . Now, let us call  $L$  *strongly Kripke complete*<sup>7</sup> if this implication holds for arbitrary sets  $\Gamma$ . In other words,  $L$  is strongly complete if every  $L$ -consistent set of formulas holds at some point in a model based on a Kripke frame for  $L$ . Another reformulation:  $L$  is strongly complete iff  $L$  is Kripke complete and the relation  $\bigcap\{\models_{\mathfrak{F}} : \mathfrak{F} \text{ is a Kripke frame for } L\}$  is finitary. It follows from the construction of the canonical models that every canonical (in particular,  $\mathcal{D}$ -persistent) logic is strongly complete, which provides us with many examples of such logics in  $\text{NExt}\mathbf{K}$ .

By Theorem 33, all logics characterized by first order definable classes of Kripke frames are strongly complete. The converse does not hold: there exist strongly complete logics which are not canonical. The simplest is the bimodal logic of the frame  $\langle \mathbb{R}, <, > \rangle$ ; see Example 144 below. By applying the Thomason simulation (to be introduced in Section 2.3) to this logic we obtain a logic in  $\text{NExt}\mathbf{K}$  with the same properties; see Theorem 123. Moreover, in contrast to  $\mathcal{D}$ -persistence, strong Kripke completeness is not preserved under finite sums of logics (see [Wolter 1996b]). It is an open problem, however, whether such logics exist in  $\text{NExt}\mathbf{K4}$ .

<sup>7</sup>Fine [1974c] calls such logics *compact*, which does not agree with the use of this term by Thomason [1972].

Perhaps the simplest examples of Kripke complete logics which are not strongly complete are **GL** and **Grz** (use Theorem 58 and the fact that these logics are not elementary; see *Correspondence Theory*). It is much more difficult to prove that the McKinsey logic  $\mathbf{K} \oplus \Box\Diamond p \rightarrow \Diamond\Box p$  is not strongly complete; the proof can be found in [Wang 1992]. For other examples of modal logics that are not strongly complete see Section 3.4. It is worth noting also that, as was shown in [Fine 1974c], every finite width logic in a *finite* language turns out to be strongly Kripke complete, though this is not the case for logics in an infinite language, witness

$$\mathbf{GL.3} = \mathbf{GL} \oplus \Box(\Box^+ p \rightarrow q) \vee \Box(\Box^+ q \rightarrow p).$$

For the consequence relation  $\vdash_L^*$ , we should take the “global” version  $\models_{\mathfrak{F}}^*$  of  $\models_{\mathfrak{F}}$ . Namely, we put  $\Gamma \models_{\mathfrak{F}}^* \varphi$  if  $\mathfrak{M} \models \Gamma$  implies  $\mathfrak{M} \models \varphi$  for any model  $\mathfrak{M}$  based on  $\mathfrak{F}$ . A modal logic  $L$  is called *globally Kripke complete* if for any finite set of formulas  $\Gamma$  and any formula  $\varphi$ ,  $\Gamma \not\vdash_L^* \varphi$  only if there is a frame  $\mathfrak{F}$  for  $L$  such that  $\Gamma \not\models_{\mathfrak{F}}^* \varphi$ .  $L$  is *strongly globally complete* if this holds for arbitrary (not only finite)  $\Gamma$ . We also say that  $L$  has the *global finite model property* if for every finite  $\Gamma$  and every  $\varphi$ ,  $\Gamma \not\vdash_L^* \varphi$  only if there is a finite frame  $\mathfrak{F}$  for  $L$  such that  $\Gamma \not\models_{\mathfrak{F}}^* \varphi$ .

The global finite model property (FMP, for short) of many standard logics can be proved by filtration. Say that a logic  $L$  *strongly admits filtration* if for every generated submodel  $\mathfrak{M}$  of the canonical model  $\mathfrak{M}_L$  and every finite set of formulas  $\Sigma$  closed under subformulas, there is a filtration of  $\mathfrak{M}$  through  $\Sigma$  based on a frame for  $L$ .

**PROPOSITION 37** (Goranko and Passy 1992). *If  $L$  strongly admits filtration then  $L$  has global FMP.*

**Proof.** Suppose that  $\Gamma \not\vdash_L^* \varphi$ ,  $\Gamma$  finite. Then  $\Box^{<\omega} \wedge \Gamma \not\vdash_L \varphi$  and so the set  $\Delta = \Box^{<\omega} \wedge \Gamma \cup \{\neg\varphi\}$  is  $L$ -consistent. It remains to filtrate through  $\mathbf{Sub}\Gamma \cup \mathbf{Sub}\varphi$  the submodel of  $\mathfrak{M}_L$  generated by a maximal  $L$ -consistent set containing  $\Delta$ . ■

It follows in particular that **K**, **T**, **D**, **KB** have global FMP.

**PROPOSITION 38.** *Suppose  $L$  is globally complete (has global FMP) and  $\Gamma$  is a finite set of variable free formulas. Then  $L \oplus \Gamma$  is globally complete (has global FMP) as well.*

**Proof.** Let  $L' = L \oplus \Gamma$  and  $\Delta \not\vdash_{L'}^* \varphi$ ,  $\Delta$  finite. Then  $\Gamma \cup \Delta \not\vdash_L^* \varphi$  and so there exists a (finite) Kripke frame  $\mathfrak{F}$  for  $L$  such that  $\Gamma \cup \Delta \not\models_{\mathfrak{F}}^* \varphi$ . Since  $\Gamma$  contains no variables,  $\mathfrak{F} \models L'$ . ■

For  $n$ -transitive logics  $L$  the global consequence relation  $\vdash_L^*$  is reducible to the “local”  $\vdash_L$  and so  $L$  is Kripke complete (has FMP, is strongly complete)

iff  $L$  is globally complete (has global FMP, is strongly globally complete). In general the global properties are stronger than the “local” ones. Although  $L$  is globally complete (has global FMP) only if  $L$  is complete (has FMP), the converse does not hold (see [Wolter 1994a] and [Kracht 1999]).

**EXAMPLE 39.** Let  $L = \mathbf{Alt}_3 \oplus p \rightarrow \Box \Diamond p \oplus (\Box p \wedge \neg p) \rightarrow \neg(\Diamond q \wedge \Diamond \neg q)$ . A Kripke frame  $\mathfrak{F}$  validates  $L$  iff no point in  $\mathfrak{F}$  has more than three successors,  $\mathfrak{F}$  is symmetric, and irreflexive points in it have at most one successor. By Proposition 22,  $L$  is Kripke complete. The class of Kripke frames for  $L$  is closed under (not necessarily generated) subframes. So, by Proposition 59 to be proved below,  $L$  has FMP. We show now that it does not have global FMP. To this end we require the formulas:

$$\alpha_1 = q_1 \wedge \neg q_2 \wedge \neg q_3, \quad \alpha_2 = \neg q_1 \wedge q_2 \wedge \neg q_3, \quad \alpha_3 = \neg q_1 \wedge \neg q_2 \wedge q_3,$$

$$\varphi = \Box p \wedge \neg p \wedge \alpha_1, \quad \psi = \bigwedge \{\alpha_i \rightarrow \Diamond \alpha_{i+1} : i = 1, 2\} \wedge \alpha_3 \rightarrow \Diamond \alpha_1.$$

Let  $\mathfrak{F} = \langle W, R \rangle$ , where  $W = \omega$  and

$$R = \{\langle m, m \rangle : m > 0\} \cup \{\langle m, m+1 \rangle : m < \omega\} \cup \{\langle m, m-1 \rangle : m > 0\}.$$

We then have  $\psi \not\models_{\mathfrak{F}}^* \neg \varphi$ . In fact,  $\varphi$  is true at 0 and  $\psi$  is true everywhere under the valuation  $\mathfrak{V}$  defined by  $\mathfrak{V}(p) = W - \{0\}$  and  $\mathfrak{V}(q_i) = \{3n + i : n < \omega\}$ . Clearly,  $\mathfrak{F} \models L$  and so  $\psi \not\models_L^* \neg \varphi$ . Suppose now that  $(\mathfrak{N}, x_0) \models \varphi$  and  $\mathfrak{N} \models \psi$ , for a model  $\mathfrak{N}$  based on a Kripke frame  $\mathfrak{G} = \langle V, S \rangle$  for  $L$ . Then we can find a sequence  $x_j$ ,  $j < \omega$ , such that  $x_j S x_{j+1}$  and  $x_{3j+i} \models \alpha_{i+1}$ , for  $j < \omega$  and  $i = 1, 2, 3$ . The reader can verify that all points  $x_j$  are distinct.

Let us consider now the algebraic meaning of the notions introduced above. A logic  $L$  is Kripke complete iff the variety  $\text{Alg}L$  of modal algebras for  $L$  is generated by the class  $\text{Kr}L = \{\mathfrak{F}^+ : \mathfrak{F} \text{ is a Kripke frame for } L\}$ . By Birkhoff’s Theorem (see e.g. [Mal’cev 1973]), this means that

$$\text{Alg}L = \text{HSPKr}L,$$

(i.e.,  $\text{Alg}L$  is obtained by taking the closure of  $\text{Kr}L$  under direct products, then the closure of the result under (isomorphic copies of) subalgebras and finally under homomorphic images). Clearly,  $L$  is globally complete iff precisely the same quasi-identities hold in  $\text{Kr}L$  and  $\text{Alg}L$ . And since the quasi-variety generated by a class of algebras  $\mathcal{C}$  is  $\text{SPP}_U \mathcal{C}$  (where  $\text{P}_U$  denotes the closure under ultraproducts; see [Mal’cev 1973]),  $L$  is globally complete iff

$$\text{Alg}L = \text{SPP}_U \text{Kr}L.$$

Goldblatt [1989] calls the variety  $\text{Alg}L$  *complex* if  $\text{Alg}L = \text{SKr}L$ , or, equivalently, if  $\text{Alg}L = \text{SPKr}L$  (this follows from the fact that the dual of the disjoint union of a family of Kripke frames  $\{\mathfrak{F}_i : i \in I\}$  is isomorphic

to the product  $\prod_{i \in I} \mathfrak{F}_i^+$ ). We say a logic  $L$  is  $\varkappa$ -complex,  $\varkappa$  a cardinal, if every modal algebra for  $L$  with  $\leq \varkappa$  generators is a subalgebra of  $\mathfrak{F}^+$  for some Kripke frame  $\mathfrak{F} \models L$ . As was shown in [Wolter 1993], this notion turns out to be the algebraic counterpart of both strong completeness and strong global completeness of logics in *infinite languages* with  $\varkappa$  variables.

**THEOREM 40.** *For every normal modal logic  $L$  in an infinite language with  $\varkappa$  variables the following conditions are equivalent:*

- (i)  $L$  is strongly Kripke complete;
- (ii)  $L$  is globally strongly complete;
- (iii)  $L$  is  $\varkappa$ -complex.

**Proof.** (i)  $\Rightarrow$  (iii) Suppose the cardinality of  $\mathfrak{A} \in \text{Alg}L$  does not exceed  $\varkappa$ . Denote by  $\mathfrak{L}$  the algebra of modal formulas over  $\varkappa$  propositional variables and take some homomorphism  $h$  from  $\mathfrak{L}$  onto  $\mathfrak{A}$ . For each ultrafilter  $\nabla$  in  $\mathfrak{A}$ , the set  $h^{-1}(\nabla)$  is maximal  $L$ -consistent. Since  $L$  is strongly complete, there is a model  $\mathfrak{M}_\nabla = \langle \mathfrak{F}_\nabla, \mathfrak{V}_\nabla \rangle$  with root  $x_\nabla$  based on a Kripke frame  $\mathfrak{F}_\nabla$  for  $L$  and such that  $(\mathfrak{M}_\nabla, x_\nabla) \models h^{-1}(\nabla)$ . Without loss of generality we may assume that the frames  $\mathfrak{F}_\nabla$  for distinct  $\nabla$  are disjoint. Let  $\mathfrak{F}$  be the disjoint union of all of them. Define a homomorphism  $\mathfrak{V}$  from  $\mathfrak{L}$  into  $\mathfrak{F}^+$  by taking

$$\mathfrak{V}(p) = \bigcup \{ \mathfrak{V}_\nabla(p) : \nabla \text{ is an ultrafilter in } \mathfrak{A} \}.$$

Then  $\mathfrak{V}(\mathfrak{L})$  is a subalgebra of  $\mathfrak{F}^+ \in \text{Alg}L$  isomorphic to  $\mathfrak{A}$ .

The implication (iii)  $\Rightarrow$  (ii) is trivial. To prove (ii)  $\Rightarrow$  (i), consider an  $L$ -consistent set of formulas  $\Gamma$  of cardinality  $\leq \varkappa$  and put

$$\Delta = \{p\} \cup \{ \Box^n(p \rightarrow \varphi) : n < \omega, \varphi \in \Gamma \},$$

where the variable  $p$  does not occur in formulas from  $\Gamma$ . It is easily checked that all finite subsets of  $\Delta$  are  $L$ -consistent, so  $\Delta$  is  $L$ -consistent too. It follows that  $\{p \rightarrow \varphi : \varphi \in \Gamma\} \not\vdash_L^* \neg p$ . And since  $L$  is globally strongly complete, there exists a model  $\mathfrak{M}$  based on a Kripke frame for  $L$  such that  $\mathfrak{M} \models \{p \rightarrow \varphi : \varphi \in \Gamma\}$  and  $(\mathfrak{M}, x) \models p$ , for some  $x$ . But then  $(\mathfrak{M}, x) \models \Gamma$ . ■

### 1.6 Canonical formulas

The main problem of completeness theory in modal logic is not only to find a sufficiently simple class of frames with respect to which a given logic  $L$  is complete but also to characterize the constitution of frames for  $L$  (in this class). The first order approach to the characterization problem, discussed in Section 1.3 in connection with Sahlqvist's Theorem, comes across two obstacles. First, there are formulas whose Kripke frames cannot be described in the first order language with  $R$  and  $=$ . The best known example

is probably the *Löb axiom*

$$\mathbf{la} = \Box(\Box p \rightarrow p) \rightarrow \Box p.$$

$\mathfrak{F} \models \mathbf{la}$  iff  $\mathfrak{F}$  is transitive, irreflexive (i.e., a strict partial order) and *Noetherian* in the sense that it contains no infinite ascending chain of distinct points. And as is well known, the condition of Noetherianness is not a first order one. The second obstacle is that this approach deals only with logics that are Kripke complete; it does not take into account sets of possible values.

There is another, purely frame-theoretic method of characterizing the structure of frames. For instance, a frame  $\mathfrak{G}$  validates  $\mathbf{K}/\mathfrak{F}$  iff  $\mathfrak{G}$  does not contain a generated subframe reducible to  $\mathfrak{F}$ . It was shown in [Zakharyashev 1984, 1988, 1992] that in a similar manner one can describe *transitive* frames validating an arbitrary modal formula. It is not clear whether characterizations of this sort can be extended to the class of all frames (an important step in this direction would be a generalization to  $n$ -transitive frames). That is why all frames in this section are assumed to be transitive. First we illustrate this method by a simple example.

EXAMPLE 41. Suppose a frame  $\mathfrak{F} = \langle W, R, P \rangle$  refutes  $\mathbf{la}$  under some valuation. Then the set  $V = \{x \in W : x \not\models \mathbf{la}\}$  is in  $P$  and  $V \subseteq V\downarrow$ . It follows from the former that  $\mathfrak{G} = \langle V, R \upharpoonright V, \{X \cap V : X \in P\} \rangle$  is a frame—we call it the *subframe of  $\mathfrak{F}$  induced by  $V$* . And the latter condition means that  $\mathfrak{G}$  is reducible to the single reflexive point  $\circ$  which is the simplest refutation frame for  $\mathbf{la}$ . Moreover, one can readily check that the converse also holds: if there is a subframe  $\mathfrak{G}$  of  $\mathfrak{F}$  reducible to  $\circ$  then  $\mathfrak{F} \not\models \mathbf{la}$ .

This example motivates the following definitions. Given frames  $\mathfrak{F} = \langle W, R, P \rangle$  and  $\mathfrak{G} = \langle V, S, Q \rangle$ , a partial (i.e., not completely defined, in general) map  $f$  from  $W$  onto  $V$  is called a *subreduction* of  $\mathfrak{F}$  to  $\mathfrak{G}$  if it satisfies the reduction conditions (R1)–(R3) for all  $x$  and  $y$  in the domain of  $f$  and all  $X \in Q$ . The domain of  $f$  will be denoted by  $\text{dom}f$ . In other words, an  $f$ -subreduct of  $\mathfrak{F}$  is a reduct of the subframe of  $\mathfrak{F}$  induced by  $\text{dom}f$ . A frame  $\mathfrak{G} = \langle V, S, Q \rangle$  is a *subframe* of  $\mathfrak{F} = \langle W, R, P \rangle$  if  $V \subseteq W$  and the identity map on  $V$  is a subreduction of  $\mathfrak{F}$  to  $\mathfrak{G}$ , i.e., if  $S = R \upharpoonright V$  and  $Q \subseteq P$ . Note that a generated subframe  $\mathfrak{G}$  of  $\mathfrak{F}$  is not in general a subframe of  $\mathfrak{F}$ , since  $V$  may be not in  $P$ .

Thus, the result of Example 41 can be reformulated like this:  $\mathfrak{F} \not\models \mathbf{la}$  iff  $\mathfrak{F}$  is subreducible to  $\circ$ .

A subreduction  $f$  of  $\mathfrak{F}$  to  $\mathfrak{G}$  is called *cofinal* if

$$\text{dom}f \uparrow \subseteq \text{dom}f \downarrow.$$

This important notion can be motivated by the following observation:  $\mathfrak{F}$  refutes  $\Diamond \top$  iff  $\mathfrak{F}$  is cofinally subreducible to  $\bullet$  (a plain subreduction is not enough).

**THEOREM 42.** *Every refutation frame  $\mathfrak{F} = \langle W, R, P \rangle$  for  $\varphi(p_1, \dots, p_n)$  is cofinally subreducible to a finite rooted refutation frame for  $\varphi$  containing at most  $c_\varphi = 2^n \cdot (c_n(1) + \dots + c_n(2^{|\mathbf{Sub}\varphi|}))$  points.<sup>8</sup>*

**Proof.** Suppose  $\varphi$  is refuted in  $\mathfrak{F}$  under a valuation  $\mathfrak{V}$ . Without loss of generality we can assume  $\mathfrak{F}$  to be generated by  $\mathfrak{V}(p_1), \dots, \mathfrak{V}(p_n)$ . Let  $X_1, \dots, X_m$  be all distinct maximal 0-cyclic sets in  $\mathfrak{F}$ . Clearly,  $m \leq c_n(1)$  but unlike Theorem 8,  $\mathfrak{F}$  is not in general refined and so these sets are not necessarily clusters of depth 1. However, they can be easily reduced to such clusters. Define an equivalence relation  $\sim$  on  $W$  by putting  $x \sim y$  iff  $x = y$  or  $x, y \in X_i$ , for some  $i \in \{1, \dots, m\}$ , and  $x \sim_\Sigma y$  (as before  $\Sigma = \{p_1, \dots, p_n\}$ ). Let  $[x]$  be the equivalence class under  $\sim$  generated by  $x$  and  $[X] = \{[x] : x \in X\}$ , for  $X \in P$ . By the definition of cyclic sets,  $xRy$  iff  $[x] \subseteq [y] \downarrow$ . So the map  $x \mapsto [x]$  is a reduction of  $\mathfrak{F}$  to the frame  $\mathfrak{F}'_1 = \langle W'_1, R'_1, P'_1 \rangle$  which results from  $\mathfrak{F}$  by “folding up” the 0-cyclic sets  $X_i$  into clusters of depth 1 and leaving the other points untouched:  $W'_1 = [W]$ ,  $[x]R'_1[y]$  iff  $[x] \subseteq [y] \downarrow$  and  $P'_1 = \{[X] : X \in P\}$ . (Roughly, we refine that part of  $\mathfrak{F}$  which gives points of depth 1.) Put  $\mathfrak{V}'_1(p_i) = [\mathfrak{V}(p_i)]$ . Then by the Reduction (or P-morphism) Theorem, we have  $x \models \psi$  iff  $[x] \models \psi$ , for every  $\psi \in \mathbf{Sub}\varphi$ .

Let  $X$  be the set of all points in  $\mathfrak{F}'_1$  of depth  $> 1$  having  $\mathbf{Sub}\varphi$ -equivalent successors of depth 1. It is not hard to see that  $X \in P'_1$ . Denote by  $\mathfrak{F}_1 = \langle W_1, R_1, P_1 \rangle$  the subframe of  $\mathfrak{F}'_1$  induced by  $W'_1 - X$  and let  $\mathfrak{V}_1$  be the restriction of  $\mathfrak{V}'_1$  to  $\mathfrak{F}_1$ . By induction on the construction of  $\psi \in \mathbf{Sub}\varphi$  one can readily show that  $\psi$  has the same truth-values at common points in  $\mathfrak{F}'_1$  and  $\mathfrak{F}_1$  (under  $\mathfrak{V}'_1$  and  $\mathfrak{V}_1$ , respectively) and so  $\mathfrak{F}_1 \not\models \varphi$ . The partial map  $x \mapsto [x]$ , for  $[x] \in W_1$ , is a cofinal subreduction of  $\mathfrak{F}$  to  $\mathfrak{F}_1$ .

Then we take the maximal 1-cyclic sets in  $\mathfrak{F}_1$ , “fold” them up into clusters of depth 2 and remove those points of depth  $> 2$  that have  $\mathbf{Sub}\varphi$ -equivalent successors of depth 2. The resulting frame  $\mathfrak{F}_2$  will be a cofinal subreduct of  $\mathfrak{F}_1$  and so of  $\mathfrak{F}$  as well. After that we form clusters of depth 3, and so forth. In at most  $2^{|\mathbf{Sub}\varphi|}$  steps of that sort we shall construct a cofinal subreduct of  $\mathfrak{F}$  refuting  $\varphi$  and containing  $\leq c_\varphi$  points. It remains to select in it a suitable rooted generated subframe. ■

For the majority of standard modal axioms the converse also holds. However, not for all. The simplest counterexample is the density axiom  $\mathbf{den} = \Box\Box p \rightarrow \Box p$ . It is refuted by the chain  $\mathfrak{H}$  of two irreflexive points but becomes valid if we insert between them a reflexive one. In fact,  $\mathfrak{F} \not\models \mathbf{den}$  iff there is a subreduction  $f$  of  $\mathfrak{F}$  to  $\mathfrak{H}$  such that  $f(x\uparrow) = \{a\}$ , for no point  $x$  in  $\text{dom}f\uparrow - \text{dom}f$ , where  $a$  is the final point in  $\mathfrak{H}$ .

Loosely, every refutation frame for formulas like  $\mathbf{la}$  can be constructed by adding new points to a frame  $\mathfrak{G}$  that is reducible to some finite refutation

<sup>8</sup>The function  $c_n(m)$  was defined in Section 1.2.

frame of fixed size. For formulas like  $\diamond\top$  we have to take into account the cofinality condition and do not put new points “above”  $\mathfrak{G}$ . And formulas like *den* impose another restriction: some places inside  $\mathfrak{G}$  may be “closed” for inserting new points. These “closed domains” can be singled out in the following way.

Suppose  $\mathfrak{M} = \langle \mathfrak{H}, \mathfrak{U} \rangle$  is a model and  $\mathfrak{a}$  an antichain in  $\mathfrak{H}$ . Say that  $\mathfrak{a}$  is an *open domain* in  $\mathfrak{M}$  relative to a formula  $\varphi$  if there is a pair  $t_{\mathfrak{a}} = (\Gamma_{\mathfrak{a}}, \Delta_{\mathfrak{a}})$  such that  $\Gamma_{\mathfrak{a}} \cup \Delta_{\mathfrak{a}} = \mathbf{Sub}\varphi$ ,  $\bigwedge \Gamma_{\mathfrak{a}} \rightarrow \bigvee \Delta_{\mathfrak{a}} \notin \mathbf{K4}$  and

- $\Box\psi \in \Gamma_{\mathfrak{a}}$  implies  $\psi \in \Gamma_{\mathfrak{a}}$ ,
- $\Box\psi \in \Gamma_{\mathfrak{a}}$  iff  $a \models \Box^+\psi$  for all  $a \in \mathfrak{a}$ .

Otherwise  $\mathfrak{a}$  is called a *closed domain* in  $\mathfrak{M}$  relative to  $\varphi$ . A reflexive singleton  $\mathfrak{a} = \{a\}$  is always open: just take

$$t_{\mathfrak{a}} = (\{\psi \in \mathbf{Sub}\varphi : a \models \psi\}, \{\psi \in \mathbf{Sub}\varphi : a \not\models \psi\}).$$

It is easy to see also that antichains consisting of points from the same clusters are open or closed simultaneously; we shall not distinguish between such antichains.

For a frame  $\mathfrak{H}$  and a (possibly empty) set  $\mathfrak{D}$  of antichains in  $\mathfrak{H}$ , we say a subreduction  $f$  of  $\mathfrak{F}$  to  $\mathfrak{H}$  satisfies the *closed domain condition* for  $\mathfrak{D}$  if

$$(\text{CDC}) \quad \neg\exists x \in \text{dom}f \uparrow - \text{dom}f \exists \mathfrak{d} \in \mathfrak{D} f(x \uparrow) = \mathfrak{d} \uparrow.$$

Notice that the cofinal subreduction  $f$  of  $\mathfrak{F}$  to the resulting finite rooted frame  $\mathfrak{H}$  in the proof of Theorem 42 satisfies (CDC) for the set  $\mathfrak{D}$  of closed domains in the corresponding model  $\mathfrak{M}$  on  $\mathfrak{H}$  refuting  $\varphi$ . Indeed, every  $x \in \text{dom}f \uparrow - \text{dom}f$  has a  $\mathbf{Sub}\varphi$ -equivalent successor  $y \in \text{dom}f$ , and so an antichain  $\mathfrak{d}$  such that  $f(x \uparrow) = \mathfrak{d} \uparrow$  is open, since we can take

$$t_{\mathfrak{d}} = (\{\psi \in \mathbf{Sub}\varphi : y \models \psi\}, \{\psi \in \mathbf{Sub}\varphi : y \not\models \psi\}).$$

On the other hand, we have

**PROPOSITION 43.** *Suppose  $\mathfrak{M} = \langle \mathfrak{H}, \mathfrak{U} \rangle$  is a finite countermodel for  $\varphi$  and  $\mathfrak{D}$  the set of all closed domains in  $\mathfrak{M}$  relative to  $\varphi$ . Then  $\mathfrak{F} \not\models \varphi$  whenever there is a cofinal subreduction  $f$  of  $\mathfrak{F}$  to  $\mathfrak{H}$  satisfying (CDC) for  $\mathfrak{D}$ . Moreover, if  $\varphi$  is negation free (i.e., contains no  $\perp$ ,  $\neg$ ,  $\diamond$ ) then a plain subreduction satisfying (CDC) for  $\mathfrak{D}$  is enough.*

**Proof.** If  $f$  is cofinal and  $\mathfrak{F} = \langle W, R, P \rangle$  then we can assume  $\text{dom}f \uparrow = W$ . Define a valuation  $\mathfrak{V}$  in  $\mathfrak{F}$  as follows. If  $x \in \text{dom}f$  then we take  $x \models p$  iff  $f(x) \models p$ , for every variable  $p$  in  $\varphi$ . If  $x \notin \text{dom}f$  then  $f(x \uparrow) \neq \emptyset$ , since  $f$  is cofinal. Let  $\mathfrak{a}$  be an antichain in  $\mathfrak{H}$  such that  $\mathfrak{a} \uparrow = f(x \uparrow)$ . By (CDC),  $\mathfrak{a}$  is an open domain in  $\mathfrak{M}$ , and we put  $y \models p$  iff  $p \in \Gamma_{\mathfrak{a}}$ , for every  $y \notin \text{dom}f$  such that  $f(y \uparrow) = f(x \uparrow)$ . One can show that  $\mathfrak{V}$  is really a valuation in  $\mathfrak{F}$  and,

for every  $\psi \in \mathbf{Sub}\varphi$ ,  $x \models \psi$  iff  $f(x) \models \psi$  in the case  $x \in \text{dom}f$ , and  $x \models \psi$  iff  $\psi \in \Gamma_{\mathfrak{a}}$ , where  $\mathfrak{a}$  is the open domain in  $\mathfrak{N}$  associated with  $x$ , in the case  $x \notin \text{dom}f$ .

If  $\varphi$  is negation free and  $f$  is a plain subreduction then  $f(x\uparrow)$  may be empty. In such a case we just put  $x \models p$ , for all variables  $p$ . ■

Now let us summarize what we have got. Given an arbitrary formula  $\varphi$ , we can effectively construct a finite collection of finite rooted frames  $\mathfrak{F}_1, \dots, \mathfrak{F}_n$  (underlying all possible rooted countermodels for  $\varphi$  with  $\leq c_\varphi$  points) and select in them sets  $\mathfrak{D}_1, \dots, \mathfrak{D}_n$  of antichains (open domains in those countermodels) such that, for any frame  $\mathfrak{F}$ ,  $\mathfrak{F} \not\models \varphi$  iff there is a cofinal subreduction of  $\mathfrak{F}$  to  $\mathfrak{F}_i$ , for some  $i$ , satisfying (CDC) for  $\mathfrak{D}_i$ . If  $\varphi$  is negation free then a plain subreduction satisfying (CDC) is enough.

This general characterization of the constitution of refutation transitive frames can be presented in a more convenient form if with every finite rooted frame  $\mathfrak{F} = \langle W, R \rangle$  and a set  $\mathfrak{D}$  of antichains in  $\mathfrak{F}$  we associate formulas  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  and  $\alpha(\mathfrak{F}, \mathfrak{D})$  such that  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  ( $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathfrak{D})$ ) iff there is a cofinal (respectively, plain) subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ . For instance, one can take

$$\alpha(\mathfrak{F}, \mathfrak{D}, \perp) = \bigwedge_{a_i R a_j} \varphi_{ij} \wedge \bigwedge_{i=0}^n \varphi_i \wedge \bigwedge_{\mathfrak{d} \in \mathfrak{D}} \varphi_{\mathfrak{d}} \wedge \varphi_{\perp} \rightarrow p_0$$

where  $a_0, \dots, a_n$  are all points in  $\mathfrak{F}$  and  $a_0$  is its root,

$$\begin{aligned} \varphi_{ij} &= \Box^+(\Box p_j \rightarrow p_i), \\ \varphi_i &= \Box^+(\bigwedge_{\neg a_i R a_k} \Box p_k \wedge \bigwedge_{j=0, j \neq i}^n p_j \rightarrow p_i) \rightarrow p_i, \\ \varphi_{\mathfrak{d}} &= \Box^+(\bigwedge_{a_i \in W - \mathfrak{d}\uparrow} \Box p_j \wedge \bigwedge_{i=0}^n p_i \rightarrow \bigvee_{a_j \in \mathfrak{d}} \Box p_j), \\ \varphi_{\perp} &= \Box^+(\bigwedge_{i=0}^n \Box^+ p_i \rightarrow \perp). \end{aligned}$$

$\alpha(\mathfrak{F}, \mathfrak{D})$  results from  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  by deleting the conjunct  $\varphi_{\perp}$ .  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  and  $\alpha(\mathfrak{F}, \mathfrak{D})$  are called the *canonical* and *negation free canonical formulas* for  $\mathfrak{F}$  and  $\mathfrak{D}$ , respectively. It is not hard to check that if  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  is refuted in  $\mathfrak{G} = \langle V, S, Q \rangle$  under some valuation then the partial map defined by  $x \mapsto a_i$  if the premise of  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  is true at  $x$  and  $p_i$  false is a cofinal subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ ; and conversely, if  $f$  is such a subreduction then the valuation  $\mathfrak{U}$  defined by  $\mathfrak{U}(p_i) = V - f^{-1}(a_i)$  refutes  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  at any point in  $f^{-1}(a_0)$ .



THEOREM 44. *There is an algorithm which, given a formula  $\varphi$ , returns canonical formulas  $\alpha(\mathfrak{F}_1, \mathfrak{D}_1, \perp), \dots, \alpha(\mathfrak{F}_n, \mathfrak{D}_n, \perp)$  such that*

$$\mathbf{K4} \oplus \varphi = \mathbf{K4} \oplus \alpha(\mathfrak{F}_1, \mathfrak{D}_1, \perp) \oplus \dots \oplus \alpha(\mathfrak{F}_n, \mathfrak{D}_n, \perp).$$

*So the set of canonical formulas is complete for the class  $\text{NExtK4}$ . If  $\varphi$  is negation free then one can use negation free canonical formulas.*

It is not hard to see that  $\mathbf{K4} \oplus \varphi$  is a splitting of  $\text{NExtK4}$  iff  $\varphi$  is deductively equivalent in  $\text{NExtK4}$  to a formula of the form  $\alpha(\mathfrak{F}, \mathfrak{D}^\sharp, \perp)$ , where  $\mathfrak{D}^\sharp$  is the set of all antichains in  $\mathfrak{F}$  (in this case  $\mathbf{K4}/\mathfrak{F} = \mathbf{K4} \oplus \alpha(\mathfrak{F}, \mathfrak{D}^\sharp, \perp)$ ). Such formulas are known as *Jankov formulas* (Jankov [1963] introduced them for intuitionistic logic), or *frame formulas* (cf. [Fine 1974a]), or *Jankov–Fine formulas*. Since  $\mathbf{GL}$  is not a union-splitting of  $\text{NExtK4}$ , this class of logics has no axiomatic basis.

We conclude this section by showing in Table 2 canonical axiomatizations of some standard modal logics in the field of  $\mathbf{K4}$ . For brevity we write  $\alpha(\mathfrak{F}, \perp)$  instead of  $\alpha(\mathfrak{F}, \emptyset, \perp)$  and  $\alpha^\sharp(\mathfrak{F}, \perp)$  instead of  $\alpha(\mathfrak{F}, \mathfrak{D}^\sharp, \perp)$ . Each  $*$  in the table is to be replaced by both  $\circ$  and  $\bullet$ .

For more information about the canonical formulas the reader is referred to [Zakharyashev 1992, 1997b].

### 1.7 Decidability via the finite model property

Although, for cardinality reason, there are “much more” undecidable logics than decidable ones, almost all “natural” propositional systems close to those we deal with in this chapter turn out to be decidable. Relevant and linear logics are probably the best known among very few exceptions (see [Urquhart 1984], [Lincoln *et al.* 1992]).

The majority of decidability results in modal logic was obtained by means of establishing the finite model property. FMP by itself does not ensure yet decidability (there is a continuum of logics with FMP); some additional conditions are required to be satisfied. For instance, to prove the decidability of  $\mathbf{S4}$  McKinsey [1941] used two such conditions: that the logic under consideration is characterized by an effective class of finite frames (or algebras, matrices, models, etc.) and that there is an effective (exponential in the case of  $\mathbf{S4}$ ) upper bound for the size of minimal refutation frames. Under these conditions, a formula belongs to the logic iff it is validated by (finite) frames in a finite family which can be effectively constructed. Another sufficient condition of decidability is provided by the following well known

THEOREM 45 (Harrop 1958). *Every finitely axiomatizable logic with FMP is decidable.*

Here we need not to know a priori anything about the structure of frames for a given logic. This information is replaced by checking the validity of its

Table 2. Canonical axioms of standard modal logics

---

<b>D4</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\bullet, \perp)$	
<b>S4</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\bullet)$	
<b>GL</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\circ)$	
<b>Grz</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\bullet) \oplus \alpha(\overline{\circ\circ})$	
<b>K4.1</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\bullet, \perp) \oplus \alpha(\overline{\circ\circ}, \perp)$	
<b>Triv</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\bullet) \oplus \alpha(\overline{\circ\circ}) \oplus \alpha(\overset{\circ}{\circ})$	
<b>Verum</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\circ) \oplus \alpha(\overset{\bullet}{\circ})$	
<b>S5</b>	<b>=</b>	<b>S4</b>	$\oplus$	$\alpha(\overset{\circ}{\circ})$	
<b>K4B</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\overset{*}{\bullet})$ (4 axioms)	
<b>A*</b>	<b>=</b>	<b>GL</b>	$\oplus$	$\alpha(\overset{1 \bullet}{\bullet} \overset{2 \bullet}{\bullet}, \{\{1\}, \{1, 2\}\})$	
<b>K4.2</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\overset{\bullet}{\bullet}, \perp) \oplus \alpha(\overset{\circ}{\circ}, \perp) \oplus \alpha(\overset{* \bullet}{\bullet}, \perp)$ (8 axioms)	
<b>K4.3</b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\overset{* \bullet}{\bullet})$ (6 axioms)	
<b>Dum</b>	<b>=</b>	<b>S4</b>	$\oplus$	$\alpha(\overset{\overline{\circ\circ}}{\circ} \overset{\circ}{\circ}) \oplus \alpha(\overset{\circ}{\overline{\circ\circ}})$	
<b>K4BW<sub>n</sub></b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\overset{n+1}{\overset{* \bullet}{\bullet} \cdots \overset{* \bullet}{\bullet}})$ (2n + 4 axioms)	
<b>K4BD<sub>n</sub></b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\overset{* n}{\vdots} \overset{* 1}{\bullet} \overset{* 0}{\bullet})$ (2 <sup>n+1</sup> axioms)	
<b>K4<sub>n,m</sub></b>	<b>=</b>	<b>K4</b>	$\oplus$	$\alpha(\overset{\bullet m}{\vdots} \overset{\bullet 1}{\bullet}, \mathfrak{D}^{\sharp})$	

---

axioms in finite frames, and the restriction of the size of refutation frames is replaced by constructing all possible derivations: in a finite number of steps we either separate a tested formula from the logic or derive it. Note that unlike the previous case now we cannot estimate the time required to complete this algorithm.

The condition of finite axiomatizability in Harrop's Theorem cannot be weakened to that of recursive axiomatizability. For there is a logic of depth 3 in  $\text{NExt}\mathbf{K4}$  (i.e., a logic in  $\text{NExt}\mathbf{K4BD}_3$ ) with an infinite set of independent axioms; so the logic of depth 3 is axiomatizable by some recursively enumerable but not recursive sequence of formulas in this set is undecidable and has FMP. On the other hand there are examples of undecidable logics characterized by decidable classes of finite frames (see e.g. [Chagro and Zakharyashev 1997]). Yet one can generalize Harrop's Theorem in the following way. A logic is decidable iff it is recursively enumerable and characterized by a recursive class of recursive algebras. However, this criterion is absolutely useless in its generality. In this connection we note two open problems posed by Kuznetsov [1979]. Is every finitely axiomatizable logic characterized by recursive algebras? Is every finitely axiomatizable logic, characterized by recursive algebras, decidable? (That *finite* axiomatizability is essential here is explained by the following fact: if a lattice of logics contains a logic with a continuum of immediate predecessors then there is no countable sequence of algebras such that every logic in the lattice is characterized by one of its subsequences. For details see [Chagro and Zakharyashev 1997].)

FMP of almost all standard systems was proved using various forms of filtration (consult Section 12 *Basic Modal Logic* and [Gabbay 1976]). However, the method of filtration is rather capricious; one needs a special craft to apply it in each particular case (for instance, to find a suitable "filter"). In this and two subsequent sections we discuss other methods of proving FMP which are applicable to families of logics and provide in fact sufficient conditions of FMP. (It is to be noted that the families of Kripke complete logics considered in Section 1.3 contain logics without FMP.) A pair of such conditions was already presented in *Basic Modal Logic*:

**THEOREM 46** (Segherberg 1971). *Each logic in  $\text{NExt}\mathbf{K4}$  characterized by a frame of finite depth (or, which is equivalent, containing  $\mathbf{K4BD}_n$ , for some  $n < \omega$ ) has FMP.*

**THEOREM 47** (Bull 1966b, Fine 1971). *Each logic in  $\text{NExt}\mathbf{S4.3}$  has FMP and is finitely axiomatizable (and so decidable).*

The former result, covering a continuum of logics, follows immediately from the description of finitely generated refined frames for  $\mathbf{K4}$  in Section 1.2 and the latter is a consequence of Theorem 52 and Example 54 below. It is worth noting also that since  $\mathfrak{F}_L(n)$  is finite for every logic  $L \in \text{NExt}\mathbf{K4}$  of finite depth and every  $n < \omega$ , there are only finitely many pairwise

non-equivalent in  $L$  formulas with  $n$  variables. Logics with this property are called *locally tabular* (or *locally finite*). Moreover, as was observed by Maksimova [1975a], the converse is also true: if  $L \in \text{NExt}\mathbf{K4}$  has frames of any depth  $< \omega$  then the formulas in the sequence  $\varphi_1 = p$ ,  $\varphi_{n+1} = p \vee \Box(p \rightarrow \Box\varphi_n)$  are not equivalent in  $L$ . Thus, a logic in  $\text{NExt}\mathbf{K4}$  is locally tabular iff it is of finite depth. For  $L \in \text{NExt}\mathbf{S4}$  this criterion can be reformulated in the following way:  $L$  is not locally tabular iff  $L \subseteq \mathbf{Grz.3}$ , where  $\mathbf{Grz.3} = \mathbf{S4.3} \oplus \mathbf{Grz}$ . Likewise,  $L \in \text{NExt}\mathbf{GL}$  is not locally tabular iff  $L \subseteq \mathbf{GL.3}$ . Nagle and Thomason [1985] showed that all normal extensions of  $\mathbf{K5}$  are locally tabular.

**Uniform logics** Fine [1975a] used a modal analog of the full disjunctive normal form for constructing finite models and proving FMP of a family of logics in  $\text{NExt}\mathbf{D}$  (containing in particular the McKinsey system  $\mathbf{K} \oplus \Box\Diamond p \rightarrow \Diamond\Box p$  which had resisted all attempts to prove its completeness by the method of canonical models and filtration). Let us notice first that every formula  $\varphi(p_1, \dots, p_m)$  is equivalent in  $\mathbf{K}$  either to  $\perp$  or to a disjunction of normal forms (in the variables  $p_1, \dots, p_m$ ) of degree  $md(\varphi)$ , which are defined inductively in the following way.  $\mathbf{NF}_0$ , the set of *normal forms of degree 0*, contains all formulas of the form  $\neg_1 p_1 \wedge \dots \wedge \neg_m p_m$ , where each  $\neg_i$  is either blank or  $\neg$ .  $\mathbf{NF}_{n+1}$ , the set of *normal forms of degree  $n+1$* , consists of formulas of the form

$$\theta \wedge \neg_1 \Diamond \theta_1 \wedge \dots \wedge \neg_k \Diamond \theta_k,$$

where  $\theta \in \mathbf{NF}_0$  and  $\theta_1, \dots, \theta_k$  are all distinct normal forms in  $\mathbf{NF}_n$ . Put  $\mathbf{NF} = \bigcup_{n < \omega} \mathbf{NF}_n$ . Using the fact that  $\bigvee \{\Diamond \theta : \theta \in \mathbf{NF}_n\} \in \mathbf{D}$  it is not hard to see also that in  $\mathbf{D}$  every formula  $\varphi$  with  $md(\varphi) \leq n$  is equivalent either to  $\perp$  or to a disjunction of normal forms of degree  $n$  such that at least one of  $\neg_1, \dots, \neg_k$  in the inductive step of the definition above is blank. Such normal forms are called  *$\mathbf{D}$ -suitable*.

It should be clear that, for any distinct  $\theta', \theta'' \in \mathbf{NF}_n$ ,  $\neg(\theta' \wedge \theta'') \in \mathbf{K}$ . Consequently, for every  $\theta \in \mathbf{NF}_n$  and every  $\varphi(p_1, \dots, p_m)$  with  $md(\varphi) \leq n$ , we have either  $\theta \rightarrow \varphi \in \mathbf{K}$  or  $\theta \rightarrow \neg\varphi \in \mathbf{K}$ .

With each  $\mathbf{D}$ -suitable normal form  $\theta$  we associate a model  $\mathfrak{M}_\theta = \langle \mathfrak{F}_\theta, \mathfrak{B}_\theta \rangle$  on a frame  $\mathfrak{F}_\theta = \langle W_\theta, R_\theta \rangle$  by taking

$$W_\theta = \{\top\} \cup \{\theta' \in \mathbf{NF} : \theta' <^n \theta, \text{ for some } n \geq 0\},$$

$$\theta' < \theta'' \text{ iff } \Diamond \theta' \text{ is a conjunct of } \theta'',$$

$$\theta' R_\theta \theta'' \text{ iff either } \theta' > \theta'' \text{ or } md(\theta') = 0 \text{ and } \theta'' = \top,$$

$$\mathfrak{B}_\theta(p) = \{\theta' \in W_\theta : p \text{ is a conjunct of } \theta'\}.$$

According to the definition,  $\top$  is the reflexive last point in  $\mathfrak{F}_\theta$  and so  $\mathfrak{F}_\theta$  is serial. By a straightforward induction on the degree of  $\theta' \in W_\theta$  one can

readily show that  $(\mathfrak{M}_\theta, \theta') \models \theta'$ . It follows immediately that  $\mathbf{D}$  has FMP. Indeed, given  $\varphi \notin \mathbf{D}$ , we reduce  $\neg\varphi$  to a disjunction of  $\mathbf{D}$ -suitable normal forms with at least one disjunct  $\theta$ , and then  $(\mathfrak{M}_\theta, \theta) \models \theta$ .

It turns out that in the same way we can prove FMP of all logics in  $\text{NExt}\mathbf{D}$  axiomatizable by *uniform formulas*, which are defined as follows. Every  $\varphi$  without modal operators is a *uniform formula of degree 0*; and if  $\varphi = \psi(\bigcirc_1\chi_1, \dots, \bigcirc_m\chi_m)$ , where  $\bigcirc_i \in \{\Box, \Diamond\}$ ,  $md(\psi(p_1, \dots, p_m)) = 0$  and  $\chi_1, \dots, \chi_m$  are uniform formulas of degree  $n$ , then  $\varphi$  is a *uniform formula of degree  $n+1$* . A remarkable property of uniform formulas is the following:

**PROPOSITION 48.** *Suppose  $\varphi$  is a uniform formula of degree  $n$  and  $\mathfrak{M}, \mathfrak{N}$  are models based upon the same frame and such that, for some point  $x$ ,  $(\mathfrak{M}, y) \models p$  iff  $(\mathfrak{N}, y) \models p$  for every  $y \in x\uparrow^n$  and every variable  $p$  in  $\varphi$ . Then  $(\mathfrak{M}, x) \models \varphi$  iff  $(\mathfrak{N}, x) \models \varphi$ .*

Given a logic  $L$ , we call a normal form  $\theta$   *$L$ -suitable* if  $\mathfrak{F}_\theta \models L$ .

**THEOREM 49** (Fine 1975a). *Every logic  $L \in \text{NExt}\mathbf{D}$  axiomatizable by uniform formulas has FMP.*

**Proof.** It suffices to prove that each formula  $\varphi$  with  $md(\varphi) \leq n$  is equivalent in  $L$  either to  $\perp$  or to a disjunction of  $L$ -suitable normal forms of degree  $n$ . And this fact will be established if we show that every  $\mathbf{D}$ -suitable normal form  $\theta$  such that  $\theta \rightarrow \perp \notin L$  is  $L$ -suitable. Suppose otherwise. Let  $\theta$  be an  $L$ -consistent and  $\mathbf{D}$ -suitable normal form of the least possible degree under which it is not  $L$ -suitable. Then there are a uniform formula  $\psi \in L$  of some degree  $m$  and a model  $\mathfrak{M} = \langle \mathfrak{F}_\theta, \mathfrak{M} \rangle$  such that  $(\mathfrak{M}, \theta) \not\models \psi$ .

For every variable  $p$  in  $\psi$ , let  $\Gamma_p = \{\theta' \in \theta\uparrow^m : (\mathfrak{M}, \theta') \models p\}$  and let  $\delta_p = \bigvee \Gamma_p$  (if  $\Gamma_p = \emptyset$  then  $\delta_p = \perp$ ). Observe that for every  $\theta' \in \theta\uparrow^m$  we have  $(\mathfrak{M}_\theta, \theta') \models \delta_p$  iff  $\theta' \in \Gamma_p$  iff  $(\mathfrak{M}, \theta') \models p$ . Therefore, by Proposition 48, the formula  $\psi'$  which results from  $\psi$  by replacing each  $p$  with  $\delta_p$  is false at  $\theta$  in  $\mathfrak{M}_\theta$ . Now, if  $md(\psi') > n$  then  $m > n$  and so  $\delta_p = \perp$  for every  $p$  in  $\psi$ , i.e.,  $\psi'$  is variable free. But then  $\psi'$  is equivalent in  $\mathbf{D}$  to  $\top$  or  $\perp$ , contrary to  $\mathfrak{F}_\theta \not\models \psi'$  and  $L$  being consistent. And if  $md(\psi') \leq n$  then either  $\theta \rightarrow \psi' \in \mathbf{K}$ , which is impossible, since  $(\mathfrak{M}_\theta, \theta) \not\models \theta \rightarrow \psi'$ , or  $\theta \rightarrow \neg\psi' \in \mathbf{K}$ , from which  $\psi' \rightarrow \neg\theta \in \mathbf{K}$  and so  $\neg\theta \in L$ , contrary to  $\theta$  being  $L$ -consistent.  $\blacksquare$

**Logics with  $\Box\Diamond$ -axioms** Another result, connecting FMP of logics with the distribution of  $\Box$  and  $\Diamond$  over their axioms, is based on the following

**LEMMA 50.** *For any  $\varphi$  and  $\psi$ ,  $\Diamond\varphi \leftrightarrow \Diamond\psi \in \mathbf{S5}$  iff  $\Box\Diamond\varphi \leftrightarrow \Box\Diamond\psi \in \mathbf{K4}$ .*

**Proof.** Suppose  $\Box\Diamond\varphi \rightarrow \Box\Diamond\psi \notin \mathbf{K4}$ . Then there is a finite model  $\mathfrak{M}$ , based on a transitive frame, and a point  $x$  in it such that  $x \models \Box\Diamond\varphi$  and  $x \not\models \Box\Diamond\psi$ . It follows from the former that every final cluster accessible from  $x$ , if any, is non-degenerate and contains a point where  $\varphi$  is true. The latter means

that  $x$  sees a final cluster  $C$  at all points of which  $\psi$  is false. Now, taking the generated submodel of  $\mathfrak{M}$  based on  $C$ , we obtain a model for **S5** refuting  $\Diamond\varphi \rightarrow \Diamond\psi$ . The rest is obvious, since  $\Diamond p \leftrightarrow \Diamond\Box p$  is in **S5** and **K4**  $\subseteq$  **S5**. ■

Formulas in which every occurrence of a variable is in the scope of a modality  $\Box\Diamond$  will be called  $\Box\Diamond$ -formulas.

**THEOREM 51** (Rybakov 1978). *If a logic  $L \in \text{NExtK4}$  is decidable (or has FMP) and  $\psi$  is a  $\Box\Diamond$ -formula then  $L \oplus \psi$  is also decidable (has FMP).*

**Proof.** Let  $\psi = \psi'(\Box\Diamond\chi_1, \dots, \Box\Diamond\chi_n)$ , for some formula  $\psi'(q_1, \dots, q_n)$ . If  $\varphi(p_1, \dots, p_m) \in L \oplus \psi$  then there exists a derivation of  $\varphi$  in  $L \oplus \psi$  in which substitution instances of  $\psi$  contain no variables different from  $p_1, \dots, p_m$ . Each of these instances has the form  $\psi'(\Box\Diamond\chi'_1, \dots, \Box\Diamond\chi'_n)$ , where every  $\chi'_i$  is some substitution instance of  $\chi_i$  containing only  $p_1, \dots, p_m$ . By Lemma 50 and in view of the local tabularity of **S5** (it is of depth 1), there are finitely many pairwise non-equivalent in **K4** substitution instances of  $\Box\Diamond\chi_i$  of that sort (the reader can easily estimate the number of them). So there exist only finitely many pairwise non-equivalent in **K4** substitution instances of  $\psi$  containing  $p_1, \dots, p_m$ , say  $\psi_1, \dots, \psi_k$ , and we can effectively construct them. Then, by the Deduction Theorem,

$$\varphi \in L \oplus \psi \text{ iff } \psi_1, \dots, \psi_k \vdash_L^* \varphi \text{ iff } \Box^+(\psi_1 \wedge \dots \wedge \psi_k) \rightarrow \varphi \in L$$

and so  $L \oplus \psi$  is decidable (or has FMP) whenever  $L$  is decidable (has FMP). ■

It should be noted that by adding to  $L$  with FMP infinitely many  $\Box\Diamond$ -formulas we can construct an incomplete logic. For a concrete example see [Rybakov 1977]. By adding a variable free formula to a logic in  $\text{NExtK}$  with FMP one can get a logic without FMP. However,  $\mathbf{K} \oplus \varphi$ ,  $\varphi$  variable free, has FMP, as can be easily shown by the standard filtration through the set  $\mathbf{Sub}\varphi \cup \mathbf{Sub}\psi$ , where  $\psi \notin \mathbf{K} \oplus \varphi$ . Infinitely many variable free formulas can axiomatize a normal extension of **K4** without FMP (for a concrete example see [Chagrov and Zakharyashev 1997]).

### 1.8 Subframe and cofinal subframe logics

A very useful source of information for investigating various properties of logics in  $\text{NExtK4}$  is their canonical axioms. Notice, for instance, that the canonical axioms of all logics in Table 2, save **A\*** and **K4<sub>n,m</sub>**, contain no closed domains. Canonical and negation free canonical formulas of the form  $\alpha(\mathfrak{F})$  and  $\alpha(\mathfrak{F}, \perp)$  are called *subframe* and *cofinal subframe formulas*, respectively, and logics in  $\text{NExtK4}$  axiomatizable by them are called *subframe* and *cofinal subframe logics*. The classes of such logics will be denoted by  $\mathcal{SF}$

and  $\mathcal{CSF}$ . Subframe and cofinal subframe logics in  $\text{NExt}\mathbf{K4}$  were studied by Fine [1985] and Zakharyashev [1984, 1988, 1996].

**THEOREM 52.** *All logics in  $\mathcal{SF}$  and  $\mathcal{CSF}$  have FMP.*

**Proof.** Suppose  $L = \mathbf{K4} \oplus \{\alpha(\mathfrak{F}_i, \perp) : i \in I\}$  and  $\varphi \notin L$ . By Theorem 44, without loss of generality we may assume that  $\varphi$  is a canonical formula, say,  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ . Now consider two cases. (1) For no  $i \in I$ ,  $\mathfrak{F}$  is cofinally subreducible to  $\mathfrak{F}_i$ . Then  $\mathfrak{F} \models L$ ,  $\mathfrak{F} \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ , and we are done. (2)  $\mathfrak{F}$  is cofinally subreducible to  $\alpha(\mathfrak{F}_i, \perp)$ , for some  $i \in I$ . In this case we have  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in \mathbf{K4} \oplus \alpha(\mathfrak{F}_i, \perp) \subseteq L$ , which is a contradiction. Indeed, suppose  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ . Then there is a cofinal subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$ . And since the composition of (cofinal) subreductions is again a (cofinal) subreduction,  $\mathfrak{G}$  is cofinally subreducible to  $\mathfrak{F}_i$ , which means that  $\mathfrak{G} \not\models \alpha(\mathfrak{F}_i, \perp)$ . Subframe logics are treated analogously. ■

The names “subframe logic” and “cofinal subframe logic” are explained by the following frame-theoretic characterization of these logics. A subframe  $\mathfrak{G} = \langle V, S, Q \rangle$  of a frame  $\mathfrak{F}$  is called *cofinal* if  $V \uparrow \subseteq V \downarrow$  in  $\mathfrak{F}$ . Say that a class  $\mathcal{C}$  of frames is *closed under (cofinal) subframes* if every (cofinal) subframe of  $\mathfrak{F}$  is in  $\mathcal{C}$  whenever  $\mathfrak{F} \in \mathcal{C}$ .

**THEOREM 53.**  *$L \in \text{NExt}\mathbf{K4}$  is a (cofinal) subframe logic iff it is characterized by a class of frames that is closed under (cofinal) subframes.*

**Proof.** Suppose  $L \in \mathcal{CSF}$ . We show that the class of all frames for  $L$  is closed under cofinal subframes. Let  $\mathfrak{G} \models L$  and  $\mathfrak{H}$  be a cofinal subframe of  $\mathfrak{G}$ . If  $\mathfrak{H} \not\models \alpha(\mathfrak{F}, \perp)$ , for some  $\alpha(\mathfrak{F}, \perp) \in L$ , then (since  $\mathfrak{G}$  is cofinally subreducible to  $\mathfrak{H}$ )  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \perp)$ , which is a contradiction. So  $\mathfrak{H} \models L$ .

Now suppose that  $L$  is characterized by some class of frames  $\mathcal{C}$  closed under cofinal subframes. We show that  $L = L'$ , where

$$L' = \mathbf{K4} \oplus \{\alpha(\mathfrak{F}, \perp) : \mathfrak{F} \not\models L\}.$$

If  $\mathfrak{F}$  is a finite rooted frame and  $\mathfrak{F} \not\models L$  then  $\alpha(\mathfrak{F}, \perp) \in L$ , for otherwise  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \perp)$  for some  $\mathfrak{G} \in \mathcal{C}$ , and hence there is a cofinal subframe  $\mathfrak{H}$  of  $\mathfrak{G}$  which is reducible to  $\mathfrak{F}$ ; but  $\mathfrak{H} \in \mathcal{C}$  and so, by the Reduction Theorem,  $\mathfrak{F}$  is a frame for  $L$ , which is a contradiction. Thus,  $L' \subseteq L$ . To prove the converse, suppose  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in L$ . Then  $\mathfrak{F} \not\models L$ , and hence  $\alpha(\mathfrak{F}, \perp) \in L'$ , from which  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in L'$ .

Subframe logics are considered in the same way. ■

It follows in particular that  $\mathcal{SF} \subset \mathcal{CSF}$  (**K4.1** and **K4.2** are cofinal subframe logics but not subframe ones). One can easily show also that  $\mathcal{CSF}$  is a complete sublattice of  $\text{NExt}\mathbf{K4}$  and  $\mathcal{SF}$  a complete sublattice of  $\mathcal{CSF}$ .

EXAMPLE 54. Every normal extension of **S4.3** is axiomatizable by canonical formulas which are based on chains of non-degenerate clusters and so have no closed domains. Therefore,  $\text{NExt}\mathbf{S4.3} \subset \mathcal{CSF}$ .

The classes  $\mathcal{SF}$  and  $\mathcal{CSF} - \mathcal{SF}$  contain a continuum of logics. And yet, unlike  $\text{NExt}\mathbf{K}$  or  $\text{NExt}\mathbf{K4}$ , their structure and their logics are not so complex. For instance, it is not hard to see that every logic in  $\mathcal{CSF}$  is uniquely axiomatizable by an independent set of cofinal subframe formulas and so these formulas form an axiomatic basis for  $\mathcal{CSF}$ .

The concept of subframe logic was extended in [Wolter 1993] to the class  $\text{NExt}\mathbf{K}$  by taking the frame-theoretic characterization of Theorem 53 as the definition. Namely, we say that  $L \in \text{NExt}\mathbf{K}$  is a *subframe logic* if the class of frames for  $L$  is closed under subframes. In other words, subframe logics are precisely those logics whose axioms “do not force the existence of points”. For example, **K**, **KB**, **K5**, **T**, and **Alt<sub>n</sub>** are subframe logics. To give a syntactic characterization of subframe logics we require the following formulas.

For a formula  $\varphi$  and a variable  $p$  not occurring in  $\varphi$ , define a formula  $\varphi^p$  inductively by taking

$$\begin{aligned} q^p &= q \wedge p, \quad q \text{ an atom,} \\ (\psi \odot \chi)^p &= \psi^p \odot \chi^p, \quad \text{for } \odot \in \{\wedge, \vee, \rightarrow\}, \\ (\Box\psi)^p &= \Box(p \rightarrow \psi^p) \wedge p \end{aligned}$$

and put  $\varphi^{sf} = p \rightarrow \varphi^p$ .

LEMMA 55. *For any frame  $\mathfrak{F}$ ,  $\mathfrak{F} \models \varphi^{sf}$  iff  $\varphi$  is valid in all subframes of  $\mathfrak{F}$ .*

**Proof.** It suffices to notice that if  $\mathfrak{M}$  is a model based on  $\mathfrak{F}$ ,  $\mathfrak{M}'$  a model based on the subframe of  $\mathfrak{F}$  induced by  $\{y : (\mathfrak{M}, y) \models p\}$  and  $(\mathfrak{M}, x) \models q$  iff  $(\mathfrak{M}', x) \models q$ , for all variables  $q$ , then  $(\mathfrak{M}, x) \models \varphi^p$  iff  $(\mathfrak{M}', x) \models \varphi$ . ■

PROPOSITION 56. *The following conditions are equivalent for any modal logic  $L$ :*

- (i)  $L$  is a subframe logic;
- (ii)  $L = \mathbf{K} \oplus \{\varphi^{sf} : \varphi \in \Gamma\}$ , for some set of formulas  $\Gamma$ ;
- (iii)  $L$  is characterized by a class of frames closed under subframes.

**Proof.** The implication (i)  $\Rightarrow$  (iii) is trivial; (iii)  $\Rightarrow$  (ii) and (ii)  $\Rightarrow$  (i) are consequences of Lemma 55. ■

It follows that the class of subframe logics forms a complete sublattice of  $\text{NExt}\mathbf{K}$ . However, not all of them have FMP and even are Kripke complete.

EXAMPLE 57. Let  $L$  be the logic of the frame  $\mathfrak{F}$  constructed in Example 7. Since every rooted subframe  $\mathfrak{G}$  of  $\mathfrak{F}$  is isomorphic to a generated subframe of  $\mathfrak{F}$ ,  $L$  is a subframe logic. We show that  $L$  has the same Kripke frames



as **GL.3**. Suppose  $\mathfrak{G}$  is a rooted Kripke frame for **GL.3** refuting  $\varphi \in L$ . Then clearly  $\mathfrak{G}$  contains a finite subframe  $\mathfrak{H}$  refuting  $\varphi$ . Since  $\mathfrak{H}$  is a finite chain of irreflexive points, it is isomorphic to a generated subframe of  $\mathfrak{F}$ , contrary to  $\mathfrak{F} \not\models \varphi$ . Thus  $\mathfrak{G} \models L$ . Conversely, suppose  $\mathfrak{G}$  is a Kripke frame for  $L$ . Then  $\mathfrak{G}$  is irreflexive. For otherwise  $\mathfrak{G}$  refutes the formula  $\varphi = \Box^2(\Box p \rightarrow p) \wedge \Box(\Box p \rightarrow p) \rightarrow \Box p$ , which is valid in  $\mathfrak{F}$ . Let us show now that  $\mathfrak{G}$  is transitive. Suppose otherwise. Then  $\mathfrak{G}$  refutes the formula  $\Box p \rightarrow \Box(\Box p \vee (\Box q \rightarrow q))$ , which is valid in  $\mathfrak{F}$  because  $\omega$  is a reflexive point. Finally, since  $\mathfrak{G} \models \varphi$ ,  $\mathfrak{G}$  is Noetherian and since  $\mathfrak{F}$  is of width 1, we may conclude that  $\mathfrak{G} \models \mathbf{GL.3}$ . It follows that the subframe logic  $L$  is Kripke incomplete. Indeed, it shares the same class of Kripke frames with **GL.3** but  $\Box p \rightarrow \Box \Box p \in \mathbf{GL.3} - L$ .

The following theorem provides a frame-theoretic characterization of those complete subframe logics in **NExtK** that are elementary,  $\mathcal{D}$ -persistent and strongly complete. Say that a logic  $L$  has the *finite embedding property* if a Kripke frame  $\mathfrak{F}$  validates  $L$  whenever all finite subframes of  $\mathfrak{F}$  are frames for  $L$ .

**THEOREM 58** (Fine 1985). *For each Kripke complete subframe logic  $L$  the following conditions are equivalent:*

- (i)  $L$  is universal;<sup>9</sup>
- (ii)  $L$  is elementary;
- (iii)  $L$  is  $\mathcal{D}$ -persistent;
- (iv)  $L$  is strongly Kripke complete;
- (v)  $L$  has the finite embedding property.

**Proof.** The implications (i)  $\Rightarrow$  (ii) and (iii)  $\Rightarrow$  (iv) are trivial; (ii)  $\Rightarrow$  (iii) follows from Fine's [1975b] Theorem formulated in Section 1.3 and (v)  $\Rightarrow$  (i) from [Tarski 1954]. Thus it remains to show that (iv)  $\Rightarrow$  (v). Suppose  $\mathfrak{F}$  is a Kripke frame with root  $r$  such that  $\mathfrak{F} \not\models L$  but all finite subframes of  $\mathfrak{F}$  validate  $L$ . Then it is readily checked that all finite subsets of  $\Gamma = \{p_r\} \cup \Box^{<\omega} \Delta_{\mathfrak{F}}$  are  $L$ -consistent. Hence the whole set  $\Gamma$  is  $L$ -consistent. On the other hand, similarly to the proof of Lemma 13 one can show that  $\Gamma$  is satisfiable in a Kripke frame iff the frame is subreducible to  $\mathfrak{F}$ . So  $\Gamma$  cannot be satisfied in a Kripke frame for  $L$  and  $L$  is not strongly complete. ■

A similar criterion for the cofinal subframe logics in **NExtK4** can be found in [Zakharyashev 1996]. Note, however, that they are not in general universal and certainly do not have the finite embedding property, but (ii), (iii) and (iv) are still equivalent.

**PROPOSITION 59.** *Every subframe logic  $L \in \mathbf{NExtAlt}_n$  has FMP.*

<sup>9</sup>I.e., universal is the class of Kripke frames for  $L$  considered as models of the first order language with  $R$  and  $=$ .

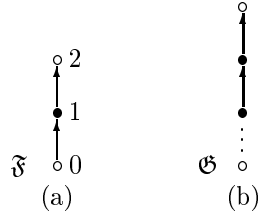


Figure 5.

**Proof.** Suppose  $\varphi \notin L$ . By Theorem 22, there is a Kripke frame  $\mathfrak{F}$  for  $L$  refuting  $\varphi$  at a point  $x$ . Denote by  $X$  the set of points in  $\mathfrak{F}$  accessible from  $x$  by  $\leq md(\varphi)$  steps. Clearly,  $X$  is finite and the subframe of  $\mathfrak{F}$  induced by  $X$  validates  $L$  and refutes  $\varphi$ . ■

To understand the place of incomplete logics in the lattice of subframe logics we call a subframe logic  $L$  *strictly sf-complete* if it is Kripke complete and no other subframe logic has the same Kripke frames as  $L$ . Example 57 shows that **GL.3** is not strictly sf-complete. However, the logics **T**, **S4** and **Grz** turn out to be strictly sf-complete. The following result clarifies the situation. It is proved by applying the splitting technique to lattices of subframe logics.

**THEOREM 60.** *A subframe logic  $L$  containing **K4** is strictly sf-complete iff  $L \not\subseteq \mathbf{GL.3}$ . All subframe logics in  $\mathbf{NExtAlt}_n$  are strictly sf-complete. A subframe logic is tabular iff there are only finitely many subframe logics containing it.*

### 1.9 More sufficient conditions of FMP

As follows from Theorem 52, a logic in  $\mathbf{NExtK4}$  does not have FMP only if at least one of its canonical axioms contains closed domains. We illustrate their role by a simple example.

**EXAMPLE 61.** Consider the logic  $L = \mathbf{K4.3} \oplus \alpha^\sharp(\mathfrak{F}, \perp)$  and the formula  $\alpha(\mathfrak{F}, \perp)$ , where  $\mathfrak{F}$  is the frame depicted in Fig. 5 (a). The frame  $\mathfrak{G}$  in Fig. 5 (b) separates  $\alpha(\mathfrak{F}, \perp)$  from  $L$ . Indeed,  $\mathfrak{F}$  is a cofinal subframe of  $\mathfrak{G}$  and so  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \perp)$ . To show that  $\mathfrak{G} \models \alpha^\sharp(\mathfrak{F}, \perp)$ , suppose  $f$  is a cofinal subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$ . Then  $f^{-1}(1)$  contains only one point, say  $x$ ;  $f^{-1}(0)$  also contains only one point, namely the root of  $\mathfrak{G}$ . So the infinite set of points between  $x$  and the root is outside  $\text{dom}f$ , which means that  $f$  does not satisfy (CDC) for  $\{\{1\}\}$ . On the other hand, if  $\mathfrak{H}$  is a finite refutation frame of width 1 for  $\alpha(\mathfrak{F}, \perp)$  then  $\mathfrak{H}$  contains a generated subframe reducible to  $\mathfrak{F}$ , from which  $\mathfrak{H} \not\models L$ . Thus,  $L$  fails to have FMP. In the same manner the reader can prove that **A\*** in Table 2 does not have FMP either.

We show now two methods developed in [Zakharyashev 1997a] for establishing FMP of logics whose canonical axioms contain closed domains. One of them uses the following lemma, which is an immediate consequence of the refutability criterion for the canonical formulas.

**LEMMA 62.** *Suppose  $\alpha(\mathfrak{F}, \mathfrak{D})$  and  $\alpha(\mathfrak{G}, \mathfrak{E})$  ( $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  and  $\alpha(\mathfrak{G}, \mathfrak{E}, \perp)$ ) are canonical formulas such that there is a (cofinal) subreduction  $f$  of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$  and an antichain  $\epsilon \subseteq \text{dom} f \uparrow$  is in  $\mathfrak{E}$  whenever  $f(\epsilon \uparrow) = \mathfrak{d} \uparrow$  for some  $\mathfrak{d} \in \mathfrak{D}$ . Then  $\alpha(\mathfrak{G}, \mathfrak{E}) \in \mathbf{K4} \oplus \alpha(\mathfrak{F}, \mathfrak{D})$  (respectively,  $\alpha(\mathfrak{G}, \mathfrak{E}, \perp) \in \mathbf{K4} \oplus \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ ).*

**THEOREM 63.**  *$L = \mathbf{K4} \oplus \{\alpha(\mathfrak{F}_i, \mathfrak{D}_i, \perp) : i \in I\} \oplus \{\alpha(\mathfrak{F}_j, \mathfrak{D}_j) : j \in J\}$  has FMP provided that either all frames  $\mathfrak{F}_i$ , for  $i \in I \cup J$ , are irreflexive or all of them are reflexive.*

**Proof.** Suppose all  $\mathfrak{F}_i$  are irreflexive and  $\alpha(\mathfrak{G}, \mathfrak{E}, \perp)$  is an arbitrary canonical formula. We construct from  $\mathfrak{G}$  a new finite frame  $\mathfrak{H}$  by inserting into it new *reflexive* points. Namely, suppose  $\epsilon$  is an antichain in  $\mathfrak{G}$  such that  $\epsilon \not\subseteq \mathfrak{E}$ . Suppose also that  $C_1, \dots, C_n$  are all clusters in  $\mathfrak{G}$  such that  $\epsilon \subseteq C_i \uparrow$  and  $\epsilon \cap C_i = \emptyset$ , for  $i = 1, \dots, n$ , but no successor of  $C_i$  possesses this property. Then we insert in  $\mathfrak{G}$  new reflexive points  $x_1, \dots, x_n$  so that each  $x_i$  could see only the points in  $\epsilon$  and their successors and could be seen only from the points in  $C_i$  and their predecessors. The same we simultaneously do for all antichains  $\epsilon$  in  $\mathfrak{G}$  of that sort. The resulting frame is denoted by  $\mathfrak{H}$ . Since no new point was inserted just below an antichain in  $\mathfrak{E}$ ,  $\mathfrak{H} \not\models \alpha(\mathfrak{G}, \mathfrak{E}, \perp)$ .

Suppose now that  $\alpha(\mathfrak{G}, \mathfrak{E}, \perp) \notin L$  and show that  $\mathfrak{H} \models L$ . If this is not so then either  $\mathfrak{H} \not\models \alpha(\mathfrak{F}_i, \mathfrak{D}_i, \perp)$ , for some  $i \in I$ , or  $\mathfrak{H} \not\models \alpha(\mathfrak{F}_j, \mathfrak{D}_j)$ , for some  $j \in J$ . We consider only the former case, since the latter one is treated similarly. Thus, we have a cofinal subreduction  $f$  of  $\mathfrak{H}$  to  $\mathfrak{F}_i$  satisfying (CDC) for  $\mathfrak{D}_i$ . Since  $\mathfrak{F}_i$  is irreflexive, no point that was added to  $\mathfrak{G}$  is in  $\text{dom} f$ . So  $f$  may be regarded as a cofinal subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}_i$  satisfying (CDC) for  $\mathfrak{D}_i$ . We clearly may assume also that the subframe of  $\mathfrak{G}$  generated by  $\text{dom} f$  is rooted. Let  $\epsilon$  be an antichain in  $\mathfrak{G}$  belonging to  $\text{dom} f \uparrow$  and such that  $f(\epsilon \uparrow) = \mathfrak{d} \uparrow$  for some  $\mathfrak{d} \in \mathfrak{D}_i$ . If  $\epsilon \not\subseteq \mathfrak{E}$  then there is a reflexive point  $x$  in  $\mathfrak{H}$  such that  $x \in \text{dom} f \uparrow$  and  $x$  sees only  $\epsilon \uparrow$  and, of course, itself. But then  $f(x \uparrow) = f(\epsilon \uparrow) = \mathfrak{d} \uparrow$  and so, by (CDC),  $x \in \text{dom} f$ , which is impossible. Therefore,  $\epsilon \subseteq \mathfrak{E}$  and so, by Lemma 62,  $\alpha(\mathfrak{G}, \mathfrak{E}, \perp) \in L$ , contrary to our assumption.

In the case of reflexive frames *irreflexive* points are inserted. ■

**EXAMPLE 64.** According to Theorem 63, the logic

$$L = \mathbf{K4} \oplus \alpha\left( \begin{array}{c} 1 \bullet \quad \bullet 2 \\ \searrow \quad \nearrow \\ \bullet \end{array}, \{\{1\}, \{1, 2\}\} \right)$$

has FMP. However, Artemov's logic  $\mathbf{A}^* = L \oplus \mathbf{GL}$  does not enjoy this property. So FMP is not in general preserved under sums of logics.

The scope of the method of inserting points is not bounded only by canonical axioms associated with homogeneous (irreflexive or reflexive) frames. It can be applied, for instance, to normal extensions of  $\mathbf{K4}$  with modal reduction principles, i.e., formulas of the form  $\mathbf{M}p \rightarrow \mathbf{N}p$ , where  $\mathbf{M}$  and  $\mathbf{N}$  are strings of  $\Box$  and  $\Diamond$  (for first order equivalents of modal reduction principles see [van Benthem 1976]). One can show that each such logic is either of finite depth, or can be axiomatized by  $\Box\Diamond$ -formulas and canonical formulas based upon almost homogeneous frames (containing at most one reflexive point), for which the method works as well. So we have

**THEOREM 65.** *All logics in  $\text{NExt}\mathbf{K4}$  axiomatizable by modal reduction principles have FMP and are decidable.*

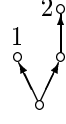
One of the most interesting open problems in completeness theory of modal logic is to prove an analogous theorem for logics in  $\text{NExt}\mathbf{K}$  or to construct a counter-example. It is unknown, in particular, whether the logics of the form  $\mathbf{K} \oplus \Box^m p \rightarrow \Box^n p$  have FMP; the same concerns the logics  $\mathbf{K} \oplus \text{tra}_n$ .

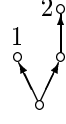
The second method of proving FMP uses the more conventional technique of removing points. Suppose that  $L = \mathbf{K4} \oplus \{\alpha(\mathfrak{G}_i, \mathfrak{D}_i, \perp) : i \in I\}$  and  $\alpha = \alpha(\mathfrak{H}, \mathfrak{E}, \perp) \notin L$ . Then there exists a frame  $\mathfrak{F}$  for  $L$  such that  $\mathfrak{F} \not\models \alpha$ , i.e., there is a cofinal subreduction  $h$  of  $\mathfrak{F}$  to  $\mathfrak{H}$  satisfying (CDC) for  $\mathfrak{E}$ . Construct the countermodel  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  for  $\alpha$  as it was done in Section 1.6. Without loss of generality we may assume that  $\text{dom}h\uparrow = \text{dom}h\downarrow = \mathfrak{F}$  and that  $\mathfrak{F}$  is generated by the sets  $\mathfrak{V}(p_i)$ ,  $p_i$  a variable in  $\alpha$ .

Actually, the step-wise refinement procedure with deleting points having **Sub** $\alpha$ -equivalent successors, used in the proof of Theorem 42, establishes FMP of  $L$  when all  $\mathfrak{D}_i$  are empty, i.e.,  $L$  is a cofinal subframe logic. To tune it for  $L$  with non-empty  $\mathfrak{D}_i$ , we should follow a subtler strategy of deleting points, preserving those that are “responsible” for validating the axioms of  $L$ . Suppose we have already constructed a model  $\mathfrak{M}'_n = \langle \mathfrak{F}'_n, \mathfrak{V}'_n \rangle$  by “folding up”  $n - 1$ -cyclic sets into clusters of depth  $n$  (we use the same notations as in the proof of Theorem 42). Now we throw away points of two sorts.

First, for every proper cluster  $C$  of depth  $n$  such that some  $x \in C$  has a **Sub** $\alpha$ -equivalent successor of depth  $< n$ , we remove from  $C$  all points except  $x$ . Second, call a point  $x$  of depth  $> n$  *redundant* in  $\mathfrak{M}'_n$  if it has a **Sub** $\alpha$ -equivalent successor of depth  $\leq n$  and, for every  $i \in I$  and every cofinal subreduction  $g$  of  $(\mathfrak{F}'_n)^{\leq n}$  to the subframe of  $\mathfrak{G}_i$  generated by some  $\mathfrak{d} \in \mathfrak{D}_i$  such that  $\mathfrak{d} \subseteq g(x\uparrow)$  and  $g$  satisfies (CDC) for  $\mathfrak{D}_i$ , there is a point  $y \in x\uparrow$  of depth  $\leq n$  such that  $g(y\uparrow) = \mathfrak{d}\uparrow$ . Let  $X$  be the maximal set of redundant points in  $\mathfrak{M}'_n$  which is upward closed in  $(W'_n)^{>n}$ . We define  $\mathfrak{M}_{n+1} = \langle \mathfrak{F}_{n+1}, \mathfrak{V}_{n+1} \rangle$  as the submodel of  $\mathfrak{M}'_n$  resulting from it by removing all points in  $X$  as well. Since all deleted points have **Sub** $\alpha$ -equivalent successors,  $\mathfrak{M}_{n+1} \not\models \alpha$ . And since we keep in  $\mathfrak{F}_{n+1}$  points which

violate (CDC) for  $\mathfrak{D}_i$  of possible cofinal subreductions to  $\mathfrak{G}_i$ ,  $\mathfrak{F}_{n+1} \models L$ . So FMP of  $L$  will be established if we manage to prove that this process eventually terminates.



EXAMPLE 66. Let  $L = \mathbf{S4} \oplus \alpha(\mathfrak{G}, \{\{1, 2\}\}, \perp)$ , where  $\mathfrak{G}$  is , and assume that our “algorithm”, when being applied to  $\mathfrak{F}$ ,  $\alpha$  and  $L$ , works infinitely long. Then the frame  $\mathfrak{F}_\omega = \langle W_\omega, R_\omega \rangle$ , where

$$W_\omega = \bigcup_{0 < i < \omega} W_i^{\leq i}, \quad R_\omega = \bigcup_{0 < i < \omega} R_i^{\leq i}, \quad \mathfrak{F}_i = \langle W_i, R_i, P_i \rangle,$$

is of infinite depth. By König’s Lemma, there is an infinite descending chain  $\dots x_i R_\omega x_{i-1} \dots R_\omega x_2 R_\omega x_1$  in  $\mathfrak{F}_\omega$  such that  $x_i$  is of depth  $i$ . Since there are only finitely many pairwise non-**Sub** $\alpha$ -equivalent points, there must be some  $n > 0$  such that, for every  $k \geq n$ , each point in  $C(x_k)$  has a **Sub** $\alpha$ -equivalent successor in  $\mathfrak{F}_k^{\leq k}$ . And since  $\mathfrak{F}_1^{\leq 1}$  is finite, there is  $m \geq n$  starting from which all  $x_i$  see the same points of depth 1. Let us consider now  $\mathfrak{F}_m$  and ask why points in the  $m$ -cyclic set  $X$ , folded at step  $m + 1$  into  $C(x_{m+1})$ , were not removed at step  $m$ .  $X$  is upward closed in  $W_m^{> m}$  and every point in it has a **Sub** $\alpha$ -equivalent successor in  $\mathfrak{F}_m^{\leq m}$ . So the only reason for keeping some  $x \in X$  is that  $\mathfrak{F}_m^{\leq m}$  is cofinally subreducible to  $\mathfrak{G}^{\leq 1}$ ,  $x$  sees inverse images of both points in  $\mathfrak{G}^{\leq 1}$  but none of its successors in  $\mathfrak{F}_m^{\leq m}$  does. By the cofinality condition, these inverse images can be taken from  $\mathfrak{F}_1^{\leq 1}$ . But then they are also seen from  $x_m$ , which is a contradiction. Thus sooner or later our algorithm will construct a finite frame separating  $L$  from  $\alpha$ , which proves that  $L$  has FMP.

The reason why we succeeded in this example is that inverse images of points in the closed domain  $\{1, 2\}$  can be found at a fixed finite depth in  $\mathfrak{F}_\omega$ , and so points violating (CDC) for it can also be found at finite depth (that was not the case in Example 61). The following definitions describe a big family of frames and closed domains of that sort.

A point  $x$  in a frame  $\mathfrak{G}$  is called a *focus* of an antichain  $\mathfrak{a}$  in  $\mathfrak{G}$  if  $x \notin \mathfrak{a}$  and  $x \uparrow = \{x\} \cup \mathfrak{a} \uparrow$ . Suppose  $\mathfrak{G}$  is a finite frame and  $\mathfrak{D}$  a set of antichains in  $\mathfrak{G}$ . Define by induction on  $n$  notions of *n-stable point* in  $\mathfrak{G}$  (relative to  $\mathfrak{D}$ ) and *n-stable antichain* in  $\mathfrak{D}$ . A point  $x$  is *1-stable* in  $\mathfrak{G}$  iff either  $x$  is of depth 1 in  $\mathfrak{G}$  or the cluster  $C(x)$  is proper. A point  $x$  is *n + 1-stable* in  $\mathfrak{G}$  (relative to  $\mathfrak{D}$ ) iff it is not *m-stable*, for any  $m \leq n$ , and either there is an *n-stable point* in  $\mathfrak{G}$  (relative to  $\mathfrak{D}$ ) which is not seen from  $x$  or  $x$  is a focus of an antichain in  $\mathfrak{D}$  containing an *n – 1-stable point* and no *n-stable point*. And we say an antichain  $\mathfrak{d}$  in  $\mathfrak{D}$  is *n-stable* iff it contains an *n-stable point* in the subframe  $\mathfrak{G}'$  of  $\mathfrak{G}$  generated by  $\mathfrak{d}$  (relative to  $\mathfrak{D}$ ) and no *m-stable point* in  $\mathfrak{G}'$  (relative to  $\mathfrak{D}$ ), for  $m > n$ . A point or an antichain is *stable* if

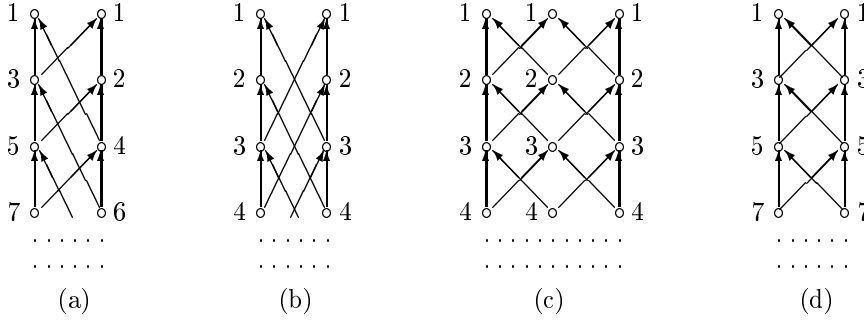


Figure 6.

it is  $n$ -stable for some  $n$ . It should be clear that if a point in an antichain is stable then the rest points in the antichain are also stable.

EXAMPLE 67.

- (1) Suppose  $\mathfrak{G}$  is a finite rooted generated subframe of one of the frames shown in Fig. 6 (a)–(c). Then, regardless of  $\mathfrak{D}$ , each point in  $\mathfrak{G}$  different from its root is  $n$ -stable, where  $n$  is the number located near the point. Every antichain  $\mathfrak{d}$  in  $\mathfrak{G}$ , containing at least two points, is also  $n$ -stable, with  $n$  being the maximal degree of stability of points in  $\mathfrak{d}$ .
- (2) If  $\mathfrak{G}$  is a rooted generated subframe of the frame depicted in Fig. 6 (d) and  $\mathfrak{D}$  is the set of all two-point antichains in  $\mathfrak{G}$  then every point in  $\mathfrak{G}$  is  $n$ -stable (relative to  $\mathfrak{D}$ ), where  $n$  stays near the point. However, for  $\mathfrak{D} = \emptyset$  no point in  $\mathfrak{G}$ , save those of depth 1, is stable.
- (3) If  $\mathfrak{G}$  is a finite tree of clusters then every antichain in  $\mathfrak{G}$ , different from a non-final singleton, is either 1- or 2-stable in  $\mathfrak{G}$  regardless of  $\mathfrak{D}$ . Every antichain containing a point  $x$  with proper  $C(x)$  is 1- or 2-stable as well, whatever  $\mathfrak{G}$  and  $\mathfrak{D}$  are.
- (4) Every antichain is stable in every irreflexive frame  $\mathfrak{G}$  relative to the set  $\mathfrak{D}^\#$  of all antichains in  $\mathfrak{G}$ . However, this is not so if  $\mathfrak{G}$  contains reflexive points (for reflexive singletons are open domains and do not belong to  $\mathfrak{D}^\#$ ).

The sufficient condition of FMP below is proved by arguments that are similar to those we used in Example 66.

**THEOREM 68.** *If  $L = \mathbf{K4} \oplus \{\alpha(\mathfrak{G}_i, \mathfrak{D}_i, \perp) : i \in I\}$  and there is  $d > 0$  such that, for any  $i \in I$ , every closed domain  $\mathfrak{d} \in \mathfrak{D}_i$  is  $n$ -stable in  $\mathfrak{G}_i$  (relative to  $\mathfrak{D}_i$ ), for some  $n \leq d$ , then  $L$  has FMP.*

Example 67 shows many applications of this condition. Moreover, using it one can prove the following

**THEOREM 69.** *Every normal extension of  $\mathbf{S4}$  with a formula in one variable has FMP and is decidable.*

Note that, as was shown by Shehtman [1980], a formula with two variables or an infinite set of one-variable formulas can axiomatize logics in  $\text{NExt}\mathbf{S4}$  without FMP (and even Kripke incomplete).

### 1.10 The reduction method

That a logic does not have FMP (or is Kripke incomplete) is not yet an evidence of its undecidability: it is enough to recall that the majority of decidability results for classical theories was proved without using any analogues of the finite model property (see e.g. [Rabin 1977], [Ershov 1980]). The first example of a decidable finitely axiomatizable modal logic without FMP was constructed by Gabbay [1971].

It seems unlikely that the methods of classical model theory can be applied directly for proving the decidability of propositional modal logics. However, sometimes it is possible to *reduce* the decision problem for a given modal logic  $L$  to that for a knowingly decidable first or higher order theory whose language is expressive enough for describing the structure of frames characterizing  $L$ . The most popular tools used for this purpose are Büchi's [1962] Theorem on the decidability of the weak monadic second order theory of the successor function on natural numbers and Rabin's [1969] Tree Theorem. Below we illustrate the use of Rabin's Theorem following [Gabbay 1975] and [Cresswell 1984].

Let  $\omega^*$  be the set of all finite sequences of natural numbers and  $\prec$  the lexicographic order on it. For  $x \in \omega^*$  and  $i < \omega$ , put  $r_i(x) = x * i$ , where  $*$  denotes the usual concatenation operation. Besides, define the following predicates  $<_i$  on  $\omega^*$ , for  $0 \leq i \leq 2$ ,

$$x <_i y \text{ iff } y = x * (3n + i) \text{ for some } n < \omega.$$

It follows from [Rabin 1969] that the monadic second order theory  $\text{S}\omega\text{S}$  of the model  $\langle \omega^*, \{r_i : i < \omega\}, \{<_i : 0 \leq i \leq 2\}, \prec, \emptyset \rangle$  ( $\emptyset$  denotes the empty sequence) is decidable.

The theory  $\text{S}\omega\text{S}$  has a very strong expressive power which makes it possible to effectively describe semantical definitions of many modal (as well as some other) logics and thereby prove their decidability. In this way Gabbay [1975] established the decidability of, for instance,

$$\mathbf{K} \oplus \Box^m \Diamond p \rightarrow \Diamond p, \quad \mathbf{K} \oplus \Diamond^m \Box p \rightarrow \Box p,$$

$$\mathbf{K} \oplus \Box^m p \rightarrow \Diamond^n p, \quad \mathbf{K} \oplus \Diamond^m p \rightarrow \Box^n p.$$

By Sahlqvist's Theorem, all these logics are Kripke complete; however, we do not know whether they have FMP. General frames can also be described by means of  $\text{S}\omega\text{S}$ .

EXAMPLE 70. The frame  $\mathfrak{F} = \langle W, R, P \rangle$  constructed in Example 7 can be represented in the language of  $\mathcal{S}\omega\mathcal{S}$  as follows. Let us encode each  $n < \omega$  by the sequence  $\langle 3n \rangle$ , while  $\omega$  and  $\omega + 1$  by  $r_1(\emptyset)$  and  $r_2(\emptyset)$ , respectively. Then we have

$$\begin{aligned} x \in W & \text{ iff } \emptyset <_0 x \vee x = r_1(\emptyset) \vee x = r_2(\emptyset), \\ xRy & \text{ iff } (\emptyset <_0 x \wedge \emptyset <_0 y \wedge y \prec x \wedge x \neq y) \vee \\ & (x = r_1(\emptyset) \wedge \emptyset <_0 y) \vee x = y = r_1(\emptyset) \vee \\ & (x = r_2(\emptyset) \wedge y = r_1(\emptyset)), \\ X \in P & \text{ iff } \forall x (x \in X \rightarrow x \in W) \wedge ((Fin(X) \wedge r_1(\emptyset) \notin X) \vee \\ & \forall Y (\forall y (y \in Y \leftrightarrow (y \in W \wedge y \notin X)) \rightarrow Fin(Y) \wedge r_1(\emptyset) \notin Y)), \end{aligned}$$

where  $x = y$  means  $x \prec y \wedge y \prec x$  and

$$Fin(X) = \exists x \forall y (y \in X \rightarrow y \prec x).$$

It follows that the logic  $\text{Log}\mathfrak{F}$  is decidable. Indeed, for every formula  $\varphi(p_1, \dots, p_n)$ , we have  $\varphi \in \text{Log}\mathfrak{F}$  iff the second order formula

$$\forall x \forall X_1, \dots, X_n (X_1 \in P \wedge \dots \wedge X_n \in P \wedge x \in W \rightarrow ST(\varphi(X_1, \dots, X_n)))$$

belongs to  $\mathcal{S}\omega\mathcal{S}$ . Here  $ST(\varphi(X_1, \dots, X_n))$ , the *standard translation* of  $\varphi$ , is defined inductively in the following way (see also *Correspondence Theory*):

$$ST(X) = x \in X, \quad ST(\perp) = \perp,$$

$$ST(X \odot Y) = ST(X) \odot ST(Y), \text{ for } \odot \in \{\wedge, \vee, \rightarrow\},$$

$$ST(\Box X) = \forall y (xRy \rightarrow ST(X)\{y/x\}).$$

Recall that, as was shown in Example 57,  $\text{Log}\mathfrak{F}$  is Kripke incomplete.

Also, it is not hard to find examples of applications of this technique for proving the decidability of finitely axiomatizable quasi-normal unimodal and normal polymodal (in particular, tense) logics which do not have Kripke frames at all; perhaps, the simplest one is Solovay's logic **S**.

Sobolev [1977a] found another way of proving decidability by applying methods of automata theory on infinite sequences. Using the results of [Büchi and Siefkes 1973] he showed that all finitely axiomatizable superintuitionistic logics of finite width (see Section 3.4) containing the formula

$$(((p \rightarrow q) \rightarrow p) \rightarrow p) \vee (((q \rightarrow p) \rightarrow q) \rightarrow q).$$

are decidable. By the preservation theorem of Section 3.3, this result can be transferred to the corresponding extensions of **S4**.

If a logic is known to be complete with respect to a suitable class of frames, the methods discussed above are usually applicable to it in a rather



straightforward manner. A relative disadvantage of this approach is that the resulting decision algorithms inherit the extremely high complexity of the decision algorithms for  $S\omega S$  or other “rich theories” used to prove decidability. On the other hand, the logic  $\mathbf{S}$ , for instance, turns out to be decidable by an algorithm of the same complexity as that for  $\mathbf{GL}$  (see Example 75), in particular, the derivability problem in  $\mathbf{S}$  is *PSPACE*-complete. The logic of the frame  $\mathfrak{F}$  in Example 7 is “almost trivial”—it is polynomially equivalent to classical propositional logic, which follows from the fact that every formula  $\varphi$  refutable by  $\mathfrak{F}$  can be also refuted in  $\mathfrak{F}$  under a valuation giving the same truth-value to all variables in  $\varphi$  at all points  $i$  such that  $|\mathbf{Sub}\varphi| < i < \omega$  (see Section 4.6). Actually, this sort of decidability proofs (ignoring “inessential” parts of infinite frames) was used already by Kuznetsov and Gerchiu [1970] for studying some superintuitionistic logics.

Recently more general semantical methods of obtaining decidability results without turning to “rich theories” have been developed. We demonstrate them in the next section by establishing the decidability of all finitely axiomatizable logics in  $\mathbf{NExtK4.3}$ , which according to Example 61 do not in general have FMP. We show, however, that those logics are complete with respect to recursively enumerable classes of recursive frames in which the validity of formulas can be effectively checked—it was this rather than the finiteness of frames that we used in the proof of Harrop’s Theorem. In Section 2.5 this result will be extended to linear tense logics which in general are not even Kripke complete. Our presentation follows [Zakharyashev and Alekseev 1995].

### 1.11 Logics containing $\mathbf{K4.3}$

Each logic in  $L \in \mathbf{NExtK4.3}$  is represented in the form

$$L = \mathbf{K4.3} \oplus \{\alpha(\mathfrak{F}_i, \mathfrak{D}_i, \perp) : i \in I\},$$

where all  $\mathfrak{F}_i$  are chains of clusters. So our decidability problem reduces to finding an algorithm which, given such a representation with finite  $I$  and a canonical formula  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  built on a chain of clusters  $\mathfrak{F}$ , could decide whether  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in L$ . Recall also that, by Fine’s [1974c] Theorem, logics of width 1 are characterized by Kripke frames having the form of Noetherian chains of clusters.

**LEMMA 71.** *For any Noetherian chain of clusters  $\mathfrak{G}$  and any canonical formula  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ ,  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  iff there is an injective<sup>10</sup> cofinal subreduction  $g$  of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ .*

**Proof.** If  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  then there is a cofinal subreduction  $f$  of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ . Clearly,  $f^{-1}(x)$  is a singleton if  $x$  is irreflexive.

<sup>10</sup>That is  $g(x) \neq g(y)$ , for every distinct  $x, y \in \text{dom}g$ .

Suppose now that  $x$  is a reflexive point in  $\mathfrak{F}$ . Since  $\mathfrak{G}$  contains no infinite ascending chains,  $f^{-1}(x)$  has a finite cover and so there is a reflexive point  $u_x \in f^{-1}(x)$  such that  $f^{-1}(x) \subseteq u_x \downarrow$ . Fix such a  $u_x$  for each reflexive  $x$  and define a partial map  $g$  by taking

$$g(y) = \begin{cases} f(y) & \text{if either } f(y) \text{ is irreflexive or} \\ & f(y) \text{ is reflexive and } y = u_{f(y)} \\ \text{undefined} & \text{otherwise.} \end{cases}$$

One can readily check that  $g$  is the injective cofinal subreduction we need. The converse is trivial.  $\blacksquare$

Roughly, every Noetherian chain of clusters refuting  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  results from  $\mathfrak{F}$  by inserting some Noetherian chains of clusters just below clusters  $C(x)$  in  $\mathfrak{F}$  such that  $\{x\} \notin \mathfrak{D}$ . We show now that if  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  is not in  $L \in \text{NExtK4.3}$  then it can be separated from  $L$  by a frame constructed from  $\mathfrak{F}$  by inserting in open domains between its adjacent clusters either finite descending chains of irreflexive points possibly ending with a reflexive one or infinite descending chains of irreflexive points.

Let  $C(x_0), \dots, C(x_n)$  be all distinct clusters in  $\mathfrak{F}$  ordered in such a way that  $C(x_0) \subset C(x_1) \downarrow \subset \dots \subset C(x_n) \downarrow$ . Say that an  $n$ -tuple  $t = \langle \xi_1, \dots, \xi_n \rangle$  is a *type* for  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  if either  $\xi_i = m$  or  $\xi_i = m+$ , for some  $m < \omega$ , or  $\xi_i = \omega$ , with  $\xi_i = 0$  if  $\{x_i\} \in \mathfrak{D}$ . Given a type  $t = \langle \xi_1, \dots, \xi_n \rangle$  for  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ , we define the *t-extension* of  $\mathfrak{F}$  to be the frame  $\mathfrak{G}$  that is obtained from  $\mathfrak{F}$  by inserting between each pair  $C(x_{i-1}), C(x_i)$  either a descending chain of  $m$  irreflexive points, if  $\xi_i = m < \omega$ , or a descending chain of  $m + 1$  points of which only the last (lowest) one is reflexive, if  $\xi_i = m+$ , or an infinite descending chain of irreflexive points, if  $\xi_i = \omega$ . It should be clear that  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ .

**LEMMA 72.** *If  $L \in \text{NExtK4.3}$  and  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \notin L$  then  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  is separated from  $L$  by the  $t$ -extension of  $\mathfrak{F}$ , for some type  $t$  for  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ .*

**Proof.** By Lemma 71, we have a Noetherian chain of clusters  $\mathfrak{G}$  for  $L$  and an injective cofinal subreduction  $f$  of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ . By the Generation Theorem, we may assume that  $f$  maps the root of  $\mathfrak{G}$  to the root of  $\mathfrak{F}$ . Let  $\mathfrak{G}_0$  be the subframe of  $\mathfrak{G}$  obtained by removing from  $\mathfrak{G}$  all those points that are not in  $\text{dom} f$  but belong to clusters containing some points in  $\text{dom} f$ . The very same map  $f$  is an injective cofinal subreduction of  $\mathfrak{G}_0$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ , and so  $\mathfrak{G}_0 \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ . Since  $\mathfrak{G}_0$  is a reduct of  $\mathfrak{G}$ ,  $\mathfrak{G}_0 \models L$ .

Let  $C(x_0), \dots, C(x_n)$  be all distinct clusters in  $\mathfrak{G}_0$  such that

$$\text{dom} f = \bigcup_{i=0}^n C(x_i), \quad C(x_0) \subset C(x_1) \downarrow \subset \dots \subset C(x_n) \downarrow.$$

By induction on  $i$  we define a sequence of frames  $\mathfrak{G}_0 \supseteq \dots \supseteq \mathfrak{G}_n$  such that (a)  $f$  is an injective cofinal subreduction of  $\mathfrak{G}_i$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ , (b) between  $C(x_{i-1})$  and  $C(x_i)$  the frame  $\mathfrak{G}_i$  contains either a finite descending chain of irreflexive points possibly ending with a reflexive one or an infinite descending chain of irreflexive points, and (c)  $\mathfrak{G}_i \models L$ .

Suppose  $\mathfrak{G}_{i-1}$  has been already constructed and  $\mathfrak{C}_i$  is the chain of clusters located between  $C(x_{i-1})$  and  $C(x_i)$ . Three cases are possible. (1)  $\mathfrak{C}_i$  is a finite chain of irreflexive points. Then we put  $\mathfrak{G}_i = \mathfrak{G}_{i-1}$ . (2)  $\mathfrak{C}_i$  contains a non-degenerate cluster  $C(x)$  having finitely many distinct successors in  $\mathfrak{C}_i$  and all of them are irreflexive. Then  $\mathfrak{G}_i$  results from  $\mathfrak{G}_{i-1}$  by removing from  $\mathfrak{C}_i$  all points save  $x$  and those successors.  $\mathfrak{G}_i$  is a reduct of  $\mathfrak{G}_{i-1}$  and so conditions (a)–(c) are satisfied. (3) Suppose (1) and (2) do not hold. Then  $\mathfrak{C}_i$  contains an infinite descending chain  $Y$  of irreflexive points accessible from all other points in  $\mathfrak{C}_i$ . In this case  $\mathfrak{G}_i$  is obtained from  $\mathfrak{G}_{i-1}$  by removing all points in  $\mathfrak{C}_i$  save those in  $Y$ . Clearly,  $\mathfrak{G}_i$  satisfies (a) and (b). To prove (c) suppose  $\mathfrak{G}_i \not\models \alpha(\mathfrak{H}, \mathfrak{E}, \perp)$  for some  $\alpha(\mathfrak{H}, \mathfrak{E}, \perp) \in L$ . Then there is an injective cofinal subreduction  $g$  of  $\mathfrak{G}_i$  to  $\mathfrak{H}$  satisfying (CDC) for  $\mathfrak{E}$ . Consider  $g$  as a cofinal subreduction of  $\mathfrak{G}_{i-1}$  to  $\mathfrak{H}$  and show that it also satisfies (CDC) for  $\mathfrak{E}$ . Indeed, (CDC) could be violated only by a point in  $z \in \mathfrak{C}_i - Y$  such that  $g(z\uparrow) = w\uparrow$ , for some  $\{w\} \in \mathfrak{E}$ . Since  $g^{-1}(w)$  is a singleton and  $Y \subseteq z\uparrow$ , there is  $y \in Y$  such that  $g(y\uparrow) = w\uparrow$  and  $y \notin \text{dom}g$ , contrary to  $g$  satisfying (CDC) for  $\mathfrak{E}$  as a subreduction of  $\mathfrak{G}_i$  to  $\mathfrak{H}$ . ■

Thus, a frame separating  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \notin L$  from  $L \in \text{NExt}\mathbf{K4.3}$  can be found in the recursively enumerable class of  $t$ -extensions of  $\mathfrak{F}$ ,  $t$  being a type for  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ . Moreover, given a formula  $\alpha(\mathfrak{H}, \mathfrak{E}, \perp)$  and a type  $t$  for  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ , one can effectively check whether  $\alpha(\mathfrak{H}, \mathfrak{E}, \perp)$  is valid in the  $t$ -extension of  $\mathfrak{F}$ . Indeed, let  $k$  be the number of irreflexive points in  $\mathfrak{H}$ ,  $t = \langle \xi_1, \dots, \xi_n \rangle$ , and  $\mathfrak{G}$  the  $t$ -extension of  $\mathfrak{F}$ . Construct a cofinal subframe  $\mathfrak{G}_k$  of  $\mathfrak{G}$  by “cutting off” the infinite descending chains inserted in  $\mathfrak{F}$  (if any) just below their  $k + 1$ th points, and let  $X$  be the set of all these  $k + 1$ th points. Clearly,  $\mathfrak{G}_k$  is finite. It is now an easy exercise to prove the following

**LEMMA 73.**  $\mathfrak{G} \not\models \alpha(\mathfrak{H}, \mathfrak{E}, \perp)$  iff there is an injective cofinal subreduction  $f$  of  $\mathfrak{G}_k$  to  $\mathfrak{H}$  satisfying (CDC) for  $\mathfrak{E}$  and such that  $X \cap \text{dom}f = \emptyset$ .

As a consequence we obtain

**THEOREM 74.** All finitely axiomatizable normal extensions of  $\mathbf{K4.3}$  are decidable.

### 1.12 Quasi-normal modal logics

All logics we have considered so far were *normal*, i.e., closed under the rule of necessitation  $\varphi/\Box\varphi$ . McKinsey and Tarski [1948] noticed, however, that by adding to  $\mathbf{S4}$  the McKinsey axiom  $\mathbf{ma} = \Box\Diamond p \rightarrow \Diamond\Box p$  and taking

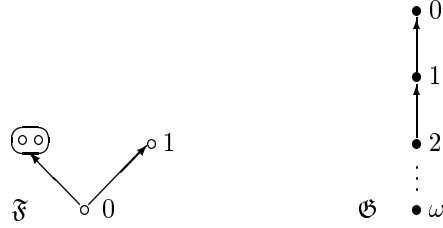


Figure 7.

the closure under modus ponens and substitution we obtain a logic—let us denote it by **S4.1'**—which is not normal in that sense. To understand why this is so, consider the frame  $\mathfrak{F}$  shown in Fig. 7. One can easily construct a model on  $\mathfrak{F}$  such that  $0 \not\models \Box ma$  (0 sees a final proper cluster). On the other hand,  $ma$  and all its substitution instances are true at 0 (0 sees a final simple cluster), from which **S4.1'**  $\subseteq \{\varphi : 0 \models \varphi\}$  and so  $\Box ma \notin \mathbf{S4.1'}$ .

A set of modal formulas containing **K** and closed under modus ponens and substitution was called by Segerberg [1971] a *quasi-normal logic*. The minimal quasi-normal extension of a logic  $L$  with formulas  $\varphi_i, i \in I$ , will be denoted by  $L + \{\varphi_i : i \in I\}$  (i.e., the operation  $+$  presupposes taking the closure under modus ponens and substitution only).  $\text{Ext}L$  is the class of all quasi-normal logics above  $L$ . It is easy to see that a quasi-normal logic is normal iff it is closed under the congruence rule  $p \leftrightarrow q / \Box p \leftrightarrow \Box q$ .

Quasi-normal logics, introduced originally as some abstract (though natural) generalization of normal ones, attracted modal logicians' attention after Solovay [1976] constructed his provability logics **GL** and **S**. The former one treats  $\Box$  as “it is provable in Peano Arithmetic” and describes those properties of Gödel’s provability predicate that are provable in PA; it is normal. The latter characterizes the properties of the provability predicate that are true in the standard arithmetic model, and in view of Gödel’s Incompleteness Theorem it cannot be normal. (For a detailed discussion of provability logic consult *Modal Logic and Self-reference*.) Solovay showed in fact that

$$\mathbf{S} = \mathbf{GL} + \Box p \rightarrow p.$$

At first sight **S** may appear to be inconsistent: Löb’s axiom requires frames to be irreflexive, while  $\Box p \rightarrow p$  is refuted in them. And indeed, no Kripke frame validates both these axioms (in particular no consistent extension of **S** is normal).

Having the algebraic semantics for normal modal logics, it is fairly easy to construct an adequate algebraic semantics for a consistent  $L \in \text{Ext}\mathbf{K}$ . Let  $M$  be a normal logic contained in  $L$  (for instance the greatest one, which is called the *kernel* of  $L$ ) and  $\mathfrak{A}_M$  its Tarski–Lindenbaum algebra (in Section 11 of *Basic Modal Logic* it was called the canonical modal algebra for  $M$ ).

The set

$$\nabla = \{[\varphi]_M : \varphi \in L\}$$

is clearly a filter in  $\mathfrak{A}_M$ . By the well known properties of the Tarski–Lindenbaum algebras, we then obtain the following completeness result:  $\varphi \in L$  iff under every valuation in  $\mathfrak{A}_M$  the value of  $\varphi$  belongs to  $\nabla$ . Structures of the form  $\langle \mathfrak{A}, \nabla \rangle$ , where  $\mathfrak{A}$  is a modal algebra and  $\nabla$  a filter in  $\mathfrak{A}$ , are known as *modal matrices*. Thus, every quasi-normal logic is characterized by a suitable class of modal matrices. It is not hard to see that  $L$  is normal iff it is characterized by a class of modal matrices with unit filters.

Now, going over to the dual (Stone–Jónsson–Tarski representation)  $\mathfrak{A}_+$  of  $\mathfrak{A}$  in a modal matrix  $\langle \mathfrak{A}, \nabla \rangle$  and taking  $\nabla_+$  to be the set of ultrafilters in  $\mathfrak{A}$  containing  $\nabla$ , we arrive at the general frame  $\mathfrak{A}_+$  with the set of *distinguished points* (or *actual worlds*)  $\nabla_+$ . A formula  $\varphi$  is regarded to be valid in  $\langle \mathfrak{A}_+, \nabla_+ \rangle$  iff under any valuation in  $\mathfrak{A}_+$ ,  $\varphi$  is true at all points in  $\nabla_+$ .

Taking into account the Generation Theorem, we can conclude that every quasi-normal modal logic is characterized by a suitable class of rooted general frames in which the root is regarded to be the only actual world. It follows in particular that, as was first observed by McKinsey and Tarski [1948],

$$\mathbf{K4} + \{\Box\varphi_i : i \in I\} = \mathbf{K4} \oplus \{\Box\varphi_i : i \in I\}.$$

However, one cannot replace here  $\mathbf{K4}$  by  $\mathbf{K}$  or  $\mathbf{T}$ . Note also that as was shown by Segerberg [1971],  $\mathbf{K}$ ,  $\mathbf{T}$  and some other standard normal logics are not finitely axiomatizable with modus ponens and substitution as the only postulated inference rules. Duality theory between modal matrices and frames with distinguished points can be developed along with duality theory for normal logics (for details see [Chagrov and Zakharyashev 1997]). Kripke frames with distinguished points were used for studying quasi-normal logics by Segerberg [1971]. Modal matrices were considered by Blok and Köhler [1983] (under the name of filtered algebras), Chagrov [1985b], and Shum [1985].

**EXAMPLE 75.** Consider the (transitive) frame  $\mathfrak{G} = \langle V, S, Q \rangle$  whose underlying Kripke frame is shown in Fig. 7 and  $Q$  consists of  $\emptyset, V$ , all finite sets of natural numbers and the complements to them in the space  $V$  (so  $\omega \in X \in Q$  iff there is  $n < \omega$  such that  $m \in X$  for all  $m \geq n$ ). Since  $\mathfrak{G}$  is irreflexive and Noetherian, it validates  $\mathbf{GL}$ . Moreover, we have  $\langle \mathfrak{G}, \omega \rangle \models \Box p \rightarrow p$ ; for if under some valuation  $\omega \models \Box p$  then  $p$  must be true at every point. It follows that  $\mathfrak{G}$  with actual world  $\omega$  validates  $\mathbf{S}$ . (The reader can check that by making  $\omega$  reflexive we again obtain a frame for  $\mathbf{S}$ .)

By inserting the “tail”  $\mathfrak{G}$  as in Fig. 7 into finite rooted frames for  $\mathbf{GL}$  below their roots and using the fact that  $\mathbf{GL}$  has FMP, one can readily

show that, for every formula  $\varphi$ ,

$$\varphi \in \mathbf{S} \text{ iff } \bigwedge_{\Box\psi \in \mathbf{Sub}\varphi} (\Box\psi \rightarrow \psi) \rightarrow \varphi \in \mathbf{GL}.$$

It follows in particular that  $\mathbf{S}$  is decidable.

This example shows that the concepts of Kripke completeness and FMP do not play so important role in the quasi-normal case: even simple logics require infinite general frames. One possible way to cope with them at least in the transitive case is to extend the frame-theoretic language of the canonical formulas to the class  $\mathbf{ExtK4}$ .

Notice first that the canonical formulas, introduced in Section 1.6, cannot axiomatize all logics in  $\mathbf{ExtK4}$ . Indeed,  $\langle \mathfrak{G}, w \rangle \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  iff there is a cofinal subreduction  $f$  of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$  and the following *actual world condition* as well:

$$\text{(AWC)} \quad f(w) \text{ is the root of } \mathfrak{F}.$$

Now, consider the frame  $\langle \mathfrak{G}, \omega \rangle$  constructed in Example 75. Since each set  $X \in Q$  containing  $\omega$  is infinite and has a dead end, it is impossible to reduce  $X$  to  $\circ$  or  $\bullet$ , and so  $\langle \mathfrak{G}, \omega \rangle$  validates all normal canonical formulas. On the other hand, we clearly have  $\langle \mathfrak{G}, \omega \rangle \not\models B_n$  for every  $n \geq 1$ . So the logics  $\mathbf{K4BD}_n$  cannot be axiomatized by normal canonical formulas without the postulated necessitation.

To get over this obstacle we have to modify the definition of subreduction so that such sets as  $X$  above may be “reduced” at least to irreflexive roots of frames. Given a frame  $\mathfrak{G} = \langle V, S, Q \rangle$  with an *irreflexive* root  $u$  and a frame  $\mathfrak{F} = \langle W, R, P \rangle$ , we say a partial map  $f$  from  $W$  onto  $V$  is a *quasi-subreduction* of  $\mathfrak{F}$  to  $\mathfrak{G}$  if it satisfies (R1) for all  $x, y \in \text{dom}f$  such that  $f(x) \neq u$  or  $f(y) \neq u$ , (R2) and (R3).<sup>11</sup> Thus, we may map all points in the frame  $\mathfrak{G}$  in Fig. 7 to  $\bullet$ , and this map will be a quasi-reduction of  $\mathfrak{G}$  to  $\bullet$  satisfying (AWC). Actually, every frame is quasi-reducible to  $\bullet$ .

Now, given a finite frame  $\mathfrak{F}$  with an irreflexive root  $a_0$  and a set  $\mathfrak{D}$  of antichains in  $\mathfrak{F}$ , we define the *quasi-normal canonical formula*  $\alpha^\bullet(\mathfrak{F}, \mathfrak{D}, \perp)$  as the result of deleting  $\Box p_0$  from  $\varphi_0$  in  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  (which says that  $a_0$  is not self-accessible); the *quasi-normal negation free canonical formula*  $\alpha^\bullet(\mathfrak{F}, \mathfrak{D})$  is defined in exactly the same way, starting from  $\alpha(\mathfrak{F}, \mathfrak{D})$ . It is not hard to see that  $\alpha^\bullet(\mathfrak{F}, \mathfrak{D}, \perp)$  (or  $\alpha^\bullet(\mathfrak{F}, \mathfrak{D})$ ) is refuted in a frame  $\langle \mathfrak{G}, w \rangle$  iff there is a cofinal (respectively, plain) quasi-subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$  and (AWC). The following result is obtained by an obvious generalization of the proof of Theorem 44 to frames with distinguished points (for details see [Zakharyashev 1992]).

<sup>11</sup>Another possibility is to allow “reductions” of  $X$  to reflexive points by relaxing (R2); cf. Section 2.6.

**THEOREM 76.** *There is an algorithm which, given a modal (negation free) formula  $\varphi$ , constructs a finite set  $\Delta$  of normal and quasi-normal (negation free) canonical formulas such that  $\mathbf{K4} + \varphi = \mathbf{K4} + \Delta$ .*

For example,  $\mathbf{S} = \mathbf{K4} + \alpha(\circ) + \alpha(\bullet)$ . Since frames for  $\mathbf{S4}$  are reflexive, we have

**COROLLARY 77.** *There is an algorithm which, given a modal formula  $\varphi$ , constructs a finite set  $\Delta$  of normal canonical formulas built on reflexive frames such that  $\mathbf{S4} + \varphi = \mathbf{S4} + \Delta$ .*

As a consequence we obtain

**THEOREM 78** (Segerberg 1975).  $\text{Ext}\mathbf{S4.3} = \text{NExt}\mathbf{S4.3}$ .

**Proof.** We must show that every logic  $L \in \text{Ext}\mathbf{S4.3}$  is normal, i.e.,  $\varphi \in L$  only if  $\Box\varphi \in L$ , for every  $\varphi$ . Suppose otherwise. Then by Corollary 77, there exists  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in L$  such that  $\Box\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \notin L$ . Let  $\langle \mathfrak{G}, w \rangle$  be a frame validating  $L$  and refuting  $\Box\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ . Since  $\mathfrak{G} \models \mathbf{S4.3}$ ,  $\mathfrak{G}$  is a chain of non-degenerate clusters. And since it refutes  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  there is a cofinal subreduction  $f$  of  $\mathfrak{G}$  to  $\mathfrak{F}$ . It follows, in particular, that  $\mathfrak{F}$  is also a chain of non-degenerate clusters and so  $\mathfrak{D} = \emptyset$ . Let  $a$  be the root of  $\mathfrak{F}$ . Define a map  $g$  by taking

$$g(x) = \begin{cases} f(x) & \text{if } x \in \text{dom } f \\ a & \text{if } x \in f^{-1}(a) \downarrow - \text{dom } f \\ \text{undefined} & \text{otherwise.} \end{cases}$$

It should be clear that  $g$  cofinally subreduces  $\mathfrak{G}$  to  $\mathfrak{F}$  and  $g(w) = a$ . Consequently,  $\langle \mathfrak{G}, w \rangle \not\models \alpha(\mathfrak{F}, \perp)$ , which is a contradiction.  $\blacksquare$

Let us now briefly consider quasi-normal analogues of subframe and cofinal subframe logics in  $\text{NExt}\mathbf{K4}$ . Those logics that can be represented in the form

$$(\mathbf{K4} \oplus \{\alpha(\mathfrak{F}_i) : i \in I\}) + \{\alpha(\mathfrak{F}_j) : j \in J\} + \{\alpha^\bullet(\mathfrak{F}_k) : k \in K\}$$

are called (*quasi-normal*) *subframe logics* and those of the form

$$(\mathbf{K4} \oplus \{\alpha(\mathfrak{F}_i, \perp) : i \in I\}) + \{\alpha(\mathfrak{F}_j, \perp) : j \in J\} + \{\alpha^\bullet(\mathfrak{F}_k, \perp) : k \in K\}$$

are called (*quasi-normal*) *cofinal subframe logics*. The classes of quasi-normal subframe and cofinal subframe logics are denoted by  $QS\mathcal{F}$  and  $QCS\mathcal{F}$ , respectively. The example of  $\mathbf{S}$  shows that Theorem 52 cannot be extended to  $QS\mathcal{F}$  and  $QCS\mathcal{F}$ . Yet one can show that all finitely axiomatizable logics in  $QS\mathcal{F}$  and  $QCS\mathcal{F}$  are decidable. We omit almost all proofs and confine ourselves mainly to formulations of relevant results. For details the reader is referred to [Zakharyashev 1996].

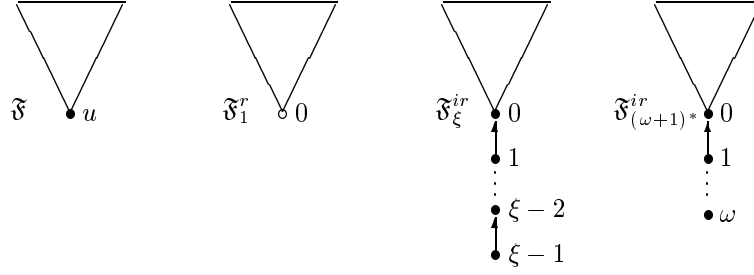


Figure 8.

We use the following notation. For a frame  $\mathfrak{F} = \langle W, R \rangle$  with irreflexive root  $u$  and  $0 < \xi < \omega$ ,  $\mathfrak{F}_\xi^{ir}$  and  $\mathfrak{F}_\xi^r$  denote the frames obtained from  $\mathfrak{F}$  by replacing  $u$  with the descending chains  $0, \dots, \xi - 1$  of irreflexive and reflexive points, respectively;  $\mathfrak{F}_{(\omega+1)^*}^{ir} = \langle W_{(\omega+1)^*}, R_{(\omega+1)^*}, P_{(\omega+1)^*} \rangle$  is the frame that results from  $\mathfrak{F}$  by replacing  $u$  with the infinite descending chain  $0, 1, \dots$  of irreflexive points and then adding irreflexive root  $\omega$ , with  $P_{(\omega+1)^*}$  containing all subsets of  $W - \{u\}$ , all finite subsets of natural numbers  $\{0, 1, \dots\}$ , all (finite) unions of these sets and all complements to them in the space  $W_{(\omega+1)^*}$  (see Fig. 8). Note that  $\mathfrak{F}$  is a quasi-reduct of every frame of the form  $\mathfrak{F}_\xi^{ir}$ ,  $\mathfrak{F}_\xi^r$  or  $\mathfrak{F}_{(\omega+1)^*}^{ir}$ .

The following theorem characterizes the canonical formulas belonging to logics in  $QSF$  and  $QCSF$ .

**THEOREM 79.** *Suppose  $L$  is a subframe or cofinal subframe quasi-normal logic. Then*

- rm (i) for every finite frame  $\mathfrak{F}$  with root  $u$ ,  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in L$  iff  $\langle \mathfrak{F}, u \rangle \not\models L$ ;*  
*rm (ii) for every finite frame  $\mathfrak{F}$  with irreflexive root  $u$ ,  $\alpha^\bullet(\mathfrak{F}, \mathfrak{D}, \perp) \in L$  iff  $\langle \mathfrak{F}, u \rangle \not\models L$ ,  $\langle \mathfrak{F}_1^r, 0 \rangle \not\models L$  and  $\langle \mathfrak{F}_{(\omega+1)^*}^{ir}, \omega \rangle \not\models L$ .*

**Proof.** We prove only  $(\Leftarrow)$  of (ii). Let  $\mathfrak{G} = \langle V, S, Q \rangle$  refute  $\alpha^\bullet(\mathfrak{F}, \mathfrak{D}, \perp)$  at its root  $w$  and show that  $\langle \mathfrak{G}, w \rangle \not\models L$ . We have a cofinal quasi-subreduction  $f$  of  $\mathfrak{G}$  to  $\mathfrak{F}$  such that  $f(w) = u$ . Consider the set  $U = f^{-1}(u) \in Q$ . Without loss of generality we may assume that  $U = U\downarrow$ . There are three possible cases.

*Case 1.* The point  $w$  is irreflexive and  $\{w\} \in Q$ . Then the restriction of  $f$  to  $\text{dom}f - (U - \{w\})$  is a cofinal subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (AWC) and so  $\langle \mathfrak{G}, w \rangle \not\models L$ .

*Case 2.* There is  $X \subseteq U$  such that  $w \in X \in Q$  and, for every  $x \in X$ , there exists  $y \in X \cap x\uparrow$ . Then the restriction of  $f$  to  $\text{dom}f - (U - X)$  is a cofinal subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}_1^r$  satisfying (AWC) and so again  $\langle \mathfrak{G}, w \rangle \not\models L$ .



*Case 3.* If neither of the preceding cases holds then, for every  $X \subseteq U$  such that  $w \in X \in Q$ , the set  $D_X = X - X\downarrow$  of dead ends in  $X$  is a cover for  $X$ , i.e.,  $X \subseteq D_X\downarrow$ , and  $w \in X - D_X \in Q$ . Put

$$X_0 = D_U, \dots, X_{n+1} = D_{U - (X_0 \cup \dots \cup X_n)}, \dots, X_\omega = U - \bigcup_{\xi < \omega} X_\xi.$$

Each of these sets, save possibly  $X_\omega$ , is an antichain of irreflexive points and belongs to  $Q$ . Besides,  $X_\zeta \subset X_n\downarrow = \bigcup_{n < \xi \leq \omega} X_\xi$  for every  $n < \zeta \leq \omega$ . Therefore, the map  $g$  defined by

$$g(x) = \begin{cases} f(x) & \text{if } x \in V - U \\ \xi & \text{if } x \in X_\xi, 0 \leq \xi \leq \omega \end{cases}$$

is a cofinal quasi-subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}_{(\omega+1)^*}^{ir}$  satisfying (AWC).

Now using the fact that  $\langle \mathfrak{F}_{(\omega+1)^*}^{ir}, \omega \rangle \not\models L$  and that the composition of (cofinal) (quasi-) subreductions is again a (cofinal) (quasi-) subreduction, it is not hard to see that  $\langle \mathfrak{G}, w \rangle \not\models L$ .  $\blacksquare$

**COROLLARY 80.** *All subframe and cofinal subframe quasi-normal logics above **S4** have FMP.*

**EXAMPLE 81.** As an illustration let us use Theorem 79 to characterize those normal and quasi-normal canonical formulas that belong to **S**. Clearly, either  $\alpha(\circ)$  or  $\alpha(\bullet)$  is refuted at the root of every rooted Kripke frame. So all normal canonical formulas are in **S**. Every quasi-normal formula  $\alpha^\bullet(\mathfrak{F}, \mathfrak{D}, \perp)$  associated with  $\mathfrak{F}$  containing a reflexive point is also in **S**, since  $\Box\alpha(\circ)$  is refuted at the roots of  $\mathfrak{F}$ ,  $\mathfrak{F}_1^r$  and  $\mathfrak{F}_{(\omega+1)^*}^{ir}$ . But no quasi-normal formula  $\alpha^\bullet(\mathfrak{F}, \mathfrak{D}, \perp)$  built on irreflexive  $\mathfrak{F}$  belongs to **S**, because  $\mathfrak{F}_{(\omega+1)^*}^{ir} \models \alpha(\circ)$  and  $\langle \mathfrak{F}_{(\omega+1)^*}^{ir}, \omega \rangle \models \alpha(\bullet)$ , since  $\{\omega\} \notin P_{(\omega+1)^*}$ . Notice that incidentally we have proved the following completeness theorem for **S**.

**THEOREM 82.** ***S** is characterized by the class*

$$\{\langle \mathfrak{F}_{(\omega+1)^*}^{ir}, \omega \rangle : \mathfrak{F} \text{ is a finite rooted irreflexive frame}\}.$$

Theorem 79 reduces the decision problem for a logic  $L$  in  $QS\mathcal{F}$  or  $QCS\mathcal{F}$  to the problem of verifying, given a finite frame  $\mathfrak{F}$  with root  $u$ , whether  $\langle \mathfrak{F}, u \rangle$ ,  $\langle \mathfrak{F}_1^r, 0 \rangle$  and  $\langle \mathfrak{F}_{(\omega+1)^*}^{ir}, \omega \rangle$  refute an axiom of  $L$ . The two former frames present no difficulties: they are finite. As to the latter, it is not hard to see that, for instance,  $\langle \mathfrak{F}_{(\omega+1)^*}^{ir}, \omega \rangle \not\models \alpha^\bullet(\mathfrak{G}, \perp)$  iff  $\langle \mathfrak{F}_\xi^{ir}, \xi - 1 \rangle$ , for some  $\xi \leq |\mathfrak{G}|$ , is cofinally quasi-subreducible to  $\mathfrak{G}$ . Thus we obtain

**THEOREM 83.** *All finitely axiomatizable subframe and cofinal subframe quasi-normal logics are decidable.*

One can also give a frame-theoretic characterization of the classes  $QSF$  and  $QCSF$  similar to Theorem 53. Let us say that a frame  $\mathfrak{F}$  with actual world  $u$  is a (*cofinal*) *subframe* of a frame  $\mathfrak{G}$  with actual world  $w$  if  $\mathfrak{F}$  is a (*cofinal*) subframe of  $\mathfrak{G}$  and  $u = w$ .

**THEOREM 84.** *L is a (cofinal) subframe quasi-normal logic iff L is characterized by a class of frames with actual worlds that is closed under (cofinal) subframes.*

### 1.13 Tabular logics

Every logic  $L$  having the finite model property can be represented as the intersection of some *tabular logics*, that is logics characterized by finite frames (or models, algebras, matrices, etc.):

$$L = \bigcap \{ \text{Log} \mathfrak{F} : \mathfrak{F} \text{ is a finite frame for } L \}.$$

(It follows in particular that every fragment of  $L$  containing only those formulas whose length does not exceed some fixed  $n < \omega$  is determined by a finite frame; for that reason logics with FMP are also called *finitely approximable*.) In many respects tabular logics are very easy to deal with. For instance, the key problem of recognizing whether a formula  $\varphi$  belongs to a tabular  $L$  is trivially decided by the direct inspection of all possible valuations of  $\varphi$ 's variables in the finite frame characterizing  $L$ . That is why the question “is it tabular?” is one of the first items in the standard “questionnaire” for every new logical system.

First results concerning the tabularity of modal logics were obtained by Gödel [1932] and Dugundji [1940] who showed that intuitionistic propositional logic and all Lewis' modal systems **S1–S5** are not tabular. (Note that using the same method Drabbé [1967] proved that the three non-normal Lewis' systems **S1–S3** cannot be characterized by a matrix with a finite number of distinguished elements). For arbitrary logics in  $\text{Ext}\mathbf{K}$  one can easily prove the following syntactical criterion of tabularity, which uses the formulas

$$\begin{aligned} \alpha_n &= \neg(\varphi_1 \wedge \diamond(\varphi_2 \wedge \diamond(\varphi_3 \wedge \dots \wedge \diamond\varphi_n) \dots)), \\ \beta_n &= \bigwedge_{m=0}^{n-1} \neg\diamond^m(\diamond\varphi_1 \wedge \dots \wedge \diamond\varphi_n), \\ \mathbf{tab}_n &= \alpha_n \wedge \beta_n, \end{aligned}$$

where  $\varphi_i = p_1 \wedge \dots \wedge p_{i-1} \wedge \neg p_i \wedge p_{i+1} \wedge \dots \wedge p_n$ .

**THEOREM 85.** *L ∈ ExtK is tabular iff  $\mathbf{tab}_n \in L$ , for some  $n < \omega$ .*

**Proof.** A frame  $\mathfrak{F} = \langle W, R \rangle$  refutes  $\alpha_n$  at a point  $x_1$  iff a chain of length  $n$  starts from  $x_1$ , and  $\mathfrak{F}$  refutes  $\beta_n$  at  $x_1$  iff there is a chain  $x_1 R x_2 R \dots R x_m$

of length  $m < n$  such that  $x_m$  is of branching  $n$ , i.e.,  $x_m R y_1, \dots, x_m R y_n$  for some distinct  $y_1, \dots, y_n$ . It follows that every rooted generated (by an actual world) subframe of the canonical frame for  $L$  containing  $\mathbf{tab}_n$  has at most  $1 + (n-1) + \dots + (n-1)^{n-2}$  points. ■

As a consequence we immediately obtain

**COROLLARY 86.** *Every tabular modal logic has finitely many extensions and all of them are also tabular.*

The next theorem follows from general algebraic results of [Blok and Köhler 1983]; equally easy it can be proved using the characterization above.

**THEOREM 87.** *Every tabular logic  $L \in \text{Ext}\mathbf{K}$  is finitely axiomatizable.*

**Proof.** According to Theorem 85,  $L$  is an extension of  $\mathbf{K} + \mathbf{tab}_n$ , for some  $n < \omega$ . By Corollary 86, we have a chain

$$\mathbf{K} + \mathbf{tab}_n = L_1 \subset L_2 \subset \dots \subset L_{k-1} \subset L_k = L$$

of quasi-normal logics such that  $\{L' \in \text{Ext}\mathbf{K} : L_i \subset L' \subset L_{i+1}\} = \emptyset$ , for every  $i = 1, \dots, k-1$ . It remains to notice that if  $L'$  is finitely axiomatizable,  $L' \subset L''$  and there is no logic located properly between  $L'$  and  $L''$  then  $L''$  is also finitely axiomatizable (e.g.  $L'' = L' + \varphi$ , for any  $\varphi \in L'' - L'$ ). ■

Theorem 12 provides us in fact with an algorithm to decide, given a tabular logic  $L \in \text{NExt}\mathbf{K4}$  and an arbitrary formula  $\varphi$ , whether  $\mathbf{K4} \oplus \varphi = L$ . Indeed, notice first that we have

**THEOREM 88.** *Each finitely axiomatizable logic  $L \in \text{NExt}\mathbf{K4}$  of finite depth is a finite union-splitting, i.e., can be represented in the form*

$$L = \mathbf{K4} \oplus \{\alpha^\sharp(\mathfrak{F}_i, \perp) : i \in I\}$$

with finite  $I$ .

**Proof.** Let  $L = \mathbf{K4} \oplus \varphi$  be a logic of depth  $n$  and let  $m$  be the number of variables in  $\varphi$ . We show that  $L$  coincides with the logic

$$L' = \mathbf{K4} \oplus \{\alpha^\sharp(\mathfrak{G}, \perp) : |\mathfrak{G}| \leq \sum_{i=1}^{n+1} 2^m c_m(i), \mathfrak{G} \not\models \varphi\}$$

( $c_m(i)$  was defined in Section 1.2). The inclusion  $L \supseteq L'$  is obvious. Suppose  $\varphi \notin L'$ . Then there is a rooted refined  $m$ -generated frame  $\mathfrak{F}$  for  $L'$  refuting  $\varphi$ . Clearly,  $\mathfrak{F}$  is of depth  $\leq n$ , since otherwise  $\alpha^\sharp(\mathfrak{G}, \perp)$  is an axiom of  $L'$  for every rooted generated subframe  $\mathfrak{G}$  of  $\mathfrak{F}$  of depth  $n+1$  and so  $\mathfrak{F} \models L'$ , which is a contradiction. But then  $\alpha^\sharp(\mathfrak{F}, \perp)$  is an axiom of  $L'$ , contrary to our assumption. ■

Thus, all tabular logics in  $\text{NExt}\mathbf{K4}$  are finite union-splittings and so, by Theorem 12, we obtain the following

**THEOREM 89.** *Let  $L$  be a tabular logic in  $\text{NExt}\mathbf{K4}$ .*

- (i) (Blok 1980c)  *$L$  has finitely many immediate predecessors and they are also tabular.*
- (ii) *The axiomatizability problem for  $L$  above  $\mathbf{K4}$  is decidable.*

For logics in  $\text{NExt}\mathbf{K}$  this is not the case, witness Theorems 36 and 205.

The tabularity criterion of Theorem 85 is not effective. Moreover, as we shall see in Section 4.4, no effective tabularity criterion exists in general. However, if we restrict attention to sufficiently strong logics, e.g. to the class  $\text{NExt}\mathbf{S4}$ , the tabularity problem turns out to be decidable. The key idea, proposed by Kuznetsov [1971], is to consider the so called pre-tabular logics.

A logic  $L \in (\text{N})\text{Ext}L_0$  is said to be *pre-tabular* in the lattice  $(\text{N})\text{Ext}L_0$ , if  $L$  is not tabular but every proper extension of  $L$  in  $(\text{N})\text{Ext}L_0$  is tabular. In other words, a pre-tabular logic in  $(\text{N})\text{Ext}L_0$  is a maximal non-tabular logic in  $(\text{N})\text{Ext}L_0$ .

**THEOREM 90.** *In the lattices  $\text{Ext}\mathbf{K}$  and  $\text{NExt}\mathbf{K}$  every non-tabular logic is contained in a pre-tabular one.*

**Proof.** By Theorem 85, a logic is non-tabular iff it does not contain the formula  $\mathbf{tab}_n$ , for any  $n < \omega$ . It follows that the union of an ascending chain of non-tabular logics is a non-tabular logic as well. The standard use of Zorn's Lemma completes the proof. ■

If there is a simple description of all pre-tabular logics in a lattice, we obtain an effective (modulo the description) tabularity criterion for the lattice. Indeed, take for definiteness the lattice  $\text{NExt}\mathbf{K4}$ . How to determine, given a formula  $\varphi$ , whether  $\mathbf{K4} \oplus \varphi$  is tabular? We may launch two parallel processes: one of them generates all derivations in  $\mathbf{K4} \oplus \varphi$  and stops after finding a derivation of  $\mathbf{tab}_n$ , for some  $n < \omega$ ; another process checks if  $\varphi$  belongs to a pre-tabular logic in  $\text{NExt}\mathbf{K4}$  and stops if this is the case. The termination of the first process means that  $\mathbf{K4} \oplus \varphi$  is tabular, while that of the second one shows that it is not tabular.

Unfortunately, it is impossible to describe in an effective way all pre-tabular logics in  $(\text{N})\text{Ext}\mathbf{K}$  and even  $(\text{N})\text{Ext}\mathbf{K4}$ : Blok [1980c] and Chagrov [1989] constructed a continuum of them. However, for smaller lattices like  $\text{NExt}\mathbf{S4}$  or  $\text{NExt}\mathbf{GL}$  such descriptions were found by Maksimova [1975b], Esakia and Meskhi [1977] and Blok [1980c]. The five pre-tabular logics in  $\text{NExt}\mathbf{S4}$  were presented in Section 17 of *Basic Modal Logic*. In  $\text{NExt}\mathbf{GL}$  the picture is much more complicated.

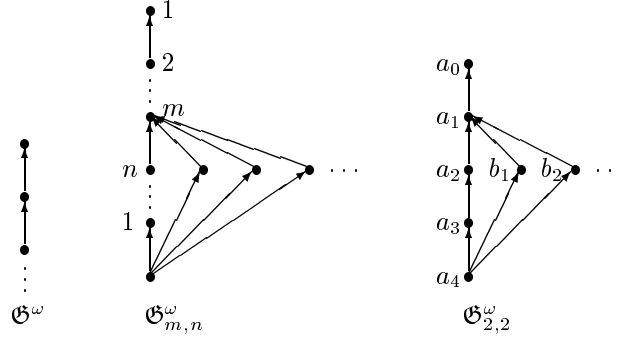


Figure 9.

**THEOREM 91** (Blok 1980c, Chagrov 1989). *The set of pretabular logics in  $\text{NExtGL}$  is denumerable. It consists of the logics  $\mathbf{GL.3} = \text{Log}\mathfrak{G}^\omega$  and  $\text{Log}\mathfrak{G}_{m,n}^\omega$ , for  $m \geq 0, n \geq 1$ , where  $\mathfrak{G}^\omega$  and  $\mathfrak{G}_{m,n}^\omega$  are the frames depicted in Fig. 9. If  $\langle m, n \rangle \neq \langle k, l \rangle$  then  $\text{Log}\mathfrak{G}_{m,n}^\omega \neq \text{Log}\mathfrak{G}_{k,l}^\omega$ .*

Using this semantic description of pretabular logics in  $\text{NExtGL}$ , it is not hard to find finite sets of formulas axiomatizing them. Moreover, all of them turn out to be decidable. For we have

**THEOREM 92.** *Every non-tabular logic  $L \in \text{NExtK4}$  has a non-tabular extension with FMP, and so every pretabular logic in  $\text{NExtK4}$  has FMP.*

**Proof.** Since  $L$  is non-tabular and characterized by the class of its rooted finitely generated refined frames, we have either a sequence  $\mathfrak{F}_i, i = 1, 2, \dots$ , of rooted finite frames for  $L$  of depth  $i$ , or a sequence  $\mathfrak{F}_i$  of rooted finite frames for  $L$  of width  $\geq i$ . In both cases the logic  $\text{Log}\{\mathfrak{F}_i : i < \omega\} \supseteq L$  is non-tabular and has FMP. ■

So we obtain the following result on the decidability of tabularity.

**THEOREM 93.** *The property of tabularity is decidable in  $\text{NExtS4}$ ,  $\text{ExtS4}$ ,  $\text{NExtGL}$ ,  $\text{ExtGL}$ .*

Since a logic in  $\text{ExtK4}$  is locally tabular iff it is determined by a frame of finite depth, the property of local tabularity is decidable in the lattices mentioned in Theorem 93 as well. However, this is not the case for  $\text{ExtK4}$  itself.

### 1.14 Interpolation

One of the fundamental properties of logics is their capability to provide explicit definitions of implicitly definable terms, which is known as the Beth property (Beth [1953] proved it for classical logic). In the modal case we

say a logic  $L$  has the *Beth property* if, for any formula  $\varphi(p_1, \dots, p_n, p_{n+1})$  and variables  $p$  and  $q$  different from  $p_1, \dots, p_n$ ,

$$\varphi(p_1, \dots, p_n, p) \wedge \varphi(p_1, \dots, p_n, q) \rightarrow (p \leftrightarrow q) \in L$$

only if there is a formula  $\psi(p_1, \dots, p_n)$  such that

$$\varphi(p_1, \dots, p_n, p) \rightarrow (p \leftrightarrow \psi(p_1, \dots, p_n)) \in L.$$

The Beth property turns out to be closely related to the interpolation property which was introduced by Craig [1957] for classical logic. Namely, we say that a logic  $L$  has the *interpolation property* if, for every implication  $\alpha \rightarrow \beta \in L$ , there exists a formula  $\gamma$ , called an *interpolant* for  $\alpha \rightarrow \beta$  in  $L$ , such that  $\alpha \rightarrow \gamma \in L$ ,  $\gamma \rightarrow \beta \in L$  and every variable in  $\gamma$ , if any, occurs in both  $\alpha$  and  $\beta$ . While in abstract model theory interpolation is weaker than Beth definability, for modal logics we have

**THEOREM 94** (Maksimova 1992). *A normal modal logic has interpolation iff it has the Beth property.*

Say also that a normal modal logic  $L$  has the *interpolation property for the consequence relation*  $\vdash_L^*$ ,  $\vdash^*$ -interpolation for short, if every time when  $\alpha \vdash_L^* \beta$ , there is a formula  $\gamma$  such that  $\alpha \vdash_L^* \gamma$ ,  $\gamma \vdash_L^* \beta$  and  $\mathbf{Var}\gamma \subseteq \mathbf{Var}\alpha \cap \mathbf{Var}\beta$ . (Here  $\mathbf{Var}\varphi$  is the set of all variables in  $\varphi$ .) It should be clear that interpolation implies  $\vdash^*$ -interpolation.

By the end of the 1970s interpolation had been established for a good many standard modal systems. The semantical proofs, sometimes rather sophisticated, resemble the Henkin construction of the canonical models. Here are two examples of such proofs (which are due to Maksimova [1982b] and Smoryński [1978]).

**THEOREM 95** (Gabbay 1972). *The logics **K**, **K4**, **T**, **S4** have the interpolation property.*

**Proof.** We consider only **S4**; for the other logics the proofs are similar. Suppose  $\alpha \rightarrow \gamma \notin \mathbf{S4}$  and  $\gamma \rightarrow \beta \notin \mathbf{S4}$  for any  $\gamma$  whose variables occur in both  $\alpha$  and  $\beta$ , and show that in this case  $\alpha \rightarrow \beta \notin \mathbf{S4}$ .

Let  $t = (\Gamma, \Delta)$  be a pair of sets of formulas such that  $\mathbf{Var}\varphi \subseteq \mathbf{Var}\alpha$  if  $\varphi \in \Gamma$  and  $\mathbf{Var}\varphi \subseteq \mathbf{Var}\beta$  if  $\varphi \in \Delta$ . Say that  $t$  is *inseparable* if there are no formulas  $\varphi_i \in \Gamma$ ,  $\psi_j \in \Delta$  and  $\gamma$  with  $\mathbf{Var}\gamma \subseteq \mathbf{Var}\alpha \cap \mathbf{Var}\beta$  such that  $\bigwedge_{i=1}^n \varphi_i \rightarrow \gamma \in \mathbf{S4}$ ,  $\gamma \rightarrow \bigvee_{i=1}^m \psi_i \in \mathbf{S4}$ . The pair  $t$  is called *complete* if for every  $\varphi$  and  $\psi$  with  $\mathbf{Var}\varphi \subseteq \mathbf{Var}\alpha$  and  $\mathbf{Var}\psi \subseteq \mathbf{Var}\beta$ , one of the formulas  $\varphi$  and  $\neg\varphi$  is in  $\Gamma$  and one of  $\psi$  and  $\neg\psi$  is in  $\Delta$ .

**LEMMA 96.** *Every inseparable pair  $t_0 = (\Gamma_0, \Delta_0)$  can be extended to a complete inseparable pair.*

**Proof.** Let  $\varphi_1, \varphi_2, \dots$  and  $\psi_1, \psi_2, \dots$  be enumerations of all formulas whose variables occur in  $\alpha$  and  $\beta$ , respectively. Define pairs  $t'_n = (\Gamma'_n, \Delta'_n)$  and  $t_{n+1} = (\Gamma_{n+1}, \Delta_{n+1})$  inductively by taking

$$t'_n = \begin{cases} (\Gamma_n \cup \{\varphi_n\}, \Delta_n) & \text{if this pair is inseparable} \\ (\Gamma_n \cup \{\neg\varphi_n\}, \Delta_n) & \text{otherwise,} \end{cases}$$

$$t_{n+1} = \begin{cases} (\Gamma'_n, \Delta'_n \cup \{\psi_n\}) & \text{if this pair is inseparable} \\ (\Gamma'_n, \Delta'_n \cup \{\neg\psi_n\}) & \text{otherwise} \end{cases}$$

and put  $t^* = (\Gamma^*, \Delta^*)$ , where  $\Gamma^* = \bigcup_{n < \omega} \Gamma_n$ ,  $\Delta^* = \bigcup_{n < \omega} \Delta_n$ . Clearly  $t^*$  is complete. Suppose it is separable, i.e., for some  $\varphi_1, \dots, \varphi_n \in \Gamma^*$ ,  $\psi_1, \dots, \psi_m \in \Delta^*$  and some  $\gamma$  containing only those variables that occur in both  $\alpha$  and  $\beta$ , we have  $\bigwedge_{i=1}^n \varphi_i \rightarrow \gamma \in \mathbf{S4}$  and  $\gamma \rightarrow \bigvee_{i=1}^m \psi_i \in \mathbf{S4}$ . Then there is  $k < \omega$  such that  $\varphi_1, \dots, \varphi_n \in \Gamma_k$  and  $\psi_1, \dots, \psi_m \in \Delta_k$ , which means that  $t_k$  is separable. So it remains to show that if  $t = (\Gamma, \Delta)$  is inseparable,  $\mathbf{Var}\varphi \subseteq \mathbf{Var}\alpha$  and  $\mathbf{Var}\psi \subseteq \mathbf{Var}\beta$  then

- one of the pairs  $(\Gamma \cup \{\varphi\}, \Delta)$  or  $(\Gamma \cup \{\neg\varphi\}, \Delta)$  is inseparable and
- one of the pairs  $(\Gamma, \Delta \cup \{\psi\})$  or  $(\Gamma, \Delta \cup \{\neg\psi\})$  is inseparable.

We prove only the former claim. Suppose, on the contrary, that both pairs are separable, i.e., there are formulas  $\gamma_1, \gamma_2$  in variables occurring in both  $\alpha$  and  $\beta$  such that, for some  $\varphi_1, \dots, \varphi_n \in \Gamma$ ,  $\psi_1, \dots, \psi_m \in \Delta$ , we have

$$\varphi_1 \wedge \dots \wedge \varphi_n \wedge \varphi \rightarrow \gamma_1 \in \mathbf{S4}, \quad \gamma_1 \rightarrow \psi_1 \vee \dots \vee \psi_m \in \mathbf{S4},$$

$$\varphi_1 \wedge \dots \wedge \varphi_n \wedge \neg\varphi \rightarrow \gamma_2 \in \mathbf{S4}, \quad \gamma_2 \rightarrow \psi_1 \vee \dots \vee \psi_m \in \mathbf{S4}.$$

Then we obtain  $(\varphi_1 \wedge \dots \wedge \varphi_n \wedge \varphi) \vee (\varphi_1 \wedge \dots \wedge \varphi_n \wedge \neg\varphi) \rightarrow \gamma_1 \vee \gamma_2 \in \mathbf{S4}$ ,  $\gamma_1 \vee \gamma_2 \rightarrow \psi_1 \vee \dots \vee \psi_m \in \mathbf{S4}$ , from which

$$\varphi_1 \wedge \dots \wedge \varphi_n \rightarrow \gamma_1 \vee \gamma_2 \in \mathbf{S4}, \quad \gamma_1 \vee \gamma_2 \rightarrow \psi_1 \vee \dots \vee \psi_m \in \mathbf{S4},$$

contrary to  $t$  being inseparable. ■

Now we define a frame  $\mathfrak{F} = \langle W, R \rangle$  by taking  $W$  to be the set of all complete and inseparable pairs and, for  $t_1 = (\Gamma_1, \Delta_1)$ ,  $t_2 = (\Gamma_2, \Delta_2)$  in  $W$ ,  $t_1 R t_2$  iff  $\Box\varphi \in \Gamma_1$  implies  $\varphi \in \Gamma_2$ . Using the axioms  $\Box p \rightarrow p$  and  $\Box p \rightarrow \Box\Box p$  of  $\mathbf{S4}$ , one can readily check that  $R$  is a quasi-order on  $W$ , i.e.,  $\mathfrak{F} \models \mathbf{S4}$ .

Define a valuation  $\mathfrak{V}$  in  $\mathfrak{F}$  by taking for every variable  $p \in \mathbf{Var}(\alpha \rightarrow \beta)$ ,  $\mathfrak{V}(p) = \{(\Gamma, \Delta) \in W : \text{either } p \in \Gamma \text{ or } p \in \mathbf{Var}\beta \text{ and } p \notin \Delta\}$ . Put  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$ . By induction on the construction of formulas  $\varphi$  and  $\psi$  with  $\mathbf{Var}\varphi \subseteq \mathbf{Var}\alpha$ ,  $\mathbf{Var}\psi \subseteq \mathbf{Var}\beta$  one can show that for every  $t = (\Gamma, \Delta)$  in  $\mathfrak{F}$

$$(\mathfrak{M}, t) \models \varphi \text{ iff } \varphi \in \Gamma, \quad (\mathfrak{M}, t) \not\models \psi \text{ iff } \psi \in \Delta.$$

Indeed, the basis of induction follows from the definition of  $\mathfrak{B}$  and the completeness and inseparability of  $t$ . The cases of the Boolean connectives present no difficulty. So suppose  $\varphi = \Box\varphi_1$ . If  $t \models \Box\varphi_1$  then, for every  $t' = (\Gamma', \Delta') \in t \uparrow$ , we have  $t' \models \varphi_1$  and so  $\varphi_1 \in \Gamma'$ . Suppose  $\Box\varphi_1 \notin \Gamma$ . Then  $\neg\Box\varphi_1 \in \Gamma$ . Consider the pair  $t_0 = (\Gamma_0, \Delta_0)$ , where

$$\Gamma_0 = \{\neg\varphi_1\} \cup \{\chi : \Box\chi \in \Gamma\}, \quad \Delta_0 = \{\neg\chi : \neg\Box\chi \in \Delta\},$$

and show that it is inseparable. Assume otherwise. Then there is  $\gamma$  with  $\mathbf{Var}\gamma \subseteq \mathbf{Var}\alpha \cap \mathbf{Var}\beta$  such that, for some formulas  $\Box\chi_1, \dots, \Box\chi_n \in \Gamma$ ,  $\neg\Box\chi_{n+1}, \dots, \neg\Box\chi_m \in \Delta$ ,

$$\neg\varphi_1 \wedge \chi_1 \wedge \dots \wedge \chi_n \rightarrow \gamma \in \mathbf{S4}, \quad \gamma \rightarrow \neg\chi_{n+1} \vee \dots \vee \neg\chi_m \in \mathbf{S4}.$$

It follows that

$$\neg\Box\varphi_1 \wedge \Box\chi_1 \wedge \dots \wedge \Box\chi_n \rightarrow \Diamond\gamma \in \mathbf{S4},$$

$$\Diamond\gamma \rightarrow \neg\Box\chi_{n+1} \vee \dots \vee \neg\Box\chi_m \in \mathbf{S4},$$

contrary to  $t$  being inseparable. Let  $t' = (\Gamma', \Delta')$  be a complete inseparable extension of  $t_0$ . By the definition of  $t_0$ , we have  $tRt'$  and so  $\varphi_1 \in \Gamma'$ , contrary to  $\neg\varphi_1 \in \Gamma_0 \subseteq \Gamma'$  and  $t'$  being inseparable.

Suppose now that  $\Box\varphi_1 \in \Gamma$ . Then for every  $t' = (\Gamma', \Delta')$  such that  $tRt'$ , we have  $\varphi_1 \in \Gamma$  and so  $t' \models \varphi_1$ . Consequently,  $t \models \Box\varphi_1$ . The formula  $\psi$  is treated in the dual way.

To complete the proof it remains to observe that  $\mathfrak{M} \not\models \alpha \rightarrow \beta$ . ■

This proof does not always go through for different kinds of logics. However, sometimes suitable modifications are possible.

**THEOREM 97.** ***GL** has the interpolation property.*

**Proof.** Suppose  $\alpha \rightarrow \beta$  has no interpolant in **GL**. Our goal is to construct a finite irreflexive transitive frame refuting  $\alpha \rightarrow \beta$ .

This time we consider finite pairs  $t = (\Gamma, \Delta)$  such that all formulas in  $\Gamma$  and  $\Delta$  are constructed from variables and their negations using  $\wedge, \vee, \Box, \Diamond$ . Without loss of generality we will assume  $\alpha$  and  $\beta$  to be formulas of that sort. Say that  $t$  is *separable* if there is a formula  $\gamma$  with  $\mathbf{Var}\gamma \subseteq \mathbf{Var}\alpha \cap \mathbf{Var}\beta$  such that  $\bigwedge \Gamma \rightarrow \gamma \in \mathbf{GL}$  and  $\gamma \rightarrow \bigvee \Delta \in \mathbf{GL}$ . It should be clear that if  $t = (\Gamma, \Delta)$  is a finite inseparable pair then in the same way as in the proof of Theorem 95 but taking only subformulas of  $\alpha$  and  $\beta$  we can obtain a finite inseparable pair  $t^* = (\Gamma^*, \Delta^*)$  satisfying the conditions: for every  $\varphi \in \mathbf{Sub}\alpha$  and  $\psi \in \mathbf{Sub}\beta$ , one of the formulas  $\varphi$  and  $\neg\varphi$  (an equivalent formula of the form under consideration, to be more precise) is in  $\Gamma^*$  and one of  $\psi$  and  $\neg\psi$  is in  $\Delta^*$ .



Now we construct by induction a finite rooted model for **GL** refuting  $\alpha \rightarrow \beta$ . As its root we take  $(\{\alpha\}^*, \{\beta\}^*)$ . If we have already put in our model a pair  $t = (\Gamma, \Delta)$  and it has not been considered yet, then for every  $\diamond\varphi \in \Gamma$  and every  $\Box\psi \in \Delta$ , we add to the model the pairs

$$t_1 = (\{\chi, \Box\chi, \Box\neg\varphi, \varphi : \Box\chi \in \Gamma\}^*, \{\chi, \diamond\chi : \diamond\chi \in \Delta\}^*),$$

$$t_2 = \{\chi, \Box\chi : \Box\chi \in \Gamma\}^*, \{\chi, \diamond\chi, \diamond\neg\psi, \psi : \diamond\chi \in \Delta\}^*.$$

One can readily show that if  $t$  is inseparable then  $t_1$  and  $t_2$  are also inseparable. Put  $tR't_1$  and  $tR't_2$ . The process of adding new pairs must eventually terminate, since each step reduces the number of formulas of the form  $\diamond\varphi$  and  $\Box\psi$  in the left and right parts of pairs. Let  $W$  be the set of all pairs constructed in this way and  $R$  the transitive closure of  $R'$ . Clearly, the resulting frame  $\mathfrak{F} = \langle W, R \rangle$  validates **GL**. Define a valuation  $\mathfrak{V}$  in  $\mathfrak{F}$  by taking, for each variable  $p$ ,

$$\mathfrak{V}(p) = \{(\Gamma, \Delta) \in W : p \in \Gamma\}.$$

As in the proof of Theorem 95, it is easily shown that  $\alpha \rightarrow \beta$  is refuted in  $\mathfrak{F}$  under  $\mathfrak{V}$ . ■

To clarify the algebraic meaning of interpolation we require the following well known proposition.

**PROPOSITION 98.** *If  $\nabla$  is a normal filter<sup>12</sup> in a modal algebra  $\mathfrak{A}$  then the relation  $\sim_\nabla$ , defined by  $a \sim_\nabla b$  iff  $a \leftrightarrow b \in \nabla$ , is a congruence relation. The map  $\nabla \mapsto \sim_\nabla$  is an isomorphism from the lattice of normal filters in  $\mathfrak{A}$  onto the lattice of congruences in  $\mathfrak{A}$ .*

Denote by  $\mathfrak{A}/\nabla$  the quotient algebra  $\mathfrak{A}/\sim_\nabla$  and let  $\|a\|_\nabla = \{b : a \sim_\nabla b\}$ .

Say that a class  $\mathcal{C}$  of algebras is *amalgamable* if for all algebras  $\mathfrak{A}_0, \mathfrak{A}_1, \mathfrak{A}_2$  in  $\mathcal{C}$  such that  $\mathfrak{A}_0$  is embedded in  $\mathfrak{A}_1$  and  $\mathfrak{A}_2$  by isomorphisms  $f_1$  and  $f_2$ , respectively, there exist  $\mathfrak{A} \in \mathcal{C}$  and isomorphisms  $g_1$  and  $g_2$  of  $\mathfrak{A}_1$  and  $\mathfrak{A}_2$  into  $\mathfrak{A}$  with  $g_1(f_1(x)) = g_2(f_2(x))$ , for any  $x$  in  $\mathfrak{A}_0$ . If in addition we have

$$g_i(x) \leq g_j(y) \text{ implies } \exists z \in A_0 (x \leq_i f_i(z) \text{ and } f_j(z) \leq_j y)$$

for all  $x \in A_i, y \in A_j$  such that  $\{i, j\} = \{1, 2\}$ , then  $\mathcal{C}$  is called *superamalgamable*. Here  $A_i$  is the universe of  $\mathfrak{A}_i$  and  $\leq_i$  its lattice order.

**THEOREM 99** (Maksimova 1979).  *$L$  has the interpolation property iff the variety  $\text{Alg}L$  of modal algebras for  $L$  is superamalgamable.  $L$  has the  $\vdash^*$ -interpolation property iff  $\text{Alg}L$  is amalgamable.*

<sup>12</sup>A filter  $\nabla$  is *normal* (or open, as in Section 10 of *Basic Modal Logic*) if  $\Box a \in \nabla$  whenever  $a \in \nabla$ .

**Proof.** We prove only the former claim. ( $\Rightarrow$ ) Suppose  $L$  has the interpolation property and  $\mathfrak{A}_0, \mathfrak{A}_1, \mathfrak{A}_2$  are modal algebras for  $L$  such that  $\mathfrak{A}_0$  is a subalgebra of both  $\mathfrak{A}_1$  and  $\mathfrak{A}_2$ . With each element  $a \in A_i, i = 0, 1, 2$ , we associate a variable  $p_a^i$  in such a way that, for  $a \in A_0, p_a^0 = p_a^1 = p_a^2$ . Denote by  $\mathcal{L}_i$  the language with the variables  $p_a^i, \text{ for } a \in A_i, i = 0, 1, 2$ , and let  $\mathcal{L} = \mathcal{L}_1 \cup \mathcal{L}_2$ . We will assume that  $\mathcal{L}$  is the language of  $L$ .

Fix the valuation  $\mathfrak{V}_i$  of  $\mathcal{L}_i$  in  $\mathfrak{A}_i$ , defined by  $\mathfrak{V}_i(p_a^i) = a$ , and put

$$\Sigma_i = \{\varphi \in \mathbf{For}\mathcal{L}_i : \mathfrak{V}_i(\varphi) = \top\}.$$

Let  $\Sigma$  be the closure of  $\Sigma_1 \cup \Sigma_2 \cup L$  under modus ponens. We show that, for every  $\varphi \in \mathbf{For}\mathcal{L}_i, \psi \in \mathbf{For}\mathcal{L}_j$  such that  $\{i, j\} = \{1, 2\}$ ,

$$(13) \quad \varphi \rightarrow \psi \in \Sigma \text{ iff } \exists \chi \in \mathbf{For}\mathcal{L}_0 \ (\varphi \rightarrow \chi \in \Sigma_i \text{ and } \chi \rightarrow \psi \in \Sigma_j).$$

Suppose  $\varphi \rightarrow \psi \in \Sigma$ . Then there exist finite sets  $\Gamma_i \subseteq \Sigma_i$  and  $\Gamma_j \subseteq \Sigma_j$  such that

$$\bigwedge \Gamma_i \wedge \varphi \rightarrow (\bigwedge \Gamma_j \rightarrow \psi) \in L.$$

Since  $L$  has interpolation, there is a formula  $\chi \in \mathbf{For}\mathcal{L}_0$  such that

$$\bigwedge \Gamma_i \wedge \varphi \rightarrow \chi \in L, \quad \bigwedge \Gamma_j \rightarrow (\chi \rightarrow \psi) \in L,$$

from which  $\varphi \rightarrow \chi \in \Sigma_i$  and  $\chi \rightarrow \psi \in \Sigma_j$ . The converse implication is obvious.

Now construct an algebra  $\mathfrak{A}$  by taking the set  $\{\|\varphi\| : \varphi \in \Sigma\}$  as its universe, where  $\|\varphi\| = \{\psi : \varphi \leftrightarrow \psi \in \Sigma\}$ ,  $\|\varphi\| \wedge \|\psi\| = \|\varphi \wedge \psi\|$  and  $\odot\|\varphi\| = \|\odot\varphi\|$ , for  $\odot \in \{\neg, \Box\}$ . One can readily prove that  $\mathfrak{A} \in \text{Alg}L$ . Define maps  $g_i$  from  $\mathfrak{A}_i$  into  $\mathfrak{A}$  by taking  $g_i(a) = \|p_a^i\|$ . It is not difficult to show that  $g_i$  is an embedding of  $\mathfrak{A}_i$  in  $\mathfrak{A}$ . And for  $a \in A_0$ , we have

$$g_1(a) = \|p_a^0\| = g_2(a).$$

It remains to check the condition for superamalgamability: Suppose  $a \in A_i, b \in A_j, \{i, j\} = \{1, 2\}$ , and  $g_i(a) \leq g_j(b)$ . Then  $g_i(a) \rightarrow g_j(b) = \top$  and so  $\|p_a^i \rightarrow p_b^j\| = \top$ , i.e.,  $p_a^i \rightarrow p_b^j \in \Sigma$ . By (13), we have  $\chi \in \mathbf{For}\mathcal{L}_0$  with  $\mathfrak{V}(\chi) = c$  such that  $a \leq_i c \leq_j b$ .

( $\Leftarrow$ ) Assuming  $\text{Alg}L$  to be superamalgamable, we show that  $L$  has the interpolation property. To this end we require

**LEMMA 100.** *Suppose  $\mathfrak{A}_0$  is a subalgebra of modal algebras  $\mathfrak{A}_1$  and  $\mathfrak{A}_2$ ,  $a \in A_1, b \in A_2$  and there is no  $c \in A_0$  such that  $a \leq_1 c \leq_2 b$ . Then there are ultrafilters  $\nabla_1$  in  $\mathfrak{A}_1$  and  $\nabla_2$  in  $\mathfrak{A}_2$  such that  $a \in \nabla_1, b \notin \nabla_2$  and  $\nabla_1 \cap A_0 = \nabla_2 \cap A_0$ .*

Suppose  $\varphi(p_1, \dots, p_m, q_1, \dots, q_n)$  and  $\psi(q_1, \dots, q_n, r_1, \dots, r_l)$  are formulas for which there is no  $\chi(q_1, \dots, q_n)$  such that  $\varphi \rightarrow \chi \in L$  and  $\chi \rightarrow \psi \in L$ . We show that in this case there exists an algebra  $\mathfrak{A} \in \text{Var}L$  refuting  $\varphi \rightarrow \psi$ .

Let  $\mathfrak{A}'_0$ ,  $\mathfrak{A}'_1$  and  $\mathfrak{A}'_2$  be the free algebras in  $\text{Alg}L$  generated by the sets  $\{c_1, \dots, c_n\}$ ,  $\{a_1, \dots, a_m, c_1, \dots, c_n\}$  and  $\{c_1, \dots, c_n, b_1, \dots, b_l\}$ , respectively. According to this definition,  $\mathfrak{A}'_0$  is a subalgebra of both  $\mathfrak{A}'_1$  and  $\mathfrak{A}'_2$ . By Lemma 100, there are ultrafilters  $\nabla_1$  in  $\mathfrak{A}'_1$  and  $\nabla_2$  in  $\mathfrak{A}'_2$  such that we have  $\varphi(a_1, \dots, a_m, c_1, \dots, c_n) \in \nabla_1$  and  $\psi(c_1, \dots, c_n, b_1, \dots, b_l) \notin \nabla_2$ . Define normal filters

$$\nabla_i^* = \{a \in A'_i : \forall m < \omega \ \square^m a \in \nabla_i\}$$

and put  $\mathfrak{A}_1 = \mathfrak{A}'_1 / \nabla_1^*$ ,  $\mathfrak{A}_2 = \mathfrak{A}'_2 / \nabla_2^*$ . Construct an algebra  $\mathfrak{A}_0$  by taking  $A_0 = \{\|a\|_{\nabla_1^*} : a \in A'_0\}$ . By the definition,  $\mathfrak{A}_0$  is a subalgebra of  $\mathfrak{A}_1$ , i.e., is embedded in  $\mathfrak{A}_1$  by the map  $f_1(x) = x$ . One can show that  $\mathfrak{A}_0$  is embedded in  $\mathfrak{A}_2$  by the map  $f_2(\|x\|_{\nabla_1^*}) = \|x\|_{\nabla_2^*}$ . Then there are an algebra  $\mathfrak{A}$  for  $L$  and isomorphisms  $g_1$  and  $g_2$  of  $\mathfrak{A}_1$  and  $\mathfrak{A}_2$  into  $\mathfrak{A}$  satisfying the conditions of superamalgamability. Define a valuation  $\mathfrak{V}$  in  $\mathfrak{A}$  by taking  $\mathfrak{V}(p_i) = g_1(\|a_i\|_{\nabla_1^*})$ ,  $\mathfrak{V}(q_j) = g_1(\|c_j\|_{\nabla_1^*}) = g_2(\|c_j\|_{\nabla_2^*})$  and  $\mathfrak{V}(r_k) = g_2(\|b_k\|_{\nabla_2^*})$ . Then  $\mathfrak{V}(\varphi) \not\leq \mathfrak{V}(\psi)$  because otherwise there would exist  $\{i, j\} = \{1, 2\}$  and  $z \in A_0$  such that  $\mathfrak{V}(\varphi) \leq_i f_i(z)$  and  $f_j(z) \leq_j \mathfrak{V}(\psi)$ . Thus,  $\mathfrak{A} \not\models \varphi \rightarrow \psi$  and so  $\varphi \rightarrow \psi \notin L$ . ■

Using this theorem Maksimova [1979] discovered a surprising fact: there are only finitely many logics in  $\text{NExt}\mathbf{S4}$  with the interpolation property (not more than 38, to be more exact) and all of them turned out to be union-splittings. By Theorem 12, we obtain then

**THEOREM 101** (Maksimova 1979). *There is an algorithm which, given a modal formula  $\varphi$ , decides whether  $\mathbf{S4} \oplus \varphi$  has interpolation.*

We illustrate this result by considering a much simpler class of logics.

**THEOREM 102.** *Only four logics in  $\text{NExt}\mathbf{S5}$  have the interpolation property:  $\mathbf{S5}$  itself, the logic of the two-point cluster,  $\mathbf{Triv}$  and  $\mathbf{For}$ .*

**Proof.** We have already demonstrated how to prove that a logic has interpolation. So now we show only that no logic  $L$  in  $\text{NExt}\mathbf{S5}$  different from those mentioned in the formulation has the interpolation property. Suppose on the contrary that  $L$  has interpolation. We use the amalgamability of the variety of modal algebras for  $L$  to show that an arbitrary big finite cluster is a frame for  $L$ , from which it will follow that  $L = \mathbf{S5}$ .

Figure 10 demonstrates two ways of reducing the three-point cluster to the two-point one. By the amalgamation property, there must exist a cluster reducible to the two depicted copies of the two-point cluster, with the reductions satisfying the amalgamation condition. It should be clear from Fig. 10 that such a cluster contains at least four points. By the same scheme one can prove now that every  $n$ -point cluster validates  $L$ . ■

It would be naive to expect that such a simple picture can be extended to classes like  $\text{NExt}\mathbf{K4}$  or  $\text{NExt}\mathbf{K}$ . Even in  $\text{NExt}\mathbf{GL}$  the situation is quite

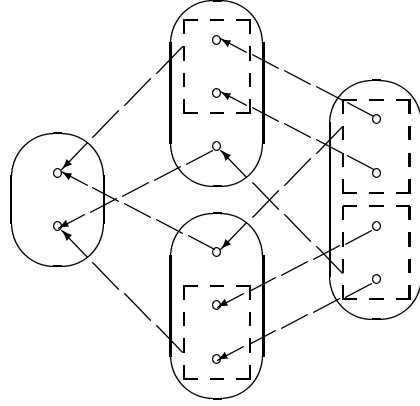


Figure 10.

different from that in  $\text{NExtS4}$ : Maksimova [1989] discovered that there is a continuum of logics in  $\text{NExtGL}$  having the interpolation property. This result is based upon the following observation. For  $L \in \text{NExtK4}$ , we call a formula  $\alpha(p)$  *conservative* in  $\text{NExtL}$  if

$$\Box^+(\alpha(\perp) \wedge \alpha(p) \wedge \alpha(q)) \rightarrow \alpha(p \rightarrow q) \wedge \alpha(\Box p) \in L.$$

For example, in  $\text{NExtS4}$  conservative are  $\Box\Diamond p \rightarrow \Diamond\Box p$ ,  $\Box\Diamond p \leftrightarrow \Diamond\Box p$ , and  $\Box p \leftrightarrow \Diamond p$ .

**THEOREM 103** (Maksimova 1987). *If  $L \in \text{NExtK4}$  has the interpolation property and formulas  $\alpha_i$ , for  $i \in I$ , are conservative in  $\text{NExtL}$ , then the logic  $L \oplus \{\alpha_i : i \in I\}$  also has the interpolation property.*

**Proof.** Suppose  $\varphi \rightarrow \psi \in L \oplus \{\alpha_i : i \in I\}$ . Then there is a finite  $J \subseteq I$ , say  $J = \{1, \dots, l\}$ , such that  $\varphi \rightarrow \psi \in L \oplus \{\alpha_i : i \in J\}$  and so, as follows from the definition of conservative formulas and the Deduction Theorem for  $\mathbf{K4}$ ,

$$\Box^+ \bigwedge_{j=1}^l (\alpha_j(\perp) \wedge \alpha_j(p_1) \wedge \dots \wedge \alpha_j(p_n)) \rightarrow (\varphi \rightarrow \psi) \in L,$$

where  $p_1, \dots, p_m, p_{m+1}, \dots, p_k$  and  $p_{m+1}, \dots, p_k, p_{k+1}, \dots, p_n$  are all the variables in  $\varphi$  and  $\psi$ , respectively. Consequently

$$\begin{aligned} & \Box^+ \bigwedge_{j=1}^l (\alpha_j(\perp) \wedge \alpha_j(p_1) \wedge \dots \wedge \alpha_j(p_k)) \wedge \varphi \rightarrow \\ & (\Box^+ \bigwedge_{j=1}^l (\alpha_j(p_{m+1}) \wedge \dots \wedge \alpha_j(p_n)) \rightarrow \psi) \in L. \end{aligned}$$

Since  $L$  has the interpolation property, there is  $\chi(p_{m+1}, \dots, p_k)$  such that

$$\begin{aligned} \Box^+ \bigwedge_{j=1}^l (\alpha_j(\perp) \wedge \alpha_j(p_1) \wedge \dots \wedge \alpha_j(p_k)) \wedge \varphi &\rightarrow \chi \in L, \\ \Box^+ \bigwedge_{j=1}^l (\alpha_j(p_{m+1}) \wedge \dots \wedge \alpha_j(p_n)) &\rightarrow (\chi \rightarrow \psi) \in L. \end{aligned}$$

Then we obtain  $\varphi \rightarrow \chi \in L \oplus \{\alpha_i : i \in I\}$  and  $\chi \rightarrow \psi \in L \oplus \{\alpha_i : i \in I\}$ , i.e.,  $\chi$  is an interpolant for  $\varphi \rightarrow \psi$  in  $L \oplus \{\alpha_i : i \in I\}$ .  $\blacksquare$

Using the formulas

$$\alpha_i = \Box^+(\Diamond^{i+1}\top \wedge \Box^{i+2}\perp \rightarrow \Box^{i+1}p \vee \Box^{i+1}\neg p)$$

which are conservative in  $\text{NExtGL}$ , one can readily construct a continuum of logics in this class with the interpolation property. The set of logics in  $\text{NExtGL}$  without interpolation is also continual.

In general, an interpolant  $\gamma$  for an implication  $\alpha \rightarrow \beta \in L$  depends on both  $\alpha$  and  $\beta$ . Say that a logic  $L$  has *uniform interpolation* if, for any finite set of variables  $\Xi$  and any formula  $\alpha$ , there exists a formula  $\gamma$  such that  $\mathbf{Var}\gamma \subseteq \Xi$  and  $\alpha \rightarrow \gamma \in L$ ,  $\gamma \rightarrow \beta \in L$  whenever  $\mathbf{Var}\alpha \cap \mathbf{Var}\beta \subseteq \Xi$  and  $\alpha \rightarrow \beta \in L$ . In this case  $\gamma$  is called a *post-interpolant* for  $\alpha$  and  $\Xi$ . Roughly speaking, a logic has uniform interpolation if we can choose an interpolant for  $\alpha \rightarrow \beta \in L$  independently from the actual shape of  $\beta$ . Uniform interpolation was first investigated by Pitts [1992] who proved that intuitionistic logic enjoys it. It is fairly easy to find multiple examples of modal logics with uniform interpolation by observing that any locally tabular logic with interpolation has uniform interpolation as well. Indeed, for every formula  $\alpha$  and every set of variables  $\Xi$ , we can define a post-interpolant  $\gamma$  as the conjunction of a maximal set of pairwise non-equivalent in  $L$  formulas  $\gamma'$  such that  $\mathbf{Var}\gamma' \subseteq \Xi$  and  $\alpha \rightarrow \gamma' \in L$  (which is finite in view of the local tabularity of  $L$ ). It follows, for instance, that **S5** has uniform interpolation. In general, however, interpolation does not imply uniform interpolation: [Ghilardi and Zawadowski 1995] showed that **S4** does not enjoy the latter, witness the following formula without a post-interpolant for  $\{r\}$  in **S4**

$$p \wedge \Box(p \rightarrow \Diamond q) \wedge \Box(q \rightarrow \Diamond p) \wedge \Box(p \rightarrow r) \wedge \Box(q \rightarrow \neg r).$$

Only a few positive results on the uniform interpolation of modal logics are known: Shavrukov [1993] proved it for **GL**, Ghilardi [1995] for **K**, and Visser [1996] for **Grz**.

A property closely related to interpolation is so called Halldén completeness. A logic  $L$  is said to be *Halldén complete* if  $\varphi \vee \psi \in L$  and

$\mathbf{Var}\varphi \cap \mathbf{Var}\psi = \emptyset$  imply  $\varphi \in L$  or  $\psi \in L$ . Since every variable free formula is equivalent in  $\mathbf{D}$  either to  $\top$  or to  $\perp$ ,  $L \in \text{Ext}\mathbf{D}$  is Halldén complete whenever it has interpolation.  $\mathbf{K}$ ,  $\mathbf{K4}$ ,  $\mathbf{GL}$  are examples of Halldén incomplete logics with interpolation: each of them contains  $\diamond\top \vee \neg\diamond\top$  but not  $\diamond\top$  and  $\neg\diamond\top$ . On the other hand,  $\mathbf{S4.3}$  is a Halldén complete logic (see [van Benthem and Humberstone 1983]) without interpolation (see [Maksimova 1982a]). Actually, there is a continuum of Halldén complete logics in  $\text{NExt}\mathbf{S4}$  (see [Chagrov and Zakharyashev 1993]).

Halldén completeness has an interesting lattice-theoretic characterization.

**THEOREM 104** (Lemmon 1966c). *A logic  $L \in \text{Ext}\mathbf{K}$  is Halldén complete iff it is  $\cap$ -irreducible in  $\text{Ext}L$ .*

Since the lattice  $\text{Ext}\mathbf{S5}$  is linearly ordered by inclusion, all logics above  $\mathbf{S5}$  are Halldén complete. There are various semantic criteria for Halldén completeness (see e.g. [Maksimova 1995]). Here we note only the following generalization of the result of [van Benthem and Humberstone 1983].

**THEOREM 105.** *Suppose a logic  $L \in \text{Ext}\mathbf{K}$  is characterized by a class  $\mathcal{C}$  of descriptive rooted frames with distinguished roots. Then  $L$  is Halldén complete iff, for all frames  $\langle \mathfrak{F}_1, d_1 \rangle$  and  $\langle \mathfrak{F}_2, d_2 \rangle$  in  $\mathcal{C}$ , there is a frame  $\langle \mathfrak{F}, d \rangle$  for  $L$  reducible<sup>13</sup> to both  $\langle \mathfrak{F}_1, d_1 \rangle$  and  $\langle \mathfrak{F}_2, d_2 \rangle$ .*

For more results and references on Halldén completeness consult [Chagrov and Zakharyashev 1991].

## 2 POLYMODAL LOGICS

So far we have confined ourselves to considering modal logics with only one necessity operator. From a theoretical point of view this restriction is not such a great loss as it may seem at first sight. In fact, really important concepts of modal logic do not depend on the number of boxes and can be introduced and investigated on the basis of just one. We shall give a precise meaning to this claim in Section 2.3 below where it is shown that polymodal logic is reduced in a natural way to unimodal logic. However, there are at least two reasons for a detailed discussion of polymodal logic in this chapter.

First, a number of interesting phenomena are easily missed in unimodal logic and actually appear in a representative form only in the polymodal case. For example, with the exception of  $\text{NExt}\mathbf{K4.3}$  and  $QCS\mathcal{F}$  all known general decidability results in unimodal logic have been obtained by proving the finite model property. In fact, nearly all natural classes of logics in  $\text{NExt}\mathbf{K}$  turned out to be describable by their finite frames. The situation

---

<sup>13</sup>By reductions that map  $d$  to  $d_i$ .

drastically changes with the addition of just one more box. Even in the case of linear tense logics or bimodal provability logics one has to start with a thorough investigation of their infinite frames: FMP becomes a rather rare guest. While the result on NExt**K4.3** indicated the need for general methods of establishing decidability without FMP, this need becomes of vital importance only in the context of polymodal logic.

The second reason is that various applications of modal logic require polymodal languages. For example, in tense logic we have two necessity-like operators  $\Box_1$  and  $\Box_2$ . One of them, say the former, is interpreted as “it will always be true” and the other as “it was always true”. Kripke frames for tense logics are structures  $\langle W, R_1, R_2 \rangle$  with two binary relations  $R_1$  and  $R_2$  such that  $R_2$  coincides with the converse  $R_1^{-1}$  of  $R_1$  (which reflects the fact that a moment  $x$  is earlier than  $y$  iff  $y$  is later than  $x$ ). The characteristic axioms connecting the two tense operators are

$$p \rightarrow \Box_1 \Diamond_2 p \text{ and } p \rightarrow \Box_2 \Diamond_1 p.$$

For more information about tense systems consult *Basic Tense Logic*.

Another example is basic temporal logic in which we have two necessity-like operators: one of them—usually called **Next**—is interpreted by the successor relation in  $\omega$  and the other by its transitive and reflexive closure. Details can be found in [Seegerberg 1989]. Propositional dynamic logic **PDL** and its extensions, like deterministic **PDL**, can also be regarded as polymodal logics (see *Dynamic Logic*).

A number of provability logics use two or more modal operators; see e.g. Boolos [1993]. In **GLB**, for instance, we have one operator  $\Box_1$  understood as provability in PA and another operator  $\Box_2$  interpreted as  $\omega$ -provability in PA. The unimodal fragments of **GLB** coincide with **GL**. The axioms connecting  $\Box_1$  and  $\Box_2$  are

$$\Box_1 p \rightarrow \Box_2 p \text{ and } \Diamond_1 p \rightarrow \Box_2 \Diamond_1 p.$$

In epistemic logics we need an operator  $\Box_i$  for each agent  $i$ ;  $\Box_i \varphi$  is interpreted as “agent  $i$  believes (or knows)  $\varphi$ ”. One possible way to axiomatize the logic of knowledge with  $m$  agents is to take the axioms of **S5** for each agent without any principles connecting different  $\Box_i$  and  $\Box_j$ . We denote the resultant logic by  $\bigotimes_{i=1}^m \mathbf{S5}$ . Often  $\bigotimes_{i=1}^m \mathbf{S5}$  is extended by the common knowledge operator **C** with the intended meaning

$$C\varphi = E\varphi \wedge E^2\varphi \wedge \dots \wedge E^n\varphi \wedge \dots, \text{ where } E\varphi = \bigwedge_{i=1}^m \Box_i \varphi$$

(see e.g. [Halpern and Moses 1992] and [Meyer and van der Hoek 1995]).

The reader will find more items for this list in other chapters of the *Handbook*.

From the semantical point of view, many standard polymodal logics can be obtained by applying Boolean or various natural closure operators to the accessibility relations of Kripke frames. For instance, in frames  $\langle W, R_1, \dots, R_n \rangle$  for epistemic logic the common knowledge operator is interpreted by the transitive closure of  $R_1 \cup \dots \cup R_n$ . Tense frames result from usual  $\langle W, R \rangle$  by adding the converse of  $R$ . Humberstone [1983] and Goranko [1990a] study the bimodal logic of *inaccessible worlds* determined by frames of the form  $\langle W, R, W^2 - R \rangle$ . This list of examples can be continued; for a general approach and related topics consult [Goranko 1990b; Gargov *et al.* 1987; Gargov and Passy 1990].

Let us see now how polymodal logics in general fit into the theory developed so far. We begin by demonstrating how the concepts introduced in the unimodal case transfer to polymodal logic and showing that a few general results—like Sahlqvist’s and Blok’s Theorems—have natural analogues in polymodal logic. We hope to convince the reader that up to this point no new difficulties arise when one switches from the unimodal language to the polymodal one. After that, in Section 2.2, we start considering subtler features of polymodal logics.

### 2.1 From unimodal to polymodal

Let  $\mathcal{L}_I$  be the propositional language with a finite number of necessity operators  $\Box_i$ ,  $i \in I$ . A *normal polymodal logic* in  $\mathcal{L}_I$  is a set of  $\mathcal{L}_I$ -formulas containing all classical tautologies, the axioms  $\Box_i(p \rightarrow q) \rightarrow (\Box_i p \rightarrow \Box_i q)$  for all  $i \in I$ , and closed under substitution, modus ponens and the rule of necessitation  $\varphi/\Box_i \varphi$  for every  $i \in I$ . If the language is clear from the context, we call these logics just (*normal*) *modal logics* and denote by  $\text{NExt}L$  the family of all normal extensions of  $L$  (in the language  $\mathcal{L}_I$ ). The smallest normal modal logic with  $n$  necessity operators is denoted by  $\mathbf{K}_n$  ( $\mathbf{K} = \mathbf{K}_1$ , of course).

Given a logic  $L_0$  in  $\mathcal{L}_I$  and a set of  $\mathcal{L}_I$ -formulas  $\Gamma$ , we again denote by  $L_0 \oplus \Gamma$  the smallest normal logic (in  $\mathcal{L}_I$ ) containing  $L_0 \cup \Gamma$ . A number of other notions and results also transfer in a rather straightforward way, e.g. Theorems 4 and 6, Proposition 5 and all concepts involved in their formulations. More care has to be taken to generalize Theorems 1, 2 and 3. Denote by  $\mathbf{M}_I^*$  the set of non-empty strings (words) over  $\{\Box_i : i \in I\}$  which do not contain any  $\Box_i$  twice and put

$$\Box_I \varphi = \bigwedge \{ \mathbf{M} \varphi : \mathbf{M} \in \mathbf{M}_I^* \}, \quad \Box_I^{\leq m} \varphi = \bigwedge \{ \Box_I^n \varphi : n \leq m \}.$$

In the language  $\mathcal{L}_I$  the operator  $\Box_I$  serves as a sort of surrogate for  $\Box$  in  $\mathbf{K}$ . For example, the following polymodal version of Theorem 1 holds.

**THEOREM 106 (Deduction).** *For every modal logic  $L$  in  $\mathcal{L}_I$ , every set of*



$\mathcal{L}_I$ -formulas  $\Gamma$ , and all  $\mathcal{L}_I$ -formulas  $\varphi$  and  $\psi$ ,

$$\Gamma, \psi \vdash_L^* \varphi \text{ iff } \exists m \geq 0 \Gamma \vdash_L^* \Box_I^{\leq m} \psi \rightarrow \varphi.$$

Theorems 2 and 3 can be reformulated analogously by replacing  $\Box$  with  $\Box_I$  (a logic  $L$  in  $\mathcal{L}_I$  is *n-transitive* if it contains  $\Box_I^{\leq n} p \rightarrow \Box_I^{n+1} p$ ).

Basic semantic concepts are lifted to the polymodal case in a straightforward manner. The algebraic counterpart of  $L \in \text{NExt}\mathbf{K}_n$  is the variety of Boolean algebras with  $n$  unary operators validating  $L$ . A structure  $\mathfrak{F} = \langle W, \langle R_i : i \in I \rangle, P \rangle$  is called a (*general polymodal*) *frame* whenever every  $\langle W, R_i, P \rangle$ , for  $i \in I$ , is a unimodal frame. We then put

$$\Box_i X = \{x \in W : \forall y (xR_i y \rightarrow y \in X)\}.$$

*Differentiated, refined* and *descriptive frames* and the truth-preserving operations can also be defined in the same component-wise way. For instance, a frame  $\mathfrak{F} = \langle W, \langle R_i : i \in I \rangle, P \rangle$  is differentiated if all the unimodal frames  $\langle W, R_i, P \rangle$ , for  $i \in I$ , are differentiated.  $\mathfrak{F} = \langle W, \langle R_i : i \in I \rangle, P \rangle$  is a (*generated*) *subframe* of  $\mathfrak{G} = \langle V, \langle S_i : i \in I \rangle, Q \rangle$  if all  $\langle W, R_i, P \rangle$  are (generated) subframes of  $\langle V, S_i, Q \rangle$ , and  $f$  is a *reduction* of  $\mathfrak{F}$  to  $\mathfrak{G}$  if  $f$  is a reduction of  $\langle W, R_i, P \rangle$  to  $\langle V, S_i, Q \rangle$ , for every  $i \in I$ .

There are some exceptions to this rule. A point  $r$  is called a root of  $\mathfrak{F}$  if it is a root of the unimodal frame  $\langle W, \bigcup_{i \in I} R_i \rangle$ . This does not mean that  $r$  is a root of all unimodal reducts of  $\mathfrak{F}$ . Another important exception: as before, a polymodal frame is  *$\varkappa$ -generated* if the algebra  $\mathfrak{F}^+$  is  $\varkappa$ -generated; however, this does not mean that the unimodal reducts of  $\mathfrak{F}$  are  $\varkappa$ -generated.

**Splittings and the degree of Kripke incompleteness** The semantic criterion of splittings by finite frames given in Theorem 15 transfers to polymodal logics by replacing  $\Box$  with  $\Box_I$ . Again, all finite rooted frames split  $\text{NExt}L_0$ , if  $L_0$  is an  $n$ -transitive logic in  $\mathcal{L}_I$ . Notice, however, that  $n$ -transitivity is a rather strong condition in the polymodal case. For example, it is easily checked that the fusion  $\mathbf{S5} \otimes \mathbf{S5}$  as well as the minimal tense logic  $\mathbf{K4.t}$  containing  $\mathbf{K4}$  are not  $n$ -transitive, for any  $n < \omega$  (see Sections 2.2 and 2.4 for precise definitions). In fact, only  $\circ$  splits the lattice  $\text{NExt}(\mathbf{S5} \otimes \mathbf{S5})$  and only  $\bullet$  splits  $\text{NExt}\mathbf{K4.t}$  (see [Wolter 1993] and [Kracht 1992], respectively).

Call a frame  $\langle W, \langle R_i : i \in I \rangle \rangle$  *cycle free* if the unimodal frame  $\langle W, \bigcup_{i \in I} R_i \rangle$  is cycle free. Kracht [1990] showed that precisely the finite cycle free frames split  $\text{NExt}\mathbf{K}_n$ .

It is not difficult now to extend Blok's result on the degree of Kripke incompleteness to the polymodal case. Note, however, that the degree of incompleteness of  $\mathbf{For}$  in  $\text{NExt}\mathbf{K}_n$  is  $2^{\aleph_0}$  whenever  $n \geq 2$ . So, we do not have a polymodal analog of Makinson's Theorem. (An example of an incomplete

maximal consistent logic in  $\text{NExt}\mathbf{K}_2$  is the logic determined by the tense frame  $\mathfrak{C}(0, \circ)$  introduced in Section 2.5).

**THEOREM 107.** *Let  $n > 1$ . If  $L$  is a union-splitting of  $\text{NExt}\mathbf{K}_n$ , then  $L$  is strictly Kripke complete. Otherwise  $L$  has degree of Kripke incompleteness  $2^{\aleph_0}$  in  $\text{NExt}\mathbf{K}_n$ .*

**Sahlqvist’s Theorem and persistence** The proof of the following polymodal version of Sahlqvist’s Theorem is a straightforward extension of the proof in the unimodal case. Say that  $\varphi$  is a *Sahlqvist formula* (in  $\mathcal{L}_I$ ) if the result of replacing all  $\Box_i$  and  $\Diamond_i$ ,  $i \in I$ , in  $\varphi$  with  $\Box$  and  $\Diamond$ , respectively, is a unimodal Sahlqvist formula.

**THEOREM 108.** *Suppose that  $\varphi$  is equivalent in  $\text{NExt}\mathbf{K}_n$  to a Sahlqvist formula. Then  $\mathbf{K}_n \oplus \varphi$  is  $\mathcal{DF}$ -persistent, and one can effectively construct a first order formula  $\phi(x)$  in  $R_1, \dots, R_n$  and  $=$  such that, for every descriptive or Kripke frame  $\mathfrak{F}$  and every point  $a$  in  $\mathfrak{F}$ ,  $(\mathfrak{F}, a) \models \varphi$  iff  $\mathfrak{F} \models \phi(x)[a]$ .*

Bellissima’s result on the  $\mathcal{DF}$ -persistence of all logics in  $\text{NExt}\mathbf{Alt}_n$  has a polymodal analog as well. Denote by  $\bigotimes_{i \in I} \mathbf{Alt}_n$  the smallest polymodal logic in  $\mathcal{L}_I$  containing  $\mathbf{Alt}_n$  in all its unimodal fragments. It is easy to see that every  $L \in \text{NExt} \bigotimes_{i \in I} \mathbf{Alt}_n$  is  $\mathcal{DF}$ -persistent and so Kripke complete. However, in contrast to the lattice  $\text{NExt}\mathbf{Alt}_1$ —which is countable and all logics in which have FMP (see [Seegerberg 1986] and [Bellissima 1988])—the lattice  $\text{NExt}(\mathbf{Alt}_1 \otimes \mathbf{Alt}_1)$  is rather complex: as was shown by Grefe [1994], it contains logics without FMP (even without finite frames at all) and uncountably many maximal consistent logics.

**Some FMP results** Fine’s Theorem on uniform logics can be extended to a suitable class of polymodal logics in  $\mathcal{L}_I$ , namely those logics that contain  $\Diamond_i \top$ , for all  $i \in I$ , and are axiomatizable by formulas  $\varphi$  in which all maximal sequences of nested modal operators coincide with respect to the distribution of the indices  $i$  of  $\Box_i$  and  $\Diamond_i$ ,  $i \in I$ .

Now consider a result of Lewis [1974] which we have not proved in its unimodal formulation. Call a normal polymodal logic *non-iterative* if it is axiomatizable by formulas without nested modalities. Examples of non-iterative logics are  $\mathbf{T} = \mathbf{K} \oplus \Box p \rightarrow p$ ,  $\mathbf{Alt}_m \otimes \mathbf{Alt}_n$  and  $\mathbf{K}_2 \oplus \Box_2 p \rightarrow \Box_1 p$ .

**THEOREM 109** (Lewis 1974). *All non-iterative normal logics have FMP.*

**Proof.** Suppose the axioms of  $L = \mathbf{K}_n \oplus \Gamma$  have no nested modal operators and  $\varphi \notin L$ . By a  $\varphi$ -description we mean any set of subformulas of  $\varphi$  together with the negations of the remaining formulas in  $\mathbf{Sub}\varphi$ . For each  $L$ -consistent  $\varphi$ -description  $\Theta$  select a maximal  $L$ -consistent set  $\Delta_\Theta$  containing  $\Theta$ . Denote by  $W$  the (finite) set of the selected  $\Delta_\Theta$  and define

$\mathfrak{F} = \langle W, \langle R_i : i \in I \rangle \rangle$  and  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{W} \rangle$  by taking

$$\Delta_\Theta R_i \Delta_\Psi \text{ iff } \diamond_i \bigwedge \Psi \in \Delta_\Theta$$

and  $\mathfrak{W}(p) = \{\Delta_\Theta \in W : p \in \Delta_\Theta\}$ . It is easily proved that  $(\mathfrak{M}, \Delta_\Theta) \models \psi$  iff  $\psi \in \Delta_\Theta$ , for all subformulas  $\psi$  of  $\varphi$  and  $\Delta_\Theta \in W$ . Hence  $\mathfrak{F} \not\models \varphi$ . It is also easy to see that for all truth-functional compounds  $\psi$  of subformulas in  $\varphi$ ,

$$(14) \quad (\mathfrak{M}, \Delta_\Theta) \models \diamond_i \psi \text{ iff } \diamond_i \psi \in \Delta_\Theta.$$

Consider now a model  $\mathfrak{M}' = \langle \mathfrak{F}, \mathfrak{W}' \rangle$  and  $\chi \in \Gamma$ . For each variable  $p$  put

$$\psi_p = \bigvee \left\{ \bigwedge \Theta : \Delta_\Theta \in \mathfrak{W}(p) \right\}$$

and denote by  $\chi'$  the result of substituting  $\psi_p$  for  $p$ , for each  $p$  in  $\chi$ . Then  $\mathfrak{M}' \models \chi$  iff  $\mathfrak{M} \models \chi'$ . In view of (14), we have  $\mathfrak{M} \models \chi'$  because  $\chi'$  has no nested modalities. Therefore,  $\mathfrak{F} \models \chi$  and so  $\mathfrak{F} \models L$ .  $\blacksquare$

**Tabular Logics** Needless to say that all polymodal tabular logics are finitely axiomatizable and have only finitely many extensions. (The proof is the same as in the unimodal case.) A more interesting observation concerns the complexity of polymodal logics whose unimodal fragments are tabular or pretabular. In fact, it is not difficult to construct two tabular unimodal logics  $L_1$  and  $L_2$  such that their fusion  $L_1 \otimes L_2$  has uncountably many normal extensions (see e.g. [Grefe 1994]). However, those logics are  $\mathcal{DF}$ -persistent and so Kripke complete. Wolter [1994b] showed that the lattice

$\text{NExt}\mathbf{T}$  can be embedded into the lattice  $\text{NExt}(\text{Log} \overset{\circ}{\circ} \otimes \mathbf{S5})$  in such a way that properties like FMP, decidability and Kripke completeness are reflected under this embedding. It follows that almost all “negative” phenomena of modal logic are exhibited by bimodal logics one unimodal fragment of which is tabular and the other pretabular.

## 2.2 Fusions

The simplest way of constructing polymodal logics from unimodal ones is to form the *fusions* (alias *independent joins*) of them. Namely, given two unimodal logics  $L_1$  and  $L_2$  in languages with the same set of variables and distinct modal operators  $\Box_1$  and  $\Box_2$ , respectively, the *fusion*  $L_1 \otimes L_2$  of  $L_1$  and  $L_2$  is the smallest bimodal logic to contain  $L_1 \cup L_2$ . If  $\Gamma_1$  and  $\Gamma_2$  axiomatize  $L_1$  and  $L_2$ , then  $L_1 \otimes L_2$  is axiomatized by  $\Gamma_1 \cup \Gamma_2$ , i.e.,  $L_1 \otimes L_2 = \mathbf{K}_2 \oplus \Gamma_1 \oplus \Gamma_2$ . So the fusions are precisely those bimodal logics that are axiomatizable by sets of formulas each of which contains only one of  $\Box_1, \Box_2$ . From the model-theoretic point of view this means that a frame  $\langle W, R_1, R_2, P \rangle$  validates  $L_1 \otimes L_2$  iff  $\langle W, R_i, P \rangle \models L_i$  for  $i = 1, 2$ .

PROPOSITION 110 (Thomason 1980). *If logics  $L_1$  and  $L_2$  are consistent, then  $L_1 \otimes L_2$  is a conservative extension of both  $L_1$  and  $L_2$ .*

**Proof.** Suppose for definiteness that  $\varphi \notin L_1$ , for some formula  $\varphi$  in the language of  $L_1$ , and consider the Tarski–Lindenbaum algebras

$$\mathfrak{A}_{L_1}(\omega) = \langle A, \wedge^A, \neg^A, \Box_1 \rangle \text{ and } \mathfrak{A}_{L_2}(\omega) = \langle B, \wedge^B, \neg^B, \Box_2 \rangle.$$

The Boolean reducts of them are countably infinite atomless Boolean algebras which are known to be isomorphic (see e.g. [Koppelberg 1988]). So we may assume that  $A = B$ ,  $\wedge^A = \wedge^B$ ,  $\neg^A = \neg^B$ . Since the algebra  $\mathfrak{A}_{L_1}(\omega)$  refutes  $\varphi$ ,  $\langle A, \wedge^A, \neg^A, \Box_1, \Box_2 \rangle$  is then an algebra for  $L_1 \otimes L_2$  refuting  $\varphi$ . ■

Having constructed the fusion of logics, it is natural to ask which of their properties it inherits. For example, the first order theory of a single equivalence relation has the finite model property and is decidable, but the theory of two equivalence relations is undecidable and so does not have the finite model property (see [Janiczak 1953]). So neither decidability nor the finite model property is preserved under joins of first order theories. On the other hand, as was shown by Pigozzi [1974], decidability is preserved under fusions of equational theories in languages with mutually disjoint sets of operation symbols.

For modal logics we have:

THEOREM 111. *Suppose  $L_1$  and  $L_2$  are normal unimodal consistent logics and  $\mathcal{P}$  is one of the following properties: FMP, (strong) Kripke completeness, decidability, Halldén completeness, interpolation, uniform interpolation. Then  $L = L_1 \otimes L_2$  has  $\mathcal{P}$  iff both  $L_1$  and  $L_2$  have  $\mathcal{P}$ .*

**Proof.** We outline proofs of some claims in this theorem; the reader can consult [Fine and Schurz 1996], [Kracht and Wolter 1991], and [Wolter 1997b] for more details.

The implication  $(\Rightarrow)$  presents no difficulties. So let us concentrate on  $(\Leftarrow)$ . With each formula  $\varphi$  of the form  $\Box_i\psi$  we associate a new variable  $q_\varphi$  which will be called the *surrogate* of  $\varphi$ . For a formula  $\varphi$  containing no surrogate variables, denote by  $\varphi^1$  the formula that results from  $\varphi$  by replacing all occurrences of formulas  $\Box_2\psi$ , which are not within the scope of another  $\Box_2$ , with their surrogate variables  $q_{\Box_2\psi}$ . So  $\varphi^1$  is a unimodal formula containing only  $\Box_1$ . Denote by  $\Theta^1(\varphi)$  the set of variables in  $\varphi$  together with all subformulas of  $\Box_2\psi \in \mathbf{Sub}\varphi$ . The formula  $\varphi^2$  and the set  $\Theta^2(\varphi)$  are defined symmetrically.

Suppose now that both  $L_1$  and  $L_2$  are Kripke complete and  $\varphi \notin L$ . To prove the completeness of  $L$  we construct a Kripke frame for  $L$  refuting  $\varphi$ . Since we know only how to build refutation frames for the unimodal fragments of  $L$ , the frame is constructed by steps alternating between  $\Box_1$  and  $\Box_2$ . First, since  $L_1$  is complete, there is a unimodal model  $\mathfrak{M}$  based

on a Kripke frame for  $L_1$  and refuting  $\varphi^1$  at its root  $r$ . Our aim now is to ensure that the formulas of the form  $\Box_2\psi$  have the same truth-values as their surrogates  $q_{\Box_2\psi}$ . To do this, with each point  $x$  in  $\mathfrak{M}$  we can associate the formula

$$\varphi_x = \bigwedge \{ \psi \in \Theta^1(\varphi) : (\mathfrak{M}, x) \models \psi^1 \} \wedge \bigwedge \{ \neg\psi : \psi \in \Theta^1(\varphi), (\mathfrak{M}, x) \not\models \psi^1 \},$$

construct a model  $\mathfrak{M}_x$  based on a frame for  $L_2$  and satisfying  $\varphi_x^2$  at its root  $y$ , and then hook  $\mathfrak{M}_x$  to  $\mathfrak{M}$  by identifying  $x$  and  $y$ . After that we can switch to  $\Box_1$  and in the same manner ensure that formulas  $\Box_1\psi$  have the same truth-values as  $q_{\Box_1\psi}$  at all points in every  $\mathfrak{M}_x$ . And so forth.

However, to realize this quite obvious scheme we must be sure that  $\varphi_x$  is really satisfiable in a frame for  $L_2$ , which may impose some restrictions on the models we choose. First, one can show that in the construction above it is enough to deal with points  $x$  accessible from  $r$  by at most  $m = md(\varphi)$  steps. Let  $X$  be the set of all such points. Now, a sufficient and necessary condition for  $\varphi_x$  to be  $L$ - (and so  $L_2$ -) consistent can be formulated as follows. Call a  $\Theta^1(\varphi)$ -description the conjunction of formulas in any maximal  $L$ -consistent subset of  $\Theta^1(\varphi) \cup \{ \neg\psi : \psi \in \Theta^1(\varphi) \}$ . It should be clear that  $\varphi_x$  is  $L$ -consistent iff it is a  $\Theta^1(\varphi)$ -description. Denote by  $\Sigma_1(\varphi)$  the set of all  $\Theta^1(\varphi)$ -descriptions. It follows that all  $\varphi_x$ , for  $x \in X$ , are  $L$ -consistent iff  $(\mathfrak{M}, r) \models \Box_1^{\leq m}(\bigvee \Sigma_1(\varphi))^1$ . In other words, we should start with a model  $\mathfrak{M}$  satisfying  $\varphi^1 \wedge \Box_1^{\leq m}(\bigvee \Sigma_1(\varphi))^1$  at its root  $r$ . Of course, the subsequent models  $\mathfrak{M}_x$ , for  $x \in X$ , must satisfy  $\varphi_x^2 \wedge \Box_2^{\leq m}(\bigvee \Sigma_2(\varphi_x))^2$ , where  $\Sigma_2(\varphi_x)$  is the set of all  $\Theta^2(\varphi_x)$ -descriptions, etc.

In this way we can prove that Kripke completeness is preserved under fusions. The preservation of strong completeness and FMP can be established in a similar manner. The following lemma plays the key role in the proof of the preservation of the four remaining properties.

LEMMA 112. *The following conditions are equivalent for every  $\varphi$ :*

- (i)  $\varphi \in L_1 \otimes L_2$ ;
- (ii)  $\Box_1^{\leq m}(\bigvee \Sigma_1(\varphi))^1 \rightarrow \varphi^1 \in L_1$ , where  $m = md(\varphi)$ ;
- (iii)  $\Box_2^{\leq m}(\bigvee \Sigma_2(\varphi))^2 \rightarrow \varphi^2 \in L_2$ .

For Kripke complete  $L_1$  and  $L_2$  this lemma was first proved by Fine and Schurz [1996] and Kracht and Wolter [1991]; actually, it is an immediate consequence of the consideration above. The proof for the arbitrary case is also based upon a similar construction combined with the algebraic proof of Proposition 110; for details see [Wolter 1997b].

Now we show how one can use this lemma to prove the preservation of the remaining properties. Define  $a^1(\varphi)$  to be the length of the longest

sequence  $\Box_2, \Box_1, \Box_2, \dots$  of boxes starting with  $\Box_2$  such that a subformula of the form  $\Box_2(\dots \Box_1(\dots \Box_2(\dots)))$  occurs in  $\varphi$ . The function  $a^2(\varphi)$  is defined analogously by exchanging  $\Box_1$  and  $\Box_2$ , and  $a(\varphi) = a^1(\varphi) + a^2(\varphi)$ . It is easy to see that

$$a(\varphi) > a(\bigvee \Sigma_1(\varphi)) \quad \text{or} \quad a(\varphi) > a(\bigvee \Sigma_2(\varphi)).$$

The preservation of decidability, Halldén completeness, interpolation, and uniform interpolation can be proved by induction on  $a(\varphi)$  with the help of Lemma 112. We illustrate the method only for Halldén completeness. Notice first that, modulo the Boolean equivalence, we have

$$\bigvee \Sigma_1(\varphi \vee \psi) = \bigvee \Sigma_1(\varphi) \wedge \bigvee \Sigma_1(\psi) \wedge \bigwedge \Delta(\varphi, \psi),$$

where

$$\Delta(\varphi, \psi) = \{\chi_1 \rightarrow \neg \chi_2 : \chi_1 \in \Sigma_1(\varphi), \chi_2 \in \Sigma_1(\psi), \chi_1 \rightarrow \neg \chi_2 \in L\}.$$

Suppose both  $L_1$  and  $L_2$  are Halldén complete. By induction on  $n = a(\varphi \vee \psi)$  we prove that  $\varphi \vee \psi \in L$  implies  $\varphi \in L$  or  $\psi \in L$  whenever  $\varphi$  and  $\psi$  have no common variables. The basis of induction is trivial. So suppose  $a(\varphi \vee \psi) = n > 0$  and  $\varphi \vee \psi \in L$ . We may also assume that  $a(\varphi \vee \psi) > a(\bigvee \Sigma_1(\varphi \vee \psi))$ . By the induction hypothesis, it follows that  $\Delta(\varphi, \psi) = \emptyset$ . Hence, up to the Boolean equivalence,  $\bigvee \Sigma_1(\varphi \vee \psi) = \bigvee \Sigma_1(\varphi) \wedge \bigvee \Sigma_1(\psi)$  and, by Lemma 112,

$$\Box_1^{\leq m}(\bigvee \Sigma_1(\varphi))^1 \wedge \Box_1^{\leq m}(\bigvee \Sigma_1(\psi))^1 \rightarrow (\varphi \vee \psi)^1 \in L_1,$$

for  $m = md(\varphi \vee \psi)$ . Then

$$(\Box_1^{\leq m}(\bigvee \Sigma_1(\varphi))^1 \rightarrow \varphi^1) \vee (\Box_1^{\leq m}(\bigvee \Sigma_1(\psi))^1 \rightarrow \psi^1) \in L_1$$

and, by the Halldén completeness of  $L_1$ , one of the disjuncts in this formula belongs to  $L_1$ . By Lemma 112, this means that  $\varphi \in L$  or  $\psi \in L$ . ■

**REMARK.** This theorem can be generalized to fusions of polymodal logics with polyadic modalities.

Note that in languages with finitely many variables both **GL.3** and **K** are strongly complete but **GL.3**  $\otimes$  **K** is not strongly complete even in the language with one variable (see [Kracht and Wolter 1991]).

It is natural now to ask whether there exist interesting axioms  $\varphi$  containing both  $\Box_1$  and  $\Box_2$  and such that  $(L_1 \otimes L_2) \oplus \varphi$  inherits basic properties of  $L_1, L_2 \in \text{NExtK}$ . Let us start with the observation that even such a simple axiom as  $\Box_1 p \leftrightarrow \Box_2 p$  destroys almost all “good” properties because (i) we can identify the logic  $(L_1 \otimes L_2) \oplus \Box_1 p \leftrightarrow \Box_2 p$  with the sum of the translation of  $L_1$  and  $L_2$  into a common unimodal language and (ii) such properties as

FMP, decidability, and Kripke completeness are not preserved under sums of unimodal logics (see Example 64 and [Chagrov and Zakharyashev 1997]). Even for the simpler formula  $\Box_2 p \rightarrow \Box_1 p$  no general results are available. To demonstrate this we consider the following way of constructing a bimodal logic  $L_u$  for a given  $L \in \text{NExt}\mathbf{K}$ :

$$L_u = (L \otimes \mathbf{S5}) \oplus \Box_2 p \rightarrow \Box_1 p.$$

The modal operator  $\Box_2$  in  $L_u$  is called the *universal modality*. Its meaning is explained by the following lemma:

LEMMA 113 (Goranko and Passy 1992). *For every normal unimodal logic  $L$  and all unimodal formulas  $\varphi$  and  $\psi$ ,*

$$\varphi \vdash_L^* \psi \text{ iff } \vdash_{L_u} \Box_2 \varphi \rightarrow \psi.$$

**Proof.** Follows immediately from Theorem 19 (ii), since

$$\langle W, R, P \rangle \models L \text{ iff } \langle W, R, W \times W, P \rangle \models L_u,$$

for every frame  $\langle W, R, P \rangle$  and every unimodal logic  $L$ . ■

The universal modality is used to express those properties of frames  $\mathfrak{F} = \langle W, R, W \times W \rangle$  that cannot be expressed in the unimodal language. For example,  $\mathfrak{F}$  validates  $\Box_2(p \rightarrow \Diamond_1 p) \rightarrow \neg p$  iff it contains no infinite  $R$ -chains. Recall that there is no corresponding unimodal axiom, since  $\mathbf{K}$  is determined by the class of frames without infinite  $R$ -chains. We refer the reader to [Goranko and Passy 1992] for more information on this matter.

THEOREM 114 (Goranko and Passy 1992). *For any  $L \in \text{NExt}\mathbf{K}$ ,*

- (i)  *$L$  is globally Kripke complete iff  $L_u$  is Kripke complete;*
- (ii)  *$L$  has global FMP iff  $L_u$  has FMP.*

**Proof.** We prove only (i). Suppose that  $L_u$  is Kripke complete and  $\varphi \not\vdash_L^* \psi$ . Then by Lemma 113,  $\Box_2 \varphi \rightarrow \psi \notin L_u$  and so  $\Box_2 \varphi \rightarrow \psi$  is refuted in a Kripke frame  $\mathfrak{F} = \langle W, R_1, R_2 \rangle$  for  $L_u$ . We may assume that  $R_2 = W \times W$ . But then  $\varphi \vdash_L^* \psi$  is refuted in  $\langle W, R_1 \rangle$ . Conversely, suppose that  $L$  is globally Kripke complete and  $\varphi \notin L_u$ , for a (possibly bimodal) formula  $\varphi$ . Using the properties of  $\mathbf{S5}$  it is readily checked that  $\varphi$  is (effectively) equivalent in  $\mathbf{K}_u$  to a formula  $\varphi'$  which is a conjunction of formulas  $\psi$  of the form

$$\psi = \chi_0 \vee \Diamond_2 \chi_1 \vee \Box_2 \chi_2 \vee \Box_2 \chi_3 \vee \cdots \vee \Box_2 \chi_n$$

such that  $\chi_0, \dots, \chi_n$  are unimodal formulas in the language with  $\Box_1$ . Let  $\psi$  be a conjunct of  $\varphi'$  such that  $\psi \notin L_u$ . Then  $\neg \chi_1 \not\vdash_L^* \chi_i$ , for every  $i \in \{0, 2, 3, \dots, n\}$ . Since  $L$  is globally Kripke complete, we have Kripke frames  $\langle W_i, R_i \rangle$  for  $L$  refuting  $\neg \chi_1 \vdash_L^* \chi_i$ , for  $i \in \{0, 2, \dots, n\}$ . Denote by  $\langle W, R \rangle$  the disjoint union of those frames. Then  $\langle W, R, W \times W \rangle$  is a Kripke frame for  $L_u$  refuting  $\varphi$ . ■

We have seen in Section 1.5 that there are Kripke complete logics (logics with FMP) which do not enjoy the corresponding global property. In view of Theorem 114, we conclude that neither FMP nor Kripke completeness is preserved under the map  $L \mapsto L_u$ .

Another interesting way of adding to fusions new axioms mixing the necessity operators is to use the so called *inductive* (or *Segerberg's*) *axioms*. First, we extend the language  $\mathcal{L}_I$  with  $m$  necessity operators by introducing the operators **E** and **C** and then let

$$\mathit{ind} = \{ \mathbf{E}p \leftrightarrow \bigwedge_{i \in I} \Box_i p, \mathbf{C}p \rightarrow \mathbf{E} \mathbf{C}p; \mathbf{C}(p \rightarrow \mathbf{E}p) \rightarrow (p \rightarrow \mathbf{C}p) \}.$$

Now, given  $L \in \mathbf{NExtK}_m$ , we put

$$\mathbf{LEC}_m = (L \otimes \mathbf{K}_E \otimes \mathbf{S4}_C) \oplus \mathit{ind},$$

where  $\mathbf{K}_E$  and  $\mathbf{S4}_C$  are just  $\mathbf{K}$  and  $\mathbf{S4}$  in the languages with **E** and **C**, respectively. The following proposition explains the meaning of the inductive axioms.

**PROPOSITION 115.** *A frame  $\langle W, R_1, \dots, R_m, R_E, R_C \rangle$  validates  $\mathbf{LEC}_m$  iff  $\langle W, R_1, \dots, R_m \rangle \models L$ ,  $R_E = R_1 \cup \dots \cup R_m$  and  $R_C$  is the transitive reflexive closure of  $R_E$ .*

**EXAMPLE 116.** The logic  $(\mathbf{Alt}_1 \oplus \mathbf{D})\mathbf{EC}_1$  is determined by the frame  $\langle \omega, S, \leq \rangle$  in which  $S$  is the successor relation in  $\omega$ . (Here we omit writing  $R_E$  because  $R_E = S$ .) For details consult [Segerberg 1989].<sup>14</sup>

No general results are known about the preservation properties of the map  $L \mapsto \mathbf{LEC}_m$ . In fact, it is easy to extend the counter-examples for the map  $L \mapsto L_u$  to the present case (see [Hemaspaandra 1996]). However, at least in some cases—especially those that are of importance for epistemic logic—the logic  $\mathbf{LEC}_m$  enjoys a number of desirable properties.

**THEOREM 117** (Halpern and Moses 1992). *For every  $m \geq 1$ , the logics  $(\bigotimes_{i=1}^m \mathbf{K})\mathbf{EC}_m$ ,  $(\bigotimes_{i=1}^m \mathbf{S4})\mathbf{EC}_m$  and  $(\bigotimes_{i=1}^m \mathbf{S5})\mathbf{EC}_m$  have FMP.*

**Proof.** We consider only  $L = (\bigotimes_{i=1}^m \mathbf{S5})\mathbf{EC}_m$ . The proof is by filtration and so the main difficulty is to find a suitable “filter”. Suppose that  $\varphi \notin L$  and let  $\mathfrak{M} = \langle \langle W, R_1, \dots, R_m, R_E, R_C \rangle, \mathfrak{U} \rangle$  be the canonical model for  $L$ . Denote by  $\Gamma^\neg$  the closure of a set of formulas  $\Gamma$  under negations and define a filter  $\Phi = \Phi_1^\neg \cup \Phi_2^\neg \cup \Phi_3^\neg$ , where  $\Phi_1 = \mathbf{Sub}\varphi$ ,  $\Phi_2 = \{ \Box_i \psi : \mathbf{E}\psi \in \Phi_1^\neg \}$  and  $\Phi_3 = \{ \mathbf{E} \mathbf{C}\psi, \Box_i \mathbf{C}\psi : \mathbf{C}\psi \in \Phi_1^\neg \}$ . Certainly,  $\Phi$  is finite and closed under subformulas. Now, we filter  $\mathfrak{M}$  through  $\Phi$ , i.e., put  $W^* = \{ [x] : x \in W \}$ ,

<sup>14</sup>Krister Segerberg kindly informed us that this result was independently obtained by D. Scott, H. Kamp, K. Fine and himself.



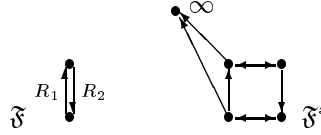


Figure 11.

where  $[x]$  consists of all points that validate the same formulas in  $\Phi$  as  $x$ , and

$$[x]R_i[y] \text{ iff } \forall \Box_i \psi \in \Phi ((\mathfrak{M}, x) \models \Box_i \psi \rightarrow (\mathfrak{M}, y) \models \Box_i \psi),$$

$$R_E^* = R_1^* \cup \dots \cup R_m^*,$$

and  $R_C^*$  is the transitive and reflexive closure of  $R_E^*$ . A rather tedious inductive proof shows that  $\langle W^*, R_1^*, \dots, R_m^*, R_E^*, R_C^* \rangle$  refutes  $\varphi$  under the valuation  $\mathfrak{U}^*(p) = \{[x] : x \models p\}$ ,  $p$  a variable in  $\varphi$ . For details we refer the reader to [Halpern and Moses 1992] and [Meyer and van der Hoek 1995]. ■

It would be of interest to look for big classes of logics  $L$  for which  $LE\mathbf{C}_m$  inherits basic properties of  $L$ .

### 2.3 Simulation

In the preceding section we saw how results concerning logics in  $\mathbf{NExtK}$  can be extended to a certain class of polymodal logics. More generally, we may ask whether—at least theoretically—polymodal logics are reducible to unimodal ones. The first to attack this problem was Thomason [1974b, 1975c] who proved that each polymodal logic  $L$  can be embedded into a unimodal logic  $L^s$  in such a way that  $L$  inherits almost all interesting properties of  $L^s$ . Using this result one can construct unimodal logics with various “negative” properties by presenting first polymodal logics with the corresponding properties, which is often much easier. It was in this way that Thomason [1975c] constructed Kripke incomplete and undecidable unimodal calculi. Kracht [1996] strengthened Thomason’s result by showing that his embedding not only reflects but also (i) preserves almost all important properties and (ii) induces an isomorphism from the lattice  $\mathbf{NExtK}_2$  onto the interval  $[\mathbf{Sim}, \mathbf{K} \oplus \Box \perp]$ , for some normal unimodal logic  $\mathbf{Sim}$ . Thus indeed, in many respects polymodal logics turn out to be reducible to unimodal ones.

Below we outline Thomason’s construction following [Kracht 1999] and [Kracht and Wolter 1999]. To define the unimodal “simulation”  $L^s$  of a bimodal logic  $L$ , let us first transform each bimodal frame into a unimodal one.

So suppose  $\mathfrak{F} = \langle W, R_1, R_2, P \rangle$  is a bimodal frame. Construct a unimodal frame  $\mathfrak{F}^s = \langle W^s, R^s, P^s \rangle$ —the *simulation* of  $\mathfrak{F}$ —by taking

$$\begin{aligned} W^s &= W \times \{1, 2\} \cup \{\infty\}, \\ R^s &= \{ \langle \langle x, 1 \rangle, \langle x, 2 \rangle \rangle : x \in W \} \cup \\ &\quad \{ \langle \langle x, 2 \rangle, \langle x, 1 \rangle \rangle : x \in W \} \cup \\ &\quad \{ \langle \langle x, 1 \rangle, \infty \rangle : x \in W \} \cup \\ &\quad \{ \langle \langle x, 1 \rangle, \langle y, 1 \rangle \rangle : x, y \in W, xR_1y \} \cup \\ &\quad \{ \langle \langle x, 2 \rangle, \langle y, 2 \rangle \rangle : x, y \in W, xR_2y \}, \\ P^s &= \{ (X \times \{2\}) \cup (Y \times \{1\}) \cup Z : X, Y \in P, Z \subseteq \{\infty\} \}. \end{aligned}$$

This construction is illustrated by Fig. 11. One can easily prove that  $\mathfrak{F}^s$  is a Kripke (differentiated, refined, descriptive) frame whenever  $\mathfrak{F}$  is so. Notice also that if  $W = \emptyset$  then  $\mathfrak{F}^s \cong \bullet$ . Now, given a bimodal logic  $L$ , define the *simulation*  $L^s$  of  $L$  to be the unimodal logic

$$\text{Log}\{\mathfrak{F}^s : \mathfrak{F} \models L\}.$$

To formulate the translation which embeds  $L$  into  $L^s$  we require the following formulas and notations:

$$\begin{aligned} \gamma &= \Box \perp & \Box_\gamma \varphi &= \Box(\gamma \rightarrow \varphi) \\ \alpha &= \Diamond \Box \perp & \Box_\alpha \varphi &= \Box(\alpha \rightarrow \varphi) \\ \beta &= \neg \gamma \wedge \neg \Diamond \gamma & \Box_\beta \varphi &= \Box(\beta \rightarrow \varphi). \end{aligned}$$

$\Diamond_\gamma$ ,  $\Diamond_\alpha$  and  $\Diamond_\beta$  are defined dually. Observe that the formula  $\gamma$  is true in  $\mathfrak{F}^s$  only at  $\infty$ ,  $\alpha$  is true precisely at the points in the set  $\{\langle x, 1 \rangle : x \in W\}$ , and  $\beta$  is true at the points  $\{\langle x, 2 \rangle : x \in W\}$  and only at them. Put

$$\begin{aligned} p^s &= p, \\ (\neg \varphi)^s &= \alpha \wedge \neg \varphi^s, \\ (\varphi \wedge \psi)^s &= \varphi^s \wedge \psi^s, \\ (\Box_1 \varphi)^s &= \Box_\alpha \varphi^s, \\ (\Box_2 \varphi)^s &= \Box_\beta \Box_\beta \Box_\alpha \varphi^s. \end{aligned}$$

By an easy induction on the construction of  $\varphi$  one can prove

LEMMA 118. *Let  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{B} \rangle$  be a bimodal model,  $X = \{x : x \models \alpha\}$  and let  $\mathfrak{M}^s = \langle \mathfrak{F}^s, \mathfrak{B}^s \rangle$  be a model such that  $\mathfrak{B}^s(p) \cap X = \mathfrak{B}(p) \times \{1\}$ , for all variables  $p$ . Then for every bimodal formula  $\varphi$ ,*

$$\begin{aligned} (\mathfrak{M}, x) \models \varphi &\text{ iff } (\mathfrak{M}^s, \langle x, 1 \rangle) \models \varphi^s, \\ \mathfrak{M} \models \varphi &\text{ iff } \mathfrak{M}^s \models \alpha \rightarrow \varphi^s, \\ \mathfrak{F} \models \varphi &\text{ iff } \mathfrak{F}^s \models \alpha \rightarrow \varphi^s. \end{aligned}$$

Using this lemma, both consequence relations  $\vdash_L$  and  $\vdash_L^*$  can be reduced to the corresponding consequence relations for  $L^s$ .

PROPOSITION 119. *Let  $L$  be a bimodal logic,  $\Delta$  a set of bimodal formulas and  $\varphi$  a bimodal formula. Then*

$$\begin{aligned} \Delta \vdash_L \varphi & \text{ iff } \alpha \rightarrow \Delta^s \vdash_{L^s} \alpha \rightarrow \varphi^s, \\ \Delta \vdash_L^* \varphi & \text{ iff } \alpha \rightarrow \Delta^s \vdash_{L^s}^* \alpha \rightarrow \varphi^s, \end{aligned}$$

where  $\alpha \rightarrow \Delta^s = \{\alpha \rightarrow \delta : \delta \in \Delta^s\}$ .

To axiomatize  $L^s$ , given an axiomatization of  $L$ , we require the following formulas:

- (a)  $\alpha \rightarrow (\diamond_\gamma p \leftrightarrow \square_\gamma p)$ ,  $\alpha \wedge \diamond_\gamma p \rightarrow \square_\alpha \diamond_\gamma p$ ,
- (b)  $\alpha \rightarrow (\diamond_\beta p \leftrightarrow \square_\beta p)$ ,
- (c)  $\beta \rightarrow (\diamond_\alpha p \leftrightarrow \square_\alpha p)$ ,
- (d)  $\alpha \wedge p \rightarrow \square_\beta \square_\alpha p$ ,  $\beta \wedge p \rightarrow \square_\alpha \square_\beta p$ ,
- (e)  $\alpha \wedge \diamond_\gamma p \rightarrow \square_\beta \square_\alpha \square_\beta \square_\alpha \diamond_\gamma p$ .

Let  $\mathbf{Sim} = \mathbf{K} \oplus \{(a), \dots, (e)\}$ . Obviously,  $\mathfrak{F}^s$  is a frame for  $\mathbf{Sim}$  whenever  $\mathfrak{F}$  is a bimodal frame. Consider now a differentiated frame  $\mathfrak{F} = \langle W, R, P \rangle$  for  $\mathbf{Sim}$  which contains only one point where  $\gamma$  is true. (Actually, every rooted differentiated frame for  $\mathbf{Sim}$  satisfies this condition.) Construct a bimodal frame  $\mathfrak{F}_s = \langle V, R_1, R_2, Q \rangle$ , called the *unsimulation* of  $\mathfrak{F}$ , in the following way. Put  $V = \{x \in W : x \models \alpha\}$ ,  $V^\bullet = \{x \in W : x \models \beta\}$  and  $U = \{x \in W : x \models \gamma\}$ . Since  $\gamma \vee \alpha \vee \beta \in \mathbf{K}$ , we have  $W = V \cup V^\bullet \cup U$ . It is not hard to verify using (b) and (c) (and the differentiatedness of  $\mathfrak{F}$ ) that for every  $x \in V$  there exists a unique  $x^\bullet \in V^\bullet$  such that  $xRx^\bullet$ , and for every  $y \in V^\bullet$  there exists  $y^\circ \in V$  such that  $yRy^\circ$ . By (d),  $x = x^{\bullet\circ}$ . Finally, we put  $R_1 = R \cap V^2$ ,  $R_2 = \{(x, y) \in V^2 : x^\bullet Ry^\bullet\}$  and  $Q = \{X \cap V : X \in P\}$ . It is easily proved that  $\mathfrak{F}_s$  is a bimodal frame. The name *unsimulation* is justified by the following lemma.

LEMMA 120. *For every differentiated bimodal frame  $\mathfrak{F}$ ,  $(\mathfrak{F}^s)_s \cong \mathfrak{F}$ .*

Now we have:

THEOREM 121. *For every bimodal logic  $L = \mathbf{K}_2 \oplus \Delta$ ,*

$$L^s = \mathbf{Sim} \oplus \alpha \rightarrow \Delta^s.$$

**Proof.** Clearly,  $\mathbf{Sim} \oplus \alpha \rightarrow \Delta^s \subseteq L^s$ . Assume that the converse inclusion does not hold. Then there exists a rooted differentiated  $\mathfrak{F}$  such that  $\mathfrak{F} \not\models L^s$  but  $\mathfrak{F} \models \mathbf{Sim} \oplus \alpha \rightarrow \Delta^s$ . By Lemma 120,  $(\mathfrak{F}_s)^s \not\models L^s$ . By the definition of  $L^s$ , we then conclude that  $\mathfrak{F}_s \not\models L$ . And by Proposition 119, we have  $(\mathfrak{F}_s)^s \not\models \alpha \rightarrow \Delta^s$ , from which  $\mathfrak{F} \not\models \alpha \rightarrow \Delta^s$ . ■

Given  $L \in [\mathbf{Sim}, \mathbf{K} \oplus \square\perp]$ , the logic  $L_s = \{\varphi : \alpha \rightarrow \varphi^s \in L\}$  is called the *unsimulation* of  $L$ .

LEMMA 122. *If  $L$  is determined by a class  $\mathcal{C}$  of frames in which  $\gamma$  is true only at one point then  $L_s = \text{Log}\{\mathfrak{F}_s : \mathfrak{F} \in \mathcal{C}\}$ .*

We are in a position now to formulate the main result of this section.

THEOREM 123 (Kracht 1999). *The map  $L \mapsto L^s$  is an isomorphism from the lattice  $\text{NExt}\mathbf{K}_2$  onto the interval  $[\mathbf{Sim}, \mathbf{K}_1 \oplus \Box\perp]$ . The inverse map is  $L \mapsto L_s$ . Both these maps preserve tabularity, (global) FMP, (global) Kripke completeness, decidability, interpolation, strong completeness,  $\mathcal{R}$ - and  $\mathcal{D}$ -persistence, elementarity.*

**Proof.** To prove the first claim it suffices to show that  $(L_s)^s = L$  for every  $L \in [\mathbf{Sim}, \mathbf{K} \oplus \Box\perp]$ . That  $L \subseteq (L_s)^s$  is clear. Consider the set  $\mathcal{C}$  of all differentiated frames  $\mathfrak{F}_s$  such that  $\mathfrak{F} \models L$  and  $\gamma$  is true only at one point in  $\mathfrak{F}$ . By Lemma 122,  $\mathcal{C}$  characterizes  $L_s$ . It is not difficult to show now that the class  $\{\mathfrak{F}_s^+ : \mathfrak{F} \in \mathcal{C}\}$  is closed under subalgebras, homomorphic images and direct products; so it is a variety. Consequently,  $\mathcal{C}$  is (up to isomorphic copies) the class of all differentiated frames for  $L_s$ .

Take a differentiated frame  $\mathfrak{F}$  for  $(L_s)^s$ . Then  $\mathfrak{F}_s \models L_s$ . So there exists  $\mathfrak{G}_s \in \mathcal{C}$  which is isomorphic to  $\mathfrak{F}_s$ . Hence  $(\mathfrak{F}_s)^s \cong (\mathfrak{G}_s)^s$  and  $\mathfrak{F} \models L$ , since  $\mathfrak{G} \models L$ . It follows that  $L^s$  is determined by  $\{\mathfrak{F}^s : \mathfrak{F} \in \mathcal{C}\}$  whenever  $L$  is determined by  $\mathcal{C}$ .

The preservation of tabularity, (global) FMP, (global) Kripke completeness, and strong completeness under both maps is proved with the help of Lemma 122 and the observation above. It is also clear that  $L$  is decidable whenever  $L^s$  is decidable. For the remaining (rather technical) part of the proof the reader is referred to [Kracht 1999] and [Kracht and Wolter 1999]. ■

Besides its theoretical significance, this theorem can be used to transfer rather subtle counter-examples from polymodal logic to unimodal logic. For instance, Kracht [1996] constructs a polymodal logic which has FMP and is globally Kripke incomplete. By Theorem 123, we obtain a unimodal logic with the same properties.

#### 2.4 Minimal tense extensions

Now let us turn to *tense logics* which may be regarded as normal bimodal logics containing the axioms  $p \rightarrow \Box_1\Diamond_2p$  and  $p \rightarrow \Box_2\Diamond_1p$ . Usually studies in Tense Logic concern some special systems representing various models of time, like cyclic time, discrete or dense linear time, branching time, relativistic time, etc. Such systems are discussed in *Basic Tense Logic*, volume 6 of this *Handbook* (see also [Gabbay *et al.* 1994], [Goldblatt 1987] and [van Benthem 1991]). However, as before our concern is general methods which make it possible to obtain results not only for this or that particular system but for wide classes of logics. This direction of studies in Tense Logic is quite

new and actually not so many general results are available. In this and the next section we consider two natural families of tense logics—the minimal tense extensions of unimodal logics and tense logics of linear frames. Our aim is to find out to what extent the theory developed for unimodal logics in  $\text{NExt}\mathbf{K}$  and especially  $\text{NExt}\mathbf{K4}$  can be “lifted” to these families.

The smallest tense logic  $\mathbf{K}.t$  is determined by the class of bimodal Kripke frames  $\langle W, R, R^{-1} \rangle$  in which  $R$  is the accessibility relation for  $\Box_1$  and  $R^{-1}$  for  $\Box_2$ . Frames of this type are known as *tense Kripke frames*; general frames of the form  $\langle W, R, R^{-1}, P \rangle$  will be called just *tense frames*. Notice that not all unimodal general frames  $\langle W, R, P \rangle$  can be converted into tense frames  $\langle W, R, R^{-1}, P \rangle$  because  $P$  is not necessarily closed under the operation

$$\Diamond_2 X = \{x \in W : \exists y \in X \ xR^{-1}y\}.$$

For instance, in the frame  $\mathfrak{F}$  of Example 7 we have  $\Diamond_2\{\omega + 1\} = \{\omega\} \notin P$ .

Each normal unimodal logic  $L = \mathbf{K} \oplus \Gamma$  in the language with  $\Box_1$  gives rise to its *minimal tense extension*  $L.t = \mathbf{K}.t \oplus \Gamma$ . From the semantical point of view  $L.t$  is the logic determined by the class of tense frames  $\langle W, R, R^{-1}, P \rangle$  such that  $\langle W, R, P \rangle \models L$ . The formation of the minimal tense extensions is the simplest way of constructing tense logics from unimodal ones. Of “natural” tense logics, minimal tense extensions are, for instance, the logics of (converse) transitive trees, (converse) well-founded frames, (converse) transitive directed frames, etc. The main aim of this section is to describe conditions under which various properties of  $L$  are inherited by  $L.t$ .

Notice first that unlike fusions,  $L.t$  is not in general a conservative extension of  $L$ , witness  $L = \text{Log}\mathfrak{F}$  where  $\mathfrak{F}$  is again the frame constructed in Example 7: one can easily check that  $\mathbf{K4}.t \subseteq L.t$ . However, if  $L$  is Kripke complete then  $L.t$  is a conservative extension of  $L$  and so  $L'.t = L.t$  implies  $L' \subseteq L$ . This example may appear to be accidental (as the first examples of Kripke incomplete logics in  $\text{NExt}\mathbf{K}$ ). However, we can repeat (with a slight modification) Blok’s construction of Theorem 35 and prove the following

**THEOREM 124.** *If  $L$  is a union-splitting of  $\text{NExt}\mathbf{K}$  or  $L = \mathbf{For}$ , then  $L'.t = L.t$  implies  $L' = L$ . Otherwise there is a continuum of logics in  $\text{NExt}\mathbf{K}$  having the same minimal tense extension as  $L$ .*

It is not known whether there exists  $L \in \text{NExt}\mathbf{K4}$  such that  $L.t$  is not a conservative extension of  $L$ .

Theorem 124 leaves us little hope to obtain general positive results for the whole family of minimal tense extensions. As in the case of unimodal logics we can try our luck by considering logics with transitive frames. So in the rest of this section it is assumed that the unimodal and tense logics we deal with contain  $\mathbf{K4}$  and  $\mathbf{K4}.t$ , respectively, and that frames are transitive. But even in this case we do not have general preservation results: Wolter [1996b] constructed a logic  $L \in \text{NExt}\mathbf{K4}$  having FMP and such that  $L.t$  is not Kripke complete. However, the situation turns out to be not so hopeless

if we restrict attention to the well-behaved classes of logics in  $\text{NExt}\mathbf{K4}$ , namely logics of finite width, finite depth and cofinal subframe logics. First, we have the following results of [Wolter 1997a].

**THEOREM 125.** *If  $L \in \text{NExt}\mathbf{K4}$  is a logic of finite depth then  $L.t$  has FMP. If  $L \in \text{NExt}\mathbf{K4}$  is a logic of finite width then  $L.t$  is Kripke complete.*

It is to be noted that tense logics of finite depth are much more complex than their unimodal counterparts. For example, there exists an undecidable finitely axiomatizable logic containing  $\mathbf{K4}.t \oplus \Box_1 \Box_1 \perp$  (for details see [Kracht and Wolter 1999]).

The minimal tense extensions of cofinal subframe logics were investigated in [Wolter 1995, 1997a].

**THEOREM 126.** *If  $L \in \text{NExt}\mathbf{K4}$  is a cofinal subframe logic then*

- (i)  *$L.t$  is Kripke complete;*
- (ii)  *$L.t$  has FMP iff  $L$  is canonical;*
- (iii)  *$L.t$  is decidable whenever  $L$  is finitely axiomatizable.*

Before outlining the idea of the proof we note some immediate consequences for a few standard tense logics.

**EXAMPLE 127.** (i) The logic of the converse well-founded tense frames is  $\mathbf{GL}.t$ ; it does not have FMP but is decidable. (ii) The logic of the converse transitive trees is  $\mathbf{K4.3}.t$ ; it has FMP and is decidable. (iii) The logic of the converse well-founded directed tense frames is  $\mathbf{GL}.t \oplus \mathbf{K4.2}.t$ ; it does not have FMP and is decidable.

**Proof.** The proof of the negative part, i.e., that  $L.t$  does not have FMP if  $L$  is not canonical, is rather technical; it is based on the characterization of the canonical cofinal subframe logics of [Zakharyashev 1996]. The reader can get some intuition from the following example: neither  $\mathbf{Grz}.t$  nor  $\mathbf{GL}.t$  has FMP. Indeed, the Grzegorzcyk axiom

$$\Box_2(\Box_2(p \rightarrow \Box_2 p) \rightarrow p) \rightarrow p$$

is refuted in  $\langle \omega, \geq, \leq \rangle$  and so does not belong to  $\mathbf{Grz}.t$ ; however, it is valid in all finite partial orders. The argument for  $\mathbf{GL}.t$  is similar: take the Löb axiom in  $\Box_2$  and the frame  $\langle \omega, >, < \rangle$ .

We sketch now the proof of the positive part. For a tense Kripke frame  $\mathfrak{F} = \langle W, R, R^{-1} \rangle$ , let  $\text{rp}$  be a partial function associating with some clusters in  $\mathfrak{F}$  one of the frames

$$\langle \omega, >, < \rangle \text{ or } \langle \omega, \geq, \leq \rangle.$$

We call it a *replacement function* for  $\mathfrak{F}$  and define  $\mathfrak{F}^{rp}$  to be the result of replacing in  $\mathfrak{F}$  all clusters  $C$  in the domain of  $\text{rp}$  by (disjoint copies of)  $\text{rp}C$ . Our first observation is that for each cofinal subframe logic  $L$ ,  $L.t$  is

determined by a set of frames of the form  $\mathfrak{F}^{rp}$  such that  $\mathfrak{F}$  is of finite depth. Indeed, suppose  $\varphi \notin L.t$  and consider a countermodel  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  for  $\varphi$  based on a descriptive finitely generated tense frame  $\mathfrak{F} = \langle W, R, R^{-1}, P \rangle$  for  $L.t$ . Say that a point  $x \in W$  is *non-eliminable* (relative to  $\varphi$ ) if there are a subformula  $\psi$  of  $\varphi$  and  $S \in \{R, R^{-1}\}$  such that  $x \in \max_S \{y \in W : y \models \psi\}$  or  $x \in \max_S \{y \in W : y \models \neg\psi\}$ . Denote by  $W_e$  the set of non-eliminable points in  $W$  and construct a new model  $\mathfrak{M}_e$  on the frame  $\mathfrak{F}_e = \langle W_e, R \upharpoonright W_e, R^{-1} \upharpoonright W_e \rangle$  by taking  $\mathfrak{V}_e(p) = \mathfrak{V}(p) \cap W_e$  for all variables  $p$  in  $\varphi$ . Clearly, the Kripke frame  $\mathfrak{F}_e$  is of finite depth ( $d(\mathfrak{F}_e) \leq 2l(\varphi)$ , to be more precise). Besides, using Theorem 23 one can easily show that  $(\mathfrak{M}_e, y) \models \psi$  iff  $(\mathfrak{M}, y) \models \psi$ , for all  $\psi \in \mathbf{Sub}\varphi$  and  $y \in W_e$ . (Note that Theorem 23 is applicable in this case, since  $\langle W, R, P \rangle$  is descriptive whenever  $\langle W, R, R^{-1}, P \rangle$  is descriptive.) Moreover, the  $R$ -reduct  $\langle W_e, R \upharpoonright W_e \rangle$  of  $\mathfrak{F}_e$  is a cofinal subframe of the  $R$ -reduct  $\langle W, R \rangle$  of the underlying Kripke frame of  $\mathfrak{F}$ . So  $\mathfrak{F}_e$  is a frame for  $L.t$  whenever  $L$  is canonical (=  $\mathcal{D}$ -persistent). However, this is not so if  $L$  is not canonical.

EXAMPLE 128. Consider the frame  $\mathfrak{F} = \langle W, R, R^{-1}, P \rangle$ , where  $\langle W, R \rangle$  is the reflexive point  $\infty$  followed by the chain  $\langle \omega, > \rangle$  and  $P$  consists of all cofinite sets containing  $\infty$  and their complements. Then  $\mathfrak{F} \models \mathbf{GL}.t$  but (for an arbitrary  $\varphi$ )  $\mathfrak{F}_e$  contains  $\infty$  and so  $\mathfrak{F}_e \not\models \mathbf{GL}.t$ .

A rather tedious proof (see [Wolter 1997a]) shows, however, that there exists a replacement function  $\text{rp}$  for  $\mathfrak{F}_e$  such that  $\mathfrak{F}_e^{rp}$  validates  $L.t$  and all points in clusters from  $\text{domrp}$  are eliminable relative to  $R$  in  $\mathfrak{F}$ . (In the example above we put  $\text{rp}\{\infty\} = \langle \omega, >, < \rangle$  and  $\infty$  is eliminable relative to  $R$ .) So let us assume that such  $\text{rp}$  is given and that its domain is empty if  $L$  is canonical. Define a model  $\mathfrak{M}_e^{rp} = (\mathfrak{F}_e^{rp}, \mathfrak{V}^{rp})$  as follows. First we put  $y \in \mathfrak{V}^{rp}(p)$  whenever  $y \in \mathfrak{V}_e(p)$  and  $y \notin \text{domrp}$ . Consider now a cluster  $C = \{a_0, \dots, a_{m-1}\}$  in  $\text{domrp}$ .  $\mathfrak{V}^{rp}$  is defined in  $\text{rp}C$  by unravelling  $C$  into the chain  $\text{rp}C$ ; more precisely, we put

$$\mathfrak{V}^{rp}(p) \cap \text{rp}C = \{mj + i : j < \omega, a_i \in \mathfrak{V}(p)\}.$$

Using the fact that  $\text{domrp}$  contains only  $R$ -eliminable points, one can show by induction that, for every  $\psi \in \mathbf{Sub}\varphi$ ,  $(\mathfrak{M}_e, y) \models \psi$  iff  $(\mathfrak{M}_e^{rp}, y) \models \psi$ , if  $C(y)$  does not belong to  $\text{domrp}$ , and

$$\{n \in \text{rp}C : (\mathfrak{M}_e^{rp}, n) \models \psi\} = \{mj + i : j < \omega, (\mathfrak{M}_e, a_i) \models \psi\},$$

if a cluster  $C = \{a_0, \dots, a_{m-1}\}$  is in  $\text{domrp}$ . Thus  $\mathfrak{F}_e^{rp}$  refutes  $\varphi$ , which proves that  $L.t$  is Kripke complete.

To show that all canonical logics  $L.t$  do have FMP we reduce  $\mathfrak{F}_e^{rp}$  once again. Define an equivalence relation  $\sim$  on  $W_e$  by induction on the  $R$ -depth  $d_R(x)$  of a point  $x$  in  $\mathfrak{F}_e$ . Suppose that  $d_R(x) = d_R(y)$  and  $\sim$  is already defined for all points of  $R$ -depth  $< d_R(x)$  and put  $x \sim y$  if the following

conditions are satisfied: (a)  $x \models \psi$  iff  $y \models \psi$ , for all  $\psi \in \mathbf{Sub}\varphi$  ( $x \sim_\varphi y$ , for short), (b) if  $z$  is an  $R$ -successor of  $y$  and  $C(z) \neq C(y)$  then there exists an  $R$ -successor  $z'$  of  $x$  with  $C(z') \neq C(x)$  such that  $z \sim z'$  and vice versa, (c) the cluster  $C(x)$  is degenerate iff  $C(y)$  is degenerate, (d)  $\text{rp}C(x) = \text{rp}C(y)$ , (e) for each  $z \in C(x)$  there exists  $z' \in C(y)$  such that  $z \sim_\varphi z'$  and vice versa.

Let  $[x]$  denote the equivalence class generated by  $x$ . Define a frame  $\mathfrak{G} = \langle V, S, S^{-1} \rangle$  by taking  $V = \{[x] : x \in W_e\}$ , and  $[x]S[y]$  iff there are  $x' \in [x]$  and  $y' \in [y]$  such that  $x'Ry'$ . Since  $\mathfrak{F}_e$  is of finite depth,  $V$  is finite. Moreover, the map  $x \mapsto [x]$  is a reduction of the unimodal frame  $\langle W_e, R \upharpoonright W_e \rangle$  to  $\langle V, S \rangle$ . It follows that  $\mathfrak{G}$  is a frame for  $L.t$  whenever  $L$  is canonical. Define a valuation in  $\mathfrak{G}$  by putting  $[x] \models p$  iff  $x \models p$ , for all  $x \in W_e$  and all variables  $p$  in  $\varphi$ . Then one can show that  $[x] \models \psi$  iff  $x \models \psi$ , for all  $\psi \in \mathbf{Sub}\varphi$ . So  $\mathfrak{G} \not\models \varphi$ , as required, which means that  $L.t$  has FMP.

To prove the decidability of a finitely axiomatizable  $L.t$  we first show its completeness with respect to a rather simple class of frames.

Define a replacement function  $\text{rf}$  for  $\mathfrak{G}$  as follows. For each cluster  $C$  in  $\mathfrak{F}_e$  the set  $[C] = \{[x] : x \in C\}$  is a cluster in  $\mathfrak{G}$ , and moreover, every cluster in  $\mathfrak{G}$  can be presented in this way. So we put  $\text{rf}[C] = \text{rp}C$ , for all clusters  $[C]$  in  $\mathfrak{G}$ . Notice that by (d),  $\text{rf}$  is well-defined. It is easily shown now that the  $R$ -reduct of  $\mathfrak{F}_e^{\text{rp}}$  is reducible to the  $R$ -reduct of  $\mathfrak{G}^{\text{rf}}$  and that  $\mathfrak{G}^{\text{rf}}$  refutes  $\varphi$ . Thus we obtain

LEMMA 129. *For each cofinal subframe logic  $L$ ,*

$$L.t = \text{Log}\{\mathfrak{G}^{\text{rp}} : \mathfrak{G}^{\text{rp}} \models L.t, \mathfrak{G} \text{ finite, rp a replacement function for } \mathfrak{G}\}.$$

So, to establish the decidability of a finitely axiomatizable  $L.t$  it is enough now to present an algorithm which is capable of deciding, given an  $\text{rp}$  for a finite  $\mathfrak{G}$  and  $\varphi$ , whether  $\mathfrak{G}^{\text{rp}} \models \varphi$ . To this end we require the notion of a *cluster assignment*  $\mathbf{t} = \langle \mathbf{t}_1, \mathbf{t}_2 \rangle$  in a tense frame  $\mathfrak{G}$ , which is any function from the set of clusters in  $\mathfrak{G}$  into the set  $\{m, j\} \times \{m, j\}$  such that  $\mathbf{t}C = (m, m)$  if  $C$  is degenerate (here  $m$  and  $j$  are just two symbols;  $m$  stands for “maximal” and  $j$  for “joker”). A valuation  $\mathfrak{V}$  in  $\mathfrak{G}$  is called  $\varphi$ -good for  $(\mathfrak{G}, \mathbf{t})$  if the following conditions hold:

- if  $\mathbf{t}_1C = j$  then  $C \cap \max_R(\mathfrak{V}(\psi)) = \emptyset$ , for all  $\psi \in \mathbf{Sub}\varphi$ ;
- if  $\mathbf{t}_2C = j$  then  $C \cap \max_{R^{-1}}(\mathfrak{V}(\psi)) = \emptyset$ , for all  $\psi \in \mathbf{Sub}\varphi$ .

EXAMPLE 130. Let  $\mathfrak{F}$  be the frame constructed in Example 128 and suppose that  $\mathbf{t}\{\infty\} = (j, m)$ . Then each valuation  $\mathfrak{V}$  in  $\mathfrak{F}$  is  $\varphi$ -good for  $(\mathfrak{G}, \mathbf{t})$  no matter what  $\varphi$  is, because  $\infty$  is eliminable relative to  $R$ . The point  $\infty$  is not  $R^{-1}$ -eliminable, since  $\infty \in \max_{R^{-1}}(\top)$ .

Given a formula  $\varphi$ , a finite frame  $\mathfrak{F}$  and a replacement function  $\text{rp}$  for  $\mathfrak{F}$ , we construct a finite frame  $\mathfrak{G} = \langle V, S, S^{-1} \rangle$  with a cluster assignment



$\mathfrak{t}$  as follows. Let  $k$  be the number of variables in  $\varphi$ . Then  $\mathfrak{G}$  is obtained from  $\mathfrak{F}^{rp}$  by replacing every  $\text{rp}C = \langle \omega, >, < \rangle$  with a non-degenerate cluster  $C'$  of cardinality  $2^k$ ,  $S$ -followed by a chain of  $2l(\varphi)$  irreflexive points, and by replacing every  $\text{rp}C = \langle \omega, \geq, \leq \rangle$  with a non-degenerate cluster  $C'$  of cardinality  $2^k$ ,  $S$ -followed by a chain of  $2l(\varphi)$  reflexive points. The cluster assignment  $\mathfrak{t}$  in  $\mathfrak{G}$  is defined by putting  $\mathfrak{t}C' = (j, m)$ , for all new clusters  $C'$  of cardinality  $2^k$ , and  $\mathfrak{t}C' = (m, m)$ , for all the other clusters. It is not difficult now to prove that  $\mathfrak{F}^{rp} \models \varphi$  iff  $(\mathfrak{G}, \mathfrak{U}) \models \varphi$ , for all  $\varphi$ -good for  $(\mathfrak{G}, \mathfrak{t})$  valuations  $\mathfrak{U}$  in  $\mathfrak{G}$ . This equivalence provides an effective procedure for deciding whether  $\mathfrak{F}^{rp} \models \varphi$ . ■

Note that a similar technique can be used to prove completeness and decidability of various tense logics that are not minimal tense extensions. For instance, all logics of the form  $L.t \oplus \diamond_2 \square_2 p \rightarrow \square_2 \diamond_2 p$ , where  $L$  is a cofinal subframe logic, are complete and decidable if finitely axiomatizable.

### 2.5 Tense logics of linear frames

One of the most important types of tense logics are logics characterized by linear tense frames, i.e., transitive frames  $\langle W, R, R^{-1}, P \rangle$  such that, for all  $x, y \in W$ ,  $xRy$  or  $xR^{-1}y$  or  $x = y$ . For example, Bull [1968] and Segerberg [1970] axiomatized the logics of the frames,  $\langle \mathbb{Z}, <, > \rangle$ ,  $\langle \mathbb{Q}, <, > \rangle$  and  $\langle \mathbb{R}, <, > \rangle$  ( $\mathbb{Z}$ ,  $\mathbb{Q}$  and  $\mathbb{R}$  are the sets of integer, rational and real numbers, respectively).

Linear tense logics form the lattice  $\text{NExtLin}$ , where

$$\text{Lin} = \mathbf{K4}.t \oplus \diamond_1 \diamond_2 p \vee \diamond_2 \diamond_1 p \rightarrow p \vee \diamond_1 p \vee \diamond_2 p$$

is the tense logic determined by the class of all linearly ordered Kripke frames  $\langle W, R, R^{-1} \rangle$ . As we saw in Section 1.11, even unimodal logics of linear orders are rather non-trivial (for instance, they do not always enjoy FMP). Yet they can be characterized by Kripke frames with a transparent structure, which yields a decision algorithm for those of them that are finitely axiomatizable. Tense logics of linear frames turn out to be even more complicated. In fact, one can find almost all kinds of “monsters” among them: uncountably many logics without Kripke frames, strongly complete logics that are not canonical, canonical logics that are not  $\mathcal{R}$ -persistent, incomplete subframe logics, etc. Nevertheless, in this section we show that these logics are quite manageable. Our exposition follows [Wolter 1996b, c], where the reader can find the omitted details. All frames in this section are assumed to be linear.

Given a finite sequence  $\overline{\mathfrak{F}} = \langle \mathfrak{F}_i = \langle W_i, R_i, P_i \rangle : 1 \leq i \leq n \rangle$  of disjoint frames, we denote by  $[\overline{\mathfrak{F}}] = \mathfrak{F}_1 \triangleleft \dots \triangleleft \mathfrak{F}_n$  the ordered sum of them, i.e., the

frame  $\langle W, R, R^{-1}, P \rangle$  in which

$$W = \bigcup_{i=1}^n W_i, \quad R = \bigcup_{i=1}^n R_i \cup \bigcup_{1 \leq i < j \leq n} (W_i \times W_j)$$

and  $P = \{X_1 \cup \dots \cup X_n : X_i \in P_i\}$ . Each finite frame can be represented then as the ordered sum  $C_1 \triangleleft \dots \triangleleft C_n$  of its clusters.

We begin our study by developing a language of “canonical formulas” for axiomatizing logics in **NExtLin** and characterizing the constitution of their frames. It will play the same role as the language of canonical formulas for **K4**. With every finite frame  $\mathfrak{F} = \langle W, R, R^{-1} \rangle = C_1 \triangleleft \dots \triangleleft C_n$  and a cluster assignment  $\mathbf{t} = (\mathbf{t}_1, \mathbf{t}_2)$  in it we associate the formula

$$\alpha(\mathfrak{F}, \mathbf{t}) = \delta(\mathfrak{F}, \mathbf{t}) \wedge \Box_1 \delta(\mathfrak{F}, \mathbf{t}) \wedge \Box_2 \delta(\mathfrak{F}, \mathbf{t}) \rightarrow \neg p_r,$$

where  $r$  is an arbitrary fixed point in  $\mathfrak{F}$  and

$$\begin{aligned} \delta(\mathfrak{F}, \mathbf{t}) = & \bigwedge \{p_x \rightarrow \Diamond_1 p_y : xRy, \neg(yRx)\} \wedge \\ & \bigwedge \{p_x \rightarrow \Diamond_2 p_y : xR^{-1}y, \neg(xRy)\} \wedge \\ & \bigwedge \{p_x \rightarrow \neg p_y : x \neq y\} \wedge \bigwedge \{p_x \rightarrow \neg \Diamond_2 p_y : \neg(xRy)\} \wedge \\ & \bigwedge \{p_x \rightarrow \Diamond_1 p_y : \exists i \leq n (\mathbf{t}_1 C_i = \mathbf{m} \wedge x, y \in C_i \wedge xRy)\} \wedge \\ & \bigwedge \{p_x \rightarrow \Diamond_2 p_y : \exists i \leq n (\mathbf{t}_2 C_i = \mathbf{m} \wedge x, y \in C_i \wedge xR^{-1}y)\} \wedge \\ & \bigvee \{p_y : y \in W\}. \end{aligned}$$

To explain the semantical meaning of these formulas, notice first that if  $\mathbf{t}C = (\mathbf{m}, \mathbf{m})$  for all clusters  $C$  then  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathbf{t})$  iff  $\mathfrak{G}$  is reducible to  $\mathfrak{F}$ ; so  $\mathbf{Lin} \oplus \alpha(\mathfrak{F}, \mathbf{t})$  is a splitting of **NExtLin**. Suppose now that  $\mathbf{t}_i C = \mathbf{j}$  for some  $i \in \{1, 2\}$  and some cluster  $C$  in  $\mathfrak{F}$ . In this case  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathbf{t})$  iff there exist frames  $\mathfrak{G}_i$ , for  $1 \leq i \leq n$ , such that  $\mathfrak{G} = \mathfrak{G}_1 \triangleleft \dots \triangleleft \mathfrak{G}_n$  and  $\mathfrak{G}_i \not\models \alpha(C_i, \mathbf{t} \upharpoonright C_i)$  for all  $1 \leq i \leq n$ . So it suffices to examine the situation when  $\mathfrak{G} \not\models \alpha(C, \mathbf{t})$  for a cluster  $C$ . Assume for simplicity that  $\mathfrak{G}$  is a Kripke frame. *Case 1:*  $\mathbf{t}C = (\mathbf{j}, \mathbf{j})$ . Then  $\mathfrak{G} \not\models \alpha(C, \mathbf{t})$  iff  $|\mathfrak{G}| \geq |C|$ . *Case 2:*  $\mathbf{t}C = (\mathbf{m}, \mathbf{j})$ . Then  $C$  is non-degenerate and  $\mathfrak{G} \not\models \alpha(C, \mathbf{t})$  iff either  $\mathfrak{G}$  contains an  $R$ -final cluster of cardinality  $\geq |C|$  or it has no  $R$ -final point at all. *Case 3:*  $\mathbf{t}C = (\mathbf{j}, \mathbf{m})$ . This is the mirror image of Case 2. *Case 4:*  $\mathbf{t}C = (\mathbf{m}, \mathbf{m})$ . If  $C$  is an irreflexive point then  $\mathfrak{G}$  is an irreflexive point as well whenever  $\mathfrak{G} \not\models \alpha(C, \mathbf{t})$ . If  $C$  is non-degenerate and  $\mathfrak{G} \not\models \alpha(C, \mathbf{t})$  then  $\mathfrak{G}$  satisfies the conditions of Cases 2 and 3.

**EXAMPLE 131.** Let  $\alpha = \alpha(\overset{a}{\circ} \rightarrow \overset{b}{\circ}, \mathbf{t})$  where  $\mathbf{t}a = (\mathbf{m}, \mathbf{j})$  and  $\mathbf{t}b = (\mathbf{j}, \mathbf{m})$ . Then  $\mathfrak{F} \not\models \alpha$  iff there exists a non-empty upward closed set  $X \in P$  such that  $\forall x \in X \exists y \in X yRx$ ,  $W - X \neq \emptyset$  and  $\forall x \in W - X \exists y \in W - X xRy$ .

Hence  $\langle \mathbb{Q}, <, > \rangle \not\models \alpha$  (take  $X = \{y \in \mathbb{Q} : \sqrt{2} < y\}$ ) but  $\langle \mathbb{R}, <, > \rangle \models \alpha$ , since the real line contains no gaps.

**THEOREM 132.** *There is an algorithm which, given a formula  $\varphi$ , returns formulas  $\alpha(\mathfrak{F}_1, \mathbf{t}_1), \dots, \alpha(\mathfrak{F}_n, \mathbf{t}_n)$  such that*

$$\mathbf{Lin} \oplus \varphi = \mathbf{Lin} \oplus \alpha(\mathfrak{F}_1, \mathbf{t}_1) \oplus \dots \oplus \alpha(\mathfrak{F}_n, \mathbf{t}_n).$$

**Proof.** Let  $(\mathfrak{F}_i, \mathbf{t}_i)$ ,  $1 \leq i \leq n$ , be the collection of all finite frames with type assignments such that, for each  $i$ , (a) there is a countermodel  $\mathfrak{M}_i = \langle \mathfrak{F}_i, \mathfrak{V}_i \rangle$  for  $\varphi$  in which  $\mathfrak{V}_i$  is  $\varphi$ -good for  $(\mathfrak{F}_i, \mathbf{t}_i)$ , (b) the depth of  $\mathfrak{F}_i$  does not exceed  $4l(\varphi) + 1$ , and (c) no cluster in  $\mathfrak{F}_i$  contains more than  $2^{v(\varphi)}$  points, where  $v(\varphi)$  is the number of variables in  $\varphi$ .

Let  $\mathfrak{F}$  refute  $\alpha(\mathfrak{G}_i, \mathbf{t}_i)$  under a valuation  $\mathfrak{U}$ . By the definition of  $(\mathfrak{F}_i, \mathbf{t}_i)$ , the model  $\mathfrak{M}_i$  refutes  $\varphi$ . Define a valuation  $\mathfrak{U}'$  in  $\mathfrak{F}$  by taking, for all variables  $p$  in  $\varphi$ ,

$$\mathfrak{U}'(p) = \bigcup \{ \mathfrak{U}(p_x) : x \in \mathfrak{V}_i(p) \}.$$

It is not hard to show by induction that  $\mathfrak{U}'(\psi) = \bigcup \{ \mathfrak{U}(p_x) : x \in \mathfrak{V}_i(\psi) \}$  for all  $\psi \in \mathbf{Sub}\varphi$ , and so  $\mathfrak{F}$  refutes  $\varphi$  under  $\mathfrak{U}'$ . Thus  $\mathfrak{F} \models \varphi$  implies  $\mathfrak{F} \models \alpha(\mathfrak{F}_i, \mathbf{t}_i)$  for every  $i$ . The converse direction is rather technical; we refer the reader to [Wolter 1996c].  $\blacksquare$

“Canonical” axiomatizations of some standard linear tense logics are shown in Table 3, where we use the following abbreviations. Given a finite frame  $\mathfrak{F} = C_1 \triangleleft \dots \triangleleft C_n$ , we write  $\alpha((C_1, \mathbf{t}C_1) \triangleleft \dots \triangleleft (C_n, \mathbf{t}C_n))$  instead of  $\alpha(\mathfrak{F}, \mathbf{t})$  and  $\alpha(-, (C_1, \mathbf{t}C_1) \triangleleft \dots \triangleleft (C_n, \mathbf{t}C_n))$  instead of

$$\alpha((C_1, \mathbf{t}C_1) \triangleleft \dots \triangleleft (C_n, \mathbf{t}C_n)) \oplus \alpha((\circ, (j, j)) \triangleleft (C_1, \mathbf{t}C_1) \triangleleft \dots \triangleleft (C_n, \mathbf{t}C_n)).$$

$\alpha((C_1, \mathbf{t}C_1) \triangleleft \dots \triangleleft (C_n, \mathbf{t}C_n), -)$  is defined analogously.

Now we exploit the formulas  $\alpha(\mathfrak{F}, \mathbf{t})$  to characterize the  $\bigcap$ -irreducible logics in  $\mathbf{NExtLin}$ . Recall that every logic  $L \in \mathbf{NExt}L_0$  is represented as

$$L = \bigcap \{ L' \supseteq L : L' \text{ is } \bigcap\text{-irreducible} \}.$$

So such a characterization can open the door to a better understanding of the structure of the lattice  $\mathbf{NExtLin}$ . The  $\bigcap$ -irreducible logics will be described semantically as the logics determined by certain descriptive frames.

**DEFINITION 133.**

- (1) Denote by  $\textcircled{k}$  the non-degenerate cluster with  $k > 0$  points.
- (2) Let  $\omega^{<}(0)$  be the strictly ascending chain  $\langle \omega, <, > \rangle$  of natural numbers,  $\omega^{<}(1)$  the chain  $\langle \omega, \leq, \geq \rangle$ ,  $\omega^{<}(2)$  the ascending chain of natural numbers in which precisely the even points are reflexive,  $\omega^{<}(3)$  the chain in which precisely the multiples of 3 are reflexive, and so on;  $\omega^{>}(n)$  is the mirror image of  $\omega^{<}(n)$ .

Table 3. Axiomatizations of standard tense logics

---

<b>Ord<sub>t</sub></b>	$= \text{Log}\{\langle \xi, <, > \rangle : \xi \text{ an ordinal}\} =$ $\mathbf{Lin} \oplus \alpha(-, (\circ, (j, m)))$
<b>E<sub>t</sub></b>	$= \mathbf{Lin} \oplus \diamond_1 \top \oplus \diamond_2 \top =$ $\mathbf{Lin} \oplus \alpha(-, (\bullet, (m, m))) \oplus \alpha((\bullet, (m, m)), -)$
<b>O<sub>n</sub></b>	$= \text{Log}\langle \omega n, <, > \rangle =$ $\mathbf{Ord}_t \oplus \alpha(\underbrace{(\circ, (m, j)) \triangleleft \cdots \triangleleft (\circ, (m, j))}_{n+1}) \oplus \alpha(-, (\bullet, (m, m)))$
<b>RD</b>	$= \text{Log}\{\mathfrak{G} : \forall x(\neg xRx \rightarrow \exists y(xRy \wedge \{z : xRzRy\} = \emptyset))\} =$ $\mathbf{Lin} \oplus \alpha(-, (\bullet, (m, m))) \oplus \alpha(-, (\bullet, (m, m)) \triangleleft (\circ, (m, j)))$
<b>LD</b>	$= \text{the mirror image of RD}$
<b>Z<sub>t</sub></b>	$= \text{Log}\langle \mathbb{Z}, <, > \rangle =$ $\mathbf{RD} \oplus \mathbf{LD} \oplus \alpha((\circ, (j, j)) \triangleleft (\circ, (j, m))) \oplus$ $\alpha((\circ, (m, j)) \triangleleft (\circ, (j, j)))$
<b>Ds<sub>n</sub></b>	$= \mathbf{Lin} \oplus \Box_1^{n+1} p \rightarrow \Box_1^n p =$ $\mathbf{Lin} \oplus \alpha(-, \underbrace{(\bullet, (m, m)) \triangleleft \cdots \triangleleft (\bullet, (m, m))}_{n+1}), -)$
<b>Q<sub>t</sub></b>	$= \text{Log}\langle \mathbb{Q}, <, > \rangle =$ $\mathbf{Ds}_1 \oplus \mathbf{E}_t$
<b>R<sub>t</sub></b>	$= \text{Log}\langle \mathbb{R}, <, > \rangle =$ $\mathbf{Q}_t \oplus \alpha((\circ, (m, j)) \triangleleft (\circ, (j, m)))$
<b>Rd<sub>t</sub></b>	$= \text{Log}\{\langle \xi, \leq, \geq \rangle : \xi \text{ an ordinal}\} =$ $\mathbf{Lin} \oplus \alpha(-, (\odot, (j, m)))$

---

- (3)  $\mathfrak{C}(0, \mathbb{1})$  is the mirror image of the frame introduced in Example 128, i.e.,  $\mathfrak{C}(0, \mathbb{1}) = \langle \omega^{<}(0) \triangleleft \mathbb{1}, P \rangle$ , where  $P$  consists of all cofinite sets containing  $\mathbb{1}$  and their complements. We generalize this construction to chains  $\omega^{<}(n)$  and clusters  $\mathbb{k}$ . Namely, for  $n < \omega$ ,  $k > 1$  and  $\mathbb{k} = \{a_0, \dots, a_{k-1}\}$ , we put

$$\mathfrak{C}(n, \mathbb{k}) = \langle \omega^{<}(n) \triangleleft \mathbb{k}, P \rangle,$$

where  $P$  is the set of possible values generated by  $\{X_i : 0 \leq i \leq k-1\}$ , for  $X_i = \{a_i\} \cup \{kj + i : j \in \omega\}$ ,  $0 \leq i \leq k-1$ .  $\mathfrak{C}(\mathbb{k}, n)$  denotes the mirror image of  $\mathfrak{C}(n, \mathbb{k})$ .

- (4)  $\mathfrak{C}(0, \mathbb{1}, 0) = \langle \omega^{<}(0) \triangleleft \mathbb{1} \triangleleft \omega^{>}(0), P \rangle$ , where  $P$  consists of all cofinite sets containing  $\mathbb{1}$  and their complements.

It is easy to check that the frames defined in (3) and (4) are descriptive and a singleton  $\{x\}$  is in  $P$  iff  $x \notin \mathbb{k}$ .

For a class of frames  $\mathcal{C}$ , we denote by  $\mathcal{C}^*$  the class of finite sequences of frames from  $\mathcal{C}$  and let  $[\mathcal{C}^*] = \{\overline{\mathfrak{F}} : \mathfrak{F} \in \mathcal{C}^*\}$ . The class of finite clusters and the frames of the form (3) in Definition 133 is denoted by  $\mathcal{B}_0$ ; put also  $\mathcal{B} = \{\mathfrak{C}(0, \mathbb{1}, 0)\} \cup \mathcal{B}_0$ .

**THEOREM 134.** *Each logic  $L \in \text{NExtLin}$  is determined by a set  $\mathcal{C} \subseteq [\mathcal{B}^*]$ . If  $L$  is finitely axiomatizable then  $L = \text{Log}\mathcal{C}$  for some set  $\mathcal{C} \subseteq [\mathcal{B}_0^*]$ .*

**Proof.** We explain the idea of the proof of the first claim. Suppose that  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{M} \rangle$  is a countermodel for  $\alpha = \alpha((C_1, \mathfrak{t}C_1) \triangleleft \dots \triangleleft (C_n, \mathfrak{t}C_n))$  based on a descriptive frame  $\mathfrak{F} = \langle W, R, R^{-1}, P \rangle$ . We must show that there exists  $\mathfrak{G} \in [\mathcal{B}^*]$  refuting  $\alpha$  and such that  $\text{Log}\mathfrak{G} \supseteq \text{Log}\mathfrak{F}$ . Consider the sets

$$W_i = \{y \in W : (\mathfrak{M}, y) \models \bigvee \{p_x : x \in C_i\}\}.$$

One can easily show that  $W_i$  are intervals in  $\mathfrak{F}$  and  $\mathfrak{F} = \mathfrak{F}_1 \triangleleft \dots \triangleleft \mathfrak{F}_n$ , for the subframes  $\mathfrak{F}_i$  of  $\mathfrak{F}$  induced by  $W_i$ . Moreover,  $\mathfrak{G} = \overline{\mathfrak{G}}$  is as required if  $\overline{\mathfrak{G}} = \langle \mathfrak{G}_1, \dots, \mathfrak{G}_n \rangle$  is a sequence in  $\mathcal{B}^*$  such that  $\text{Log}\mathfrak{G}_i \supseteq \text{Log}\mathfrak{F}_i$ , and  $\mathfrak{G}_i \not\models \alpha(C_i, \mathfrak{t}C_i)$ , for  $1 \leq i \leq n$ . Frames  $\mathfrak{G}_i$  with those properties are constructed in [Wolter96d]. ■

**EXAMPLE 135.** The logic  $\mathbf{Q}_t$  is determined by the frames  $\mathfrak{F} \in [\mathcal{B}^*]$  which contain no pair of adjacent irreflexive points, and  $\mathbf{R}_t$  is determined by the frames  $\mathfrak{F} \in [\mathcal{B}^*]$  which contain neither a pair of adjacent irreflexive points nor a pair of adjacent non-degenerate clusters.

It is not difficult to show now that the logics  $\text{Log}\mathfrak{F}$ , for  $\mathfrak{F} \in [\mathcal{B}^*]$ , coincide with the  $\bigcap$ -irreducible logics in  $\text{NExtLin}$ . Our first aim is achieved, and

in the remaining part of this section we shall draw consequences of this result. Using the same sort of arguments as in the proof of Theorem 126 and Kruskal's [1960] Tree Theorem one can prove

**COROLLARY 136.** (i) *All finitely axiomatizable logics in  $\text{NExtLin}$  are decidable.*

(ii) *A logic  $L$  is finitely axiomatizable whenever there exists  $n < \omega$  such that  $L \in \text{NExtDs}_n$ .*

It follows in particular that all logics in  $\text{NExtQ}_t$  and all logics of reflexive frames are finitely axiomatizable and decidable.

Now we formulate two corollaries concerning the Kripke completeness of linear tense logics. First, it is not hard to see that every logic in  $\text{NExtLin}$  characterized by an infinite frame in  $[\mathcal{B}^*]$  is Kripke incomplete. Using this observation one can prove

**COROLLARY 137.** *Suppose  $L \in \text{NExtLin}$  and there is a Kripke frame of infinite depth for  $L$ . Then there exists a Kripke incomplete logic in  $\text{NExtL}$ .*

This result means in particular that in Tense Logic we do not have analogues of the unimodal completeness results of Bull [1966b] and Fine [1974c]. However, if a logic is complete then it is determined by a simple class of frames. Let  $\mathcal{K}$  be the class frames containing finite clusters and frames of the form (2) in Definition 133.

**THEOREM 138.** *Each Kripke complete logic in  $\text{NExtLin}$  is determined by a subset of  $[\mathcal{K}^*]$ .*

One of the main types of logics considered in conventional Tense Logic are logics determined by strict linear orders, known also as *time-lines*. We call them *t-line logics*. All logics in Table 3, save  $\mathbf{Rd}_t$ , are t-line logics. T-line logics were defined semantically, and now we are going to determine a necessary syntactic condition for a linear tense logic to be a t-line logic.

Given a frame  $\mathfrak{F}$ , we denote by  $\mathfrak{F}^\circ$  the frame that results from  $\mathfrak{F}$  by replacing its proper clusters with reflexive points. Call  $L \in \text{NExtLin}$  a *t-axiom logic* if  $L$  is axiomatizable by a set of formulas of the form  $\alpha(\mathfrak{F}, \mathbf{t})$  in which  $\mathfrak{F}$  contains no proper clusters.

**PROPOSITION 139.** *The following conditions are equivalent for all logics  $L \in \text{NExtLin}$ :*

- (i)  *$L$  is a t-axiom logic;*
- (ii)  *$\mathfrak{F}^\circ \models L$  implies  $\mathfrak{F} \models L$ , for every  $\mathfrak{F} \in [\mathcal{B}^*]$ .*
- (iii)  *$\alpha(\mathfrak{G}, \mathbf{t}) \in L$  implies  $\alpha(\mathfrak{G}^\circ, \mathbf{t}) \in L$ ,<sup>15</sup> for every finite  $\mathfrak{G}$ .*

---

<sup>15</sup>We assume that  $\mathbf{t}C = \mathbf{t}o$  whenever  $o$  replaces  $C$  in  $\mathfrak{G}$ .

**Proof.** The implications (i)  $\Rightarrow$  (ii) and (iii)  $\Rightarrow$  (i) are clear. To prove that (ii)  $\Rightarrow$  (iii), suppose  $\alpha(\mathfrak{G}^\circ, \mathfrak{t}) \notin L$ . Then there exists a frame  $\mathfrak{F} \in [\mathcal{B}^*]$  for  $L$  refuting  $\alpha(\mathfrak{G}^\circ, \mathfrak{t})$ . Without loss of generality we may assume that  $\mathfrak{F}$  contains no proper clusters. By enlarging some clusters in  $\mathfrak{F}$  we can construct a frame  $\mathfrak{H} \in [\mathcal{B}^*]$  such that  $\mathfrak{H}^\circ = \mathfrak{F}$  and  $\mathfrak{H} \models \alpha(\mathfrak{G}, \mathfrak{t})$ . In view of (ii),  $\mathfrak{H} \models L$  and so  $\alpha(\mathfrak{G}, \mathfrak{t}) \notin L$ .  $\blacksquare$

It follows that the t-axiom logics form a complete sublattice of the lattice  $\text{NExtLin}$ .

THEOREM 140.

- (i) *All finitely axiomatizable t-axiom logics are Kripke complete.*
- (ii) *All t-line logics are t-axiom logics.*

**Proof.** (i) Suppose that  $L = \mathbf{Lin} \oplus \{\alpha(\mathfrak{G}_i^\circ, \mathfrak{t}_i) : i \in I\}$ , for some finite set  $I$ . By Theorem 134,  $L$  is determined by a subset of  $[\mathcal{B}_0^*]$ . For  $\mathfrak{F} \in [\mathcal{B}_0^*]$ , let  $k\mathfrak{F}$  be the Kripke frame that results from  $\mathfrak{F}$  by replacing all  $\mathfrak{C}(n, \mathbb{k})$  and  $\mathfrak{C}(\mathbb{k}, n)$  with  $\omega^{<}(n)$  and  $\omega^{>}(n)$ , respectively. Then we clearly have  $\text{Log}k\mathfrak{F} \subseteq \text{Log}\mathfrak{F}$ , and  $\mathfrak{F} \models \alpha(\mathfrak{G}^\circ, \mathfrak{t})$  iff  $k\mathfrak{F} \models \alpha(\mathfrak{G}^\circ, \mathfrak{t})$ . It follows that  $L$  is Kripke complete. (ii) Suppose that  $L$  is a t-line logic. By Proposition 139 (3), it suffices to observe that  $\mathfrak{F} \models \alpha(\mathfrak{G}^\circ, \mathfrak{t})$  iff  $\mathfrak{F} \models \alpha(\mathfrak{G}, \mathfrak{t})$ , for all time-lines  $\mathfrak{F}$  and all finite  $\mathfrak{G}$ .  $\blacksquare$

So the fact that in Table 3 all t-line logics are axiomatized by canonical formulas of the form  $\alpha(\mathfrak{G}^\circ, \mathfrak{t})$  is no accident. Finding and verifying axiomatizations of t-line logics becomes almost trivial now.

EXAMPLE 141. Let us check the axiomatization of  $\mathbf{Z}_t$  in Table 3. Put

$$L = \mathbf{RD} \oplus \mathbf{LD} \oplus \alpha((\circ, (j, j)) \triangleleft (\circ, (j, m))) \oplus \alpha((\circ, (m, j)) \triangleleft (\circ, (j, j))).$$

By Theorem 140,  $L$  is complete. By Theorem 138,  $L$  is then determined by a subset of  $[\mathcal{K}^*]$ . Clearly this set contains  $\langle \mathbb{Z}, <, > \rangle$ , possibly  $\mathbb{k}$  for  $k > 0$ , and nothing else. But the logic of  $\mathbb{k}$  contains  $\mathbf{Z}_t$ , for all  $k > 0$ .

We conclude this section by discussing the decidability of properties of logics in  $\text{NExtLin}$ . In Section 4.4 it will be shown that almost all interesting properties of calculi are undecidable in  $\text{NExtK}$  and even in  $\text{NExtS4}$ . In  $\text{NExtLin}$  the situation is different, as was proved in [Wolter 1996c, 1997c].

THEOREM 142. (i) *There are algorithms which, given a formula  $\varphi$ , decide whether  $\mathbf{Lin} \oplus \varphi$  has FMP, interpolation, whether it is Kripke complete, strongly complete, canonical,  $\mathcal{R}$ -persistent.*

(ii) *A linear tense logic is canonical iff it is  $\mathcal{D}$ -persistent iff it is complete and its frames are first order definable.*

(iii) *If a logic in NExtLin has a frame of infinite depth then it does not have interpolation.*

So NExtLin provides an interesting example of a rather complex lattice of modal logics for which almost all important properties of calculi are decidable. We shall not go into details of the proof here but discuss quite natural criteria for canonicity and strong completeness of logics in NExtLin required to prove this theorem. Denote by  $\mathcal{B}_+$  the class of frames containing  $\mathcal{B}$  together with frames  $\mathfrak{C}(n_1, \mathbb{k}, n_2)$  defined as follows. Suppose  $k > 1$ ,  $n_1, n_2 < \omega$  are such that  $n_1 + n_2 > 0$  and  $\mathbb{k} = \{a_0, \dots, a_{k-1}\}$ . Then

$$\mathfrak{C}(n_1, \mathbb{k}, n_2) = \langle \omega^{<}(n_1) \triangleleft \mathbb{k} \triangleleft \omega^{>}(n_2), P \rangle,$$

where  $P$  is the set of possible values generated by  $\{X_i : 0 \leq i \leq k-1\}$ , for

$$X_i = \{a_i\} \cup \{kj + i : j \in \omega\} \cup \{k^*j^* + i^* : j \in \omega\}$$

and  $\{0^*, 1^*, \dots, n^*, \dots\}$  being the points in  $\omega^{>}(n_2)$ .

Let  $\mathcal{F}$  be the class of frames of the form

$$\langle \{0, \dots, n_1\}, <, > \rangle \triangleleft \mathbb{1} \triangleleft \langle \{0, \dots, n_2\}, <, > \rangle \text{ or } \langle \{0, \dots, n\}, <, > \rangle.$$

**THEOREM 143.** (i) *A logic  $L \in \text{NExtLin}$  is canonical iff the underlying Kripke frame of each frame  $\mathfrak{F} \in [\mathcal{B}_+^*]$  for  $L$  validates  $L$  as well.*

(ii) *A logic  $L \in \text{NExtLin}$  is strongly complete iff for each frame  $\mathfrak{F} \in [\mathcal{B}_+^*]$  validating  $L$ , there exists a Kripke frame  $\mathfrak{G}$  for  $L$  which results from  $\mathfrak{F}$  by replacing*

- every  $\mathfrak{C}(n, \mathbb{k})$  with  $\omega^{<}(n)$  or  $\omega^{<}(n) \triangleleft \mathfrak{H} \triangleleft \mathbb{k}$ , for some  $\mathfrak{H} \in \mathcal{F}$ , and
- every  $\mathfrak{C}(\mathbb{k}, n)$  with  $\omega^{>}(n)$  or  $\mathbb{k} \triangleleft \mathfrak{H} \triangleleft \omega^{>}(n)$ , for some  $\mathfrak{H} \in \mathcal{F}$ , and
- every  $\mathfrak{C}(n_1, \mathbb{k}, n_2)$  with  $\omega^{<}(n_1) \triangleleft \mathfrak{H} \triangleleft \omega^{>}(n_2)$ , for some  $\mathfrak{H} \in \mathcal{F}$ .

**EXAMPLE 144.** The logic  $\mathbf{R}_t$  is not canonical because  $\mathfrak{C}(2, \mathbb{2}) \models \mathbf{R}_t$  but  $\omega^{<}(2) \triangleleft \mathbb{2} \not\models \mathbf{R}_t$ . However,  $\mathbf{R}_t$  is strongly complete, since  $\mathfrak{F} \models \mathbf{R}_t$  whenever  $\mathfrak{G} \in [\mathcal{B}_+^*]$  validates  $\mathbf{R}_t$  and  $\mathfrak{F}$  is obtained from  $\mathfrak{G}$  as in the formulation of Theorem 143 with  $\mathfrak{H} = \bullet \in \mathcal{F}$ .

One can also use Theorem 143 to construct two strongly complete logics  $L_1, L_2 \in \text{NExtLin}$  whose sum  $L_1 \oplus L_2$  is not strongly complete (see [Wolter 1996b]).

## 2.6 Bimodal provability logics

Bimodal provability logics emerge when combinations of two different provability predicates are investigated, for example, if  $\Box_1$  is understood as “it



is provable in PA” and  $\Box_2$  as “it is provable in ZF”. In contrast to the situation in unimodal provability logic, where almost all provability predicates behave like the necessity operator  $\Box$  in **GL**, there exist quite a lot of different types of bimodal provability logics. Various completeness results extending Solovay’s completeness theorem for **GL** to the bimodal case were established by Smoryński [1985], Montagna [1987], Beklemishev [1994, 1996] and Visser [1995]. Here we will not deal with the interpretation of modal operators as provability predicates but sketch some results on modal logics containing the bimodal provability logic

$$\mathbf{CSM}_0 = (\mathbf{GL} \otimes \mathbf{GL}) \oplus \Box_1 p \rightarrow \Box_2 p \oplus \Box_2 p \rightarrow \Box_1 \Box_2 p$$

(named so by Visser [1995] after Carlson, Smoryński and Montagna). A number of provability logics is included in this class, witness the list below. (As in unimodal provability logic we have quasi-normal logics among them, i.e., sets of formulas containing  $\mathbf{K}_2$  and closed under modus ponens and substitutions (but not necessarily under  $\varphi/\Box_i\varphi$ ). Recall that we denote by  $L + \Gamma$  the smallest quasi-normal logic containing  $L$  and  $\Gamma$ .)

- $\mathbf{CSM}_1 = \mathbf{CSM}_0 \oplus \Box_2(\Box_1 p \rightarrow p)$ . (This is  $\mathbf{PRL}_{ZF}$  in [Smoryński 1985] and  $\mathbf{F}$  in [Montagna 1987].)
- $\mathbf{NB}_1 = \mathbf{CSM}_0 \oplus (\neg\Box_1 p \wedge \Box_2 p) \rightarrow \Box_2(\Box_1 q \rightarrow q)$ .
- $\mathbf{CSM}_2 = \mathbf{CSM}_1 + \Box_1 p \rightarrow p$ . (This is  $\mathbf{PRL}_{ZF} + \text{Reflection}_{\Box_1}$  in [Smoryński 1985] and  $\mathbf{F}_1$  in [Montagna 1987].)
- $\mathbf{CSM}_3 = \mathbf{CSM}_2 + \Box_2 p \rightarrow p$ . (This is  $\mathbf{PRL}_{ZF} + \text{Reflection}_{\Box_2}$  in [Smoryński 1985].)
- $\mathbf{NB}_2 = \mathbf{NB}_1 + \Box_2 p \rightarrow p + \Box_2 p \rightarrow \Box_1 p$ .

A remarkable feature of  $\mathbf{CSM}_0$  is that—like in **GL**—we have uniquely determined definable fixed points.

**THEOREM 145** (Smoryński 1985). *Let  $\varphi(p)$  be a formula in which every occurrence of  $p$  lies within the scope of some  $\Box_1$  or some  $\Box_2$ . Then*

- (i) *there exists a formula  $\psi$  containing only the propositional variables of  $\varphi(p)$  different from  $p$  such that  $\psi \leftrightarrow \varphi(\psi) \in \mathbf{CSM}_0$ ;*
- (ii)  $\Box_1((p \leftrightarrow \varphi(p)) \wedge (q \leftrightarrow \varphi(q))) \rightarrow (p \leftrightarrow q) \in \mathbf{CSM}_0$ .

In the remaining part of this section we are concerned with subframe logics containing  $\mathbf{CSM}_0$ , the main result stating that those of them that are finitely axiomatizable are decidable. All the provability logics introduced above turn out to be subframe logics, so we obtain a uniform proof of their decidability. An interesting trait of subframe logics in  $\text{Ext}\mathbf{CSM}_0$  is that (as a rule) they are Kripke incomplete; in the list above such are  $\mathbf{CSM}_i$ ,

$i = 1, 2, 3$ , and  $\mathbf{NB}_i$ ,  $i = 1, 2$ . The proof extends the techniques introduced by Visser [1995]; for details we refer the reader to [Wolter 1998].

First we develop—as was done for  $\mathbf{NExtK4}$  and  $\mathbf{NExtLin}$ —a frame theoretic language for axiomatizing subframe logics in the lattice  $\mathbf{ExtCSM}_0$ . A finite frame  $\mathfrak{G} = \langle W, R_1, R_2 \rangle$  validates  $\mathbf{CSM}_0$  iff both  $R_1$  and  $R_2$  are transitive, irreflexive,  $R_2 \subseteq R_1$  and

$$\forall x, y, z (xR_1y \wedge yR_2z \rightarrow xR_2z).$$

In this section all (not only finite) frames are assumed to satisfy these conditions, *save irreflexivity*.

A finite frame  $\mathfrak{F}$  is called a *surrogate frame* if it has precisely one root  $r$  and all points different from  $r$  are  $R_2$ -irreflexive. Surrogate frames will provide the language to axiomatize subframe logics in  $\mathbf{ExtCSM}_0$ . A *normal surrogate frame*  $\langle W, R_1, R_2 \rangle$  is a surrogate frame in which the root  $r$  is  $R_1$ -irreflexive. We write  $xR_i^p y$  iff  $xR_i y$  and  $\neg yR_i x$ . Given a frame  $\mathfrak{G} = \langle V, S_1, S_2, Q \rangle$  for  $\mathbf{CSM}_0$  and a surrogate frame  $\mathfrak{F} = \langle W, R_1, R_2 \rangle$ , a map  $h$  from  $V$  onto  $W$  is called a *weak reduction* of  $\mathfrak{G}$  to  $\mathfrak{F}$  if for  $i \in \{1, 2\}$  and all  $x, y \in V$ ,

- $xS_i y$  implies  $f(x)R_i f(y)$ ,
- $f(x)R_i^p f(y)$  implies  $\exists z \in V (xS_i z \wedge f(z) = f(y))$ ,
- $f^{-1}(X) \in Q$  for all  $X \subseteq W$ .

(The standard definition of reduction is relaxed here in the second condition.) Each weak reduction to a  $\mathbf{CSM}_0$ -frames is a usual reduction, since in this case  $R_i^p = R_i$ . A frame  $\mathfrak{G}$  is said to be *weakly subreducible* to a surrogate frame  $\mathfrak{F}$  if a subframe of  $\mathfrak{G}$  is weakly reducible to  $\mathfrak{F}$ . To describe weak subreducibility syntactically, with each surrogate frame  $\mathfrak{F} = \langle W, R_1, R_2 \rangle$  we associate the formula

$$\alpha(\mathfrak{F}) = \delta(\mathfrak{F}) \wedge \Box_1 \delta(\mathfrak{F}) \rightarrow \neg p_r,$$

where  $r$  is the root of  $\mathfrak{F}$  and

$$\begin{aligned} \delta(\mathfrak{F}) = & \bigwedge \{p_x \rightarrow \Diamond_1 p_y : xR_1^p y, x, y \in W\} \wedge \\ & \bigwedge \{p_x \rightarrow \Diamond_2 p_y : xR_2^p y, x, y \in W\} \wedge \\ & \bigwedge \{p_x \rightarrow \neg p_y : x \neq y, x, y \in W\} \wedge \\ & \bigwedge \{p_x \rightarrow \neg \Diamond_1 p_y : \neg(xR_1 y), x, y \in W\} \wedge \\ & \bigwedge \{p_x \rightarrow \neg \Diamond_2 p_y : \neg(xR_2 y), x, y \in W\}. \end{aligned}$$

**LEMMA 146.** *For every surrogate frame  $\mathfrak{F}$  and every  $\mathbf{CSM}_0$ -frame  $\mathfrak{G}$ , we have  $\mathfrak{G} \not\models \alpha(\mathfrak{F})$  iff  $\mathfrak{G}$  is weakly subreducible to  $\mathfrak{F}$ .*

Table 4. Axiomatizations of provability logics

---

$\mathbf{CSM}_1$	$=$	$\mathbf{CSM}_0 \oplus \alpha(\overset{\bullet}{\downarrow})$
$\mathbf{CSM}_0 + \Box_1 p \rightarrow p$	$=$	$\mathbf{CSM}_0 + \alpha(\bullet)$
$\mathbf{CSM}_0 + \Box_2 p \rightarrow p$	$=$	$\mathbf{CSM}_0 + \alpha(\overset{1}{\circ})$
$\mathbf{CSM}_0 + \Box_2 p \rightarrow \Box_1 p$	$=$	$\mathbf{CSM}_0 + \alpha(\overset{\circ}{\circ} \overset{1}{1})$
$\mathbf{NB}_1$	$=$	$\mathbf{CSM}_0 \oplus \alpha(\begin{array}{c} \bullet \xrightarrow{1} \bullet \\ \swarrow \quad \searrow \\ \bullet \end{array}) \oplus \alpha(\begin{array}{c} \bullet \quad \bullet \\ \swarrow \quad \searrow \\ \bullet \end{array}) \oplus$ $\alpha(\begin{array}{c} \bullet \xrightarrow{1} \bullet \\ \swarrow \quad \searrow \\ \bullet \end{array}) \oplus \alpha(\begin{array}{c} \bullet \quad \bullet \\ \swarrow \quad \searrow \\ \bullet \end{array})$

---

It follows immediately that  $\mathbf{CSM}_0 \oplus \alpha(\mathfrak{F})$  and  $\mathbf{CSM}_0 + \alpha(\mathfrak{F})$  are subframe logics. Conversely, we have the following completeness result.

**THEOREM 147.**

- (i) *There is an algorithm which, given a formula  $\varphi$  such that  $\mathbf{CSM}_0 + \varphi$  is a subframe logic, returns surrogate frames  $\mathfrak{F}_1, \dots, \mathfrak{F}_n$  for which*

$$\mathbf{CSM}_0 + \varphi = \mathbf{CSM}_0 + \alpha(\mathfrak{F}_1) + \dots + \alpha(\mathfrak{F}_n).$$

- (ii) *There is an algorithm which, given a formula  $\varphi$  such that  $\mathbf{CSM}_0 \oplus \varphi$  is a subframe logic, returns normal surrogate frames  $\mathfrak{F}_1, \dots, \mathfrak{F}_n$  such that*

$$\mathbf{CSM}_0 \oplus \varphi = \mathbf{CSM}_0 \oplus \alpha(\mathfrak{F}_1) \oplus \dots \oplus \alpha(\mathfrak{F}_n).$$

Table 4 shows axiomatizations of the logics introduced above by means of formulas of the form  $\alpha(\mathfrak{F})$ . In this section we adopt the convention that in figures we place the number 1 nearby an arrow from  $x$  to  $y$  if  $xR_1y$  and  $\neg xR_2y$ . An arrow without a number means that  $xR_2y$  (and therefore  $xR_1y$  as well).

The proof of decidability is based on the completeness of subframe logics in  $\text{ExtCSM}_0$  with respect to rather simple descriptive frames. With every surrogate frame  $\mathfrak{F}$  we associate a finite set of frames

$$E(\mathfrak{F}) = \{\mathfrak{F}_{\overline{A}} : \overline{A} \in \text{Seq}\mathfrak{F}\}.$$

Loosely, it is defined as follows. Let us first assume that the root  $r$  of  $\mathfrak{F}$  is  $R_2$ -irreflexive. Then the frames in  $E(\mathfrak{F})$  are the results of inserting an infinite strictly descending  $R_1$ -chain, denoted by  $C(\omega)$ , between each non-degenerate  $R_1$ -cluster  $C$  and its  $R_1$ -successors. This defines  $R_1$  uniquely. However,  $R_2$  may be defined in different ways, since a point  $R_2$ -seeing a point in  $C$  need not (but may)  $R_2$ -see certain points in the chain  $C(\omega)$ .

To be more precise, the set  $\text{Seq}\mathfrak{F}$  consists of all sequences  $\overline{A}$  of the form

$$\overline{A} = \langle A_x : xR_1x, x \in W \rangle.$$

where  $A_x$  is a subset of  $\{y \in W - C : yR_2x\}$  such that for all  $y$  and  $z$ ,  $y \in A_x$  and  $zR_1y$  imply  $z \in A_x$ . For each non-degenerate  $R_1$ -cluster  $C$ , denote by  $C(\omega)$  the set  $\{(n, C) : n \in \omega\}$ . Finally, given  $\overline{A} \in \text{Seq}\mathfrak{F}$ , we construct  $\mathfrak{F}_{\overline{A}} = \langle V, S_0, S_1 \rangle$  as the frame satisfying the following conditions:

- $V = W \cup \bigcup \{C(\omega) : C \text{ a non-degenerate } R_1\text{-cluster in } \mathfrak{F}\}$ ;
- $R_i = S_i \cap (W \times W)$ , for  $i \in \{1, 2\}$ ;
- $S_1$  is defined so that  $C(\omega)$  becomes an infinite descending chain between  $C$  and its immediate successors;
- for every non-degenerate  $R_1$ -cluster  $C$ ,
  - $((C(\omega) \cup C) \times (C(\omega) \cup C)) \cap S_2 = \emptyset$ ,
  - for all  $y \in W - C$  and  $x \in C(\omega)$ ,  $xS_2y$  iff  $CR_2y$ ,
  - for all  $y \in W - C$ ,  $C = \{j : 0 \leq j \leq m - 1\}$  and  $x \in C(\omega)$ ,  $yS_2x$  iff  $\exists i \in \omega \exists j \leq m - 1 (x = (im + j, C) \wedge y \in A_j)$ ,
  - for all  $x \in C(\omega)$  and  $y \in V - C$ ,  $xS_2y$  iff  $CS_2y$ .

We illustrate this technical definition by a simple example.

**EXAMPLE 148.** Construct  $E(\mathfrak{F})$  for the frame  $\mathfrak{F}$  in Fig. 12 (a). In this case we have two  $R_1$ -reflexive points, namely  $c$  and  $d$ . So,  $\text{Seq}\mathfrak{F}$  consists of pairs  $\langle A_c, A_d \rangle$ . There are four different pairs and so we have four frames in  $E(\mathfrak{F})$ : the frame in Fig. 12 (b) is  $\mathfrak{F}_{\langle \emptyset, \emptyset \rangle}$  and that in (c) is  $\mathfrak{F}_{\langle \{a\}, \{b\} \rangle}$ .  $\mathfrak{F}_{\langle \emptyset, \{b\} \rangle}$  is obtained from  $\mathfrak{F}_{\langle \{a\}, \{b\} \rangle}$  by omitting the  $R_2$ -arrows starting from  $a$ , save the arrow to  $c$ , and  $\mathfrak{F}_{\langle \{a\}, \emptyset \rangle}$  is obtained from  $\mathfrak{F}_{\langle \{a\}, \{b\} \rangle}$  by omitting the  $R_2$ -arrows starting from  $b$ , save the arrow to  $d$ .

Suppose now that the root  $r$  of  $\mathfrak{F} = \langle W, R_1, R_2 \rangle$  is  $R_2$ -reflexive. We define  $\mathfrak{F}_{\overline{A}}$  as in the previous case, but this time we also insert an infinite strictly descending  $R_2$ -chain  $C(\omega)$  between  $r$  and its  $R_1$ -successors.

We have defined the relational component of our frames and now turn to their sets of possible values. Given  $\mathfrak{F}_{\overline{A}} = \langle V, S_1, S_2 \rangle$  and a non-degenerate  $R_1$ -cluster  $C = \{j : 0 \leq j \leq m - 1\}$  in  $\mathfrak{F}$ , let

$$P_C = \{\{j\} \cup \{(im + j, C) : i \in \omega\} : j = 0, \dots, m - 1\}$$

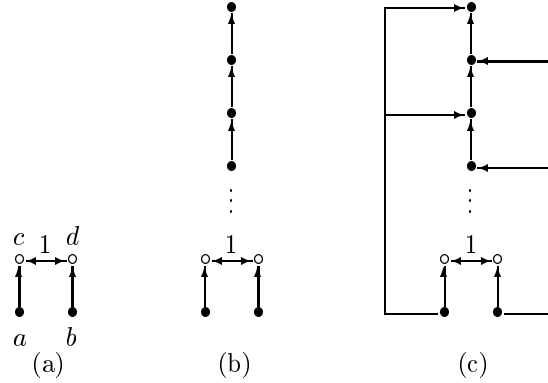


Figure 12.

and denote by  $P$  the closure of

$$\{\{x\} : x \in V, \neg xS_1x\} \cup \{P_C : C \text{ is a non-degenerate } R_1\text{-cluster in } \mathfrak{F}\}$$

under intersections and complements in  $V$ . The resultant general frame is denoted by  $\mathfrak{G}(\mathfrak{F}_{\bar{A}}) = \langle V, S_1, S_2, P \rangle$ . One can check that it is a descriptive frame for  $\mathbf{CSM}_0$ . The following completeness result is proved similarly to that in Section 2.4.

**THEOREM 149.**

- (i) *Each subframe logic in  $\mathbf{NExtCSM}_0$  is determined by a set of frames of the form  $\mathfrak{G}(\mathfrak{F}_{\bar{A}})$ , in which  $\mathfrak{F}$  is a normal surrogate frame and  $\bar{A} \in \text{Seq}\mathfrak{F}$ .*
- (ii) *Each subframe logic in  $\mathbf{ExtCSM}_0$  is determined by a set of frames with distinguished worlds of the form  $\langle \mathfrak{G}(\mathfrak{F}_{\bar{A}}), r \rangle$  in which  $\mathfrak{F}$  is a surrogate frame with root  $r$  and  $\bar{A} \in \text{Seq}\mathfrak{F}$ .*

As a consequence of Theorem 149 and the fact that, for each surrogate frame  $\mathfrak{F}$  with root  $r$  and each  $\bar{A} \in \text{Seq}\mathfrak{F}$ , both the logics of  $\mathfrak{G}(\mathfrak{F}_{\bar{A}})$  and  $\langle \mathfrak{G}(\mathfrak{F}_{\bar{A}}), r \rangle$  are decidable, we obtain

**THEOREM 150.** *All finitely axiomatizable subframe logics in  $\mathbf{ExtCSM}_0$  are decidable.*

We conjecture that the method above can be extended to logics without the **GL**-axioms, i.e., all finitely axiomatizable subframe logics containing

$$(\mathbf{K4} \otimes \mathbf{K4}) \oplus \Box_1 p \rightarrow \Box_2 p \oplus \Box_2 p \rightarrow \Box_1 \Box_2 p$$

are decidable.

### 2.7 Cartesian products of modal logics

Polymodal logics can be used for talking about multi-dimensional relational structures such as Cartesian products of Kripke frames. The formation of products is probably the most natural way of introducing a concept of dimension in modal logic in order to reflect interactions between modal operators representing time, space, knowledge, actions, etc. Products of modal logics (i.e., the sets of polymodal formulas that are valid in Cartesian products of Kripke frames for those logics) have been studied in both pure modal logic (see e.g. [Seegerberg 1973], [Shehtman 1978a], [Marx and Venema 1997], [Gabbay and Shehtman 1998], [Kurucz 2000], [Wolter 2000]) and applications in computer science and artificial intelligence (see e.g. [Reif and Sistla 1985], [Fagin *et al.* 1995], [Baader and Ohlbach 1995], [Finger and Reynolds 1999], [Wolter and Zakharyashev 1998, 1999b, 1999c, 2000]) since the 1970s. (Products of modal logics are also relevant to finite variable fragments of modal and intermediate predicate logics; see [Gabbay and Shehtman 1993].)

The (*Cartesian*) *product* of two frames  $\mathfrak{F}_1 = \langle W_1, R_1 \rangle$  and  $\mathfrak{F}_2 = \langle W_2, R_2 \rangle$  is the bimodal frame of the form

$$\mathfrak{F}_1 \times \mathfrak{F}_2 = \langle W_1 \times W_2, R_h, R_v \rangle$$

in which, for all  $u_1, u_2 \in W_1$  and  $v_1, v_2 \in W_2$ ,

$$\begin{aligned} \langle u_1, v_1 \rangle R_h \langle u_2, v_2 \rangle &\text{ iff } u_1 R_1 u_2 \text{ and } v_1 = v_2, \\ \langle u_1, v_1 \rangle R_v \langle u_2, v_2 \rangle &\text{ iff } u_1 = u_2 \text{ and } v_1 R_2 v_2. \end{aligned}$$

The subscripts  $h$  and  $v$  appeal to the geometrical intuition of considering  $R_h$  as the “horizontal” accessibility relation in  $\mathfrak{F}_1 \times \mathfrak{F}_2$  and  $R_v$  as the “vertical” one.

Let  $\mathcal{L}_2$  be the bimodal language with boxes  $\Box$  and  $\sqcup$  (and their duals  $\Diamond$  and  $\diamond$ ). Frames for this language will be denoted by  $\mathfrak{F} = \langle W, R_h, R_v \rangle$ , so that  $\Box$  and  $\sqcup$  are interpreted by the relations  $R_h$  and  $R_v$ , respectively. Products are just special frames of this form.

Every product  $\mathfrak{F} = \mathfrak{F}_1 \times \mathfrak{F}_2$  satisfies the following two important properties:  $R_v \circ R_h = R_h \circ R_v$  and  $R_v^{-1} \circ R_h \subseteq R_h \circ R_v^{-1}$ , known as *commutativity* and the *Church–Rosser property*, respectively.  $\mathfrak{F}$  is commutative iff it validates the formula

$$\mathbf{com} = \diamond \Diamond p \leftrightarrow \Diamond \diamond p,$$

and  $\mathfrak{F}$  is Church–Rosser iff it validates

$$\mathbf{chr} = \diamond \sqcup p \rightarrow \sqcup \diamond p.$$

It is to be noted, however, that these properties are not characteristic for products: there are bimodal commutative and Church–Rosser frames that are not (isomorphic to) products of any two frames.

Given two classes of Kripke frames  $\mathcal{C}_1$  and  $\mathcal{C}_2$ , we define their (*Cartesian*) *product*  $\mathcal{C}_1 \times \mathcal{C}_2$  by taking

$$\mathcal{C}_1 \times \mathcal{C}_2 = \{\mathfrak{F}_1 \times \mathfrak{F}_2 : \mathfrak{F}_1 \in \mathcal{C}_1, \mathfrak{F}_2 \in \mathcal{C}_2\}.$$

Let  $\mathcal{C}_1$  and  $\mathcal{C}_2$  be the classes of all Kripke frames for complete unimodal logics  $L_1$  and  $L_2$ . Their (*Cartesian*) *product*  $L_1 \times L_2$  is defined as the bimodal logic  $\text{Log}(\mathcal{C}_1 \times \mathcal{C}_2)$  in the language  $\mathcal{L}_2$ . It is easy to see that  $L_1 \times L_2$  is a conservative extension of both  $L_1$  and  $L_2$ , and that

$$L_1 \times L_2 \supseteq (L_1 \otimes L_2) \oplus \mathbf{com} \oplus \mathbf{chr}.$$

In some important cases the converse inclusion also holds:

**THEOREM 151** (Gabbay and Shehtman 1998). *Suppose both  $L_1$  and  $L_2$  are axiomatized by variable free formulas and formulas of the form*

$$\diamond^n \Box p \rightarrow \Box^m p.$$

*Then*

$$L_1 \times L_2 = (L_1 \otimes L_2) \oplus \mathbf{com} \oplus \mathbf{chr}.$$

This theorem yields, for instance, axiomatizations for products of **K**, **D**, **K4**, **T**, **S4**, **S5**. However there are many products of standard logics which cannot be axiomatized in this canonical way, for instance, **Grz**  $\times$  **Grz** (see [Gabbay and Shehtman 1998]). Moreover, products of logics of linear frames may be even not recursively enumerable, e.g. **GL.3**  $\times$  **GL.3** (see [Reynolds and Zakharyashev 2000]). (On the other hand, as was observed in [Gabbay and Shehtman 1998], if classes  $\mathcal{C}_1, \dots, \mathcal{C}_n$  are elementary and recursive then  $\text{Log}(\mathcal{C}_1 \times \dots \times \mathcal{C}_n)$  is recursively axiomatizable.)

In contrast to the unimodal case, usually it is rather difficult to prove positive results about products of even simple standard logics. Here we illustrate one of the methods of establishing FMP and decidability developed in [Wolter and Zakharyashev 1998, 1999b] by applying it to **S5**  $\times$  **S5**. Other techniques—filtration, finite depth method, and mosaic—can be found in [Gabbay and Shehtman 1998] and [Marx and Venema 1997].

**S5**  $\times$  **S5** is clearly determined by the class of products of universal frames which will be called **S5-rectangles**. Suppose we are given a formula  $\varphi$  and want to find out whether it is satisfiable in some **S5**-rectangle.

Let us call a *type* for  $\varphi$  any subset  $t$  of **Sub** $\varphi$  such that

- $\psi \wedge \chi \in t$  iff  $\psi \in t$  and  $\chi \in t$ , for every  $\psi \wedge \chi \in \mathbf{Sub}\varphi$ ,
- $\neg\psi \in t$  iff  $\psi \notin t$ , for every  $\neg\psi \in \mathbf{Sub}\varphi$ .

A *type-cluster* for  $\varphi$  is a set  $T$  of distinct types for  $\varphi$  such that

$$\forall t \in T \forall \psi \in \mathbf{Sub}\varphi (\psi \in t \leftrightarrow \exists t' \in T \psi \in t').$$

Let  $\Omega$  be a non-empty set elements  $a$  in which are labelled by type-clusters  $\mathbf{T}$  for  $\varphi$ . In other words, we can think of  $\Omega$  as consisting of pairs of the form  $\mathbf{T}_a$  with pairwise distinct  $a$ .

A *run through*  $\Omega$  (or a  $\Omega$ -*run*, for short) is a function  $\mathbf{r}$  from  $\Omega$  to the set of types for  $\varphi$  such that

- $\mathbf{r}(\mathbf{T}_a) \in \mathbf{T}$  for every  $\mathbf{T}_a \in \Omega$ , and
- $\forall \mathbf{T}_a \in \Omega \forall \diamond\psi \in \mathbf{Sub}\varphi (\diamond\psi \in \mathbf{r}(\mathbf{T}_a) \leftrightarrow \exists sc_b \in \Omega \psi \in \mathbf{r}(sc_b))$ .

Say that  $\Omega$  is a *quasimodel* for  $\varphi$  if, for every  $\mathbf{T}_a \in \Omega$  and every  $t \in \mathbf{T}$ , there is a  $\Omega$ -run  $\mathbf{r}$  coming through  $t$ , i.e.,  $\mathbf{r}(\mathbf{T}_a) = t$ .  $\Omega$  *satisfies*  $\varphi$  if  $\varphi$  belongs to a type occurring in a type-cluster in  $\Omega$ .

One can readily show that a formula  $\varphi$  is satisfied in an **S5**-rectangle iff  $\varphi$  is satisfied in some quasimodel for  $\varphi$ .

We prove now that every satisfiable formula  $\varphi$  is satisfied in a quasimodel of some bounded size. Let  $\Omega$  be a quasimodel satisfying  $\varphi$ . Construct a subquasimodel  $\Omega'$  of  $\Omega$  in the following way. To begin with, we put in  $\Omega'$  an element  $a_0$  from  $\Omega$  labelled by a type-cluster  $\mathbf{T}$  containing a type with  $\varphi$ . Then, for every  $t \in \mathbf{T}$  we take a run  $\mathbf{r}$  coming through  $t$ , and for each  $\diamond\psi \in t$  select  $\mathbf{r}(a)$  containing  $\psi$  and put in  $\Omega'$  the element  $a$  together with its copy  $a'$  (labelled by the same type-cluster as  $a$ ). Thus the resulting  $\Omega'$  contains at most  $2^{|\mathbf{Sub}\varphi|} \cdot 2 \cdot |\mathbf{Sub}\varphi|$  elements. It is now easy to see that  $\Omega'$  is a quasimodel satisfying  $\varphi$ . For suppose we have an element  $a \in \Omega'$  labelled by  $\mathbf{T}$  and a  $t \in \mathbf{T}$ . If  $a = a_0$  then, by the construction, we have a  $\Omega'$ -run coming through  $t$ . Assume now that  $a \neq a_0$ . We know that there is a  $\Omega$ -run  $\mathbf{r}'$  through  $t$ . Let  $\mathbf{r}'(a_0) = t'$ . By the construction we have a  $\Omega'$ -run  $\mathbf{r}''$  through  $t'$ . But then the function  $\mathbf{r}$  defined by

$$\mathbf{r}(b) = \begin{cases} \mathbf{r}'(b) & \text{if } b = a \\ \mathbf{r}''(b) & \text{otherwise,} \end{cases}$$

for  $b \in \Omega'$ , is a run in  $\Omega'$  coming through  $t$ . Note that this gives us  $2^{2^{|\mathbf{Sub}\varphi|}} \cdot 2 \cdot |\mathbf{Sub}\varphi|$  runs in  $\Omega'$  coming through all its types. As a consequence we obtain:

**THEOREM 152.** *Every formula satisfiable in an **S5**-rectangle is satisfied in an **S5**-rectangle containing at most  $2^{3^{|\mathbf{Sub}\varphi|}} \cdot 4 \cdot |\mathbf{Sub}\varphi|^2$  points. Thus **S5**  $\times$  **S5** has FMP and is decidable.*

Unfortunately, there is no general transfer theorem that could guarantee the preservation of such properties of logics as decidability, axiomatizability, or interpolation under the formation of products. If we consider only products of standard modal logics, then the results obtained so far can roughly be described as follows (for more details consult [Spaan 1993], [Marx and Venema 1997], [Gabbay and Shehtman 1998], [Marx and Areces 1998], [Marx 1999], [Wolter 2000], [Reynolds and Zakharyashev 2000]):



- logics of the form  $\mathbf{K} \times L$  and  $\mathbf{S5} \times L$  are usually decidable (in particular, for  $L \in \{\mathbf{K}, \mathbf{T}, \mathbf{K4}, \mathbf{K4.3}, \mathbf{S4}, \mathbf{S5}\}$ );
- products of logics determined by infinite linear orders are undecidable,
- the computational complexity of a decidable product is usually substantially higher than the complexity of its components; for example, the satisfiability problem for  $\mathbf{S5} \times \mathbf{S5}$  is *NEXPTIME*-complete.

The decidability and FMP of logics like  $\mathbf{K4} \times \mathbf{K4}$ ,  $\mathbf{S4} \times \mathbf{S4}$ ,  $\mathbf{S4} \times \mathbf{S4.3}$  remain challenging open problems of the field.

In higher dimensions—for  $n \geq 3$ —the first results related to products of modal logics were obtained in the framework of algebraic logic: as follows from [Maddux 1980] and [Johnson 1969],  $\mathbf{S5}^n$  is undecidable and not finitely axiomatizable. However, as we mentioned above, products like  $\mathbf{S5}^n$  and  $\mathbf{K}^n$  are recursively enumerable. It is worth noting that although  $\mathbf{K}^n$  have FMP for all  $n < \omega$  [Gabbay and Shehtman 1998], this could imply the decidability of  $\mathbf{K}^n$  only if the class of finite frames for  $\mathbf{K}^n$  were recursive. We could have such a test if  $\mathbf{K}^n$  were finitely axiomatizable; however this is not the case [Kurucz 2000]. To prove the decidability of  $\mathbf{K}^n$ , it would also be enough to show that it has the product FMP, i.e., it is characterized by the class of products of  $n$ -many finite frames. But this approach does not work either: as has been shown by Hirsch *et al.* [2000], all logics  $L$  such that  $\mathbf{K}^n \subseteq L \subseteq \mathbf{S5}^n$  are undecidable, non finitely axiomatizable, do not have the product FMP, and it is undecidable whether a finite frame is a frame for  $L$ . (The only known example of a decidable higher dimensional product of non-tabular logics is  $\mathbf{Alt}^n$  [Gabbay and Shehtman 1998].)

### 3 SUPERINTUITIONISTIC LOGICS

Although C.I. Lewis constructed his first modal calculus  $\mathbf{S3}$  in 1918, it was Gödel’s [1933] two page note that attracted serious attention of mathematical logicians to modal systems. While Lewis [1918] used an abstract necessity operator to avoid paradoxes of material implication, Gödel [1933] and earlier Orlov [1928]<sup>16</sup> treated  $\Box$  as “it is provable” to give a classical interpretation of intuitionistic propositional logic  $\mathbf{Int}$  by means of embedding it into a modal “provability” system which turned out to be equivalent to Lewis’  $\mathbf{S4}$ .

Approximately at the same time Gödel [1932] observed that there are infinitely many logics located between  $\mathbf{Int}$  and classical logic  $\mathbf{Cl}$ , which—together with the creation of constructive (proper) extensions of  $\mathbf{Int}$  by Kleene [1945] and Rose [1953] (realizability logic), Medvedev [1962] (logic

<sup>16</sup>Orlov’s paper remained unnoticed till the end of the 1980s. It is remarkable also for constructing the first system of relevant logic.

of finite problems), Kreisel and Putnam [1957]—gave an impetus to studying the class of logics intermediate between **Int** and **Cl**, started by Umezawa [1955, 1959]. Gödel’s embedding of **Int** into **S4**, presented in an algebraic form by McKinsey and Tarski [1948] and extended to all intermediate logics by Dummett and Lemmon [1959], made it possible to develop the theories of modal and intermediate logics in parallel ways. And the structural results of Blok [1976] and Esakia [1979a,b], establishing an isomorphism between the lattices **ExtInt** and **NExtGrz**, along with preservation results of Maksimova and Rybakov [1974] and Zakharyashev [1991], transferring various properties from modal to intermediate logics and back, showed that in many respects the theory of intermediate logics is reducible to the theory of logics in **NExtS4**.

To demonstrate this as well as some features of intermediate logics is the main aim of this part. We will use the same system of notations as in the modal case. In particular, **ExtInt** is the lattice of all logics of the form **Int** +  $\Gamma$  (where  $\Gamma$  is an arbitrary set of formulas in the language of **Int** and + as before means taking the closure under modus ponens and substitution); we call them *superintuitionistic logics* or *si-logics* for short. Basic facts about the syntax and semantics of **Int** and relevant references can be found in *Intuitionistic Logic*, see volume 7 of this *Handbook*. A list of some “standard” si-logics is given in Table 5.

### 3.1 Intuitionistic frames

As in the case of modal logics, the adequate relational semantics for si-logics can be constructed on the base of the Stone representation of the algebraic “models” for **Int**, known as *Heyting* (or *pseudo-Boolean*) *algebras*. It is hard to trace now who was the first to introduce intuitionistic general frames—the earliest references we know are [Esakia 1974] and [Rautenberg 1979]—but in any case, having at hand [Jónsson and Tarski 1951] and [Goldblatt 1976a], the construction must have been clear.

An *intuitionistic (general) frame* is a triple  $\mathfrak{F} = \langle W, R, P \rangle$  in which  $R$  is a partial order on  $W \neq \emptyset$  and  $P$ , the *set of possible values* in  $\mathfrak{F}$ , is a collection of upward closed subsets (cones) in  $W$  containing  $\emptyset$  and closed under the Boolean  $\cap, \cup$ , and the operation  $\supset$  (for  $\rightarrow$ ) defined by

$$X \supset Y = \{x \in W : \forall y \in x \uparrow (y \in X \rightarrow y \in Y)\}.$$

If  $P$  contains all upward closed subsets in  $W$  then we call  $\mathfrak{F}$  a *Kripke frame* and denote it by  $\mathfrak{F} = \langle W, R \rangle$ . An important feature of intuitionistic models  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  ( $\mathfrak{V}$ , a *valuation* in  $\mathfrak{F}$ , maps propositional variables to sets in  $P$ ) is that  $\mathfrak{V}(\varphi)$ , the *truth-value* of a formula  $\varphi$ , is always upward closed.

Every intuitionistic frame  $\mathfrak{F} = \langle W, R, P \rangle$  gives rise to the Heyting algebra  $\mathfrak{F}^+ = \langle P, \cap, \cup, \supset, \emptyset \rangle$  called the *dual* of  $\mathfrak{F}$ . Conversely, given a Heyting algebra

Table 5. A list of standard superintuitionistic logics

---

<b>For</b>	= <b>Int</b> + $p$
<b>CI</b>	= <b>Int</b> + $p \vee \neg p$
<b>SmL</b>	= <b>Int</b> + $(\neg q \rightarrow p) \rightarrow (((p \rightarrow q) \rightarrow p) \rightarrow p)$
<b>KC</b>	= <b>Int</b> + $\neg p \vee \neg \neg p$
<b>LC</b>	= <b>Int</b> + $(p \rightarrow q) \vee (q \rightarrow p)$
<b>SL</b>	= <b>Int</b> + $((\neg \neg p \rightarrow p) \rightarrow \neg p \vee p) \rightarrow \neg p \vee \neg \neg p$
<b>KP</b>	= <b>Int</b> + $(\neg p \rightarrow q \vee r) \rightarrow (\neg p \rightarrow q) \vee (\neg p \rightarrow r)$
<b>BD<sub>n</sub></b>	= <b>Int</b> + $\mathbf{bd}_n$ , where $\mathbf{bd}_1 = p_1 \vee \neg p_1$ , $\mathbf{bd}_{n+1} = p_{n+1} \vee (p_{n+1} \rightarrow \mathbf{bd}_n)$
<b>BW<sub>n</sub></b>	= <b>Int</b> + $\bigvee_{i=0}^n (p_i \rightarrow \bigvee_{j \neq i} p_j)$
<b>BTW<sub>n</sub></b>	= <b>Int</b> + $\bigwedge_{0 \leq i < j \leq n} \neg(\neg p_i \wedge \neg p_j) \rightarrow \bigvee_{i=0}^n (\neg p_i \rightarrow \bigvee_{j \neq i} \neg p_j)$
<b>T<sub>n</sub></b>	= <b>Int</b> + $\bigwedge_{i=0}^n ((p_i \rightarrow \bigvee_{i \neq j} p_j) \rightarrow \bigvee_{i \neq j} p_j) \rightarrow \bigvee_{i=0}^n p_i$
<b>B<sub>n</sub></b>	= <b>Int</b> + $\bigwedge_{i=0}^n (\neg p_i \leftrightarrow \bigvee_{i \neq j} p_j) \rightarrow \bigvee_{i=0}^n p_i$
<b>NL<sub>n</sub></b>	= <b>Int</b> + $\mathbf{nf}_n$ , where $\mathbf{nf}_0 = \perp$ , $\mathbf{nf}_1 = p$ , $\mathbf{nf}_2 = \neg p$ , $\mathbf{nf}_\omega = \top$ $\mathbf{nf}_{2m+3} = \mathbf{nf}_{2m+1} \vee \mathbf{nf}_{2m+2}$ , $\mathbf{nf}_{2m+4} = \mathbf{nf}_{2m+3} \rightarrow \mathbf{nf}_{2m+1}$

---

$\mathfrak{A} = \langle A, \wedge, \vee, \rightarrow, \perp \rangle$ , we construct its relational representation  $\mathfrak{A}_+ = \langle W, R \rangle$  by taking  $W$  to be the set of all prime filters in  $\mathfrak{A}$  (a filter  $\nabla$  is *prime* if it is proper and  $a \vee b \in \nabla$  implies  $a \in \nabla$  or  $b \in \nabla$ ),  $R$  to be the set-theoretic inclusion  $\subseteq$  and

$$P = \{ \{ \nabla \in W : a \in \nabla \} : a \in A \}.$$

It is readily checked that  $\mathfrak{A}_+$ , the *dual* of  $\mathfrak{A}$ , is an intuitionistic frame,  $\mathfrak{A} \cong (\mathfrak{A}_+)^+$  and  $\mathfrak{A}_+$  is differentiated, *tight* in the sense that

$$xRy \text{ iff } \forall X \in P (x \in X \rightarrow y \in X),$$

and *compact*, i.e., for any families  $\mathcal{X} \subseteq P$  and  $\mathcal{Y} \subseteq \{W - X : X \in P\}$ ,

$$\bigcap (\mathcal{X} \cup \mathcal{Y}) = \{x \in W : \forall X \in \mathcal{X} \forall Y \in \mathcal{Y} (x \in X \wedge x \in Y)\} \neq \emptyset$$

whenever  $\bigcap (\mathcal{X}' \cup \mathcal{Y}') \neq \emptyset$  for every finite subfamilies  $\mathcal{X}' \subseteq \mathcal{X}$ ,  $\mathcal{Y}' \subseteq \mathcal{Y}$ . Frames with these three properties (actually differentiatedness follows from tightness) are called *descriptive*. In the same way as in the modal case one can prove that  $\mathfrak{F}$  is descriptive iff  $\mathfrak{F} \cong (\mathfrak{F}^+)_+$ . Duality between the

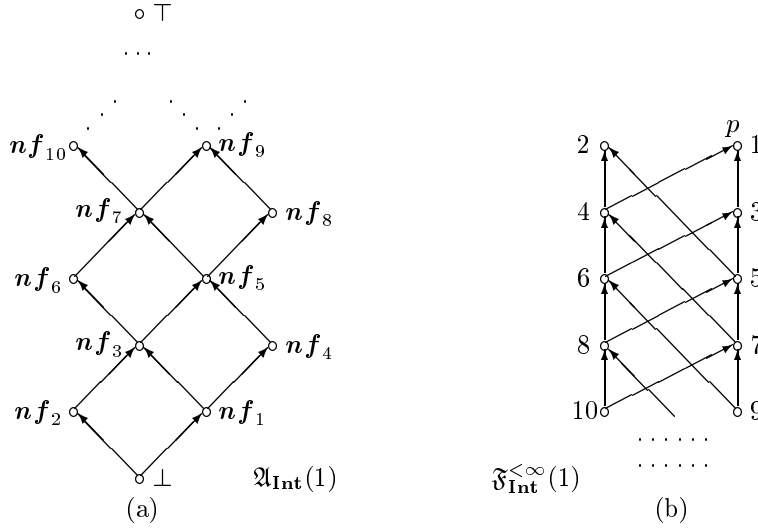


Figure 13.

basic truth-preserving operations on algebras and descriptive frames (the definitions of generated subframes, reductions and disjoint unions do not change) is also established by the same technique.

Since every consistent si-logic  $L$  is characterized by its Tarski–Lindenbaum algebra  $\mathfrak{A}_L$ , we conclude that  $L$  is characterized also by a class of intuitionistic frames, say by the dual of  $\mathfrak{A}_L$ .

Refined finitely generated frames for **Int** look similarly to those for **K4**: the only difference is that now all clusters are simple and the truth-sets must be upward closed. Fig. 13 showing (a) the free 1-generated Heyting algebra  $\mathfrak{A}_{\mathbf{Int}}(1)$  and (b) its dual  $\mathfrak{F}_{\mathbf{Int}}^{\leq \infty}(1)$  will help the reader to restore the details.  $\mathfrak{A}_{\mathbf{Int}}(1)$  was first constructed by Rieger [1949] and Nishimura [1960]; it is called the *Rieger–Nishimura lattice*. The formulas  $nf_n$  defined in Table 5 and used for the construction are known as *Nishimura formulas* (see also Section 3 of *Intuitionistic Logic*, in volume 7 of this *Handbook*).

At the algebraic level the connection between **Int** and **S4** discovered by Gödel is reflected by the fact, established in [Mckinsey and Tarski 1946], that the algebra of open elements (i.e., elements  $a$  such that  $\Box a = a$ ) of every modal algebra for **S4** (known as a *topological Boolean algebra*; see [Rasiowa and Sikorski 1963]) is a Heyting algebra and conversely, every Heyting algebra is isomorphic to the algebra of open elements of a suitable algebra for **S4**. We explain this result in the frame-theoretic language.

Given a frame  $\mathfrak{F} = \langle W, R, P \rangle$  for **S4** (which means that  $R$  is a quasi-order on  $W$ ), we denote by  $\rho W$  the set of clusters in  $\mathfrak{F}$ —more generally,

$\rho X = \{C(x) : x \in X\}$ —and put  $C(x)\rho C(y)$  iff  $xRy$ ,

$$\rho P = \{\rho X : X \in P \wedge X = \Box X\} = \{\rho X : X \in P \wedge X = X\uparrow\}.$$

It is readily checked that the structure  $\rho\mathfrak{F} = \langle \rho W, \rho R, \rho P \rangle$  is an intuitionistic frame (for instance,  $\rho(X) \supset \rho(Y) = \rho(\Box(-X \cup Y))$ ); we call it the *skeleton* of  $\mathfrak{F}$ . The *skeleton* of a model  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  for **S4** is the intuitionistic model  $\rho\mathfrak{M} = \langle \rho\mathfrak{F}, \rho\mathfrak{V} \rangle$ , where  $\rho\mathfrak{V}(p) = \mathfrak{V}(\Box p)$ .

Denote by  $T$  the *Gödel translation* prefixing  $\Box$  to all subformulas of a given intuitionistic formula.<sup>17</sup> By induction on the construction of  $\varphi$  one can easily prove the following

LEMMA 153 (Skeleton). *For every model  $\mathfrak{M}$  for **S4**, every intuitionistic formula  $\varphi$  and every point  $x$  in  $\mathfrak{M}$ ,*

$$(\rho\mathfrak{M}, C(x)) \models \varphi \text{ iff } (\mathfrak{M}, x) \models T(\varphi).$$

It follows that  $\varphi \in \mathbf{Int}$  implies  $T(\varphi) \in \mathbf{S4}$ . To prove the converse we should be able to convert intuitionistic frames  $\mathfrak{F}$  into modal ones with the skeleton (isomorphic to)  $\mathfrak{F}$ . This is trivial if  $\mathfrak{F}$  is a Kripke frame—we can just regard it to be a frame for **S4**, which in view of the Kripke completeness of both **Int** and **S4**, shows that  $T$  really embeds the former into the latter, i.e.,

$$\varphi \in \mathbf{Int} \text{ iff } T(\varphi) \in \mathbf{S4}.$$

In general, the most obvious way of constructing a modal frame from an intuitionistic frame  $\mathfrak{F} = \langle W, R, P \rangle$  is to take the closure  $\sigma P$  of  $P$  under the Boolean operations  $\cap$ ,  $\cup$  and  $\rightarrow$ . It is well known in the theory of Boolean algebras (see [Rasiowa and Sikorski 1963]) that for every  $X \subseteq W$ ,  $X$  is in  $\sigma P$  iff

$$X = (-X_1 \cup Y_1) \cap \cdots \cap (-X_n \cup Y_n)$$

for some  $X_1, Y_1, \dots, X_n, Y_n \in P$  and  $n \geq 1$ . It follows that if  $X \in \sigma P$  then

$$\Box X = (X_1 \supset Y_1) \cap \cdots \cap (X_n \supset Y_n) \in P \subseteq \sigma P,$$

and so  $\sigma P$  is closed under  $\Box$  in  $\langle W, R \rangle$  and  $P$  coincides with the set of upward closed sets in  $\sigma P$ . Thus,  $\langle W, R, \sigma P \rangle$  is a partially ordered modal frame; we shall denote it by  $\sigma\mathfrak{F}$ . Moreover, we clearly have  $\mathfrak{F} \cong \rho\sigma\mathfrak{F}$ . If  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  is an intuitionistic model then  $\sigma\mathfrak{M} = \langle \sigma\mathfrak{F}, \mathfrak{V} \rangle$  is a modal model having  $\mathfrak{M}$  as its skeleton. So by the Skeleton Lemma,

$$(\mathfrak{M}, x) \models \varphi \text{ iff } (\sigma\mathfrak{M}, x) \models T(\varphi),$$

<sup>17</sup>The translation defined in [Gödel 1933] does not prefix  $\Box$  to conjunctions and disjunctions. However this difference is of no importance as far as embeddings into logics in **NExtS4** are concerned.

for every intuitionistic formula  $\varphi$  and every point  $x$  in  $\mathfrak{F}$ .

It is worth noting that if  $\mathfrak{F} = \langle W, R \rangle$  is a finite intuitionistic Kripke frame then  $\sigma\mathfrak{F}$  is also a Kripke frame. However, for an infinite  $\mathfrak{F}$ ,  $\sigma\mathfrak{F}$  is not in general a Kripke frame, witness  $\langle \omega, \leq \rangle$ .

The operator  $\sigma$  is not the only one which, given an intuitionistic frame  $\mathfrak{F}$ , returns a modal frame whose skeleton is isomorphic to  $\mathfrak{F}$ . As an example, we define now an infinite class of such operators. For Kripke frames  $\mathfrak{F} = \langle W, R \rangle$  and  $\mathfrak{G} = \langle V, S \rangle$ , denote by  $\mathfrak{F} \times \mathfrak{G}$  the *direct product* of  $\mathfrak{F}$  and  $\mathfrak{G}$ , i.e., the frame  $\langle W \times V, R \times S \rangle$  in which the relation  $R \times S$  is defined component-wise:

$$\langle x_1, y_1 \rangle (R \times S) \langle x_2, y_2 \rangle \text{ iff } x_1 R x_2 \text{ and } y_1 S y_2.$$

Let  $0 < k \leq \omega$ . We will regard  $k$  to be the set  $\{0, \dots, k-1\}$  if  $k < \omega$  and  $\{0, 1, \dots\}$  if  $k = \omega$ . Denote by  $\tau_k$  an operator which, given an intuitionistic frame  $\mathfrak{F} = \langle W, R, P \rangle$ , returns a modal frame  $\tau_k\mathfrak{F} = \langle kW, kR, kP \rangle$  such that

(i)  $\langle kW, kR \rangle$  is the direct product of the  $k$ -point cluster  $\langle k, k^2 \rangle$  and  $\langle W, R \rangle$  (in other words,  $\langle kW, kR \rangle$  is obtained from  $\langle W, R \rangle$  by replacing its every point with a  $k$ -point cluster);

(ii)  $\rho\tau_k\mathfrak{F} \cong \mathfrak{F}$ ;

(iii)  $I \times X \in kP$ , for every  $I \subseteq k$  and  $X \in \sigma P$ .

For instance, we can take  $kP$  to be the Boolean closure of the set

$$\{I \times X : I \subseteq k, X \in \sigma P\}.$$

For a Kripke frame  $\mathfrak{F} = \langle W, R, \text{Up}W \rangle$  we can, of course, take  $kP = 2^{kW}$  and then  $\tau_k\mathfrak{F} = \langle kW, kR, 2^{kW} \rangle$ .

### 3.2 Canonical formulas

The language of canonical formulas, axiomatizing all si-logics and characterizing the structure of their frames, can be easily developed following the scheme of constructing the canonical formulas for **K4** outlined in Section 1.6 and using the connection between modal and intuitionistic frames established above. We confine ourselves here only to pointing out the differences from the modal case and some interesting peculiarities; details can be found in [Zakharyashev 1983, 1989] and [Chagrova and Zakharyashev 1997].

Actually, there are two important differences. First, in the definition of subreduction of  $\mathfrak{F} = \langle W, R, P \rangle$  to  $\mathfrak{G}$  the condition (R3) does not correspond to the fact that all sets in  $P$  are upward closed. We replace it by the following condition

$$(R3') \quad \forall X \in \overline{Q} \ f^{-1}(X) \downarrow \in \overline{P},$$

where  $\overline{Q} = \{V - X : X \in Q\}$  and  $\overline{P} = \{W - X : X \in P\}$ . For a completely defined  $f$  satisfying (R1) and (R2) the condition (R3') is clearly

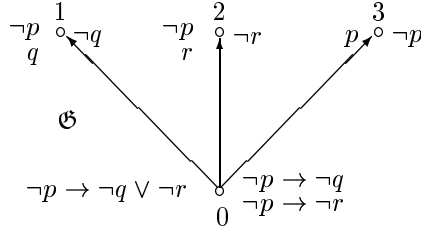


Figure 14.

equivalent to (R3) and so every reduction is also a subreduction. If  $\mathfrak{G}$  is a finite Kripke frame then (R3') is equivalent to  $\forall z \in V f^{-1}(z) \downarrow \in \overline{P}$ .  $\mathfrak{G}$  is a *subframe* of  $\mathfrak{F}$  if  $\kappa\mathfrak{G}$  is a subframe of  $\kappa\mathfrak{F}$  and the identity map on  $V$  is a subreduction of  $\mathfrak{F}$  to  $\mathfrak{G}$ . It is of interest to note that in the intuitionistic case (cofinal) subreductions are dual to IC(N)-subalgebras of Heyting algebras which preserve only implication, conjunction (and negation or  $\perp$ ) but do not necessarily preserve disjunction.

Second, we have to change the definition of open domains. Now we say an antichain  $\mathfrak{a}$  (of at least two points) is an *open domain* in an intuitionistic model  $\mathfrak{M}$  relative to a formula  $\varphi$  if there is a pair  $t_{\mathfrak{a}} = (\Gamma_{\mathfrak{a}}, \Delta_{\mathfrak{a}})$  such that  $\Gamma_{\mathfrak{a}} \cup \Delta_{\mathfrak{a}} = \mathbf{Sub}\varphi$ ,  $\bigwedge \Gamma_{\mathfrak{a}} \rightarrow \bigvee \Delta_{\mathfrak{a}} \notin \mathbf{Int}$  and

- $\psi \in \Gamma_{\mathfrak{a}}$  iff  $a \models \psi$  for all  $a \in \mathfrak{a}$ .

It is worth noting that in any intuitionistic model every antichain  $\mathfrak{a}$  is open relative to every disjunction free formula  $\varphi$ . Indeed, let  $\Gamma_{\mathfrak{a}}$  be defined by condition above and  $\Delta_{\mathfrak{a}} = \mathbf{Sub}\varphi - \Gamma_{\mathfrak{a}}$ . It should be clear that  $\psi \wedge \chi \in \Gamma_{\mathfrak{a}}$  iff  $\psi \in \Gamma_{\mathfrak{a}}$  and  $\chi \in \Gamma_{\mathfrak{a}}$ . And if  $\psi \rightarrow \chi \in \Gamma_{\mathfrak{a}}$ ,  $\psi \in \Gamma_{\mathfrak{a}}$  but  $\chi \in \Delta_{\mathfrak{a}}$  then  $a \models \psi$  for every  $a \in \mathfrak{a}$  and  $b \not\models \chi$  for some  $b \in \mathfrak{a}$ , whence  $b \not\models \psi \rightarrow \chi$ , which is a contradiction. It follows that  $\bigwedge \Gamma_{\mathfrak{a}} \rightarrow \bigvee \Delta_{\mathfrak{a}} \notin \mathbf{Int}$ .

EXAMPLE 154. Let us try to characterize the class of intuitionistic refutation frames for the *Weak Kreisel–Putnam Formula*

$$\mathbf{wkp} = (\neg p \rightarrow \neg q \vee \neg r) \rightarrow (\neg p \rightarrow \neg q) \vee (\neg p \rightarrow \neg r).$$

First we construct its simplest countermodel; it is depicted in Fig. 14, where by putting a formula to the left (right) of a point we mean that it is true (not true) at the point. Then we observe that every frame  $\mathfrak{F}$  refuting  $\mathbf{wkp}$  is cofinally subreducible to the frame  $\mathfrak{G}$  underlying this countermodel by

the map  $f$  defined as follows:

$$f(x) = \begin{cases} 0 & \text{if } x \models \neg p \rightarrow \neg q \vee \neg r, x \not\models (\neg p \rightarrow \neg q) \vee (\neg p \rightarrow \neg r) \\ 1 & \text{if } x \models \neg p \rightarrow \neg q \vee \neg r, x \models \neg p \text{ and } x \models q \\ 2 & \text{if } x \models \neg p \rightarrow \neg q \vee \neg r, x \models \neg p \text{ and } x \models r \\ 3 & \text{if } x \models p \text{ or } x \models \neg p \wedge \neg q \wedge \neg r \\ \text{undefined} & \text{otherwise.} \end{cases}$$

However, the cofinal subreducibility to  $\mathfrak{G}$  is only a necessary condition for  $\mathfrak{F} \not\models \mathbf{wkp}$ , witness the frame having the form of the three-dimensional Boolean cube with the top point deleted. The reason for this is that the antichain  $\{1, 2\}$  is a closed domain in  $\mathfrak{N}$ : it is impossible to insert a point  $a$  between 0 and  $\{1, 2\}$  and extend to it consistently the truth-sets for the depicted formulas. Indeed, otherwise we would have  $a \models \neg p \rightarrow \neg q \vee \neg r$ ,  $a \not\models \neg q \vee \neg r$  and so  $a \not\models \neg p$ , i.e., there must be a point  $x \in a \uparrow$  such that  $x \models p$ , but such a point does not exist. In fact,  $\mathfrak{F} \not\models \mathbf{wkp}$  iff there is a cofinal subreduction of  $\mathfrak{F}$  to  $\mathfrak{G}$  satisfying (CDC) for  $\{\{1, 2\}\}$ .

Now, as in the modal case, with every finite rooted intuitionistic frame  $\mathfrak{F} = \langle W, R \rangle$  and a set  $\mathfrak{D}$  of antichains in it we can associate two formulas  $\beta(\mathfrak{F}, \mathfrak{D}, \perp)$  and  $\beta(\mathfrak{F}, \mathfrak{D})$ , called the *canonical* and *negation free canonical formulas*, respectively, so that  $\mathfrak{G} \not\models \beta(\mathfrak{F}, \mathfrak{D}, \perp)$  ( $\mathfrak{G} \not\models \beta(\mathfrak{F}, \mathfrak{D})$ ) iff there is a (cofinal) subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ . For instance, if  $a_0, \dots, a_n$  are all points in  $\mathfrak{F}$  and  $a_0$  is its root, then one can take

$$\beta(\mathfrak{F}, \mathfrak{D}, \perp) = \bigwedge_{a_i R a_j} \psi_{ij} \wedge \bigwedge_{\mathfrak{d} \in \mathfrak{D}} \psi_{\mathfrak{d}} \wedge \psi_{\perp} \rightarrow p_0,$$

where

$$\begin{aligned} \psi_{ij} &= \left( \bigwedge_{\neg a_j R a_k} p_k \rightarrow p_j \right) \rightarrow p_i, \\ \psi_{\mathfrak{d}} &= \bigwedge_{a_i \in W - \mathfrak{d} \uparrow} \left( \bigwedge_{\neg a_i R a_k} p_k \rightarrow p_i \right) \rightarrow \bigvee_{a_j \in \mathfrak{d}} p_j, \\ \psi_{\perp} &= \bigwedge_{i=0}^n \left( \bigwedge_{\neg a_i R a_k} p_k \rightarrow p_i \right) \rightarrow \perp. \end{aligned}$$

$\beta(\mathfrak{F}, \mathfrak{D})$  is obtained from  $\beta(\mathfrak{F}, \mathfrak{D}, \perp)$  by deleting the conjunct  $\psi_{\perp}$ .

**THEOREM 155.** *There is an algorithm which, given an intuitionistic  $\varphi$ , returns canonical formulas  $\beta(\mathfrak{F}_1, \mathfrak{D}_1, \perp), \dots, \beta(\mathfrak{F}_n, \mathfrak{D}_n, \perp)$  such that*

$$\mathbf{Int} + \varphi = \mathbf{Int} + \beta(\mathfrak{F}_1, \mathfrak{D}_1, \perp) + \dots + \beta(\mathfrak{F}_n, \mathfrak{D}_n, \perp).$$

*So the set of intuitionistic canonical formulas is complete for  $\mathbf{ExtInt}$ . If  $\varphi$  is negation free then one can use only negation free canonical formulas. And if  $\varphi$  is disjunction free then all  $\mathfrak{D}_i$  are empty.*



Table 6 and Theorem 156 show canonical axiomatizations of the si-logics in Table 5. Using this “geometrical” representation it is not hard to see, for instance, that **SmL**, known as the *Smetanich logic*, is the greatest consistent extension of **Int** different from **Cl**; it is the logic of the two-point rooted frame. **KC**, the logic of the *Weak Law of the Excluded Middle*, is characterized by the class of directed frames. It is the greatest si-logic containing the same negation free formulas as **Int** (see [Jankov 1968a]). **LC**, the *Dummett* or *chain logic*, is characterized by the class of linear frames (see [Dummett 1959]). **BD<sub>n</sub>** and **BW<sub>n</sub>** are the minimal logics of depth  $n$  and width  $n$ , respectively (see [Hosoi 1967] and [Smoryński 1973]). Finite frames for **BTW<sub>n</sub>** contain  $\leq n$  top points [Smoryński 1973] and finite frames for **T<sub>n</sub>** are of branching  $\leq n$ , i.e., no point has more than  $n$  immediate successors. THEOREM 156 (Nishimura 1960, Anderson 1972). *Every extension  $L$  of **Int** by formulas in one variable can be represented either as*

$$L = \mathbf{Int} + \mathbf{nf}_{2n} = \mathbf{Int} + \beta^\sharp(\mathfrak{H}_n, \perp)$$

or as

$$L = \mathbf{Int} + \mathbf{nf}_{2n-1} = \mathbf{Int} + \beta^\sharp(\mathfrak{H}_{n+1}, \perp) + \beta^\sharp(\mathfrak{H}_{n+2}, \perp),$$

where  $\mathfrak{H}_n, \mathfrak{H}_{n+1}, \mathfrak{H}_{n+2}$  are the subframes of the frame in Fig. 13 generated by the points  $n, n+1$  and  $n+2$ , respectively, and  $\beta^\sharp(\mathfrak{F}, \perp)$  is an abbreviation for  $\beta(\mathfrak{F}, \mathfrak{D}^\sharp, \perp)$ ,  $\mathfrak{D}^\sharp$  the set of all antichains in  $\mathfrak{F}$ .

Jankov [1969] proved in fact that logics of the form  $\mathbf{Int} + \beta^\sharp(\mathfrak{F}, \perp)$  and only them are splittings of **ExtInt**. However, not every si-logic is a union-splitting of **ExtInt** which means that this class has no axiomatic basis.

### 3.3 Modal companions and preservation theorems

The fact that the Gödel translation  $T$  embeds **Int** into **S4** and the relationship between intuitionistic and modal frames established in Section 3.1 can be used to reduce various problems concerning **Int** (e.g. proving completeness or FMP) to those for **S4** and vice versa. Moreover, it turns out that each logic in **ExtInt** is embedded by  $T$  into some logics in **NExtS4**, and for each logic in **NExtS4** there is one in **ExtInt** embeddable in it.

We say a modal logic  $M \in \mathbf{NExtS4}$  is a *modal companion* of a si-logic  $L$  if  $L$  is embedded in  $M$  by  $T$ , i.e., if for every intuitionistic formula  $\varphi$ ,

$$\varphi \in L \text{ iff } T(\varphi) \in M.$$

If  $M$  is a modal companion of  $L$  then  $L$  is called the *si-fragment* of  $M$  and denoted by  $\rho M$ . The reason for denoting the operator “modal logic  $\mapsto$  its si-fragment” by the same symbol we used for the skeleton operator is explained by the following

Table 6. Canonical axioms of standard superintuitionistic logics

---

<b>For</b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\circ)$	
<b>CI</b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\circ \uparrow \circ)$	
<b>SmL</b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\circ \swarrow \circ \searrow \circ) + \beta(\circ \uparrow \circ)$	
<b>KC</b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\circ \swarrow \circ \searrow \circ, \perp)$	
<b>LC</b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\circ \swarrow \circ \searrow \circ)$	
<b>SL</b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta^\sharp(\circ \swarrow \circ \searrow \circ, \perp)$	
<b>KP</b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\circ \swarrow \circ \searrow \circ, \{\{1, 2\}\}, \perp) + \beta(\circ \swarrow \circ \searrow \circ, \{\{1, 2\}\}, \perp)$	
<b>BD<sub>n</sub></b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\circ \uparrow \circ \uparrow \dots \uparrow \circ, 0)$	
<b>BW<sub>n</sub></b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\overbrace{\circ \dots \circ}^{n+1} \swarrow \circ \searrow \circ)$	
<b>BTW<sub>n</sub></b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta(\overbrace{\circ \dots \circ}^{n+1} \swarrow \circ \searrow \circ, \perp)$	
<b>T<sub>n</sub></b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta^\sharp(\overbrace{\circ \dots \circ}^{n+1} \swarrow \circ \searrow \circ)$	
<b>B<sub>n</sub></b>	<b>=</b>	<b>Int</b>	<b>+</b>	$\beta^\sharp(\overbrace{\circ \dots \circ}^{n+1} \swarrow \circ \searrow \circ, \perp)$	

---

**THEOREM 157.** *For every  $M \in \text{NExtS4}$ ,  $\rho M = \{\varphi : T(\varphi) \in M\}$ . Moreover, if  $M$  is characterized by a class  $\mathcal{C}$  of modal frames then  $\rho M$  is characterized by the class  $\rho\mathcal{C} = \{\rho\mathfrak{F} : \mathfrak{F} \in \mathcal{C}\}$  of intuitionistic frames.*

**Proof.** It suffices to show that  $\{\varphi : T(\varphi) \in M\} = \text{Log}\rho\mathcal{C}$ . Suppose that  $T(\varphi) \in M$ . Then  $\mathfrak{F} \models T(\varphi)$  and so, by the Skeleton Lemma,  $\rho\mathfrak{F} \models \varphi$  for every  $\mathfrak{F} \in \mathcal{C}$ , i.e.,  $\varphi \in \text{Log}\rho\mathcal{C}$ . Conversely, if  $\rho\mathfrak{F} \models \varphi$  for all  $\mathfrak{F} \in \mathcal{C}$  then, by the same lemma,  $T(\varphi)$  is valid in all frames in  $\mathcal{C}$  and so  $T(\varphi) \in M$ . ■

Thus,  $\rho$  maps  $\text{NExtS4}$  into  $\text{ExtInt}$ . The following simple observation shows that actually  $\rho$  is a surjection. Given a logic  $L \in \text{ExtInt}$ , we put

$$\tau L = \mathbf{S4} \oplus \{T(\varphi) : \varphi \in L\}.$$

**THEOREM 158** (Dummett and Lemmon 1959). *For every si-logic  $L$ ,  $\tau L$  is a modal companion of  $L$ .*

**Proof.** Clearly,  $L \subseteq \rho\tau L$ . To prove the converse inclusion, suppose  $\varphi \notin L$ , i.e., there is a frame  $\mathfrak{F}$  for  $L$  refuting  $\varphi$ . Since  $\mathfrak{F} \cong \rho\sigma\mathfrak{F}$ , by the Skeleton Lemma we have  $\sigma\mathfrak{F} \models \tau L$  and  $\sigma\mathfrak{F} \not\models T(\varphi)$ . Therefore,  $T(\varphi) \notin \tau L$  and so  $\varphi \notin \rho\tau L$ . ■

Now we use the language of canonical formulas to obtain a general characterization of all modal companions of a given si-logic  $L$ . Our presentation follows [Zakharyashev 1989, 1991]. Notice first that for every modal frame  $\mathfrak{G}$  and every intuitionistic canonical formula  $\beta(\mathfrak{F}, \mathfrak{D}, \perp)$ ,  $\mathfrak{G} \models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  iff  $\rho\mathfrak{G} \models \beta(\mathfrak{F}, \mathfrak{D}, \perp)$  and so  $\mathbf{S4} \oplus T(\beta(\mathfrak{F}, \mathfrak{D}, \perp)) = \mathbf{S4} \oplus \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ . The same concern, of course, the negation free canonical formulas.

**THEOREM 159.** *A logic  $M \in \text{NExtS4}$  is a modal companion of a si-logic  $L = \text{Int} + \{\beta(\mathfrak{F}_i, \mathfrak{D}_i, \perp) : i \in I\}$  iff  $M$  can be represented in the form*

$$M = \mathbf{S4} \oplus \{\alpha(\mathfrak{F}_i, \mathfrak{D}_i, \perp) : i \in I\} \oplus \{\alpha(\mathfrak{F}_j, \mathfrak{D}_j, \perp) : j \in J\},$$

where every frame  $\mathfrak{F}_j$ , for  $j \in J$ , contains a proper cluster.

**Proof.** ( $\Leftarrow$ ) We must show that for every intuitionistic formula  $\varphi$ ,  $\varphi \in L$  iff  $T(\varphi) \in M$ . Suppose that  $\varphi \notin L$  and  $\mathfrak{F} = \langle W, R, P \rangle$  is a frame separating  $\varphi$  from  $L$ . We prove that  $\sigma\mathfrak{F}$  separates  $T(\varphi)$  from  $M$ . As was observed above,  $\sigma\mathfrak{F} \not\models T(\varphi)$  and  $\sigma\mathfrak{F} \models \alpha(\mathfrak{F}_i, \mathfrak{D}_i, \perp)$  for any  $i \in I$ . So it remains to show that  $\sigma\mathfrak{F} \models \alpha(\mathfrak{F}_j, \mathfrak{D}_j, \perp)$  for every  $j \in J$ .

Suppose otherwise. Then, for some  $j \in J$ , we have a subreduction  $f$  of  $\sigma\mathfrak{F}$  to  $\mathfrak{F}_j$ . Let  $a_1$  and  $a_2$  be distinct points belonging to the same proper cluster in  $\mathfrak{F}_j$ . By the definition of subreduction,  $f^{-1}(a_1) \subseteq f^{-1}(a_2)\downarrow$  and  $f^{-1}(a_2) \subseteq f^{-1}(a_1)\downarrow$ , and so there is an infinite chain  $x_1 R y_1 R x_2 R y_2 R \dots$

in  $\sigma\mathfrak{F}$  such that  $\{x_1, x_2, \dots\} \subseteq f^{-1}(a_1)$  and  $\{y_1, y_2, \dots\} \subseteq f^{-1}(a_2)$ . And since  $R$  is a partial order, all the points  $x_i$  and  $y_i$  are distinct.

Since  $f^{-1}(a_1) \in \sigma P$ , there are  $X_i, Y_i \in P$  such that

$$f^{-1}(a_1) = (-X_1 \cup Y_1) \cap \dots \cap (-X_n \cup Y_n).$$

And since  $f^{-1}(a_1) \cap f^{-1}(a_2) = \emptyset$ , for every point  $y_i$  there is some number  $n_i$  such that  $y_i \in X_{n_i}$  and  $y_i \notin Y_{n_i}$ . But then, for some distinct  $l$  and  $m$ , the numbers  $n_l$  and  $n_m$  must coincide, and so if, say,  $y_l R y_m$  then  $x_m \notin Y_{n_m}$  and  $x_m \in X_{n_l}$  (for  $y_l R x_m R y_m$ ,  $X_i = X_i \uparrow$ ,  $Y_i = Y_i \uparrow$ ). Therefore,  $x_m \notin f^{-1}(a_1)$ , which is a contradiction.

The rest of the proof presents no difficulties. ■

This proof does not touch upon the cofinality condition. So along with canonical formulas in Theorem 159 we can use negation free canonical formulas. Thus, we have:

$$\rho\mathbf{S4} = \rho\mathbf{S4.1} = \rho\mathbf{Dum} = \rho\mathbf{Grz} = \mathbf{Int},$$

$$\rho\mathbf{S4.2} = \rho(\mathbf{S4.2} \oplus \mathbf{Grz}) = \mathbf{KC},$$

$$\rho\mathbf{S4.3} = \rho(\mathbf{S4.3} \oplus \mathbf{Grz}) = \mathbf{LC},$$

$$\rho\mathbf{S5} = \rho(\mathbf{S5} \oplus \mathbf{Grz}) = \mathbf{Cl}.$$

**COROLLARY 160.** *The set of modal companions of every consistent si-logic  $L$  forms the interval*

$$\rho^{-1}(L) = [\tau L, \tau L \oplus \alpha(\textcircled{\circ})] = \{M \in \mathbf{NExtS4} : \tau L \subseteq M \subseteq \tau L \oplus \mathbf{Grz}\}$$

*and contains an infinite descending chain of logics.*

**Proof.** Notice first that  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  and  $\alpha(\mathfrak{F}, \mathfrak{D})$  are in  $\mathbf{Grz}$  iff  $\mathfrak{F}$  contains a proper cluster. So  $\rho^{-1}(L) \subseteq [\tau L, \tau L \oplus \alpha(\textcircled{\circ})]$ . On the other hand, the si-fragments of all logics in the interval are the same, namely  $L$ . Therefore,  $\rho^{-1}(L) = [\tau L, \tau L \oplus \alpha(\textcircled{\circ})]$ . Now, if  $L$  is consistent then  $\beta(\circ) \notin L$  and so we have

$$\tau L \subset \dots \subset \tau L \oplus \alpha(\mathfrak{C}_n) \subset \dots \subset \tau L \oplus \alpha(\mathfrak{C}_2) \subset \tau L \oplus \alpha(\mathfrak{C}_1) = \mathbf{For},$$

where  $\mathfrak{C}_i$  is the non-degenerate cluster with  $i$  points. ■

This result is due to Maksimova and Rybakov [1974], Blok [1976] and Esakia [1979b].

Thus, all modal companions of every si-logic  $L$  are contained between the least companion  $\tau L$  and the greatest one, viz.,  $\tau L \oplus \alpha(\textcircled{\circ})$ , which will be denoted by  $\sigma L$ . Using Theorems 159 and 44, we obtain

**COROLLARY 161.** *There is an algorithm which, given a modal formula  $\varphi$ , returns an intuitionistic formula  $\psi$  such that  $\rho(\mathbf{S4} \oplus \varphi) = \mathbf{Int} + \psi$ .*

The following theorem, which is also a consequence of Theorem 159, describes lattice-theoretic properties of the maps  $\rho$ ,  $\tau$  and  $\sigma$ . Items (i), (ii) and (iv) in it were first proved by Maksimova and Rybakov [1974], and (iii) is due to Blok [1976] and Esakia [1979b] and known as the Blok–Esakia Theorem.

**THEOREM 162.**

- (i) *The map  $\rho$  is a homomorphism of the lattice  $\mathbf{NExtS4}$  onto the lattice  $\mathbf{ExtInt}$ .*
- (ii) *The map  $\tau$  is an isomorphism of  $\mathbf{ExtInt}$  into  $\mathbf{NExtS4}$ .*
- (iii) *The map  $\sigma$  is an isomorphism of  $\mathbf{ExtInt}$  onto  $\mathbf{NExtGrz}$ .*
- (iv) *All these maps preserve infinite sums and intersections of logics.*

Now we give frame-theoretic characterizations of the operators  $\tau$  and  $\sigma$ . Note first that the following evident relations between frames for si-logics and their modal companions hold:

$$\begin{aligned} \mathfrak{F} \models \rho M \text{ iff } \sigma \mathfrak{F} \models M, \quad \mathfrak{F} \models L \text{ iff } \sigma \mathfrak{F} \models \sigma L, \\ \rho \mathfrak{F} \models L \text{ iff } \mathfrak{F} \models \tau L, \quad \mathfrak{F} \models L \text{ iff } \tau_k \mathfrak{F} \models \tau L. \end{aligned}$$

**THEOREM 163** (Maksimova and Rybakov 1974). *A si-logic  $L$  is characterized by a class  $\mathcal{C}$  of intuitionistic frames iff  $\sigma L$  is characterized by the class  $\sigma \mathcal{C} = \{\sigma \mathfrak{F} : \mathfrak{F} \in \mathcal{C}\}$ .*

**Proof.** ( $\Rightarrow$ ) It suffices to show that any canonical formula  $\alpha(\mathfrak{F}, \mathcal{D}, \perp) \notin \sigma L$  is refuted by some frame in  $\sigma \mathcal{C}$ . Since  $\mathfrak{F}$  is partially ordered,  $\beta(\mathfrak{F}, \mathcal{D}, \perp) \notin L$ , i.e., there is  $\mathfrak{F} \in \mathcal{C}$  refuting  $\beta(\mathfrak{F}, \mathcal{D}, \perp)$  and so  $\sigma \mathfrak{F} \not\models \alpha(\mathfrak{F}, \mathcal{D}, \perp)$ . ( $\Leftarrow$ ) is straightforward.  $\blacksquare$

To characterize  $\tau$  we require

**LEMMA 164.** *For any canonical formula  $\alpha(\mathfrak{F}, \mathcal{D}, \perp)$  built on a quasi-ordered frame  $\mathfrak{F}$ ,  $\alpha(\mathfrak{F}, \mathcal{D}, \perp) \in \mathbf{S4} \oplus \alpha(\rho \mathfrak{F}, \rho \mathcal{D}, \perp)$ , where  $\rho \mathcal{D} = \{\rho \mathfrak{d} : \mathfrak{d} \in \mathcal{D}\}$  and  $\rho \mathfrak{d} = \{C(x) : x \in \mathfrak{d}\}$ .*

**Proof.** Let  $\mathfrak{G}$  be a quasi-ordered frame refuting  $\alpha(\mathfrak{F}, \mathcal{D}, \perp)$ . Then there is a cofinal subreduction  $f$  of  $\mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathcal{D}$ . The map  $h$  from  $\mathfrak{F}$  onto  $\rho \mathfrak{F}$  defined by  $h(x) = C(x)$ , for every  $x$  in  $\mathfrak{F}$ , is clearly a reduction of  $\mathfrak{F}$  to  $\rho \mathfrak{F}$ . So the composition  $hf$  is a cofinal subreduction of  $\mathfrak{G}$  to  $\rho \mathfrak{F}$ , and it is easy to verify that it satisfies (CDC) for  $\rho \mathcal{D}$ .  $\blacksquare$

**THEOREM 165.** *A si-logic  $L$  is characterized by a class  $\mathcal{C}$  of frames iff  $\tau L$  is characterized by the class  $\bigcup_{0 < k < \omega} \tau_k \mathcal{C}$ , where  $\tau_k \mathcal{C} = \{\tau_k \mathfrak{F} : \mathfrak{F} \in \mathcal{C}\}$ .*

**Proof.** ( $\Rightarrow$ ) As was noted above, if  $\mathfrak{F}$  is a frame for  $L$  then  $\tau_k \mathfrak{F}$  is a frame for  $\tau L$ . So suppose that a formula  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ , built on a quasi-ordered frame  $\mathfrak{F} = \langle W, R \rangle$ , does not belong to  $\tau L$  and show that it is refuted by some frame in  $\bigcup_{0 < k < \omega} \tau_k \mathcal{C}$ . By Lemma 164,  $\alpha(\rho \mathfrak{F}, \rho \mathfrak{D}, \perp) \notin \tau L$  and so  $\beta(\rho \mathfrak{F}, \rho \mathfrak{D}, \perp) \notin L$ . Hence there is a frame  $\mathfrak{G} = \langle V, S, Q \rangle$  in  $\mathcal{C}$  which refutes  $\beta(\rho \mathfrak{F}, \rho \mathfrak{D}, \perp)$ . But then  $\sigma \mathfrak{G} \models \tau L$  and  $\sigma \mathfrak{G} \not\models \alpha(\rho \mathfrak{F}, \rho \mathfrak{D}, \perp)$ . Let  $f$  be a subreduction of  $\sigma \mathfrak{G}$  to  $\rho \mathfrak{F}$  satisfying (CDC) for  $\rho \mathfrak{D}$  and let  $k = \max\{|C(x)| : x \in W\}$ . Define a partial map  $h$  from  $\tau_k \mathfrak{G} = \langle kV, kS, kQ \rangle$  onto  $\mathfrak{F}$  as follows: if  $x \in V$ ,  $y_0 \in W$ ,  $f(x) = C(y_0)$  and  $C(y_0) = \{y_0, \dots, y_n\}$  then we put  $h(\langle i, x \rangle) = y_i$ , for  $i = 0, \dots, n$ . By the definition of  $\tau_k$ , for any  $i \in \{0, \dots, n\}$  we have

$$h^{-1}(y_i) = \{\langle i, x \rangle : x \in f^{-1}(C(y_0))\} = \{i\} \times f^{-1}(C(y_0)) \in kQ.$$

Now, one can readily prove that  $h$  is a cofinal subreduction of  $\tau_k \mathfrak{G}$  to  $\mathfrak{F}$  satisfying (CDC) for  $\mathfrak{D}$ . So  $\tau_k \mathfrak{G} \not\models \alpha(\mathfrak{F}, \mathfrak{D}, \perp)$ . ( $\Leftarrow$ ) is obvious. ■

It is worth noting that this proof will not change if we put in it  $k = \omega$ .

**COROLLARY 166.** *A logic  $L \in \text{ExtInt}$  is characterized by a class  $\mathcal{C}$  of frames iff  $\tau L$  is characterized by the class  $\tau_\omega \mathcal{C}$ .*

The following theorem provides a deductive characterization of the maps  $\tau$  and  $\sigma$ .

**THEOREM 167.** *For every si-logic  $L$  and every modal canonical formula  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp)$  built on a quasi-ordered frame  $\mathfrak{F}$ ,*

- (i)  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in \tau L$  iff  $\beta(\rho \mathfrak{F}, \rho \mathfrak{D}, \perp) \in L$ ;
- (ii)  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in \sigma L$  iff either  $\mathfrak{F}$  is partially ordered and  $\beta(\mathfrak{F}, \mathfrak{D}, \perp) \in L$  or  $\mathfrak{F}$  contains a proper cluster.

**Proof.** (i) The implication ( $\Rightarrow$ ) was actually established in the proof of Theorem 165, and the converse one follows from Lemma 164.

(ii) Suppose  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in \sigma L$ . Then either  $\mathfrak{F}$  is partially ordered, and so  $\beta(\mathfrak{F}, \mathfrak{D}, \perp) \in L$ , or  $\mathfrak{F}$  contains a proper cluster. The converse implication follows from (i) and the fact that  $\alpha(\mathfrak{F}, \mathfrak{D}, \perp) \in \mathbf{Grz}$  for every frame  $\mathfrak{F}$  with a proper cluster. ■

The results obtained in this section not only establish some structural correspondences between logics in  $\text{ExtInt}$  and  $\text{NExtS4}$  and their frames, but may be also used for transferring various properties of modal logics to their si-fragments and back. A few results of that sort are collected in

Table 7. Preservation Theorem

Property of logics	Preserved under		
	$\rho$	$\tau$	$\sigma$
Decidability	Yes	Yes	Yes
Kripke completeness	Yes	Yes	No
Strong completeness	Yes	Yes	No
Finite model property	Yes	Yes	Yes
Tabularity	Yes	No	Yes
Pretabularity	Yes	No	Yes
$\mathcal{D}$ -persistence	Yes	Yes	No
Local tabularity	Yes	No	No
Disjunction property	Yes	Yes	Yes
Halldén completeness	Yes	No	No
Interpolation property	Yes	No	No
Elementarity	Yes	Yes	No
Independent axiomatizability	No	Yes	Yes

Table 7; we shall cite them as the Preservation Theorem. The preservation of decidability follows from the definition of  $\rho$  and Theorem 167. That  $\rho$  preserves Kripke completeness, FMP and tabularity is a consequence of Theorem 157. The map  $\tau$  preserves Kripke completeness and FMP, since we can define  $\tau_k$  in Theorem 165 so that  $\tau_k \langle W, R \rangle = \langle kW, kR \rangle$ ; however,  $\tau$  does not in general preserve the tabularity, because  $\tau \mathbf{C1} = \mathbf{S5}$  is not tabular. The preservation of FMP and tabularity under  $\sigma$  follows from Theorem 163. On the other hand, Shehtman [1980] proved that  $\sigma$  does not preserve Kripke completeness (since  $\tau$  preserves it and  $\mathbf{Grz}$  is complete, this means in particular that Kripke completeness is not preserved under sums of logics in  $\mathbf{NExtS4}$ ). Some other preservation results in Table 7 will be discussed later. For references see [Chagrova and Zakharyashev 1992, 1997].

### 3.4 Completeness

In this section we briefly discuss the most important results concerning completeness of si-logics with respect to various classes of Kripke frames.

**Kripke completeness** That not all si-logics are complete with respect to Kripke frames was discovered by Shehtman [1977], who found a way

to adjust Fine’s [1974b] idea to the intuitionistic case (which was not so easy because intuitionistic formulas do not “feel” infinite ascending chains essential in Fine’s construction; see Section 20 of *Basic Modal Logic*). Note however that Kuznetsov’s [1975] question whether all si-logics are complete with respect to the topological semantics (see *Intuitionistic Logic*, volume 7 of this *Handbook*) is still open.

As to general positive results, notice first that the Preservation Theorem yields the following translation of Fine’s [1974c] Theorem on finite width logics (si-logics of finite width were studied by Sobolev [1977a]).

**THEOREM 168.** *Every si-logic of width  $n$  (i.e., a logic in  $\text{Ext}\mathbf{BW}_n$ ; see Table 5) is characterized by a class of Noetherian Kripke frames of width  $\leq n$ .*

The translation of Sahlqvist’s Theorem gives nothing interesting for si-logics. A sort of intuitionistic analog of this theorem has been recently proved by Ghilardi and Meloni [1997]. Here is a somewhat simplified variant of their result in which  $\bar{p}, \bar{q}, \bar{r}, \bar{s}$  denote tuples of propositional variables and  $\bar{\psi}, \bar{\chi}$  tuples of formulas of the same length as  $\bar{r}$  and  $\bar{s}$ , respectively.

**THEOREM 169** (Ghilardi and Meloni 1997). *Suppose  $\varphi(\bar{p}, \bar{q}, \bar{r}, \bar{s})$  is an intuitionistic formula in which the variables  $\bar{r}$  occur positively and the variables  $\bar{s}$  occur negatively, and which does not contain any  $\rightarrow$ , except for negations and double negations of atoms, in the premise of a subformula of the form  $\varphi' \rightarrow \varphi''$ . Assume also that  $\bar{\psi}(\bar{p}, \bar{q})$  and  $\bar{\chi}(\bar{p}, \bar{q})$  are formulas such that  $\bar{p}$  occur positively in  $\bar{\psi}$  and negatively in  $\bar{\chi}$ , while  $\bar{q}$  occur negatively in  $\bar{\psi}$  and positively in  $\bar{\chi}$ . Then the logic*

$$\mathbf{Int} + \varphi(\bar{p}, \bar{q}, \bar{\psi}(\bar{p}, \bar{q}), \bar{\chi}(\bar{p}, \bar{q}))$$

*is canonical.*

The preservation of  $\mathcal{D}$ -persistence under  $\rho$  (see [Zakharyashev 1996]) and the fact (discovered by Chagrova [1990]) that  $\tau L$  is characterized by an elementary class of Kripke frames whenever  $L$  is determined by such a class provide us with an intuitionistic variant of the Fine–van Benthem Theorem.

**THEOREM 170.** *If a si-logic is characterized by an elementary class of Kripke frames then it is  $\mathcal{D}$ -persistent.*

As in the modal case, it is unknown whether the converse of this theorem holds. All known non-elementary si-logics, for instance the Scott logic  $\mathbf{SL}$  and the logics  $\mathbf{T}_n$  of finite  $n$ -ary trees (see [Rodenburg 1986]) are not canonical and even strongly complete either, as was shown by Shimura [1995]. (Actually he proved that no logic in the intervals  $[\mathbf{SL}, \mathbf{SL} + \mathbf{bd}_3]$  and  $[\mathbf{Int}, \mathbf{T}_2]$ , save of course  $\mathbf{Int}$ , is strongly complete.)

As far as we know, there are no examples of si-logics separating canonicity,  $\mathcal{D}$ -persistence and strong completeness. (Ghilardi, Meloni and Miglioli have



recently showed that **SL** in any language with finitely many variables is canonical). Theorem 40 which holds in the intuitionistic case as well gives an algebraic counterpart of strong Kripke completeness.

**The finite model property.** The first example of an infinitely axiomatizable si-logic without FMP was constructed by Jankov [1968b]—that was in fact the starting point of a long series of “negative” results in modal logic. A finitely axiomatizable logic without FMP appeared two years later in [Kuznetsov and Gerchiu 1970]. The reader can get some impression about this and other examples of that sort by proving (it is really not hard) that

$$\varphi = \beta\left(\begin{array}{c} 1 \quad 2 \\ \circ \quad \circ \quad \circ \\ \swarrow \quad \downarrow \quad \searrow \\ \circ \end{array}\right) \notin L = \mathbf{Int} + \mathbf{bw}_4 + \beta\left(\begin{array}{c} 1 \quad 2 \\ \circ \quad \circ \quad \circ \\ \swarrow \quad \downarrow \quad \searrow \\ \circ \end{array}\right), \{\{1, 2\}\}$$

but no finite frame can separate  $\varphi$  from  $L$ . (Notice by the way that  $\tau L$  is axiomatizable by Sahlqvist formulas; see [Chagrov and Zakharyashev 1995b].)

FMP of a good many si-logics was proved using various forms of filtration; see e.g. [Gabbay 1970], [Ono 1972], [Smoryński 1973], [Ferrari and Miglioli 1993]. As an illustration of a rather sophisticated selective filtration we present here the following

**THEOREM 171** (Gabbay and de Jongh 1974). *The logic  $\mathbf{T}_n$  (see Table 5) is characterized by the class of finite  $n$ -ary trees.*

**Proof.** First we prove that  $\mathbf{T}_n$  is characterized by the class of finite frames of branching  $\leq n$ . Suppose  $\varphi \notin \mathbf{T}_n$  and  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  is a model for  $\mathbf{T}_n$  refuting  $\varphi$ . Without loss of generality we may assume that  $\mathfrak{F} = \langle W, R \rangle$  is a tree. Let  $\Sigma = \mathbf{Sub}\varphi$  and  $\Gamma_x = \{\psi \in \Sigma : x \models \psi\}$ , for every point  $x$  in  $\mathfrak{F}$ .

Given  $x$  in  $\mathfrak{F}$ , put  $rg(x) = \{[y] : y \in x\uparrow\}$  and say that  $x$  is of *minimal range* if  $rg(x) = rg(y)$  for every  $y \in [x] \cap x\uparrow$ . Since there are only finitely many distinct  $\Sigma$ -equivalence classes in  $\mathfrak{M}$ , every  $y \in [x]$  sees a point  $z \in [x]$  of minimal range. Now we extract from  $\mathfrak{M}$  a finite refutation frame  $\mathfrak{G} = \langle V, S \rangle$  for  $\varphi$  of branching  $\leq n$ . To begin with, we select some point  $x$  of minimal range at which  $\varphi$  is refuted and put  $V_0 = \{x\}$ .

Suppose  $V_k$  has already been defined. If  $|rg(x)| = 1$  for every  $x \in V_k$ , then we put  $\mathfrak{G} = \langle V, S \rangle$ , where  $V = \bigcup_{i=0}^k V_i$  and  $S$  is the restriction of  $R$  to  $V$ . Otherwise, for each  $x \in V_k$  with  $|rg(x)| > 1$  and each  $[y] \in rg(x)$  different from  $[x]$  and such that  $\Gamma_z \subset \Gamma_y$  for no  $[z] \in rg(x) - \{[x]\}$ , we select a point  $u \in [y] \cap x\uparrow$  of minimal range. Let  $U_x$  be the set of all selected points for  $x$  and  $V_{k+1} = \bigcup_x U_x$ . It should be clear that  $\Gamma_x \subset \Gamma_u$  (and  $rg(x) \supset rg(u)$ ), for every  $u \in U_x$ , and so the inductive process must terminate. Consequently  $\mathfrak{G} \not\models \varphi$ .

It remains to establish that  $\mathfrak{G} \models \mathbf{T}_n$ , i.e.,  $\mathfrak{G}$  is of branching  $\leq n$ . Suppose otherwise. Then there is a point  $x$  in  $\mathfrak{G}$  with  $m \geq n+1$  immediate successors  $x_0, \dots, x_m$ , which are evidently in  $U_x$  because  $\mathfrak{F}$  is a tree. We are going to construct a substitution instance of  $\mathbf{T}_n$ 's axiom  $\mathbf{bb}_n$  which is refuted at  $x$  in  $\mathfrak{M}$ .

Denote by  $\delta_i$  the conjunction of the formulas in  $\Gamma_{x_i}$ . Since all of them are true at  $x_i$  in  $\mathfrak{M}$ , we have  $x_i \models \delta_i$ ; and since  $\Gamma_i \subseteq \Gamma_j$  for no distinct  $i$  and  $j$ , we have  $x_j \not\models \delta_i$  if  $i \neq j$ . Put  $\chi_i = \delta_i$ , for  $0 \leq i < n$ ,  $\chi_n = \delta_n \vee \dots \vee \delta_m$  and consider the truth-value of the formula  $\psi = \mathbf{bb}_n\{\chi_0/p_0, \dots, \chi_n/p_n\}$  at  $x$  in  $\mathfrak{M}$ .

Since  $xR x_i$  for every  $i = 0, \dots, m$ , we have  $x \not\models \bigvee_{i=0}^n \chi_i$ . Suppose that  $x \not\models \bigwedge_{i=0}^n ((\chi_i \rightarrow \bigvee_{i \neq j} \chi_j) \rightarrow \bigvee_{i \neq j} \chi_j)$ . Then  $y \models \chi_i \rightarrow \bigvee_{i \neq j} \chi_j$  and  $y \not\models \bigvee_{i \neq j} \chi_j$ , for some  $y \in x \uparrow$  and some  $i \in \{0, \dots, n\}$ , and hence  $y \not\models \chi_i$ . Since  $x_i \models \chi_i$  and  $x_i \not\models \bigvee_{i \neq j} \chi_j$ ,  $y$  sees no point in  $[x_i]$  and so  $y \not\sim_{\Sigma} x$  (for otherwise  $x$  would not be of minimal range). Therefore,  $\Gamma_{x_j} \subseteq \Gamma_y$  for some  $j \in \{0, \dots, m\}$ , and then  $y \models \chi_j$  if  $j < n$  and  $y \models \chi_n$  if  $j \geq n$ , which is a contradiction.

It follows that  $x \models \bigwedge_{i=0}^n ((\chi_i \rightarrow \bigvee_{i \neq j} \chi_j) \rightarrow \bigvee_{i \neq j} \chi_j)$ , from which  $x \models \psi$ , contrary to  $\mathfrak{M}$  being a model for  $\mathbf{bb}_n$ . It remains to notice that every finite frame of branching  $\leq n$  is a reduct of a finite  $n$ -ary tree, which clearly validates  $\mathbf{T}_n$ . ■

Another way of obtaining general results on FMP of si-logics is to translate the corresponding results in modal logic with the help of the Preservation Theorem.

**THEOREM 172.** *Every si-logic of finite depth (i.e., every logic in  $\text{Ext}\mathbf{BD}_n$ , for  $n < \omega$ ) is locally tabular.*

Note, however, that unlike  $\text{NExt}\mathbf{K4}$ , the converse does not hold: the Dummett logic  $\mathbf{LC}$ , characterized by the class of finite chains (or by the infinite ascending chain), is locally tabular. As we saw in Section 1.7, every non-locally tabular in  $\text{NExt}\mathbf{S4}$  logic is contained in  $\mathbf{Grz.3}$ , the only *pre-locally tabular logic* in  $\text{NExt}\mathbf{S4}$ . But in  $\text{Ext}\mathbf{Int}$  this way of determining local tabularity does not work:

**THEOREM 173** (Mardaev 1984). *There is a continuum of pre-locally tabular logics in  $\text{Ext}\mathbf{Int}$ .*

Besides, it is not clear whether every locally tabular logic in  $\text{Ext}\mathbf{Int}$  (or  $\text{NExt}\mathbf{K4}$ ) is contained in a pre-locally tabular one.

An intuitionistic formula is said to be *essentially negative* if every occurrence of a variable in it is in the scope of some  $\neg$ . If  $\varphi$  is essentially negative then  $T(\varphi)$  is a  $\Box\Diamond$ -formula, which yields

THEOREM 174 (McKay 1971, Rybakov 1978). *If a si-logic  $L$  is decidable (or has FMP) and  $\varphi$  is an essentially negative formula then  $L+\varphi$  is decidable (has FMP).*

Originally this result was proved with the help of Glivenko's Theorem (see Section 7 in *Intuitionistic Logic*). Say that an occurrence of a variable in a formula is *essential* if it is not in the scope of any  $\neg$ . A formula  $\varphi$  is *mild* if every two essential occurrences of the same variable in  $\varphi$  are either both positive or both negative. Kuznetsov [1972] claimed (we have not seen the proof) that all si-logics whose extra axioms do not contain negative occurrences of essential variables have FMP. And Wroński [1989] announced that if  $L$  is a decidable si-logic and  $\varphi$  a mild formula then  $L + \varphi$  is also decidable.

Subframe and cofinal subframe si-logics—that is logics axiomatizable by canonical formulas of the form  $\beta(\mathfrak{F})$  and  $\beta(\mathfrak{F}, \perp)$ , respectively—can be characterized both syntactically and semantically (see [Zakharyashev 1996]).

THEOREM 175. *The following conditions are equivalent for every si-logic  $L$ :*

- (i)  *$L$  is a (cofinal) subframe logic;*
- (ii)  *$L$  is axiomatizable by implicative (respectively, disjunction free) formulas;*
- (iii)  *$L$  is characterized by a class of finite frames closed under the formation of (cofinal) subframes.*

That all si-logics with disjunction free axioms have FMP was first proved by McKay [1968] with the help of Diego's [1966] Theorem according to which there are only finitely many pairwise non-equivalent in **Int** disjunction free formulas in variables  $p_1, \dots, p_n$  (see also [Urquhart 1974]).

Since frames for **Int** contain no clusters, Theorem 58 and its analog for cofinal subframe logics reduce in the intuitionistic case to the following result which is due to Chagrova [1986], Rodenburg [1986], Shimura [1993] and Zakharyashev [1996].

THEOREM 176. *All si-logics with disjunction free axioms are elementary (definable by  $\forall\exists$ -sentences) and  $\mathcal{D}$ -persistent.*

Theorem 68 is translated into the intuitionistic case simply by replacing **K4** with **Int**,  $\oplus$  with  $+$  and  $\alpha$  with  $\beta$ . As a consequence we obtain, for instance, that Ono's [1972]  $\mathbf{B}_n$  and all other logics whose canonical axioms are built on trees have FMP. Moreover, we also have

THEOREM 177 (Sobolev 1977b, Nishimura 1960). *All si-logics with extra axioms in one variable have FMP and are decidable.*

In fact Sobolev [1977b] proved a more general (but rather complicated) syntactical sufficient condition of FMP and constructed a formula in two variables axiomatizing a si-logic without FMP (Shehtman's [1977] incomplete si-logic has also axioms in two variables).

**Tabularity** By the Blok–Esakia and Preservation Theorems, the situation with tabular logics in  $\text{ExtInt}$  is the same as in  $\text{NExtGrz}$ . In particular,  $L \in \text{ExtInt}$  is tabular iff  $\mathbf{BD}_n + \mathbf{BW}_n \subseteq L$  for some  $n < \omega$  iff  $L$  is not a sublogic of one of the three pretabular logics in  $\text{ExtInt}$ , namely  $\mathbf{LC}$ ,  $\mathbf{BD}_2$  and  $\mathbf{KC} + \mathbf{bd}_3$ . (The pretabular si-logics were described by Maksimova [1972].) The tabularity problem is decidable in  $\text{ExtInt}$ .

### 3.5 Disjunction property

One of the aims of studying extensions of  $\mathbf{Int}$ , which may be of interest for applications in computer science, is to describe the class of constructive si-logics. At the propositional level a consistent logic  $L \in \text{ExtInt}$  is regarded to be constructive if it has the *disjunction property* (DP, for short) which means that for all formulas  $\varphi$  and  $\psi$ ,

$$\varphi \vee \psi \in L \text{ implies } \varphi \in L \text{ or } \psi \in L.$$

That intuitionistic logic itself is constructive in this sense was proved in a syntactic way by Gentzen [1934–1935]. However, Łukasiewicz (1952) conjectured that no proper consistent extension of  $\mathbf{Int}$  has DP.

A similar property was introduced for modal logics (see e.g. [Lemmon and Scott 1977]):  $L \in \text{NExtK}$  has the (*modal*) *disjunction property* if, for every  $n \geq 1$  and all formulas  $\varphi_1, \dots, \varphi_n$ ,

$$\Box\varphi_1 \vee \dots \vee \Box\varphi_n \in L \text{ implies } \varphi_i \in L, \text{ for some } i \in \{1, \dots, n\}.$$

The following theorem (in a somewhat different form it was proved in [Hughes and Cresswell 1984] and [Maksimova 1986]) provides a semantic criterion of DP.

**THEOREM 178.** *Suppose a modal or si-logic  $L$  is characterized by a class  $\mathcal{C}$  of descriptive rooted frames closed under the formation of rooted generated subframes. Then  $L$  has DP iff, for every  $n \geq 1$  and all  $\mathfrak{F}_1, \dots, \mathfrak{F}_n \in \mathcal{C}$  with roots  $x_1, \dots, x_n$ , there is a frame  $\mathfrak{F}$  for  $L$  with root  $x$  such that the disjoint union  $\mathfrak{F}_1 + \dots + \mathfrak{F}_n$  is a generated subframe of  $\mathfrak{F}$  with  $\{x_1, \dots, x_n\} \subseteq x\uparrow$ .*

**Proof.** We consider only the modal case. ( $\Rightarrow$ ) Let  $\mathfrak{F}_L = \langle W_L, R_L, P_L \rangle$  be a universal frame for  $L$ , big enough to contain  $\mathfrak{F}_1 + \dots + \mathfrak{F}_n$  as its generated subframe. Assuming that  $\mathfrak{F}_L$  is associated with a suitable canonical model for  $L$ , we show that there is a point  $x$  in  $\mathfrak{F}_L$  such that  $x\uparrow = W_L$ . The set

$$\Delta' = \{\neg\Box\varphi : \exists y \in W_L \ y \not\models \varphi\}$$

is  $L$ -consistent (for otherwise  $\Box\varphi_1 \vee \dots \vee \Box\varphi_n \in L$  for some  $\varphi_1, \dots, \varphi_n \notin L$ ). Let  $\Delta$  be a maximal  $L$ -consistent extension of  $\Delta'$  and  $x$  the point in  $\mathfrak{F}_L$  where  $\Delta$  is true. Then  $xR_L y$ , for every  $y \in W_L$ .

( $\Leftarrow$ ) Suppose otherwise. Then there are formulas  $\varphi_1, \dots, \varphi_n \notin L$  such that  $\Box\varphi_1 \vee \dots \vee \Box\varphi_n \in L$ . Take frames  $\mathfrak{F}_1, \dots, \mathfrak{F}_n \in \mathcal{C}$  refuting  $\varphi_1, \dots, \varphi_n$  at their roots, respectively, and let  $\mathfrak{F}$  be a rooted frame for  $L$  containing  $\mathfrak{F}_1 + \dots + \mathfrak{F}_n$  as a generated subframe and such that its root  $x$  sees the roots of  $\mathfrak{F}_1, \dots, \mathfrak{F}_n$ . Then all the formulas  $\Box\varphi_1, \dots, \Box\varphi_n$  are refuted at  $x$  and so  $\Box\varphi_1 \vee \dots \vee \Box\varphi_n \notin L$ , which is a contradiction.  $\blacksquare$

It should be clear that if we use only the sufficient condition of Theorem 178, the requirement that frames in  $\mathcal{C}$  are descriptive is redundant. Furthermore, it is easy to see that for  $L \in \text{NExt}\mathbf{K4}$  we may assume  $n \leq 2$ . And clearly a logic  $L \in \text{NExt}\mathbf{S4}$  has DP iff, for all  $\varphi$  and  $\psi$ ,  $\Box\varphi \vee \Box\psi \in L$  implies  $\Box\varphi \in L$  or  $\Box\psi \in L$ .

As a direct consequence of the proof above we obtain

**COROLLARY 179.** *A modal or si-logic  $L$  has DP iff the canonical frame  $\mathfrak{F}_L = \langle W_L, R_L \rangle$  contains a point  $x$  such that  $x\uparrow = W_L$ .*

Using the semantic criterion above it is not hard to show that DP is preserved under  $\rho$ ,  $\tau$  and  $\sigma$ . It is also a good tool for proving and disproving DP of logics with transparent semantics.

**EXAMPLE 180.**

- (i) Let  $\mathfrak{F}_1, \dots, \mathfrak{F}_n$  be serial rooted Kripke frames. Then the frame obtained by adding a root to  $\mathfrak{F}_1 + \dots + \mathfrak{F}_n$  is also serial. Therefore,  $\mathbf{D}$  has DP. In the same way one can show that  $\mathbf{K}$ ,  $\mathbf{K4}$ ,  $\mathbf{T}$ ,  $\mathbf{S4}$ ,  $\mathbf{Grz}$ ,  $\mathbf{GL}$  and many other modal logics have DP.
- (ii) Since no rooted symmetrical frame can contain a proper generated subframe, no consistent logic in  $\text{NExt}\mathbf{KB}$  has DP.

The first proper extensions of  $\mathbf{Int}$  with DP were constructed by Kreisel and Putnam [1957]: these were  $\mathbf{KP}$  (now called the *Kreisel–Putnam logic*) and  $\mathbf{SL}$  (known as the *Scott logic*). We present here Gabbay's [1970] proof that  $\mathbf{KP}$  has DP.

**THEOREM 181** (Kreisel and Putnam 1957).  $\mathbf{KP}$  has DP.

**Proof.** Using filtration one can show that  $\mathbf{KP}$  is characterized by the class of finite rooted frames  $\mathfrak{F} = \langle W, R \rangle$  satisfying the condition

$$(15) \quad \forall x, y, z (xRy \wedge xRz \wedge \neg yRz \wedge \neg zRy \rightarrow \exists u (xRu \wedge uRy \wedge uRz \wedge \forall v (uRv \rightarrow \exists w (vRw \wedge (yRw \vee zRw)))))).$$

If  $\mathfrak{F}$  is such a frame then for each non-empty  $X \subseteq W^{\leq 1}$ , the generated subframe of  $\mathfrak{F}$  based on the set  $W - (W^{\leq 1} - X)_{\downarrow}$  is rooted; we denote its root by  $r(X)$ .

Let  $\mathfrak{F}_1 = \langle W_1, R_1 \rangle$  and  $\mathfrak{F}_2 = \langle W_2, R_2 \rangle$  be finite rooted frames satisfying (15). We construct from them a frame  $\mathfrak{F} = \langle W, R \rangle$  by taking

$$W = W_1 \cup W_2 \cup U,$$

where  $U = \{X_1 \cup X_2 : X_1 \subseteq W_1^{\leq 1}, X_2 \subseteq W_2^{\leq 1}, X_1, X_2 \neq \emptyset\}$ , and

$$\begin{aligned} xRy \quad \text{iff} \quad & (x, y \in W_i \wedge xR_i y) \vee (x, y \in U \wedge x \supseteq y) \vee \\ & (x = X_1 \cup X_2 \in U \wedge y \in W_i \wedge r(X_i)R_i y). \end{aligned}$$

It follows from the given definition that  $\mathfrak{F}_1 + \mathfrak{F}_2$  is a generated subframe of  $\mathfrak{F}$ ,  $W_1 \cup W_2$  is a cover for  $\mathfrak{F}$  and  $W_1^{\leq 1} \cup W_2^{\leq 1}$  is its root. So our theorem will be proved if we show that (15) holds.

Suppose  $x, y, z \in W$  satisfy the premise of (15). Since (15) holds for  $\mathfrak{F}_1$  and  $\mathfrak{F}_2$ , we can assume that  $x = X_1 \cup X_2 \in U$ . Let  $Y_1 \cup Y_2$  and  $Z_1 \cup Z_2$  be the sets of final points in  $y \uparrow$  and  $z \uparrow$ , respectively, with  $Y_i, Z_i \subseteq W_i$ . By the definition of  $R$ , we have  $Y_i, Z_i \subseteq X_i$ . Consider  $u = (Y_1 \cup Z_1) \cup (Y_2 \cup Z_2)$ . Clearly,  $xRu, uRy$  and  $uRz$ . Suppose now that  $v \in u \uparrow$ . Let  $w$  be any final point in  $v \uparrow$ . Then  $v \in (Y_1 \cup Z_1) \cup (Y_2 \cup Z_2)$  and so either  $yRw$  or  $zRw$ . ■

Other examples of constructive si-logics were constructed by Ono [1972] and Gabbay and de Jongh [1974], namely,  $\mathbf{B}_n$  and  $\mathbf{T}_n$ . Anderson [1972] proved that among the consistent si-logics with extra axioms in one variable only those of the form  $\mathbf{Int} + \mathbf{nf}_{2n+2}$ , for  $n \geq 5$ , have DP (for  $n = 6$  the proof was found by Wroński [1974]; see also [Sasaki 1992]). Finally, Wroński [1973] showed that there is a continuum of si-logics with DP.

The additional axioms of logics in all these examples contained occurrences of  $\vee$ ; on the other hand, known examples of si-logics with disjunction free extra axioms, say  $\mathbf{LC}$ ,  $\mathbf{KC}$ ,  $\mathbf{Cl}$ ,  $\mathbf{BW}_n$  or  $\mathbf{BD}_n$ , were not constructive. This observation led Hosoi and Ono [1973] to the conjecture that the disjunction free fragment of every consistent si-logic with DP coincides with that of  $\mathbf{Int}$ . We present a proof of this conjecture following [Zakharyashev 1987].

First we describe the cofinal subframe logics in  $\mathbf{NExtS4}$  with DP, assuming that every such logic  $L$  is represented by its independent canonical axiomatization

$$(16) \quad L = \mathbf{S4} \oplus \{\alpha(\mathfrak{F}_i, \perp) : i \in I\}.$$

All frames in the rest of this section are assumed to be quasi-ordered.

Say that a finite rooted frame  $\mathfrak{F}$  with  $\geq 2$  points is *simple* if its root cluster and at least one of the final clusters are simple. Suppose  $\mathfrak{F} = \langle W, R \rangle$  is a

simple frame,  $a_0, a_1, \dots, a_m, a_{m+1}, \dots, a_n$  are all its points, with  $a_0$  being the root,  $C(a_1), \dots, C(a_m)$  all the distinct immediate cluster-successors of  $a_0$ , and  $a_n$  a final point with simple  $C(a_n)$ . For every  $k = 1, \dots, n$ , define a formula  $\psi_k$  by taking

$$\psi_k = \bigwedge_{a_i R a_j, i \neq 0} \varphi_{ij} \wedge \bigwedge_{i=1}^n \varphi_i \wedge \varphi'_\perp \rightarrow p_k$$

where  $\varphi_{ij}$ ,  $\varphi_i$  were defined in Section 3.2 and  $\varphi'_\perp = \Box(\bigwedge_{i=1}^n \Box p_i \rightarrow \perp)$ . Now we associate with  $\mathfrak{F}$  the formula  $\gamma(\mathfrak{F}) = \Box p_0 \vee \Box \psi_1$  if  $m = 1$ , and the formula  $\gamma(\mathfrak{F}) = \Box \psi_1 \vee \dots \vee \Box \psi_m$  if  $m > 1$ .

LEMMA 182. *For every simple frame  $\mathfrak{F}$ ,  $\gamma(\mathfrak{F}) \in \mathbf{S4} \oplus \alpha(\mathfrak{F}, \perp)$ .*

**Proof.** It is enough to show that  $\mathfrak{G} \not\models \gamma(\mathfrak{F})$  implies  $\mathfrak{G} \not\models \alpha(\mathfrak{F}, \perp)$ , for any finite  $\mathfrak{G}$ . So suppose  $\gamma(\mathfrak{F})$  is refuted in a finite frame  $\mathfrak{G}$  under some valuation. Define a partial map  $f$  from  $\mathfrak{G}$  onto  $\mathfrak{F}$  by taking

$$f(x) = \begin{cases} a_0 & \text{if } x \not\models \gamma(\mathfrak{F}) \\ a_i & \text{if } x \not\models \psi_i, 1 \leq i \leq n \\ \text{undefined} & \text{otherwise.} \end{cases}$$

One can readily check that  $f$  is a subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$ . However it is not necessarily cofinal. So we extend  $f$  by putting  $f(x) = a_n$ , for every  $x$  of depth 1 in  $\mathfrak{G}$  such that  $f(x \downarrow) = \{a_0\}$ . Clearly, the improved map is still a subreduction of  $\mathfrak{G}$  to  $\mathfrak{F}$ , and  $\varphi'_\perp$  ensures its cofinality.  $\blacksquare$

Using the semantical properties of the canonical formulas it is a matter of routine to prove the following

LEMMA 183. *Suppose  $i \in \{1, \dots, m\}$  and  $\mathfrak{G}$  is the subframe of  $\mathfrak{F}$  generated by  $a_i$ . Then  $\alpha(\mathfrak{G}, \perp) \in \mathbf{S4} \oplus \psi_i$ .*

We are in a position now to prove a criterion of DP for the cofinal subframe logics in  $\mathbf{NExtS4}$ .

THEOREM 184. *A consistent cofinal subframe logic  $L \in \mathbf{NExtS4}$  has the disjunction property iff no frame  $\mathfrak{F}_i$  in its independent axiomatization (16) is simple, for  $i \in I$ .*

**Proof.** ( $\Rightarrow$ ) Suppose, on the contrary, that  $\mathfrak{F}_i$  is simple, for some  $i \in I$ . Since the axiomatization (16) is independent, every proper generated subframe of  $\mathfrak{F}_i$  validates  $L$ . By Lemma 182,  $\gamma(\mathfrak{F}_i) \in L$  and so either  $p_0 \in L$  or  $\psi_j \in L$ . However, both alternatives are impossible: the former means that  $L$  is inconsistent, while the latter, by Lemma 183, implies  $\alpha(\mathfrak{G}, \perp) \in L$ , where  $\mathfrak{G}$  is the subframe of  $\mathfrak{F}_i$  generated by an immediate successor of  $\mathfrak{F}_i$ 's root.

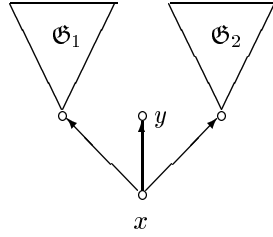


Figure 15.

( $\Leftarrow$ ) Given two finite rooted frames  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  for  $L$ , we construct the frame  $\mathfrak{F}$  as shown in Fig. 15 and prove that  $\mathfrak{F} \models L$ . Suppose otherwise, i.e., there exists a cofinal subreduction  $f$  of  $\mathfrak{F}$  to  $\mathfrak{F}_i$ , for some  $i \in I$ . Let  $x_i$  be the root of  $\mathfrak{F}_i$ . Since  $\mathfrak{G}_1$  and  $\mathfrak{G}_2$  are not cofinally subreducible to  $\mathfrak{F}_i$  and since  $L$  is consistent,  $f^{-1}(x_i) = \{x\}$ . By the cofinality condition, it follows in particular that  $y \in \text{dom} f$ . But then  $\mathfrak{F}_i$  is simple, which is a contradiction. Thus, by Theorem 178,  $L$  has DP. ■

Note that in fact the proof of ( $\Rightarrow$ ) shows that if  $L \in \text{NExtS4}$ ,  $\mathfrak{F}$  is a simple frame,  $\alpha(\mathfrak{F}, \perp) \in L$  and  $\alpha(\mathfrak{G}, \perp) \notin L$  for any proper generated subframe  $\mathfrak{G}$  of  $\mathfrak{F}$  then  $L$  does not have DP. Transferring this observation to the intuitionistic case, we obtain

**THEOREM 185** (Minari 1986, Zakharyashev 1987). *If a si-logic is consistent and has DP then the disjunction free fragments of  $L$  and **Int** are the same.*

Sufficient conditions of DP in terms of canonical formulas can be found in [Chagrova and Zakharyashev 1993, 1997].

Since classical logic is not constructive, it is of interest to find maximal consistent si-logics with DP. That they exist follows from Zorn's Lemma. Here is a concrete example of such a logic.

Trying to formalize the proof interpretation of intuitionistic logic, Medvedev [1962] proposed to treat intuitionistic formulas as finite problems. Formally, a *finite problem* is a pair  $\langle X, Y \rangle$  of finite sets such that  $Y \subseteq X$  and  $X \neq \emptyset$ ; elements in  $X$  are called *possible solutions* and elements in  $Y$  *solutions* to the problem. The operations on finite problems, corresponding to the logical connectives, are defined as follows:

$$\begin{aligned} \langle X_1, Y_1 \rangle \wedge \langle X_2, Y_2 \rangle &= \langle X_1 \times X_2, Y_1 \times Y_2 \rangle, \\ \langle X_1, Y_1 \rangle \vee \langle X_2, Y_2 \rangle &= \langle X_1 \sqcup X_2, Y_1 \sqcup Y_2 \rangle, \\ \langle X_1, Y_1 \rangle \rightarrow \langle X_2, Y_2 \rangle &= \left\langle X_2^{X_1}, \{f \in X_2^{X_1} : f(Y_1) \subseteq Y_2\} \right\rangle, \end{aligned}$$



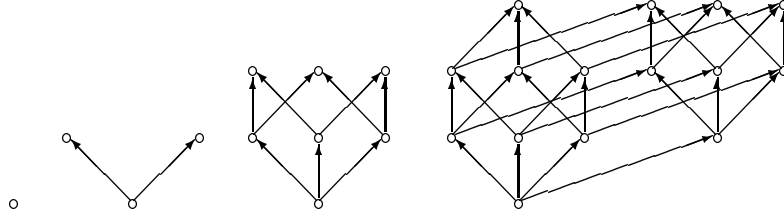


Figure 16.

$$\perp = \langle X, \emptyset \rangle.$$

Here  $X \sqcup Y = (X \times \{1\}) \cup (Y \times \{2\})$  and  $X^Y$  is the set of all functions from  $X$  into  $Y$ . Note that in the definition of  $\perp$  the set  $X$  is fixed, but arbitrary; for definiteness one can take  $X = \{\emptyset\}$ .

Now we can interpret formulas by finite problems. Namely, given a formula  $\varphi$ , we replace its variables by arbitrary finite problems and perform the operations corresponding to the connectives in  $\varphi$ . If the result is a problem with a non-empty set of solutions no matter what finite problems are substituted for the variables in  $\varphi$ , then  $\varphi$  is called *finitely valid*. One can show that the set of all finitely valid formulas is a si-logic; it is called *Medvedev's logic* and denoted by **ML**.

In fact, **ML** can be defined semantically. Medvedev [1966] showed that **ML** coincides with the set of formulas that are valid in all frames  $\mathfrak{B}_n$  having the form of the  $n$ -ary Boolean cubes with the topmost point deleted; for  $n = 1, 2, 3, 4$ , the Medvedev frames are shown in Fig. 16. Since  $\mathfrak{B}_n + \mathfrak{B}_m$  is a generated subframe of  $\mathfrak{B}_{n+m}$ , **ML** has DP. Moreover, Levin [1969] proved that it has no proper consistent extension with DP. The following proof of this result is due to Maksimova [1986].

**THEOREM 186** (Levin 1969). *ML is a maximal si-logic with DP.*

**Proof.** Suppose, on the contrary, that there exists a proper consistent extension  $L$  of **ML** having DP. Then we have a formula  $\varphi \in L - \mathbf{ML}$ . We show first that there is an essentially negative substitution instance  $\varphi^*$  of  $\varphi$  such that  $\varphi^* \notin \mathbf{ML}$ . Since  $\varphi(p_1, \dots, p_n) \notin \mathbf{ML}$ , there is a Medvedev frame  $\mathfrak{B}_m$  refuting  $\varphi$  under some valuation  $\mathfrak{V}$ . With every point  $x$  in  $\mathfrak{B}_m$  we associate a new variable  $q_x$  and extend  $\mathfrak{V}$  to these variables by taking  $\mathfrak{V}(q_x)$  to be the set of final points in  $\mathfrak{B}_m$  that are not accessible from  $x$ . By the construction of  $\mathfrak{B}_m$ , we have  $y \models \neg q_x$  iff  $y \in x\uparrow$ , from which

$$\mathfrak{V}\left(\bigvee_{x \in \mathfrak{V}(p_i)} \neg q_x\right) = \mathfrak{V}(p_i).$$

Let  $\varphi^* = \varphi(\bigvee_{x \in \mathfrak{B}(p_1)} \neg q_x, \dots, \bigvee_{x \in \mathfrak{B}(p_n)} \neg q_x)$ . It follows that  $\mathfrak{B}(\varphi^*) = \mathfrak{B}(\varphi)$  and so  $\varphi^* \notin \mathbf{ML}$ .

Thus, we may assume that  $\varphi$  is an essentially negative formula. Since  $\mathbf{KP} \subseteq \mathbf{ML}$ ,  $\mathbf{ML}$  contains the formulas

$$\mathbf{nd}_k = (\neg p \rightarrow \neg q_1 \vee \dots \vee \neg q_k) \rightarrow (\neg p \rightarrow \neg q_1) \vee \dots \vee (\neg p \rightarrow \neg q_k)$$

which, as is easy to see, belong to  $\mathbf{KP}$ . Let us consider the logic

$$\mathbf{ND} = \mathbf{Int} + \{\mathbf{nd}_k : k \geq 1\}.$$

Using the fact that the outermost  $\rightarrow$  in  $\mathbf{nd}_k$  can be replaced with  $\leftrightarrow$  and that  $(\neg p \rightarrow \neg q) \leftrightarrow \neg(\neg p \wedge q) \in \mathbf{Int}$ , one can readily show that every essentially negative formula is equivalent in  $\mathbf{ND}$  to the conjunction of formulas of the form  $\neg\chi_1 \vee \dots \vee \neg\chi_l$ . So  $L\text{-ML}$  contains a formula of the form  $\neg\chi_1 \vee \dots \vee \neg\chi_l$ . Since  $L$  has DP,  $\neg\chi_i \in L$  for some  $i$ . But then, by Glivenko's Theorem,  $\neg\chi_i \in \mathbf{ML}$ , which is a contradiction. ■

**REMARK.**  $\mathbf{ML}$  is not finitely axiomatizable, as was shown by Maksimova *et al.* [1979]. Nobody knows whether it is decidable.

It turns out, however, that  $\mathbf{ML}$  is not the unique maximal logic with DP in  $\text{ExtInt}$ . Kirk [1982] noted that there is no greatest consistent si-logic with DP. Maksimova [1984] showed that there are infinitely many maximal constructive si-logics, and Chagrov [1992a] proved that in fact there are a continuum of them; see also Ferrari and Miglioli [1993, 1995a, 1995b]. Galanter [1990] claims that each si-logic characterized by the class of frames of the form

$$\langle \{W : W \subseteq \{1, \dots, n\}, W \neq \emptyset, |W| \notin N\}, \supseteq \rangle,$$

where  $n = 1, 2, \dots$  and  $N$  is some fixed infinite set of natural numbers, is a maximal si-logic with DP.

### 3.6 Intuitionistic Modal Logics

All modal logics we have dealt with so far were constructed on the classical non-modal basis. It can be replaced by logics of other types. For instance, one can consider modal logics based on relevant logic (see e.g. [Fuhrmann 1989]) or many-valued logics (see e.g. [Seegerberg 1967], [Morikawa 1989], [Ostermann 1988]), and many others. In this section we briefly discuss modal logics with the intuitionistic basis.

Unlike the classical case, the intuitionistic  $\Box$  and  $\Diamond$  are not supposed to be dual, which provides more possibilities for defining intuitionistic modal logics. For a non-empty set  $M$  of modal operators, let  $\mathcal{L}_M$  be the standard propositional language augmented by the connectives in  $M$ . By an

*intuitionistic modal logic* in the language  $\mathcal{L}_M$  we understand any subset of  $\mathcal{L}_M$  containing **Int** and closed under modus ponens, substitution and the regularity rule  $\varphi \rightarrow \psi / \bigcirc \varphi \rightarrow \bigcirc \psi$ , for every  $\bigcirc \in M$ .

There are three ways of defining intuitionistic analogues of (classical) normal modal logics. First, one can take the family of logics extending the basic system **IntK** $_{\square}$  in the language  $\mathcal{L}_{\square}$  which is axiomatized by adding to **Int** the standard axioms of **K**

$$\square(p \wedge q) \leftrightarrow \square p \wedge \square q \text{ and } \square \top.$$

An example of a logic in this family is Kuznetsov's [1985] intuitionistic provability logic **I** $^{\Delta}$  (Kuznetsov used  $\Delta$  instead of  $\square$ ), the intuitionistic analog of the provability logic **GL**. It can be obtained by adding to **IntK** $_{\square}$  (and even to **Int**) the axioms

$$p \rightarrow \square p, (\square p \rightarrow p) \rightarrow p, ((p \rightarrow q) \rightarrow p) \rightarrow (\square q \rightarrow p).$$

A model theory for logics in **NExtIntK** $_{\square}$  was developed by Ono [1977], Božić and Došen [1984], Došen [1985a], Sotirov [1984] and Wolter and Zakharyashev [1997, 1999a]; we discuss it below. Font [1984, 1986] considered these logics from the algebraic point of view, and Luppi [1996] investigated their interpolation property by proving, in particular, that the superamalgamability of the corresponding varieties of algebras is equivalent to interpolation.

A possibility operator  $\diamond$  in logics of this sort can be defined in the classical way by taking  $\diamond \varphi = \neg \square \neg \varphi$ . Note, however, that in general this  $\diamond$  does not distribute over disjunction and that the connection via negation between  $\square$  and  $\diamond$  is too strong from the intuitionistic standpoint (actually, the situation here is similar to that in intuitionistic predicate logic where  $\exists$  and  $\forall$  are not dual.)

Another family of “normal” intuitionistic modal logics can be defined in the language  $\mathcal{L}_{\diamond}$  by taking as the basic system the smallest logic in  $\mathcal{L}_{\diamond}$  to contain the axioms

$$\diamond(p \vee q) \leftrightarrow \diamond p \vee \diamond q \text{ and } \neg \diamond \perp;$$

it will be denoted by **IntK** $_{\diamond}$ . Logics in **NExtIntK** $_{\diamond}$  were studied by Božić and Došen [1984], Došen [1985a], Sotirov [1984] and Wolter [1997e].

Finally, we can define intuitionistic modal logics with independent  $\square$  and  $\diamond$ . These are extensions of **IntK** $_{\square \diamond}$ , the smallest logic in the language  $\mathcal{L}_{\square \diamond}$  containing both **IntK** $_{\square}$  and **IntK** $_{\diamond}$ . Fischer Servi [1980, 1984] constructed a logic in **NExtIntK** $_{\square \diamond}$  by imposing a weak connection between the necessity and possibility operators:

$$\mathbf{FS} = \mathbf{IntK}_{\square \diamond} \oplus \diamond(p \rightarrow q) \rightarrow (\square p \rightarrow \diamond q) \oplus (\diamond p \rightarrow \square q) \rightarrow \square(p \rightarrow q).$$

A remarkable feature of **FS** is that the standard translation  $ST$  of modal formulas into first order ones (see *Correspondence Theory*) not only embeds **K** into classical predicate logic but also **FS** into intuitionistic first order logic:  $\varphi$  belongs to the former iff  $ST(\varphi)$  is a theorem of the latter. According to Simpson [1994], this result was proved by C. Stirling; see also Grefe [1997].

Various extensions of **FS** were studied by Bull [1966a], Ono [1977], Fischer Servi [1977, 1980, 1984], Amati and Pirri [1994], Ewald [1986], Wolter and Zakharyashev [1997], Wolter [1997e]. The best known one is probably the logic

$$\begin{aligned} \mathbf{MIPC} = \mathbf{FS} \oplus & \Box p \rightarrow p \oplus \Box p \rightarrow \Box \Box p \oplus \Diamond p \rightarrow \Box \Diamond p \oplus \\ & p \rightarrow \Diamond p \oplus \Diamond \Diamond p \rightarrow \Diamond p \oplus \Diamond \Box p \rightarrow \Box p \end{aligned}$$

introduced by Prior [1957]. Bull [1966a] noticed that the translation  $*$  defined by

$$\begin{aligned} (p_i)^* &= P_i(x), \quad \perp^* = \perp, \\ (\psi \odot \chi)^* &= \psi^* \odot \chi^*, \text{ for } \odot \in \{\wedge, \vee, \rightarrow\}, \\ (\Box \psi)^* &= \forall x \psi^*, \quad (\Diamond \psi)^* = \exists x \psi^* \end{aligned}$$

is an embedding of **MIPC** into the monadic fragment of intuitionistic predicate logic. Ono [1977], Ono and Suzuki [1988], Suzuki [1990], and Bezhanishvili [1998] investigated the relations between logics in **NExtMIPC** and superintuitionistic predicate logics induced by that translation.

In what follows we restrict attention only to the classes of intuitionistic modal logics introduced above. An interesting example of a system not covered here was constructed by Wijesekera [1990]. A general model theory for such logics is developed by Sotirov [1984] and Wolter and Zakharyashev [1997].

Let us consider first the algebraic and relational semantics for the logics introduced above. All the semantical concepts to be defined below turn out to be natural combinations of the corresponding notions developed for classical modal and si-logics. For details and proofs we refer the reader to Wolter and Zakharyashev [1997, 1999a].

From the algebraic point of view, every logic  $L \in \mathbf{NExtIntK}_M$ , for  $M \subseteq \{\Box, \Diamond\}$ , corresponds to the variety of Heyting algebras with one or two operators validating  $L$ . The variety of algebras for **IntK<sub>M</sub>** will be called the *variety of M-algebras*.

To construct the relational representations of  $M$ -algebras, we define a  $\Box$ -frame to be a structure of the form  $\langle W, R, R_\Box, P \rangle$  in which  $\langle W, R, P \rangle$  is an intuitionistic frame,  $R_\Box$  a binary relation on  $W$  such that

$$R \circ R_\Box \circ R = R_\Box$$

and  $P$  is closed under the operation

$$\Box X = \{x \in W : \forall y \in W (xR_\Box y \rightarrow y \in X)\}.$$

A  $\diamond$ -frame has the form  $\langle W, R, R_\diamond, P \rangle$ , where  $\langle W, R, P \rangle$  is again an intuitionistic frame,  $R_\diamond$  a binary relation on  $W$  satisfying the condition

$$R^{-1} \circ R_\diamond \circ R^{-1} = R_\diamond$$

and  $P$  is closed under

$$\diamond X = \{x \in W : \exists y \in X \ x R_\diamond y\}.$$

Finally, a  $\Box\diamond$ -frame is a structure  $\langle W, R, R_\Box, R_\diamond, P \rangle$  the unimodal reducts  $\langle W, R, R_\Box, P \rangle$  and  $\langle W, R, R_\diamond, P \rangle$  of which are  $\Box$ - and  $\diamond$ -frames, respectively. (To see why the intuitionistic and modal accessibility relations are connected by the conditions above the reader can construct in the standard way the canonical models for the logics under consideration. The important point here is that we take the Leibnizean definition of the truth-relation for the modal operators. Other definitions may impose different connecting conditions; see below.)

Given a  $\Box\diamond$ -frame  $\mathfrak{F} = \langle W, R, R_\Box, R_\diamond, P \rangle$ , it is easy to check that its *dual*

$$\mathfrak{F}^+ = \langle P, \cap, \cup, \rightarrow, \emptyset, \Box, \diamond \rangle$$

is a  $\Box\diamond$ -algebra. Conversely, for each  $\Box\diamond$ -algebra  $\mathfrak{A} = \langle A, \wedge, \vee, \rightarrow, \perp, \Box, \diamond \rangle$  we can define the *dual frame*

$$\mathfrak{A}_+ = \langle W, R, R_\Box, R_\diamond, P \rangle$$

by taking  $\langle W, R, P \rangle$  to be the dual of the Heyting algebra  $\langle A, \wedge, \vee, \rightarrow, \perp \rangle$  and putting

$$\nabla_1 R_\Box \nabla_2 \text{ iff } \forall a \in A \ (\Box a \in \nabla_1 \rightarrow a \in \nabla_2),$$

$$\nabla_1 R_\diamond \nabla_2 \text{ iff } \forall a \in A \ (a \in \nabla_2 \rightarrow \diamond a \in \nabla_1).$$

$\mathfrak{A}_+$  is a  $\Box\diamond$ -frame and, moreover,  $\mathfrak{A} \cong (\mathfrak{A}_+)^+$ . Using the standard technique of the model theory for classical modal and si-logics, one can show that a  $\Box\diamond$ -frame  $\mathfrak{F}$  is isomorphic to its bidual  $(\mathfrak{F}^+)_+$  iff  $\mathfrak{F} = \langle W, R, R_\Box, R_\diamond, P \rangle$  is *descriptive*, i.e.,  $\langle W, R, P \rangle$  is a descriptive intuitionistic frame and, for all  $x, y \in W$ ,

$$x R_\Box y \text{ iff } \forall X \in P \ (x \in \Box X \rightarrow y \in X),$$

$$x R_\diamond y \text{ iff } \forall X \in P \ (y \in X \rightarrow x \in \diamond X).$$

Thus we get the following completeness theorem.

**THEOREM 187.** *Every logic  $L \in \text{NExtIntK}_{\Box\diamond}$  is characterized by a suitable class of (descriptive)  $\Box\diamond$ -frames, e.g. by the class  $\{\mathfrak{A}_+ : \mathfrak{A} \models L\}$ .*

Similar results hold for logics in  $\text{NExtIntK}_\Box$  and  $\text{NExtIntK}_\diamond$ .

As usual, by a *Kripke frame* we understand a frame  $\langle W, R, R_\square, R_\diamond, P \rangle$  in which  $P$  consists of all  $R$ -cones; in this case we omit  $P$ . An intuitionistic modal logic  $L$  is  $\mathcal{D}$ -persistent if the underlying Kripke frame of each descriptive frame for  $L$  validates  $L$ . For example, **FS** as well as the logics

$$\mathbf{L}(k, l, m, n) = \mathbf{IntK}_{\square\Diamond} \oplus \Diamond^k \Box^l p \rightarrow \Box^m \Diamond^n p, \text{ for } k, l, m, n \geq 0$$

are  $\mathcal{D}$ -persistent and so Kripke complete (see Wolter and Zakharyashev [1997]). Descriptive frames validating **FS** satisfy the conditions

$$xR_\diamond y \rightarrow \exists z (yRz \wedge xR_\square z \wedge xR_\diamond z),$$

$$xR_\square y \rightarrow \exists z (xRz \wedge zR_\square y \wedge zR_\diamond y),$$

and those for  $\mathbf{L}(k, l, m, n)$  satisfy

$$xR_\diamond^k y \wedge xR_\square^m y \rightarrow \exists u (yR_\square^l u \wedge zR_\diamond^n u).$$

It follows, in particular, that **MIPC** is  $\mathcal{D}$ -persistent; its Kripke frames have the properties:  $R_\square$  is a quasi-order,  $R_\diamond = R_\square^{-1}$  and  $R_\square = R \circ (R_\square \cap R_\diamond)$ . On the contrary,  $\mathbf{I}^\Delta$  is not  $\mathcal{D}$ -persistent, although it is complete with respect to the class of Kripke frames  $\langle W, R, R_\square \rangle$  such that  $\langle W, R_\square \rangle$  is a frame for **GL** and  $R$  the reflexive closure of  $R_\square$ .

The next step in constructing duality theory of  $\mathbf{M}$ -algebras and  $\mathbf{M}$ -frames is to find relational counterparts of the algebraic operations of forming homomorphisms, subalgebras and direct products. Let  $\mathfrak{F} = \langle W, R, R_\square, R_\diamond, P \rangle$  be a  $\square\Diamond$ -frame and  $V$  a non-empty subset of  $W$  such that

$$\forall x \in V \forall y \in W (xR_\square y \vee xRy \rightarrow y \in V),$$

$$\forall x \in V \forall y \in W (xR_\diamond y \rightarrow \exists z \in V (xR_\diamond z \wedge yRz)).$$

Then  $\mathfrak{G} = \langle V, R \upharpoonright V, R_\square \upharpoonright V, R_\diamond \upharpoonright V, \{X \cap V : X \in P\} \rangle$  is also a  $\square\Diamond$ -frame which is called the *subframe of  $\mathfrak{F}$  generated by  $V$* . The former of the two conditions above is standard: it requires  $V$  to be upward closed with respect to both  $R$  and  $R_\square$ . However, the latter one does not imply that  $V$  is upward closed with respect to  $R_\diamond$ : the frame  $\mathfrak{G}$  in Fig. 17 is a generated subframe of  $\mathfrak{F}$ , although the set  $\{x, z\}$  is not an  $R_\diamond$ -cone in  $\mathfrak{F}$ . This is one difference from the standard (classical modal or intuitionistic) case. Another one arises when we define the relational analog of subalgebras.

Given  $\square\Diamond$ -frames  $\mathfrak{F} = \langle W, R, R_\square, R_\diamond, P \rangle$  and  $\mathfrak{G} = \langle V, S, S_\square, S_\diamond, Q \rangle$ , we say a map  $f$  from  $W$  onto  $V$  is a *reduction* of  $\mathfrak{F}$  to  $\mathfrak{G}$  if  $f^{-1}(X) \in P$  for every  $X \in Q$  and, for all  $x, y \in W$  and  $u \in V$ ,

$$xRy \text{ implies } f(x)Sf(y),$$

$$xR_\circ y \text{ implies } f(x)S_\circ f(y), \text{ for } \circ \in \{\square, \diamond\},$$

$$f(x)Su \text{ implies } \exists z \in f^{-1}(u) xRz,$$

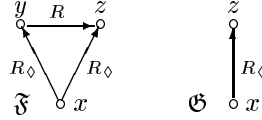


Figure 17.

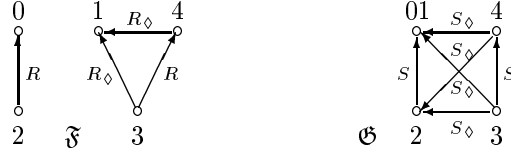


Figure 18.

$f(x)S_{\square}u$  implies  $\exists z \in f^{-1}(u) xR_{\square}z$ ,  
 $f(x)S_{\diamond}u$  implies  $\exists z \in W (xR_{\diamond}z \wedge uSf(z))$ .

Again, the last condition differs from the standard one: given  $f(x)S_{\diamond}f(y)$ , in general we do not have a point  $z$  such that  $xR_{\diamond}z$  and  $f(y) = f(z)$ , witness the map gluing 0 and 1 in the frame  $\mathfrak{F}$  in Fig. 18 and reducing it to  $\mathfrak{G}$ .

Note that both these concepts coincide with the standard ones in classical modal frames, where  $R$  and  $S$  are the diagonals. The relational counterpart of direct products—disjoint unions of frames—is defined as usual.

**THEOREM 188.**

- (i) *If  $\mathfrak{G}$  is the subframe of a  $\square\diamond$ -frame  $\mathfrak{F}$  generated by  $V$  then the map  $h$  defined by  $h(X) = X \cap V$ , for  $X$  an element in  $\mathfrak{F}^+$ , is a homomorphism from  $\mathfrak{F}^+$  onto  $\mathfrak{G}^+$ .*
- (ii) *If  $h$  is a homomorphism from a  $\square\diamond$ -algebra  $\mathfrak{A}$  onto a  $\square\diamond$ -algebra  $\mathfrak{B}$  then the map  $h_+$  defined by  $h_+(\nabla) = h^{-1}(\nabla)$ ,  $\nabla$  a prime filter in  $\mathfrak{B}$ , is an isomorphism from  $\mathfrak{B}_+$  onto a generated subframe of  $\mathfrak{A}_+$ .*
- (iii) *If  $f$  is a reduction of a  $\square\diamond$ -frame  $\mathfrak{F}$  to a  $\square\diamond$ -frame  $\mathfrak{G}$  then the map  $f^+$  defined by  $f^+(X) = f^{-1}(X)$ ,  $X$  an element in  $\mathfrak{G}^+$ , is an embedding of  $\mathfrak{G}^+$  into  $\mathfrak{F}^+$ .*
- (iv) *If  $\mathfrak{B}$  is a subalgebra of a  $\square\diamond$ -algebra  $\mathfrak{A}$  then the map  $f$  defined by  $f(\nabla) = \nabla \cap B$ ,  $\nabla$  a prime filter in  $\mathfrak{A}$  and  $B$  the universe of  $\mathfrak{B}$ , is a reduction of  $\mathfrak{A}_+$  to  $\mathfrak{B}_+$ .*

This duality can be used for proving various results on modal definability. For instance, a class  $\mathcal{C}$  of  $\square\diamond$ -frames is of the form  $\mathcal{C} = \{\mathfrak{F} : \mathfrak{F} \models \Gamma\}$ , for

some set  $\Gamma$  of  $\mathcal{L}_{\square\lozenge}$ -formulas, iff  $\mathcal{C}$  is closed under the formation of generated subframes, reducts, disjoint unions, and both  $\mathcal{C}$  and its complement are closed under the operation  $\mathfrak{F} \mapsto (\mathfrak{F}^+)_+$  (see Wolter and Zakharyashev [1997]). Moreover, one can extend Fine's Theorem connecting the first order definability and  $\mathcal{D}$ -persistence of classical modal logics to the intuitionistic modal case:

**THEOREM 189.** *If a logic  $L \in \text{NExtIntK}_{\square\lozenge}$  is characterized by an elementary class of Kripke frames then  $L$  is  $\mathcal{D}$ -persistent.*

These results may be regarded as a justification for the relational semantics introduced in this section. However, it is not the only possible one. For example, Božić and Došen [1984] impose a weaker condition on the connection between  $R$  and  $R_{\square}$  in  $\square$ -frames. Fisher Servi [1980] interprets **FS** in birelational Kripke frames of the form  $\langle W, R, S \rangle$  in which  $R$  is a partial order,  $R \circ S \subseteq S \circ R$ , and

$$xRy \wedge xSz \rightarrow \exists u (ySu \wedge zRu).$$

The intuitionistic connectives are interpreted by  $R$  and the truth-conditions for  $\square$  and  $\lozenge$  are defined as follows

$$\square X = \{x \in W : \forall y, z (xRySz \rightarrow z \in X)\},$$

$$\lozenge X = \{x \in W : \exists y \in X xSy\}.$$

In birelational frames for **MIPC**  $S$  is an equivalence relation and

$$xSyRz \rightarrow \exists u xRuSz.$$

These frames were independently introduced by L. Esakia who also established duality between them and “monadic Heyting algebras”.

There are two ways of investigating various properties of intuitionistic modal logics. One is to continue extending the classical methods to logics in  $\text{NExtIntK}_M$ . Another one uses those methods indirectly via embeddings of intuitionistic modal logics into classical ones. That such embeddings are possible was noticed by Shehtman [1979], Fischer Servi [1980, 1984], and Sotirov [1984]. Our exposition here follows Wolter and Zakharyashev [1997, 1999a]. For simplicity we confine ourselves only to considering the class  $\text{NExtIntK}_{\square}$  and refer the reader to the cited papers for information about more general embeddings.

Let  $T$  be the translation of  $\mathcal{L}_{\square}$  into  $\mathcal{L}_{\square_I\square}$  prefixing  $\square_I$  to every subformula of a given  $\mathcal{L}_{\square}$ -formula. Thus, we are trying to embed intuitionistic modal logics in  $\text{NExtIntK}_{\square}$  into classical bimodal logics with the necessity operators  $\square_I$  (of **S4**) and  $\square$ . Say that  $T$  embeds  $L \in \text{NExtIntK}_{\square}$  into  $M \in \text{NExt}(\mathbf{S4} \otimes \mathbf{K})$  (**S4** in  $\mathcal{L}_{\square_I}$  and **K** in  $\mathcal{L}_{\square}$ ) if, for every  $\varphi \in \mathcal{L}_{\square}$ ,

$$\varphi \in L \text{ iff } T(\varphi) \in M.$$



In this case  $M$  is called a *bimodal* (or BM-) *companion* of  $L$ .

For every logic  $M \in \text{NExt}(\mathbf{S4} \otimes \mathbf{K})$  put

$$\rho M = \{\varphi \in \mathcal{L}_\square : T(\varphi) \in M\},$$

and let  $\sigma$  be the map from  $\text{NExtIntK}_\square$  into  $\text{NExt}(\mathbf{S4} \otimes \mathbf{K})$  defined by

$$\sigma(\text{IntK}_\square \oplus \Gamma) = (\mathbf{Grz} \otimes \mathbf{K}) \oplus \mathit{mix} \oplus T(\Gamma),$$

where  $\Gamma \subseteq \mathcal{L}_\square$  and  $\mathit{mix} = \square_I \square \square_I p \leftrightarrow \square p$ . (The axiom  $\mathit{mix}$  reflects the condition  $R \circ R_\square \circ R = R_\square$  of  $\square$ -frames.) Then we have the following extension of the embedding results of Maksimova and Rybakov [1974], Blok [1976] and Esakia [1979a,b]:

THEOREM 190.

(i) *The map  $\rho$  is a lattice homomorphism from the lattice  $\text{NExt}(\mathbf{S4} \otimes \mathbf{K})$  onto  $\text{NExtIntK}_\square$  preserving decidability, Kripke completeness, tabularity and the finite model property.*

(ii) *Each logic  $\text{IntK}_\square \oplus \Gamma$  is embedded by  $T$  into any logic  $M$  in the interval*

$$(\mathbf{S4} \otimes \mathbf{K}) \oplus T(\Gamma) \subseteq M \subseteq (\mathbf{Grz} \otimes \mathbf{K}) \oplus \mathit{mix} \oplus T(\Gamma).$$

(iii) *The map  $\sigma$  is an isomorphism from the lattice  $\text{NExtIntK}_\square$  onto the lattice  $\text{NExt}(\mathbf{Grz} \otimes \mathbf{K}) \oplus \mathit{mix}$  preserving FMP and tabularity.*

Note that Fischer Servi [1980] used another generalization of the Gödel translation. She defined

$$T(\diamond\varphi) = \diamond T(\varphi),$$

$$T(\square\varphi) = \square_I \square T(\varphi)$$

and showed that this translation embeds  $\mathbf{FS}$  into the logic

$$(\mathbf{S4} \otimes \mathbf{K}) \oplus \diamond \square_I p \rightarrow \square_I \diamond p \oplus \diamond \diamond \square_I p \rightarrow \diamond \square_I \diamond p.$$

It is not clear, however, whether all extensions of  $\mathbf{FS}$  can be embedded into classical bimodal logics via this translation.

Let us turn now to completeness theory of intuitionistic modal logics. As to the standard systems  $\mathbf{I}^\Delta$ ,  $\mathbf{FS}$ , and  $\mathbf{MIPC}$ , their FMP can be proved by using (sometimes rather involved) filtration arguments; see Muravitskij [1981], Simpson [1994] and Grefe [1997], and Ono [1977], respectively. Further results based on the filtration method were obtained by Sotirov [1984] and Ono [1977]. However, in contrast to classical modal logic, only a few general completeness results covering interesting classes of intuitionistic modal logics are known. The proofs of the following two theorems are based on the translation into classical bimodal logics discussed above.

**THEOREM 191.** *Suppose that a si-logic  $\mathbf{Int} + \Gamma$  has one of the properties: decidability, Kripke completeness, FMP. Then the logics  $\mathbf{IntK}_\square \oplus \Gamma$  and  $\mathbf{IntK}_\square \oplus \Gamma \oplus \Box p \rightarrow p$  also have the same property.*

**Proof.** It suffices to show that there is a BM-companion of each of these systems satisfying the corresponding property. Notice that

$$\rho((\mathbf{S4} \oplus T(\Gamma)) \otimes \mathbf{K}) = \mathbf{IntK}_\square \oplus \Gamma,$$

$$\rho((\mathbf{S4} \oplus T(\Gamma)) \otimes (\mathbf{K} \oplus \Box p \rightarrow p)) = \mathbf{IntK}_\square \oplus \Gamma \oplus \Box p \rightarrow p.$$

So it remains to use the fact that if  $\mathbf{Int} + \Gamma$  has one of the properties under consideration then its smallest modal companion  $\mathbf{S4} \oplus T(\Gamma)$  has this property as well (Table 7), and if  $L_1, L_2$  are unimodal logics having one of those properties then the fusion  $L_1 \otimes L_2$  also enjoys the same property (Theorem 111). ■

Such a simple reduction to known results in classical modal logic is not available for logics containing  $\mathbf{IntK4}_\square = \mathbf{IntK}_\square \oplus \Box p \rightarrow \Box \Box p$ . However, by extending Fine's [1974] method of maximal points to bimodal companions of extensions of  $\mathbf{IntK4}_\square$  Wolter and Zakharyashev [1999a] proved the following:

**THEOREM 192.** *Suppose  $L \supseteq \mathbf{IntK4}_\square$  has a  $\mathcal{D}$ -persistent BM-companion  $M \supseteq (\mathbf{S4} \otimes \mathbf{K4}) \oplus \mathit{mix}$  whose Kripke frames are closed under the formation of substructures. Then*

- (i) *for every set  $\Gamma$  of intuitionistic negation and disjunction free formulas,  $L \oplus \Gamma$  has FMP;*
- (ii) *for every set  $\Gamma$  of intuitionistic disjunction free formulas and every  $n \geq 1$ ,*

$$L \oplus \Gamma \oplus \bigvee_{i=0}^n (p_i \rightarrow \bigvee_{j \neq i} p_j)$$

*has the finite model property.*

One can use this result to show that the following (and many other) intuitionistic modal logics enjoy FMP:

- (1)  $\mathbf{IntK4}_\square$ ;
- (2)  $\mathbf{IntS4}_\square = \mathbf{IntK4}_\square \oplus \Box p \rightarrow p$  ( $R_\square$  is reflexive);
- (3)  $\mathbf{IntS4.3}_\square = \mathbf{IntS4}_\square \oplus \Box(\Box p \rightarrow q) \vee \Box(\Box q \rightarrow p)$  ( $R_\square$  is reflexive and connected);
- (4)  $\mathbf{IntK4}_\square \oplus p \vee \Box \neg \Box p$  ( $R_\square$  is symmetrical);

(5)  $\mathbf{IntK4}_\Box \oplus \Box p \vee \Box \neg \Box p$  ( $R_\Box$  is Euclidean);

(6)  $\mathbf{IntK4}_\Box \oplus \Box p \vee \neg \Box p$  ( $xRy \wedge xR_\Box z \rightarrow yR_\Box z$ ).

We conclude this section with some remarks on lattices of intuitionistic modal logics. Wolter [1997e] uses duality theory to study splittings of lattices of intuitionistic modal logics. For example, he showed that each finite rooted frame splits  $\mathbf{NExt}(L \oplus \Box^{\leq n} p \rightarrow \Box^{n+1} p)$ , for  $L = \mathbf{IntK}_\Box$  and  $L = \mathbf{FS}$ , and each  $R_\Box$ -cycle free finite rooted frame splits the lattices of extensions of  $\mathbf{IntK}_\Box$  and  $\mathbf{FS}$ . No positive results are known, however, for the lattice  $\mathbf{NExtIntK}_\Diamond$ . In fact, the behavior of  $\Diamond$ -frames is quite different from that of frames for  $\mathbf{FS}$ . For instance, in classical modal logic we have  $\mathbf{RG}\mathcal{F} = \mathbf{GR}\mathcal{F}$ , for each class of frames (or even  $\Box$ -frames)  $\mathcal{F}$ , where  $\mathbf{G}$  and  $\mathbf{R}$  are the operations of forming generated subframes and reducts, respectively. But this does not hold for  $\Diamond$ -frames. More precisely, there exists a finite  $\Diamond$ -frame  $\mathfrak{G}$  such that  $\mathbf{RG}\{\mathfrak{G}\} \not\cong \mathbf{GR}\{\mathfrak{G}\}$ . In other terms, the variety of modal algebras for  $\mathbf{K}$  has the *congruence extension property* (i.e., each congruence of a subalgebra of a modal algebra can be extended to a congruence of the algebra itself) but this is not the case for the variety of  $\Diamond$ -algebras.

Vakarelov [1981, 1985] and Wolter [1997e] investigate how logics having  $\mathbf{Int}$  as their non-modal fragment are located in the lattices of intuitionistic modal logics. It turns out, for instance, that in  $\mathbf{NExtIntK}_\Diamond$  the inconsistent logic has a continuum of immediate predecessors all of which have  $\mathbf{Int}$  as their non-modal fragment, but no such logic exists in the lattice of extensions of  $\mathbf{IntK}_\Box$ .

For a recent methodological approach to combining logics, see [Gabbay, 1988].

## 4 ALGORITHMIC PROBLEMS

All algorithmic results considered in the previous sections were positive: we presented concrete procedures for deciding whether an arbitrary given formula belongs to a given logic in some class or whether it axiomatizes a logic with a certain property. What is the complexity of those decision algorithms? Do there exist undecidable calculi<sup>18</sup> and properties? These are the main questions we address in this chapter.

### 4.1 Undecidable calculi

The first undecidable modal and si-calculi were constructed by Thomason [1975c] (polymodal and unimodal), Isard [1977] (unimodal) and Shehtman

<sup>18</sup>By a calculus we mean a logic with finitely many axioms (inference rules in our case are fixed).

[1978c] (superintuitionistic). However, we begin with the very simple example of [Shehtman 1982] which is a modal reformulation of the undecidable associative calculus  $T$  of [Tseitin 1958]. The axioms of  $T$  are

$$\begin{aligned} ac = ca, & & ad = da, \\ bc = cb, & & bd = db, \\ edb = be, & & eca = ae, \\ abac = abacc. \end{aligned}$$

The reader will notice immediately an analogy between them and the axioms of the following modal calculus with five necessity operators:

$$\begin{aligned} L = \mathbf{K}_5 \oplus & \square_1 \square_3 p \leftrightarrow \square_3 \square_1 p \oplus \square_1 \square_4 p \leftrightarrow \square_4 \square_1 p \oplus \\ & \square_2 \square_3 p \leftrightarrow \square_3 \square_2 p \oplus \square_2 \square_4 p \leftrightarrow \square_4 \square_2 p \oplus \\ & \square_5 \square_4 \square_2 p \leftrightarrow \square_2 \square_5 p \oplus \square_5 \square_3 \square_1 p \leftrightarrow \square_1 \square_5 p \oplus \\ & \square_1 \square_2 \square_1 \square_3 p \leftrightarrow \square_1 \square_2 \square_1 \square_3 \square_3 p. \end{aligned}$$

Moreover, it is not hard to see that words  $x, y$  in the alphabet  $\{a, b, c, d, e\}$  are equivalent in  $T^{19}$  iff  $f(x)p \leftrightarrow f(y)p \in \mathbf{K}_5$ , where  $f$  is the natural one-to-one correspondence between such words and modalities in language  $\{\square_1, \dots, \square_5\}$  under which, for instance,  $f(cadedb) = \square_3 \square_1 \square_4 \square_5 \square_4 \square_2$ . It follows immediately that  $L$  is undecidable. Using the undecidable associative calculus of Matiyasevich [1967], one can construct in the same way an undecidable bimodal calculus having three reductions of modalities as its axioms. It is unknown whether there is an undecidable unimodal calculus axiomatizable by reductions of modalities.

Another simple way of proving undecidability, known as the *domino* or *tiling technique*, was suggested by Harel [1983]. It is particularly useful in the case of multi-dimensional modal logics, say Cartesian products.

*Tiles* can be thought of as 4-tuples of colours

$$t = \langle left(t), right(t), up(t), down(t) \rangle.$$

A finite set  $T$  of tiles is said to *tile*  $\mathbb{N} \times \mathbb{N}$  if there is a map  $\tau : \mathbb{N} \times \mathbb{N} \mapsto T$  such that for all  $i, j \in \mathbb{N}$ ,

- $up(\tau(i, j)) = down(\tau(i, j + 1))$  and
- $right(\tau(i, j)) = left(\tau(i + 1, j))$ .

If we think of a tile as a physical  $1 \times 1$ -square with colours along its four edges, then a tiling of  $\mathbb{N} \times \mathbb{N}$  is just a way of placing an infinite number of

---

<sup>19</sup>I.e., they can be obtained from each other by a finite number of transformations of the form  $w_1 w w_2 \rightarrow w_1 v w_2$ , where  $w = v$  or  $v = w$  is an axiom of  $T$ .

tiles, each of a type from  $T$ , together to cover the first quarter of the infinite plane, with no rotation of the tiles allowed and the colours on adjacent edges of adjacent tiles matching.

The *tiling problem* for  $\mathbb{N} \times \mathbb{N}$  is formulated as follows: “given a finite set  $T$  of tiles, does  $T$  tile  $\mathbb{N} \times \mathbb{N}$ ?” Robinson [1971] proved that this problem is undecidable (in fact, co-r.e.-complete).

We will demonstrate the use of tiling to show the undecidability of the logic  $(\mathbf{K} \times \mathbf{K})_u$ , i.e., the square of  $\mathbf{K}$  (with boxes  $\Box$  and  $\sqcup$ ) extended with the universal modality  $\square$  (see Section 2.2); this result is due to Spaan [1993].

Given a finite set  $T$  of tiles, construct a formula  $\varphi_T$  as the conjunction of the following formulas:

$$\begin{aligned} & \square \bigvee_{t \in T} p_t, \\ & \square \bigwedge_{t \neq t'} \neg(p_t \wedge p_{t'}), \\ & \square \bigwedge_{t \in T} (p_t \rightarrow \bigvee_{up(t)=down(t')} \square p_{t'}), \\ & \square \bigwedge_{t \in T} (p_t \rightarrow \bigvee_{right(t)=left(t')} \square p_{t'}), \\ & \square(\Diamond \top \wedge \Diamond \top). \end{aligned}$$

It is easily seen (see e.g. [Spaan 1993] or [Marx 1999]) that  $\varphi_T$  is satisfiable in the product of two frames iff  $T$  tiles  $\mathbb{N} \times \mathbb{N}$ . It follows that  $(\mathbf{K} \times \mathbf{K})_u$  is undecidable.

Thomason’s simulation and the undecidable polymodal calculi mentioned above provide us with examples of undecidable calculi in  $\text{NExt}\mathbf{K}$ . However, to find axioms of undecidable unimodal calculi with transitive frames, as well as undecidable si-calculi, a more sophisticated construction is required.

Instead of associative calculi, let us use now Minsky machines with two tapes (or register machines with two registers). A *Minsky machine* is a finite set (program) of instructions for transforming triples  $\langle s, m, n \rangle$  of natural numbers, called *configurations*. The intended meaning of the current configuration  $\langle s, m, n \rangle$  is as follows:  $s$  is the number (label) of the current machine state and  $m, n$  represent the current state of information. Each instruction has one of the four possible forms:

$$\begin{aligned} & s \rightarrow \langle t, 1, 0 \rangle, \quad s \rightarrow \langle t, 0, 1 \rangle, \\ & s \rightarrow \langle t, -1, 0 \rangle (\langle t', 0, 0 \rangle), \quad s \rightarrow \langle t, 0, -1 \rangle (\langle t', 0, 0 \rangle). \end{aligned}$$

The last of them, for instance, means: transform  $\langle s, m, n \rangle$  into  $\langle t, m, n - 1 \rangle$  if  $n > 0$  and into  $\langle t', m, n \rangle$  if  $n = 0$ . For a Minsky machine  $\mathbf{P}$ , we shall write  $\mathbf{P} : \langle s, m, n \rangle \rightarrow \langle t, k, l \rangle$  if starting with  $\langle s, m, n \rangle$  and applying the instructions in  $\mathbf{P}$ , in finitely many steps (possibly, in 0 steps) we can reach  $\langle t, k, l \rangle$ .

We shall use the well known fact (see e.g. [Mal’cev 1970]) that the following *configuration problem* is undecidable: given a program  $\mathbf{P}$  and configurations  $\langle s, m, n \rangle, \langle t, k, l \rangle$ , determine whether  $\mathbf{P} : \langle s, m, n \rangle \rightarrow \langle t, k, l \rangle$ .

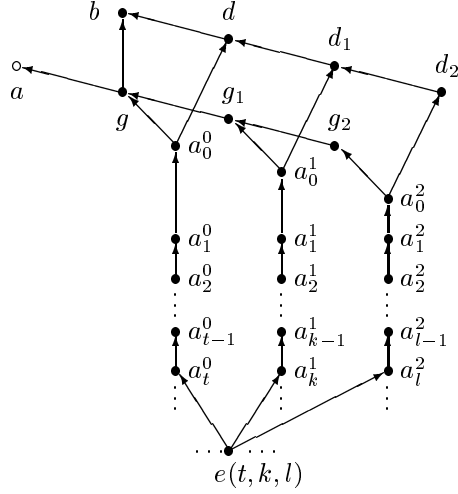


Figure 19.

With every program  $\mathbf{P}$  and configuration  $\langle s, m, n \rangle$  we associate the transitive frame  $\mathfrak{F}$  depicted in Fig. 19. Its points  $e(t, k, l)$  represent configurations  $\langle t, k, l \rangle$  such that  $\mathbf{P} : \langle s, m, n \rangle \rightarrow \langle t, k, l \rangle$ ;  $e(t, k, l)$  sees the points  $a_t^0, a_k^1, a_l^2$  representing the components of  $\langle t, k, l \rangle$ . The following variable free formulas characterize points in  $\mathfrak{F}$  in the sense that each of these formulas, denoted by Greek letters with subscripts and/or superscripts, is true in  $\mathfrak{F}$  only at the point denoted by the corresponding Roman letter with the same subscript and/or superscript:

$$\begin{aligned} \alpha &= \Diamond \top \wedge \Box \Diamond \top, \quad \beta = \Box \perp, \quad \gamma = \Diamond \alpha \wedge \Diamond \beta \wedge \neg \Diamond^2 \beta, \\ \delta &= \neg \gamma \wedge \Diamond \beta \wedge \neg \Diamond^2 \beta, \quad \delta_1 = \Diamond \delta \wedge \neg \Diamond^2 \delta, \quad \delta_2 = \Diamond \delta_1 \wedge \neg \Diamond^2 \delta_1, \\ \gamma_1 &= \Diamond \gamma \wedge \neg \Diamond^2 \gamma \wedge \neg \Diamond \delta, \quad \gamma_2 = \Diamond \gamma_1 \wedge \neg \Diamond^2 \gamma_1 \wedge \neg \Diamond \delta, \\ \alpha_0^0 &= \Diamond \gamma \wedge \Diamond \delta \wedge \neg \Diamond^2 \gamma \wedge \neg \Diamond^2 \delta, \\ \alpha_0^1 &= \Diamond \gamma_1 \wedge \Diamond \delta_1 \wedge \neg \Diamond^2 \gamma_1 \wedge \neg \Diamond^2 \delta_1, \\ \alpha_0^2 &= \Diamond \gamma_2 \wedge \Diamond \delta_2 \wedge \neg \Diamond^2 \gamma_2 \wedge \neg \Diamond^2 \delta_2, \\ \alpha_{j+1}^i &= \Diamond \alpha_j^i \wedge \neg \Diamond^2 \alpha_j^i \wedge \bigwedge_{i \neq k} \neg \Diamond \alpha_0^k, \end{aligned}$$

where  $i \in \{0, 1, 2\}$ ,  $j \geq 0$ . The formulas characterizing  $e(t, k, l)$  are denoted by  $\epsilon(t, \alpha_k^1, \alpha_l^2)$ , where

$$\epsilon(t, \varphi, \psi) = \bigwedge_{i=0}^t \Diamond \alpha_i^0 \wedge \neg \Diamond \alpha_{t+1}^0 \wedge \Diamond \varphi \wedge \neg \Diamond^2 \varphi \wedge \Diamond \psi \wedge \neg \Diamond^2 \psi.$$

We require also formulas characterizing not only fixed but arbitrary configurations:

$$\begin{aligned}\pi_1 &= (\Diamond\alpha_0^1 \vee \alpha_0^1) \wedge \neg\Diamond\alpha_0^0 \wedge \neg\Diamond\alpha_0^2 \wedge p_1 \wedge \neg\Diamond p_1, \\ \pi_2 &= \Diamond\alpha_0^1 \wedge \neg\Diamond\alpha_0^0 \wedge \neg\Diamond\alpha_0^2 \wedge \Diamond p_1 \wedge \neg\Diamond^2 p_1, \\ \tau_1 &= (\Diamond\alpha_0^2 \vee \alpha_0^2) \wedge \neg\Diamond\alpha_0^0 \wedge \neg\Diamond\alpha_0^1 \wedge p_2 \wedge \neg\Diamond p_2, \\ \tau_2 &= \Diamond\alpha_0^2 \wedge \neg\Diamond\alpha_0^0 \wedge \neg\Diamond\alpha_0^1 \wedge \Diamond p_2 \wedge \neg\Diamond^2 p_2.\end{aligned}$$

Now we are fully equipped to simulate the behavior of Minsky machines by means of modal formulas. Let us consider for simplicity only tense logics and observe that  $\mathfrak{F}$  satisfies the condition

$$\forall x\forall y\exists z (xRzR^{-1}y \vee xR^{-1}zRy \vee xRy \vee xR^{-1}y \vee x = y).$$

So, for every valuation in  $\mathfrak{F}$ , a formula  $\varphi$  is true at some point in  $\mathfrak{F}$  iff the formula

$$\bigcirc\varphi = \Diamond\Diamond^{-1}\varphi \vee \Diamond^{-1}\Diamond\varphi \vee \Diamond\varphi \vee \Diamond^{-1}\varphi \vee \varphi$$

is true at all points in  $\mathfrak{F}$ , i.e., the modal operator  $\bigcirc$  can be understood as “omniscience”. Let  $\chi$  be a formula which is refuted in  $\mathfrak{F}$  and does not contain  $p_1$  and  $p_2$ . With each instruction  $I$  in  $\mathbf{P}$  we associate a formula  $AxI$  by taking:

$$AxI = \neg\chi \wedge \bigcirc\epsilon(t, \pi_1, \tau_1) \rightarrow \neg\chi \wedge \bigcirc\epsilon(t', \pi_2, \tau_1)$$

if  $I$  has the form  $t \rightarrow \langle t', 1, 0 \rangle$ ,

$$AxI = \neg\chi \wedge \bigcirc\epsilon(t, \pi_1, \tau_1) \rightarrow \neg\chi \wedge \bigcirc\epsilon(t', \pi_1, \tau_2)$$

if  $I$  is  $t \rightarrow \langle t', 0, 1 \rangle$ ,

$$\begin{aligned}AxI &= (\neg\chi \wedge \bigcirc\epsilon(t, \pi_2, \tau_1) \rightarrow \neg\chi \wedge \bigcirc\epsilon(t', \pi_1, \tau_1)) \wedge \\ &\quad (\neg\chi \wedge \bigcirc\Diamond\epsilon(t, \alpha_0^1, \tau_1) \rightarrow \neg\chi \wedge \bigcirc\epsilon(t'', \alpha_0^1, \tau_1))\end{aligned}$$

if  $I$  is  $t \rightarrow \langle t', -1, 0 \rangle (\langle t'', 0, 0 \rangle)$ ,

$$\begin{aligned}AxI &= (\neg\chi \wedge \bigcirc\epsilon(t, \pi_1, \tau_2) \rightarrow \neg\chi \wedge \bigcirc\epsilon(t', \pi_1, \tau_1)) \wedge \\ &\quad (\neg\chi \wedge \bigcirc\epsilon(t, \pi_1, \alpha_0^2) \rightarrow \neg\chi \wedge \bigcirc\epsilon(t'', \pi_1, \alpha_0^2))\end{aligned}$$

if  $I$  is  $t \rightarrow \langle t', 0, -1 \rangle (\langle t'', 0, 0 \rangle)$ . The formula simulating  $\mathbf{P}$  as a whole is

$$AxP = \bigwedge_{I \in \mathbf{P}} AxI.$$

Now, by induction on the length of computations and using the frame  $\mathfrak{F}$  in Fig. 19 one can show that for every program  $\mathbf{P}$  and configurations  $\langle s, m, n \rangle$ ,  $\langle t, k, l \rangle$ , we have  $\mathbf{P} : \langle s, m, n \rangle \rightarrow \langle t, k, l \rangle$  iff

$$\neg\chi \wedge \bigcirc\epsilon(s, \alpha_m^1, \alpha_n^2) \rightarrow \neg\chi \wedge \bigcirc\epsilon(t, \alpha_k^1, \alpha_l^2) \in \mathbf{K4}.t \oplus AxP.$$

Thus, if the configuration problem is undecidable for  $\mathbf{P}$  then the tense calculus  $\mathbf{K4.t} \oplus AxP$  is undecidable too. In the same manner (but using somewhat more complicated frames and formulas) one can construct undecidable calculi in  $\mathbf{NExtK4}$  and even  $\mathbf{ExtInt}$ ; for details consult [Chagrova, 1991] and [Chagrov and Zakharyashev, 1997]. The following table presents some "quantitative characteristics" of known undecidable calculi in various classes of logics. Its first line, for instance, means that there is an undecidable si-calculus with axioms in 4 variables and the derivability problem in it is undecidable in the class of formulas in 2 variables; = means that the number of variables is optimal, and  $\leq$  indicates that the optimal number is still unknown.

Class of logics	The number of variables in	
	undecidable calculi	separated formulas
$\mathbf{ExtInt}$	$\leq 4, \geq 2$	$= 2$
$\mathbf{NExtS4}$	$\leq 3, \geq 2$	$= 1$
$\mathbf{ExtS4}$	$\leq 3$	$= 1$
$\mathbf{NExtGL}$	$= 1$	$= 1$
$\mathbf{ExtGL}$	$= 1$	$= 1$
$\mathbf{ExtS}$	$= 1$	$= 1$
$\mathbf{NExtK4}$	$= 1$	$= 0$
$\mathbf{ExtK4}$	$= 1$	$= 0$

These observations follow from [Anderson, 1972; Chagrov, 1994; Sobolev, 1977a] and [Zakharyashev, 1997a]. Say that a formula  $\psi$  is *undecidable* in  $(\mathbf{N})\mathbf{ExtL}$  if no algorithm can determine for an arbitrary given  $\varphi$  whether  $\psi \in L + \varphi$  (respectively,  $\psi \in L \oplus \varphi$ ). For example, formulas in one variable, the axioms of  $\mathbf{BW}_n$  and  $\mathbf{BD}_n$  are decidable in  $\mathbf{ExtInt}$ . On the other hand, there are purely implicative undecidable formulas in  $\mathbf{ExtInt}$ , and

$$\neg(p \wedge q) \vee \neg(\neg p \wedge q) \vee \neg(p \wedge \neg q) \vee \neg(\neg p \wedge \neg q)$$

is the shortest known undecidable formula in this class. Here are some modal examples: the formula  $\Box(\Box^2 \perp \rightarrow \Box p \vee \Box \neg p)$  is undecidable in  $\mathbf{NExtGL}$ ,  $\Box^+ \neg \Box^+ p \vee \Box^+ \neg \Box^+ \neg \Box^+ p$  in  $\mathbf{ExtS}$ ,  $\perp$  in  $\mathbf{ExtK4}$  and  $\mathbf{NExtK4.t}$ ; in  $\mathbf{NExtK}$  and  $\mathbf{NExtK4.t}$  undecidable is the conjunction of axioms of any consistent tabular logic in these classes. However, no non-trivial criteria are known for a formula to be decidable; it is unclear also whether one can effectively recognize the decidability of formulas in the classes  $\mathbf{ExtInt}$ ,  $(\mathbf{N})\mathbf{ExtS4}$ ,  $(\mathbf{N})\mathbf{ExtGL}$ ,  $\mathbf{ExtS}$ ,  $(\mathbf{N})\mathbf{ExtK4}$ .

#### 4.2 Admissibility and derivability of inference rules

Another interesting algorithmic problem for a logic  $L$  is to determine whether an arbitrary given inference rule  $\varphi_1, \dots, \varphi_n / \varphi$  is *derivable* in  $L$ , i.e.,  $\varphi$  is



derivable in  $L$  from the assumptions  $\varphi_1, \dots, \varphi_n$ , and whether it is *admissible* in  $L$ , i.e., for every substitution  $\mathbf{s}$ ,  $\varphi \mathbf{s} \in L$  whenever  $\varphi_1 \mathbf{s}, \dots, \varphi_n \mathbf{s} \in L$ . (Note that derivability depends on the postulated inference rules in  $L$ , while admissibility depends only on the set of formulas in  $L$ .) Admissible and derivable rules are used for simplifying the construction of derivations. Derivable rules, like the well known rule of syllogism

$$\frac{\varphi \rightarrow \psi, \psi \rightarrow \chi}{\varphi \rightarrow \chi},$$

may replace some fragments of fixed length in derivations, thereby shortening them linearly. Admissible rules in principle may reduce derivations more drastically. Since  $\varphi \in L$  iff the rule  $\top/\varphi$  is derivable (or admissible) in  $L$ , the derivability and admissibility problems for inference rules may be regarded as generalizations of the decidability problem.

If the only postulated rules in  $L$  are substitution and modus ponens, the Deduction Theorem reduces the derivability problem for inference rules in  $L$  to its decidability:

$$\frac{\varphi_1, \dots, \varphi_n}{\psi} \text{ is derivable in } L \text{ iff } \varphi_1 \wedge \dots \wedge \varphi_n \rightarrow \psi \in L.$$

However, if the rule of necessitation  $\varphi/\Box\varphi$  is also postulated in  $L$ , we have only

$$\frac{\varphi_1, \dots, \varphi_n}{\psi} \text{ is derivable in } L \text{ iff } \varphi_1, \dots, \varphi_n \vdash_L^* \psi.$$

For  $n$ -transitive  $L$  this is equivalent to  $\Box^{\leq n}(\varphi_1 \wedge \dots \wedge \varphi_n) \rightarrow \psi \in L$ , and so the derivability problem for inference rules in  $n$ -transitive logics is decidable iff the logics themselves are decidable. In general, in view of the existential quantifier in Theorem 1, the situation is much more complicated.

Notice first that similarly to Harrop's Theorem, a sufficient condition for the derivability problem to be decidable in a calculus is its global FMP (see Section 1.5). Thus we have

**THEOREM 193.** *The derivability problem for inference rules in  $\mathbf{K}$ ,  $\mathbf{T}$ ,  $\mathbf{D}$ ,  $\mathbf{KB}$  is decidable.*

Moreover, sometimes we can obtain an upper bound for the parameter  $m$  in the Deduction Theorem, which also ensures the decidability of the derivability problem for inference rules. One can prove, for instance, that for  $\mathbf{K}$  it is enough to take  $m = 2^{|\text{Sub}\varphi \cup \text{Sub}\psi|}$ . In general, however, the derivability problem for inference rules in a logic  $L$  turns out to be more complex than the decidability problem for  $L$ . (Recall, by the way, that there are logics with FMP but not global FMP.)

**THEOREM 194** (Spaan 1993). *There is a decidable calculus in  $\text{NExt}\mathbf{K}$  the derivability problem for inference rules in which is undecidable.*

Spaan proves this result by simulating in  $\vdash_L^*$ ,  $L$  the decidable logic defined below, the tiling problem for  $\mathbb{N} \times \mathbb{N}$ . The logic  $L$  is surprisingly simple:

$$L = \mathbf{Alt}_2 \oplus \bigwedge_{1 \leq i \leq 4} \diamond \diamond p_i \rightarrow \bigvee_{1 \leq i < j \leq 4} \diamond \diamond (p_i \wedge p_j).$$

It is a subframe logic, so it is  $\mathcal{D}$ -persistent and has FMP (because  $\mathbf{Alt}_2 \subseteq L$ ; see Theorem 22 and Proposition 59). Note also that the bimodal logic  $L_u$  (see Section 2.2) is a complete and elementary subframe logic which is undecidable because  $\vdash_L^*$  is undecidable. Using this observation one can construct a unimodal subframe logic in  $\mathbf{NExtK}$  with the same properties.

Let us turn now to the admissibility problem. It is not hard to see that the rules

$$\frac{(\neg \neg p \rightarrow p) \rightarrow p \vee \neg p}{\neg p \vee \neg \neg p} \quad \text{and} \quad \frac{\neg p \rightarrow q \vee r}{(\neg p \rightarrow q) \vee (\neg p \rightarrow r)}$$

are admissible but not derivable in  $\mathbf{Int}$  and  $\diamond p \wedge \diamond \neg p / \perp$  is admissible but not derivable in any extension of **S4.3** save those containing  $\Box \diamond p \rightarrow \diamond \Box p$ , in which it is derivable. (Recall that a logic  $L$  is said to be *structurally complete* if every admissible inference rule in  $L$  is derivable in  $L$ . We have just seen that  $\mathbf{Int}$  as well as **S4.3** are not structurally complete. For more information on structural completeness see e.g. [Tsytkin 1978, 1987] and [Rybakov 1995].) The following result strengthens Fine's [1971] Theorem according to which all logics in  $\mathbf{ExtS4.3}$  are decidable.

**THEOREM 195** (Rybakov 1984a). *The admissibility problem for inference rules is decidable in every logic containing **S4.3**.*

An impetus for investigations of admissible inference rules in various logics was given by Friedman's [1975] problem 40 asking whether one can effectively recognize admissible rules in  $\mathbf{Int}$ . This problem turned out to be closely connected to the admissibility problem in suitable modal logics. We demonstrate this below for the logic **GL** following [Rybakov 1987, 1989].

First we show that dealing with logics in  $\mathbf{NExtK}$ , it is sufficient to consider inference rules of a rather special form. Let  $\varphi(q_1, \dots, q_{2n+2})$  be a formula containing no  $\Box$  and  $\diamond$  and represented in the full disjunctive normal form. Say that an inference rule is *reduced* if it has the form

$$\varphi(p_0, \dots, p_n, \diamond p_0, \dots, \diamond p_n) / p_0.$$

**THEOREM 196.** *For every rule  $\varphi/\psi$  one can effectively construct a reduced rule  $\varphi'/\psi'$  such that  $\varphi/\psi$  is admissible in a logic  $L \in \mathbf{NExtK}$  iff  $\varphi'/\psi'$  is admissible in  $L$ .*

**Proof.** Observe first that if  $\varphi$  and  $\psi$  do not contain  $p$  then  $\varphi/\psi$  is admissible in  $L$  iff  $\varphi \wedge (\psi \leftrightarrow p) / p$  is admissible in  $L$ . So we can consider only rules of

the form  $\varphi/p_0$ . Besides, without loss of generality we may assume that  $\varphi$  does not contain  $\Box$ . With every non-atomic subformula  $\chi$  of  $\varphi$  we associate the new variable  $p_\chi$ . For convenience we also put  $p_\chi = p_i$  if  $\chi = p_i$  and  $p_\chi = \perp$  if  $\chi = \perp$ . We show now that the rule

$$p_\varphi \wedge \bigwedge \{p_\chi \leftrightarrow p_{\chi_1} \odot p_{\chi_2} : \chi = \chi_1 \odot \chi_2 \in \mathbf{Sub}\varphi, \odot \in \{\wedge, \vee, \rightarrow\}\} \wedge \\ \bigwedge \{p_\chi \leftrightarrow \Diamond p_{\chi_1} : \chi = \Diamond \chi_1 \in \mathbf{Sub}\varphi\} / p_0$$

is admissible in  $L$  iff  $\varphi/p_0$  is admissible in  $L$ . For brevity we denote the antecedent of that rule by  $\varphi''$ .

( $\Rightarrow$ ) Since every substitution instance of  $\varphi''/p_0$  is admissible in  $L$ , the rule  $\varphi \wedge \bigwedge_{\chi \in \mathbf{Sub}\varphi} (\chi \leftrightarrow \chi) / p_0$  and so  $\varphi/p_0$  are also admissible in  $L$ .

( $\Leftarrow$ ) Suppose  $\varphi/p_0$  is admissible in  $L$  and  $\varphi''s$  is in  $L$ , for some substitution  $s = \{\alpha_\chi/p_\chi : \chi \in \mathbf{Sub}\varphi\}$ . By induction on the construction of  $\chi$  one can readily show that  $\alpha_\chi \leftrightarrow \chi s \in L$ . Therefore,  $\alpha_\varphi \leftrightarrow \varphi s \in L$ . Since  $\varphi''s \in L$ , we must have  $p_\varphi s = \alpha_\varphi \in L$ , from which  $\varphi s \in L$  and so  $p_0 s \in L$ . Thus  $\varphi''/p_0$  is admissible in  $L$ .

The rule  $\varphi''/p_0$  is not reduced, but it is easy to make it so simply by representing  $\varphi''$  in its full disjunctive normal form  $\varphi'$ , treating subformulas  $\Diamond p_i$  as variables.  $\blacksquare$

From now on we will deal with only reduced rules different from  $\perp/p_0$  (which is clearly admissible in any logic). Let  $\bigvee_j \varphi_j/p_0$  be a reduced rule in which every disjunct  $\varphi_j$  is the conjunction of the form

$$(17) \quad \neg_0 p_0 \wedge \cdots \wedge \neg_m p_m \wedge \neg^0 \Diamond p_0 \wedge \cdots \wedge \neg^m \Diamond p_m,$$

where each  $\neg_i$  and  $\neg^j$  is either blank or  $\neg$ . We will identify such conjunctions with the sets of their conjuncts. Now, given a non-empty set  $W$  of conjunctions of the form (17), we define a frame  $\mathfrak{F} = \langle W, R \rangle$  and a model  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  by taking

$$\varphi_i R \varphi_j \quad \text{iff} \quad \forall k \in \{0, \dots, m\} (\neg \Diamond p_k \in \varphi_i \rightarrow \neg \Diamond p_k \in \varphi_j \wedge \neg p_k \in \varphi_j) \wedge \\ \exists k \in \{0, \dots, m\} (\neg \Diamond p_k \in \varphi_j \wedge \Diamond p_k \in \varphi_i), \\ \mathfrak{V}(p_k) = \{\varphi_i \in W : p_k \in \varphi_i\}.$$

It should be clear that  $\mathfrak{F}$  is finite, transitive and irreflexive.

**THEOREM 197.** *A reduced rule  $\bigvee_j \varphi_j/p_0$  is not admissible in  $\mathbf{GL}$  iff there is a model  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  defined as above on a set  $W$  of conjunctions of the form (17) and such that*

- (i)  $\neg p_0 \in \varphi_i$  for some  $\varphi_i \in W$ ;
- (ii)  $\varphi_i \models \varphi_i$  for every  $\varphi_i \in W$ ;

- (iii) for every antichain  $\mathfrak{a}$  in  $\mathfrak{F}$  there is  $\varphi_j \in W$  such that, for every  $k \in \{0, \dots, m\}$ ,  $\varphi_j \models \Diamond p_k$  iff  $\varphi_i \models \Diamond^+ p_k$  for some  $\varphi_i \in \mathfrak{a}$ .

**Proof.** ( $\Rightarrow$ ) We are given that there are formulas  $\psi_0, \dots, \psi_m$  in variables  $q_1, \dots, q_n$  such that  $\bigvee_j \varphi_j^* \in \mathbf{GL}$  and  $p_0^* \notin \mathbf{GL}$ , where by  $\chi^*$  we denote  $\chi\{\psi_0/p_0, \dots, \psi_m/p_m\}$ . This is equivalent to  $\mathfrak{M}_{\mathbf{GL}}(n) \models \bigvee_j \varphi_j^*$  and  $\mathfrak{M}_{\mathbf{GL}}(n) \not\models p_0^*$ . Define  $W$  to be the set of those disjuncts  $\varphi_j$  in  $\bigvee_j \varphi_j$  whose substitution instances  $\varphi_j^*$  are satisfied in  $\mathfrak{M}_{\mathbf{GL}}(n)$ . Clearly  $W \neq \emptyset$ . Let us check (i) – (iii).

- (i) Take a point  $x$  in  $\mathfrak{M}_{\mathbf{GL}}(n)$  at which  $p_0^*$  is false. As  $\mathfrak{M}_{\mathbf{GL}}(n) \models \bigvee_j \varphi_j^*$ , we must have  $x \models \varphi_i^*$  for some  $i$ . One of the formulas  $p_0^*$  or  $\neg p_0^*$  is a conjunct of  $\varphi_i^*$ . Clearly it is not  $p_0^*$ . Therefore,  $\neg p_0 \in \varphi_i$ .
- (ii) It suffices to show that, for all  $\varphi_i \in W$  and  $k \in \{0, \dots, m\}$ ,  $\varphi_i \models \Diamond p_k$  iff  $\Diamond p_k \in \varphi_i$ . Suppose  $\varphi_i \models \Diamond p_k$ . Then there is  $\varphi_j \in W$  such that  $\varphi_i R \varphi_j$  and  $\varphi_j \models p_k$ . By the definition of  $\mathfrak{B}$  and  $R$ , this means that  $p_k \in \varphi_j$  and  $\Diamond p_k \in \varphi_i$ . Conversely, suppose  $\Diamond p_k \in \varphi_i$ . Then  $x \models \varphi_i^*$  and in particular  $x \models \Diamond p_k^*$  for some  $x$  in  $\mathfrak{M}_{\mathbf{GL}}(n)$ . Let  $y$  be a final point in the set  $\{z \in x \uparrow : z \models p_k^*\}$ . Since  $\mathfrak{M}_{\mathbf{GL}}(n)$  is irreflexive, we have  $y \models p_k^*$ ,  $y \not\models \Diamond p_k^*$  and  $y \models \varphi_j^*$  for some  $\varphi_j \in W$ . It follows that  $\varphi_i R \varphi_j$  and  $\varphi_j \models p_k$ , from which  $\varphi_i \models \Diamond p_k$ .
- (iii) Let  $\mathfrak{a}$  be an antichain in  $\mathfrak{F}$ . For every  $\varphi_i \in \mathfrak{a}$ , let  $x_i$  be a final point in the set  $\{y \in W_{\mathbf{GL}}(n) : y \models \varphi_i^*\}$ . It should be clear that the points  $\{x_i : \varphi_i \in \mathfrak{a}\}$  form an antichain  $\mathfrak{b}$  in  $\mathfrak{F}_{\mathbf{GL}}(n)$  and so, by the construction of  $\mathfrak{F}_{\mathbf{GL}}(n)$ , there is a point  $y$  in  $\mathfrak{F}_{\mathbf{GL}}(n)$  such that  $y \uparrow = \mathfrak{b} \uparrow$ . Then the formula  $\varphi_j \in W$  we are looking for is any one satisfying the condition  $y \models \varphi_j^*$ , as can be easily checked by a straightforward inspection.

( $\Leftarrow$ ) The proof in this direction is rather technical; we confine ourselves to just a few remarks. Let  $\mathfrak{M}$  be a model satisfying (i)–(iii). To prove that  $\bigvee_j \varphi_j/p_0$  is not admissible in  $\mathbf{GL}$  we require once again the  $n$ -universal model  $\mathfrak{M}_{\mathbf{GL}}(n)$ , but this time we take  $n$  to be the number of symbols in the rule. By induction on the depth of points in  $\mathfrak{M}$  one can show that  $\mathfrak{M}$  is a generated submodel of  $\mathfrak{M}_{\mathbf{GL}}(n)$ .

Our aim is to find formulas  $\psi_0, \dots, \psi_m$  such that  $\mathfrak{M}_{\mathbf{GL}}(n) \models \bigvee_j \varphi_j^*$  and  $\mathfrak{M}_{\mathbf{GL}}(n) \not\models p_0^*$  (here again  $\chi^* = \chi\{\psi_0/p_0, \dots, \psi_m/p_m\}$ ). Loosely, we need to extend the properties of  $\mathfrak{M}$  to the whole model  $\mathfrak{M}_{\mathbf{GL}}(n)$ . To this end we can take the sets  $\{\varphi_i\}$  in  $\mathfrak{F}_{\mathbf{GL}}(n)$  and augment them inductively in such a way that we could embrace all points in  $\mathfrak{F}_{\mathbf{GL}}(n)$ . At the induction step we use the condition (iii), and the required  $\psi_0, \dots, \psi_m$  are constructed with the help of (i) and (ii); roughly, they describe in  $\mathfrak{M}_{\mathbf{GL}}(n)$  the analogues of the truth-sets in  $\mathfrak{M}$  of the variables in our rule. ■

A remarkable feature of this criterion is that it can be effectively checked. Thus we have

**THEOREM 198.** *There is an algorithm which, given an inference rule, can decide whether it is admissible in **GL**.*

In a similar way one can prove

**THEOREM 199** (Rybakov 1987). *The admissibility problem in **Grz** is decidable.*

We show now that the admissibility problem in **Int** can be reduced to the same problem in **Grz** and so is also decidable. To this end we require the following

**THEOREM 200** (Rybakov 1984b). *A rule  $\varphi/\psi$  is admissible in **Int** iff the rule  $T(\varphi)/T(\psi)$  is admissible in **Grz**.*

As a consequence of Theorems 199 and 200 we obtain

**THEOREM 201** (Rybakov 1984b). *The admissibility problem in **Int** is decidable.*

Although there are many other examples of logics in which the admissibility problem is decidable and the scheme of establishing decidability is quite similar to the argument presented above,<sup>20</sup> proofs are rather difficult and only in few cases they work for big families of logics as in [Rybakov 1994]. Besides, all these results hold only for extensions of **K4** and **Int**. For logics with non-transitive frames, even for **K**, the admissibility problem is still waiting for a solution. The same concerns polymodal, in particular tense logics. Chagrov [1992b] constructed a decidable infinitely axiomatizable logic in **NExtK4** for which the admissibility problem is undecidable. It would be of interest to find modal and si-calculi of that sort.

A close algorithmic problem for a logic  $L$  is to determine, given an arbitrary formula  $\varphi(p_1, \dots, p_n)$ , whether there exist formulas  $\psi_1, \dots, \psi_n$  such that  $\varphi(\psi_1, \dots, \psi_n) \in L$ . Note that an “equation”  $\varphi(p_1, \dots, p_n)$  has a solution in  $L$  iff the rule  $\varphi(p_1, \dots, p_n)/\perp$  is not admissible in  $L$ . This observation and Theorem 195 provide us with examples of logics in which the substitution problem is decidable (see e.g. [Rybakov 1993]). We do not know, however, if there is a logic such that the substitution problem in it is decidable, while the admissibility one is not.

The inference rules we have dealt with so far were *structural* in the sense that they were “closed” under substitution. An interesting example of a

---

<sup>20</sup>Quite recently S. Ghilardi [1999a,b] has found another way of recognizing admissibility of inference rules. He showed that certain si- and modal logics  $L$  (in particular, **Int**, **K4**, **S4**, **GL**, **Grz**) have the following property. Given an  $L$ -consistent formula  $\varphi$ , one can effectively compute substitutions  $\sigma_1, \dots, \sigma_n$  such that  $\sigma_i \varphi \in L$  for every  $i = 1, \dots, n$ , and if  $\sigma \varphi \in L$  for some substitution  $\sigma$ , then  $\sigma$  is, up to provable equivalence, an instantiation of some of the  $\sigma_i$ . A rule  $\varphi/\psi$  is then admissible in  $L$  iff  $\sigma_i \psi \in L$  for all  $i = 1, \dots, n$ .

nonstructural rule was considered by Gabbay [1981a]:

$$\frac{\varphi \vee (\Box p \rightarrow p), \text{ where } p \notin \mathbf{Sub}\varphi}{\varphi}$$

It is readily seen that this rule holds in a frame  $\mathfrak{F}$  (in the sense that for every formula  $\varphi$  and every variable  $p$  not occurring in  $\varphi$ ,  $\varphi$  is valid in  $\mathfrak{F}$  whenever  $(\Box p \rightarrow p) \vee \varphi$  is valid in  $\mathfrak{F}$ ) iff  $\mathfrak{F}$  is irreflexive and that  $\mathbf{K}$  is closed under it (since  $\mathbf{K}$  is characterized by the class of irreflexive frames). We refer the reader to [Venema 1991] and [Marx and Venema 1997] for more information about rules of this type.

### 4.3 Properties of recursively axiomatizable logics

Dealing with infinite classes of logics, we can regard questions like “Is a logic  $L$  decidable?”, “Does  $L$  have FMP?”, etc., as mass algorithmic problems. But to formulate such problems properly we should decide first how to represent the input data of algorithms recognizing properties of logics. One can, for instance, consider the class of recursively axiomatizable logics (which, by Craig’s [1953] Theorem, coincides with that of recursively enumerable ones) and represent them as programs generating their axioms. However, this approach turns out to be too general because the following analog of the Rice–Uspenskij Theorem holds.

**THEOREM 202** (Kuznetsov). *No nontrivial property of recursively axiomatizable si-logics is decidable.*

Of course, nothing will change if we take some other family of logics, say **NExtK4**. The proof of this theorem (Kuznetsov left it unpublished) is very simple; we give it even in a more general form than required.

**PROPOSITION 203.** *Suppose  $L_1$  and  $L_2$  are logics in some family  $\mathcal{L}$ ,  $L_1$  is recursively axiomatizable,  $L_1 \subset L_2$ ,  $L_2$  is finitely axiomatizable (say, by a formula  $\gamma$ ), and a property  $\mathcal{P}$  holds for only one of  $L_1, L_2$ . Then no algorithm can recognize  $\mathcal{P}$ , given a program enumerating axioms of a logic in  $\mathcal{L}$ .*

**Proof.** Let  $\alpha_0, \alpha_1, \dots$  be a recursive sequence of axioms for  $L_1$ . Given an arbitrary (Turing, Minsky, Pascal, etc.) program  $\mathbf{P}$  having natural numbers as its input, we define the following recursive sequence of formulas (where  $(n)_1$  and  $(n)_2$  are the first and second components of the pair of natural numbers with code  $n$  under some fixed effective encoding):

$$\beta_n = \begin{cases} \alpha_n & \text{if } \mathbf{P} \text{ does not come to a stop on input } (n)_1 \text{ in } (n)_2 \text{ steps} \\ \gamma & \text{otherwise.} \end{cases}$$

This sequence axiomatizes  $L_1$  if  $\mathbf{P}$  does not come to a stop on any input and  $L_2$  otherwise. It is well known in recursion theory that the halting problem is undecidable, and so the property  $\mathcal{P}$  is undecidable in  $\mathcal{L}$  as well. ■

The reader must have already noticed that this proof has nothing to do with modal and si-logics; it is rather about effective computations. To avoid this unpleasant situation let us confine ourselves to the smaller class of *finitely axiomatizable* modal and si-logics and try to find algorithms recognizing properties of the corresponding calculi. However, even in this case we should be very careful. If arbitrary finite axiomatizations are allowed then we come across the following

**THEOREM 204** (Kuznetsov 1963). *For every finitely axiomatizable si-logic  $L$  (in particular, **Int**, **Cl**, inconsistent logic), there is no algorithm which, given an arbitrary finite list of formulas, can determine whether its closure under substitution and modus ponens coincides with  $L$ .*

Needless to say that the same holds for (normal) modal logics as well. Fortunately, the situation is not so hopeless if we consider finite axiomatizations over some basic logics. For instance, by Makinson's Theorem, one can effectively recognize, given a formula  $\varphi$ , whether the logic  $\mathbf{K} \oplus \varphi$  is consistent. Other examples of decidable properties in various lattices of modal logics were presented in Theorems 89, 93, 101, and 142. In the next section we consider those properties that turn out to be undecidable in various classes of modal and si-calculi.

#### 4.4 Undecidable properties of calculi

The first “negative” algorithmic results concerning properties of modal calculi were obtained by Thomason [1982] who showed that FMP and Kripke completeness are undecidable in  $\text{NExt}\mathbf{K}$ , and consistency is undecidable in  $\text{NExt}\mathbf{K}.t$ . Later Thomason's discovery has been extended to other properties and narrower classes of logics. In fact, a good many standard properties of modal and si-calculi (in reasonably big classes) proved to be undecidable; decidable ones are rather exceptional.

In this section we present three known schemes of proving such kind of undecidability results. Each of them has its advantages (as well as disadvantages) and can be adjusted for various applications. The first one is due to Thomason [1982].

Let  $L(n)$  be a recursive sequence of normal bimodal calculi such that no algorithm can decide, given  $n$ , whether  $L(n)$  is consistent. Such sequences, as we shall see a bit later, exist even in  $\text{NExt}\mathbf{K4}.t$ . Suppose also that  $L^*$  is a normal unimodal calculus which does not have some property, say, FMP, decidability or Kripke completeness. Consider now the recursive sequence of logics  $L(n) \otimes L^*$  with three necessity operators. If  $L(n)$  is inconsistent then the fusion  $L(n) \otimes L^*$  is inconsistent too and so has the properties mentioned above. And if  $L(n)$  is consistent then, in accordance with Proposition 110,  $L(n) \otimes L^*$  is a conservative extension of both  $L(n)$  and  $L^*$ , which means that it is Kripke incomplete, undecidable and does not have FMP whenever

$L^*$  is so. Consequently, the three properties under consideration cannot be decidable in the class  $\text{NExt}\mathbf{K}_3$ , for otherwise the consistency of  $L(n)$  would be decidable. By Theorem 123, these properties are undecidable in  $\text{NExt}\mathbf{K}$  as well. Note however that, since Thomason’s simulation embeds polymodal logics only into “non-transitive” unimodal ones, this very simple scheme does not work if we want to investigate algorithmic aspects of properties of calculi in  $\text{NExt}\mathbf{K4}$  and  $\text{ExtInt}$ .

To illustrate the second scheme let us recall the construction of the undecidable calculus in  $\text{NExt}\mathbf{K4.t}$  discussed in Section 4.1. First, we choose a Minsky program  $\mathbf{P}$  and a configuration  $\mathbf{a} = \langle s, m, n \rangle$  so that no algorithm can decide, given a configuration  $\mathbf{b}$ , whether  $\mathbf{P} : \mathbf{a} \rightarrow \mathbf{b}$ . (That they exist is shown in [Chagrov 1990b].) Then we put  $\chi = \perp$  and add to  $\mathbf{K4.t} \oplus \text{AxP}$  one more axiom

$$(\neg\chi \wedge \text{O}\epsilon(s, \alpha_m^1, \alpha_n^2) \rightarrow \neg\chi \wedge \text{O}\epsilon(t, \alpha_k^1, \alpha_l^2)) \rightarrow \chi,$$

where  $\mathbf{c} = \langle t, k, l \rangle$  is an arbitrary fixed configuration. The resulting calculus is denoted by  $L(\mathbf{c})$ . Suppose that  $\mathbf{P} : \mathbf{a} \not\rightarrow \mathbf{c}$ . Then one can readily check that the new axiom is valid in the frame  $\mathfrak{F}$  shown in Fig. 19 and prove that  $\mathbf{P} : \langle s, m, n \rangle \rightarrow \langle t', k', l' \rangle$  iff

$$\neg\chi \wedge \text{O}\epsilon(s, \alpha_m^1, \alpha_n^2) \rightarrow \neg\chi \wedge \text{O}\epsilon(t', \alpha_{k'}^1, \alpha_{l'}^2) \in L(\mathbf{c}).$$

Therefore,  $L(\mathbf{c})$  is undecidable, consistent and does not have FMP. And if  $\mathbf{P} : \mathbf{a} \rightarrow \mathbf{c}$  then  $L(\mathbf{c})$  is clearly inconsistent. It follows by the choice of  $\mathbf{P}$  and  $\mathbf{a}$  that consistency, decidability and FMP are undecidable in  $\text{NExt}\mathbf{K4.t}$ . In fact, the argument will change very little if we take as  $\chi$  the axiom of some tabular logic in  $\text{NExt}\mathbf{K4.t}$ . So we obtain

**THEOREM 205.** *The properties of tabularity and coincidence with an arbitrary fixed tabular logic (in particular, inconsistent) are undecidable in  $\text{NExt}\mathbf{K4.t}$*

Moreover, these results (except the consistency problem, of course) can be transferred to logics in  $\text{NExt}\mathbf{K}$ . We demonstrate this by an example; complete proofs can be found in [Chagrov 1996].

We require the frame which results from that in Fig. 19 by adding to it a reflexive point  $c_0$  and an irreflexive one  $c_1$  so that  $c_1$  sees all other points save  $a$  and  $b$  and is seen itself only from  $a$  and  $b$ . As before, we denote the frame by  $\mathfrak{F}$ .

**PROPOSITION 206.** *Let  $\chi$  be a formula refutable at some point in  $\mathfrak{F}$  different from  $c_0$  and  $\diamond\top \in \mathbf{K} \oplus \chi$ . Then the problem of deciding, for an arbitrary formula  $\varphi$ , whether  $\mathbf{K} \oplus \varphi = \mathbf{K} \oplus \chi$  is undecidable.*

**Proof.** It should be clear that  $\chi$  contains at least one variable, say  $r$ , and there are points in  $\mathfrak{F}$  at which  $r$  has distinct truth-values (under the



valuation refuting  $\chi$ );  $c_0$  and  $c_1$  are then the only points in  $\mathfrak{F}$  where the formulas  $\sigma_0 = \Box^3 r \vee \Box^3 \neg r$  and

$$\sigma_1 = \Diamond \sigma_0 \wedge (r \vee \Diamond r \vee \Diamond^2 r) \wedge (\neg r \vee \Diamond \neg r \vee \Diamond^2 \neg r)$$

are true, respectively. Observe that from every point in  $\mathfrak{F}$  save  $c_0$  we can reach all points in  $\mathfrak{F}$  by  $\leq 3$  steps. So we can take  $\bigcirc = \Diamond^{\leq 3}$ . The formulas  $\alpha$  and  $\beta$  should be replaced with  $\alpha = \Diamond \sigma_1 \wedge \Diamond^2 \sigma_1$ ,  $\beta = \Diamond \sigma_1 \wedge \neg \Diamond^2 \sigma_1$  which (under the valuation refuting  $\chi$ ) are true only at  $a$  and  $b$ , respectively. Now consider the logic

$$L(\mathfrak{c}) = \mathbf{K} \oplus AxP \oplus (\neg \chi \wedge \bigcirc \epsilon(s, \alpha_m^1, \alpha_n^2) \rightarrow \neg \chi \wedge \bigcirc \epsilon(t, \alpha_k^1, \alpha_l^2)) \rightarrow \chi.$$

If  $\mathbf{P} : \mathfrak{a} \rightarrow \mathfrak{c}$  then  $L(\mathfrak{c}) = \mathbf{K} \oplus \chi$ . And if  $\mathbf{P} : \mathfrak{a} \not\rightarrow \mathfrak{c}$  then, using the fact that the set of points in  $\mathfrak{F}$  where  $\chi$  is refutable coincides with the set of points from which every point of the form  $e(x, y, z)$  is accessible by three steps, one can show that  $\mathfrak{F} \models L(\mathfrak{c})$  and so  $L(\mathfrak{c}) \neq \mathbf{K} \oplus \chi$ . ■

Putting, for instance,  $\chi = \Box p \leftrightarrow p$ , we obtain then that the problem of coincidence with  $\text{Logo}$  is undecidable in  $\text{NExt}\mathbf{K}$ . Likewise one can prove the following

**THEOREM 207.**

- (i) *If a consistent finitely axiomatizable logic  $L$  is not a union-splitting of  $\text{NExt}\mathbf{K}$  then the axiomatization problem for  $L$  above  $\mathbf{K}$  is undecidable.*
- (ii) *The properties of tabularity and coincidence with an arbitrary fixed consistent tabular logic are undecidable in  $\text{NExt}\mathbf{K}$ .*
- (iii) *The problem of coincidence with an arbitrary fixed consistent calculus in  $\text{NExt}\mathbf{D4}$  or in  $\text{NExt}\mathbf{GL}$  is undecidable in  $\text{NExt}\mathbf{K}$ .*
- (iv) *The properties of tabularity and coincidence with an arbitrary fixed tabular (in particular, inconsistent) logic are undecidable in  $\text{Ext}\mathbf{K4}$ .*

Of the algorithmic problems concerning tabularity that remain open the most intriguing are undoubtedly the tabularity and local tabularity problems in  $\text{NExt}\mathbf{K4}$ . Note that a positive solution to the former implies a positive solution to the latter.

Now we present the second scheme in a more general form used in [Chagro 1990b] and [Chagro and Zakharyashev 1993]. Assume again that the second configuration problem is undecidable for  $\mathbf{P}$  and  $\mathfrak{a}$ , and let  $\chi$  be a formula such that  $L_0 \oplus \chi$  has some property  $\mathcal{P}$ , where  $L_0$  is the minimal logic in the class under consideration. Associate with  $\mathbf{P}$ ,  $\mathfrak{a}$  and a configuration  $\mathfrak{b}$  formulas  $AxP$  and  $\psi(\mathfrak{a}, \mathfrak{b})$  such that  $\psi(\mathfrak{a}, \mathfrak{b}) \in L_0 \oplus AxP$  iff  $\mathbf{P} : \mathfrak{a} \rightarrow \mathfrak{b}$ .

Besides,  $\chi$  and  $AxP$  are chosen so that  $AxP \in L_0 \oplus \chi$ . Now consider the calculus

$$L(\mathfrak{b}) = L_0 \oplus AxP \oplus \psi(\mathfrak{a}, \mathfrak{b}) \rightarrow \chi \oplus \gamma,$$

where  $\gamma$  is some formula such that  $\gamma \in L_0 \oplus \chi$ . If  $\mathbf{P} : \mathfrak{a} \rightarrow \mathfrak{b}$  then we clearly have  $L(\mathfrak{b}) = L_0 \oplus \chi$  and so  $L(\mathfrak{b})$  has  $\mathcal{P}$ ; but if  $\mathbf{P} : \mathfrak{a} \not\rightarrow \mathfrak{b}$  then the fact that  $L(\mathfrak{b})$  does not have  $\mathcal{P}$  must be ensured by an appropriate choice of  $\gamma$ . (In the considerations above we did not need  $\gamma$ , i.e., it was sufficient to put  $\gamma = \top$ ). With the help of this scheme one can prove the following

**THEOREM 208.**

- (i) *The properties of decidability, Kripke completeness as well as FMP are undecidable in the classes ExtInt, (N)ExtGrz, (N)ExtGL.*
- (ii) *The interpolation property is undecidable in (N)ExtGL.*
- (iii) *Halldén completeness is undecidable in ExtInt, (N)ExtGrz, ExtS.*

These and some other results of that sort can be found in [Chagrov 1990b,c, 1994, 1996], [Chagrova 1991], [Chagrov and Zakharyashev 1993, 1995b].

The third scheme was developed in [Chagrova 1989, 1991] and [Chagrov and Chagrova 1995] for establishing the undecidability of certain first order properties of modal calculi (or formulas). The difference of this scheme from the previous one is that now we use calculi of the form

$$L(\mathfrak{b}) = L_0 \oplus AxP \oplus \psi(\mathfrak{a}, \mathfrak{b}) \vee \gamma,$$

where  $AxP$  satisfies one more condition besides those mentioned above: it must be first order definable on Kripke frames for  $L_0$ . If  $P : \mathfrak{a} \rightarrow \mathfrak{b}$  then the formula  $AxP \wedge (\psi(\mathfrak{a}, \mathfrak{b}) \vee \gamma)$  is equivalent to  $AxP$  in the class of Kripke frames for  $L_0$  and so is first order definable on that class or its any subclass. And if  $P : \mathfrak{a} \not\rightarrow \mathfrak{b}$  then by choosing an appropriate  $\gamma$  one can show that  $AxP \wedge (\psi(\mathfrak{a}, \mathfrak{b}) \vee \gamma)$  is not first order definable on, say, countable Kripke frames for  $L_0$ , as in [Chagrova 1989], or on finite frames for  $L_0$ , as in [Chagrov and Chagrova 1995]. In this way the following theorem is proved:

**THEOREM 209.**

- (i) *No algorithm is able to recognize the first order definability of modal formulas on the class of Kripke frames for S4 and even the first order definability on countable (finite) Kripke frames for S4. The properties of first order definability and definability on countable (finite) Kripke frames of intuitionistic formulas are undecidable as well.*
- (ii) *The set of modal or intuitionistic formulas that are first order definable on countable (finite) frames but are not first order definable on the*

*class of all (respectively, countable) Kripke frames mentioned in (i) is undecidable.*

We conclude this section with two remarks. First, all undecidability results above can be formulated in the stronger form of recursive inseparability. For instance, the set of inconsistent calculi in  $\text{NExt}\mathbf{K4}.t$  and the set of calculi without FMP are recursively inseparable. And second, some properties are not only undecidable but the families of calculi having them are not recursively enumerable; for example, the set of consistent calculi in  $\text{NExt}\mathbf{K4}.t$  is not enumerable. However, for the majority of other properties the problem of enumerability of the corresponding calculi is open.

#### 4.5 Semantical consequence

So far we have dealt with only syntactical formalizations of logical entailment. However, sometimes a semantical approach is preferable. Say that a formula  $\varphi$  is a *semantical consequence* of a formula  $\psi$  in a class of frames  $\mathcal{C}$  if  $\varphi$  is valid in all frames in  $\mathcal{C}$  validating  $\psi$ . (One can consider also the local, i.e., point-wise variant of this relation.) Note that  $\varphi$  is a consequence of  $\psi$  in the class of, say, Kripke frames for  $\mathbf{S4}$  iff  $\varphi$  is a consequence of  $(\Box p \rightarrow \Box^2 p) \wedge (\Box p \rightarrow p) \wedge \psi$  in the class of all Kripke frames. But the consequence relation on finite frames is not expressible by modal formulas (as was shown in [Chagrov 1995], if  $(\Box p \rightarrow \Box^2 p) \wedge \varphi$  is valid in arbitrarily large finite rooted frames then it is valid in some infinite rooted frame as well).

In parallel with constructing and proving the undecidability of modal and si-calculi we can obtain the following

**THEOREM 210.** *The semantical consequence relation in the class of all ( $\mathbf{K4}$ -,  $\mathbf{S4}$ -,  $\mathbf{Int}$ -) Kripke frames is undecidable. Moreover, if  $\models$  denotes one of these relations then there is a formula  $\psi$  (a formula  $\varphi$ ) such that the set  $\{\varphi : \psi \models \varphi\}$  is undecidable.*

In a sense, formulas  $\psi$  and  $\varphi$ , for which  $\{\varphi : \psi \models \varphi\}$  is undecidable are analogous to undecidable calculi and formulas, respectively. However, this analogy is far from being perfect: for every formula  $\psi$ , the sets  $\{\varphi : \psi \vdash \varphi\}$  and  $\{\varphi : \psi \vdash^* \varphi\}$  are recursively enumerable, which contrasts with

**THEOREM 211** (Thomason 1975a). *There exists a formula  $\psi$  such that  $\{\varphi : \psi \models \varphi\}$  is a  $\Pi_1^1$ -complete set.*

Unfortunately, Thomason's [1974b, 1975b, 1975c] results have not been transferred so far to transitive frames, although this does not seem to be absolutely impossible.

Chagrov [1990a] (see also [Chagrov and Chagrova 1995]) developed a technique for proving the analog of Theorem 210 for the consequence relation

on all (**K4**-, **S4**-, **GL**-, **Int**-) finite frames. Moreover, since this relation is clearly enumerable, instead of “undecidable” one can use “not enumerable”.

#### 4.6 Complexity problems

Having proved that a given logic is decidable, we are facing the problem of finding an optimal (in one sense or another) decision algorithm for it. The complexity of decision algorithms for many standard modal and si-logics is determined by the size of minimal frames separating formulas from those logics. For instance, as was shown by Jaśkowski (1936) and McKinsey (1941), for every  $\varphi \notin \mathbf{S4}$  (or  $\varphi \notin \mathbf{Int}$ ) there is a frame  $\mathfrak{F} \models \mathbf{S4}$  with  $\leq 2^{|\text{Sub}\varphi|}$  points such that  $\mathfrak{F} \not\models \varphi$ . The same upper bound is usually obtained by the standard filtration. Is it possible to reduce the exponential upper bound to the polynomial one? This question was raised by Kuznetsov [1975] for **Int**. It turned out, however, that it concerns not only **Int**. First, Kuznetsov observed (for the proof see [Kuznetsov 1979]) that if the answer to his question is positive, i.e., **Int** has polynomial FMP, then the problem “Are **Int** and **Cl** polynomially equivalent?” has a positive solution as well. (Logics  $L_1$  and  $L_2$  are *polynomially equivalent* if there are polynomial time transformations  $f$  and  $g$  of formulas such that  $\varphi \in L_1$  iff  $f(\varphi) \in L_2$  and  $\varphi \in L_2$  iff  $g(\varphi) \in L_1$ .) Then Statman [1979] showed that the problem “ $\varphi \in \mathbf{Int}$ ?” is *PSPACE*-complete and so Kuznetsov’s problem is equivalent to one of the “hopeless” complexity problems, namely “ $NP = PSPACE$ ?”.

#### Complexity function

For a logic  $L$  with FMP, we introduce the *complexity function*

$$f_L(n) = \max_{\substack{l(\varphi) \leq n \\ \varphi \notin L}} \min_{\substack{\mathfrak{F} \models L \\ \mathfrak{F} \not\models \varphi}} |\mathfrak{F}|,$$

where  $l(\varphi)$ , the *length* of  $\varphi$ , is the number of subformulas in  $\varphi$  and  $|\mathfrak{F}|$  the number of points in  $\mathfrak{F}$ . If there is a constant  $c$  such that

$$f_L(n) \leq 2^{c \cdot n} \text{ (or } f_L(n) \leq n^c \text{ or } f_L(n) \leq c \cdot n),$$

$L$  is said to have the *exponential* (respectively, *polynomial* or *linear*) *finite model property*. The following result shows that **Int** does not have polynomial FMP.

**THEOREM 212** (Zakharyashev and Popov 1979).  $\log_2 f_{\mathbf{Int}}(n) \asymp n$ .

**Proof.** The exponential upper bound is well known and to establish the lower one it is sufficient to use the formulas

$$\beta_n = \bigwedge_{i=1}^{n-1} ((\neg p_{i+1} \rightarrow q_{i+1}) \vee (p_{i+1} \rightarrow q_{i+1}) \rightarrow q_i) \rightarrow (\neg p_1 \rightarrow q_1) \vee (p_1 \rightarrow q_1).$$

It is not hard to see that  $\beta_n \notin \mathbf{Int}$  and every refutation frame for  $\beta_n$  contains the full binary tree of depth  $n$  as a subframe. ■

Likewise the same result can be proved for many other standard super-intuitionistic and modal logics whose FMP is established by the usual filtration and whose frames contain full binary trees of arbitrary finite depth. Such are, for instance, **KC**, **SL**, **K4**, **S4**, **GL**. In the case of **K** the length of formulas that play the role of  $\beta_n$  is not a linear but a square function of  $n$ , which means that  $f_{\mathbf{K}}(n) \geq 2^{\sqrt{c \cdot n}}$ , for some constant  $c > 0$ , and so **K** does not have polynomial FMP either. As was shown in [Zakharyashev 1996], all cofinal subframe modal and si-logics have exponential FMP. It seems plausible that  $\log_2 f_L(n) \asymp n$  for every consistent si-logic  $L$  different from **CI** and axiomatizable by formulas in one variable.

The construction of Theorem 212 does not work for logics whose frames do not contain arbitrarily large full binary trees. Such are, for instance, logics of finite width or of finite depth, and the following was proved in [Chagrov 1983].

**THEOREM 213.**

- (i) *The minimal logics of width  $n < \omega$  in the classes **NExtK4**, **NExtS4**, **NExtGrz**, **NExtGL**, **ExtInt** have polynomial FMP.*
- (ii) **Lin** and all logics containing **S4.3** have linear FMP.
- (iii) *The minimal logics of depth  $n$  in **NExtGrz**, **NExtGL**, **ExtInt** have polynomial FMP, with the power of the corresponding polynomial  $\leq n - 1$ .*
- (iv) *The minimal logics of depth  $n$  in **NExtK4**, **NExtS4** have polynomial FMP, with the power of the corresponding polynomial  $\leq n$ .*

**Proof.** (i) is proved by two filtrations. First, with the help of the standard filtration one constructs a finite frame separating a formula  $\varphi$  from the given logic  $L$  and then, using the selective filtration, extracts from it a polynomial separation frame: it suffices to take a point refuting  $\varphi$  and all maximal points at which  $\psi$  is false, for some  $\Box\psi \in \mathbf{Sub}\varphi$  (in the intuitionistic case  $\psi \rightarrow \chi \in \mathbf{Sub}\varphi$  should be considered). (ii) is proved analogously.

To illustrate the proof of (iii) and (iv), we consider the minimal logic  $L$  of depth 3 in **NExtGL**. Suppose  $\varphi \notin L$ . Then there is a transitive irreflexive model  $\mathfrak{M}$  of depth  $\leq 3$  refuting  $\varphi$  at its root  $r$ . Let  $\Box\psi_i$ , for  $1 \leq i \leq m$ , be all “boxed” subformulas of  $\varphi$ . For every  $i \in \{1, \dots, m\}$ , we choose a point refuting  $\psi_i$ , if it exists. And then we do the same in the set  $x\uparrow$ , for every chosen point  $x$ . Let  $\mathfrak{M}'$  be the submodel formed by the selected points and  $r$ . Clearly, it contains at most  $1 + m + m^2$  points. And by induction on the

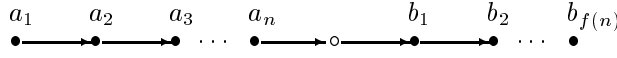


Figure 20.

construction of formulas in **Sub** $\varphi$  one can easily show that  $\mathfrak{M}'$  refutes  $\varphi$  at  $r$ .

To prove the lower bound one can use the formulas

$$\alpha_n = \neg \left( \bigwedge_{i=1}^n \Box(p_{i+1} \rightarrow p_i) \wedge \bigwedge_{i=1}^n \Box(q_{i+1} \rightarrow q_i) \wedge \bigwedge_{i=1}^n \Diamond(\Diamond \top \wedge \Box^+(\neg p_{i+1} \wedge p_i)) \wedge \Box(\Diamond \perp \rightarrow \bigwedge_{i=1}^n \Diamond(\neg q_{i+1} \wedge q_i)) \right)$$

which are not in  $L$  and every separation frame for which contains the full  $n$ -ary tree of depth 3, i.e., at least  $1 + n + n^2$  points. ■

However, even if frames for a logic with FMP do not contain full finite binary trees its complexity function can grow very fast, witness the following result of [Chagrov 1985a].

**THEOREM 214.** *For every arithmetic function  $f(n)$ , there are logics  $L$  of width 1 in **NExtK4** and of width 2 in **ExtInt**, **NExtGrz**, **NExtGL** having FMP and such that  $f_L(n) \geq f(n)$ .*

**Proof.** We construct a logic  $L \in \mathbf{NExtK4.3}$  whose complexity function grows faster than a given increasing arithmetic function  $f(n)$ . Define  $L$  to be the logic of all frames of the form shown in Fig. 20. To see that  $L$  satisfies the property we need, consider the sequence of formulas

$$\beta_1 = p_1 \vee \Box(\Box p_1 \rightarrow (\Box(\Box p \rightarrow p) \rightarrow p)),$$

$$\beta_{i+1} = p_{i+1} \vee \Box(\Box p_{i+1} \rightarrow \beta_i).$$

Since these formulas are refuted at points of the form  $a_j$  in sufficiently large frames depicted in Fig. 20, they are not in  $L$ . And since  $L$  contains the formulas

$$\neg \beta_n \rightarrow \Diamond(\Diamond^{f(n)-1} \top \wedge \Box^{f(n)} \perp),$$

$\beta_n$  cannot be separated from  $L$  by a frame with  $\leq f(n)$  points. ■

For logics of finite depth this theorem does not hold, since according to the description of finitely generated universal frames in Section 1.2, for every  $L \in \mathbf{NExtK4BD}_k$  ( $k \geq 3$ ), we have

$$f_L(n) \leq 2^2 \left. \begin{matrix} 2^{c \cdot n} \\ \dots \\ 2^{c \cdot n} \end{matrix} \right\} k-2$$

for some constant  $c > 0$ . And as was shown in [Chagrov 1985a], one cannot in general reduce this upper bound.

**THEOREM 215.** *For every  $k \geq 3$ , there are logics  $L$  of depth  $k$  in  $\text{NExtGrz}$ ,  $\text{NExtGL}$ ,  $\text{ExtInt}$  such that*

$$f_L(n) \geq 2^{\left\{ \begin{array}{l} 2^{\dots 2^n \\ \dots \end{array} \right\}^{k-2}}.$$

**Proof.** We illustrate the proof for  $k = 3$  in  $\text{NExtGL}$ . Let  $L$  be the logic characterized by the class of rooted frames  $\mathfrak{F}_m$  for  $\mathbf{GL}$  of depth 3 defined as follows.  $\mathfrak{F}_m$  contains  $m$  dead ends, every non-empty set of them has a focus, i.e., a point that sees precisely the dead ends in this set, and besides the root there are no other points in  $\mathfrak{F}_m$ . It should be clear that  $L$  does not contain the formulas

$$\gamma_m = \bigwedge_{i=1}^n \Box(p_{i+1} \rightarrow p_i) \rightarrow \bigwedge_{i=1}^n \Box\Box(p_i \rightarrow p_{i+1}).$$

On the other hand  $\gamma_n$  is not refutable in a frame for  $L$  with  $< 2^m$  points because the following formulas are in  $L$ :

$$\neg\gamma_m \rightarrow \bigwedge_{X \subseteq \{1, \dots, m\}, X \neq \emptyset} \Diamond \left( \bigwedge_{i \in X} \Diamond \delta_i \wedge \bigwedge_{i \notin X, 1 \leq i \leq m} \neg \Diamond \delta_i \right),$$

where  $\delta_i = p_1 \wedge \dots \wedge p_i \wedge \neg p_{i+1} \wedge \dots \wedge \neg p_{m+1}$ . ■

Note, however, that the logics constructed in the proofs of the last two theorems are not finitely axiomatizable. We know of only one “very complex” calculus with FMP.

**THEOREM 216.**  $\log_2 \log_2 f_{\mathbf{KIP}}(n) \asymp n$ .

For the proof see [Chagrov and Zakharyashev 1997], where the reader can find also some other results in this direction.

#### *Relation to complexity classes*

Let us return to the original problem of optimizing decision algorithms for the logics under consideration. First of all, it is to be noted that there is a natural lower bound for decision algorithms which cannot be reduced—we mean the complexity of decision procedures for  $\mathbf{CI}$ . This is clear for (consistent) modal logics on the classical base; and by Glivenko’s Theorem, every si-logic “contains”  $\mathbf{CI}$  in the form of the negated formulas. Thus, if we manage to construct an effective decision procedure for some of our logics then  $\mathbf{CI}$  can be decided by an equally effective algorithm. (We remind the reader that all existing decision algorithms for  $\mathbf{CI}$  require exponential

time (of the number of variables in the tested formulas). On the other hand, only polynomial time algorithms are regarded to be acceptable in complexity theory.)

So, when analyzing the complexity of decision algorithms for modal and si-logics, it is reasonable to compare them with decision algorithms for **CI**. For example, if a logic  $L$  is polynomially equivalent to **CI** then we can regard these two logics to be of the same complexity. Moreover, provided that somebody finds a polynomial time decision procedure for **CI**, a polynomial time decision algorithm can be constructed for  $L$  as well. The following theorem lists results obtained by [Ladner 1977], [Ono and Nakamura 1980], [Chagrova 1983], and [Spaan 1993].

**THEOREM 217.** *All logics mentioned in the formulation of Theorem 213 are polynomially equivalent to **CI**.*

**Proof.** We illustrate the proof only for the minimal logic  $L$  of depth 3 in **NExtGL** using the method of [Kuznetsov 1979]. Suppose  $\varphi$  is a formula of length  $n$ . By Theorem 213, the condition  $\varphi \notin L$  means that  $\mathfrak{M} \not\models \varphi$ , for some model  $\mathfrak{M} = \langle \mathfrak{F}, \mathfrak{V} \rangle$  based on a frame  $\mathfrak{F}$  for **GL** of depth  $\leq 3$  and cardinality  $\leq c \cdot n^2$ . We describe this observation by means of classical formulas, understanding their variables as follows. Let  $x, y, z$  be names (numbers) of points in  $\mathfrak{F}$ , for  $1 \leq x, y, z \leq c \cdot n^2$ . With every pair  $\langle x, y \rangle$  of points in  $\mathfrak{F}$  we associate a variable  $p_{xy}$  whose meaning is “ $x$  sees  $y$ ”. And with every  $\psi \in \mathbf{Sub}\varphi$  and every  $x$  we associate a variable  $q_x^\psi$  which means “ $\psi$  is true at  $x$ ”. Denote by  $\alpha$  the conjunction

$$q_1^\varphi \wedge q_2^\varphi \wedge \cdots \wedge q_{c \cdot n^2}^\varphi.$$

It means that  $\varphi$  is true in  $\mathfrak{M}$ . And let  $\beta$  be the conjunction of the following formulas under all possible values of their subscripts:

$$\neg p_{xx}, \quad p_{xy} \wedge p_{yz} \rightarrow p_{xz}, \quad q_x^{\neg\psi} \leftrightarrow \neg q_x^\psi,$$

$$q_x^{\psi \wedge \chi} \leftrightarrow q_x^\psi \wedge q_x^\chi, \quad q_x^{\psi \vee \chi} \leftrightarrow q_x^\psi \vee q_x^\chi, \quad q_x^{\Box\psi} \leftrightarrow \bigwedge_{y=1}^{c \cdot n^2} (p_{xy} \rightarrow q_y^\psi).$$

(The first two formulas say that  $R$  is irreflexive and transitive and the rest simulate the truth-relation in  $\mathfrak{M}$ .) Finally, we define a formula saying that our frame is of depth  $\leq 3$ :

$$\gamma = \bigwedge_{1 \leq x, y, z, u \leq c \cdot n^2} \neg(p_{xy} \wedge p_{yz} \wedge p_{zu}).$$

The formula  $\beta \wedge \gamma \wedge \neg\alpha$  is of length  $\leq 50(c \cdot n^2)^5$  and can be clearly constructed by an algorithm working at most polynomial time in the length of  $\varphi$ . It is readily seen that  $\varphi \notin L$  iff  $\beta \wedge \gamma \wedge \neg\alpha$  is satisfiable in **CI**. Thus we have



polynomially reduced the derivability problem in  $L$  to that in  $\mathbf{Cl}$ . Since the converse reduction is trivial,  $L$  and  $\mathbf{Cl}$  are polynomially equivalent. ■

The reader must have noticed that Theorem 217 lists almost all logics known to have polynomial FMP. Kuznetsov [1975] conjectured that every calculus having polynomial FMP is polynomially equivalent to  $\mathbf{Cl}$ . This conjecture is closely related to some problems in the complexity theory of algorithms. We remind the reader that  $\mathbf{NP}$  is the class of problems that can be solved by polynomial time algorithms on nondeterministic (Turing) machines. An  $\mathbf{NP}$ -complete problem is a problem in  $\mathbf{NP}$  to which all other problems in  $\mathbf{NP}$  are polynomially reducible. (For more detailed definitions consult [Garey and Johnson 1979].) The most popular  $\mathbf{NP}$ -complete problem is the satisfiability problem for Boolean formulas, i.e., the nonderivability problem for  $\mathbf{Cl}$ . So the nonderivability problem for all logics listed Theorem 217 is  $\mathbf{NP}$ -complete and Kuznetsov's conjecture is equivalent to a positive solution to the problem whether the nonderivability problem for every calculus with polynomial FMP is  $\mathbf{NP}$ -complete.

Note that if  $\mathbf{coNP} = \mathbf{NP}$  (for the definition of the class  $\mathbf{coNP}$  see [Garey and Johnson 1979]; we just mention that the derivability problem in  $\mathbf{Cl}$  is  $\mathbf{coNP}$ -complete) then Kuznetsov's conjecture does hold. But since " $\mathbf{coNP} = \mathbf{NP}$ ?" belongs to the list of "unsolvable" problems under the current state of knowledge, it may be of interest to find out whether Kuznetsov's conjecture implies  $\mathbf{coNP} = \mathbf{NP}$ .

Another complexity class we consider here is the class  $\mathbf{PSPACE}$  of problems that can be solved by polynomial space algorithms. A typical example of a  $\mathbf{PSPACE}$ -complete problem is the truth problem for quantified Boolean formulas. The following theorem (which summarizes results obtained by Ladner [1977], Statman [1979], Chagrov [1985a], Halpern and Moses [1992] and Spaan [1993]) lists some  $\mathbf{PSPACE}$ -complete logics.

**THEOREM 218.** *The nonderivability problem (and so the derivability problem) in the following logics is  $\mathbf{PSPACE}$ -complete:  $\mathbf{Int}$ ,  $\mathbf{KC}$ ,  $\mathbf{K}$ ,  $\mathbf{K} \otimes \mathbf{K}$ ,  $\mathbf{S4}$ ,  $\mathbf{S4} \otimes \mathbf{S4}$ ,  $\mathbf{S5} \otimes \mathbf{S5}$ ,  $\mathbf{GL}$ ,  $\mathbf{Grz}$ ,  $\mathbf{K.t}$  and  $\mathbf{K4.t}$ .*

It follows in particular that complexity is not preserved under the formation of fusions of logics (under the assumption  $\mathbf{NP} \neq \mathbf{PSPACE}$ ), since nonderivability in  $\mathbf{S5}$  is  $\mathbf{NP}$ -complete. For more information on the preservation of complexity under fusions consult [Spaan 1993].

Finally we note that the nonderivability problem in logics with the universal modality or common knowledge operator is mostly even  $\mathbf{EXPTIME}$ -complete, witness  $\mathbf{K}_u$  [Spaan 1993] and  $\mathbf{S4EC}_2$  [Halpern and Moses 1992]. The complexity of the nonderivability problem for Cartesian products of many standard modal logics is  $\mathbf{NEXPTIME}$ -hard;  $\mathbf{S5} \times \mathbf{S5}$  and  $\mathbf{K} \times \mathbf{S5}$  are examples of  $\mathbf{NEXPTIME}$ -complete logics (see [Marx 1999]). (Note, by the way, that the known upper bound for  $\mathbf{K} \times \mathbf{K}$  is non-elementary.)

## 5 APPENDIX

We conclude this chapter with a (by no means complete) list of references for those directions of research in modal logic that were not considered above:

- *Congruential logics.* These are modal logics that do not necessarily contain the distribution axiom  $\Box(p \rightarrow q) \rightarrow (\Box p \rightarrow \Box q)$  but are closed under modus ponens and the congruence rule  $p \leftrightarrow q / \Box p \leftrightarrow \Box q$ . Segerberg [1971] and Chellas [1980] define a semantics for these logics; Lewis [1974] proves FMP of all congruential non-iterative logics and Surendonk [1996] shows that they are canonical. Došen [1988] considers duality between algebras and neighbourhood frames and Kracht and Wolter [1999] study embeddings into normal bimodal logics.
- *Modal logics with graded modalities.* The truth-relation for their possibility operators  $\Diamond_n$  is defined as follows:  $x \models \Diamond_n p$  iff there exist at least  $n$  points accessible from  $x$  at which  $p$  holds. An early reference is [Fine 1972]; more recent are [van der Hoek 1992] (applications to epistemic logic) and [Cerrato 1994] (FMP and decidability).
- *Modal logics with the difference operator* or with *nominals* (or *names*). The semantics of nominals is similar to that of propositional variables; the difference is that a nominal is true at exactly one point in a frame. For the difference operator  $[\neq]$ , we have  $x \models [\neq]p$  iff  $p$  is true everywhere except  $x$ . De Rijke [1993], Blackburn [1993] and Goranko and Gargov [1993] study the completeness and expressive power of systems of that sort. Closely related to the difference operator is the modal operator  $[i]$  for inaccessible worlds:  $x \models [i]p$  iff  $p$  is true in all worlds which are not accessible from  $x$ , see [Humberstone 1983] and [Goranko 1990a].
- *Modal logics with dyadic or even polyadic operators.* For duality theory in this case see [Goldblatt 1989]. An extensive study of Sahlqvist-type theorems with applications to polyadic logics is [Venema 1991]. For connections with the theory of relational algebras see [Mikulas 1995] and [Marx 1995]. In those dissertations the reader can find also recent results on arrow logic, i.e., a certain type of polyadic logic which is interpreted in Kripke frames built from arrows. An embedding of polyadic logics into polymodal logics is discussed in [Kracht and Wolter 1997].
- *Bisimulations.* Bisimulations were introduced in modal logic by van Benthem [1983] to characterize its expressive power; see also [de Rijke 1996]. Visser [1996] used bisimulations to prove uniform interpolation. Recently, bisimulations have attracted attention because they form a

common tool in modal logic and process theory. We refer the reader to collection [Ponse *et al.* 1996] for information on this subject.

- *Modal logics with fixed point operators*, i.e., modal logics enriched by operators forming the least and greatest fixed points of monotone formulas. These systems are also called *modal  $\mu$ -calculi*. Under this name they were introduced and studied by Kozen [1983, 1988]; see also [Walukiewicz 1993, 1996] and [Bosangue and Kwiatkowska 1996].
- *Proof theory*. Early references to studies of sequent calculi and natural deduction systems for a few modal logics can be found in *Basic Modal Logic*. More recently, (non-standard) sequent calculi for modal logics have been considered by Došen [1985b], Masini [1992] and Avron [1996]; see also collection [Wansing 1996] and the chapter *Sequent systems for modal logics* later in this *Handbook*. For natural deduction systems see Borghuis [1993]; tableau systems for modal and tense logics were constructed in [Fitting 1983], [Rautenberg 1983], [Gore 1994] and [Kashima 1994]. Orłowska [1996] develops *relational proof systems*. Display calculi for modal logics were introduced by Belnap [1982]; see also [Wansing 1994] and collection [Wansing 1996].
- *Description logic*, a formalism closely related to modal logic, was designed in artificial intelligence by Brachman and Schmoltz [1985] as a means for knowledge representation and reasoning (for a survey see [Donini *et al.* 1996]). Schild [1991] was the first to observe that the basic description logic *ALC* is just a terminological variant of the polymodal **K**. Recently, in order to represent dynamic and intensional knowledge, combinations of description and modal logics have been introduced, see e.g. Baader and Ohlbach [1995], Baader and Laux [1995], and Wolter and Zakharyashev [1998, 1999b,c].

#### ACKNOWLEDGMENTS

First of all, we are indebted to our friend and colleague Marcus Kracht who not only helped us with numerous advices but also supplied us with some material for this chapter. We are grateful to Hiroakira Ono and the members of his Logic Group in Japan Advanced Institute of Science and Technology for the creative and stimulating atmosphere that surrounded the first two authors during their stay in JAIST in 1996–97, where the bulk of this chapter was written. Thanks are also due to Johan van Benthem, Wim Blok, Dov Gabbay, Silvio Ghilardi, Agnes Kurucz, Krister Segerberg, Valentin Shehtman, Dimiter Vakarelov, and Heinrich Wansing for their helpful comments and stimulating discussions. And certainly our work would be impossible without constant support and love of our wives: Olga, Imke and Lilia.

The work of the first author was partly financed by the Alexander von Humboldt Foundation.

A. Chagrov  
*Tver State University, Russia*

F. Wolter  
*Institute of Information Science, Leipzig University, Germany*

M. Zakharyashev  
*King's College London, UK*

## BIBLIOGRAPHY

- [Amati and Pirri, 1994] G. Amati and F. Pirri. A uniform tableau method for intuitionistic modal logics I. *Studia Logica*, 53:29–60, 1994.
- [Anderson, 1972] J.G. Anderson. Superconstructive propositional calculi with extra axiom schemes containing one variable. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 18:113–130, 1972.
- [Avron, 1996] A. Avron. The method of hypersequents in the proof theory of propositional non-classical logics. In W. Hodges, M. Hyland, C. Steinhorn, and J. Truss, editors, *Logic: from Foundations to Applications*, pages 1–32. Clarendon Press, Oxford, 1996.
- [Baader and Laux, 1995] F. Baader and A. Laux. Terminological logics with modal operators. In *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, pages 808–814, Montreal, Canada, 1995. Morgan Kaufman.
- [Baader and Ohlbach, 1995] F. Baader and H.J. Ohlbach. A multi-dimensional terminological knowledge representation language. *Journal of Applied Non-Classical Logic*, 5:153–197, 1995.
- [Barwise and Moss, 1996] J. Barwise and L. Moss. *Vicious Circles*. CSLI Publications, Stanford, 1996.
- [Beklemishev, 1994] L.D. Beklemishev. On bimodal logics of provability. *Annals of Pure and Applied Logic*, 68:115–159, 1994.
- [Beklemishev, 1996] L.D. Beklemishev. Bimodal logics for extensions of arithmetical theories. *Journal of Symbolic Logic*, 61:91–124, 1996.
- [Bellissima, 1984] F. Bellissima. Atoms in modal algebras. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 30:303–312, 1984.
- [Bellissima, 1985] F. Bellissima. An effective representation for finitely generated free interior algebras. *Algebra Universalis*, 20:302–317, 1985.
- [Bellissima, 1988] F. Bellissima. On the lattice of extensions of the modal logic  $K.Alt_n$ . *Archive of Mathematical Logic*, 27:107–114, 1988.
- [Bellissima, 1991] F. Bellissima. Atoms of tense algebras. *Algebra Universalis*, 28:52–78, 1991.
- [Belnap, 1982] N.D. Belnap. Display logic. *Journal of Philosophical Logic*, 11:375–417, 1982.
- [Beth, 1953] E.W. Beth. On Padua's method in the theory of definitions. *Indagationes Mathematicae*, 15:330–339, 1953.
- [Bezhanishvili, 1998] G. Bezhanishvili. Varieties of monadic algebras. Part I. *Studia Logica*, 61:367–402, 1998.
- [Blackburn, 1993] P. Blackburn. Nominal tense logic. *Notre Dame Journal of Formal Logic*, 34:56–83, 1993.
- [Blok and Köhler, 1983] W.J. Blok and P. Köhler. Algebraic semantics for quasi-classical modal logics. *Journal of Symbolic Logic*, 48:941–964, 1983.
- [Blok and Pigozzi, 1982] W. Blok and D. Pigozzi. On the structure of varieties with equationally definable principal congruences I. *Algebra Universalis*, 15:195–227, 1982.

- [Blok, 1976] W.J. Blok. *Varieties of interior algebras*. PhD thesis, University of Amsterdam, 1976.
- [Blok, 1978] W.J. Blok. On the degree of incompleteness in modal logics and the covering relation in the lattice of modal logics. Technical Report 78-07, Department of Mathematics, University of Amsterdam, 1978.
- [Blok, 1980a] W.J. Blok. The lattice of modal algebras is not strongly atomic. *Algebra Universalis*, 11:285–294, 1980.
- [Blok, 1980b] W.J. Blok. The lattice of modal logics: an algebraic investigation. *Journal of Symbolic Logic*, 45:221–236, 1980.
- [Blok, 1980c] W.J. Blok. Pretabular varieties of modal algebras. *Studia Logica*, 39:101–124, 1980.
- [Boolos, 1993] G. Boolos. *The Logic of Provability*. Cambridge University Press, 1993.
- [Borghuis, 1993] T. Borghuis. Interpreting modal natural deduction in type theory. In M. de Rijke, editor, *Diamonds and Defaults*, pages 67–102. Kluwer Academic Publishers, 1993.
- [Bosangue and Kwiatkowska, 1996] M. Bosangue and M. Kwiatkowska. Re-interpreting the modal  $\mu$ -calculus. In A. Ponse, M. de Rijke, and Y. Venema, editors, *Modal Logic and Process Algebra*, pages 65–83. CSLI publications, Stanford, 1996.
- [Božić and Došen, 1984] M. Božić and K. Došen. Models for normal intuitionistic logics. *Studia Logica*, 43:217–245, 1984.
- [Brachman and Schmolze, 1985] R.J. Brachman and J.G. Schmolze. An overview of the KL-ONE knowledge representation system. *Cognitive Science*, 9:171–216, 1985.
- [Büchi and Siefkes, 1973] J.R. Büchi and D. Siefkes. *The monadic second order theory of all countable ordinals*. Number 328 in Lecture Notes in Mathematics. Springer, 1973.
- [Büchi, 1962] J.R. Büchi. On a decision method in restricted second order arithmetic. In *Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress*, pages 1–11. Stanford University Press, 1962.
- [Bull, 1966a] R.A. Bull. *MIPC* as the formalization of an intuitionistic concept of modality. *Journal of Symbolic Logic*, 31:609–616, 1966.
- [Bull, 1966b] R.A. Bull. That all normal extensions of *S4.3* have the finite model property. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 12:341–344, 1966.
- [Bull, 1968] R.A. Bull. An algebraic study of tense logic with linear time. *Journal of Symbolic Logic*, 33:27–38, 1968.
- [Cerrato, 1994] C. Cerrato. Decidability by filtrations for graded normal logics. *Studia Logica*, 53:61–73, 1994.
- [Chagrov and Chagrova, 1995] A.V. Chagrov and L.A. Chagrova. Algorithmic problems concerning first order definability of modal formulas on the class of all finite frames. *Studia Logica*, 55:421–448, 1995.
- [Chagrov and Zakharyashev, 1991] A.V. Chagrov and M.V. Zakharyashev. The disjunction property of intermediate propositional logics. *Studia Logica*, 50:63–75, 1991.
- [Chagrov and Zakharyashev, 1992] A.V. Chagrov and M.V. Zakharyashev. Modal companions of intermediate propositional logics. *Studia Logica*, 51:49–82, 1992.
- [Chagrov and Zakharyashev, 1993] A.V. Chagrov and M.V. Zakharyashev. The undecidability of the disjunction property of propositional logics and other related problems. *Journal of Symbolic Logic*, 58:49–82, 1993.
- [Chagrov and Zakharyashev, 1995a] A.V. Chagrov and M.V. Zakharyashev. On the independent axiomatizability of modal and intermediate logics. *Journal of Logic and Computation*, 5:287–302, 1995.
- [Chagrov and Zakharyashev, 1995b] A.V. Chagrov and M.V. Zakharyashev. Sahlqvist formulas are not so elementary even above *S4*. In L. Csirmaz, D.M. Gabbay, and M. de Rijke, editors, *Logic Colloquium '92*, pages 61–73. CSLI Publications, Stanford, 1995.
- [Chagrov and Zakharyashev, 1997] A.V. Chagrov and M.V. Zakharyashev. *Modal Logic*. Oxford Logic Guides 35. Clarendon Press, Oxford, 1997.

- [Chagrov, 1983] A.V. Chagrov. On the polynomial approximability of modal and superintuitionistic logics. In *Mathematical Logic, Mathematical Linguistics and Algorithm Theory*, pages 75–83. Kalinin State University, Kalinin, 1983. (Russian).
- [Chagrov, 1985a] A.V. Chagrov. On the complexity of propositional logics. In *Complexity Problems in Mathematical Logic*, pages 80–90. Kalinin State University, Kalinin, 1985. (Russian).
- [Chagrov, 1985b] A.V. Chagrov. Varieties of logical matrices. *Algebra and Logic*, 24:278–325, 1985.
- [Chagrov, 1989] A.V. Chagrov. Nontabularity—pretabularity, antitabularity, co-antitabularity. In *Algebraic and Logical Constructions*, pages 105–111. Kalinin State University, Kalinin, 1989. (Russian).
- [Chagrov, 1990a] A.V. Chagrov. Undecidability of the finitary semantical consequence. In *Proceedings of the XXth USSR Conference on Mathematical Logic, Alma-Ata*, page 162, 1990. (Russian).
- [Chagrov, 1990b] A.V. Chagrov. Undecidable properties of extensions of provability logic. I. *Algebra and Logic*, 29:231–243, 1990.
- [Chagrov, 1990c] A.V. Chagrov. Undecidable properties of extensions of provability logic. II. *Algebra and Logic*, 29:406–413, 1990.
- [Chagrov, 1992a] A.V. Chagrov. Continuity of the set of maximal superintuitionistic logics with the disjunction property. *Mathematical Notes*, 51:188–193, 1992.
- [Chagrov, 1992b] A.V. Chagrov. A decidable modal logic with the undecidable admissibility problem for inference rules. *Algebra and Logic*, 31:53–55, 1992.
- [Chagrov, 1994] A.V. Chagrov. Undecidable properties of superintuitionistic logics. In S.V. Jablonskij, editor, *Mathematical Problems of Cybernetics*, volume 5, pages 67–108. Physmatlit, Moscow, 1994. (Russian).
- [Chagrov, 1995] A.V. Chagrov. One more first-order effect in Kripke semantics. In *Proceedings of the 10th International Congress of Logic, Methodology and Philosophy of Science*, page 124, Florence, Italy, 1995.
- [Chagrov, 1996] A.V. Chagrov. Tabular modal logics: algorithmic problems. Manuscript, 1996.
- [Chagrova, 1986] L.A. Chagrova. On the first order definability of intuitionistic formulas with restrictions on occurrences of the connectives. In M.I. Kanovich, editor, *Logical Methods for Constructing Effective Algorithms*, pages 135–136. Kalinin State University, Kalinin, 1986. (Russian).
- [Chagrova, 1989] L.A. Chagrova. *On the problem of definability of propositional formulas of intuitionistic logic by formulas of classical first order logic*. PhD thesis, Kalinin State University, 1989. (Russian).
- [Chagrova, 1990] L.A. Chagrova. On the preservation of first order properties under the embedding of intermediate logics into modal logics. In *Proceedings of the Xth USSR Conference for Mathematical Logic*, page 163, 1990. (Russian).
- [Chagrova, 1991] L.A. Chagrova. An undecidable problem in correspondence theory. *Journal of Symbolic Logic*, 56:1261–1272, 1991.
- [Chellas and Segerberg, 1994] B. Chellas and K. Segerberg. Modal logics with the MacIntosh-rule. *Journal of Philosophical Logic*, 23:67–86, 1994.
- [Chellas, 1980] B.F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.
- [Craig, 1953] W. Craig. On axiomatizability within a system. *Journal of Symbolic Logic*, 18:30–32, 1953.
- [Craig, 1957] W. Craig. Three uses of the Herbrandt–Gentzen theorem in relating model theory and proof theory. *Journal of Symbolic Logic*, 22:269–285, 1957.
- [Cresswell, 1984] M.J. Cresswell. An incomplete decidable modal logic. *Journal of Symbolic Logic*, 49:520–527, 1984.
- [Day, 1977] A. Day. Splitting lattices generate all lattices. *Algebra Universalis*, 7:163–170, 1977.
- [de Rijke, 1993] M. de Rijke. *Extending Modal Logic*. PhD thesis, Universiteit van Amsterdam, 1993.

- [de Rijke, 1996] M. de Rijke. A Lindström theorem for modal logic. In A. Ponse, M. de Rijke, and Y. Venema, editors, *Modal Logic and Process Algebra*, pages 217–230. CSLI Publications, Stanford, 1996.
- [Diego, 1966] A. Diego. *Sur les algèbres de Hilbert*. Gauthier-Villars, Paris, 1966.
- [Doets, 1987] K. Doets. *Completeness and definability*. PhD thesis, Universiteit van Amsterdam, 1987.
- [Donini *et al.*, 1996] F. Donini, M. Lenzerini, D. Nardi, and A. Schaerf. Reasoning in description logics. In G. Brewka, editor, *Principles of Knowledge Representation*, pages 191–236. CSLI Publications, 1996.
- [Došen, 1985a] K. Došen. Models for stronger normal intuitionistic modal logics. *Studia Logica*, 44:39–70, 1985.
- [Došen, 1985b] K. Došen. Sequent-systems for modal logic. *Journal of Symbolic Logic*, 50:149–159, 1985.
- [Došen, 1988] K. Došen. Duality between modal algebras and neighbourhood frames. *Studia Logica*, 48:219–234, 1988.
- [Drabbé, 1967] J. Drabbé. Une propriété des matrices caractéristiques des systèmes  $S_1$ ,  $S_2$ , et  $S_3$ . *Comptes Rendus de l'Académie des Sciences, Paris*, 265:A1, 1967.
- [Dugundji, 1940] J. Dugundji. Note on a property of matrices for Lewis and Langford's calculi of propositions. *Journal of Symbolic Logic*, 5:150–151, 1940.
- [Dummett and Lemmon, 1959] M.A.E. Dummett and E.J. Lemmon. Modal logics between  $S_4$  and  $S_5$ . *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 5:250–264, 1959.
- [Dummett, 1959] M.A.E. Dummett. A propositional calculus with denumerable matrix. *Journal of Symbolic Logic*, 24:97–106, 1959.
- [Ershov, 1980] Yu.L. Ershov. *Decision problems and constructivizable models*. Nauka, Moscow, 1980. (Russian).
- [Esakia and Meskhi, 1977] L.L. Esakia and V.Yu. Meskhi. Five critical systems. *Theoria*, 40:52–60, 1977.
- [Esakia, 1974] L.L. Esakia. Topological Kripke models. *Soviet Mathematics Doklady*, 15:147–151, 1974.
- [Esakia, 1979a] L.L. Esakia. On varieties of Grzegorzczuk algebras. In A. I. Mikhailov, editor, *Studies in Non-classical Logics and Set Theory*, pages 257–287. Moscow, Nauka, 1979. (Russian).
- [Esakia, 1979b] L.L. Esakia. To the theory of modal and superintuitionistic systems. In V.A. Smirnov, editor, *Logical Inference. Proceedings of the USSR Symposium on the Theory of Logical Inference*, pages 147–172. Nauka, Moscow, 1979. (Russian).
- [Ewald, 1986] W.B. Ewald. Intuitionistic tense and modal logic. *Journal of Symbolic Logic*, 51:166–179, 1986.
- [Fagin *et al.*, 1995] R. Fagin, J. Halpern, Y. Moses, and M. Vardi. *Reasoning about Knowledge*. MIT Press, 1995.
- [Ferrari and Miglioli, 1993] M. Ferrari and P. Miglioli. Counting the maximal intermediate constructive logics. *Journal of Symbolic Logic*, 58:1365–1408, 1993.
- [Ferrari and Miglioli, 1995a] M. Ferrari and P. Miglioli. A method to single out maximal propositional logics with the disjunction property. I. *Annals of Pure and Applied Logic*, 76:1–46, 1995.
- [Ferrari and Miglioli, 1995b] M. Ferrari and P. Miglioli. A method to single out maximal propositional logics with the disjunction property. II. *Annals of Pure and Applied Logic*, 76:117–168, 1995.
- [Fine and Schurz, 1996] K. Fine and G. Schurz. Transfer theorems for stratified modal logics. In J. Copeland, editor, *Logic and Reality, Essays in Pure and Applied Logic. In memory of Arthur Prior*, pages 169–213. Oxford University Press, 1996.
- [Fine, 1971] K. Fine. The logics containing  $S_4.3$ . *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 17:371–376, 1971.
- [Fine, 1972] K. Fine. In so many possible worlds. *Notre Dame Journal of Formal Logic*, 13:516–520, 1972.
- [Fine, 1974a] K. Fine. An ascending chain of  $S_4$  logics. *Theoria*, 40:110–116, 1974.
- [Fine, 1974b] K. Fine. An incomplete logic containing  $S_4$ . *Theoria*, 40:23–29, 1974.

- [Fine, 1974c] K. Fine. Logics containing  $K4$ , part I. *Journal of Symbolic Logic*, 39:229–237, 1974.
- [Fine, 1975a] K. Fine. Normal forms in modal logic. *Notre Dame Journal of Formal Logic*, 16:31–42, 1975.
- [Fine, 1975b] K. Fine. Some connections between elementary and modal logic. In S. Kanger, editor, *Proceedings of the Third Scandinavian Logic Symposium*, pages 15–31. North-Holland, Amsterdam, 1975.
- [Fine, 1985] K. Fine. Logics containing  $K4$ , part II. *Journal of Symbolic Logic*, 50:619–651, 1985.
- [Finger and Reynolds, 2000] M. Finger and M. Reynolds. Imperative history: two-dimensional executable temporal logic. In U. Reyle and H.J. Ohlbach, editors, *Logic, Language and Reasoning: Essays in Honour of Dov Gabbay*, pages 73–98. Kluwer Academic Publishers, 2000.
- [Fischer-Servi, 1977] G. Fischer-Servi. On modal logics with an intuitionistic base. *Studia Logica*, 36:141–149, 1977.
- [Fischer-Servi, 1980] G. Fischer-Servi. Semantics for a class of intuitionistic modal calculi. In M. L. Dalla Chiara, editor, *Italian Studies in the Philosophy of Science*, pages 59–72. Reidel, Dordrecht, 1980.
- [Fischer-Servi, 1984] G. Fischer-Servi. Axiomatizations for some intuitionistic modal logics. *Rend. Sem. Mat. Univers. Polit.*, 42:179–194, 1984.
- [Fitting, 1983] M. Fitting. *Proof Methods for Modal and Intuitionistic Logics*. Reidel, Dordrecht, 1983.
- [Font, 1984] J. Font. Implication and deduction in some intuitionistic modal logics. *Reports on Mathematical Logic*, 17:27–38, 1984.
- [Font, 1986] J. Font. Modality and possibility in some intuitionistic modal logics. *Notre Dame Journal of Formal Logic*, 27:533–546, 1986.
- [Friedman, 1975] H. Friedman. One hundred and two problems in mathematical logic. *Journal of Symbolic Logic*, 40:113–130, 1975.
- [Fuhrmann, 1989] A. Fuhrmann. Models for relevant modal logics. *Studia Logica*, 49:502–514, 1989.
- [Gabbay and de Jongh, 1974] D.M. Gabbay and D.H.J. de Jongh. A sequence of decidable finitely axiomatizable intermediate logics with the disjunction property. *Journal of Symbolic Logic*, 39:67–78, 1974.
- [Gabbay and Shehtman, 1993] D. Gabbay and V. Shehtman. Undecidability of modal and intermediate first-order logics with two individual variables. *Journal of Symbolic Logic*, 58:800–823, 1993.
- [Gabbay and Shehtman, 1998] D. Gabbay and V. Shehtman. Products of modal logics. Part I. *Journal of the IGPL*, 6:73–146, 1998.
- [Gabbay et al., 1994] D. Gabbay, I. Hodkinson, and M. Reynolds. *Temporal Logic: Mathematical Foundations and Computational Aspects, Volume 1*. Oxford University Press, 1994.
- [Gabbay, 1970] D.M. Gabbay. The decidability of the Kreisel–Putnam system. *Journal of Symbolic Logic*, 35:431–436, 1970.
- [Gabbay, 1971] D.M. Gabbay. On decidable, finitely axiomatizable modal and tense logics without the finite model property. I, II. *Israel Journal of Mathematics*, 10:478–495, 496–503, 1971.
- [Gabbay, 1972] D.M. Gabbay. Craig’s interpolation theorem for modal logics. In W. Hodges, editor, *Proceedings of logic conference, London 1970*, volume 255 of *Lecture Notes in Mathematics*, pages 111–127. Springer-Verlag, Berlin, 1972.
- [Gabbay, 1975] D.M. Gabbay. Decidability results in non-classical logics. *Annals of Mathematical Logic*, 8:237–295, 1975.
- [Gabbay, 1976] D.M. Gabbay. *Investigations into Modal and Tense Logics, with Applications to Problems in Linguistics and Philosophy*. Reidel, Dordrecht, 1976.
- [Gabbay, 1981a] D.M. Gabbay. An irreflexivity lemma with application to axiomatizations of conditions on linear frames. In U. Mönnich, editor, *Aspects of Philosophical Logic*, pages 67–89. Reidel, Dordrecht, 1981.
- [Gabbay, 1981b] D.M. Gabbay. *Semantical Investigations in Heyting’s Intuitionistic Logic*. Reidel, Dordrecht, 1981.



- [Gabbay, 1988] D.M. Gabbay. *Fibring Logics*. Oxford University Press, 1998.
- [Galanter, 1990] G.I. Galanter. A continuum of intermediate logics which are maximal among the logics having the intuitionistic disjunctionless fragment. In *Proceedings of 10th USSR Conference for Mathematical Logic*, page 41, Alma-Ata, 1990. (Russian).
- [Garey and Johnson, 1979] M.R. Garey and D.S. Johnson. *Computers and intractability. A guide to the theory of NP-completeness*. Freeman, San Francisco, 1979.
- [Gargov and Passy, 1990] G. Gargov and S. Passy. A note on Boolean modal logic. In P. Petkov, editor, *Mathematical Logic*, pages 299–309. Plenum Press, 1990.
- [Gargov et al., 1987] G. Gargov, S. Passy, and T. Tinchev. Modal environment for Boolean speculations. In D. Skordev, editor, *Mathematical Logic and its Applications*, pages 253–263. Plenum Press, 1987.
- [Gentzen, 1934–35] G. Gentzen. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, 39:176–210, 405–431, 1934–35.
- [Ghilardi and Meloni, 1997] S. Ghilardi and G. Meloni. Constructive canonicity in non-classical logics. *Annals of Pure and Applied Logic*, 86:1–32, 1997.
- [Ghilardi and Zawadowski, 1995] S. Ghilardi and M. Zawadowski. Undefinability of propositional quantifiers in modal system  $S4$ . *Studia Logica*, 55:259–271, 1995.
- [Ghilardi, 1995] S. Ghilardi. An algebraic theory of normal forms. *Annals of Pure and Applied Logic*, 71:189–245, 1995.
- [Ghilardi, 1999a] S. Ghilardi. Best solving modal equations. Università degli Studi di Milano, Manuscript, 1999.
- [Ghilardi, 1999b] S. Ghilardi. Unification in intuitionistic logic. *Journal of Symbolic Logic*, 64:859–880, 1999.
- [Gödel, 1932] K. Gödel. Zum intuitionistischen Aussagenkalkül. *Anzeiger der Akademie der Wissenschaften in Wien*, 69:65–66, 1932.
- [Gödel, 1933] K. Gödel. Eine Interpretation des intuitionistischen Aussagenkalküls. *Ergebnisse eines mathematischen Kolloquiums*, 4:39–40, 1933.
- [Goldblatt and Thomason, 1974] R.I. Goldblatt and S.K. Thomason. Axiomatic classes in propositional modal logic. In J. Crossley, editor, *Algebraic Logic, Lecture Notes in Mathematics vol. 450*, pages 163–173. Springer, Berlin, 1974.
- [Goldblatt, 1976a] R.I. Goldblatt. Metamathematics of modal logic, Part I. *Reports on Mathematical Logic*, 6:41–78, 1976.
- [Goldblatt, 1976b] R.I. Goldblatt. Metamathematics of modal logic, Part II. *Reports on Mathematical Logic*, 7:21–52, 1976.
- [Goldblatt, 1987] R.I. Goldblatt. *Logics of Time and Computation*. Number 7 in CSLI Lecture Notes, Stanford. CSLI, 1987.
- [Goldblatt, 1989] R.I. Goldblatt. Varieties of complex algebras. *Annals of Pure and Applied Logic*, 38:173–241, 1989.
- [Goldblatt, 1995] R.I. Goldblatt. Elementary generation and canonicity for varieties of boolean algebras with operators. *Algebra Universalis*, 34:551–607, 1995.
- [Goranko and Gargov, 1993] V. Goranko and G. Gargov. Modal logic with names. *Journal of Philosophical Logic*, 22:607–636, 1993.
- [Goranko and Passy, 1992] V. Goranko and S. Passy. Using the universal modality: Gains and questions. *Journal of Logic and Computation*, 2:5–30, 1992.
- [Goranko, 1990a] V. Goranko. Completeness and incompleteness in the bimodal base  $L(R, -R)$ . In P. Petkov, editor, *Mathematical Logic*, pages 311–326. Plenum Press, 1990.
- [Goranko, 1990b] V. Goranko. Modal definability in enriched languages. *Notre Dame Journal of Formal Logic*, 31:81–105, 1990.
- [Gore, 1994] R. Gore. Cut-free sequent and tableau systems for propositional Diodorian modal logics. *Studia Logica*, 53:433–458, 1994.
- [Grefe, 1994] C. Grefe. Modale Logiken funktionaler Frames. Master's thesis, Department of Mathematics, Freie Universität Berlin, 1994.
- [Grefe, 1997] C. Grefe. Fischer Servi's intuitionistic modal logic has the finite model property. In M. Kracht, M. de Rijke, H. Wansing, and M. Zakharyashev, editors, *Advances in Modal Logic*, pages 85–98. CSLI, Stanford, 1997.

- [Halpern and Moses, 1992] J. Halpern and Yo. Moses. A guide to completeness and complexity for modal logics of knowledge and belief. *Artificial Intelligence*, 54:319–379, 1992.
- [Harel, 1983] D. Harel. Recurring dominoes: Making the highly undecidable highly understandable. In *Conference on Foundations of Computing Theory*, volume 158 of *Lecture Notes in Computer Science*, pages 177–194. Springer, Berlin, 1983.
- [Harrop, 1958] R. Harrop. On the existence of finite models and decision procedures for propositional calculi. *Proceedings of the Cambridge Philosophical Society*, 54:1–13, 1958.
- [Hemaspaandra, 1996] E. Hemaspaandra. The price of universality. *Notre Dame Journal of Formal Logic*, 37:174–203, 1996.
- [Hirsch *et al.*, 2000] R. Hirsch, I. Hodkinson, and A. Kurucz. On modal logics between  $\mathbf{K} \times \mathbf{K} \times \mathbf{K}$  and  $\mathbf{S5} \times \mathbf{S5} \times \mathbf{S5}$ . Submitted. Available at <http://www.doc.ic.ac.uk/~kuag>.
- [Hosoi and Ono, 1973] T. Hosoi and H. Ono. Intermediate propositional logics (A survey). *Journal of Tsuda College*, 5:67–82, 1973.
- [Hosoi, 1967] T. Hosoi. On intermediate logics. *Journal of the Faculty of Science, University of Tokyo*, 14:293–312, 1967.
- [Hughes and Cresswell, 1968] G.E. Hughes and M.J. Cresswell. *A Companion to Modal Logic*. Methuen, London, 1968.
- [Humberstone, 1983] I.L. Humberstone. Inaccessible worlds. *Notre Dame Journal of Formal Logic*, 24:346–352, 1983.
- [Isard, 1977] S. Isard. A finitely axiomatizable undecidable extension of  $K$ . *Theoria*, 43:195–202, 1977.
- [Janiczak, 1953] A. Janiczak. Undecidability of some simple formalized theories. *Fundamenta Mathematicae*, 40:131–139, 1953.
- [Jankov, 1963] V.A. Jankov. The relationship between deducibility in the intuitionistic propositional calculus and finite implicational structures. *Soviet Mathematics Doklady*, 4:1203–1204, 1963.
- [Jankov, 1968a] V.A. Jankov. The calculus of the weak “law of excluded middle”. *Mathematics of the USSR, Izvestiya*, 2:997–1004, 1968.
- [Jankov, 1968b] V.A. Jankov. The construction of a sequence of strongly independent superintuitionistic propositional calculi. *Soviet Mathematics Doklady*, 9:806–807, 1968.
- [Jankov, 1969] V.A. Jankov. Conjunctively indecomposable formulas in propositional calculi. *Mathematics of the USSR, Izvestiya*, 3:17–35, 1969.
- [Jaškowski, 1936] S. Jaškowski. Recherches sur le système de la logique intuitioniste. In *Actes Du Congrès Intern. De Phil. Scientifique. VI. Phil. Des Mathématiques, Act. Sc. Et Ind 393, Paris*, pages 58–61, 1936.
- [Jipsen and Rose, 1993] P. Jipsen and H. Rose. *Varieties of Lattices*. 1993.
- [Johnson, 1969] J.S. Johnson. Nonfinitizability of classes of representable polyadic algebras. *Journal of Symbolic Logic*, 34:344–352, 1969.
- [Jónsson and Tarski, 1951] B. Jónsson and A. Tarski. Boolean algebras with operators. I. *American Journal of Mathematics*, 73:891–939, 1951.
- [Jónsson, 1994] B. Jónsson. On the canonicity of Sahlqvist identities. *Studia Logica*, 53:473–491, 1994.
- [Kashima, 1994] R. Kashima. Cut-free sequent calculi for some tense logics. *Studia Logica*, 53:119–136, 1994.
- [Kirk, 1982] R.E. Kirk. A result on propositional logics having the disjunction property. *Notre Dame Journal of Formal Logic*, 23:71–74, 1982.
- [Kleene, 1945] S. Kleene. On the interpretation of intuitionistic number theory. *Journal of Symbolic Logic*, 10:109–124, 1945.
- [Kleyman, 1984] Yu.G. Kleyman. Some questions in the theory of varieties of groups. *Mathematics of the USSR, Izvestiya*, 22:33–65, 1984.
- [Koppelberg, 1988] S. Koppelberg. General theory of Boolean algebras. In J. Monk, editor, *Handbook of Boolean Algebras*, volume 1. North-Holland, Amsterdam, 1988.
- [Kozen, 1983] D. Kozen. Results on the propositional  $\mu$ -calculus. *Theoretical Computer Science*, 27:333–354, 1983.

- [Kozen, 1988] D. Kozen. A finite model theorem for the propositional  $\mu$ -calculus. *Studia Logica*, 47:234–241, 1988.
- [Kracht and Wolter, 1991] M. Kracht and F. Wolter. Properties of independently axiomatizable bimodal logics. *Journal of Symbolic Logic*, 56:1469–1485, 1991.
- [Kracht and Wolter, 1997] M. Kracht and F. Wolter. Simulation and transfer results in modal logic: A survey. *Studia Logica*, 59:229–259, 1997.
- [Kracht and Wolter, 1999] M. Kracht and F. Wolter. Normal monomodal logics can simulate all others. *Journal of Symbolic Logic*, 64:99–138, 1999.
- [Kracht, 1990] M. Kracht. An almost general splitting theorem for modal logic. *Studia Logica*, 49:455–470, 1990.
- [Kracht, 1992] M. Kracht. Even more about the lattice of tense logics. *Archive of Mathematical Logic*, 31:243–357, 1992.
- [Kracht, 1993] M. Kracht. How completeness and correspondence theory got married. In M. de Rijke, editor, *Diamonds and Defaults*, pages 175–214. Kluwer Academic Publishers, 1993.
- [Kracht, 1999] M. Kracht. *Tools and Techniques in Modal Logic*. Elsevier, North-Holland, 1999.
- [Kreisel and Putnam, 1957] G. Kreisel and H. Putnam. Eine Unableitbarkeitsbeweismethode für den intuitionistischen Aussagenkalkül. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 3:74–78, 1957.
- [Kruskal, 1960] J. B. Kruskal. Well-quasi-ordering, the tree theorem and Vazsonyi's conjecture. *Transactions of the American Mathematical Society*, 95:210–225, 1960.
- [Kurucz, 2000] A. Kurucz. On axiomatising products of Kripke frames. *Journal of Symbolic Logic*, 65:923–945, 2000.
- [Kuznetsov and Gerchik, 1970] A.V. Kuznetsov and V.Ya. Gerchik. Superintuitionistic logics and the finite approximability. *Soviet Mathematics Doklady*, 11:1614–1619, 1970.
- [Kuznetsov, 1963] A.V. Kuznetsov. Undecidability of general problems of completeness, decidability and equivalence for propositional calculi. *Algebra and Logic*, 2:47–66, 1963. (Russian).
- [Kuznetsov, 1971] A.V. Kuznetsov. Some properties of the structure of varieties of pseudo-Boolean algebras. In *Proceedings of the XIth USSR Algebraic Colloquium*, pages 255–256. Kishinev, 1971. (Russian).
- [Kuznetsov, 1972] A.V. Kuznetsov. The decidability of certain superintuitionistic calculi. In *Proceedings of the IIth USSR Conference on Mathematical Logic*, Moscow, 1972. (Russian).
- [Kuznetsov, 1975] A.V. Kuznetsov. On superintuitionistic logics. In *Proceedings of the International Congress of Mathematicians*, pages 243–249, Vancouver, 1975.
- [Kuznetsov, 1979] A.V. Kuznetsov. Tools for detecting non-derivability or non-expressibility. In V.A. Smirnov, editor, *Logical Inference. Proceedings of the USSR Symposium on the Theory of Logical Inference*, pages 5–23. Nauka, Moscow, 1979. (Russian).
- [Kuznetsov, 1985] A.V. Kuznetsov. Proof-intuitionistic propositional calculus. *Doklady Akademii Nauk SSSR*, 283:27–30, 1985. (Russian).
- [Ladner, 1977] R.E. Ladner. The computational complexity of provability in systems of modal logic. *SIAM Journal on Computing*, 6:467–480, 1977.
- [Lemmon and Scott, 1977] E.J. Lemmon and D.S. Scott. *An Introduction to Modal Logic*. Oxford, Blackwell, 1977.
- [Lemmon, 1966a] E.J. Lemmon. Algebraic semantics for modal logic. I. *Journal of Symbolic Logic*, 31:46–65, 1966.
- [Lemmon, 1966b] E.J. Lemmon. Algebraic semantics for modal logic. II. *Journal of Symbolic Logic*, 31:191–218, 1966.
- [Lemmon, 1966c] E.J. Lemmon. A note on Halldén-incompleteness. *Notre Dame Journal of Formal Logic*, 7:296–300, 1966.
- [Levin, 1969] V.A. Levin. Some syntactic theorems on the calculus of finite problems of Yu.T. Medvedev. *Soviet Mathematics Doklady*, 10:288–290, 1969.
- [Lewis, 1918] C.I. Lewis. *A Survey of Symbolic Logic*. University of California Press, Berkeley, 1918.

- [Lewis, 1974] D. Lewis. Intensional logics without iterative axioms. *Journal of Philosophical Logic*, 3:457–466, 1974.
- [Lincoln *et al.*, 1992] P.D. Lincoln, J. Mitchell, A. Scedrov, and N. Shankar. Decision problems for propositional linear logic. *Annals of Pure and Applied Logic*, 56:239–311, 1992.
- [Lukasiewicz, 1952] J. Lukasiewicz. On the intuitionistic theory of deduction. *Indagationes Mathematicae*, 14:202–212, 1952.
- [Luppi, 1996] C. Luppi. On the interpolation property of some intuitionistic modal logics. *Archive for Mathematical Logic*, 35:173–189, 1996.
- [Maddux, 1980] R. Maddux. The equational theory of  $CA_3$  is undecidable. *Journal of Symbolic Logic*, 45:311–315, 1980.
- [Makinson, 1971] D.C. Makinson. Some embedding theorems for modal logic. *Notre Dame Journal of Formal Logic*, 12:252–254, 1971.
- [Maksimova and Rybakov, 1974] L.L. Maksimova and V.V. Rybakov. Lattices of modal logics. *Algebra and Logic*, 13:105–122, 1974.
- [Maksimova *et al.*, 1979] L.L. Maksimova, V.B. Shehtman, and D.P. Skvortsov. The impossibility of a finite axiomatization of Medvedev's logic of finitary problems. *Soviet Mathematics Doklady*, 20:394–398, 1979.
- [Maksimova, 1972] L.L. Maksimova. Pretabular superintuitionistic logics. *Algebra and Logic*, 11:308–314, 1972.
- [Maksimova, 1975a] L.L. Maksimova. Modal logics of finite slices. *Algebra and Logic*, 14:188–197, 1975.
- [Maksimova, 1975b] L.L. Maksimova. Pretabular extensions of Lewis  $S_4$ . *Algebra and Logic*, 14:16–33, 1975.
- [Maksimova, 1979] L.L. Maksimova. Interpolation theorems in modal logic and amalgamable varieties of topological Boolean algebras. *Algebra and Logic*, 18:348–370, 1979.
- [Maksimova, 1982a] L.L. Maksimova. Failure of the interpolation property in modal companions of Dummett's logic. *Algebra and Logic*, 21:690–694, 1982.
- [Maksimova, 1982b] L.L. Maksimova. Lyndon's interpolation theorem in modal logics. In *Mathematical Logic and Algorithm Theory*, pages 45–55. Institute of Mathematics, Novosibirsk, 1982. (Russian).
- [Maksimova, 1984] L.L. Maksimova. On the number of maximal intermediate logics having the disjunction property. In *Proceedings of the 7th USSR Conference for Mathematical Logic*, page 95. Institute of Mathematics, Novosibirsk, 1984. (Russian).
- [Maksimova, 1986] L.L. Maksimova. On maximal intermediate logics with the disjunction property. *Studia Logica*, 45:69–75, 1986.
- [Maksimova, 1987] L.L. Maksimova. On the interpolation in normal modal logics. *Non-classical Logics, Studies in Mathematics*, 98:40–56, 1987. (Russian).
- [Maksimova, 1989] L.L. Maksimova. A continuum of normal extensions of the modal provability logic with the interpolation property. *Sibirskij Matematičeskij Žurnal*, 30:122–131, 1989. (Russian).
- [Maksimova, 1992] L.L. Maksimova. Definability and interpolation in classical modal logics. *Contemporary Mathematics*, 131:583–599, 1992.
- [Maksimova, 1995] L.L. Maksimova. On variable separation in modal and superintuitionistic logics. *Studia Logica*, 55:99–112, 1995.
- [Mal'cev, 1970] A.I. Mal'cev. *Algorithms and Recursive Functions*. Wolters-Noordhoff, Groningen, 1970.
- [Mal'cev, 1973] A.I. Mal'cev. *Algebraic Systems*. Springer-Verlag, Berlin-Heidelberg, 1973.
- [Mardaev, 1984] S.I. Mardaev. The number of prelocally tabular superintuitionistic propositional logics. *Algebra and Logic*, 23:56–66, 1984.
- [Marx and Areces, 1998] M. Marx and C. Areces. Failure of interpolation in combined modal logics. *Notre Dame Journal of Formal Logic*, 39:253–273, 1998.
- [Marx and Venema, 1997] M. Marx and Y. Venema. *Multi-dimensional modal logic*. Kluwer Academic Publishers, 1997.
- [Marx, 1995] M. Marx. *Algebraic relativization and arrow logic*. PhD thesis, University of Amsterdam, 1995.

- [Marx, 1999] M. Marx. Complexity of products of modal logics. *Journal of Logic and Computation*, 9:197–214, 1999.
- [Masini, 1992] A. Masini. 2-sequent calculus: a proof theory of modality. *Annals of Pure and Applied Logic*, 58:229–246, 1992.
- [Matiyasevich, 1967] Y.V. Matiyasevich. Simple examples of undecidable associative calculi. *Soviet Mathematics Doklady*, 8:555–557, 1967.
- [McKay, 1968] C.G. McKay. The decidability of certain intermediate logics. *Journal of Symbolic Logic*, 33:258–264, 1968.
- [McKay, 1971] C.G. McKay. A class of decidable intermediate propositional logics. *Journal of Symbolic Logic*, 36:127–128, 1971.
- [McKenzie, 1972] R. McKenzie. Equational bases and non-modular lattice varieties. *Transactions of the American Mathematical Society*, 174:1–43, 1972.
- [McKinsey and Tarski, 1946] J.C.C. McKinsey and A. Tarski. On closed elements in closure algebras. *Annals of Mathematics*, 47:122–162, 1946.
- [McKinsey and Tarski, 1948] J.C.C. McKinsey and A. Tarski. Some theorems about the sentential calculi of Lewis and Heyting. *Journal of Symbolic Logic*, 13:1–15, 1948.
- [McKinsey, 1941] J.C.C. McKinsey. A solution of the decision problem for the Lewis systems  $S_2$  and  $S_4$ , with an application to topology. *Journal of Symbolic Logic*, 6:117–134, 1941.
- [Medvedev, 1962] Yu.T. Medvedev. Finite problems. *Soviet Mathematics Doklady*, 3:227–230, 1962.
- [Medvedev, 1966] Yu.T. Medvedev. Interpretation of logical formulas by means of finite problems. *Soviet Mathematics Doklady*, 7:857–860, 1966.
- [Meyer and van der Hoek, 1995] J. Meyer and W. van der Hoek. *Epistemic Logic for AI and Computer Science*. Cambridge University Press, 1995.
- [Mikulas, 1995] S. Mikulas. *Taming Logics*. PhD thesis, University of Amsterdam, 1995.
- [Minari, 1986] P. Minari. Intermediate logics with the same disjunctionless fragment as intuitionistic logic. *Studia Logica*, 45:207–222, 1986.
- [Montagna, 1987] F. Montagna. Provability in finite subtheories of PA and relative interpretability: a modal investigation. *Journal of Symbolic Logic*, 52:494–511, 1987.
- [Morikawa, 1989] O. Morikawa. Some modal logics based on three-valued logic. *Notre Dame Journal of Formal Logic*, 30:130–137, 1989.
- [Muravitskij, 1981] A.Yu. Muravitskij. On finite approximability of the calculus  $I^\Delta$  and non-modelability of some of its extensions. *Mathematical Notes*, 29:907–916, 1981.
- [Nagle and Thomason, 1985] M.C. Nagle and S.K. Thomason. The extensions of the modal logic  $K_5$ . *Journal of Symbolic Logic*, 50:102–108, 1985.
- [Nishimura, 1960] I. Nishimura. On formulas of one variable in intuitionistic propositional calculus. *Journal of Symbolic Logic*, 25:327–331, 1960.
- [Ono and Nakamura, 1980] H. Ono and A. Nakamura. On the size of refutation Kripke models for some linear modal and tense logics. *Studia Logica*, 39:325–333, 1980.
- [Ono and Suzuki, 1988] H. Ono and N. Suzuki. Relations between intuitionistic modal logics and intermediate predicate logics. *Reports on Mathematical Logic*, 22:65–87, 1988.
- [Ono, 1972] H. Ono. Some results on the intermediate logics. *Publications of the Research Institute for Mathematical Science, Kyoto University*, 8:117–130, 1972.
- [Ono, 1977] H. Ono. On some intuitionistic modal logics. *Publications of the Research Institute for Mathematical Science, Kyoto University*, 13:55–67, 1977.
- [Orlov, 1928] I.E. Orlov. The calculus of compatibility of propositions. *Mathematics of the USSR, Sbornik*, 35:263–286, 1928. (Russian).
- [Ostermann, 1988] P. Ostermann. Many-valued modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 34:343–354, 1988.
- [Pigozzi, 1974] D. Pigozzi. The join of equational theories. *Colloquium Mathematicum*, 30:15–25, 1974.
- [Pitts, 1992] A.M. Pitts. On an interpretation of second order quantification in first order intuitionistic propositional logic. *Journal of Symbolic Logic*, 57:33–52, 1992.
- [Ponse et al., 1996] A. Ponse, M. de Rijke, and Y. Venema. *Modal Logic and Process Algebra*. CSLI Publications, Stanford, 1996.
- [Prior, 1957] A. Prior. *Time and Modality*. Clarendon Press, Oxford, 1957.

- [Rabin, 1969] M.O. Rabin. Decidability of second order theories and automata on infinite trees. *Transactions of the American Mathematical Society*, 141:1–35, 1969.
- [Rabin, 1977] M.O. Rabin. Decidable theories. In J. Barwise, editor, *Handbook of Mathematical Logic*, pages 595–630. Elsevier, North-Holland, 1977.
- [Rasiowa and Sikorski, 1963] H. Rasiowa and R. Sikorski. *The Mathematics of Metamathematics*. Polish Scientific Publishers, 1963.
- [Rautenberg, 1977] W. Rautenberg. Der Verband der normalen verzweigten Modallogiken. *Mathematische Zeitschrift*, 156:123–140, 1977.
- [Rautenberg, 1979] W. Rautenberg. *Klassische und nichtklassische Aussagenlogik*. Vieweg, Braunschweig–Wiesbaden, 1979.
- [Rautenberg, 1980] W. Rautenberg. Splitting lattices of logics. *Archiv für Mathematische Logik*, 20:155–159, 1980.
- [Rautenberg, 1983] W. Rautenberg. Modal tableau calculi and interpolation. *Journal of Philosophical Logic*, 12:403–423, 1983.
- [Reif and Sistla, 1985] J. Reif and A. Sistla. A multiprocess network logic with temporal and spatial modalities. *Journal of Computer and System Sciences*, 30:41–53, 1985.
- [Reynolds and Zakharyashev, 2000] M. Reynolds and M. Zakharyashev. On the products of linear modal logics. *Journal of Logic and Computation*, 2000. (To appear.)
- [Rieger, 1949] L. Rieger. On the lattice of Brouwerian propositional logics. *Acta Universitatis Carolinae. Mathematica et Physica*, 189, 1949.
- [Robinson, 1971] R. Robinson. Undecidability and nonperiodicity for tilings of the plane. *Inventiones Math.*, 12:177–209, 1971.
- [Rodenburg, 1986] P.H. Rodenburg. *Intuitionistic correspondence theory*. PhD thesis, University of Amsterdam, 1986.
- [Rose, 1953] G.F. Rose. Propositional calculus and realizability. *Transactions of the American Mathematical Society*, 75:1–19, 1953.
- [Rybakov, 1977] V.V. Rybakov. Noncompact extensions of the logic  $S_4$ . *Algebra and Logic*, 16:321–334, 1977.
- [Rybakov, 1978] V.V. Rybakov. Modal logics with LM-axioms. *Algebra and Logic*, 17:302–310, 1978.
- [Rybakov, 1984a] V.V. Rybakov. Admissible rules for logics containing  $S_4.3$ . *Siberian Mathematical Journal*, 25:795–798, 1984.
- [Rybakov, 1984b] V.V. Rybakov. A criterion for admissibility of rules in the modal system  $S_4$  and intuitionistic logic. *Algebra and Logic*, 23:369–384, 1984.
- [Rybakov, 1987] V.V. Rybakov. The decidability of admissibility of inference rules in the modal system  $Grz$  and intuitionistic logic. *Mathematics of the USSR, Izvestiya*, 28:589–608, 1987.
- [Rybakov, 1989] V.V. Rybakov. Admissibility of inference rules in the modal system  $G$ . *Mathematical Logic and Algorithmical Problems, Mathematical Institute, Novosibirsk*, 12:120–138, 1989. (Russian).
- [Rybakov, 1993] V.V. Rybakov. Rules of inference with parameters for intuitionistic logic. *Journal of Symbolic Logic*, 58:1803–1834, 1993.
- [Rybakov, 1994] V.V. Rybakov. Criteria for admissibility of inference rules. Modal and intermediate logics with the branching property. *Studia Logica*, 53:203–226, 1994.
- [Rybakov, 1995] V.V. Rybakov. Hereditarily structurally complete modal logics. *Journal of Symbolic Logic*, 60:266–288, 1995.
- [Sahlqvist, 1975] H. Sahlqvist. Completeness and correspondence in the first and second order semantics for modal logic. In S. Kanger, editor, *Proceedings of the Third Scandinavian Logic Symposium*, pages 110–143. North-Holland, Amsterdam, 1975.
- [Sambin and Vaccaro, 1989] G. Sambin and V. Vaccaro. A topological proof of Sahlqvist's theorem. *Journal of Symbolic Logic*, 54:992–999, 1989.
- [Sasaki, 1992] K. Sasaki. The disjunction property of the logics with axioms of only one variable. *Bulletin of the Section of Logic*, 21:40–46, 1992.
- [Schild, 1991] K. Schild. A correspondence theory for terminological logics: preliminary report. In *Proc. of the 12th Int. Joint Conf. on Artificial Intelligence (IJCAI-91)*, pages 466–471, Sydney, 1991.
- [Scroggs, 1951] S.J. Scroggs. Extensions of the Lewis system  $S_5$ . *Journal of Symbolic Logic*, 16:112–120, 1951.

- [Segerberg, 1967] K. Segerberg. Some modal logics based on three valued logic. *Theoria*, 33:53–71, 1967.
- [Segerberg, 1970] K. Segerberg. Modal logics with linear alternative relations. *Theoria*, 36:301–322, 1970.
- [Segerberg, 1971] K. Segerberg. An essay in classical modal logic. *Philosophical Studies*, Uppsala, 13, 1971.
- [Segerberg, 1973] K. Segerberg. Two-dimensional modal logic. *Journal of Philosophical Logic*, 2:77–96, 1973.
- [Segerberg, 1975] K. Segerberg. That all extensions of  $S4.3$  are normal. In S. Kanger, editor, *Proceedings of the Third Scandinavian Logic Symposium*, pages 194–196. North-Holland, Amsterdam, 1975.
- [Segerberg, 1986] K. Segerberg. Modal logics with functional alternative relations. *Notre Dame Journal of Formal Logic*, 27:504–522, 1986.
- [Segerberg, 1989] K. Segerberg. Von Wright's tense logic. In P. Schilpp and L. Hahn, editors, *The Philosophy of Georg Henrik von Wright*, pages 603–635. La Salle, IL: Open Court, 1989.
- [Shavrukov, 1991] V.Yu. Shavrukov. On two extensions of the provability logic  $GL$ . *Mathematics of the USSR, Sbornik*, 69:255–270, 1991.
- [Shavrukov, 1993] V.Yu. Shavrukov. Subalgebras of diagonalizable algebras of theories containing arithmetic. *Dissertationes Mathematicae (Rozprawy Matematyczne, Polska Akademia Nauk, Instytut Matematyczny)*, Warszawa, 323, 1993.
- [Shehtman, 1977] V.B. Shehtman. On incomplete propositional logics. *Soviet Mathematics Doklady*, 18:985–989, 1977.
- [Shehtman, 1978a] V.B. Shehtman. Rieger–Nishimura lattices. *Soviet Mathematics Doklady*, 19:1014–1018, 1978.
- [Shehtman, 1978b] V.B. Shehtman. Two-dimensional modal logics. *Mathematical Notices of the USSR Academy of Sciences*, 23:417–424, 1978. (Translated from Russian).
- [Shehtman, 1978c] V.B. Shehtman. An undecidable superintuitionistic propositional calculus. *Soviet Mathematics Doklady*, 19:656–660, 1978.
- [Shehtman, 1979] V.B. Shehtman. Kripke type semantics for propositional modal logics with the intuitionistic base. In V.A. Smirnov, editor, *Modal and Tense Logics*, pages 108–112. Institute of Philosophy, USSR Academy of Sciences, 1979. (Russian).
- [Shehtman, 1980] V.B. Shehtman. Topological models of propositional logics. *Semiotics and Information Science*, 15:74–98, 1980. (Russian).
- [Shehtman, 1982] V.B. Shehtman. Undecidable propositional calculi. In *Problems of Cybernetics. Nonclassical logics and their application*, volume 75, pages 74–116. USSR Academy of Sciences, 1982. (Russian).
- [Shimura, 1993] T. Shimura. Kripke completeness of some intermediate predicate logics with the axiom of constant domain and a variant of canonical formulas. *Studia Logica*, 52:23–40, 1993.
- [Shimura, 1995] T. Shimura. On completeness of intermediate predicate logics with respect to Kripke semantics. *Bulletin of the Section of Logic*, 24:41–45, 1995.
- [Shum, 1985] A.A. Shum. Relative varieties of algebraic systems, and propositional calculi. *Soviet Mathematics Doklady*, 31:492–495, 1985.
- [Simpson, 1994] A.K. Simpson. *The proof theory and semantics of intuitionistic modal logic*. PhD thesis, University of Edinburgh, 1994.
- [Smoryński, 1973] C. Smoryński. *Investigations of Intuitionistic Formal Systems by means of Kripke Frames*. PhD thesis, University of Illinois, 1973.
- [Smoryński, 1978] C. Smoryński. Beth's theorem and self-referential sentences. In *Logic Colloquium 77*, pages 253–261. North-Holland, Amsterdam, 1978.
- [Smoryński, 1985] C. Smoryński. *Self-reference and Modal Logic*. Springer Verlag, Heidelberg & New York, 1985.
- [Sobolev, 1977a] S.K. Sobolev. On finite-dimensional superintuitionistic logics. *Mathematics of the USSR, Izvestiya*, 11:909–935, 1977.
- [Sobolev, 1977b] S.K. Sobolev. On the finite approximability of superintuitionistic logics. *Mathematics of the USSR, Sbornik*, 31:257–268, 1977.
- [Solovay, 1976] R. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 25:287–304, 1976.

- [Sotirov, 1984] V.H. Sotirov. Modal theories with intuitionistic logic. In *Proceedings of the Conference on Mathematical Logic, Sofia, 1980*, pages 139–171. Bulgarian Academy of Sciences, 1984.
- [Spaan, 1993] E. Spaan. *Complexity of Modal Logics*. PhD thesis, Department of Mathematics and Computer Science, University of Amsterdam, 1993.
- [Statman, 1979] R. Statman. Intuitionistic propositional logic is polynomial-space complete. *Theoretical Computer Science*, 9:67–72, 1979.
- [Surendonk, 1996] T. Surendonk. Canonicity of intensional logics without iterative axioms. *Journal of Philosophical Logic*, 1996. To appear.
- [Suzuki, 1990] N. Suzuki. An algebraic approach to intuitionistic modal logics in connection with intermediate predicate logics. *Studia Logica*, 48:141–155, 1990.
- [Tarski, 1954] A. Tarski. Contributions to the theory of models I, II. *Indagationes Mathematicae*, 16:572–588, 1954.
- [Thomason, 1972] S. K. Thomason. Noncompactness in propositional modal logic. *Journal of Symbolic Logic*, 37:716–720, 1972.
- [Thomason, 1974a] S. K. Thomason. An incompleteness theorem in modal logic. *Theoria*, 40:30–34, 1974.
- [Thomason, 1974b] S. K. Thomason. Reduction of tense logic to modal logic I. *Journal of Symbolic Logic*, 39:549–551, 1974.
- [Thomason, 1975a] S. K. Thomason. The logical consequence relation of propositional tense logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 21:29–40, 1975.
- [Thomason, 1975b] S. K. Thomason. Reduction of second-order logic to modal logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 21:107–114, 1975.
- [Thomason, 1975c] S. K. Thomason. Reduction of tense logic to modal logic II. *Theoria*, 41:154–169, 1975.
- [Thomason, 1980] S. K. Thomason. Independent propositional modal logics. *Studia Logica*, 39:143–144, 1980.
- [Thomason, 1982] S. K. Thomason. Undecidability of the completeness problem of modal logic. In *Universal Algebra and Applications, Banach Center Publications*, volume 9, pages 341–345, Warsaw, 1982. PNW–Polish Scientific Publishers.
- [Tseitin, 1958] G.S. Tseitin. Associative calculus with unsolvable equivalence problem. *Proceedings of the Mathematical Steklov Institute of the USSR Academy of Sciences*, 52:172–189, 1958. Translation: American Mathematical Society. Translations. Series 2. 94:73–92.
- [Tsytkin, 1978] A.I. Tsytkin. On structurally complete superintuitionistic logics. *Soviet Mathematics Doklady*, 19:816–819, 1978.
- [Tsytkin, 1987] A.I. Tsytkin. Structurally complete superintuitionistic logics and primitive varieties of pseudo-Boolean algebras. *Mathematical Studies*, 98:134–151, 1987. (Russian).
- [Umezawa, 1955] T. Umezawa. Über die Zwischensysteme der Aussagenlogik. *Nagoya Mathematical Journal*, 9:181–189, 1955.
- [Umezawa, 1959] T. Umezawa. On intermediate propositional logics. *Journal of Symbolic Logic*, 24:20–36, 1959.
- [Urquhart, 1974] A. Urquhart. Implicational formulas in intuitionistic logic. *Journal of Symbolic Logic*, 39:661–664, 1974.
- [Urquhart, 1984] A. Urquhart. The undecidability of entailment and relevant implication. *Journal of Symbolic Logic*, 49:1059–1073, 1984.
- [Vakarelov, 1981] D. Vakarelov. Intuitionistic modal logics incompatible with the law of excluded middle. *Studia Logica*, 40:103–111, 1981.
- [Vakarelov, 1985] D. Vakarelov. An application of the Rieger–Nishimura formulas to the intuitionistic modal logics. *Studia Logica*, 44:79–85, 1985.
- [van Benthem and Blok, 1978] J.A.F.K. van Benthem and W.J. Blok. Transitivity follows from Dummett’s axiom. *Theoria*, 44:117–118, 1978.
- [van Benthem and Humberstone, 1983] J.A.F.K. van Benthem and I.L. Humberstone. Halldén-completeness by gluing Kripke frames. *Notre Dame Journal of Formal Logic*, 24:426–430, 1983.



- [van Benthem, 1976] J.A.F.K. van Benthem. Modal reduction principles. *Journal of Symbolic Logic*, 41:301–312, 1976.
- [van Benthem, 1979] J.A.F.K. van Benthem. Syntactic aspects of modal incompleteness theorems. *Theoria*, 45:63–77, 1979.
- [van Benthem, 1980] J.A.F.K. van Benthem. Some kinds of modal completeness. *Studia Logica*, 39:125–141, 1980.
- [van Benthem, 1983] J.A.F.K. van Benthem. *Modal Logic and Classical Logic*. Bibliopolis, Napoli, 1983.
- [van Benthem, 1991] J.A.F.K. van Benthem. *The Logic of Time. A Model-Theoretic Investigation into the Varieties of Temporal Ontology and Temporal Discourse*. Kluwer Academic Publishers, 1991.
- [van der Hoek, 1992] W. van der Hoek. *Modalities for Reasoning about Knowledge and Quantities*. PhD thesis, University of Amsterdam, 1992.
- [Venema, 1991] Y. Venema. *Many-Dimensional Modal Logics*. PhD thesis, Universiteit van Amsterdam, 1991.
- [Visser, 1995] A. Visser. A course in bimodal provability logic. *Annals of Pure and Applied Logic*, 73:115–142, 1995.
- [Visser, 1996] A. Visser. Uniform interpolation and layered bisimulation. In P. Hayek, editor, *Gödel'96*, pages 139–164. Springer Verlag, 1996.
- [Walukiewicz, 1993] I. Walukiewicz. *A Complete Deduction system for the  $\mu$ -calculus*. PhD thesis, Warsaw, 1993.
- [Walukiewicz, 1996] I. Walukiewicz. A note on the completeness of Kozen's axiomatization of the propositional  $\mu$ -calculus. *Bulletin of Symbolic Logic*, 2:349–366, 1996.
- [Wang, 1992] X. Wang. The McKinsey axiom is not compact. *Journal of Symbolic Logic*, 57:1230–1238, 1992.
- [Wansing, 1994] H. Wansing. Sequent calculi for normal modal propositional logics. *Journal of Logic and Computation*, 4:125–142, 1994.
- [Wansing, 1996] H. Wansing. *Proof Theory of Modal Logic*. Kluwer Academic Publishers, 1996.
- [Whitman, 1943] P. Whitman. Splittings of a lattice. *American Journal of Mathematics*, 65:179–196, 1943.
- [Wijesekera, 1990] D. Wijesekera. Constructive modal logic I. *Annals of Pure and Applied Logic*, 50:271–301, 1990.
- [Williamson, 1994] T. Williamson. Non-genuine MacIntosh logics. *Journal of Philosophical Logic*, 23:87–101, 1994.
- [Wolter and Zakharyashev, 1997] F. Wolter and M. Zakharyashev. On the relation between intuitionistic and classical modal logics. *Algebra and Logic*, 36:121–155, 1997.
- [Wolter and Zakharyashev, 1998] F. Wolter and M. Zakharyashev. Satisfiability problem in description logics with modal operators. In *Proceedings of the sixth Conference on Principles of Knowledge Representation and Reasoning, KR'98, Trento, Italy*, pages 512–523, 1998. Morgan Kaufman.
- [Wolter and Zakharyashev, 1999a] F. Wolter and M. Zakharyashev. Intuitionistic modal logics as fragments of classical bimodal logics. In E. Orłowska, editor, *Logic at Work*, pages 168–186. Springer-Verlag, 1999.
- [Wolter and Zakharyashev, 1999b] F. Wolter and M. Zakharyashev. Modal description logics: modalizing roles. *Fundamenta Informaticae*, 30:411–438, 1999.
- [Wolter and Zakharyashev, 1999c] F. Wolter and M. Zakharyashev. Multi-dimensional description logics. In *Proceedings of the 16th International Joint Conference on Artificial Intelligence, IJCAI'99, Stockholm*, pages 104–109, 1999. Morgan Kaufman.
- [Wolter and Zakharyashev, 2000] F. Wolter and M. Zakharyashev. Spatio-temporal representation and reasoning based on RCC-8. In *Proceedings of the seventh Conference on Principles of Knowledge Representation and Reasoning, KR2000, Breckenridge, USA*, pages 1–12, 2000. Morgan Kaufman.
- [Wolter, 1993] F. Wolter. *Lattices of Modal Logics*. PhD thesis, Freie Universität Berlin, 1993. Parts of the thesis appeared in *Annals of Pure and Applied Logic*, 86:47–100, 1997, under the title “The structure of lattices of subframe logics”.
- [Wolter, 1994a] F. Wolter. Solution to a problem of Goranko and Passy. *Journal of Logic and Computation*, 4:21–22, 1994.

- [Wolter, 1994b] F. Wolter. What is the upper part of the lattice of bimodal logics? *Studia Logica*, 53:235–242, 1994.
- [Wolter, 1995] F. Wolter. The finite model property in tense logic. *Journal of Symbolic Logic*, 60:757–774, 1995.
- [Wolter, 1996a] F. Wolter. A counterexample in tense logic. *Notre Dame Journal of Formal Logic*, 37:167–173, 1996.
- [Wolter, 1996b] F. Wolter. Properties of tense logics. *Mathematical Logic Quarterly*, 42:481–500, 1996.
- [Wolter, 1996c] F. Wolter. Tense logics without tense operators. *Mathematical Logic Quarterly*, 42:145–171, 1996.
- [Wolter, 1997a] F. Wolter. Completeness and decidability of tense logics closely related to logics containing  $K4$ . *Journal of Symbolic Logic*, 62:131–158, 1997.
- [Wolter, 1997b] F. Wolter. Fusions of modal logics revisited. In M. Kracht, M. de Rijke, H. Wansing, and M. Zakharyashev, editors, *Advances in Modal Logic*. CSLI, Stanford, 1997.
- [Wolter, 1997c] F. Wolter. A note on atoms in polymodal algebras. *Algebra Universalis*, 37:334–341, 1997.
- [Wolter, 1997d] F. Wolter. A note on the interpolation property in tense logic. *Journal of Philosophical Logic*, 26:545–551, 1997.
- [Wolter, 1997e] F. Wolter. Superintuitionistic companions of classical modal logics. *Studia Logica*, 58:229–259, 1997.
- [Wolter, 1998] F. Wolter. All finitely axiomatizable subframe logics containing CSM are decidable. *Archive for Mathematical Logic*, 37:167–182, 1998.
- [Wolter, 2000] F. Wolter. The product of converse PDL and polymodal  $K$ . *Journal of Logic and Computation*, 10:223–251, 2000.
- [Wroński, 1973] A. Wroński. Intermediate logics and the disjunction property. *Reports on Mathematical Logic*, 1:39–51, 1973.
- [Wroński, 1974] A. Wroński. Remarks on intermediate logics with axioms containing only one variable. *Reports on Mathematical Logic*, 2:63–75, 1974.
- [Wroński, 1989] A. Wroński. Sufficient condition of decidability for intermediate propositional logics. In *ASL Logic Colloquium, Berlin'89*, 1989.
- [Zakharyashev and Alekseev, 1995] M. Zakharyashev and A. Alekseev. All finitely axiomatizable normal extensions of  $K4.3$  are decidable. *Mathematical Logic Quarterly*, 41:15–23, 1995.
- [Zakharyashev and Popov, 1979] M.V. Zakharyashev and S.V. Popov. On the complexity of Kripke countermodels in intuitionistic propositional calculus. In *Proceedings of the 2nd Soviet-Finland Logic Colloquium*, pages 32–36, 1979. (Russian).
- [Zakharyashev, 1983] M.V. Zakharyashev. On intermediate logics. *Soviet Mathematics Doklady*, 27:274–277, 1983.
- [Zakharyashev, 1984] M.V. Zakharyashev. Normal modal logics containing  $S4$ . *Soviet Mathematics Doklady*, 28:252–255, 1984.
- [Zakharyashev, 1987] M.V. Zakharyashev. On the disjunction property of superintuitionistic and modal logics. *Mathematical Notes*, 42:901–905, 1987.
- [Zakharyashev, 1988] M.V. Zakharyashev. Syntax and semantics of modal logics containing  $S4$ . *Algebra and Logic*, 27:408–428, 1988.
- [Zakharyashev, 1989] M.V. Zakharyashev. Syntax and semantics of intermediate logics. *Algebra and Logic*, 28:262–282, 1989.
- [Zakharyashev, 1991] M.V. Zakharyashev. Modal companions of superintuitionistic logics: syntax, semantics and preservation theorems. *Mathematics of the USSR, Sbornik*, 68:277–289, 1991.
- [Zakharyashev, 1992] M.V. Zakharyashev. Canonical formulas for  $K4$ . Part I: Basic results. *Journal of Symbolic Logic*, 57:1377–1402, 1992.
- [Zakharyashev, 1994] M.V. Zakharyashev. A new solution to a problem of Hosoi and Ono. *Notre Dame Journal of Formal Logic*, 35:450–457, 1994.
- [Zakharyashev, 1996] M.V. Zakharyashev. Canonical formulas for  $K4$ . Part II: Cofinal subframe logics. *Journal of Symbolic Logic*, 61:421–449, 1996.
- [Zakharyashev, 1997a] M.V. Zakharyashev. Canonical formulas for  $K4$ . Part III: The finite model property. *Journal of Symbolic Logic*, 62:950–975, 1997.

- [Zakharyashev, 1997b] M.V. Zakharyashev. Canonical formulas for modal and super-intuitionistic logics: a short outline. In M. de Rijke, editor, *Advances in Intensional Logic*, pages 191–243. Kluwer Academic Publishers, 1997.
- [Zakharyashev, 1997c] M.V. Zakharyashev. The greatest extension of  $S4$  into which intuitionistic logic is embeddable. *Studia Logica*, 59:345–358, 1997.

## QUANTIFICATION IN MODAL LOGIC

### 0 INTRODUCTION

#### *0.1 An Outline of this Chapter*

The novice may wonder why quantified modal logic (QML) is considered difficult. QML would seem to be easy: simply add the principles of first-order logic to propositional modal logic. Unfortunately, this choice does not correspond to an intuitively satisfying semantics. From the semantical point of view, we are confronted with a number of decisions concerning the quantifiers, and these in turn prompt new questions about the semantics of identity, terms, and predicates. Since most of the choices can be made independently, the number of interesting quantified modal logics seems bewilderingly large.

The main purpose of this chapter is to try to make sense of this seemingly chaotic terrain. Section 1 provides a review of the major systems. Section 2 explains the difficulties in completeness proofs for QMLs, and presents strategies for overcoming them. Section 3 shows that some systems of QML behave like second-order logics; they have strong expressive powers and so are incomplete. The Appendix lists rules, systems, and semantical conditions covered in this chapter.

Free logic serves, in one way or another, as the foundation for most of the systems we will study. We will argue in Section 1.2.1.2 that allegiance to first-order logic is a source of *ad hoc* stipulations in semantics for QML. However, when the principles of free logic are adopted, complications can be avoided. Since free logic is such a crucial foundation for QML, we will give a brief description of it here. The reader who knows about free logic, or who wants to read Bencivena's chapter (in Volume 7 of this *Handbook*) on the topic, may skip section 0.2. Since free logics are usually formulated using = in QML in any case, we will briefly discuss identity in intensional logics in Section 0.3.

#### *0.2 A Short Review of Free Logic*

One oddity of first-order logic with identity is that it seems to provide an argument for the existence of God. From the provable identity  $g = g$  we may derive,  $\exists x x = g$  by Existential Generalisation. If  $g$  abbreviates 'God', then  $\exists x x = g$  reads 'God exists'. This anomaly is connected with the basic assumption made in the semantics for quantificational logic that every constant (such as  $g$ ) refers to an object in the domain of quantification.

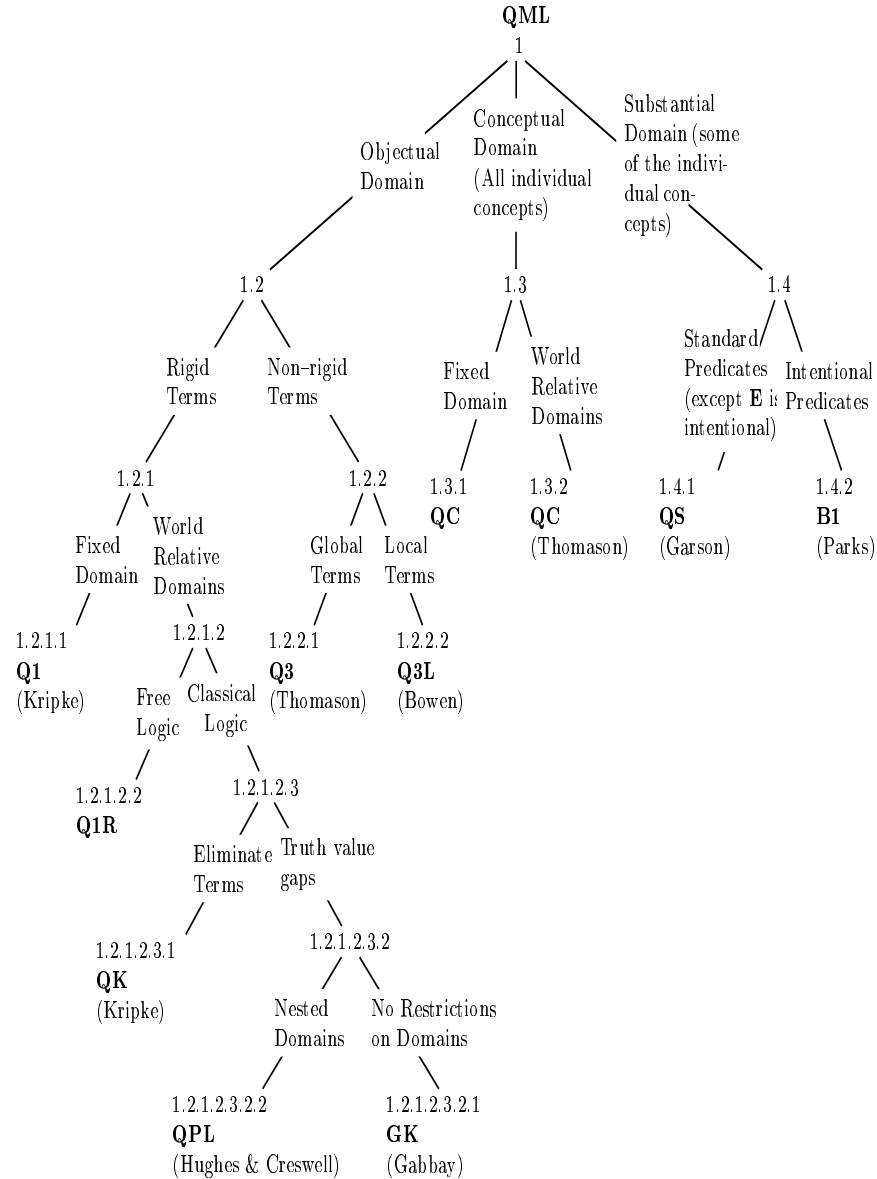


Figure 1. ROADMAP

## Explanation of the quantified Modal Logic Roadmap

This tree represents the structure of the discussion of quantified modal logic in this chapter. Each node contains a number indicating the section of this chapter where a topic is discussed. Branches from each node are labelled with the main options which one can choose at that point. The 'leaves' of the tree are labelled with the name used in this chapter of the system which results from choosing the options on all branches leading to it. Beneath the name of each system is the name of an author associated with the system. The references in the bibliography associated with his name contain a description of the system in question.

There are a number of ways for a believer in the principles of first-order logic to handle this problem. One popular tactic is to count ‘God’ as a definite description  $IxGx$ , where  $Gx$  is interpreted to be true only of God. Then ‘God exists’ translates to  $\exists yy = IxGx$ . By Russell’s theory of descriptions, this amounts to  $\exists z(\exists yy = z \wedge Gz \wedge \forall x(Gx \rightarrow x = z))$ , which is not a theorem. However, this reply depends on a debatable assumption, namely that for every name which may fail to refer, we can find a predicate (or open sentence) which picks out that referent uniquely. Kripke [1972] presents strong evidence that we cannot find such uniquely identifying predicates. Even if we could solve this problem, the use of Russell’s theory causes another problem. We want to be able to say that ‘Pegasus has wings’ is true, but that ‘Pegasus is a hippopotamus’ is false. If we translated ‘Pegasus’ away in these two sentences according to Russell’s theory of descriptions, we obtain sentences of the shapes  $W(IxPx)$  and  $H(IxPx)$ , which are both false since Pegasus does not exist. We do no better translating these sentences by  $\forall x(Px \rightarrow Wx)$  and  $\forall x(Px \rightarrow Hx)$ , because in this case both are vacuously true, since nothing satisfies the predicate  $P$ .

Free logic avoids these difficulties by dropping the assumption that every name must refer to an object in the domain of quantification. As a result, the principles for the quantifiers are somewhat weaker. Let us assume that we have a primitive predicate  $E$ , whose extension is the domain of quantification. The revised axiom of Existential Generalisation becomes:

$$(FEG) \quad (Pt \wedge Et) \rightarrow \exists xPx.$$

The proof we gave for  $\exists xx = g$  in first-order logic is now blocked. Using (FEG), we may obtain  $\exists xx = g$  from  $g = g$  only if we have already proven  $Eg$ , and  $Eg$  expresses what we are trying to prove.

A complete system **MFL** of minimal free logic with identity can be constructed by defining  $\exists x$  and  $\neg\forall x\neg$  and adding the following rules to propositional logic plus identity theory:

$$(FUI) \quad \frac{\forall xPx}{Et \rightarrow Pt} \quad \text{for any term } t$$

$$(FUG) \quad \frac{\vdash A \rightarrow (Et \rightarrow Pt)}{\vdash A \rightarrow \forall xPx} \quad t \text{ is a term that does not appear in } A \rightarrow \forall xPx.$$

In these rules, and throughout this chapter,  $A$  and  $Px$  are wffs,  $x$  is any variable, and  $Pt$  is the result of substituting the term  $t$  properly for all occurrences of  $x$  in  $Px$ . It is an easy exercise to show that  $Et$  is equivalent in **MFL** to  $\exists xx = t$  (where  $x$  is not  $t$ ). So we could have defined  $Et$  as  $\exists xx = t$ , and avoided the introduction of a special predicate letter  $E$ . However, in some intensional logics, there is no way to define  $Et$  in terms of the rest of the primitive vocabulary, and so we have prepared for this by assuming that  $E$  is primitive.

### 0.3 Identity in Intensional Logics

The failure of the substitution of identical terms is a familiar criterion for identifying intensional expressions. For example, the invalidity of the famous argument:

Scott is the author of Waverley	
King George wonders whether Scott is Scott	
King George wonders whether Scott is the author of Waverley	

serves as evidence that ‘King George wonders whether’ is intensional. It should not surprise us, then, if we need to limit the rule of substitution of identities in intensional logics. One simple way to enforce the desired restriction is to allow substitution in atomic sentences only, as in the following system **ID** for identity:

$$\begin{array}{l}
 (= \text{In}) \quad t = t \qquad (= \text{Out}) \quad \frac{t = t'}{Pt \rightarrow Pt'} \quad \text{where } Pt \text{ is an atom.}
 \end{array}$$

Although the restriction to atomic sentences may seem strong, it has no effect whatsoever in first-order logic, because (= Out) insures the substitution of identities in all extensional sentences. However, in intensional logics, it does not guarantee substitution of identical terms which lie in the scope of intensional operators.

Some may object to the view that the substitution of identicals fails. Russell, for example, gave an explanation of the invalidity of the argument about the author of Waverley which did not require any restrictions on the rule of substitution. Russell claimed that the description ‘the author of Waverley’, does not count as a term. When the description is eliminated according to his theory, the first premise of the argument no longer has the form of an identity. This tactic does not work, however, for arguments such as the following where there are no descriptions to eliminate:

Cicero is Tully.	
King George knows that Cicero is Cicero.	
King George knows that Cicero is Tully.	

One reaction to this sort of example is to argue that the failure of the rule of substitution is a sign that the expression being substituted is not really a term. So the invalidity of the last argument shows that ‘Cicero’ and ‘Tully’ are not terms, and must be translated using corresponding descriptions:  $IxCx$  and  $IxTx$ . When this is done, the first premise of the argument no longer has the form of an identity, and so does not count as a case of substitution.

Notice, however, that adherence to the principle of unrestricted substitution leads us to a position similar to the one which resulted from adherence

to the classical rules for quantifiers, we conclude that many of the expressions which we would ordinarily count as terms, must be treated instead as descriptions. We were forced before to deny the termhood of expressions which might fail to denote, and now we are compelled to deny it of expressions which might have synonyms. Since we have little guarantee that a given expression avoids either defect, we feel pressure, as Quine did, to claim that no expression of English should be rendered as a constant in first-order logic.

Given the simplicity of the alternative rules, the insistence on the classical rules for quantifiers and the unrestricted substitution of identities is, in our opinion, a prejudice, and one which blocks a natural exposition of an adequate foundation for quantified modal logics.

## 1 A TAXONOMY OF QUANTIFIED INTENSIONAL LOGIC

One of the most significant points of difference between semantical treatments of QML concerns the domain of quantification. Some systems quantify over objects, while others quantify over what Carnap [1947] called individual concepts. The second approach is more general, but it is also more abstract, and more difficult to motivate. So we will open this account of QML with systems that use the objectual interpretation.

### 1.1 *Some Semantical Preliminaries*

Before we begin, it will be helpful to define a few semantical ideas which we will use throughout this chapter. We assume that a quantified modal language is constructed from predicate letters, the primitive predicate constant  $E$ , terms (which include infinitely many variables) the logical constants  $\neg$ ,  $\rightarrow$ ,  $\Box$ ,  $=$ , and a quantifier  $\forall x$  for each of the variables  $x$ . The predicate letters come equipped with integers indicating their arity. The propositional variables are taken to be 0-ary predicate letters, and well-formed formulas are defined in the usual way. Given a set  $D$ , the extensions of terms and predicate letters are defined just as they are in first-order logic. The *extension of a term* is some member of  $D$ , and the *extension of an  $i$ -ary predicate letter* is a set of  $i$ -length sequences of members of  $D$ . Given a set  $W$  of indices (typically, possible worlds), the *intension of an expression* is simply a function which takes each member of  $W$  into an appropriate extension for that expression. Carnap's *individual concepts* are simply term intensions, that is, functions from the set of possible worlds into the domain of objects.

Throughout this chapter, a *Q-model*  $\langle W, R, D, Q, a \rangle$  will contain a set  $W$  of possible worlds, a binary relation  $R$  on  $W$ , a nonempty set  $D$  of possible objects, some item  $Q$  which determines the domain of quantification, and an assignment function  $a$ , which interprets the terms (including variables)



and predicate letters by assigning them the corresponding kind of intensions with respect to  $W$  and  $D$ . If the quantifier rules of a system are based on free logic, then there will be a predicate letter  $E$  in the language. To ensure that  $E$  receives the proper interpretation as picking out the quantifier domain, we will assume that a  $Q$ -model for a language that contains  $E$  always meets the condition that  $a(E)$  is  $Q$ .

In some semantics, the terms are *rigid designators*, that is, their extensions are the same in all possible worlds. Usually such terms are assigned no intensions, but given extensions directly. However, in order to keep the description of a model as consistent as possible, we will assume that terms always have intensions, and that terms which are rigid designators simply meet the added condition that their intensions are constant functions.

The symbol  $=$  will always be interpreted as contingent identity. This means that  $t = t'$  is ruled true in a world just in case  $t$  and  $t'$  have the same extension in that world. The truth value of a sentence  $A$  on a model  $\langle W, R, D, Q, a \rangle$  at world  $w$  of  $W$  (written  $a(A)(w)$ ) will be defined by induction on the shape of  $A$  using the standard clauses for atomic sentences,  $\neg$ ,  $\rightarrow$  and  $\Box$ . When we present a given approach to the quantifiers, we usually will need only to say what  $Q$  is like, and to give the truth clause for the quantifier.

The quantified modal logics we are going to discuss are all extensions of propositional modal logics which are adequate with respect to some class of Kripke frames. For example, we will consider extensions of **S4**, which are adequate (semantically consistent and complete) with respect to the class **R(S4)** of Kripke frames  $\langle W, R \rangle$  that are reflexive and transitive. Usually we will not care which propositional modal logic is chosen as the foundation for our quantified logic. We will assume that some propositional modal logic has already been chosen, and that the frame of any  $Q$ -model is in  $R(S)$ . When we need to be explicit, we will talk of  $S$ -models, and mean models whose Kripke frames are in the set  $R(S)$ . The notions of  $Q$ -satisfiability and  $Q$ -validity are determined by the concept of a  $Q$ -model exactly as in propositional modal logic.

## 1.2 The Objectual Interpretation

1.2.1 *Rigid Terms.* Kripke's historic paper [1963] serves as an excellent starting point for a discussion of logics with the objectual interpretation. One reason is that he made the important simplifying assumption that all terms of the language are rigid designators. Systems that allow nonrigid terms are, as we shall see, rather complicated, and so we will begin, as Kripke did, by assuming that the intension of every term is a constant function. This assumption validates the following two rules which we refer

to together as (RT) (for rigid terms).

$$(RT) \quad \frac{t = t'}{\Box t = t'} \quad \frac{\neg t = t'}{\Box \neg t = t'}$$

The rigidity condition reflects the view that proper names have extensions, but no intensions. Since (RT) guarantees the substitution of identity in all contexts, it sits well with those who object to restrictions on substitution of identities.

Kripke’s paper also lays out two important options concerning the quantifier domains. The simplest of the two, the fixed domain approach, assumes a single domain of quantification which contains, presumably, all the possible objects. The world-relative interpretation, on the other hand, assumes that the domain of quantification contains only the objects that exist in a given world, and so the domain varies from one world to another.

1.2.1.1 *Fixed Domains: The System Q1.* Although the fixed domain approach is less general, it is attractive from the semantical point of view because we need only add the familiar machinery for  $\forall x$  to the semantics of a modal logic in the following way. A *fixed domain objectual model with rigid terms* (or **Q1-model**) is a sequence  $\langle W, R, D, \mathbf{Q1}, a \rangle$ , where the domain of quantification **Q1** is  $D$ , the set of possible objects, and where  $a$  meets the condition (aRT), which guarantees that the term intensions are constant functions.

$$(aRT) \quad a(t)(w) \text{ is } a(t)(w') \text{ for all } w, w' \text{ in } W.$$

The truth value of a sentence on a model is then defined using the following clause for the quantifier:

$$(\mathbf{Q1}) \quad a(\forall xA)(w) \text{ is } T \text{ iff for all } d \text{ in } \mathbf{Q1}, a(d/x)(A)(w) \text{ is } T.$$

(Here  $a(d/x)$  is the assignment like  $a$  save that  $a(x) = d$ .)

For each propositional modal logic  $S$ , let the formal system **Q1-S** consist of the principles of  $S$ , rules for first-order logic (ID), (RT), and the Barcan formula (BF):

$$(BF) \quad \forall x \Box A \rightarrow \Box \forall x A.$$

One satisfying feature of the fixed domain account is that most propositional modal logics  $S$  for which we can show completeness with respect to a set  $R(S)$  of Kripke frames, have the feature that the system **Q1-S** is semantically consistent and complete with respect to **Q1-S**-validity. There are exceptions, however. For example, Cresswell [1995] explains that when  $R(S)$  is convergent, completeness of **Q1-S** may fail.

### 1.2.1.2 *World-Relative Domains.*

1.2.1.2.1 *The Motivation for World-relative Domains.* The fixed domain interpretation is satisfying from the formal point of view, but it is not an accurate account of the semantics of quantifier expressions of natural language. We do not think that ‘There is a man who signed the Declaration of Independence’ is true, at least not if we read ‘there is’ in the present tense. Nevertheless, this sentence was true in 1777, which shows that the domains of the present tense quantifiers changes to reflect which objects exist at different times. The domain varies along other dimensions as well. For example, when I announce to my class that everyone did well on the midterm, it is understood that I am not praising the whole human race. Time, place, speaker, and even topic of discussion play a role in determining the domain in ordinary communication. There are also strong reasons for rejecting fixed domains in modal languages. On the fixed domain interpretation, the sentence  $\forall x \Box \exists y (y = x)$  (which reads ‘everything exists necessarily’) is valid, but we would not ordinarily count this as a logical truth because we assume that different things exist in the different possible worlds.

The defender of the fixed domain interpretation can respond to these objections by insisting that the domain of  $\forall x$  contains merely possible objects. Expressions whose domain depends on the context, can then be defined using  $\forall x$  and predicate letters. For example, the present tense quantifier can be defined using  $\forall x$  and a predicate letter that reads ‘presently exists’.

One difficulty with this proposal is that it requires the invention of predicates for all the different subdomains which we may ever intend for quantifier expressions, and it forces us to represent simple expressions of natural language differently in different contexts of their use. It would be more satisfying if we could specify semantics for intensional logic which admits the context dependence of the domain.

1.2.1.2.2 *World-Relative Models: Q1R-Semantics.* Let us define a *world-relative objectual model with rigid terms* (or **Q1R-model**) as a sequence  $\langle W, R, D, \mathbf{Q1R}, a \rangle$ , where **Q1R** is a function that assigns a subset  $D(w)$  to  $D$  to each possible world  $w$ , and where  $a$  meets condition (*aRT*). The truth clause for the quantifier reads as follows:

$$(\mathbf{Q1R}) \quad a(\forall x A)(w) \text{ is } T \text{ iff for every } d \text{ in } D(w), a(d/x)(A)(w) \text{ is } T.$$

An adequate logic **Q1R** for **Q1R**-validity can generally be formulated by adding the principles **MFL** of free logic, rules **ID** for (intensional) identity, and (**RT**) to the underlying modal logic.

1.2.1.2.3 *Methods for Preserving Classical Quantifier Rules.* The world-relative interpretation of the quantifiers virtually demands the adoption of free logic. I say ‘virtually’ because there are systems which use first-order rules with the world-relative interpretation; however, they have serious

limitations. To appreciate the difficulties in trying to maintain the standard rules, notice first that the sentence  $\exists x(x = t)$  is true at a world on a model just in case the extension of  $t$  is in the domain of that world. However,  $\exists x(x = t)$  is a theorem of first-order logic, and so it follows that every term  $t$  of the language must refer to an object that exists in every possible world. This leads to two difficulties. First, there may not be any one object that exists in all the worlds. Second, the whole motivation for the world relative approach was to reflect the idea that objects in one world may not exist in another; but if standard rules are used, no terms may refer to such objects.

1.2.1.2.3.1 *Eliminate terms: the system QK.* Kripke [1963] gives an example of a system for the world-relative interpretation which keeps the classical rules. The system **QK** has no terms other than variables. On a semantics where variables are given extensions in the domain, the validity of  $\exists xx = y$  would demand that the extension of  $y$  be a member of every possible world. Kripke avoids this difficulty by giving sentences with free variables the closure interpretation. So  $\exists xx = y$  has the semantical effect of  $\forall y\exists xx = y$ , which is valid in free logic. From the semantical point of view, then, Kripke's system, has no terms at all, because the variables are really disguised universal quantifiers. Although Kripke has shown that modal extensions of first-order logic with the world-relative interpretation are possible, his system underscores a theme which we have been developing throughout this chapter, namely that adoption of the classical rules forces us into an inadequate account of terms. Another oddity of Kripke's system is that he must weaken the necessitation rule: 'if  $A$  is a theorem, then so is  $\Box A$ '. Otherwise we would be able to derive  $\Box\exists xx = y$  which, since it is given the closure interpretation, says that any object of one domain exists in all the others. The rule is repaired by restricting it to closed sentences.

1.2.1.2.3.2 *Nested domains and truth value gaps.* There is a second problem with using classical logic with the world- relative interpretation which has exerted pressure on the way semantics for quantified modal logics is formulated. The principles of classical logic, along with the (unrestricted) rule of necessitation entail (CBF), the converse of the Barcan Formula.

$$(CBF) \quad \Box\forall xA \rightarrow \forall x\Box A.$$

It is not difficult to show that every world-relative model of (CBF) must meet condition (ND) (for 'nested domains').

$$(ND) \quad \text{If } wRw' \text{ then } D(w) \text{ is a subset of } D(w').$$

To see this, notice that  $\Box\forall x\exists yy = x$  is **Q1R**-relative valid, and entails  $\forall x\Box\exists yy = x$  by (CBF). Our desire to avoid  $\forall x\Box\exists yy = x$  was one of the things which prompted the world-relative interpretation, for  $\forall x\Box\exists yy = x$  claims that any object which exists in the real world must also exist in all

worlds which are possible relative to ours. Certainly, we want to allow that there are possible worlds where at least one of the things of our world fails to exist.

If  $R$  is symmetric, then it follows from (ND) that all worlds accessible from ours have exactly the same domains. This result is reflected in the fact that the Barcan Formula (BF) is provable in systems as strong as **B** which use the standard quantifier rules. In models of **S5** where all worlds are accessible from each other, (ND) demands that all domains be the same, in direct conflict with our intention to distinguish the domains.

Despite these difficulties in using classical principles with an unrestricted necessitation rule, several authors have defined systems which preserve the classical rules. Typically, their systems simply adopt (ND). Yet other adjustments must be made, however, to preserve classical logic. The sentence  $\forall xPx \rightarrow Pt$ , for example, is not valid on a model where the extension of  $t$  at a world  $w$  is outside  $D(w)$ , and the extension of  $P$  at  $w$  is  $D(w)$ . One simple way to restore validity to the rule of Universal Instantiation is to stipulate that the terms are local, that is, the extension of a term at a world must be in the domain  $D(w)$  of that world. However, there are serious problems with this. According to this view, ‘Pegasus’ and possibly ‘God’ cannot count as terms since their extensions are not in the real world. As we have argued in Section 0.2, there are good reasons for wanting to count these as terms. Furthermore, we have been assuming that terms are rigid, so terms must have the same referent in all worlds. So the demand that terms be local entails that any term must have an extension which exists in all the worlds. In fact, the only objects at which the domains might vary are ones which are never named in any world. This undercuts the whole point of introducing world-relative domains, namely to accommodate terms that refer to things that may not exist in other possible worlds.

The consequences of having terms that are both local and rigid are disastrous. There is another related idea, however, that looks as though it might work. If we assume that *predicate letters* are local, i.e. that their extensions at a world must contain only objects that exist at that world, then we will ensure that the classical sentence  $Ft \rightarrow \exists xFx$  (hence  $\forall xPx \rightarrow Pt$ ) is valid. The reason is that from the truth of  $Ft$ , it follows that  $t$  refers to an existing object, and from this it follows that  $\exists xFx$  is true. Nevertheless, local predicates set up other anomalies, and they do not lead to the validation of the classical rules. To see why, consider  $\neg Ft \rightarrow \exists x\neg Ft$ . From the truth of  $\neg Ft$ , it does not follow that the extension of  $t$  is an existing object, and so it does not follow that  $\exists x\neg Ft$  is true. Not only do we fail to validate the rule of Existential Generalisation, but the valid principles cannot be expressed as axiom schemata. (We cannot write  $Pt \rightarrow \exists xPx$  for arbitrary sentences  $Pt$ , because some of these instances are valid, and others are not.) In case we are using axioms and a rule of substitution of formulas for atoms, the problem re-emerges in the failure of the rule of substitution. Either way,

the use of local predicates leads to serious formal difficulties.

There is a somewhat more plausible way to ensure the classical principles. A Strawsonian treatment would rule that a sentence has no truth value when it contains a term that does not refer to an existing object. Following this idea, we allow terms to refer to objects outside of the domain of a given world, but rule that sentences which contain such terms lack truth values. Valid sentences are then defined as ones which are never false. As a result,  $\forall xPx \rightarrow Pt$  is valid, since any assignment that gives  $t$  an extension outside the domain for a world leaves the whole conditional without a value, and assignments that give  $t$  an extension inside the domain will make  $Pt$  true if  $\forall xPx$  is true.

1.2.1.2.3.2.1 *The systems GKc and GKs.* When truth value gaps are introduced, we are faced with a number of options concerning the truth clause for  $\Box$ . On at least one of these options we may drop the nesting condition (ND) if we like and still obtain the classical rules. However, there are pressures that make us want to keep it. Suppose we are evaluating  $\Box Ft$  at  $w$  and the referent of  $t$  is in the domain  $D(w)$  of  $w$ . Then we expect to give  $\Box Ft$  a truth value on the basis of the values  $Ft$  has in the worlds accessible from  $w$ . Unless we adopt (ND), there is no guarantee that  $t$  refers to an existing object in all accessible worlds, and so  $Ft$  may be undefined in some of them. Adopting the nesting condition ensures that we will always determine a value for  $\Box Ft$  at  $w$  on the basis of the values which  $Ft$  is bound to have in all accessible worlds. If we drop (ND), however, there are two ways to determine the value of  $\Box Ft$  at  $w$  depending on whether the failure of  $Ft$  to be defined in an accessible world should make  $\Box Ft$  false or not. On the first option, Gabbay's **GKc** [Gabbay, 1976, pp. 75 ff.], the necessitation rule must be restricted so that we can no longer derive (CBF). On the second option, **GKs**, (CBF) is derivable, but the truth of (CBF) in a model no longer entails (ND). Either way, the rules of the underlying modal logic must be changed.

1.2.1.2.3.2.2 *The system QPL.* For these reasons, the more popular choice [Hughes and Cresswell, 1968] has been to assume (ND) and to define satisfiability as follows. A **QPL-satisfiable set** is one where none of its sentences is false in any world on some **Q1R**-model that meets (ND), and where any sentence which contains a term  $t$  with extension  $a(t)(w) \notin D(w)$  has no truth value at  $w$ .

**QPL**-semantics is attractive from a purely formal point of view because we have relatively simple completeness proofs for systems that result from adding the principles of (classical) predicate logic to certain propositional modal logics, provided, that is, that the language omits  $=$ . Proofs are available, for example, for **M** and **S4**. In case the modality is as strong as **B**, the domains become rigid, and the completeness proof is carried out using methods developed for systems that validate the Barcan Formula.

1.2.1.2.4 *Conclusion: We Should Adopt Free Logic.* The appeal of simple completeness proofs should not blind us to the fact that the stipulations required in order to preserve the classical principles do not always sit well with our intuitions. Our conclusion, then, is that there is little reason to attempt to preserve the classical rules in formulating systems with the objectual interpretation and world-relative domains. The principles of free logic are much better suited to the task. As we will see in Section 2, results for systems based on free logic are actually not that difficult, especially when identity is not present.

### 1.2.2 *Non-rigid terms and world-relative domains*

1.2.2.1 *The System Q3.* There are two important reasons why the assumption that all terms are rigid designators should be rejected. First, expressions like ‘the tallest man’ clearly refer to different objects in different worlds. If we want to count descriptions among our terms, as we do on a Strawsonian account, we cannot accept the rigidity condition. Second, David Lewis [1968] contends that it makes no sense to talk of identity of objects across possible worlds. Objects from two different worlds are never identical, although it may make sense to talk of the counterpart of an object in another world. On counterpart theory, then, it is impossible for the intension of any term to be a constant function. Since it is important that a logical theory not rule out reasonable positions, we would like to relax the restriction that terms are rigid. Let us define a **Q3-model**, then, as a **Q1R-model** which (possibly) fails to meet condition (*aRT*).

Something unexpected happens when we relax the assumption that terms are rigid. The rule (FUI) of instantiation for free logic is no longer **Q3**-valid. In order to see why, notice that the sentence  $(\Box t = t \wedge Et) \rightarrow \exists x \Box x = t$  is a consequence of (FUI). Since  $\Box t = t$  is also provable there, we obtain (*E□*).

(*E□*)  $Et \rightarrow \exists x \Box x = t.$

If  $t$  reads ‘the author of “Counterpart Theory”’, then (*E□*) says that if the author of ‘Counterpart Theory’ exists, then there is someone who is necessarily the author of ‘Counterpart Theory’. Intuitively, (*E□*) is unacceptable, and it is not difficult to back up this insight with a formal counter-example. Let us imagine a model with two worlds,  $r$  (real) and  $u$  (unreal) whose domains both contain two objects, namely David Lewis and Saul Kripke. Assume that both worlds are accessible from themselves and each other. Imagine that the extension of  $t$  at the real world  $r$  is Lewis, but that it is Kripke in the unreal world  $u$ . On this model,  $\exists x \Box x = t$  is false in  $r$  because neither Lewis nor Kripke is the extension of  $t$  in both worlds. Nevertheless,  $Et$  is true in  $r$  since the extension of  $t$  in the real world, namely David Lewis, is in the domain of  $r$ .

This counterexample helps us appreciate the subtle reason why (FUI) has broken down. There is no question that David Lewis exists, and there is no

question that the author of ‘Counterpart Theory’ is identical to the author of ‘Counterpart Theory’ in any world we choose. However, the claim that any *one* person counts as the author of ‘Counterpart Theory’ in all worlds seems false. One way to help diagnose this situation is to reformulate **Q3** semantics in an equivalent, but more complex way. Replace each object with the constant function which takes any world to that object. Seen this way, the items in our domain(s) are all intensions of rigid terms. The rule of instantiation is no longer valid because the domain of quantification includes only constant term intensions, whereas terms may have nonconstant intensions.

The rules of free logic would be **Q3**-valid if we were to interpret the primitive predicate  $E$  so that  $Et$  is true in world  $w$  iff the extension  $a(t)(w)$  of  $t \in D(w)$  and  $a(t)$  is a constant function. Notice, however, that the extension of  $E$  must contain term intensions, and not objects, if it is to do this job. As a result,  $E$  is an *intensional predicate*, which means that substitution of identity does not hold for its term position. Substitution fails because  $E$  ‘David Lewis’ is presumably true, while  $E$  ‘the author of “Counterpart Theory” ’ is not, even though ‘David Lewis’ and ‘the author of “Counterpart Theory” ’ refer to the same thing in the real world.

Aldo Bressan [1973] has championed the view that even scientific language requires intensional predicates. His more general semantics defines the *extension* of a one-place predicate at a possible world as a set of individual concepts (i.e. term intensions) not a set of objects. As a result, he has no difficulty accommodating a primitive predicate which expresses rigidity.

Hintikka [1970] chose more modest methods. He showed how to formulate a correct rule of instantiation for **Q3** that does not require an intensional existence predicate. Notice that the sentence  $\exists x \Box x = t$  is true in a model at world  $w$  iff the intension of  $t$  has the same value in all worlds accessible from  $w$ . Similarly,  $\exists x \Box \Box x = t$  is true at  $w$  just in case the intension of  $t$  is constant in all worlds accessible from those worlds. While there is no one sentence that expresses that a term is rigid, a sentence of the shape  $\exists x \Box \Box \dots \Box x = t$ , where  $\Box$  is a string of  $i$  boxes, guarantees that the intension of  $t$  is constant across enough worlds so that  $\Box \Box \dots \Box Ft$  follows from  $\forall x \Box \Box \dots \Box Fx$  when  $Ft$  is atomic. This idea is generalised in Hintikka’s formulation (HUI) of a valid rule of universal instantiation for nonrigid terms.

$$(HUI) \frac{\forall x Px}{(\exists x \Box \Box \dots \Box x = t \wedge \dots \wedge \exists x \Box \Box \dots \Box x = t) \rightarrow Pt}$$

where  $i, \dots, k$  is a list of integers which records for each occurrence of  $x$  in  $Px$ , the number of boxes whose scope includes that occurrence.

In modal logic as strong as **S4**, this rule can be simplified considerably because there  $\exists x \Box \Box \dots \Box x = t$  is equivalent to  $\exists x \Box x = t$ . Thomason [1970] demonstrates the adequacy of **Q3–S4**, using (TUI) as the instantiation rule.



$$(TUI) \frac{\forall xPx}{\exists x \Box x = t \rightarrow Pt}$$

Completeness proofs for the weaker modalities have never been published as far as I know. Perhaps researchers have been daunted by the complexity of Hintikka's rule. It is interesting to note that even in the context of **S4**, Thomason was forced to adopt other complex rules for identity and the quantifier. Parsons [1975] has given a weak completeness result for a system that uses more standard rules, but he also shows that, in general, Thomason's rules cannot be simplified in the obvious way.

1.2.2.2 *A Classical Logic with Local Terms: The System Q3L*. There is a simple way to avoid the complicated instantiation rule needed in **Q3**. If we add the assumption that terms are local, that is, that the extension of a term at a world  $w$  is always in that world's domain, then we restore the classical quantifier rules. A **Q3** model with local terms (**Q3L-model**) is a **Q3**-model which meets condition (L)

$$(L) \quad a(t)(w) \in D(w) \quad \text{for all } w \text{ in } W, \text{ and all terms } t.$$

This condition could not be sensibly imposed for systems with rigid terms because then, any object referred to by a term would have to exist in all the domains. However, when terms are nonrigid, the domains can change as long as the extension of the terms change in corresponding ways.

There is an important application of **Q3L** which Cocchiarella discusses in his chapter in Volume 3.4. If  $\Box$  is to capture logical necessity, then we may think of possible worlds  $w$  as predicate logic models  $\langle Dw, aw \rangle$ , each equipped with its own domain  $Dw$ , and assignment function  $aw$ . We expect an assignment function  $aw$  of a model  $\langle Dw, aw \rangle$  to give extensions to the terms (and predicate letters) in the corresponding domain  $Dw$ . So it is only natural in this case to adopt nonrigid terms, world-relative domains, the objectual interpretation, and local terms.

If we interpret  $\Box A$  to mean that  $A$  is true in all models, then **Q3L**-semantics cannot be axiomatised. However, if we give  $\Box A$  the generalised interpretation where  $\Box A$  is true iff it is true on all models in an arbitrarily selected set of models, then **Q3L** is axiomatised by adding the principles of predicate logic to **S5**.

A more general account stipulates that  $\Box A$  is true on a model  $U$  just in case  $A$  is true in all models  $U'$  suitably related to  $U$ . In this case the underlying modality depends on the conditions we adopt on the accessibility relation between models. If we take this option, however, and the accessibility relation is not symmetric, then we are forced to assume nested domains (ND), in order to preserve the classical quantifier rules. Bowen [1979] investigates systems of this kind. Even if we are willing to give up the nesting condition, problems arise. Suppose we are evaluating  $\forall x \Box Fx$  in a world  $w$  where object  $o$  exists, and  $w'$  is an accessible world where  $o$

does not exist. To determine the value of  $\forall x \Box Fx$ , we need to find the value of  $\Box Fx$  when  $x$  refers to  $o$ . This requires that we find the value of  $Fx$  in world  $w'$  where  $o$  does not exist. At this point we are faced with the same options we described in Section 1.2.1.2.3.2. We may use truth value gaps, or we may rule that  $Fx$  in this case is false. As we pointed out, both choices have disadvantages.

Despite its application to certain notions of logical necessity, the local term condition (L) is not usually acceptable. In ordinary reasoning, we would find the assumption that anything that exists in the real world exists in all worlds possible relative to our is quite implausible. For this reason, we are still interested in **Q3** without local terms, even though the rules may be difficult.

### 1.3 The Conceptual Interpretation

The systems we have discussed so far are not especially satisfying. We have good reasons for wanting to allow nonrigid terms in our language, and yet the rules we need for **Q3** are quite complex, unless we move to a language with a primitive intensional predicate that expresses rigidity. On the other hand, systems with local variables, like **Q3L**, have limited applications. One account of our difficulties, as we explained earlier, is that our terms can be assigned any intension, while the domain(s) of quantification contain only constant intensions. Perhaps allowing nonrigid intensions in our *domain* might result in a better match between the quantifiers and the terms, and so yield simpler rules.

Though it may seem philosophically dangerous to quantify over individual concepts, there are intuitions concerning tense and modality that support this choice. For example, imagine that our possible worlds are now states of the universe at a given time. The extension of a term at a given time will turn out to be a temporal slice of some thing, 'frozen' as it is at that instant. Notice that things, since they change, cannot be identified with term extensions. Instead, things are world-lines, or functions from times into time slices, and so they correspond to term intensions or individual concepts. Since our ontology takes things, not their slices as ontologically basic, it is only natural to quantify over term intensions in temporal logic. Our reluctance to quantify over individual concepts may be an accident of nomenclature. The so called 'objects' of a temporal semantics are not the familiar things of our world, while the formal entities that do correspond to things are misleadingly called 'individual concepts'.

#### 1.3.1. Fixed Domains: The System **QC**.

Let us now formulate what we will call the conceptual interpretation of the quantifier. A *conceptual model* (or **QC-model**) is a sequence  $\langle W, R, D, \mathbf{QC}, a \rangle$

where  $\mathbf{QC}$  is the set of functions from  $W$  into  $D$ . The truth clause for the quantifier reads as follows:

( $\mathbf{QC}$ )  $a(\forall xA)(w)$  is  $T$  iff for every  $f$  in  $\mathbf{QC}$ ,  $a(f/x)(A)(w)$  is  $T$ .

Here ‘ $a(f/x)$ ’ represents the assignment function identical to  $a$  except that the intension of  $x$  on  $a(f/x)$  is function  $f$ .

Although the conceptual interpretation is designed to satisfy reasonable intuitions, there are a number of problems with it. One formal difficulty is that no (consistent) system is complete for this semantics. Whenever we interpret the domain of any quantifier as a set of all functions, we run the risk that the language will have the expressive power of second-order arithmetic, with the result that Gödel’s Theorem applies. As we will show in Section 3, that is exactly what happens with  $\mathbf{QC}$ .

There are also intuitive difficulties. First, notice that  $\exists x \Box x = t$  is  $\mathbf{QC}$ -valid, and yet we have given an intuitive counterexample to it in Section 1.2.2.1. We do now want to say that there is something which is necessarily the author of ‘Counterpart Theory’, because no one thing is the author of that paper in all possible worlds. However, on the conceptual interpretation,  $\exists x \Box x = t$  is true as long as we can find some term intension which matches that of  $t$  in all possible worlds, and the term intension of  $t$  so qualifies. This shows that the conceptual interpretation differs from our ordinary reading of the quantifier. Another  $\mathbf{QC}$ -valid sentence which may tantalise some readers is  $\exists x \Box \exists yy = x$ , which claims that there is something (God?) which necessarily exists. However the  $\mathbf{QC}$ -validity of this sentence will do little to satisfy those who still search for an ontological argument for the existence of God. Any term intension will do to satisfy  $\Box \exists yy = x$ , simply because any term intension has the property that there is a term intension (namely itself) which agrees with it in accessible worlds.

### 1.3.2. *World-relative Domains: The System Q2.*

The reader may think that we can repair these problems by introducing world-relative domains. Let us investigate the situation, then, when a  $\mathbf{Q2}$ -model is a sequence  $\langle W, R, D, \mathbf{Q2}, a \rangle$ , where  $\mathbf{Q2}$  is a function that assigns a domain  $D(w)$  to each world  $w$ . The quantifier truth clause now reads as follows.

( $\mathbf{Q2}$ )  $a(\forall xA)(w)$  is  $T$  iff for every function  $f : W \rightarrow D$ ,  
if  $f(w) \in D(w)$ ,  $a(f/x)(A)(w)$  is  $T$ .

Unfortunately, the problems we mentioned still remain. First, the incompleteness result still applies to the new semantics. Second, although both  $\exists x \Box x = t$  and  $\exists x \Box \exists yy = x$  are no longer valid, they still do not receive their intuitive interpretations. For example,  $\exists x \Box \exists yy = x$  will turn out to be true on every model where the domains of the worlds all contain at least

one object. In that case, any function that picks a member of  $D(w)$  for each world  $w$  will satisfy  $\Box\exists yy = x$ , and so verify  $\exists x\Box\exists yy = x$ .

#### 1.4 *The Substantial Interpretation*

As we showed in the last section, the conceptual interpretation of the quantifiers does not match the interpretation which we give to quantifier expressions in ordinary language. The sentence  $\exists x\Box\exists yy = x$ , which we interpret as making the very strong claim that some thing must exist in every possible world, is valid on the conceptual interpretation as long as no possible world has an empty domain. The difference between our intuitive understanding of  $\exists x\Box\exists yy = x$ , and the conceptual interpretation is that the existence of a term intension that (say) picks out David Lewis in this world, a rock in another, a blade of grass in another, and so on, counts to verify  $\exists x\Box\exists yy = x$ . On the other hand, our intuitions demand that any term intension that verifies  $\exists x\Box\exists yy = x$  must be coherent in some sense; our concept of a thing brings with it some notion of what it would be like in other worlds. Only certain collections of objects, (and certainly not a collection consisting of David Lewis, a rock, a blade of grass, etc.) could count as the manifestations of a thing, and so only these collections should count to verify  $\exists x\Box\exists yy = x$ .

In order to do justice to these intuitions, we must restrict the domain of quantification to the term intensions that reflect ‘the way things are’ across possible worlds. Thomason [1969] suggests that the domain should contain only constant functions. The idea is that for  $\exists x\Box\exists yy = x$  to be true there must be one thing, identical across possible worlds, which exists in each one. This proposal is simply **Q3**, the objectual interpretation with non-rigid terms. We have already discussed some of the formal difficulties with this option in Section 1.2.2. There are also intuitive objections similar to the ones which we used in arguing against systems with rigid terms. First, Thomason’s account of substances is incompatible with counterpart theory, for on that view, the domains of the possible worlds are disjoint, and so there cannot be any constant term intensions to fill the domain of the quantifier. Second, in temporal logic, where objects are time slices, we do not want a thing to consist of the same time slice across different times. The slices of a thing picked out at different times may be quite different, but the world line composed of the slices still represents one unified thing.

##### 1.4.1. The System **QS**.

If we are to accommodate a variety of conceptions about what things are like, we should not assume that they are the constant term intensions (**Q3**), nor that they are all the term intensions (**Q2**). To be completely general, we introduce a set of term intensions for each world, to serve as its domain

of quantification, and we will make no stipulations about what these sets contain. Let us now give a formal account of this approach.

A *world-relative substantial model* (or **QS-model**) is a sequence  $\langle W, R, D, \mathbf{QS}, a \rangle$ , where **QS** is a function that assigns to each world  $w$  a set  $S(w)$  of functions from  $W$  into  $D$ . (We call  $S(w)$  the set of substances for world  $w$ .) The truth clause for the quantifier reads as follows:

$$(\mathbf{QS}) \quad a(\forall xA)(w) \text{ is } T \quad \text{iff} \quad \text{for every member } f \text{ of } S(w), \\ a(f/x)(A)(w) \text{ is } T.$$

It is not difficult to see that  $\exists x \Box \exists yy = x$  is not valid on this semantics, for it would only be true in world  $w$  if there were a substance  $f$  in  $S(w')$  in every world  $w'$  accessible from  $w$ .

Complete systems for **QS** can be constructed as long as we are willing to introduce the intensional predicate constant  $E$  to represent which functions count as substances in each possible world. An adequate system for this semantics very often results from adding the rules of **MFL**, and the rules **ID** for (intensional) identity to the underlying modal logic. As we will explain in Section 2.2.4, more general quantifier rules may be needed for weaker modal logics.

We should note an important restriction on the rule of substitution of identities in **QS**. The constant  $E$  is an intensional predicate, and this means that substitution of term identities does not hold in its term position. When we formulate the rule of substitution for identities, we must make it clear that we do not consider  $Et$  to be an atomic sentence, for otherwise we would be able to deduce  $Et'$  from  $t = t'$  and  $Et$ .

#### 1.4.2. *Fully Intensional Predicates: The System B1.*

During our discussion of **Q3**, we pointed out that one way to simplify the instantiation rule is to introduce an intensional predicate  $E$  to the language. A *predicate is intensional* when its extension at a world  $w$  contains term intensions, and not objects as we ordinarily expect. To be more careful, the *extension of an  $n$ -ary intensional predicate letter* at a world is a set of  $n$ -length sequences of term intensions.

Bressan [1973] presents a beautifully general modal logic, with descriptions and quantifiers for all types, which assumes that predicate letters are intensional in this sense. Clearly, such a strong language cannot be axiomatised. However, Parks [1976] has axiomatised the first-order fragment **B1** of Bressan's system, using the substantial interpretation of the quantifier. **B1** uses **S5** as its modal foundation, and a fixed domain of substances. For this reason **B1** validates classical quantifier rules and the Barcan Formula. However, more general languages with weaker modalities and world-relative domains of substances can be constructed using Bressan's more general treatment of predicates. In fact, we can add such predicate letters to **QS** without causing any major complications. All we need to do is adjust the

rule of substitution of identities for those predicate letters so that substitution of one term for another is not allowed unless we already have a sentence which informs us that their intensions (not just their extensions at a given world) are the same. In weaker modal logics, this requires that we introduce a symbol for strong identity, interpreted so that a strong identity is true just in case the flanking terms have the same intensions. Once this symbol is available, we simply adopt a rule of substitution of strong identities for term positions of the intensional predicate letters.

## 2 COMPLETENESS IN QUANTIFIED INTENSIONAL LOGIC

### 2.1 *Why Completeness is Hard to Prove in Quantified Modal Logic*

Completeness proofs in QML are quite a bit harder than completeness proofs for propositional modal logic or first-order logic. One reason that proofs are difficult is that sometimes there are none to find, as is the case of the conceptual interpretation **Q2**. Even when a system is complete, the proof may be elusive, and difficult to formulate in a simple way. Another problem is lack of generality: a proof strategy may only work when the underlying modal logic is fairly strong (for example, as strong as **S4**), or when *ad hoc* conditions are placed on the models.

One of the best ways to understand the methods used in completeness proofs for QML is to locate the main difficulty which arises if we simply try to ‘paste together’ proofs for quantificational logic and propositional modal logic. In order to uncover the problem, let us review the crucial steps in the completeness proofs in each kind of logic.

#### 2.1.1. *Completeness Proofs for Propositional Modal Logics*

The most powerful method for proving completeness of a propositional modal logic  $S$  is to use maximally consistent sets. Completeness follows if we can show that any  $S$ -consistent set is  $S$ -satisfiable. (A set is  $S$ -consistent iff there is no proof of a contradiction from the sentences in that set.) We begin by extending a given  $S$ -consistent set  $H$  to a maximally consistent set  $r$  (for real world) by Lindenbaum’s Lemma. Then we build what we will call the *standard model*  $\langle W, R, a \rangle$  for  $S$ . The set  $W$  of possible world of the model is taken to be the set of all maximally consistent sets of  $S$ , (on occasion,  $W$  contains just some of the maximally consistent sets related in some way to  $r$ ). The relation  $R$  (of accessibility) is usually defined so that  $wRw'$  iff if  $\Box A \in w$ , then  $A \in w'$ . Finally, the assignment function  $a$  is defined for propositional variables  $p$  so that  $a(p)(w)$  is  $T$  iff  $p \in w$ . The central lemma (TL) (for Truth Lemma) in the proof shows that membership in  $w$  and truth in  $w$  on the standard model amount to the same thing.

(TL)  $a(A)(w)$  is  $T$  iff  $A \in w$ .

Once (TL) is shown, it follows that all members of  $H$  are true at  $r$  on the standard model. We can also prove that  $\langle W, R \rangle \in R(S)$  (the set of Kripke frames that corresponds to  $S$ ), and so the standard model  $S$ -satisfies  $H$ .

The proof of (TL) is an induction on the construction of  $A$ , and the only really interesting case is when  $A$  has the shape  $\Box B$ . (The case for propositional variables is trivial given the definition of the standard model, and cases for  $\neg$  and  $\rightarrow$  simply depend on corresponding properties of maximally consistent sets  $w : \neg B \in w$  iff  $B \notin w$ , and  $B \rightarrow C \in w$  iff either  $B \notin w$  or  $C \in w$ .) The proof of the case for  $\Box$  takes the following form.

$$\begin{aligned} a(\Box A)(w) \text{ is } T & \text{ iff } \text{if } wRw' \text{ then } a(A)(w') \text{ is } T \\ (1) & \text{ iff } \text{if } wRw' \text{ then } A \in w' \\ (2) & \text{ iff } \Box A \in w. \end{aligned}$$

The only difficult part is to show the equivalence of (1) and (2). The inference from (2) to (1) is a simple consequence of the way we defined  $R$ . In order to show that (1) implies (2), we show  $(\neg\Box)$  instead.

$(\neg\Box)$  if  $\Box B \notin w$ , then there is a maximally consistent set  $w'$  such that  $wRw'$  and  $B \notin w'$ .

The proof of  $(\neg\Box)$  makes a second use of the Lindenbaum Lemma. Given that  $\Box B \notin w$ , we show the consistency of the set  $w^* = \{A : \Box A \in w\} \cup \{\neg B\}$ . Then we use the Lindenbaum Lemma to extend  $w^*$  to a maximally consistent set  $w'$ . The set  $w'$  is such that  $wRw'$  because for each sentence  $\Box A$  in  $w$ ,  $A \in w'$ ; it does not contain  $B$  since it is consistent and contains  $\neg B$ .

### 2.1.2. Completeness of First-order Logic

In this section we will give a quick review of a completeness proof for PL, first-order logic with identity. Again we show that any PL-consistent set is PL-satisfiable by first extending  $H$  to a maximally consistent set  $r$ , written in language  $L$ . We then construct a model  $\langle D, a \rangle$  from  $r$  as follows. The assignment function  $a$  is defined so that the extension  $a(t)$  of  $t$  is  $\{t' : t = t' \in r\}$ , the equivalence class of terms ruled identical in  $r$ . The domain  $D$  contains  $a(t)$  for each term  $t$ . The assignment function  $a$  is defined for  $i$ -ary predicate letters  $F$  so that  $\langle d_1, \dots, d_i \rangle$  is a member of  $a(F)$  just in case  $Ft_1, \dots, t_i \in r$  and  $a(t_j)$  is  $d_j$  for each of the  $t_j$  of  $t_1, \dots, t_i$ .

Given the presence of principles of identity, it is not difficult to show that (TL) holds for atomic sentences on this model. In order to establish (TL) for all sentences, we must be sure that the set  $r$  meets one further condition concerning the quantifier, namely  $(\forall x)$ .

$(\forall x)$   $a(\forall x Px)$  is  $T$  iff  $\forall x Px \in r$ .

The proof of  $(\forall x)$  will be ensured if we can show that  $r$  is omega-complete (OC).

(OC) If  $r \vdash Pt$ , for every term  $t$  of  $L$ , then  
 $r \vdash \forall xPx$ , for any variable  $x$ .

(Here we write ' $r \vdash A$ ' for ' $A$  is provable from the set of hypotheses  $r$ '.)  
 Notice that (OC) is equivalent to (OC').

(OC') If  $r \cup \{\neg\forall xPx\}$  is consistent, then  
 for some term  $t$  of  $L$ ,  $r \cup \{\neg Pt\}$  is consistent.

There are maximally consistent sets that are not omega-complete, so when we extend  $H$  to  $r$  using the Lindenbaum procedure, we must take special steps to guarantee (OC). Remember that the Lindenbaum method for extending a consistent set to a maximally consistent one begins by ordering the wffs. A series of sets  $M_0 = H, M_1, \dots$ , is then formed by letting  $M_{i+1}$  be the result of adding the  $i + 1$ th wff to  $M_i$ , iff doing so would leave  $M_{i+1}$  consistent. (Otherwise  $M_{i+1}$  is  $M_i$ .) The maximally consistent set desired is the union of all the  $M_i$ . To ensure a set is omega-complete during this construction, we do the following. If  $M_i$  is the  $i$ th set formed in that construction, and  $\neg\forall xPx$  is the  $i + 1$ th sentence in our ordering of all the well-formed formulas, and if adding  $\neg\forall xPx$  to  $M_i$  would yield a consistent set, then we form  $M_{i+1}$  from  $M_i$  by adding both  $\neg\forall xPx$ , and a sentence of the form  $\neg Pt$ , where  $t$  is a term that is new to  $\neg\forall xPx$  and  $M_i$ . It is not too hard to see that adding this second sentence to  $M_{i+1}$  cannot cause  $M_{i+1}$  to become inconsistent, as long as  $M_i$  plus  $\neg\forall xPx$  was already consistent as we have assumed. (The reason is that if  $M_{i+1} = M_i \cup \{\neg\forall xPx, \neg Pt\}$  were inconsistent, then  $M_i \cup \{\neg\forall xPx\} \vdash Pt$ . Since  $t$  is foreign to both  $M_i$  and  $\neg\forall xPx$ , it follows by the rule of Universal Generalisation that  $M_i \cup \{\neg\forall xPx\} \vdash \forall xPx$ , which entails that  $M_i \cup \{\neg\forall xPx\}$  is inconsistent, contrary to our assumption.) We can also see from the second formulation (OC') of omega-completeness that the result of the construction is omega-complete, and so a saturated set. (A saturated set is a maximally consistent set that is omega-complete.)

Now suppose we use this construction to produce a saturated extension  $r$  of  $H$ . As a result, we can show that  $(\forall x)$  holds in the model constructed from  $r$  by the following reasoning.

- |     |                         |     |  |
|-----|-------------------------|-----|--|
|     | $a(\forall xPx)$ is $T$ | iff | for all $d$ in $D$ , $a(d/x)(Px)$ is $T$   |
| (1) |                         | iff | for all terms $t$ , $a(a(t)/x)(Px)$ is $T$ |
| (2) |                         | iff | for all terms $t$ , $a(Pt)$ is $T$         |
| (3) |                         | iff | for all terms $t$ , $Pt \in w$             |
| (4) |                         | iff | $\forall xPx \in w$ .                      |

The equivalence between (1) and (2) is proven by a straightforward induction on the length of  $Px$ . The equivalence of (2) and (3) is the result of the hypothesis of the induction; (3) entails (4) because  $r$  is omega-complete; and (4) entails (3) because of the rule of Universal Instantiation.

Now that we have finished the proof of the case for  $\forall x$ , we have a proof of (TL). It follows that the PL-model we have defined satisfies all the sentences



of  $r$  and, hence, all sentences of our original set  $H$ . We conclude that any PL-consistent set is PL-satisfiable.

### 2.1.3. *The Difficulties in Quantified Modal Logics*

Notice that the method we described for constructing a saturated set for first-order logic requires that we have an infinite set of terms of  $L$  which are foreign to  $H$ . Since we may have infinitely many sentences  $\neg\forall xPx$  to add, we need infinitely many ‘instances’  $\neg Pt$  where  $t$  is new to the construction. As a result, the set  $w$  which we constructed using this method, contains an infinite set of terms of  $L$  which did not appear in  $H$ .

Now let us imagine that we hope to prove completeness of a modal logic  $Q$ , which adds principles of first-order logic to the propositional modal logic  $S$ . We begin with an  $Q$ -consistent set  $H$  which we hope to show is  $Q$ -satisfiable by extending  $H$  to a saturated set  $r$  written in language  $L$ . We then hope to construct the standard model, which will make all sentences of  $H$  true at  $r$ . Difficulties arise when we try to prove (TL), for there is a conflict between what we need to ensure  $(\forall x)$  and  $(\Box)$  together. Condition  $(\forall x)$  demands that the set  $W$  of possible worlds be the set of saturated sets in language  $L$ , for the terms of  $L$  (actually their equivalence classes) determine the domain of the quantification of our model. On the other hand, the proof of condition  $(\Box)$  requires the following. From a given possible world  $w$  which contains  $\neg\Box B$ , we must be able to construct a saturated set in language  $L$  which is an extension of  $w^* = \{A : \Box A \in w\} \cup \{\neg B\}$ . The problem is that in order to extend  $w^*$  to a saturated set in  $L$ , we must find an infinite set of terms of  $L$  that do not appear in  $w^*$ . However, the world  $w$  contains  $\Box(Pt \rightarrow Pt)$  for each term  $t$  of  $L$ , with the result that all formulas  $Pt \rightarrow Pt$  appear in  $w^*$ . So there are no terms of  $L$  foreign to  $w^*$ . If we attempt to remedy the problem at this point by constructing a world  $w'$  from  $W^*$  in a larger language  $L'$ , then we find ourselves in a vicious circle. Now we must prove property  $(\forall x)$  for  $L'$  instead of  $L$ . This forces us to define  $W$  as the set of all saturated sets in language  $L'$ , so that when we want to extend  $w^*$  to a saturated set, we must find infinitely many terms of  $L'$  foreign to  $w^*$ . However,  $w$  is now a saturated set in language  $L'$ , and contains  $\Box(Pt \rightarrow Pt)$  for all terms  $t$  of  $L'$ . Again, we have no guarantee that there are any terms of  $L'$  which do not appear in  $w^*$ .

## 2.2 *Strategies for Quantified Modal Logic Completeness Proofs*

In this section, we will illustrate four different strategies for obtaining completeness proofs in QML. Each of them has its strengths and weaknesses. Ideally, we would like to find a completely general completeness proof. The proof would demonstrate completeness of the most general semantics we have considered, namely **QS**. The proofs for all less general systems would

then fall out of the general proof just as proofs for the stronger propositional modal logics result from the completeness proof for **K**. This would help clarify and unify quantified modal logic. The strategy we present in Section 2.2.4 comes closest to providing such a general proof. However any such method will face some limitations for the reasons discussed at the end of Section 2.2.1.

*2.2.1. Strategy 1: Extend  $w^*$  to a saturated set without using any new terms (completeness of **Q1**)*

The completeness proof for **Q1** given by Thomason [1970] is worth reviewing because it illustrates an important strategy for overcoming the problem which we outlined in Section 2.1.3. Remember our difficulty was that we needed a way to extend a consistent set  $w^*$  to a saturated one, but we did not have an infinite set of terms missing from  $w^*$  in order to carry out the construction. The system **Q1** uses fixed domains, the objectual interpretation, and rigid terms. It verifies classical quantifier principles and the Barcan Formula. When these are present, it turns out that  $w^*$  is already omega-complete in the case of most modal logics. Since any consistent omega-complete set can be extended to a saturated set in the same language [Henkin, 1949], we can extend  $w^*$  to a saturated set without needing any extra terms.

The details of this reasoning are given in the following lemmas.

LEMMA 1. *If  $w$  is omega-complete, then so is  $w \cup f$ , provided  $f$  is finite.*

**Proof.** Suppose that  $w$  is omega-complete. To show that  $w \cup f$  is also omega-complete, let us assume that  $w \cup f \vdash Pt$  for all terms  $t$ . It follows that  $w \vdash \wedge f \rightarrow Pt$  for all terms  $t$ , where  $\wedge f$  is the conjunction of the members of  $f$ . Since  $w$  is omega-complete, it follows that  $w \vdash \forall x(\wedge f \rightarrow Px)$  for any choice of variable  $x$  we like. If we choose a variable  $x$  foreign to  $\wedge f$ , it follows that  $w \vdash \wedge f \rightarrow \forall xPx$ , and so  $w \cup f \vdash \forall xPx$ . By principles of quantificational logic, we can replace the variable  $x$  of  $\forall xPx$  for any other variable. It follows, then, that whenever  $w \cup f \vdash Pt$  for all terms  $t$ , then  $w \cup f \vdash \forall xPx$ , and so  $w \cup f$  is omega-complete. ■

LEMMA 2. *Any consistent omega-complete set  $w$  can be extended to a saturated set written in the same language.*

**Proof.** We construct a saturated extension of  $w$  using a variant of the method described in Section 2.1.2. Suppose that the set  $M_i$  plus  $\neg\forall xPx$  is consistent, so that we are to form  $M_{i+1}$  by adding  $\neg\forall xPx$  and an instance  $\neg Pt$  to  $M_i$ . Ordinarily, we would choose a term  $t$  foreign to both  $M_i$  and  $\neg\forall xPx$  in order to ensure that adding  $\neg Pt$  will not cause  $M_{i+1}$  to become

inconsistent. In this case, however, we must use a term  $t$  which may already appear in  $w$ . When  $w$  is omega-complete as we have assumed, it follows by Lemma 1 that  $M_i$  is omega-complete as well. ( $M_i$  is formed by adding only finitely many sentences to  $w$ .) Since  $M_i \cup \{\neg\forall xPx\}$  is consistent, it follows by formulation (OC') of omega-completeness that  $M_i \cup \{\neg Pt\}$  is consistent for some term  $t$  of  $L$ . So  $\neg Pt$  can be consistently added to  $M_i$  for this choice of  $t$ , and since  $\neg Pt$  entails  $\neg\forall xPx$ , the result of adding both these sentences to  $M_i$  remains consistent. Once we ensure that instances  $\neg Pt$  are consistently added in this way, it is a simple matter to verify that the union of the  $M_i$  is a saturated extension of  $w$ . ■

**LEMMA 3.** *If  $w$  is a saturated set which contains  $\neg\Box B$ , then  $w^* = \{A : \Box A \in w\} \cup \{\neg B\}$  is consistent and omega-complete.*

**Proof.** We can show that  $w^*$  is consistent just as we do in propositional modal logic. By Lemma 1,  $w^*$  is omega-complete if  $\{A : \Box A \in w\}$  is. Assume now that  $\{A : \Box A \in w\} \vdash Pt$  for every term  $t$ . By principles of the modal logic  $K$ ,  $w \vdash \Box Pt$  for each term  $t$ , and since  $w$  is omega-complete, it follows that  $w \vdash \forall x\Box Px$ . By the Barcan Formula, it follows that  $w \vdash \Box\forall xPx$ . Since  $w$  is maximal,  $\Box\forall xPx \in w$ , and so  $\forall xPx \in \{A : \Box A \in w\}$ . It follows that  $\{A : \Box A \in w\} \vdash \forall xPx$ . ■

**LEMMA 4.** *If  $w$  is a saturated set that contains  $\neg\Box B$  then  $w^* = \{A : \Box A \in w\} \cup \{\neg B\}$  can be extended to a saturated set written in the same language.*

**Proof.** By Lemma 3,  $w^*$  is consistent and omega-complete. By Lemma 2, it can be extended to a saturated set in the same language. ■

Now let us assume that the system **Q1** results from adding rules of classical logic, rules (**ID**) for identity, and (**RT**) for rigid terms to propositional modal logic  $S$ . To show completeness, we prove, as usual, that every **Q1**-consistent set is **Q1**-satisfiable. Given a consistent set, we extend it to a saturated set  $r$  written in language  $L$  in the usual way. We then construct the standard **Q1**-model  $\langle W, R, D, \mathbf{Q1}, a \rangle$  as follows.  $W$  is the set of all saturated sets that contain  $t = t'$  just in case  $t = t' \in r$ .  $R$  is defined in the usual way. The extension  $a(t)(w)$  of term  $t$  is  $\{t' : t = t' \in r\}$ .  $D$  is the set of all term extensions. Sequence  $\langle d_1, \dots, d_i \rangle \in a(F)(w)$  iff  $Ft_1, \dots, t_i \in w$  and  $a(t_j)(w)$  is  $d_j$  for the  $d_j$  of  $d_1, \dots, d_i$ . For most modal logics, we may show that  $\langle W, R \rangle \in R(S)$  just as we did in the completeness proof for  $S$ , and so once we prove the truth lemma (TL), we will know that the sentences of  $H$  are all true at  $r$  on this model. It will follow that  $H$  is **Q1**-satisfiable.

The interesting cases in the proof of (TL), concern  $\Box$  and  $\forall x$ . The proof of  $(\forall x)$  can be carried out along the lines we specified in Section 2.1.2. To establish  $(\Box)$ , it is crucial to show  $(\neg\Box)$ .

( $\neg\Box$ ) if  $\neg\Box B \in w$ , then there is a member  $w'$  of  $W$  such that  $wRw'$  and  $\neg B \in w'$ .

By Lemma 4, we know that we can construct a saturated extension  $w'$  of  $\{A : \Box A \in w\} \cup \{\neg B\}$ . We can show that this  $w'$  is a member of  $W$  if we can show that  $t = t' \in w'$  iff  $t = t' \in r$ . Since  $w$  is a member of  $W$ , we already know that  $t = t' \in w$  iff  $t = t' \in r$ . Notice that if  $t = t' \in w$ , then by (RT),  $\Box(t = t') \in w$ , and  $t = t' \in w'$ . If  $t = t' \notin w$ , then  $\neg t = t'$  is, and so by (RT)  $\Box\neg t = t' \in w$ , and  $\neg t = t' \in w'$ . It follows that  $w'$  contains exactly the identities of  $r$  and so is a member of  $W$ . Since  $\{A : \Box A \in w\}$  is a subset of  $w'$ , we know that  $wRw'$ , and so we have completed the proof of ( $\neg\Box$ ).

Strategy 1 has important limitations. First, the method depends on using first-order logic and the Barcan Formulas, so it is not applicable to systems that give a more general account of the quantifiers. Second, the completeness result is blocked for certain underlying modal logics  $S$ . We illustrate the problem with modal logics where  $R$  is convergent.

In proving that the standard model is convergent for propositional modal logics, one assumes  $wRw'$  and  $wRw''$ , establishes the consistency of  $\{A : \Box A \in w'\} \cup \{A : \Box A \in w''\}$ , and then employs the Lindenbaum Lemma to extend this set to a maximally consistent set  $w'''$  such that  $w'Rw'''$  and  $w''Rw'''$ . In the case of a quantified modal logic, we must know that  $\{A : \Box A \in w'\} \cup \{A : \Box A \in w''\}$  is omega-complete as well as consistent before Lemma 2 can be used to extend it to a saturated set. However, there is no guarantee that  $\{A : \Box A \in w'\} \cup \{A : \Box A \in w''\}$  will be omega-complete. It will not be, for example, if  $\{A : \Box A \in w'\}$  contains each of  $Pt_1, Pt_3, \dots$ , and  $\{A : \Box A \in w''\}$  contains  $\sim \forall xPx, Pt_2, Pt_4, \dots$ , and  $t_1, t_2, \dots$  is a list of all terms of  $L$ . Under these circumstances  $\{A : \Box A \in w'\} \cup \{A : \Box A \in w''\}$  contains  $\{\sim \forall xPx, Pt_1, Pt_2, Pt_3, \dots\}$  and so is not omega-complete.

Difficulties of this kind can be expected whenever the proof that  $\langle W, R \rangle \in R(S)$  for the propositional modal logic  $S$  rests on proving the existence of a consistent set, and then extending it to a maximally consistent set by the Lindenbaum Lemma. (Convergence and density are two conditions where this technique is typically used.) In this kind of case, the proof that  $\langle W, R \rangle \in (S)$  may fail for the quantificational logic when a consistent set formed fails to be omega-complete.

The problem does not arise for most modal logics. Strategy 1 works to show completeness, for example, for systems whose corresponding conditions on  $R$  are preserved under subsets. (Conditions are preserved under subsets iff when the conditions hold for  $\langle W, R \rangle$  they also hold for  $\langle W', R' \rangle$ , where  $W'$  is a subset of  $W$  and  $R'$  is  $R$  restricted to  $W'$ .) Conditions preserved under subsets include the universal conditions, i.e. conditions on  $R$  that can be expressed with universal quantifiers alone. However, for systems whose conditions are not preserved under subsets, strategy 1 does not necessarily

yield a completeness result. This failure is directly related to the fact that system **Q1-S** is not complete for a semantics with convergent  $R$  [Cresswell, 1995].

*2.2.2. Strategy 2: Build the set of possible worlds all in one construction (completeness for **Q1-S5**)*

Gallin [1975, p. 25 ff.] offers another strategy for proving completeness of **S5** systems that contain classical principles and the Barcan Formula. It is a clever technique which has applications to systems with weaker rules. Gallin avoids the complication we encountered in extending  $w^*$  to a saturated set by defining the set of worlds of his standard model so that all the worlds  $w$  are saturated and already satisfying condition  $(\neg\Box\mathbf{S5})$ .

$(\neg\Box\mathbf{S5})$  If  $\neg\Box A \in w$ , then there is a world  $w'$  such that  $\neg A \in w'$ .

In **S5**, this condition is sufficient for demonstrating the case of (TL) for formulas that begin with  $\Box$ .

Gallin shows how to build a whole collection  $W$  of saturated sets from a consistent set  $H$ , using a variation of the Lindenbaum construction. The sets in  $W$  are the possible worlds of the standard model. In order to co-ordinate the construction properly, let  $W$  be a sequence  $w_0, w_1, w_2, \dots$  of possible worlds.  $W$  is constructed from a consistent set  $H$ , using a series  $W_0, W_1, W_2, \dots$ . Each of the  $W_i$  contains a sequence  $w_0, w_1, w_2, \dots$  of consistent sets, each of which is on its way to becoming saturated as we move to larger  $W_j$ . The  $W_i$  are also arranged so that eventually,  $(\neg\Box\mathbf{S5})$  is met for each formula  $A$ .

To define the  $W_i$ , we need a generalisation of the notion of consistency. We say that a sequence  $W$  of sets is consistent just in case no finite subset  $f$  of any of the sets  $w$  in  $W$  is such that  $\Diamond \wedge f \vdash p \wedge \neg p$ . A formula  $A$  can be consistently added to world  $w$  of sequence  $W$  just in case doing so would leave the sequence  $W$  consistent. This definition of consistency ensures not only that adding  $A$  to a world  $w$  leaves  $w$  consistent, but that adding  $A$  is also consistent with all the facts about all the other worlds.

Now we are ready to define the series  $W_0, W_1, W_2, \dots$ . We let  $W_0$  be the sequence such that its first world  $w_0$  is  $H$ , and all the other worlds  $w_1, w_2, \dots$  are empty. We then order the pairs  $\langle i, A \rangle$  consisting of integers  $i$  and formulas  $A$ , and for each pair  $\langle i, A \rangle$ , we pick a term  $t(i, A)$ , which is foreign to  $H$ , and all sentences of previous pairs in the ordering. For each  $W_j$ , we define  $W_{j+1}$  as follows. We consider the  $j + 1$ th pair  $\langle i, A \rangle$  in the ordering and we add  $A$  to world  $w_i$  of  $W_j$  iff  $H$  can be consistently added to  $w_i$  of  $W_j$ . (Otherwise we set  $W_{j+1}$  equal to  $W_j$ .) In case  $A$  has the shape  $\neg\forall xPx$ , we also add  $\neg Pt$ , where  $t$  is  $t(i, A)$ . In case  $A$  has the shape  $\neg\Box A$ , we also find the first empty set in the sequence  $W_{j+1}$ , and we add  $\neg A$  to it. There is such an empty set in  $W_{j+1}$ , because we have only added

finitely many formulas to this point, and  $W_0$  contained an infinite sequence of empty sets. It is also clear that adding this formula could not cause  $W_{j+1}$  to become inconsistent.

Once the  $W_j$  have been defined this way, we let  $W$  be the sequence we get by letting the  $i$ th world of  $W_j$  be the union over all the sets  $w_0, w_1, w_2, \dots$ , which were the  $i$ th worlds of  $W_0, W_1, W_2, \dots$ . It is not difficult to prove that each of the worlds of  $W$  is a saturated set that meets property  $(\neg \Box \mathbf{S5})$ . Notice, however, that because of the special definition Gallin uses for consistency, the demonstration that these sets are saturated requires the Barcan Formula and classical principles for the quantifiers.

Gallin claims that this proof is significantly easier than the method we presented as strategy 1. We do not agree with Gallin's taste in simplicity. However, this strategy is quite interesting, and it can be modified for use with weaker rules as [Menzel, 1991] shows.

*2.2.3. Strategy 3: Allow the language to vary across possible worlds*

The second strategy we are going to discuss is illustrated by a completeness proof [Garson, 1978] for **QS**, the most general semantics we have described. The same idea will be used to sketch the proof of the completeness for **QPL** along the lines of Hughes and Cresswell [1968, p. 147 ff.] and Gabbay [1976, p. 46 ff.].

**2.2.3.1 Completeness of QS.** In systems with world-relative domains, the Barcan Formula is not valid, and so we no longer know that  $\{A : \Box A \in w\}$  is omega-complete. Notice, however, that since the domain of quantification varies from one possible world to the next, we are free to select a different language for each of the saturated sets which are in  $W$  in the standard model. When it comes time to construct a saturated set from  $w^*$ , we simply build a saturated set in a language larger than the one in which  $w$  is written.

Since **QS** is based on free logic, we have to readjust our definition of omega-completeness and, hence, our definition of saturation. An omega-complete set for free logic in language  $L$  is any set that meets condition (FOC).

(FOC) If  $w \vdash Et \rightarrow Pt$  for every term  $t$  of  $L$ , then  $w \vdash \forall xPx$  for any variable  $x$ .

A *free logic saturated set* for  $L$  is simply any maximally consistent set  $w$  for which (FOC) holds. It is easy to prove that a consistent set written in language  $L$  can be extended to a set which is free logic saturated for a language with infinitely many more terms than  $L$ . To provide the proof simply replace  $(Et \rightarrow Pt)$  for  $Pt$  in the corresponding proof for first-order logic (see Section 2.1.2).

Let **QS** be the logic that results from adding the principles of **MFL** and **ID** to certain propositional modal logics  $S$ . We will explain more about which logics these are later. We will demonstrate the completeness of **QS** with respect to the set of all **QS**-models (world relative substantial models for  $S$ ). See Section 1.4.1 for the definition of a **QS**-model.

As usual we assume that set  $H$  is consistent in **QS**, and we extend  $H$  to a free logic saturated set  $r$  written in a language  $L$ . At this point, however, we consider a larger language  $L^+$ , which contains infinitely many terms which are not in  $L$ . We then define the set  $W$  of possible worlds for our *standard QS-model*  $\langle W, R, D, S, a \rangle$  as the set of all free logic saturated sets written in some language  $L'$  such that there are infinitely many terms of  $L^+$  that do not appear in  $L'$ . The idea behind this is to guarantee that whenever  $w \in W$ , there will be infinitely many terms foreign to  $w^* = \{A : \Box A \in w\} \cup \{\neg B\}$  so that  $w^*$  can be extended to a saturated set in language  $L^+$ . The other parts of the definition of the standard **QS**-model are straightforward.  $R$  is defined in the usual way:  $wRw'$  iff if  $\Box A \in w$ , then  $A \in w'$ . The intension  $a(t)$  of a term  $t$  given by  $a$  is defined so that  $a(t)(w)$  is  $\{t : t = t' \in w\}$ , the equivalence class of terms ruled identical in  $w$ .  $S$  is defined so that  $s \in S(w)$  iff  $s$  is  $a(t)$  for some term  $t$  such that  $Et \in w$ . The domain of possible objects  $D$  is simply the set of all term extensions in all the possible worlds. The intension  $a(F)$  of an  $i$ -ary predicate letter  $F$  is given as one would expect:  $\langle d_1, \dots, d_i \rangle \in a(F)(w)$  iff  $Ft_1, \dots, t_i \in w$  and each of the  $a(t_j)(w)$  is  $d_j$ . The intension  $a(E)$  is  $S$ .

Because the members of  $w$  are free logic saturated sets written in different languages, we cannot prove the Truth Lemma (TL) for this standard model. If  $t$  does not appear in  $Lw$ , the language in which the saturated set  $w$  is written, then  $a(\neg Ft)(w)$  is  $T$ , but  $\neg Ft \notin w$ . However, there is a weaker formulation ( $w$ TL) which will still serve our purposes.

( $w$ TL) If  $A$  is a sentence of  $Lw$ , then  $a(A)(w)$  is  $T$  iff  $A \in w$ .

The proof of ( $w$ TL) for cases other than  $\Box$  and  $\forall x$  is straightforward. The crucial step in the case for  $\Box$  is to demonstrate ( $\neg\Box$ ).

( $\neg\Box$ ) If  $\Box B$  is a sentence of  $Lw$ , then if  $\neg\Box B \in w$  then there is a  $w'$  in  $W$  such that  $wRw'$  and  $\neg B \in w'$ .

We begin the proof by assuming that  $\Box B$  is a sentence of  $Lw$ , and that  $\neg\Box B \in w$ . We construct  $w^* = \{A : \Box A \in w\} \cup \{\neg B\}$  which we show to be consistent in the usual way. Since  $w$  is a member of  $W$ , there must be an infinite set  $N$  of terms of  $L^+$  that do not appear in  $w$ . By the definition of  $w^*$ , it is clear that none of these terms appear in  $w^*$  either. We could construct a free logic saturated set  $w'$  from  $w^*$  using these terms. However, if  $w'$  is to be a member of  $W$ , there must be an infinite set of terms of  $L^+$  foreign to  $w'$ . In order to ensure that we do not ‘use up’ all the terms in our

construction of  $w'$ , we divide  $N$  into two infinite sets  $N_1$  and  $N_2$ . We use  $N_1$  to extend  $w^*$  to a free logic saturated set  $w'$ , and we leave  $N_2$  in reserve to ensure that  $w' \in W$ . When  $w'$  is constructed in this way, we can easily prove that  $wRw'$ , and that  $\neg B \in w'$ , and so we have finished the proof of  $(\neg\Box)$ .

We would have skipped the case for  $\forall x$  if it were not for one ticklish point. Along the way, we need to show (ES).

$$(ES) \quad a(t') \in S(w) \text{ iff } Et' \in w.$$

((ES) is also needed to show the case of formulas with the shape  $Et$ .)

The proof of (ES) would seem to be trivial given our definition of  $S(w)$ , but it is not. The trouble comes in showing (ES) from left to right. Suppose that  $a(t') \in S(w)$ . Then by the definition of  $S(w)$ , there is a term  $t$  such that  $a(t')$  is at  $a(t)$  and  $Et \in w$ . For ordinary predicates, this would be enough to ensure that  $Et' \in w$ , for when  $a(t)(w)$  is  $a(t')(w)$ , we have that  $t = t' \in w$ , and so can substitute  $t'$  for  $t$ . Remember, however, that  $E$  is an intensional predicate for which the rule of substitution of identities does not hold, so this reasoning will not work. We must find some other way to ensure that  $Et' \in w$ . Things look bad when we realise that  $t'$  may not even be in the language  $Lw$ , in which case  $Et' \notin w$ . Luckily, our definition of the standard model ensures that whenever  $a(t)$  is  $a(t')$  then  $t$  and  $t'$  are the same term. The reason is that when  $t \notin Lw$ , it follows that  $a(t)(w) = \{t' : t = t' \in w\}$  is empty. For any pair of distinct terms  $t, t'$  we choose, we can always find a language  $Lw$  such that  $t$  is in  $Lw$  and  $t'$  is not. It follows that the only way that  $a(t)$  and  $a(t')$  can be identical is if  $t$  is identical to  $t'$ . We have that  $Et \in w$ , so we conclude that  $Et' \in w$  and our proof of (ES) is finished.

Once Lemma ( $wTL$ ) is established in this way, the completeness of **QS** is shown fairly easily. We have already extended the **QS**-consistent set  $H$  to a free logic saturated set  $r$ , and since there were infinitely many terms foreign to  $r$  in  $L^+$ , it turns out that  $r \in W$ . By ( $wTL$ ), it follows that all members of  $r$  (and so all members of  $H$ ) are true at  $r$  on the standard model, and so  $H$  is **QS**-satisfiable.

Although this proof is satisfying because it shows completeness for a system with a very general treatment of the quantifiers, it does not count as the general sort of completeness proof which we desire. The reason is that the strategy does not work to establish completeness of systems that use less general treatments of the quantifiers. For example, we might hope to show the completeness of the objectual interpretation with world relative domains and rigid terms by considering the system which results from adding (RT) to **QS**. We would hope that (RT) would ensure that terms are rigid on our standard model, with the result that all members of  $S(w)$  are constant functions.

However, these hopes cannot be realised using the present definition of the standard model. In order to ensure that ( $wTL$ ) holds for sentences



$t = t'$ , we are virtually forced into defining  $a(t)(w)$  as  $\{t : t = t' \in w\}$ . If any term  $t$  is rigid on this model, it would follow that  $a(t)(w)$  is  $a(t)(w')$ , and so that  $w$  and  $w'$  share exactly the same identities. Because every saturated set for  $L$  contains  $t = t$  for every term  $t$  of  $L$ , it follows that  $w$  and  $w'$  must be written in languages with the same terms. However, the strategy of this completeness proof depends on allowing our languages to shift from one saturated set to the next. Using similar reasoning, we can see that it is pointless to hope for a completeness proof for systems with fixed domains using the standard model of this section.

There is another respect in which the variable language strategy lacks generality. The method does not work for all propositional modal logics  $S$ . (Garson's [1978] claim to the contrary is an error.) The reason is that when possible worlds are written in different languages, we lose an important property ( $\diamond$ ) which is needed in showing that  $\langle W, R \rangle$  on the standard model is in  $R(S)$ .

( $\diamond$ )     If  $wRw'$  and  $A \in w'$ , then  $\diamond A \in w$ .

This property fails if term  $t$  is in the language of  $w'$ , but not the language of  $w$ , and  $A$  is (say)  $Ft$ . The sentence  $\diamond Ft$  cannot be in  $w$  because it is not in the language of  $w$ .

For many modal logics (for example, **D**, **M**, and **S4**), we do not need ( $\diamond$ ) in order to show that  $\langle W, R \rangle \in R(S)$ . However, for systems like **B**, the property seems indispensable. There are tricks one can use to overcome the difficulty for individual systems, but the changing language strategy does not provide a proof that is general with respect to the underlying modal logic.

**2.2.3.2 Completeness of QPL without identity.** When  $=$  is absent from our language, the problems we described in extending the completeness proof of **QS** to systems that use the objectual interpretation can be overcome, at least for some of the propositional modal logics. We will illustrate this by sketching the proof for **QPL** with respect to a **QPL**-semantics, where we use the objectual interpretation, world-relative domains, the nesting condition (ND), and truth value gaps. (See Section 1.2.1.2.3.2). We will be assuming that the underlying modal logic  $S$  does not require property ( $\diamond$ ) for its completeness proof. Remember that the system **QPL** simply results from adding the rules of first-order logic to  $S$ . Since we are using classical principles, we define the standard model using the ordinary definition of saturation. Since identity is absent, we may simply let the extension of a term (at any world) be itself. This ensures the rigidity of the terms, and so the objectual interpretation for the domains. It is easy to arrange that domains are nested in the standard model by defining  $R$  so that  $wRw'$  iff  $w'$  contains the terms of  $w$ , and if  $\Box A \in w$ , then  $A \in w'$ . This calls for no changes in the proof of the case for  $\Box$ .

It is particularly convenient that we are allowed truth value gaps in this semantics, since we may consider each world  $w$  as defining the class of sentences defined at  $w$ . The formal neatness of truth value gaps at this point suggests that their introduction was not designed to meet philosophical intuitions, but rather to avoid formal complications in the completeness proof.

2.2.4. *Strategy 4: Redefine Saturation*

Thomason's [1970] proof of the completeness of **Q3** is the inspiration for the next strategy we are going to present. At the risk of repetition, we will give a second completeness proof for **QS**. Once we have presented the details, we will show how to modify the proof to obtain completeness results for **Q3**, and several other systems.

Strategy 4 follows the outlines of strategy 1; however, the concept of omega-completeness is adjusted to reflect the fact that the Barcan Formula and classical principles of quantification are no longer available. As we have already pointed out,  $w^*$  is not omega-complete in logics that lack the Barcan Formula. However,  $w^*$  has a weaker property which ensures that  $w^*$  can be extended to a set that has a correspondingly weaker form of saturation, a form which nevertheless ensures a proof of the quantifier case of the Truth Lemma.

Although this strategy turns out to be quite powerful, it has the disadvantage that we must reformulate the quantificational principles in a more general, and more complex way. In order to help simplify our presentation, we will adopt a few abbreviations. We use ' $\rightarrow$ ' for strict implication, so that ' $A \rightarrow B$ ' abbreviates ' $\Box(A \rightarrow B)$ '. We will be working constantly with formulas that have the shape (GF), where parentheses are to be restored from right to left.

$$(GF) \quad A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_i \rightarrow B.$$

(For example,  $A \rightarrow B \rightarrow C \rightarrow D$  amounts to  $A \rightarrow (B \rightarrow (C \rightarrow D))$ , or  $A \rightarrow \Box(B \rightarrow \Box(C \rightarrow D))$ .) We will use ' $G(B)$ ' to represent any sentence with shape (GF), and  $G(C)$  will be the sentence that results from replacing  $C$  for  $B$  in  $G(B)$ . Using this notation, we may now present two general rules for the quantifiers.

$$(GUI) \quad \frac{G(\forall xPx)}{G(Et \rightarrow Pt)}$$

$$(GUG) \quad \frac{\vdash G(Et \rightarrow Pt)}{\vdash G(\forall xPx)} \text{ where } t \text{ does not appear in } G(\forall xPx).$$

We should make clear that  $G(A)$  may represent a sentence where any of the arrows (whether  $\rightarrow$  or  $\rightarrow\!\!\rightarrow$ ) is missing in the pattern (GF). So all of the following, for example, are instances of the rule (GUI).

$$\frac{\forall xPx}{Et \rightarrow Pt} \quad \frac{A \rightarrow \forall xPx}{A \rightarrow Et \rightarrow Pt} \quad \frac{A \rightarrow\!\!\rightarrow \forall xPx}{A \rightarrow\!\!\rightarrow Et \rightarrow Pt} \quad \frac{A \rightarrow\!\!\rightarrow B \rightarrow\!\!\rightarrow \forall xPx}{A \rightarrow\!\!\rightarrow B \rightarrow\!\!\rightarrow Et \rightarrow Pt}.$$

The reader can verify that (GUI) and (GUG) are **QS**-valid.

The system (GS) consists of (GUI), (GUG), (=In), (=Out), and principles for propositional modal logics  $S$ . The quantifier rules (GUI) and (GUG) appear to be very odd and cumbersome. However, GS has a simple and natural reformulation in natural deduction format. The propositional modal logic  $K$  may be formulated by introducing boxed subproofs:

$$\boxed{\quad}$$

Together with introduction and elimination rules for  $\boxed{\quad}$ :

$$(\boxed{\text{In}}) \quad \frac{\boxed{\begin{array}{c} \vdots \\ A \end{array}}}{\boxed{A}}$$

$$(\boxed{\text{Elim}}) \quad \frac{\boxed{A} \quad \boxed{\begin{array}{c} \vdots \\ A \end{array}}}{\vdots}$$

(See [Konyndyk, 1986, p. 34 ff].)

When natural deduction rules are employed, GS may be reformulated using the standard free logic rules (FUI) and (FUG), with the understanding that these apply within any subproof. It is a straightforward matter to show that this natural deduction formulation is equivalent to GS.

Another feature of GS is evidence for its naturalness. One would hope to construct a quantified modal logic with *fixed* domains by adding  $Et$  as an axiom, thus ensuring that the free logic rules collapse to their classical counterparts. In QS, the addition of  $Et$  entails (CBF), but (BF) is independent, and must be added as a separate axiom. However, when  $Et$  is added to GS, both the Barcan Formula (BF) and its converse (CBF) are provable. It is pleasing that the generalised rules are symmetrical with respect to the adoption of the Barcan Formula and its converse.

The concept of omega-completeness which corresponds to the rules of GS is (GOC) (for *general omega-completeness*).

(GOC) If  $w \vdash G(Et \rightarrow Pt)$  for every term  $t$  of  $L$ ,  
then  $w \vdash G(\forall xPx)$ , for any variable  $x$ .

A *GOC set* is just a set with property (GOC), and a *set is generally saturated* (for language  $L$ ) just in case it is a maximally consistent GOC set.

Our next task is to state and prove analogues of Lemmas 1–4 of Section 2.2.1 for general omega-completeness and general saturation.

LEMMA G1. *If  $w$  is GOC, then so is  $w \cup f$ , provided that  $f$  is finite.*

**Proof.** Suppose that  $w$  is GOC, and assume that for all terms  $t$ ,  $w \cup f \vdash G(Pt)$ . It follows that  $w \vdash \wedge f \rightarrow G(Pt)$ . By propositional logic, this sentence is equivalent to one with the shape (GF), so we know that  $w \vdash \wedge f \rightarrow G(\forall xPx)$ , and hence that  $w \cup f \vdash G(\forall xPx)$ . ■

LEMMA G2. *Any consistent set  $w$  with property (GOC) can be extended to a generally saturated set written in the same language.*

**Proof.** If  $\neg G(\forall xPx)$  is the candidate for addition to  $M_i$  in the Lindenbaum construction, and if  $M_i \cup \{\neg G(\forall xPx)\}$  is consistent, then we add both  $\neg G(\forall xPx)$  and  $\neg G(Et \rightarrow Pt)$  to  $M_i$  to form  $M_{i+1}$ , for some term  $t$  which leaves  $M_{i+1}$  consistent. There is such a term because  $w$  is GOC and so, by Lemma G1,  $M_{i+1}$  is GOC. This construction preserves consistency, and results in a GOC set, and so it yields a generally saturated set. ■

LEMMA G3. *If  $w$  is a generally saturated set that contains  $\neg \Box B$ , then  $w^* = \{A : \Box A \in w\} \cup \{\neg B\}$  is consistent and GOC.*

**Proof.** The consistency of  $w^*$  is proven in the standard way. To show that  $w^*$  is GOC, assume that  $w^* \vdash G(Et \rightarrow Pt)$  for any term  $t$  of  $L$ . It follows that  $\{A : \Box A \in w\} \vdash \neg B \rightarrow G(Et \rightarrow Pt)$ . By principles of propositional modal logic  $\mathbf{K}$ ,  $w \vdash \Box(\neg B \rightarrow G(Et \rightarrow Pt))$ , and so  $w \vdash \neg B \rightarrow \Box G(Et \rightarrow Pt)$  for every term  $t$  of  $L$ . Since  $w$  is GOC,  $w \vdash \neg B \rightarrow \Box G(\forall xPx)$ , and since  $w$  is maximal,  $\neg B \rightarrow \Box G(\forall xPx) \in w$ . As a result,  $\neg B \rightarrow G(\forall xPx) \in \{A : \Box A \in w\}$ , and so  $w^* \vdash G(\forall xPx)$ . ■

LEMMA G4. *If  $w$  is generally saturated and contains  $\neg \Box B$ , then  $w^* = \{A : \Box A \in w\} \cup \{\neg B\}$  can be extended to a generally saturated set written in the same language.*

**Proof.** By Lemmas G2 and G3. ■

2.2.4.1 *Completeness of GS*. Now that we have proven Lemmas G1–G4, only a few details need to be mentioned to finish a completeness proof for **GS**. We begin with a **GS**-consistent set, and we extend it to a generally saturated set  $r$  written in language  $L$ . (To do so, we merely generalise the standard construction so that when  $\neg G(\forall xPx)$  is added, then so is  $\neg G(Et \rightarrow Pt)$ , where  $t$  is new to the construction.) We define the standard **GS**-model so that  $W$  is the set of all generally saturated sets for  $L$ . Items  $R, D, S$  and  $a$  are defined in exactly the way as they were in Section 2.2.1. We may also prove the stronger truth lemma (TL) in a straightforward way. The case for  $\Box$  requires that we show that if  $\neg\Box A \in w$ , then there is a  $w'$  in  $W$  such that  $wRw'$  and  $\neg A \in w'$ , but this is easily established using Lemma G4.

To prove the case for  $\forall x$  we notice first that all generally saturated sets are free logic saturated, because free logic omega-completeness (FOC) is a special case of (GOC) when  $G(Et \rightarrow Pt)$  is  $Et \rightarrow Pt$ . So we will have no difficulty proving that  $a(\forall xPx)(w)$  is  $T$  iff  $\forall xPx \in w$  as long as we can show (ES).

$$(ES) \quad a(t) \in S(w) \text{ iff } Et \in w.$$

In order to show (ES) in Section 2.2.3.1, we proved that if  $t$  and  $t'$  are distinct, then so are their intensions  $a(t)$  and  $a(t')$ . We can show this is true of the standard **GS**-model as follows. In all the systems we are considering, the sentence  $\neg t = t'$  is consistent if  $t$  and  $t'$  are distinct. So there is a generally saturated set in  $W$  that contains  $\neg t = t'$ , and the extensions of  $t$  and  $t'$  differ there.

This method of proving completeness has a number of advantages. Since all our sets are generally saturated in the same language, we no longer face the difficulties noted in Section 2.2.3 in showing that  $\langle W, R \rangle \in R(S)$ . Property  $(\diamond)$  now holds, and so the proof proceeds exactly the way it does in propositional modal logics. However, there are still modal logics for which the method does not apply. The proof is still blocked, for example, when  $R$  is convergent for reasons similar to the ones we explained at the end of Section 2.2.1. Sets we can show to be consistent which we would hope to extend to a generally saturated set by Lemma G2 need not be GOC.

Although strategy 4 does not solve the completeness problem for all underlying propositional modal logics, it can be generalised in another way. Once a completeness proof is available for **GS**, the method may be modified to obtain completeness results for extensions of **GS** that correspond to less general treatments of the terms and the quantifiers. A number of variations on this theme will be explored in the next sections.

Despite its generality, there is another problem with this method. The systems we have proven complete use the generalised quantifier rules (GUI) and (GUG). We would like to be able to show completeness for logics which use the more modest principles (FUI) and (FUG) of free logic. However, this is not always possible. Parsons [1975] has shown that (GUI) is independent

from the free logic rules in **Q3**. One reason for the sporadic nature of published completeness results is that certain systems are complete only when the generalised quantifier rules are chosen. Determining the conditions under which the generalised rules are necessary is an interesting topic for future research.

2.2.4.2 *Rigid Terms: Completeness of **GQ1R***. One advantage of strategy 4 is that it can be used to obtain completeness proofs for a variety of logics that use the objectual interpretation, even if they contain identity. A simple formulation of a system **GQ1R** which is complete for the objectual interpretation results from adding the rules (RT) and (=E) to **GS** to ensure that all the terms are rigid.

$$(=E) \quad \frac{t = t'}{Et \rightarrow Et'}$$

Remember that E is an intensional predicate in **GS**, and so the rule of substitution does not apply to it. However, once the terms are rigid, substitution of identicals is valid in all contexts, and so (=E) is valid.

It is not difficult to show the completeness of **GQ1R** for the objectual interpretation with rigid terms and world relative domains. Only one change in the definition of the standard model is required, along with a simple adjustment to the proof of (TL). We begin with a consistent set  $H$ , which we extend to  $r$ , a generally saturated set in  $L$ . We then define the standard model as before, except we ensure the rigidity of all the terms by restricting  $W$  to sets that contain exactly the identities of  $r$ . We must adjust the proof of the case for  $\Box$  because we will need to know that  $w^*$  can be extended to a set that contains the same identities as  $r$ . However, this can be shown using virtually the same argument we gave in Section 2.2.1, using the fact that (RT) is provable in **GQ1R**. Because our terms are rigid, the proof of (ES) is simplified. Since substitution now holds in the term slot of  $E$ , the proof that  $Et \in w$  iff  $a(t) \in S(w)$  no longer requires a demonstration that the intensions of  $t$  and  $t'$  are identical only if  $t$  and  $t'$  are identical.

Since all term intensions are rigid on this standard model, and since our domains contain only term intension, we can modify the model by replacing each constant term intension in a domain  $D(w)$  with its value. The result is a **Q1R**-model which satisfies  $r$  and hence,  $H$ .

2.2.4.3 *Fixed Domains: Completeness of **GQ1***. It is a simple matter to verify that adding (CBF) (the converse of the Barcan Formula) to **GS** ensures that the standard model meets the nesting condition (ND).

$$(CBF) \quad \Box \forall x Px \rightarrow \forall x \Box Px.$$

$$(ND) \quad \text{If } wRw' \text{ then } D(w) \text{ is a subset of } D(w').$$

The relationship between (CBF) and (ND) can be appreciated better when it is pointed out that (BF) is equivalent (in free logic plus modal logic **K**) to  $(\Box E)$ .

$$(\Box E) \quad \forall x \Box Ex.$$

Objection to  $(\Box E)$  prompted our interest in logics with world relative domains. It is not hard to see that any model that satisfies  $(\Box E)$  meets the nesting condition.

Presence of the Barcan Formula (BF) forces the ‘converse’ condition (CND) on the standard model.

$$(\text{BF}) \quad \forall x \Box Px \rightarrow \Box \forall x Px$$

(CND) If  $wRw'$  then  $D(w')$  is a subset of  $D(w)$ .

Let us restrict the domain  $W$  of the standard model so that it contains only worlds such that  $rR^i w$ , where  $R^i$  is the result of composing  $R$  with itself  $i$  times, and  $R^0$  is the identity relation. It follows from the presence of both (BF) and (CBF) that the domains of the standard model are all identical, and so can be collapsed into one. So we may use strategy 4 to give a completeness proof for a semantics with a fixed domain of the quantifier, but with a possibly wider domain for the terms.

In order to prove completeness for **GQ1**, we need only ensure that the terms are all given extensions in the domain of quantification. The standard model meets this condition when (E) is added to **GS**, and so we have an easy completeness proof of **GQ1** = **GQ1R** + (E).

$$(E) \quad Et$$

It is interesting to note that both (BF) and (CBF) are derivable as soon as (E) is added to **GS**. In free logic, the addition of  $Et$  would restore the classical quantifier rules, and so allow us to prove (CBF); but (BF) is still independent. It is pleasing that the generalised rules are symmetrical with respect to the adoption of the Barcan Formula and its converse.

**2.2.4.4 Nonrigid Terms: Completeness of Q3.** Something like strategy 4 was invented by Thomason to prove completeness of **Q3–S4**. The system he showed complete is necessarily based on the generalised quantifier rules. We will use strategy 4 here to prove completeness of several kinds of **Q3** logics. In our discussion of systems with the objectual interpretation and non-rigid terms (Section 1.2.2), we pointed out that quantifier rules are quite complicated unless we introduce a primitive predicate that expresses that a term intension is a constant function. We have been presuming all along that there is a primitive predicate  $E$  in our language which is interpreted so that  $a(E)$  is  $S$ , the set of ‘real’ substances. So we will begin with proofs for systems with arbitrarily strong modal logic and a primitive

existence predicate. Later we will show how to modify the proof for systems as strong as **S4**, so that the inclusion of a primitive predicate is not needed.

There is a problem which arises when we allow non-rigid terms with the objectual interpretation which draws our attention to a step in the proof of (TL) which we have so far ignored. Let us look at the reasoning we will need to carry out the proof of the case for the quantifier.

- $$\begin{array}{ll}
 a(\forall x Px)(w) \text{ is } T & \text{iff for all } d \text{ in } D(w), a(d/x)(Px)(w) \text{ is } T \\
 (1) & \text{iff for all } t, \text{ if } a(t)(w) \in D(w), \\
 & \text{then } a(a(t)(w)/x)(Px)(w) \text{ is } T \\
 (2) & \text{iff for all } t, \text{ if } a(t)(w) \in D(w), \\
 & \text{then } a(Pt)(w) \text{ is } T \\
 & \text{iff for all } t, (Et \rightarrow Pt) \in w \\
 & \text{iff } \forall x Px \in w.
 \end{array}$$

The proof that (1) and (2) are equivalent requires the proof of (SL) (for Substitution Lemma).

$$(SL) \quad a(a(t)(w)/x)(Px)(w) \text{ is } a(Pt)(w).$$

Unfortunately, (SL) is not always true if  $t$  is non-rigid. It is false, for example, for  $Pt = \Box Ft$  on the following model. The set of worlds  $W$  contains (the real) world  $r$ , and (an unreal) world  $u$ , and they are both accessible from themselves and each other. The domain  $D$  contains two objects  $d$ , for (David Lewis) and  $s$  (for Saul Kripke). The term  $t$  (read ‘the author of “Counterpart Theory”’) has  $d$  as its extension in the real world, and  $s$  as its extension in the unreal world  $u$ . The extension of  $F$  (read ‘is author of “Counterpart Theory”’) contains  $d$  in  $r$ , and  $s$  in  $u$ . Notice now that  $a(a(t)(u)/x)(\Box Fx)(u)$  is  $a(s/x)(\Box Fx)(u)$ , which is false, since  $s$  is not in the extension of  $F$  in both worlds. However  $a(\Box Ft)(u)$  is true because the extension of  $t$  is in the extension of  $F$  in each world. We see that (SL) fails for reasons closely related to the fact that substitution of identities fails for non-rigid terms.

We did not face this problem for systems with rigid terms, because (SL) is true when  $a(t)$  is a constant function. The problem did not arise with the substantial interpretation because there the lemma we need (SSL) concerns substitution of intensions and is readily proven.

$$(SSL) \quad a(a(t)/x)(Px)(w) \text{ is } a(Pt)(w).$$

Thomason tackles the problem posed by the failure of (SL) in a direct way. He stipulates that variables are rigid designators and uses variables, not terms, to fix the domains of his standard model. The extension  $a(t)(w)$  is set to  $\{x : x = t \in w\}$ , and the domain  $D(w)$  contains the extensions of all terms  $t$  such that  $Et \in w$ . By adding the rules (RV), to the system, he can ensure that the standard model has rigid variables, using the methods we outlined in Section 2.2.4.2.



$$(RV) \quad \frac{x = y}{\Box x = y} \quad \frac{\neg x = y}{\Box \neg x = y} \quad \frac{x = y}{Ex \rightarrow Ey}$$

However, the use of rigid variables leads to further complications. In order to establish the case for identity in (TL), we need to know that if  $a(t)(w)$  is  $a(t')(w)$  then  $t = t' \in w$ . The identity of  $a(t)(w)$  only establishes that  $x = t \in w$  iff  $x = t' \in w$ , for all variables  $x$ . To show that  $t = t' \in w$ , we need to know that there is some variable  $y$  such that  $y = t \in w$ . This requires us to restrict the set  $W$  of possible worlds of our model to those that meet condition (V).

(V) For all  $w$  in  $W$ , and all terms  $t$  of  $L$ , there is a  $y$  such that  $y = t \in w$ .

In order to meet condition (V) when it comes time to extend  $w^*$  to a set in  $W$ , Thomason added the following rule to this system.

$$(G=) \quad \frac{\vdash G(\neg y = t)}{\vdash G(p \wedge \neg p)}$$

The rule (G=) ensures that we can consistently add a sentence of the form  $y = t$  for each of the terms  $t$  during the construction of a saturated set, and to do so without extending the language.

The system **Q3** which we can show to be complete using this method is composed of **GS**, (RV), and (G=).

The system Thomason [1970] showed to be complete lacked the primitive existence predicate  $E$ , and was built on **S4**. In **S4**, the sentence  $\exists x \Box x = t$  is true in the standard model just in case the intension of  $t$  is rigid. Also, the replacement of  $Et$  with  $\exists x \Box x = t$  in the rules of free logic results in valid quantifier rules. It follows that if **S** is **S4** or stronger, we can formulate a complete system for **Q3-S** without a primitive existence predicate by replacing  $Et$  with  $\exists x \Box x = t$  in the rules of **Q3-S**.

### 3 UNAXIOMATISABILITY OF SOME QUANTIFIED INTENSIONAL LOGICS

#### 3.1 Introduction

Certain quantified modal languages are capable of expressing statements of arithmetic. These systems cannot be axiomatised, for if they were, they would be adequate for arithmetic, which is impossible by Gödel's Theorem. In this section we will give examples of three quantified modal logics which are incomplete for this reason. First, we review Scott's result (reported in [Kamp, 1977]) that predicate tense logic is incomplete if time is described by the reals. Next we will discuss unaxiomatisability results [Fine, 1970] for propositional modal logics with quantifiers over propositional variables.

Finally, we will show that **Q2** cannot be formalised, at least not if the underlying modal logic is **S4.3** or weaker. (This is Kripke's result reported in [Kamp, 1977].) The rest of this section contains preliminary material which we need later. A reader with a background in mathematical logic will probably want to skip to Section 3.2.

### 3.1.1 Languages that express arithmetic

The language **PA** (for Peano Arithmetic) contains quantifiers, =, a constant 0, and function symbols ', +, ·. A model  $\langle D, a \rangle$  of **PA** consists of a non-empty domain  $D$  (of quantification), and an assignment function  $a$  that assigns to ' a unary function  $a(')$  from  $D$  to  $D$ , and to both + and ·, binary functions  $a(+)$  and  $a(\cdot)$  from  $D \times D$  to  $D$ . A model is the standard model of arithmetic iff  $D$  is the set of integers  $0, 1, 2, \dots$ ,  $a(')$  is the function that takes each integer into its successor,  $a(+)$  is the addition function, and  $a(\cdot)$  is multiplication.

Now suppose we have a language  $L$  which includes the symbols of **PA** and which contains a sentence  $SMA$  which is true on a model just in case it is the standard model of arithmetic. It follows that the valid sentences of  $L$  cannot be formalised. The reason is that the sentence  $A$  of arithmetic is true on the standard model just in case  $SMA \rightarrow A$  is a valid sentence of  $L$ . So any axiomatisation of  $L$  would provide a way to formalise the true sentences of arithmetic, and this, Gödel showed, cannot be done.

There is no need for  $SMA$  to pick out the standard model exactly. (In fact, it cannot.) It is easy to see that the same sentences are true on any pair of isomorphic models. So  $L$  will be unaxiomatisable as long as it contains a sentence  $SMA$  which is true only on models of  $PA$  that are isomorphic to the standard model. (To avoid talking all the time of isomorphic models, we will mean by a 'standard model' any model isomorphic to the standard one.)

We do not need 0, ', + and · in the language in order to obtain this kind of incompleteness result. It is well known that constants and function symbols are eliminable in favour of corresponding predicate letters. For example, we may introduce the predicate  $Z$  for zero, and the sentence  $\exists!xZx$  which ensures that the extension of  $Z$  is a singleton. (We use  $\exists!xPx$  to abbreviate  $\exists x(Px \wedge \forall y(Py \rightarrow x = y))$ , where  $y$  is chosen new to  $Px$ .) We may then conjoin  $\exists!xZx$  to  $SMA$ , and replace each sentence  $P0$  of  $SMA$  involving 0, with  $\forall x(Zx \rightarrow Px)$ , which says the same thing. To eliminate ', we introduce a binary predicate letter  $N$ , and we add  $\exists!yNxy$  to ensure that the extension of  $N$  is a unary function. We then replace axioms  $Px'$  involving ', with  $\forall y(Nxy \rightarrow Py)$ . By introducing ternary predicates, for + and ·, and performing the same manoeuvre, we can complete the elimination of function symbols. It follows that any language which contains first-order logic with identity and contains a sentence  $SMA$  which is true only on a

standard model is incomplete, (if it is consistent). (In the case of a language that uses predicate letters,  $Z, N, T, P$  for arithmetic,  $\langle D, a \rangle$  counts as a *standard model* iff  $a$  is a function over these predicate letters which assigns them extensions, and  $\langle D, a \rangle$  is isomorphic to another model of the same kind whose domain is the integers and which gives  $Z, N, T, P$  the extension zero, the successor function, plus, and times.)

### 3.2 Incompleteness of Predicate Tense Logic with Real Time

It is crucial in physics that we represent moments of time using numbers. If time is atomic, and there is a first moment, then the set of times looks like the integers of the standard model of arithmetic. We are more likely to think of time as dense, and so represent it using the rationals, or the reals.

Scott showed that if time is mathematical in any of these senses, then predicate tense logic is incomplete. (The result is reported in [Kamp, 1977].) When we assume that the Kripke frame  $\langle W, R \rangle$  of any tense logic model  $\langle W, R, D, a \rangle$  is such that  $W$  is the set of integers, and  $R$  the relation ‘less than’, then we can find a sentence  $SMA$  which is true only on standard models. Even when we consider frames  $\langle W, R \rangle$  where  $W$  is the set of rationals or reals, the same argument can be constructed.

#### 3.2.1 Syntax and Semantics of Predicate Tense Logic

Let us define **T1** (tense predicate logic like **Q1**) in the following way. *The syntax of T1* involves an alphabet which includes symbols of first-order logic, and two sentential operators  $G$  and  $H$  (read ‘it will always be that’ and ‘it was always the case that’). The more familiar operators  $F$  and  $P$  (read ‘it will be that’ and ‘it was the case that’) are defined by  $F =_{df} \neg G \neg$ , and  $P =_{df} \neg H \neg$ .

To formulate the semantics of **T1** let us define a **T1-model** as a sequence  $\langle W, R, D, a \rangle$ , where  $\langle W, R \rangle$  is like the integers in that sense that  $W$  is the set consisting of  $0, 1, 2, \dots$ , and  $R$  is ‘less than’. The quantifier of **T1** is interpreted with a fixed domain  $D$ , so its truth clause is (**Q1**).

(**Q1**)  $a(\forall x P x)(w)$  is  $T$  iff for all  $d$  in  $D$ ,  $a(d/x)(P x)(w)$  is  $T$ .

The truth clauses for  $G$  and  $H$  read as follows.

( $G$ )  $a(GA)(w)$  is  $T$  iff if  $wRw'$ , then  $a(A)(w')$  is  $T$ .

( $H$ )  $a(HA)(w)$  is  $T$  iff if  $w'Rw$ , then  $a(A)(w')$  is  $T$ .

For the moment, we will assume that terms are all rigid designators, so  $a(t)(w)$  is  $a(t)(w')$  for all  $w, w'$  in  $W$ . This restriction can be relaxed without changing the essentials of the incompleteness proof. Notice, then, that semantics for **T1** is exactly like **Q1**, except that in **T1** we have two intensional operators.

3.2.2 *The Expressive Capabilities of T1*

If we had quantifiers and predicate letters in **T1** whose domain were the set  $W$  of times, then the unaxiomatisability of **T1** would be easy to show. In that case, sentences valid in **T1** would be those that are valid on all frames  $\langle W, R \rangle$  where  $W$  is the integers. We could then construct the sentence  $Q$  consisting of the axioms of (first-order) arithmetic using predicate letters  $Z, N, P, T$ . (See [Boolos and Jeffrey, 1989, p. 161] for these axioms.) Sentence  $Q$  would serve as the sentence which expresses that a model is standard.

Our problem is, however, that  $W$  is not the domain of quantification in **T1**. The quantifiers range instead, over the domain  $D$  of objects. Nevertheless, it is possible to find a sentence of **T1** that sets up a correspondence between members of  $W$  and members of  $D$  so that sentences that express properties of the domain  $D$  reflect corresponding properties in the set of worlds  $W$ . In order to show how this correspondence is brought about, let us first give a few definitions and facts concerning the things that **T1** can express.

First, we will define two operators  $A$ , and  $S$  (read ‘it is always the case that’ and ‘it is sometimes the case that’) as follows.

$$AA = A \wedge GA \wedge HA, \quad SA = A \vee FA \vee PA.$$

Since  $W$  in every model of **T1** is the set of integers, it is easy to verify the following facts about all models of **T1**.

FACT 1.  $AA$  is true at  $w$  iff  $A$  is true at every time  $w'$  in  $W$ .

FACT 2.  $SA$  is true at  $w$  iff  $A$  is true at some time  $w'$  in  $W$ .

Now let us introduce the predicate letter  $E$  (read ‘exists’). We will use the following two sentences to ensure that every member of  $D$  is in the extension of  $E$  at some time, and that the extension of  $E$  is always either a singleton or empty.

(F1)  $\forall xS(Ex \wedge H\neg Ex \wedge G\neg Ex)$   
(Everything exists at exactly one time.)

(F2)  $A\forall x\forall y((Ex \wedge Ey) \rightarrow x = y)$   
(No two things exist at the same time.)

Any model that makes both of these sentences true sets up a function from  $D$  into  $W$ , because for each member  $d$  of  $D$ , we know there is exactly one integer  $t_d$  of  $W$  at which  $d$  exists.

Now let us introduce the following abbreviation.

( $<$ )  $x < y =_{df} S(Ex \wedge FEy)$ .

The sentence  $S(Ex \wedge FEy)$  is true at  $t$  for  $a$  just in case there is some time where  $a(x)$  exists, and a later time where  $a(y)$  exists. Since (F1) guarantees that an object exists at only one time, it follows that the pair  $d, d'$  satisfies the extension of  $<$  at any time just in case the integer  $t_d$  where  $d$  exists is less than the integer  $t_{d'}$  where  $d'$  exists. So  $<$  sets up an ordering on  $D$  that corresponds to the relation ‘less than’ on the integers. Actually,  $<$  does not express all the facts about ‘less than’ on the integers, because (F1) and (F2) do not guarantee that something exists at every time. The extension of  $<$  corresponds to ‘less than’ restricted to  $WD$  the set of those times when objects exist. We could set up a one-one correspondence between  $W$  and  $D$  by adding the sentence  $A\exists xEx$ , but this will block the proof for the case of the rationals and the reals, as we will see.

### 3.2.3 Unaxiomatisability of **T1**

Now let us introduce an equivalence that fixes the extension of predicate  $N$  as the successor function, and guarantees that every object in the ordering set up by  $<$  has a successor.

$$(F3) \quad xNy \leftrightarrow x < y \wedge \forall z((\neg z = y \wedge x < z) \rightarrow y < z),$$

$$(F4) \quad \exists y(Ey \wedge xNy).$$

If (F3) and (F4) are both true at any time  $t$  of  $W$ , then the pair  $\langle d, d' \rangle$  is in the extension of  $N$  at  $t$  just in case the corresponding times  $t_d, t_{d'}$  are such that  $t_{d'}$  is the successor of  $t_d$  in  $WD$  (the set of times where objects exist).

We may also define  $Z$  (read ‘is zero’), and guarantee that zero exists as follows.

$$(F5) \quad Zx \leftrightarrow \forall y \neg y < x$$

$$(F6) \quad \exists x(Ex \wedge Zx).$$

These two sentences ensure that there is a least member  $t_0$  in the set  $WD$  of times at which objects exist.

Let  $SMA$  be the conjunction of (F1)–(F6) and  $Q^*$ , the result of eliminating  $0, ', +$  and  $\cdot$  in favour of predicate letters  $Z, N, P$  and  $T$  in the axioms  $Q$  of first-order arithmetic. We claim that  $SMA$  expresses that a model is standard in the following sense.

**LEMMA 5.** *For any model  $\langle W, R, D, a \rangle$  of **T1**, and any  $w$  in  $W$ ,  $a(SMA)(w)$  is  $T$  only if  $\langle D, a(w) \rangle$  is standard.*

Here  $a(w)$  is the function that gives to each predicate letter  $F$ , the extension  $a(F)(w)$ , that  $F$  receives on  $a$  at world  $w$ .

**Proof.** Let us imagine a model  $\langle W, R, D, a \rangle$  that satisfies *SMA* at  $w$ . As we said, the truth of (F1) and (F2) sets up a correspondence between objects of  $D$  and a subset  $WD$  of  $W$ . (F3) ensures that a pair of objects  $\langle d, d' \rangle$  is in the extension of  $<$  just in case the corresponding numbers  $t_d, t_{d'}$  bear the relation 'less than'. (F4) and (F5) ensure that there is a successor for any number  $t_i$  of the series, and (F6) ensures that there is a least member  $t_0$  in the series. It follows from this that the objects of  $D$  correspond to a sequence  $w_0, w_1, w_2, w_3, \dots$  of numbers of  $W$  ordered by 'less than', with a first member  $w_0$ . Furthermore, the extension of  $N$  picks out the successor function on this ordering. Given that the sentence  $Q$  is satisfied, we also know the extensions of  $P$  and  $T$  at  $w$  must be plus and times. It follows then that  $\langle D, a(w) \rangle$  is a standard model of arithmetic. ■

Let us suppose that  $A$  is any sentence of arithmetic, and that  $A^*$  is the result of eliminating  $0, ', +, \cdot$  in favour of predicate letters  $Z, N, P, T$  in the usual way. We may now show that **T1** is incomplete on the basis of the following theorem.

**THEOREM 1.** *SMA*  $\rightarrow A^*$  is **T1**-valid iff  $A$  is true on the standard model of arithmetic.

**Proof.** (left to right) Let  $\langle W, R, D, a \rangle$  be a **T1**-model such that  $W$  is the integers (rationals, reals),  $R$  is the ordering 'less than' on  $W$ ,  $D$  is the integers, and  $d \in a(E)(w)$  iff  $d = w$ , and  $a(E)(w)$  is the empty set if  $w$  is not an integer. On this model  $a(SMA)(w)$  is  $T$ , for any  $w$  in  $W$ . By the **T1**-validity of  $SMA \rightarrow A^*$ , it follows that  $a(A^*)(w)$  is  $T$ . Notice that  $A^*$  contains no intensional operators, and so its value is determined by the extensions of  $Z, N, P, T$  exactly as it would be on the extensional model  $\langle D, a(w) \rangle$ . Since this is a standard model,  $A$  must be true on the standard model of arithmetic.

(right to left) Suppose that  $A$  is true on the standard model for arithmetic and suppose that  $a(SMA)(w)$  is  $T$  at any  $w$  in  $W$ , on a model of **T1**. By Lemma 5,  $\langle D, a(w) \rangle$  is standard, and so  $a(A^*)(w)$  is  $T$ . We conclude that  $SMA \rightarrow A^*$  is **T1**-valid. ■

Notice that Theorem 1 can be proved as long as we begin with any frame  $\langle W, R \rangle$  which contains a substructure which is isomorphic to the integers ordered by 'less than'. It follows that **T1** is incomplete with respect to the integers, the rationals, the reals, and virtually any other conceivable numerical account of time. It is also clear that the same argument works for logics that have world-relative domains. In this case, we are already supplied with a primitive predicate  $E$  which picks out things that exist at a given time, and the same argument can be carried out for this  $E$ .

### 3.3 Incompleteness of Second-Order Propositional Modal Logics

Fine [1970] shows that second-order propositional modal logic (**SOPML**) is incomplete when the modality is **S4.2** or less. **SOPML** is ordinary propositional modal logic, except we introduce quantifiers which bind the propositional variables. Here, we will give an incompleteness result for a somewhat different system called **SOMA** (for Second-Order Modal Arithmetic), assuming that the modality is **S4.3** or less. **SOMA** is **SOPML** supplemented with a propositional constant  $0$ , and connectives  $'$ ,  $+$ , and  $\cdot$ , of arities 1, 2, 2 respectively. The unaxiomatisability of **SOMA** is easier to prove than it is for **SOPML**, because **SOMA** already contains the notation for arithmetic. In the case of **SOPML**, we need to show how to get the effect of the binary function symbols  $+$  and  $\cdot$ . The proof of **SOMA** allows us to display the main strategy used in the proof for **SOPML**, without having to cover this less central detail. Another reason for concentrating on **SOMA** is that its relatively easy incompleteness result provides a quick proof of the incompleteness of **Q2**, which we give in Section 3.4.

#### 3.3.1 The Intuitions behind the Proof

We know that a system is incomplete as long as it contains a sentence that expresses that its models contain a standard model of arithmetic. It is well known that any model of both the axioms of first-order arithmetic **Q**, and (MI) (the second-order axiom of mathematical induction) is a standard model. (See, for example [Boolos and Jeffrey, 1989, Ch. 18].)

$$(MI) \quad \forall P((P0 \wedge \forall x(Px \rightarrow Px')) \rightarrow \forall xPx).$$

**Q** can be expressed in first-order logic; however, (MI) requires quantification over a monadic predicate letter. So any extension of first-order logic that can achieve the effect of quantification over monadic predicates will express arithmetic, and so be incomplete.

The idea behind the incompleteness proof for **SOMA**, then, is to show that quantification over propositional variables in modal logic can be used to get the effect of both quantification over worlds and quantification over predicates of worlds. To see how this is done, think of the intensions of propositional variables as truth sets. (The truth set of a propositional variable is just the set of worlds where it is true.) So quantification over propositional variables amounts to quantification over truth sets, that is, over properties of worlds. We also need to be able to quantify over objects of  $W$  if we are to express the axioms of arithmetic. This is done in **SOMA** by finding a way to say that an intension contains a single world. Then quantification over all singleton sets of worlds amounts to quantification over the individual worlds themselves.

3.3.2 *The Expressive Resources of SOMA*

We turn now to the details involved in showing that **SOMA** can express arithmetic. To fix our later discussion, we will give the semantics for **SOMA** here. A model  $\langle W, R, a \rangle$  of **SOMA** assigns to each propositional variable  $p$  a subset  $a(p)$  of  $W$ , called the truth set of  $p$ . It also assigns to  $'$ ,  $+$ , and  $\cdot$ , functions that take us from subsets of  $W$  into new subsets of  $W$  in the case of  $'$  and from subsets of  $W \times W$  to  $W$  in the case of  $+$  and  $\cdot$ . Since we are proving incompleteness for **SOMA-S4.3**, we will assume that  $\langle D, R \rangle$  is reflexive, transitive and connected.

The clauses for sentences with shapes  $A'$ ,  $A + B$ , and  $A \cdot B$  are as follows

$$\begin{aligned} a(A') &= a(')(a(A)) \\ a(A + B) &= a(+)(a(A), a(B)) \\ a(A \cdot B) &= a(\cdot)(a(A), a(B)) \end{aligned}$$

and the clauses for  $\neg$ ,  $\rightarrow$ , and  $\Box$  are given in the usual way.

For the quantifier we have the following.

$$(\forall p) \quad w \in a(\forall p P p) \text{ iff for every subset } s \text{ of } W, w \in a(s/p)(P p)$$

Here  $a(s/p)$  is the assignment just like  $a$  save that  $a(s/p)(p)$  is  $s$ .

Most of the properties we can express in **SOMA** only apply to a portion of the model  $\langle W, R, a \rangle$ . It will turn out, however, that this is enough for our purposes. We define the future of model  $\langle W, R, a \rangle$  at  $w$ , as the model  $\langle Wf, Rf, af \rangle$ , where  $Wf$  is  $\{w' : wRw'\}$  and  $Rf$  and  $af$  are  $R$  and  $a$  restricted to  $Wf$ . The future at  $w$ , then, contains just those worlds accessible from  $w$ , and the portions of  $R$  and  $a$  which concern these worlds. For convenience, we will also call  $\langle Wf, af \rangle$  the future of  $\langle W, R, a \rangle$  at  $w$ , where in this case,  $af(')$ ,  $af(+)$  and  $af(\cdot)$  are restricted to singleton sets of  $Wf$ . We are going to show how to construct a sentence which is true at  $w$  just in case the future  $\langle Wf, af \rangle$  at  $w$  is a standard model of arithmetic.

In order to get the effect of quantification over objects of  $W$ , let us present a sentence  $I_p$  of **SOMA** whose truth at  $w$  ensures that the intension of  $p$  is a singleton in the future at  $w$ .

$$(I) \quad I_p =_{af} \Diamond p \wedge \forall q (\Box(p \rightarrow q) \vee \Box(p \rightarrow \neg q)).$$

There is an interesting intuition behind this definition.  $I_p$  says that the sentences entailed by  $p$  form a maximally consistent set, for the first conjunct says that  $p$  is consistent, and the second, that  $p$  entails any sentence or its negation. Given that worlds are maximally consistent sets, it follows that  $p$  could only be true at one world.

Let us show now that  $I_p$  actually has this intended effect. The first conjunct ensures that there is a world accessible from  $w$  where  $p$  is true, so we know that  $af(p)$  contains at least one member of  $Wf$ . Notice next that



$\Box(p \rightarrow q)$  ensures that every world accessible from  $w \in af(p)$  only if it is in  $af(q)$ , and so that  $af(p)$  is a subset of  $af(q)$ . The effect of the second conjunct of  $Ip$ , then, is to ensure that for any subset of  $Wf$  we choose,  $af(p)$  will be a subset of either it or its complement. It follows that  $af(p)$  must be a singleton, for suppose there were two distinct members  $w, w'$  of  $af(p)$ . Then the subset  $s$  such that  $w \in s$  and  $w' \notin s$  would have to include all the members of  $af(p)$ , or its complement would. In either case, one of  $w, w'$  would have to be missing from  $af(p)$ .

Since we have a way to express uniqueness of an intension, we can quantify both over worlds and their properties. In order to enforce the structure of the standard model on the future of  $w$ , we will need to express properties about the relation  $R$ . The abbreviation  $(\leq)$  shows how to do this.

$$(\leq) \quad p \leq q =_{df} \Box(p \rightarrow \Diamond q)$$

It is a simple matter to verify that whenever the intensions of  $p$  and  $q$  are singletons in the future of  $w$ , then  $p \leq q$  is true at  $w$  just in case  $w_p R w_q$ , where  $w_p$  and  $w_q$  are the worlds at which  $p$  and  $q$  are true in the future of  $w$ .

Now let us define identity in **SOMA**.

$$(=) \quad p = q =_{df} \Box(p \leftrightarrow q).$$

It is easy to see that  $\Box(p \leftrightarrow q)$  is true at  $w$ , just in case  $a(p)$  and  $a(q)$  agree on members in  $Wf$ . So  $p = q$  is true at  $w$  iff  $af(p)$  is  $af(q)$ .

We may define ' $<$ ' from ' $\leq$ ' in the usual way.

$$(<) \quad p < q =_{df} p \leq q \wedge \neg p = q.$$

In order to write (MI) in **SOMA**, we need a way to express that a world has a property. We know that  $\Box(p \rightarrow q)$  is true at  $w$  just in case the truth set of  $p$  is a subset of the truth set of  $q$  in the future of  $w$ . So in case the intension of  $p$  is a singleton containing  $w_p$ ,  $\Box(p \rightarrow q)$  says that the world  $w_p$  is in the intension of  $q$ . Therefore, we will adopt the following abbreviation.

$$(\text{is}) \quad p \text{ is } q =_{df} \Box(p \rightarrow q).$$

### 3.3.3 Incompleteness of **SOMA**

Now we are ready to formulate arithmetic in **SOMA**. Let  $SMA$  be the conjunction of the following sentences.

1.  $IO \wedge 0$
2.  $\forall p Ip'$
3.  $\forall p \forall q ((Ip \wedge Iq) \rightarrow (p' = q' \rightarrow p = q))$

4.  $\forall q(\mathbf{I}q \rightarrow \neg q < 0)$
  5.  $\forall p\forall q((\mathbf{I}p \wedge \mathbf{I}q) \rightarrow (p' = q \rightarrow (p < q \wedge \forall r((p < r \wedge \neg r = q) \rightarrow q < r)))$
  6.  $\forall p\forall q\mathbf{I}(p + q)$
  7.  $\forall p\forall q\mathbf{I}(p \cdot q)$
  8.  $\forall p(\mathbf{I}p \rightarrow p + 0 = p)$
  9.  $\forall p\forall q((\mathbf{I}p \wedge \mathbf{I}q) \rightarrow p + q' = (p + q)')$
  10.  $\forall p(\mathbf{I}p \rightarrow p \cdot 0 = p)$
  11.  $\forall p\forall q((\mathbf{I}p \wedge \mathbf{I}q) \rightarrow p \cdot q' = (p \cdot q) + p)$
- (MII)  $\forall p((0 \text{ is } p \wedge \forall q(\mathbf{I}q \rightarrow (q \text{ is } p \rightarrow q' \text{ is } p))) \rightarrow \forall r(\mathbf{I}r \rightarrow r \text{ is } p))$ .

Sentences (3)–(5) establish the proper relationship between zero, the successor function and the relation  $Rf$  which is expressed by  $\leq$ . Sentences (8)–(11) are the axioms of  $\mathbf{Q}$ , with propositional quantifiers restricted to  $\mathbf{I}$ . (MII) is our formulation of the second-order axiom of mathematical induction.

Our next task is to convince you that  $SMA$  is true at  $w$  just in case the future at  $w$  contains a standard model of arithmetic, in the sense of the following lemma.

**LEMMA 6.**  *$a(SMA)(w)$  is  $T$  on  $\langle W, D, a \rangle$  iff the future  $\langle Wf, Rf, af \rangle$  of  $\langle W, D, a \rangle$  at  $w$  is such that  $\langle Wf, af \rangle$  is a standard model of arithmetic and  $af(0)$  is a singleton containing  $w$ , and  $Rf$  is ‘less than or equal to’ on  $Wf$ .*

**Proof.** (left to right) Let a numeral be a sentence  $0_i$  composed of the propositional constant  $0$ , followed by  $i$  primes ( $'$ ). Sentences (1) and (2) of  $SMA$  guarantee that the intension of any numeral is a singleton set in the future of  $w$ . The second conjunct of (1) establishes that  $w$  is in  $a(0)$ , and so  $af(0)$  is a singleton containing  $w$ . Now let  $w_i$  be the singleton which is in the extension of numeral  $0_i$ . Sentence (3) guarantees that  $w_i$  is not  $w_j$  for  $i$  not equal to  $j$ , and so  $w_i$  is not  $w_{i+1}$ . (MII) ensures that every member of  $Wf$  is  $w_i$  for some numeral  $0_i$ , and so there is a one-one mapping between numerals and  $Wf$ . By sentences (4) and (5), we know that the  $w_i$  form a sequence such that  $w_0 Rf w_1, w_1 Rf w_2 \dots$ , with  $w_0$  as the least member. By the reflexivity and transitivity of  $Rf$  and (5) we may show further that  $i$  is less than or equal to  $j$  iff  $w_i Rf w_j$ . So  $Rf$  is ‘less than or equal to’. Sentences (6) and (7) ensure that the intensions of  $(p + q)$  and  $(p \cdot q)$  are functions that range over singletons only. The remaining sentences (8)–(11) ensure that  $af(+)$  and  $af(\cdot)$  correspond to addition and multiplication. We conclude that  $\langle Wf, af \rangle$  is indeed a standard model of arithmetic, and  $af(0)$  is indeed a singleton containing zero.

(right to left) Suppose that  $\langle Wf, Rf, af \rangle$  is the future of  $\langle W, R, a \rangle$  at  $w$ , and suppose that  $\langle Wf, af \rangle$  is a standard model. Suppose that  $af(0)$  is a singleton containing  $w$ , and that  $Rf$  is ‘less than or equal to’. Since  $af$  is standard, we know that  $af(0) = \{w\}$  contains the representative of zero on this model. It follows that  $wRfw'$  for every  $w'$  in  $W$ . Therefore, the future of  $\langle W, R, a \rangle$  at  $w$  just is  $\langle W, R, a \rangle$ . Now the reader can verify that sentences (1)–(11) and (MII) are all true at  $w$  on  $\langle W, R, a \rangle$ , and so  $a(SMA)(w)$  is  $T$ . This completes the proof of Lemma 6. ■

Now we must check that sentences of arithmetic are true just in case their translations into **SOMA** are true in the future of a given  $w$ . We define  $A^*$  for sentences  $A$  of arithmetic as the result of replacing variables and quantifiers of  $A$  with prepositional quantifiers restricted to  $I$ , and replacing identity in  $A$  with the corresponding sentence of **SOMA** according to definition (=).

LEMMA 7. *If  $\langle Wf, af \rangle$  is the future of  $\langle W, R, a \rangle$  at  $w$ , and  $\langle Wf, af \rangle$  is a standard model of arithmetic, then  $A$  is true on the standard model iff  $w \in a(A^*)$  on  $\langle W, R, a \rangle$ .*

**Proof.** The proof is by induction on the structure of  $A$ . The non-trivial cases occur when  $A$  has the shapes  $\forall xPx$ , and  $t = s$ . The case for the quantifier runs as follows.

$\forall xPx$  is true in arithmetic.  
iff  $Pn$  is true on the standard model for every numeral  $n$   
iff  $w \in a(Pn)$  in the future of  $w$ , for every numeral  $n$   
iff  $w \in a(Pq)$  in the future of  $w$  for all  $q$  such that  $af(q)$  is a singleton  
iff  $w \in a(\forall p(Ip \rightarrow Pp))$ .

The case for identity requires first that we show that the extension of any term  $t$  in the standard model of arithmetic corresponds to  $af(t)$  in  $\langle Wf, af \rangle$ . This is easily shown by induction on the structure of  $t$ . The rest of this case proceeds as follows.

$t = s$  is true in arithmetic  
iff  $a'(t)$  is  $a'(s)$ , for the  $a'$  of the standard model  
iff  $af(t)$  is  $af(s)$  in the future of  $w$   
iff  $a(t)(w')$  is  $a(s)(w')$  for  $w'$  in  $Wf$   
iff  $w' \in a(t \leftrightarrow s)$  for  $w'$  in  $Wf$   
iff  $w \in a(\Box t \leftrightarrow s)$

This completes the proof of Lemma 7. ■

We are now ready to prove the incompleteness of **SOMA**, using the following theorem.

**THEOREM 2.** *A is true on the standard model of arithmetic iff  $SMA \rightarrow A^*$  is **S.3**-valid in **SOMA**.*

**Proof.** (left to right) Suppose that  $A$  is true on the standard model of arithmetic. Let  $\langle W, R, a \rangle$  be any **S4.3**-model of **SOMA**. Let  $w$  be any world in  $W$ , and suppose that  $w \in a(SMA)$ . It follows by Lemma 6 that  $\langle Wf, af \rangle$  is a standard model of arithmetic, where  $\langle Wf, Rf, af \rangle$  is the future of  $\langle W, R, a \rangle$  at  $w$ . It follows by Lemma 7 that  $w$  is in  $a(A^*)$ . We conclude that  $(SMA \rightarrow A^*)$  is **S4.3**-valid in **SOMA**.

(right to left) Suppose that  $(SMA \rightarrow A^*)$  is **S4.3**-valid in **SOMA**. Let  $\langle W, R, a \rangle$  be the **SOMA** model such that  $W$  is the integers,  $R$  is the relation ‘equal or less than’, the intension of 0 is the singleton containing zero, and the intensions of  $'$ ,  $+$ , and  $\cdot$  are the successor function, plus, and times, defined on singleton sets of members of  $W$ .  $\langle W, R \rangle$  is clearly reflexive, transitive and connected, so  $\langle W, R, a \rangle$  is and **S4.3**-model. By Lemma 6,  $0 \in a(SMA)$ , and so by the validity of  $(SMA \rightarrow A^*)$ ,  $0 \in a(A^*)$ . However, the future of  $\langle W, R, a \rangle$  at zero is a standard model of arithmetic, and so by Lemma 7,  $A$  is a true sentence of arithmetic. ■

Theorem 2 establishes that **SOMA** with modality of strength **S4.3** is incomplete. It follows also that **SOMA** is incomplete for all weaker logics (down to **K**). The reason is that when the following sentences of **SOMA** are true at  $w$  in a model of **SOMA**, then  $Rf$  in the future of  $w$  must be reflexive, transitive, and connected.

$$\begin{aligned} & \forall p(\Box p \rightarrow p) \\ & \forall p(\Box p \rightarrow \Box \Box p) \\ & \forall p \forall q((\Diamond p \wedge \Diamond q) \rightarrow (\Diamond(p \wedge q) \vee \Diamond(p \wedge \Diamond q) \vee \Diamond q \wedge \Diamond p)) \end{aligned}$$

So by adding these sentences to  $SMA$ , we may carry out the proof of Theorem 2 for any system weaker than **S4.3**.

### 3.4 Incompleteness of **Q2**

The proof of the incompleteness of **SOMA** can be used to show that **Q2** cannot be axiomatised as long as the propositional modal logic is **S4.3** or weaker. This is done by showing that there is a transformation  $*$  that takes us from sentences of **SOMA** to sentences of **Q2**, so that  $A$  is valid in **SOMA** just in case  $A^*$  is valid in **Q2**. It follows that since **SOMA** can express arithmetic for modalities **S4.3** or less, then so can **Q2**.

The idea behind the transformation is to mimic propositional variables  $p$  of **SOMA**, whose intensions amount to functions from  $W$  into  $(T, F)$ , using corresponding individual variables  $x_p$ , whose intensions take us from  $W$  to  $D$ . We will arbitrarily select a term  $t$  whose extension at a world  $w$  picks out the object of  $D$  which plays the role of the truth value  $T$  for that world.

Then we will represent that  $p$  is true at world  $w$ , using the **Q2** sentence  $x_p = t$ .

To simplify the proof, we will assume first that **Q2\*** is **Q2** with function symbols  $0, ', +, \cdot$ . These symbols can be eliminated later in favour of predicate letters of **Q2**. We define  $A^*$  for any sentence  $A$  of **SOMA**, as the **Q2\***-sentence that results from replacing each variable  $p$  of  $A$  with a corresponding variable  $x_p$  of **Q2**, and replacing each propositional variable  $p$  with the sentence  $x_p = t$ . (Of course, we must be sure that the  $x_p$  and  $t$  are distinct variables.)

Let us prove the following lemma about this transformation.

**LEMMA 8.** *If  $\langle W, R, a \rangle$  is a model of **SOMA**, and  $\langle W, R, D, \mathbf{Q2}, a' \rangle$  is a model of **Q2\***, and  $w' \in a(p)$  iff  $a'(x_p)(w') = a'(t)(w')$ , for all  $w'$  in  $W$ , and  $D(w')$  contains two members for all  $w'$  such that  $wRw'$ , then  $w \in a(A)$  iff  $a(A^*)(w)$  is  $T$ .*

**Proof.** The proof of Lemma 8 is straightforward induction on the structure of  $A$ . ■

Now we may show how to set up a correspondence between sentences of **SOMA** and **Q2\***, according to the following theorem.

**THEOREM 3.**  *$A$  is valid in **SOMA** iff  $\exists x \exists y \Box \neg x = y \rightarrow A^*$  is **Q2\***-valid.*

**Proof.** (left to right) Assume that  $A$  is valid in **SOMA**, and consider a **Q2\***-model  $\langle W, R, D, \mathbf{Q2}, a' \rangle$  such that  $\exists x \exists y \Box \neg x = y$  is true at any  $w$  in  $W$ . Then  $D(w')$  contains two objects for every  $w'$  such that  $wRw'$ . Build a **SOMA** model  $\langle W, R, a \rangle$  such that  $w'$  is in  $a(p)$  iff  $a'(x_p)(w')$  is  $a(t)(w')$ . We have met the conditions for Lemma 8, and so by the validity of  $A$  we conclude that  $a'(A^*)(w)$  is  $T$ .

(right to left) Now assume that  $\exists x \exists y \neg x = y \rightarrow A^*$  is **Q2**-valid. Let  $\langle W, R, a \rangle$  be any **SOMA**-model, and  $w$  any member of  $W$ . Now define a **Q2**-model  $\langle W, R, D, \mathbf{Q2}, a' \rangle$  as follows. Let  $D$  contain the objects  $T, F$ , and let  $D(w')$  be  $D$  for each  $w'$  in  $W$ . Let  $a'(t)(w')$  be  $T$  for all  $w'$  in  $W$ , and let  $a'(x_p)(w')$  be  $T$  if  $w'$  is in  $a(p)$ , and  $F$  otherwise. The value of  $a(\exists x \exists y \neg x = y)(w)$  is  $T$ , and so, by the validity of  $\exists x \exists y \neg x = y \rightarrow A^*$ ,  $a(A^*)(w)$  is  $T$ . Since the conditions for Lemma 8 are met, we conclude that  $w$  is in  $a(A)$ . ■

Theorem 3 establishes that **Q2\*** is incomplete for all modalities **S4.3** or weaker. It follows that **Q2** is also incomplete, because function symbols of sentences of **Q2\*** can be eliminated for corresponding predicate letters of **Q2**. The same sort of argument can be used to show the unaxiomatisability of **QC**, where we have a single domain of quantification. In fact, the proof is easier, since we need only the sentence  $\exists x \exists y \neg x = y$  to ensure that the domain contains two objects.

### 3.5 *Systems as Strong as S5*

It is interesting to ask whether these results apply to **Q2** and **SOMA** for modalities as strong as **S5**. The answer is that they do not. Kripke has shown (see [Kamp, 1977]) that **Q2S5** is axiomatisable, and Fine [1970] even shows that **SOPML-S5** is decidable. The reason that the proof strategy that we have used does not work for **S5** is that our method depends on our ability to build the structure of the standard model of arithmetic within the kinds of Kripke frames with which we were supplied. In the case of **S5**, however, the frames are equivalence classes, and there is no way to develop the ordering we need to construct the standard model.

## APPENDIX

*List of Rules*

	<i>Page</i>	
269	(FUI)	$\frac{\forall xPx}{Et \rightarrow Pt}$ for any term $t$
269	(FUG)	$\frac{\vdash A \rightarrow (Et \rightarrow Pt)}{\vdash A \rightarrow \forall xPx}$ where $t$ is any term not in $A \rightarrow \forall xPx$
270	(=In)	$t = t$
270	(=Out)	$\frac{t = t'}{Pt \rightarrow Pt'}$ where $Pt$ is an atom
273	(RT)	$\frac{t = t'}{\Box t = t'} \quad \frac{\neg t = t'}{\Box \neg t = t'}$
273	(BF)	$\forall x\Box A \rightarrow \Box\forall xA$
275	(CBF)	$\Box\forall xA \rightarrow \forall x\Box A$
278	(HUI)	$\frac{\forall xPx}{(\exists x\Box i x = t \wedge \dots \wedge \exists x\Box k x = t) \rightarrow Pt}$ where $i, \dots, k$ is a list of integers which records for each occurrence of $x$ in $Px$ , the number of boxes whose scope includes that occurrence.
280	(TUI)	$\frac{\forall xPx}{\exists x\Box x = t \rightarrow Pt}$
297	(GUI)	$\frac{G(\forall xPx)}{G(Et \rightarrow Pt)}$
297	(GUG)	$\frac{\vdash G(Et \rightarrow Pt)}{\vdash G(\forall xPx)}$ where $t$ does not appear in $G(\forall xPx)$ . Here $G(B)$ is any sentence with the shape $A_1 \rightarrow A_2 \neg\exists \dots \neg\exists A_i \neg\exists B$ and $G(C)$ is the result of replacing $C$ for $B$ in $G(B)$ .

Page

$$301 \quad (=E) \quad \frac{t = t'}{Et \rightarrow Et'}$$

$$302 \quad (E) \quad Et$$

$$303 \quad (RV) \quad \frac{x = y}{\Box x = y} \quad \frac{\neg x = y}{\Box \neg x = y} \quad \frac{x = y}{Ex \rightarrow Ey}$$

$$304 \quad (G=) \quad \frac{\vdash G(\neg y = t)}{\vdash G(p \wedge \neg p)}$$



*List of Systems*

- PL** = Predicate Logic, First-order Logic (without identity).  
**S** = an arbitrarily selected propositional modal logic as strong or stronger than **K**.
- 269    **MFL** = (FUI) + (FUG) Minimal Free Logic
- 270    **ID** = (=In) + (=Out) Intensional Identity Theory
- 273    **Q1** = **S** + **PL** + **ID** + (RT) + (BF).
- 274    **Q1R** = **S** + **MFL** + **ID** + (RT).
- 275    **QK** = **S** + **PL** with no terms. The necessitation rule is restricted to apply only to closed sentences.
- 277    **QPL** = **S** + **PL**.
- 282    **Q2** cannot be axiomatised, unless modality is as strong as **S5**.
- 284    **QS** = **S** + **MFL** + **ID**.
- 284    **B1-S5** = **S** + **PL** + (BF) + **ID** + axiom of substitution for strong identities.
- 298    **GS** = **S** + (GUI) + (GUG) + **ID**.
- 301    **GQ1R** = **GS** + (RT) + (=E).
- 302    **GQ1** = **GQ1R** + (E).
- 304    **Q3** = **S** + **GS** + (RV) + (G=).
- 304    **Q3-S4** = **S4** + **GS** with  $\exists x \Box x = t$  for  $Et$  + (RV) + (G=).  
 (Thomason's **Q3** [1970] also contains rules for descriptions, and the axiom  $\exists x \exists y x = y$  to guarantee that every domain contains at least one individual.)

*List of Conditions on Models*

A model has the form  $\langle W, R, D, Q, a \rangle$   
 $\langle W, R \rangle$  is the Kripke frame,  
 $D$  is the domain of possible objects,  
 $a$  is the assignment function, that gives terms and predicate letters  
their intensions over  $W$  and  $D$ ,  
 $Q$  is an item that details the nature of the quantifier domain(s).  
For free logic with primitive predicate  $E$  we have  $a(E)$  is  $Q$ .

The truth value  $a(A)(w)$  of formula  $A$  at world  $w$  is defined  
recursively by the following clauses.

( $\neg$ )  $a(\neg A)(w)$  is  $T$  iff  $a(A)(w)$  is not  $T$ ,  
( $\rightarrow$ )  $a(A \rightarrow B)(w)$  is  $T$  iff  $a(A)(w)$  is  $F$  or  $a(A)(w)$  is  $T$ ,  
( $\Box$ )  $a(\Box A)(w)$  is  $T$  iff if  $wRw'$  then  $a(A)(w')$  is  $T$   
and another clause for the quantifier which differs in different semantics.

To describe a semantics for quantifier modal logic, we give a  
description of  $Q$ , list any other conditions on the model and then  
give the truth clause for the quantifier.

273 **Q1**

A **Q1**-model has  $Q = \mathbf{Q1} = D$ , and meets (aRT).  
(aRT)  $a(t)(w)$  is  $a(t)(w')$  for all  $w, w'$  in  $W$ ,  
(**Q1**)  $a(\forall xA)(w)$  is  $T$  iff for all  $d$  in **Q1**,  $a(d/x)(A)(w)$  is  $T$ .

274 **Q1R**

A **Q1R**-model has  $Q = \mathbf{Q1R}$  a function that assigns subsets  $D(w)$  to  
the worlds  $w$  of  $W$ , and it meets (aRT).  
(**Q1R**)  $a(\forall xA)(w)$  is  $T$  iff for every  $d$  in  $D(w)$ ,  $a(d/x)(A)(w)$  is  $T$ .

277 **QPL**

A **QPL**-model is **Q1R**-model that meets (ND). Truth values are  
calculated using (TG) (truth value gaps).  
(ND) If  $wRw'$ , then  $D(w)$  is a subset of  $D(w')$ .  
(TG) If  $a(t)(w) \notin D(w)$ , then any sentence  $Pt$  containing  $t$  has no  
truth value.

277 **GK**

A **GK**-model is a **Q1R**-model. Truth values are calculated using (TG).  
For **GKc**, use clause ( $\Box c$ ) for  $\Box$ . For **GKs**, use ( $\Box s$ ).

( $\Box c$ )  $a(\Box A)(w)$  is  $T$  iff if  $wRw'$ , then  $A$  has a value at  $w$  and  $a(A)(w')$  is  $T$ .

( $\Box s$ )  $a(\Box A)(w)$  is  $T$  iff if  $wRw'$  and  $A$  has a value at  $w$ , then  $a(A)(w')$  is  $T$ .

278 **Q3**

A **Q3**-model has  $Q = \mathbf{Q1R}$  = a function that assigns a domain  $D(w)$  to each possible world. It need not meet condition (aRT). The quantifier clause is (**Q1R**).

280 **Q3L**

A **Q3L**-model is **Q3**-model that meets condition (L) (local terms).

(L)  $a(t)(w) \in D(w)$  for all  $w$  in  $W$ , and all terms  $t$ .

281 **QC**

A **QC**-model has  $Q = \mathbf{QC}$  = the set of all functions from  $W$  into  $D$ .

(**QC**)  $a(\forall xA)(w)$  is  $T$  iff for every  $f$  in **QC**,  $a(f/x)(A)(w)$  is  $T$ .

282 **Q2**

A **Q2**-model has  $Q = \mathbf{Q2} = \mathbf{Q1R}$  = a function that assigns a domain  $D(w)$  for each of the possible worlds  $w$ .

(**Q2**)  $a(\forall xA)(w)$  is  $T$  iff for every function  $f$  from  $W$  into  $D$ ,  $a(f/x)(A)(w)$  is  $T$ .

284 **QS**

A **QS**-model has  $Q = \mathbf{QS}$  = a function that assigns to each world  $w$  a subset  $S(w)$  of the set **QC** of all functions from  $W$  into  $D$ .

(**QS**)  $a(\forall xA)(w)$  is  $T$  iff for every member  $f$  of  $S(w)$ ,  $a(f/x)(A)(w)$  is  $T$ .

*Department of Philosophy, University of Houston, USA.*

## BIBLIOGRAPHY

- [Boolos and Jeffrey, 1989] G. Boolos and R. Jeffrey. *Computability and Logic*, 3rd edition. Cambridge University Press, (1st edition 1974), 1989.
- [Bowen, 1979] K. Bowen. *Model Theory for Modal Logic*. Reidel, Dordrecht, 1979.
- [Bressan, 1973] A. Bressan. *A General Interpreted Modal Calculus*. Yale University Press, 1973.
- [Carnap, 1947] R. Carnap. *Meaning and Necessity*. University of Chicago Press, 1947.
- [Cresswell, 1995] M. J. Cresswell. Incompleteness and the Barcan formula. *Journal of Philosophical Logic*, 24:379–403, 1995.

- [Fine, 1970] K. Fine. Propositional quantifiers in modal logic. *Theoria*, 36:336–346, 1970.
- [Gabbay, 1976] D. M. Gabbay. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. Reidel, Dordrecht, 1976.
- [Gallin, 1975] D. Gallin. *Intensional and Higher-Order Modal Logic*. North-Holland, Amsterdam, 1975.
- [Garson, 1978] J. Garson. Completeness of some quantified modal logics. *Logique et Analyse*, 21:153–164, 1978.
- [Henkin, 1949] L. Henkin. The completeness of the first-order functional calculus. *Journal of Symbolic Logic*, 14:159–166, 1949.
- [Hintikka, 1970] J. Hintikka. Existential and uniqueness presuppositions. In *Philosophical Problems in Logic*. D. Reidel, Dordrecht, 1970.
- [Hughes and Cresswell, 1968] G. Hughes and H. Cresswell. *An Introduction to Modal Logic*. Methuen, London, 1968.
- [Kamp, 1977] H. Kamp. Two related theorems by D. Scott and S. Kripke, 1977. Xerox.
- [Konyndyk, 1986] K. Konyndyk. *Introductory Modal Logic*. University of Notre Dame Press, Notre Dame, Indiana, 1986.
- [Kripke, 1963] S. Kripke. Semantical considerations in modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963.
- [Kripke, 1972] S. Kripke. Naming and necessity. In D. Davidson and G. Harman, editors, *Semantics of Natural Language*. Reidel, Dordrecht, 1972.
- [Lewis, 1968] D. Lewis. Counterpart theory and quantified modal logic. *Journal of Philosophy*, 65:113–126, 1968.
- [Menzel, 1991] C. Menzel. The true modal logic. *Journal of Philosophical Logic*, 20:331–374, 1991.
- [Parks, 1976] Z. Parks. Investigations into quantified modal logic. *Studia Logica*, 35:109–125, 1976.
- [Parsons, 1975] C. Parsons. On modal quantifier theory with contingent domains (abstract). *Journal of Symbolic Logic*, 40:302, 1975.
- [Thomason, 1969] R. Thomason. Modal logic and metaphysics. In K. Lambert, editor, *The Logical Way of Doing Things*. Yale University Press, 1969.
- [Thomason, 1970] R. Thomason. Some completeness results for modal predicate calculi. In K. Lambert, editor, *Philosophical Problems in Logic*. D. Reidel, Dordrecht, 1970.

#### EDITOR'S NOTE

The following recent book is of interest:

#### BIBLIOGRAPHY

- [Fitting and Mendelsohn, 1999] M. Fitting and R. L. Mendelsohn. *First-order Modal Logic*. Kluwer Academic Publishers, 1999.



## CORRESPONDENCE THEORY

## 1 INTRODUCTION TO THE SUBJECT

*Correspondences*

When possible worlds semantics arrived around 1960, one of its most charming features was the discovery of simple connections between existing intensional axioms and ordinary properties of the alternative relation among worlds. Decades of syntactic labour had produced a jungle of intensional axiomatic theories, for which a perspicuous semantic setting now became available. For instance, typical completeness theorems appeared such as the following:

A modal formula is a theorem of **S4** if and only if it is true in all *reflexive, transitive* Kripke frames.

Indeed, **S4** may also be shown to be the modal logic of the *partial orders*; which matches the most famous modal logic with perhaps the most basic type of classical relational structure. Such matchings extend to logics higher up in the **S4**-spectrum. For instance, **S4.2** with its additional axiom

$$\diamond\Box p \rightarrow \Box\diamond p$$

is complete with respect to those frames which are reflexive, transitive and *directed*, or *confluent*:

$$\forall xyz((Rxy \wedge Rxz) \rightarrow \exists u(Ryu \wedge Rzu))$$

Again, the latter condition is a ‘diamond property’ of classical frame.

Completeness results such as these have inspired a flourishing area of intensional *Completeness Theory*, witness the classic [Seegerberg, 1971]. It took modal logicians some time, however, to realise that there are also direct semantic equivalences involved here, having nothing to do with deduction in modal logics. Indeed, the whole present *Correspondence Theory* arose out of simple observations such as the following, made in the early seventies.

EXAMPLE 1. The *T*-axiom  $\Box p \rightarrow p$  is true in a Kripke frame  $\langle W, R \rangle$  if and only if *R* is reflexive.

Here, ‘true in a frame’ means true in all worlds, under all assignments to the proposition letters.

**Proof.** ‘ $\Rightarrow$ ’: Consider any  $w \in W$ . If  $\Box p \rightarrow p$  is true in  $\langle W, R \rangle$ , then, in particular, it is true at  $w$  under the assignment  $V$  with

$$V(p) = \{v \in W \mid Rww\}.$$

Thus,  $\Box p$  will be at  $w$  true by definition — and, hence, also  $p$ : i.e.  $Rww$ .

‘ $\Leftarrow$ ’: By reflexivity, truth at all  $R$ -alternatives implies actual truth. ■

EXAMPLE 2. The **S4**-axiom  $\Box p \rightarrow \Box\Box p$  is equivalent to transitivity.

**Proof.** By an analogous argument. ■

EXAMPLE 3. The **S4.2**-axiom  $\Diamond\Box p \rightarrow \Box\Diamond p$  defines directedness.

**Proof.** ‘ $\Rightarrow$ ’: Consider arbitrary  $w, v, u \in W$  such that  $Rwv, Rwu$ . Let the assignment  $V$  have

$$V(p) = \{s \in W \mid Rvs\}.$$

Immediately, this gives truth of  $\Box p$  at  $v$ . Therefore,  $\Diamond\Box p$  is true at  $w$ , whence  $\Box\Diamond p$  must hold as well. It follows that  $\Diamond p$  is true at  $u$ ; i.e.  $u$  has some  $R$ -successor in  $V(p)$  — whence  $v, u$  share a common  $R$ -successor.

‘ $\Leftarrow$ ’: If  $\Diamond\Box p$  is true at  $W$ , say because of some  $v$  with  $Rwv$  verifying  $\Box p$ , then  $\Diamond p$  will be true at all  $R$ -successors of  $w$ . For, all of these share at least one successor with  $v$ , by directedness. ■

Not all correspondences are equally simple. For instance, **S4.2** has a companion logic **S4.1** obtained by enriching **S4** with the ‘McKinsey Axiom’  $\Box\Diamond p \rightarrow \Diamond\Box p$ . This converse of the **S4.2** axiom turns out to be much more complex. A well-known completeness theorem says that **S4.1** axiomatises the modal theory of those Kripke frames which are reflexive, transitive as well as *atomic*:

$$\forall x\exists y(Rxy \wedge \forall z(Ryz \rightarrow z = y)).$$

(Notice that we need *identity* here, in addition to the predicate constant  $R$ .) We shall see later in Section 2.2 that the **S4.1** axioms together (just) manage to define the above threefold relational condition, but that the McKinsey Axiom does not define atomicity on its own (it is weaker). Indeed, this simple modal principle does not possess a first-order relational equivalent at all — a discovery made independently by several people around 1975.

### *Modal Formulas as Conditions on the Alternative Relation*

The general picture emerging here is that of modal axioms expressing certain ‘classical’ constraints on the alternative relation in frames where they are valid. With hindsight, this observation is hardly surprising. After all, given

some valuation, the clauses of the basic Kripke truth definition amount to a *translation* from modal formulas into classical ones involving  $R$ . Thus, e.g.,

$$\begin{aligned} \Box p \rightarrow p & \text{ becomes } \forall y(Rxy \rightarrow Py) \rightarrow Px \\ \Box p \rightarrow \Box \Box p & \text{ becomes } \forall y(Rxy \rightarrow Py) \rightarrow \\ & \rightarrow \forall y(Rxy \rightarrow \forall z(Ryz \rightarrow Pz)), \end{aligned}$$

while the McKinsey Axiom  $\Box \Diamond p \rightarrow \Diamond \Box p$  becomes

$$\forall y(Rxy \rightarrow \exists z(Ryz \wedge Pz)) \rightarrow \exists y(Rxy \wedge \forall z(Ryz \rightarrow Pz)).$$

Here the parameter ‘ $x$ ’ refers to the current world of evaluation, while unary predicate constants  $P$  ( $Q, \dots$ ) denote the sets of worlds where the corresponding proposition letter  $p$  ( $q, \dots$ ) holds.

Let us pause, to realise how, by this simple observation alone, many established results about classical predicate logic can be transferred straightaway to modal logic. For instance, for Kripke frames plus a fixed assignment (the modal ‘models’ of Section 2.1), *Compactness* and *Löwenheim–Skolem* results are immediate. If, e.g. a set of modal formulas is finitely satisfiable in Kripke models (given suitable assignments), then its classical transcription will be finitely satisfied too. Hence, by ordinary compactness, the latter set is simultaneously satisfied in some structure  $\langle W, R; P, Q, \dots \rangle$ : which forms a Kripke frame cum assignment verifying the original set.

But, this perspective is not quite the one we need.

In the evaluation of modal formulas according to the above truth definition, two factors are intermingled: the relational pattern of the worlds and the particular ‘facts’, i.e. the assignment. But the latter — the particular denotations of constants  $P, Q, \dots$  — is not relevant to the role of modal formulas as relational constraints. Indeed, these may even obscure the issue. When, e.g.  $V(p)$  equals  $W$ ,  $\Box p \rightarrow p$  holds in all worlds — but this observation is completely uninformative about the true content of this axiom (viz. reflexivity).

In order to arrive at the proper perspective, one simply abstracts from the effects of particular assignments, by means of a *universal* quantification over the unary predicates in the preceding translation. Thus, for instance,

$$\Box(p \vee q) \rightarrow (\Box p \vee \Box q)$$

now becomes

$$\forall P \forall Q (\forall y(Rxy \rightarrow (Py \vee Qy)) \rightarrow (\forall y(Rxy \rightarrow Py) \vee \forall y(Rxy \rightarrow Qy))).$$

Notice that modal formulas now get *second-order* transcriptions, as opposed to the earlier first-order ones.



The parameter ‘ $x$ ’ has remained: the present relational conditions are still ‘local’ in some actual world. A ‘global’ condition is obtained by performing one more universal quantification, this time with respect to this world parameter. The distinction is not without importance. The local version is more suitable for the original Kripke structures  $\langle W, R, w_0 \rangle$ , in which some ‘actual world’  $w_0$  figured prominently, as well as for ‘non-normal’ modal semantics, in which certain worlds are distinguished from others. The global reading is the more common one, however, which will be predominant in the sequel.

Again, the very point of view embodied in the above translation is significant — even though some of the earlier transfer phenomena are lost. What is lost, for instance, are most useful forms of compactness, as well as the Löwenheim–Skolem property. There is no automatic guarantee through second-order logic that, if a modal formula is true in some countable Kripke frame (i.e. under *all* valuations) it will be true in its countable elementary subframes (again, under all valuations). Still, this very phenomenon will be used to drive a wedge between ‘essentially first-order’ and ‘essentially second-order’ modal axioms in Section 2.2. Moreover, not all is lost. The above transcriptions are very simple second-order formulas, viz. so-called  $\Pi_1^1$ -sentences, with all second-order quantifiers occurring in a universal prefix in front of a first-order matrix. From classical logic, we still now a few things about  $\Pi_1^1$ -sentences, that will turn out useful. (Cf. the chapters on Higher Order Logic and Algorithms in Volume 1 of this *Handbook* for background.)

One such thing is involved in the following obvious question. In the light of earlier examples of correspondence, the present second-order transcriptions are exceedingly cumbersome. Compare, e.g. for the  $T$ -axiom  $\Box p \rightarrow p$ ,

$$\forall x Rxx \text{ with } \forall x \forall P (\forall y (Rxy \rightarrow Py) \rightarrow Px).$$

Yet it was the discovery of the former simple *first-order* equivalents that motivated the above investigation in the first place. Now for some modal formulas, the second-order complexity may be unavoidable — witness the example of McKinsey’s Axiom. But at least, there arises an obvious basic

QUERY: Which modal formulas define first-order relational conditions — and how do they manage it?

By the above perspective, classical sources provide one immediate answer. A  $\Pi_1^1$ -sentence is first-order definable if and only if it is preserved under the formation of *ultraproducts*, a fundamental construction in classical model theory. Through the above transcription, the same criterion applies to modal formulas. (The technical ins and outs of this point, as well as of related ones in this introduction, are postponed until the relevant sections: Sections 2.1 and 2.2 in this case.)

### *Modal Correspondence Theory*

The preceding query has been the starting point for a systematic study of classical definability of modal formulas, when viewed as relational principles. Now the mentioned ultraproduct characterisation is a very abstract, global one, rather removed from the actual business of finding correspondences. Also historically, it is a rather late development — and we shall therefore turn to more concrete themes, as they evolved.

At first sight, *proving* first-order definability seems a simple matter: just find an equivalent, and show that it works. Still, there is the question how much system there is to this activity. For instance, Examples 1–3 exhibited regularities in their proofs. And indeed, closer inspection reveals that reflexivity, transitivity and directedness may be obtained from the second-order transcriptions of the **S4.2**-axioms through certain *substitutions* of ‘minimal’ *definable assignments*.

The heuristics behind this method is simply this. If, e.g.  $\Box p \rightarrow p$  is true at  $x$ , then the most ‘parsimonious’ way of verifying the antecedent (i.e. by having  $V(p) = \{y \mid Rxy\}$ ) carries maximal information about the whole implication. This essentially, is why the substitution of  $Rxu$  for  $Pu$  in

$$\forall x \forall P (\forall y (Rxy \rightarrow Py) \rightarrow Px)$$

yields the equivalent formula

$$\forall x (\forall y (Rxy \rightarrow Rxy) \rightarrow Rxx).$$

By the universal validity of the antecedent, the latter may be simplified to the usual statement of reflexivity. A completely analogous line of thought produces transitivity from the transcription of  $\Box p \rightarrow \Box \Box p$ . Some complications arise with antecedents as in  $\Diamond \Box p \rightarrow \Box \Diamond p$ ; but the general idea remains the same. In this way, one discovers a large recursive class of modal formulas with effectively obtainable first-order equivalents.

Nevertheless, this method of substitutions also has definite limits. Notably, it does not work for all first-order definable modal formulas — as will be proved in Section 2.2 for the case of **S4.1**. In connection with this matter, the exact *combinatorial complexity* of the set of first-order definable modal formulas is still unknown — but there are reasons for fearing that it is not even arithmetically definable (let alone, recursive or recursively enumerable).

*Disproving* first-order definability is a more difficult matter. Indeed, how should one go about this at all? The common pattern in all examples in the literature comes to this: find some semantic preservation property of first-order sentences, which is lacked by the modal formula under consideration. Thus, e.g. the earliest published contribution by the present author was an example showing how the McKinsey Axiom sins against the Löwenheim–

Skolem theorem. It holds in a certain uncountable Kripke frame (to be presented in Section 2.2.) without holding in any of a certain group of its countable elementary subframes. A classical example of this phenomenon occurs when Dedekind Continuity (itself a  $\Pi_1^1$ -property) is added to the first-order ordering theory of the rationals. The resulting  $\Pi_1^1$ -sentence has uncountable models (notably, the reals); but, it even lacks countable models altogether.

The modal examples of ‘essentially second-order’ axioms to be found in Section 2.2 will serve to delimit the range of the above method of substitutions. As so often, the McKinsey Axiom again provides an illuminating example. The above heuristics of ‘minimal verification’ typically fails for antecedents such as  $\Box\Diamond p$ , expressing some dependency — and first-order failure is immediate.

Besides the modal half of the story, so to speak, there also exists the opposite direction, looking from classical formulas to modal ones. Again, this inspires a basic

QUERY. Which first-order relational conditions are modally definable?

The ‘positive’ side of this matter again concerns the establishing of valid equivalences. Thus, for instance, how does one find a modal definition for such a classical favourite as *connectedness*

$$\forall xyz((Rxy \wedge Rxz) \rightarrow (Ryz \vee Rzy))?$$

This time, the heuristics consists in imagining a situation where the property fails, together with a way of ‘maximally exploiting’ this failure through modal formulas. In the above particular case, supposing that  $Rxy, Rxz, \neg Ryz, \neg Rzy$ , one sets  $\Box p$  true at  $y$  (with  $p$  false at  $z$ ) and  $\Box q$  true at  $z$  (with  $q$  false at  $y$ ). This has the effect of verifying the following formula at  $x$ :

$$\Diamond(\Box p \wedge \neg q) \wedge \Diamond(\Box q \wedge \neg p).$$

Now, the original property itself will correspond to the negation of this modal ‘failure description’, i.e.

$$\neg(\Diamond(\Box p \wedge \neg q) \wedge \Diamond(\Box q \wedge \neg p)).$$

By some familiar equivalence transformations, this becomes

$$\Box(\Box p \rightarrow q) \vee \Box(\Box q \rightarrow p),$$

a principle known from the literature as Geach’s Axiom.

It remains to be shown, of course, that conversely, failure of this axiom implies failure of connectedness; but this is immediate. In order to cross-check, one might also apply the earlier method of substitutions to (some

suitable transform of) the Geach Axiom: and indeed, connectedness will ensue.

The ‘negative’ side again consists of disproofs. Here as well, these turn out to possess a particular interest — as we are forced to contemplate ‘typical behaviour’ of modal formulas. A standard example is the following. Although reflexivity was modally definable, *irreflexivity* turns out intractable:  $\forall x \neg Rxx$ . But, failed attempts are no definite refutations. What we need is some semantic property of modal formulas, as relational conditions on Kripke frames, which is not shared by this particular first-order sentence.

At this point, the modal model theory of Section 2.1 comes in. There, one finds that the following mappings play a fundamental role in the transmission of modal truth between Kripke frames: a *p-morphism* is a function  $f$  from a frame  $\langle W, R_1 \rangle$  to a frame  $\langle W_2, R_2 \rangle$  which

1. preserves  $R_1$ , and
2. ‘almost’ preserves  $R_2$ , in the following sense:  
‘If  $R_2 f(w)v$ , then there exists some  $u \in W_1$  such that (a)  $R_1 wu$  and (b)  $f(u) = v$ ’.

Under different names, this notion has had a career in standard logic already, e.g. the ‘Mostowski collapse’ in set theory is of this kind.

For the purposes of the present example, it need only be recorded that subjective *p*-morphisms preserve truth of modal formulas on Kripke frames. But then, irreflexivity may be dismissed: it holds in the frame of the natural numbers with the usual order, but it fails in its *p*-morphic image (!) arising from the contraction to one single reflexive point.

This example will have given a taste of the actual field-work in this area of Correspondence Theory. There also arises the more general question, of course, whether some combination of modally valid preservation requirements manages to *characterise* all and only the modally definable first-order sentences. This is indeed the case, and an elegant result to this effect — involving *p*-morphisms as well as other basic constructions, will be proved in Section 2.4.

The preceding survey by no means exhausts the range of questions that can be investigated in Correspondence Theory — but it does convey the spirit.

### *Correspondence and Completeness*

Three pillars of wisdom support the edifice of Modal Logic. There is the ubiquitous *Completeness Theory*, the present Correspondence, or, more generally, *Definability Theory* — and finally, the *Duality Theory* between Kripke frames and ‘modal algebras’ (cf. Section 2.3 below) has become an area of its own. Connections between the latter two will become apparent as Section

2 unfolds — in particular, the above-mentioned characterisation of modally definable first-order sentences will be obtained as a consequence of the classic Birkhoff Theorem of Universal Algebra, applied to modal algebra.

The relation between correspondence and completeness is less vital to subsequent developments. Moreover, it turns out to be rather complex — and indeed, only partially understood. Nevertheless, for those readers who are familiar with the basic notions of Completeness Theory, the following sketch of issues may serve to bring questions of correspondence closer to traditional concerns.

The early completeness theorems in modal logic were brought under one heading in [Seegerberg, 1971]: ‘modal logic  $\mathbf{L}$  is determined by a class  $\mathfrak{R}$  of Kripke frames’, i.e.  $\mathbf{L}$  axiomatises the modal theory of  $\mathfrak{R}$  (on the basis of the minimal logic  $\mathbf{K}$ ).

As before, two perspectives emerge here. First, one may start with a given class  $\mathfrak{R}$ , asking for a recursive axiomatisation  $\mathbf{L}$  of its modal theory. In general, there is no guarantee for success here; but there is one helpful observation involving first-order definability.

**FACT 4.** If  $\mathfrak{R}$  is elementary (i.e. defined by a single first-order sentence), then its modal theory is recursively axiomatisable.

**Proof.** Let  $\alpha = \alpha(R, =)$  define  $\mathfrak{R}$ . A modal formula  $\varphi$  belongs to the theory of  $\mathfrak{R}$  if and only if it holds in all frames in  $\mathfrak{R}$ . This may be restated as follows:

$$\alpha \models \forall x \forall P_1 \dots \forall P_n \tau(\varphi);$$

where  $\tau(\varphi)$  is the earlier first-order translation of  $\varphi$ , while  $p_1, \dots, p_n$  are the proposition letters occurring in the latter formula. Now, the predicate variables  $P_1, \dots, P_n$  do not occur in the first-order sentence  $\alpha$ , and, therefore the above implication is equivalent to  $\alpha \models \forall x \tau(\varphi)$ . But this is an ordinary first-order implication. So, since the latter notion is recursively axiomatisable, the same must be true for membership of the modal theory of  $\mathfrak{R}$ .

Axiomatisable, yes, but axiomatisable on the basis of the minimal modal logic  $\mathbf{K}$ ? Even this is true, choosing a suitable recursive set of axioms as in the proof of Craig’s Theorem in classical logic and noticing that  $\mathbf{K}$  contains *modus ponens* (which is all that is needed). ■

Thus, in retrospect, the earlier completeness theorems for reflexive, transitive orders (and other elementary classes) were quite predictable.

The direction from classes of frames to logics is not the current one in modal logic; being more appropriate to areas such as tense logic, where temporal structures often precede temporal theories. Usually, one already possesses a certain logic  $\mathbf{L}$ , asking for a class  $\mathfrak{R}$  of Kripke frames with respect to which it is complete. (Notice that, if *any* class  $\mathfrak{R}$  suffices, then the whole class of Kripke frames validating  $\mathbf{L}$  will.)

Nowadays, we know that not all modal logics are in fact *complete* in the above sense, contrary to earlier expectations. This is the content of the celebrated ‘modal incompleteness theorems’ in [Fine, 1974; Thomason, 1974]. But it has been hoped that, at least, all *first-order definable* axiom sets are complete. (Indeed, a defective proof to this effect has circulated.) Even this more modest expectation was frustrated in [van Benthem, 1978]:

FACT 5. The modal logic  $\mathbf{L}$  with characteristic axioms

$$\begin{aligned} \Box p &\rightarrow p \\ \Box \Diamond p &\rightarrow \Diamond \Box p \\ (\Diamond p \wedge \Box(p \rightarrow \Box p)) &\rightarrow p \end{aligned}$$

is first-order definable: its frames are just those satisfying the condition

$$\forall xy(Rxy \leftrightarrow x = y).$$

But the characteristic axiom of the modal theory of the latter class of frames, viz.  $\Box p \leftrightarrow p$ , is not minimally derivable from  $\mathbf{L}$ .

The relevant correspondence will be proved in Section 2.2. For the moment, it may be noticed that the third axiom defines a notion of ‘safe return’: from any  $R$ -successor of a world  $x$ , one can always return to  $x$  by following some finite  $R$ -chain of  $R$ -successors of  $x$ .

The relevant argument is highly nontrivial, far outside the range of our earlier method of substitutions. Nevertheless, even the latter has its relevance for completeness theory, as we shall see presently.

What the modal incompleteness theorems show is that the minimal modal logic  $\mathbf{K}$  is too weak to produce all modally valid inferences. But of course, there may be stronger reasonable ‘base logics’. One particular example arises from the method of substitutions. For instance, in proving the equivalence of substitution instances with more current first-order conditions, one uses an extremely natural second-order logic  $\mathbf{K}_2$  with the following deductive apparatus:

Some first-order base complete with respect to *modus ponens*,  
similar axioms for the second-order quantifiers;

with the following form of ‘first-order instantiation’ allowed for first-order formulas  $\psi$

$$\forall x\varphi(X) \rightarrow \varphi(\psi).$$

Through the earlier second-order transcription,  $\mathbf{K}_2$  may be used as a modal base logic.

Here is an example of some fame. In the metamathematics of arithmetical provability (cf. [Boolos, 1979] or Smoryński’s in a later volume of this *Handbook*), the following two modal axioms are basic:

$$\Box p \rightarrow \Box \Box p, \quad \Box(\Box p \rightarrow p) \rightarrow \Box p \quad (\text{‘Löb’s Axiom’}).$$

The semantic import of the latter will be established in Section 2.2: it holds in those Kripke frames whose alternative relation is transitive, while possessing a well-founded converse. Moreover, transitivity is  $\mathbf{K}_2$ -derivable from Löb's Axiom, by the substitution of

$$Rxy \wedge \forall y(Ruy \rightarrow Rxy) \quad \text{for } Pu.$$

(The antecedent becomes universally valid, while the consequent expresses transitivity.) An advantage of  $\mathbf{K}_2$  over  $\mathbf{K}$ ? No, around 1975, Dick de Jongh and Giovanni Sambin found a  $K$ -deduction for the first axiom from the second after all. The two deductions are related, but systematic connections between  $\mathbf{K}$ -deductions and  $\mathbf{K}_2$ -deductions have not been explored up to date.

Nevertheless,  $\mathbf{K}_2$  is non-conservative over  $\mathbf{K}$  in the modal realm. In [van Benthem, 1979b] we find the following incompleteness theorem.

FACT 6. The modal axiom

$$\diamond \Box \perp \vee \Box(\Box(\Box p \rightarrow p) \rightarrow p),$$

with  $\perp$  the falsum, defines the same class of Kripke frames as  $\diamond \Box \perp \vee \Box \perp$ . But, the latter formula is not  $\mathbf{K}$ -derivable from the former — even though it is  $\mathbf{K}_2$ -derivable.

Again, there is a correspondence involved here. But the idea is illustrated by a simple  $\mathbf{K}_2$ -deduction at the back of this result:

1.  $\forall P(\forall y(Rxy \rightarrow (\forall z(Ryz \rightarrow Pz) \rightarrow Py)) \rightarrow Px) \quad (' \Box(\Box p \rightarrow p) \rightarrow p')$ ,
2.  $\forall y(Rxy \rightarrow (\forall z(Ryz \rightarrow z \neq x) \rightarrow y \neq x)) \rightarrow x \neq x \quad (x \neq u \text{ for } Pu)$ ,
3.  $\neg \forall y(Rxy \rightarrow (\forall z(Ryz \rightarrow z \neq x) \rightarrow y \neq x))$ ,
4.  $\exists y(Rxy \wedge \forall z(Ryz \rightarrow z \neq x) \wedge y = x)$ ,
5.  $Rxx \wedge \forall z(Rxz \rightarrow z \neq x)$
6.  $x \neq x$ : a contradiction ( $\perp$ ).

That  $\mathbf{K}_2$ , in its turn, must be modally incomplete (as is any proposed recursively axiomatised base logic) follows from the *general* incompleteness results in [Thomason, 1975].

First-order definability does not imply completeness. But, *when* a modal logic is both first-order definable and complete, it enjoys a very pleasant form of the latter property — viz. with respect to the underlying frame of its own *Henkin model*. (First-order definability plus completeness imply canonicity': cf. [Fine, 1975; van Benthem, 1980].) Such *canonical* modal logics will be characterised semantically in Section 2.4: notice that many

of the familiar text book examples are of this kind. In fact, a canonical completeness proof, such as that for **S4**, often proceeds by means of first-order conditions on the Henkin model, induced by the corresponding axioms.

The relation between these familiar ‘Henkin arguments’ and the above method of substitutions is at present still rather mysterious. Sahlqvist [1975] contains many examples of parallels; but Fine [1975] presents a problem. The modal formula

$$\Diamond \Box(p \vee q) \rightarrow \Diamond(\Box p \vee \Box q)$$

axiomatises a canonical modal logic, without being first-order definable. Thus, we are still far from complete clarity in the area between completeness and correspondence.

### *Variations and Generalisations*

Logical model theory may be viewed as a marriage between ontology and language (or ‘mathematics’ and ‘linguistics’). Accordingly, the semantics of propositional modal logic, our paradigm example up till now, exhibits the familiar triangle

$$\begin{array}{ccc} \text{language} & \xrightarrow{\hspace{2cm}} & \text{structures} \\ & \text{interpretation} & \end{array}$$

Or, from the above translational point of view, the components are

$$\begin{array}{ccc} \text{prima facie language} & \xrightarrow{\hspace{2cm}} & \text{representation language} \\ & \text{translation} & \end{array}$$

All these ‘degrees of freedom’ may be varied in intensional logic — and thus there appears a whole family of ‘correspondence theories’. We shall explore some examples of recognised importance in Section 3. Here, let us just think about the various possibilities and their implications.

Even within the domain of propositional modal logic, alternatives have been proposed for Kripke-type relational semantics. Jennings, Johnstone and Schotch [1980] contains the proposal to work with *ternary* alternative relations, employing the following notion of necessity:

$$\Box \varphi \text{ is true at } x \text{ if } \forall yz(Rxyz \rightarrow \varphi(y) \vee \varphi(z)).$$

Their motivation was, amongst others, to create room for ‘non-cumulation’ of necessities: the ‘Aggregation Axiom’

$$\Box p \wedge \Box q \rightarrow \Box(p \wedge q)$$

will no longer be valid. What happens to earlier correspondences in this new light? Old boundaries start shifting; e.g.  $\Box p \rightarrow p$  remains first-order definable, but  $\Box p \rightarrow \Box \Box p$  becomes essentially second-order on this semantics.



This is compensated for by the phenomenon of formerly unexciting principles, such as the Aggregation Axiom (which was trivially valid before) springing into unexpected bloom:

EXAMPLE 7.  $\Box p \wedge \Box q \rightarrow \Box(p \wedge q)$  defines

$$\forall xyz(Rxyz \rightarrow (y = z \vee Rxyy \vee Rxzz)).$$

**Proof.** ‘ $\Rightarrow$ ’: Suppose the condition fails at  $x, y, z$ . Setting

$$V(p) = W = \{z\}, V(q) = W - \{y\},$$

will then verify  $\Box p, \Box q$  at  $x$ , while  $\Box(p \wedge q)$  is falsified (by  $Rxyz$ ).

‘ $\Leftarrow$ ’: Suppose that  $\Box p, \Box q$  hold at  $x$ , and consider  $Rxyz$ . Either  $y = z$ , whence  $y$  verifies both  $p$  and  $q$  (by  $Rxyy$  and the truth definition), or  $Rxyy$ , implying the same conclusion, or  $Rxzz$ , in which case  $z$  verifies both  $p$  and  $q$ . So,  $\Box(p \wedge q)$  holds at  $x$ . ■

As for the general theorems, forming the backbone of the subject, nothing essential changes in this ternary semantics.

This example changed both the structures and the form of the truth definition. What may not be generally realised is the variety offered even when fixing the two parameters of ‘language’ and ‘structures’. Therefore, a short digression is undertaken here.

The Kripke truth definition is not sacrosanct — other clauses would have been quite imaginable. Thus, for instance, we may make the following

OBSERVATION 8. The truth definition ‘ $\Box\varphi$  is true at  $x$  if  $\forall y((Rxy \vee Ryx) \rightarrow \varphi(y))$ ’ yields as a modal base logic **KB**; i.e. the minimal logic **K** plus the Brouwer Axiom  $p \rightarrow \Box\Diamond p$ .

**Proof.** The Brouwer Axiom defines *symmetry* of the alternative relation; as may be seen by substituting  $u = x$  for  $Pu$ . And indeed **KB** is complete with respect to the class of symmetric Kripke frames. Hence, any non-theorem  $\varphi$  of **KB** is falsified on some symmetric frame  $\langle W, R \rangle$ . But, on symmetric frames  $R$  coincides with the relation  $\lambda xy. (Rxy \vee Ryx)$  (i.e.  $R$  united with its converse  $\bar{R}$ ); whence  $\varphi$  also fails by the new evaluation.

Conversely, if  $\varphi$  has a counter-example  $\langle W, R \rangle$  under the new truth definition, then it has  $\langle W, R \cup \bar{R} \rangle$  for an ordinary symmetric counter-example; whence it is outside of **KB**. ■

Thus, there is a possible *trade-off* between truth definition and requirements on the alternative relation. The exact extent of this phenomenon remains to be investigated. Notice for example how **KB** is equally well generated by the following truth definition:

$$\Box\varphi \text{ is true at } x \text{ if } \forall y((Rxy \wedge Ryx) \rightarrow \varphi(y)).$$

The general principle behind such examples is this.

FACT 9. If  $C(R)$  is any condition on  $R$ , and  $\gamma(x, y)$  some formula in  $R$ , = such that

1. If  $C(R)$  is satisfied, then  $R$  and  $\lambda xy.\gamma(x, y)$  coincide,
2.  $\lambda xy.\gamma(x, y)$  satisfies  $C$ ,

then the modal logic determined by (the Kripke frames obeying)  $C$  may also be generated without conditions through the truth definition

$$\Box\varphi \text{ is true at } x \text{ if } \forall y(\gamma(x, y) \rightarrow \varphi(y)).$$

This rather subversive shift in perspective will not be investigated in this contribution. At this point, it merely serves to remind us that not a single aspect of the semantic enterprise is immune to revision.

Leaving the realm of modal logic, of the many intensional candidates for a correspondence perspective, only a few have been explored up to date. In Section 3, some important examples are reviewed briefly, viz. *tense logic*, *conditional logic* and *intuitionistic logic*. These illustrate, in ascending order, certain difficulties which tend to make Correspondence Theory rather more difficult (often also: more exciting) in many cases. These difficulties have to do with ‘pre-conditions’ on the alternative relation (not very serious), and the phenomenon of ‘admissible assignments’ (rather more serious), to be explained in due course. Nevertheless, for instance, Intuitionistic Correspondence Theory will turn out to possess also some elegant features lacked by its modal predecessor.

A few examples, even without proof, will render the above remarks more concrete. In tense logic, the correspondence runs between temporal axioms and properties of the temporal order (‘before’, ‘earlier than’).

EXAMPLE 10 (‘Hamblin’s Axiom’).  $(p \wedge Hp) \rightarrow FHp$  defines discreteness of Time:

$$\forall x \exists y > x \forall z < y (z = x \vee z < x).$$

In the logic of counterfactual conditionals, conditional inferences are related to the behaviour of the comparative similar ordering  $C$  among alternative worlds.

EXAMPLE 11 (Stalnaker’s Axiom of ‘Conditional Excluded Middle’).  $(p \Rightarrow q) \vee (p \Rightarrow \neg q)$  defines linearity of alternative worlds:

$$\forall xyz (y = z \vee Cxyz \vee Cxzy).$$

Finally, in intuitionistic logic, (‘intermediate’) axioms impose constraints upon the possible growth patterns of stages of knowledge.

EXAMPLE 12 ('Weak Excluded Middle').  $\neg p \vee \neg\neg p$  defines 'local convergence' of growing stages, i.e. directedness:

$$\forall xyz((x \subseteq y \wedge x \subseteq z) \rightarrow \exists u(y \subseteq u \wedge z \subseteq u)).$$

Proofs, and further explorations are postponed until the relevant sections. At this stage, the experienced reader may predict that two nuts will be especially difficult to crack for any Correspondence Theory.

The first of these concerns the earlier tacit restriction to *propositional* logic: what happens in the predicate case? In Section 2.5 we shall see that no essential problems seem to arise — although the field remains largely unexplored.

A more formidable problem arises when the truth definition for the intensional operators itself becomes of higher-order complexity. In that case, e.g. a search for possible first-order equivalents of intensional axioms seems rather pointless. This eventuality arises when disjunction is evaluated barwise in Beth semantics for intuitionistic logic (i.e.  $\varphi \vee \psi$  is true at  $x$  if the  $\varphi$ -worlds and  $\psi$ -worlds together form a barrier intersecting each branch passing through  $x$ ).

The last word has not been said here, however. Philosophically, it seems a rather unsatisfactory division of semantic labour to let the truth definition absorb structural complexity (in this case: the second-order behaviour of branches). The latter should be located where it belongs, viz. in the structures themselves. And indeed, the Beth semantics admits of a two-sorted first-order reformulation in terms of nodes and paths, which generates a Correspondence Theory of the usual kind.

All this is not to say that there are no limits to the useful application of a correspondence perspective. But, these are to be found in *philosophical* relevance, rather than technical *impossibility*. One should study correspondences only as long as they serve the purpose of semantic enlightenment — which is the shedding of light upon one conceptual framework by relating it systematically to another.

## 2 MODALITY

In this chapter, modal correspondence theory will be surveyed against the background of modal model theory and modal algebra, whose basics are explained. (Cf. the chapter by Bull and Segerberg in this volume for the necessary background.)

### 2.1 Modal Model Theory

The basic structures of modal semantics are introduced: *frames*, *models* and *general frames*. These may be studied either purely classically, or

with a specifically modal purpose. In both cases, the emphasis is not upon such structures in isolation, but upon their ‘categorical context’: what are their relations with other structures, and which of these relations are truth-preserving? Thus, we will introduce the modal preservation operations of *generated subframe*, *disjoint union*, *p-morphic image* and *ultrafilter extension*. Moreover, the fundamental classical formation of *ultraproducts* will be used as well. All these notions will appear again and again in later sections.

*Semantic structures.* The structures used in the Kripke truth definition are *models*  $M$ , i.e. triples  $\langle W, R, V \rangle$ , where  $W$  is a nonempty set of *worlds*,  $R$  is a *binary alternative* relation on  $W$ , and  $V$  is a *valuation* assigning sets of worlds  $V(p)$  to proposition letters  $p$ . The notion explicated then becomes

$$M \models \varphi[w] \quad : \text{‘}\varphi \text{ is true in } M \text{ at } w\text{’}.$$

In our correspondence theory we also want to see the bare bones: a *frame*  $F$  is a couple  $\langle W, R \rangle$  as above, but without a valuation. There is nothing intrinsically ‘modal’ about all this, of course. Frames are just the ‘directed graphs’ of Graph Theory.

In Sections 2.3 and 2.4, a third notion of modal structure will be required as well — intermediate, in a sense between models and frames. A *general frame*  $F$  is a couple  $\langle F, \mathfrak{W} \rangle$ , or alternatively, a triple  $\langle W, R, \mathfrak{W} \rangle$  such that  $F = \langle W, R \rangle$  is a frame, and  $\mathfrak{W}$  is a set of subsets of  $W$ , closed under the formation of *complements*, *unions* and *modal projections*. Formally,

$$\begin{aligned} \text{if } X \in \mathfrak{W}, & \quad \text{then } W - X \in \mathfrak{W} \\ \text{if } X, Y \in \mathfrak{W}, & \quad \text{then } X \cup Y \in \mathfrak{W} \\ \text{if } X \in \mathfrak{W}, & \quad \text{then } \pi(X) =_{\text{def}} \{w \in W \mid \exists v \in X : R w v\} \in \mathfrak{W}. \end{aligned}$$

The following example illustrates the effect of restricted sets  $\mathfrak{W}$ . Consider the frame  $\langle N, \leq \rangle$ , where  $N$  is the set of natural numbers. Its modal theory contains such principles as  $\Box p \rightarrow p$ ,  $\Box p \rightarrow \Box \Box p$  and Geach’s Axiom: together forming the logic **S4.3**. Typically left out is the McKinsey Axiom  $\Box \Diamond p \rightarrow \Diamond \Box p$ ; as it may be falsified in some infinite alternation of  $p$ ,  $\neg p$ : say by  $V(p) = \{2n \mid n \in N\}$ . But now, consider the structure  $\langle N, \leq, \mathfrak{W} \rangle$ , where  $\mathfrak{W}$  consists of all *finite* and all *cofinite* subsets of  $N$ . It is easily checked that all three closure conditions obtain for  $\mathfrak{W}$ . Thus, we have a general frame here. Its logic contains the earlier one (‘a fortiori’); but it also adds principles. Notably, the McKinsey Axiom can no longer be falsified, as the above ‘tell-tale’ valuation is no longer admissible. Thus, **S4.1** holds in this general frame, although it does not in the underlying ‘full frame’. And further increases in the modal theory are possible, by restricting  $\mathfrak{W}$  even more; e.g. there is even a most austere choice, viz.  $\mathfrak{W} = \{\emptyset, N\}$ , which yields a general frame validating the ‘classical logic’ with axiom  $\Box p \leftrightarrow p$  — which was still invalid in the previous general frame. Thus, one single underlying frame may still generate a hierarchy of modal logics.

The original algebraic motivation for this notion (due to Thomason [1972]) will be given in Section 2.3. But here already, a direct logical reason may be given. Kripke frames are so-called ‘standard models’ for modal formulas, considered as second-order  $\Pi_1^1$ -sentences: the universal predicate quantifiers range over *all* sets of possible worlds. An intermediate possibility would have been to allow also ‘general models’ in the sense of Henkin [1950]: in which this second-order range may be restricted, say to some set  $\mathfrak{W}$ . Usually, such ranges are to be closed under certain mild conditions of definability — in order to verify reasonable forms of the universal instantiation (or ‘comprehension’) axiom. This, of course, is precisely what happened in the above. The uses of this notion lie partly in modal Completeness Theory, partly in modal algebra. For the moment, it will not be a major concern.

*Semantic questions.* Given a formal language, interpreted in certain structures, a plethora of questions arises concerning the interplay between more ‘linguistic’ and more ‘structural’ (or ‘mathematical’) notions. We mention only a few fundamental ones.

Arguably the ‘first question’ of any model theory is that concerning the relation between linguistic indistinguishability (equality of modal theories) and structural indistinguishability (isomorphism) of semantic structures. How far do the webs of language and ontology diverge? In classical logic, we know that (first-order) elementary equivalence coincides with isomorphism on the *finite* structures, but no higher up: isomorphism then becomes by far the finer sieve.

Now, the modal language on *models* behaves like the first-order language of the first translation in the introduction: nothing spectacular results. But the second-order notion seems more interesting in this respect. (Equality of second-order theories is quite strong: modulo the Axiom of Constructibility, it even implies isomorphism in all *countable* frames; cf. [Ajtai, 1979]). From Van Benthem [1985], which treats the analogous question for tense logic in Chapter 2.2.1, we extract

**THEOREM 13.** *Finite Kripke frames that are generated by a single point (cf. below) are isomorphic if and only if they possess the same modal theory. But, the countable Kripke frames  $\mathbb{Z} \odot \mathbb{Z}$  (the integers, with each point replaced by a copy of the integers) and  $\mathbb{Q} \odot \mathbb{Z}$  (the rationals, treated likewise) possess the same modal theory, without being isomorphic.*

In tense logic, the latter result means that the formal language cannot distinguish between locally discrete/globally discrete and locally discrete/globally dense Time. (The latter may well be that of our World.) In the context of modal logic, no such appealing interpretation is possible, whence we forego further discussion of the above result.

From now on, we will confine attention to a single theme, which again, is characteristic for much of what goes on in Model Theory.

*Truth-preserving operations.* In evaluating the truth of a modal formula  $\varphi$  at a world  $w$  we only have to consider  $w$  itself, (possibly) its  $R$ -successors, (possibly) their  $R$ -successors, etcetera. Thus, only that part of the frame is involved which is ‘ $R$ -generated’ by  $w$ , so to speak. In general, one never has to look beyond  $R$ -closed environments of  $w$ : an observation summed up in the following notion and result.

DEFINITION 14.  $M_1 (= \langle W_1, R_1, V_1 \rangle)$  is a *generated submodel* of  $M_2 (= \langle W_2, R_2, V_2 \rangle)$  (notation:  $M_1 \xrightarrow{G} M_2$ ) if

1.  $W_1 \subseteq W_2$
2.  $R_1 = R_2$  restricted to  $W_1$ ,
3.  $V_1(p) = V_2(p) \cap W_1$ , for all proposition letters  $p$ ; i.e.  $M_1$  is an ordinary *submodel* of  $M_2$ , which has the additional feature that
4.  $W_1$  is closed under passing to  $R_2$ -successors.

The next result is the famous ‘Generation Theorem’ of Segerberg [1971].

THEOREM 15. *If  $M_1 \xrightarrow{G} M_2$ , then for all worlds  $w \in W_1$  and all modal formulas  $\varphi$ ,  $M_1 \models \varphi[w]$  iff  $M_2 \models \varphi[w]$ .*

This is what happens inside a single model. When comparisons are desired between evaluation in distinct models, a more external connection is required.

DEFINITION 16. A relation  $C$  is a *zigzag connection* between two models  $M_1, M_2$  if

1. domain  $(C) = W_1$ , range  $(C) = W_2$ ,
  - (a) if  $Cwv$  and  $w' \in W_1$  with  $R_1ww'$ , then  $Cw'v'$  for some  $v' \in W_2$  with  $R_2vv'$  (‘forth choice’)
  - (b) If  $Cwv$  and  $v' \in W_2$  with  $R_2vv'$ , then  $Cww'$  for some  $w' \in W_1$  with  $R_1ww'$  (‘back choice’)
2. if  $Cwv$ , then  $w, v$  verify the same proposition letters.

Starting from the basic case (3), the back-and-forth clauses ensure that evaluation of successive modalities in modal formulas yield the same results on either side:

THEOREM 17. *If  $M_1$  is zigzag-connected to  $M_2$  by  $C$ , then, for all worlds  $w \in W_1, v \in W_2$  with  $Cwv$ , and all modal formulas  $\varphi$ ,*

$$M_1 \models \varphi[w] \text{ iff } M_2 \models \varphi[v].$$

Notation.  $M_1 \xrightarrow{Z} M_2$  for zigzag-connected models (by some  $C$ ).

By a result in Van Benthem [1976], the Generation Theorem and the preceding ‘Zigzag Theorem’ combined are characteristic for modal formulas as first-order formulas in the sense of the introduction:

**THEOREM 18.** *A first-order formula  $\varphi(x)$  in the language with  $R, P, Q, \dots$  is logically equivalent to some modal transcription if and only if it is invariant for generated submodels and zigzag connections (in the above sense).*

For the case of pure frames, the above notions and results lead to the following three preservation results.

**DEFINITION 19.**  $F_1$  is a *generated subframe* of  $F_2$  ( $F_1 \hookrightarrow F_2$ ) if

1.  $W_1 \subseteq W_2$ ,
2.  $R_1 = R_2$  restricted to  $W_1$ ,
3.  $W_1$  is  $R_2$ -closed in  $W_2$ .

In general logic, this type of situation is often described by saying that the ‘converse frame’  $\langle W_2, \check{R}_2 \rangle$  is an *end extension* of  $\langle W_1, \check{R}_1 \rangle$ : the added worlds all come ‘at the end’.

From Theorem 15 we derive *preservation* under generated subframes:

**COROLLARY 20.** *If  $F_1 \hookrightarrow F_2$ , then  $F_2 \models \varphi$  implies  $F_1 \models \varphi$ , for all modal formulas  $\varphi$ .*

Here ‘ $F \models \varphi$ ’ means ‘ $\varphi$  is true in  $F$ ’, in the global second-order sense of the introduction: at all worlds, under all valuations.

But Theorem 15 also has an ‘upward’ directed moral.

**DEFINITION 21.** The *disjoint union*  $\oplus\{F_i | i \in I\}$  of a family of frames  $F_i = \langle W_i, R_i \rangle$  is the disjoint union of the domains  $W_i$ , with the obvious coordinate relations  $R_i$ .

Another direct application is *preservation under disjoint unions*:

**COROLLARY 22.** *If  $F_i \models \varphi$  (all  $i \in I$ ), then  $\oplus\{F_i | i \in I\} \models \varphi$ , for all modal formulas  $\varphi$ .*

Next, turning to Theorem 17, one now needs a connection between frames which can be turned into a suitable zigzag relation between models over them.

**DEFINITION 23.** A *zigzag morphism* from  $F_1$  to  $F_2$  is a function:  $W_1 \rightarrow W_2$  satisfying

1.  $R_1 w w'$  implies  $R_2 f(w) f(w')$ ,  
i.e.  $f$  is an ordinary *R-homomorphism*; which has the additional backward property that
2. if  $F_2 f(w) v$ , then there exists  $u \in W_1$  with  $R_1 w u$  and  $f(u) = v$ .

This notion was mentioned under its current, but rather uninformative name of ‘ $p$ -morphism’ in the introduction. Here is one more example:

the map from nodes to levels (counting from the top) is a zigzag morphism from the infinite binary tree (with the descendant relation) onto the natural numbers (with the usual ordering).

Notice also that injective (1-1) zigzag morphisms are even just isomorphisms.

Theorem 17 now implies the ‘ $p$ -morphism’ theorem of Segerberg [1971].

**COROLLARY 24.** *If  $f$  is a zigzag morphism from  $F_1$  onto  $F_2$ , then, for all modal formulas  $\varphi$ ,  $F_1 \vDash \varphi$  implies  $F_2 \vDash \varphi$ .*

For more ‘local’ versions of these results, the reader is referred to [van Benthem, 1983].

More examples, and applications of Corollaries 20, 22, and 24 will be found in Section 2.4. A quick impression may be gained from the following sample observation (D. C. Makinson). The modal theory of any Kripke frame is either contained in the classical modal logic (characteristic axiom  $\Box p \leftrightarrow p$ ) or the ‘absurd’ modal logic (characteristic axiom  $\Box(p \wedge \neg p)$ ). For, any frame  $F$  either contains end points without  $R$ -successors, or it is *serial* ( $\forall x \exists y Rxy$ ). In the former case, such an end point by itself forms a generated subframe, and by Corollary 20, the logic of the frame is contained in that of the subframe — which is the absurd one. In the latter case, contraction to one single reflexive point is a zigzag morphism, and by Corollary 24, the logic of the frame is contained in that of the reflexive point — which is the classical one.

We conclude by noting that these three notions are easily adapted to *general frames*, taking due precautions concerning the various sets  $\mathfrak{W}_1, \mathfrak{W}_2$ . Here are the three necessary additions:

In 19: add ‘ $\mathfrak{W}_1 = \{X \cap W_1 \mid X \in \mathfrak{W}_2\}$ ’.

In 21: add ‘the new  $\mathfrak{W}_2$  remains essentially the old  $\mathfrak{W}_1$ ’ (but for the disjointness procedure used).

In 23: add the following ‘continuity requirement’, reminiscent of topology:

‘for all  $X \in \mathfrak{W}_2$ ,  $f^{-1}[X] \in \mathfrak{W}_1$ ’.

These will be needed in the duality theory of Section 2.3.

*Propositions and possible worlds.* Another characteristic feature of modal semantics is the analogy between *propositions* and *sets of possible worlds*; as well as (moving up one stage in set-theoretic abstraction) that between *possible worlds* and *maximal sets of propositions*. Indeed, many philosophers would deny that there exist any differences here. Let us investigate.

The ideal setting here are general frames  $\langle W, R, \mathfrak{W} \rangle$ : the range is clearly identifiable with a collection of ‘propositions’ over  $W$ .



Now, if worlds are to be considered as sets of propositions, then some obvious desiderata govern the connection between a world  $w$  and propositions  $X, Y$  associated with  $w$ :

1.  $X \in w$  or  $Y \in w$  if and only if  $X \cup Y \in w$  ('analysis')
2.  $X \notin w$  if and only if  $W - X \in w$  ('decisiveness').

Accordingly, one considers only subsets  $w$  of  $\mathfrak{W}$  satisfying these two conditions. These are precisely the so-called *ultrafilters* on  $\mathfrak{W}$ .

What about the alternative relation to be imposed?

Again, a common idea is that a world  $v$  is  $R$ -accessible to  $w$  if it 'satisfies all  $w$ 's modal prejudices', i.e. whenever  $\Box\varphi$  is true at  $w$ ,  $\varphi$  should be true at  $v$ . The same idea may be expressed as follows: whenever  $\varphi$  is true at  $v$ ,  $\Diamond\varphi$  should be true at  $w$ . In the present context, this becomes the following stipulation:

$$Rwv \quad \text{if for all } X \in v, \pi(X) \in w.$$

In this process, no new propositions have been created, whence the former propositions  $X$  now reappear as sets  $\bar{X} = \{w \mid X \in w\}$ .

These considerations motivate

DEFINITION 25. The *ultrafilter extension*  $ue(G)$  of a general frame  $G = \langle W, R, \mathfrak{W} \rangle$  is the general frame  $\langle ue(W, \mathfrak{W}), ue(R, \mathfrak{W}), ue(\mathfrak{W}) \rangle$ , with

1.  $ue(W, \mathfrak{W})$  is the set of all ultrafilters on  $\mathfrak{W}$ ,
2.  $ue(R, \mathfrak{W})wv$ , if for each  $X \in \mathfrak{W}$  such that  $X \in v, \pi(X) \in w$ ,
3.  $ue(\mathfrak{W})$  is  $\{\bar{X} \mid X \in \mathfrak{W}\}$ .

What this construction has done is to re-create  $G$  one level higher up in the set-theoretic air, so to speak, and some calculation will prove

THEOREM 26.  $G$  and  $ue(G)$  verify the same modal formulas.

Still, not everything need have remained the same: the world pattern of  $\langle W, R \rangle$  may differ from that of  $\langle ue(W, \mathfrak{W}), ue(R, \mathfrak{W}) \rangle$ . First, each old world  $w \in W$  generates an ultrafilter  $\{X \in \mathfrak{W} \mid w \in X\}$  and, hence, a corresponding new world in  $ue(W, \mathfrak{W})$ . But, unless  $\mathfrak{W}$  satisfies certain separation principles for worlds, different old worlds may be identified to a single new one. (In the earlier example of  $\langle N, \leq, \{\emptyset, N\} \rangle$ , only a single new world remains, where there used to be infinitely many!) On the other hand, the construction may also introduce worlds that were not there before. For instance, on the earlier general frame  $\langle N, \leq, (\text{co-})\text{finite sets} \rangle$ , the co-finite sets form an ultrafilter which induces a 'point at infinity' in the resulting ultrafilter extension. Indeed, it is easily seen that the latter consists of  $\langle N, \leq \rangle$  followed by just that infinite point.

In Section 2.3, necessary and sufficient conditions will be formulated guaranteeing that a general frame is ‘stable’ under the construction of ultrafilter extensions. In any case, it turns out that the process stabilises after one step at the most. Now, these considerations also apply to ‘full’ Kripke frames.

**DEFINITION 27.** The *ultrafilter extension*  $ue(F)$  of a frame  $F = \langle W, R \rangle$  is the frame  $\langle ue(W), ue(R) \rangle$ , with

1.  $ue(W)$  is the set of all ultrafilters on  $W$ ,
2.  $ue(R)vw$  if for each  $X \subseteq W$  such that  $X \in v, \pi(X) \in w$ .

This time, Theorem 26 does not hold, however. For, it only says that the modal theory of the general frame  $\langle W, R$ , power set of  $W \rangle$  coincides with that of the induced general frame according to Definition 25. Now, the latter is, in general, a restriction of the full frame  $\langle ue(W), ue(R) \rangle$ . Hence, we can only conclude to *anti-preservation under ultrafilter extensions*:

**COROLLARY 28.** *If  $ue(F) \models \varphi$ , then  $F \models \varphi$ , for all modal formulas  $\varphi$ .*

Still, this structural notion can be made a little more familiar by connecting it with previous model-theoretic operations. First, the above-mentioned connection between old worlds and new worlds is 1-1 this time, and indeed isomorphic (consider suitable singleton sets):

**THEOREM 29.**  *$F$  lies isomorphically embedded in  $ue(F)$ .*

In general, this cannot be strengthened to ‘embedded as a generated subframe’. But, another connection with the earlier preservation notions may be drawn from [van Benthem, 1979a].

**THEOREM 30.**  *$ue(F)$  is a zigzag-morphic image of some frame  $F'$  which is elementarily equivalent to  $F$ .*

**Proof.** One expands  $F$  to  $(F, X)_{X \subseteq W}$ , and then passes on to a suitably saturated elementary extension, by ordinary model theory. From the latter, a canonical function from worlds to ultrafilters on  $F$  exists, which turns out to be a zigzag morphism. ■

*Ultraproducts and definability.* New, modally inspired notions concerning frames have been forged in the above. But old classical constructions may be considered as well. Of the various possibilities, only one is selected here, viz. the formation of ultraproducts. (For many other examples, cf. [van Benthem, 1985, Chapter I.2.1].) Its use has been indicated in the introduction already.

The basic theory (and heuristics) of the notion of ‘ultraproduct’ has been given in the Higher Order Logic chapter in volume 1 of this *Handbook*. (Cf. also [Chang and Keisler, 1973, Chapters 4.1 and 6.1].) We recall some of its outstanding features and uses.

DEFINITION 31. For any family of Kripke frames  $\{F_i \mid i \in I\}$  with an ultrafilter  $U$  on  $I$ , the *ultraproduct*  $\Pi_U F_i$  is the frame  $\langle W, R \rangle$  with

1.  $W$  is the set of classes  $f_\sim$ , for all functions  $f \in \Pi\{W_i \mid i \in I\}$ , where  $f_\sim$  is the equivalence class of  $f$  in the relation  $f \sim g \Leftrightarrow \{i \in I \mid f(i) = g(i)\} \in U$ ,
2.  $R$  is the set of couples  $\langle f_\sim, g_\sim \rangle$  for which  $\{i \in I \mid R_i f(i)g(i)\} \in U$ .

This definitional equivalence is lifted by induction to

THEOREM 32 ('Łoś Equivalence'). *For all ultraproducts, and all first-order formulas  $\varphi(x_1, \dots, x_n)$ ,*

$$\Pi_U F_i \models \varphi[f_\sim^1, \dots, f_\sim^n] \text{ iff } \{i \in I \mid F_i \models \varphi[f^1(i), \dots, f^n(i)]\} \in U.$$

Thus, in particular, all first-order sentences  $\varphi$  are *preserved under ultraproducts* in the following sense:

$$\text{if } F_i \models \varphi(\text{all } i \in I), \text{ then } \Pi_U F_i \models \varphi.$$

Conversely, 'Keisler's Theorem' tells us that this is also enough.

THEOREM 33. *A class of Kripke frames is elementary if and only if both that class and its complement are closed under the formation of ultraproducts and isomorphic images.*

**Proof.** Cf. [Chang and Keisler, 1973, Chapter 6.2]. ■

A somewhat more liberal notion of definability, viz. by means of arbitrary *sets* of first-order formulas, yields so-called  $\Delta$ -*elementary* classes. Here the relevant characterisation employs a special case of ultraproducts.

DEFINITION 34. An *ultrapower*  $\Pi_U F$  is an ultraproduct with in each coordinate  $i$  the same frame  $F$ .

Notice that by the Łoś Equivalence,  $\Pi_U F$  is *elementarily equivalent* to  $F$ , i.e. both frames possess the same first-order theory.

THEOREM 35. *A class of Kripke frames is  $\Delta$ -elementary if and only if it is closed under the formation of ultraproducts and isomorphic images, while its complement is closed under the formation of ultrapowers.*

All these notions will be used in the modal correspondence theory of the next section. In this connection, it should be observed that, as for the other kinds of modal semantic structure, ultraproducts of *models* and of *general frames* are easily defined using the above heuristics. These will not be used in the sequel however. (Cf. [van Benthem, 1983].)

The above definability question for classical model theory leads to a clear modal task: ‘to characterise the modally definable classes of Kripke frames’. In section 2.4 this matter will be investigated.

We have arrived at the interplay between classical and modal model theory, which lies at the heart of modal correspondence theory.

## 2.2 Correspondence I: From Modal to Classical Logic

Through the translation given in the Introduction, modal formulas may be viewed as defining constraints on the alternative relation in Kripke frames. Some of these constraints are first-order definable, others are not. Examples are presented of both, after which the former class is explored. A mathematical characterisation is given for it, in terms of ultrapowers, and methods are developed for (dis-)proving membership of the class. The limits of these methods are established as well.

*First-order definability.* The class of modal formulas to be studied here is defined as follows.

**DEFINITION 36.** **M1** consists of all modal formulas  $\varphi$  for which a first-order sentence  $\alpha$  (in  $R, =$ ) exists such that

$$F \models \varphi \text{ iff } F \models \alpha, \text{ for all Kripke frames } F.$$

Various examples of formulas in **M1** have occurred in the Introduction. For purposes of illustration, see Table 1 below.

As these are all rather easy to establish, some readers may desire a more complex example. Here it is, straight from the incompleteness Example 5 in the Introduction.

**THEOREM 37.** *The conjunction of the formulas  $\Box p \rightarrow p, \Box \Diamond p \rightarrow \Diamond \Box p$  and  $(\Diamond p \wedge \Box(p \rightarrow \Box p)) \rightarrow p$  is in **M1**.*

**Proof.** We shall show that this conjunction defines the same class as the classical axiom  $\Box p \leftrightarrow p$ , i.e.  $\forall xy(Rxy \leftrightarrow x = y)$ .

The argument requires several stages.

1.  $\Box p \rightarrow p$  imposes reflexivity,
2.  $\Diamond p \wedge \Box(p \rightarrow \Box p) \rightarrow p$  says the following:  
 $\forall xy(Rxy \rightarrow \exists n \exists z_1, \dots, z_n (Rxz_1 \wedge \dots \wedge Rxz_n \wedge$   
 $\wedge Ry z_1 \wedge \dots \wedge Rz_n x))$ .

In other words, from any  $R$ -successor  $y$  of  $x$ , one may return to  $x$  by way of some finite chain of  $R$ -successors of  $x$ . In case the chain is empty, this reduces to just:  $Ryx$ .

This (second-order!) equivalence is proved as follows (I. L. Humberstone): ‘ $\Rightarrow$ ’: Consider any  $y$  with  $Rxy$ . Let the *good points* be those  $R$ -successors  $z$

Table 1.

Modal formula	Condition
$\Box p \rightarrow p$	$\forall x Rxx$
$\Box p \rightarrow \Box \Box p$	$\forall xy (Rxy \rightarrow \forall z (Ryz \rightarrow Rxz))$
$\Diamond \Box p \rightarrow \Box \Diamond p$	$\forall xy (Rxy \rightarrow \forall z (Rxz \rightarrow \exists u (Ryu \wedge Rzu)))$
$\Box (p \vee q) \rightarrow \Box p \vee \Box q$	$\forall xy (Rxy \rightarrow \forall z (Rxz \rightarrow z = y))$
$\Box (\Box p \rightarrow q) \vee \Box (\Box q \rightarrow p)$	$\forall xy (Rxy \rightarrow \forall z (Rxz \rightarrow (Ryz \vee Rzy)))$
$p \rightarrow \Box p$	$\forall xy (Rxy \rightarrow y = x)$
$\Box \perp$	$\forall x \neg \exists y Rxy$
$p \rightarrow \Box \Diamond p$	$\forall xy (Rxy \rightarrow Ryx)$

of  $x$  which can be reached from  $y$  through some finite chain (possibly empty) of  $R$ -successors of  $x$ . Then, set  $V(p)$  equal to the set of all  $R$ -successors of good points. This assignment produces the following effects.

1.  $p$  is true at  $y$  ( $y$  being a successor of  $y$ , by reflexivity), and, hence,  $\Diamond p$  is true at  $x$ .
2. Any  $R$ -successor of  $x$  verifying  $p$  is itself a good point, whence all *its*  $R$ -successors belong to  $V(p)$ .

It follows that  $\Box(p \rightarrow \Box p)$  is true at  $x$ . Therefore,  $p$  itself must be true at  $x$ : i.e.  $x$  is  $R$ -successor of some good point, which was precisely to be proved.

‘ $\Leftarrow$ ’: Truth of  $p$  in  $x$  is discovered by merely following the relevant chain.

3. Now, having secured *reflexivity* and ‘*safe return*’, we can find out what the McKinsey Axiom says in the present context.

First, notice that all  $R$ -successors of any point  $x$  may be divided up into concentric shells  $S_n(x)$ , where  $S_n(x)$  consists of those  $R$ -successors  $y$  of  $x$  which return to  $x$  by  $n$   $R$ -arrows (between  $R$ -successors of  $x$ ) but no less. For instance,  $S_0(x)$  only consists of  $x$  itself,  $S_1(x)$  contains immediate  $R$ -predecessors. Notice also that, if  $y \in S_{n+1}(x)$ , then it must have some  $R$ -successor in  $S_n(x)$ .

The McKinsey Axiom makes this whole hierarchy collapse. Set  $V(p) = \cup \{S_{2n}(x) \mid n = 0, 1, 2, \dots\}$ . Then  $\Box \Diamond p$  will be true at  $x$ , as follows from the above picture. For, if  $Rxy$ , and  $y \in S_n(x)$ , then either  $n$  is even — whence  $p$  holds at  $y$  (by definition) and so  $\Diamond p$  (by reflexivity), or  $n$  is odd — whence  $y$  has an  $R$ -successor in  $S_{n-1}(x)$  verifying  $p$ : which again verifies  $\Diamond p$  at  $y$ .

It follows that  $\Diamond \Box p$  must be true at  $x$ . So,  $\Box p$  holds at some  $R$ -successor of  $x$ . Which one? In the present situation, this can only be

*x itself.* But then again, this means that there can be no shells  $S_n(x)$  with  $n$  odd. Thus, there is only  $S_0(w) : \forall y(Rxy \rightarrow y = x)$ .

4. Combining (1) and (3), the required conclusion follows: the three axioms together imply  $\forall xy(Rxy \leftrightarrow y = x)$ , and are obviously implied by it. ■

The very unexpectedness of this argument will have made it clear that there is a creative side to establishing correspondences.

*Global and local definability.* Originally, Kripke introduced frames  $\langle W, R, w_0 \rangle$ , with a designated ‘actual world’  $w_0$ . From that point of view, the study of ‘local’ equivalence becomes natural:

$$F \vDash \varphi[w] \text{ iff } \vDash \alpha[w],$$

where the first-order formula  $\alpha$  has one free variable now. The reader may have noticed already that previous correspondence arguments often provide local versions as well. For instance, we had

$$\begin{aligned} F \vDash \Box p \rightarrow p[w] & \quad \text{iff} \quad F \vDash Rxx[w] \\ F \vDash \Box p \rightarrow \Box \Box p[w] & \quad \text{iff} \quad F \vDash \forall y(Rxy \rightarrow \forall z(Ryz \rightarrow Rxz))[w]. \end{aligned}$$

The local notion is the more informative one, in that local correspondence of  $\varphi$  with  $\alpha(x)$  implies global correspondence of  $\varphi$  with  $\forall x\alpha(x)$ ; but not conversely. Indeed, [van Benthem, 1976] contains an example of a formula in **M1** which has no local first-order equivalent at all! On the other hand, there are also circumstances under which the distinction collapses — e.g. on the *transitive* Kripke frames (W. Dziobiak; cf. [van Benthem, 1981a]).

Finally, a word of warning. Local validity of, e.g.  $\Box p \rightarrow \Box \Box p$  means ‘local transitivity’, no more. The frame  $\langle N, \{(0, n) \mid n \in N\} \cup \{(n, n+1) \mid n \in N\} \rangle$  is locally transitive in 0, without being transitive.

*First-order undefinability.* There is a threshold of complexity below which second-order phenomena do not occur.

**THEOREM 38.** *All modal formulas without nestings of modal operators are in M1.*

**Proof.** Cf. [van Benthem, 1978]: a combinatorial classification suffices. ■

**EXAMPLE 39.** Löb’s Axiom  $\Box(\Box p \rightarrow p) \rightarrow \Box p$  is outside of **M1**.

**Proof.** It suffices to establish the following Claim: Löb’s Axiom defines transitivity plus well-foundedness of the converse of the alternative relation (i.e. there are no ascending sequences  $xRx_1Rx_2Rx_3, \dots$ ). For, by a well-known classical compactness argument, the latter combination cannot be first-order definable (e.g. notice that it holds in  $\langle N, > \rangle$ , but not in its non-isomorphic ultrapowers).

First, assume that Löb’s Axiom fails in  $F$ ; i.e. for some  $V$  and  $w$ ,

1.  $\langle F, V \rangle \models \Box(\Box p \rightarrow p)[w]$ , but
2.  $\langle F, V \rangle \not\models \Box p[w]$

Also, assume transitivity of  $R$ : we will refute the well-foundedness of  $\check{R}$ , by constructing an endless ascending sequence of worlds  $wRw_1Rw_2\dots$

Step 1: Chose any  $w_1$  with  $Rw_1w_1$  where  $p$  fails (by (2)). By (1),  $\Box p \rightarrow p$  is true at  $w_1$ , whence  $\Box p$  fails again.

Step 2: chose any  $w_2$  with  $Rw_1w_2$  where  $p$  fails. By (1) and transitivity,  $\Box p \rightarrow p$  is true at  $w_2$ , etcetera: an endless sequence is on its way.

Next, failure of either of the two relational conditions results in failure of Löb's Axiom. If transitivity fails, say  $Rwv, Rvu, \neg Rwu$ , then  $V(p) = W - \{v, u\}$  verifies  $\Box(\Box p \rightarrow p)$  at  $w$ , while falsifying  $\Box p$ .

If well-foundedness fails, say  $wRw_1Rw_2, \dots$ , then  $V(p) = W - \{w, w_1, w_2, \dots\}$  produces the same effect. ■

More complex undefinability arguments will be discussed later on.

*First-order definability and ultraproducts.* Modal formulas could be regarded as  $\Pi_1^1$ -sentences, witness the Introduction. Now, for the latter sentences, ultraproducts provide the touchstone for first-order definability:

**THEOREM 40.** *A  $\Pi_1^1$ -sentence in  $R, =$  is first-order definable if and only if it is preserved under ultraproducts.*

**Proof.** ' $\Rightarrow$ ': This follows from the Łoś Equivalence (cf. Section 2.1).

' $\Leftarrow$ ': Consider a typical such sentence:

$$\forall P_1 \dots \forall P_n \varphi(P_1, \dots, P_n, R, =) \quad (\varphi \text{ first-order}).$$

Clearly it is preserved under *isomorphisms* (and so is its negation). Moreover, its negation (a ' $\Sigma_1^1$ -sentence') is preserved under *ultraproducts* (cf. [Chang and Keisler, 1973, Chapter 4.1], for the easy argument). So, given the assumption on the sentence itself, Keisler's Theorem (33) applies. ■

**COROLLARY 41.** *A modal formula is in **M1** if and only if it is preserved under ultraproducts.*

A second application says that no generalisation of our topic is obtained by allowing arbitrary *sets* of defining first-order conditions.

**COROLLARY 42.** *If a modal formula has a  $\Delta$ -elementary definition, it has an elementary definition.*

**Proof.**  $\Delta$ -elementary classes are closed under the formation of ultraproducts. ■

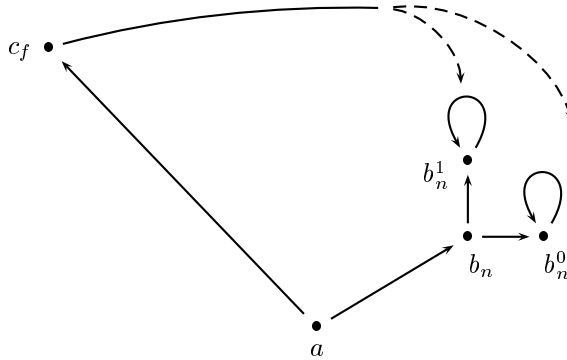
This characterisation of **M1** is rather aspecific, as it holds for *all*  $\Pi_1^1$ -sentences. Later on, we will exploit the specifically *modal* character of our formulas to do better. Moreover, the characterisation is rather abstract, as ultraproducts are hard to visualise. Therefore, we now turn to more concrete methods for separating formulas inside **M1** from those outside.

*Formulas beyond M1: Compactness and Löwenheim–Skolem arguments.* In practice, nonfirst-order definability often shows up in failure of the Compactness and Löwenheim–Skolem theorems. The first was involved in the example of Löb’s Axiom, the second will be presented now.

EXAMPLE 43 (McKinsey’s Axiom).  $\Box\Diamond p \rightarrow \Diamond\Box p$  is outside of **M1**.

**Proof.** Consider the following uncountably infinite Kripke frame

$$F = \langle W, R \rangle:$$



$$\begin{aligned} W &= \{a\} \cup \{b_n, b_n^0, b_n^1 \mid n \in N\} \cup \{c_f \mid f : N \rightarrow \{0, 1\}\} \\ R &= \{\langle a, b_n \rangle, \langle b_n, b_n^0 \rangle, \langle b_n, b_n^1 \rangle, \langle b_n^0, b_n^0 \rangle, \langle b_n^1, b_n^1 \rangle \mid n \in N\} \cup \\ &\quad \{\langle a, c_f \rangle \mid f : N \rightarrow \{0, 1\}\} \cup \{\langle c_f, b_n^{f(n)} \rangle \mid n \in N, f : N \rightarrow \{0, 1\}\}. \end{aligned}$$

We observe two things.

1.  $F \models \Box\Diamond p \rightarrow \Diamond\Box p$ .

Thanks to the presence of the reflexive endpoints  $b_n^0, b_n^1$ , the validity of the McKinsey Axiom is obvious everywhere, except for  $a$ .

So, suppose that, under some valuation  $V$ ,  $\Box\Diamond p$  is true at  $a$ . By assumption,  $\Diamond p$  is true at each  $b_n$ , and hence  $p$  is true at  $b_n^0$  or  $b_n^1$ . Now, pick any function  $f : N \rightarrow \{0, 1\}$  such that  $b_n^{f(n)}$  is a  $p$ -world (each  $n \in N$ ). Then  $\Box p$  holds at  $c_f$ , and hence  $\Diamond\Box p$  at  $a$ .

By the downward Löwenheim–Skolem theorem,  $F$  possesses a *countable* elementary substructure  $F'$  whose domain contains (at least)  $a, b_n, b_n^0, b_n^1$  (all  $n \in N$ ). As  $F$  is *uncountable*, many worlds ( $c_f$ ) must be missing in



$W'$ . Fix any one of these, say  $c_{f_0}$ . Notice, for a start, that  $c_{1-f_0}$  cannot be in  $W'$  either. (For, the existence of ‘complementary’  $c$ -worlds is first-order expressible; and  $F'$  verifies the same first-order formulas at each of its worlds as  $F$ .) Now, setting

$$V(p) = \{b_n^{f_0(n)} \mid n \in N\}$$

will verify  $\Box\Diamond p$  at  $a$ , while falsifying  $\Diamond\Box p$ . Thus, we have shown

2.  $F' \not\models \Box\Diamond p \rightarrow \Diamond\Box p$ .

We may conclude that the McKinsey Axiom is not first-order definable — not being preserved under elementary subframes. ■

In *practice*, failure of Löwenheim–Skolem or compactness properties is an infallible mark of being outside of **M1**. The reader may also think this to be the case in *theory*, by the famous Lindström Theorem. (Cf. Volume 1, chapters by Hodges or van Benthem and Doets.) But there is a little-realised problem: the Lindström Theorem does not work for languages with a fixed finite vocabulary (cf. [van Benthem, 1976]). In our case of  $R, =$ , there do exist proper extensions of predicate logic satisfying both the Löwenheim and compactness properties. These are not *modal* examples, however — and it may well be the case, for all we know, that a modal formula  $\varphi$  belongs to **M1** if and only if the logic obtained by adding  $\varphi$  to the first-order predicate logic in  $R, =$  as a propositional constant has the Löwenheim and compactness properties. Indeed, up till now, all undefinability arguments (including the above) have always been found reducible to *compactness* arguments alone.

*The final characterisation of M1.* Corollary 41 may be improved by noting the following fact about Kripke frames, connecting the modal and classical notions of Section 2.1.

LEMMA 44.  $\Pi_U F_i \xrightarrow{C} \Pi_U \oplus \{F_i \mid i \in I\}$ .

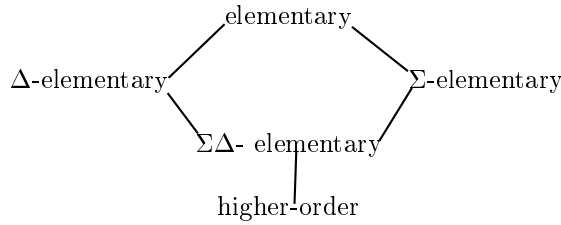
Thus, ultraproducts are generated subframes of suitable ultrapowers.

A second idea comes from the preceding section: outside of **M1**, we encountered non preservation under *elementary equivalence*, a notion tied up with ultrapowers by the Keisler–Shelah Theorem (cf. [Chang and Keisler, 1973, Chapter 6.1]). We arrive at the main result of [van Benthem, 1976].

THEOREM 45. (i) A modal formula is in **M1** if and only if (ii) it is preserved under ultrapowers if and only if (iii) it is preserved under elementary equivalence.

**Proof.** (i)  $\Rightarrow$  (iii)  $\Rightarrow$  (ii) are immediate. (ii)  $\Rightarrow$  (i): If  $\varphi$  is preserved under ultrapowers, then, by Lemma 44, it is also preserved under ultraproducts — because disjoint unions preserve modal truth (Corollary 22). Now apply Corollary 41. ■

Again, this insight saves us some spurious generalisations. Besides ‘ $\Delta$ -elementary’, there are two more levels in the definability hierarchy



A  $\Sigma$ -elementary class is defined by an infinite *disjunction* of first-order sentences ( $\Delta$ -elementary classes by infinite *conjunctions*). The prime example of this phenomenon is *finiteness*.  $\Sigma\Delta$ -elementary classes arise from infinite disjunctions of infinite conjunctions, or vice versa: both cases (and all purported ‘higher’ ones) collapse — and the hierarchy stops here, even in classical logic. The reason lies in the simple observation that a class of frames is  $\Sigma\Delta$ -elementary if and only if it is closed under *elementary equivalence*.

But the preceding result has a

**COROLLARY 46.** *Modal formulas are either elementary, or essentially higher-order.*

Unfortunately, even this better characterisation does not yield much *effective* information concerning the members of **M1**. For, there are no syntactic criteria for preservation under ultrapowers. From [van Benthem, 1983], we will cite the catalogue of what little we know.

**DIGRESSION 47.**

1.  $\Pi_1^1$ -sentences in  $R, =$  of the purely universal form

$$\forall P_1 \dots \forall P_m \forall x_1 \dots \forall x_n \varphi \quad (\varphi \text{ quantifier-free})$$

are preserved under ultrapowers. This tells us that  $p \rightarrow \Box p$ , i.e.

$$\forall P \forall x (Px \rightarrow \forall y (Rxy \rightarrow Py))$$

must be in **M1**: but that was clear without such heavy artillery.

2.  $\Pi_1^1$ -sentences in  $R, =$  of the universal-existential form

$$\forall P_1 \dots \forall P_m \exists x_1 \dots \exists x_n \varphi \quad (\varphi \text{ quantifier-free})$$

are preserved under ultrapowers. This is of no help whatsoever, as modal formulas have at least one *universal* first-order quantifier ( $\forall x$ ).

3. Further presents will not be forthcoming: *any*  $\Pi_1^1$ -sentence in  $R, =$  is logically equivalent to one of the form

$$\forall P_1 \dots \forall P_m \forall x_1 \dots \forall x_n \exists y_1 \dots \exists y_n \varphi \quad (\varphi \text{ quantifier-free})$$

So, all complexity occurs at this level already.

Thus, other ways are to be developed for describing **M1** effectively.

*The method of substitutions.* There is a common syntactic pattern to many examples of first-order definable modal formulas: certain antecedents, in combination with certain consequents enable one to ‘read off’ equivalents. Starting from the earlier examples  $\Box p \rightarrow p, \Diamond \Box p \rightarrow \Box \Diamond p$ , one may notice successively that conjunctions and disjunctions are admissible as well; as long as one avoids  $\Box \Diamond$  or  $\Box(\dots \vee \dots)$  combinations to the left.

A typical instance is the following result from [Sahlqvist, 1975]:

**THEOREM 48.** *Modal formulas  $\varphi \rightarrow \psi$  are in **M1**, provided that*

1.  $\varphi$  is constructed from the forms  $p, \Box p, \Box \Box p, \dots, \perp, \top$ , using only  $\wedge, \vee$  and  $\Diamond$ , while
2.  $\varphi$  is constructed from proposition letters,  $\perp, \top$ , using  $\wedge, \vee, \Diamond$  and  $\Box$ .

This theorem accounts for cases such as

$$\Diamond(p \wedge \Box q) \rightarrow \Box(p \vee \Diamond p \vee q)$$

which defines

$$\forall xy(Rxy \rightarrow \forall z(Rxz \rightarrow (z = y \vee Rzy \vee Ryz))).$$

**Proof.** The heuristics of the Introduction works: for each ‘minimal verification’ of the antecedent, the consequent must hold. For further technical information (e.g. the *monotonicity* of the consequent is vital too), cf. [van Benthem, 1976], which also contains generalisations of the theorem. ■

That  $\Box \Diamond$  is fatal, is shown by the McKinsey Axiom. The Fine Axiom  $\Diamond \Box(p \vee q) \rightarrow \Diamond(\Box p \vee \Box q)$  does the same for  $\Box(\dots \vee \dots)$ . Finally, the Löb Axiom (in the equivalent form  $\Diamond p \rightarrow \Diamond(p \wedge \Box \neg p)$ ) demonstrates the danger of ‘negative’ parts in the consequent. Thus, in a sense, we have a ‘best result’ here.

Notice that the class described is rather typical for modal axioms, which often assume this implicational form. Indeed, the most characteristic modal axioms are even simply *reduction principles* of the form

$$(\text{modal operators}) p \rightarrow (\text{modal operators}) p.$$

**THEOREM 49.** *A modal reduction principle is in **M1** if and only if it is of one of the following four types:*

1.  $\vec{M}p \rightarrow \Box \dots \Box \Diamond \dots \Diamond p$ ,
2.  $\Diamond \dots \Diamond \Box \dots \Box p \rightarrow \vec{M}p$ ,
3.  $\Box \dots (i \text{ times}) \dots \Box \vec{M}p \rightarrow \vec{N} \vec{M}p$  (where  $\text{length}(\vec{N}) = i$ ),

4.  $\vec{N}\vec{M}p \rightarrow \diamond \dots$  (*i times*)  $\dots \diamond \vec{M}p$  (where length  $(\vec{N}) = i$ ).

**Proof.** Cf. [van Benthem, 1976] for the rather laborious argument. ■

Thus at least, important parts of **M1** have been classified. This particular theorem finishes a project begun in [Fitch, 1973].

A general method of proof for Theorem 48 consists of the method of substitutions, introduced in the introduction. Here we shall merely illustrate how it works: a justification may be found in [van Benthem, 1983].

EXAMPLE 50. Write  $\diamond \Box p \rightarrow \Box \diamond p$  as

$$\forall P \forall x (\exists y (Rxy \wedge \forall z (Ryz \rightarrow Pz)) \rightarrow \forall u (Rxu \rightarrow \exists v (Ruv \wedge Pv))).$$

Rewrite this to the equivalent

$$\forall xy (Rxy \rightarrow \forall P (\forall z (Ryz \rightarrow Pz) \rightarrow \forall u (Rxu \rightarrow \exists v (Ruv \wedge Pv)))).$$

Substitute for  $P : \lambda z. Ryz$ , to obtain

$$\forall xy (Rxy \rightarrow (\forall z (Ryz \rightarrow Ryz) \rightarrow \forall u (Rxu \rightarrow \exists v (Ruv \wedge Ryv)))).$$

This is equivalent to

$$\forall xy (Rxy \rightarrow \forall u (Rxu \rightarrow \exists v (Ruv \wedge Ryv))),$$

i.e. *directedness (confluence)*.

Write  $\diamond (p \wedge \Box q) \rightarrow \Box (p \vee \diamond p \vee q)$  as

$$\forall xy (Rxy \rightarrow \forall P ((Py \wedge \forall z (Ryz \rightarrow Qz)) \rightarrow \forall u (Rxu \rightarrow (Pu \vee \exists v (Ruv \wedge Pv) \vee Qu)))).$$

Substitute for  $P : \lambda z. y = z$ , and for  $Q : \lambda z. Ryz$ , to obtain (an equivalent of) the earlier *connectedness*.

Write  $\diamond (p \wedge \Box p) \rightarrow p$  as

$$\forall xy (Rxy \rightarrow \forall P ((Py \wedge \forall z (Ryz \rightarrow Pz)) \rightarrow Px)).$$

Substitute for  $P : \lambda z. y = z \vee Ryz$ , to obtain (an equivalent of)

$$\forall xy (Rxy \rightarrow (Ryx \vee y = x)).$$

Write  $\Box \Box p \rightarrow \Box p$  as

$$\forall x \forall P (\forall y (Rxy \rightarrow \forall z (Ryz \rightarrow Pz)) \rightarrow \forall u (Rxu \rightarrow Pu)).$$

Substitute for  $P : \lambda z \cdot R^2xz$ ; i.e.  $\lambda z \cdot \exists v(Rxv \wedge Rvz)$ , to obtain (modulo logical equivalence)

$$\forall x \forall u (Rxu \rightarrow \exists v (Rxv \wedge Rvu)),$$

i.e., *density* of the alternative relation.

In general, substitutions will be disjunctions of forms  $R^n yz$  ( $n = 0, 1, 2, \dots$ ); the cases 0, 1 standing for  $=, R$ , respectively.

Despite these advances, the range of the method of substitutions has its limits. To see this, here is an example of a formula in **M1** with a quite different spirit.

EXAMPLE 51. The conjunction of the **K4.1** axioms, i.e.  $\Box p \rightarrow \Box \Box p$ ,  $\Box \Diamond p \rightarrow \Diamond \Box p$  is in **M1**.

**Proof.**  $\Box p \rightarrow \Box \Box p$  defined transitivity and, therefore, it suffices to prove the following

CLAIM. On the *transitive* Kripke frames, McKinsey's Axiom defines *atomicity*:

$$\forall x \exists y (Rxy \wedge \forall z (Ryz \rightarrow z = y)).$$

From right to left, the implication is clear. From left to right, however, the argument runs deeper.

Assume that  $F$  is a transitive frame, containing a world  $w \in W$  such that

$$\forall y (Rwy \rightarrow \exists z (Ryz \wedge z \neq y)).$$

Using some suitable form of the Axiom of Choice (it is as serious as this ...), find a subset  $X$  of  $w$ 's  $R$ -successors such that

1.  $\forall y \in W (Rwy \rightarrow \exists z \in X Ryz)$
2.  $\forall y \in W (Rwy \rightarrow \exists z \in (W - X) Ryz)$ .

Setting  $V(p) = X$  then falsifies the McKinsey Axiom at  $w$ . ■

This complexity is unavoidable. We can, for example, prove

THEOREM 52.  $(\Box p \rightarrow \Box \Box p) \wedge (\Box \Diamond p \rightarrow \Diamond \Box p)$  is not equivalent to any conjunction of its first-order substitution instances.

**Proof.** Here is where the earlier general frame  $\langle N, \leq$ , finite and cofinite sets) comes in. First, an ordinary model-theoretic

OBSERVATION. The finite and cofinite sets of natural numbers are precisely those first-order definable in  $\langle N, \leq \rangle$ , possibly using parameters.

Now, it was noticed already in Section 2.1 that the above formula holds in this general frame — and hence so do all its first-order substitution

instances. But the latter also hold in the full frame  $\langle N, \leq \rangle$ . So, if our formula were defined by them, it would also hold in the full frame: which it does not. ■

So, although the method of substitutions carves out a large, and important part of **M1**, it does not fully describe the latter class.

*The complexity of M1.* The method of substitutions describes a part of **M1** which may even be shown to be *recursively enumerable* (cf. [van Benthem, 1983]). But **M1** overflowed its boundaries. Indeed, there are reasons to believe that **M1** is not recursively enumerable — probably not even arithmetically definable. For, in the general case of  $\Pi_1^1$ -sentences, we know

**THEOREM 53.** *First-order definability of  $\Pi_1^1$ -sentences is not an arithmetical notion.*

**Proof.** (Cf. [van Benthem, 1983] or the Higher Order Logic Chapter in Volume 1 of this *Handbook*.) ■

*Other topics.* Various other questions had to be omitted here. At least, one example should be mentioned, viz. that of *relative correspondences*. On several occasions, a restriction to *transitive* Kripke frames produced interesting shifts: global and local first-order definability collapse, the McKinsey Axiom becomes elementary, etc. A sample result is in [van Benthem, 1976].

**THEOREM 54.** *On the transitive Kripke frames, all modal reduction principles are first-order definable.*

Thus, ‘pre-conditions’ on the alternative relation are worth considering. In areas such as tense logic, our temporal intuitions even *require* them.

### 2.3 Modal Algebra

An alternative to Kripke semantic structures is offered by so-called ‘*modal algebras*’, in which the modal language may be interpreted as well. The realm of modal algebras has its own mathematical structure, with *subalgebras*, *direct products* and *homomorphic images* as key notions. Now, back-and-forth connections may be established between these two realms, through the *Stone Representation*. A categorial parallel emerges between the above triad of notions and the basic triad of Section 2.1: zigzag-morphic images, disjoint unions and generated subframes, respectively. Moreover, the earlier ‘possible worlds construction’ for ultrafilter extensions will be seen to arise naturally from the Stone Representation.

*The algebraic perspective.* As in other areas of logic, the modal propositional language may also be interpreted in *algebraic* structures. These assume the

form of a Boolean Algebra (needed to interpret the propositional base) enriched with a unary operation, in order to capture the modal operator.

DEFINITION 55. A *modal algebra* is a tuple

$$\mathfrak{A} = \langle A, 0, 1, +, ', * \rangle,$$

where  $\langle A, 0, 1, +, ' \rangle$  is a Boolean Algebra and  $*$  is a unary operator satisfying the equations

1.  $(x + y)^* = x^* + y^*$
2.  $0^* = 0$ .

Notice that  $*$  corresponds to possibility ( $\diamond$ ): the necessity choice would have yielded equations

- 1'.  $(x \cdot y)^* = x^* \cdot y^*$
- 2'.  $1^* = 1$ .

This algebraic perspective at once yields a completeness result.

THEOREM 56. *A modal formula is derivable in the minimal modal logic  $\mathbf{K}$  if and only if it receives value 1 in all modal algebras under all assignments.*

The concept of evaluation at the back of this goes as follows. Let  $V$  assign  $A$ -values to proposition letters. Then,  $V$  may be lifted to all formulas through the recursive clauses

$$\begin{aligned} V(\neg\varphi) &= V(\varphi)' \\ V(\varphi \vee \psi) &= V(\varphi) + V(\psi) \\ V(\diamond\varphi) &= V(\varphi)^*, \text{ etc.} \end{aligned}$$

Thus, a modal formula is read as a ‘polynomial’ in  $', +, *$ .

The *proof* of the completeness Theorem 56 comes cheap. First, one shows by induction on the length of proofs that all  $\mathbf{K}$ -theorems are ‘polynomials identical to 1’. Conversely, one considers the so-called *Lindenbaum Algebra* of the modal language, whose elements are equivalence classes of  $\mathbf{K}$ -provably equivalent modal formulas, with operations defined in the obvious way through the connectives. The value 1 in this algebra is awarded to all and only the  $\mathbf{K}$ -theorems: hence non- theorems are disqualified as polynomials identical to 1.

Such uses of modal algebra are a joy to some (cf. [Rasiowa and Sikorski, 1970]); to others they show that the algebraic approach is merely ‘syntax in disguise’. After all, the above result may be viewed as a re-axiomatisation of  $\mathbf{K}$ , no more. For instance, notice that the hard work in the usual (Henkin type) model-theoretic completeness theorems consists in showing that non-theorems can be refuted in *set-theoretic* (Kripke)-models. To put this into a slogan, which will become fully comprehensible at the end of this chapter:

## HENKIN = LINDENBAUM + STONE.

Nevertheless, the algebraic perspective has further uses, which are being discovered only gradually. First, notice that it offers a more general framework than Kripke semantics. For the above Lindenbaum construction to work, one only needs the principle of Replacement of Equivalents; i.e. modally, closure under the rule

$$\text{if } \vdash \varphi \leftrightarrow \psi, \text{ then } \vdash \diamond\varphi \leftrightarrow \diamond\psi.$$

(Algebraically, this just amounts to an identity axiom.)

The above additional equations represent optional further choices.

But even in the realm of the above modal algebra, there exists a whole discipline of universal algebraic notions and results, which turn out to be applicable to modal logic in surprising ways. Two instructive references are [Goldblatt, 1979] and [Blok, 1976]. Here we shall only skim the surface, taking what is needed for the modal definability results of Section 2.4. Thus, we shall need the following three fundamental algebraic notions.

DEFINITION 57.  $\mathfrak{A}_1$  is a modal *subalgebra* of  $\mathfrak{A}_2$  if  $A_1 \subseteq A_2$ , and the operations of  $\mathfrak{A}_2$  coincide with those of  $\mathfrak{A}_1$  on  $A_1$ .

DEFINITION 58. The *direct product*  $\prod\{\mathfrak{A}_i \mid i \in I\}$  of a family of modal algebras  $\{\mathfrak{A}_i \mid i \in I\}$  consists of all functions in the Cartesian product  $\prod\{A_i \mid i \in I\}$ , with operations defined component-wise:

$$f + g = (f(i) +_i g(i))_i, \quad f^* = (f(i)^*_i)_i; \text{ etc.}$$

DEFINITION 59. A function  $f$  is a *homomorphism* from  $\mathfrak{A}_1$  to  $\mathfrak{A}_2$  if it respects all operations:

$$f(a +_1 b) = f(a) +_2 f(b), \quad f(a^{*1}) = f(a)^{*2}; \text{ etc.}$$

These three operations are fundamental in algebra because they characterise algebraic *equational definability*. This is the content of ‘Birkhoff’s Theorem’:

A class of algebras is defined by the validity of a certain set of algebraic equations (under all assignments) if and only if that class is closed under the formation of subalgebras, direct products and homomorphic images. (For a proof, cf. [Grätzer, 1968].) There is much more to Universal Algebra, of course, but this is what we shall need in the sequel.

*Kripke frames induce modal algebras.* In order to tap the above resources, a systematic connection is needed between the earlier semantic structures and modal algebras.

To begin with, each Kripke frame  $F = \langle W, R \rangle$  gives rise to the following modal algebra

$$A(F) = \langle P(W), \emptyset, W, \cup, -, \pi \rangle$$



where  $\pi$  is the *modal projection* of 2.1:

$$\pi(X) = \{w \in W \mid \exists v \in X R w v\} \quad (X \subseteq W).$$

As for truth of modal formulas, it is immediate that a modal formula  $\varphi$  is true in  $F$  if and only if its corresponding modal equation  $a(\varphi)$  is identical to 1 in the algebra  $A(F)$ . For instance, truth of

$$\diamond \square(p \vee q) \rightarrow \diamond(\square p \vee \square q),$$

or equivalently

$$\neg \diamond \neg \diamond \neg(p \vee q) \vee \diamond(\neg \diamond \neg p \vee \neg \diamond \neg q)$$

is equivalent to the validity of the identity

$$(x + y)'^{**'} + (x'^{**'} + y'^{**'})^* = 1.$$

Thus,  $A$  maps single Kripke frames to modal algebras. But what happens to the characteristic modal connections between frames, as in Section 2.1? We shall take them one by one.

First, if  $F_1$  is a *generated subframe* of  $F_2$ , then the obvious restriction map sending  $X \subseteq W_2$  to  $X \cap W_1$  is a modal *homomorphism* from  $A(F_2)$  onto  $A(F_1)$ . (The key observation is that  $R_2$ -closure of  $W_1$  guarantees homomorphic respect for the projection operator  $\pi$ .) Next, the algebra induced by a *disjoint union*  $\oplus\{F_i \mid i \in I\}$  is isomorphic, in a natural way, to the *direct product*  $\prod\{A(F_i) \mid i \in I\}$ . One simply associates a set  $X$  of worlds in the former with the function  $(X \cap W_i)_{i \in I}$ .

Finally, and this happy ending will be predictable by now, if  $F_2$  is a zigzag-morphic image of  $F_1$  through  $f$ , then the stipulation

$$A(f)(X) =_{\text{def}} f^{-1}[X]$$

defines an isomorphism between  $A(F_2)$  and a *subalgebra* of  $A(F_1)$ . (This time, the two relational clauses in the definition of ‘zigzag morphism’ ensure that  $A(f)$  respects projections.) Notice the reversal in direction in the latter case: this is a common phenomenon in these ‘categorical connections’.

*Modal algebras induce Kripke structures.* There is a road back. Conversely, modal algebras may be ‘represented’ as if they had come from an underlying base frame. The idea of this so-called *Stone Representation* is as follows. (It is due to Jónsson and Tarski around 1950.)

Worlds  $w$  are to be created such that an element  $a$  in the algebra may be thought of as the set of  $w$  ‘in  $a$ ’. But then, the desired correspondence between algebraic and set-theoretic operations becomes:

$$\begin{aligned} \text{no set } w \text{ is in } 0, \text{ all sets } w \text{ are in } 1, \\ w \text{ is in } a + b \quad \text{iff} \quad w \text{ is in } a \text{ or } w \text{ is in } b, \\ w \text{ is in } a' \quad \text{iff} \quad w \text{ is not in } a. \end{aligned}$$

Thus, as  $w$  searches through  $A$  ‘where it belongs’, it picks out a set  $X$  such that

$$\begin{aligned} 0 &\notin X, & 1 &\in X, \\ a + b \in X &\text{ iff } & a \in X &\text{ or } b \in X, \\ a' \in X &\text{ iff } & a &\notin X. \end{aligned}$$

Such sets  $X$  are called *ultrafilters* on  $\mathfrak{A}$ . Thus, let

$$W(\mathfrak{A}) = \text{all ultrafilters on } \mathfrak{A}.$$

A suitable alternative relation may be found through the same motivation as in Section 2.1.

$$\langle w, v \rangle \in R(\mathfrak{A}) \text{ iff for each } a \in A, \text{ if } a \in v, \text{ then } a^* \in w.$$

So, each modal algebra  $\mathfrak{A}$  induces a Kripke frame

$$F(\mathfrak{A}) = \langle W(\mathfrak{A}), R(\mathfrak{A}) \rangle.$$

This time, truth in  $\mathfrak{A}$  and truth in  $F(\mathfrak{A})$  need not correspond, however. For,  $F(\mathfrak{A})$  may harbour many more sets of worlds than just those corresponding to the elements  $a$  of the algebra — and hence it contains additional potential falsifiers. Thus, the implication goes only one way. The equation  $t_1 = t_2$  is valid in  $\mathfrak{A}$ , where the polynomials  $t_1, t_2$  correspond to the modal formulas  $\varphi_1, \varphi_2$ , when  $\varphi_1 \leftrightarrow \varphi_2$  is true in  $F(\mathfrak{A})$ . A complete equivalence is only restored by changing  $F(\mathfrak{A})$  to the *general frame*

$$F(\mathfrak{A}) = \langle W(\mathfrak{A}), R(\mathfrak{A}), \mathfrak{W}(\mathfrak{A}) \rangle,$$

where  $\mathfrak{W}(\mathfrak{A})$  consists of all sets of the form

$$\{w \in W(\mathfrak{A}) \mid a \in w\} \quad (a \in A).$$

So, what we now get is a two-way connection between *modal algebras* and *general frames* — and here lies the genesis of the latter notion. Two ways; for, it is easily seen that all previous insights about the mapping  $A$  apply equally well to general frames, instead of merely ‘full’ frames.

Again, the interest of the present connection may be gauged by seeing what happens to the three fundamental algebraic operations when translated through  $F$  into Kripke-semantic terms.

First, if  $\mathfrak{A}_1$  is a modal *subalgebra* of  $\mathfrak{A}_2$ , then the obvious restriction map sending ultrafilters  $w$  on  $\mathfrak{A}_2$  to ultrafilters  $w \cap A_1$  on  $\mathfrak{A}_1$  is a *zigzag morphism* from  $F(\mathfrak{A}_2)$  onto  $F(\mathfrak{A}_1)$ .

Next, the *direct product* of a family  $\{\mathfrak{A}_i \mid i \in I\}$  has an  $F$ -image containing the *disjoint union*  $\oplus \{F(\mathfrak{A}_i) \mid i \in I\}$ . No isomorphism need obtain, however: a slight flaw in our correspondence.

But finally, if  $\mathfrak{A}_2$  is a *homomorphic image* of  $\mathfrak{A}_1$  through  $f$ , then the map  $F(f)$ , defined by setting

$$F(f)(w) =_{\text{def}} f^{-1}[w],$$

sends  $\mathfrak{A}_2$ -ultrafilters to  $\mathfrak{A}_1$ -ultrafilters, in such a way that it embeds  $F(\mathfrak{A}_2)$  isomorphically as a *generated subframe* of  $F(\mathfrak{A}_1)$ .

*Back and forth.* So far, so good. Modal algebras induce general frames, and these, in their turn, induce modal algebras. But, what happens on a return-trip?

One case is simple, by construction:

**THEOREM 60.**  *$A(F(\mathfrak{A}))$  is isomorphic to  $\mathfrak{A}$ .*

The converse direction is more difficult. ( $F(A(G))$  need not be isomorphic to  $F$ , for general frames  $G$ . This is precisely what we noted in connection with ‘possible world constructions’ in Section 2.1. But, as was announced there, it can be ascertained which conditions on general frames  $G$  do guarantee such an isomorphism.

**DEFINITION 61.** A general frame  $G = \langle W, R, \mathfrak{W} \rangle$  is *descriptive* if it satisfies *Leibniz’ Principle for identity*:

$$1. \forall xy \in W (x = y \leftrightarrow \forall Z \in \mathfrak{W} (x \in Z \leftrightarrow y \in Z))$$

as well as *Leibniz’ Principle for alternatives*:

$$2. \forall xy \in W (Rxy \leftrightarrow \forall Z \in \mathfrak{W} (y \in Z \rightarrow x \in \pi(Z))).$$

Moreover, it should satisfy *Saturation*:

3. each subset of  $\mathfrak{W}$  with the finite intersection property has a non-empty total intersection.

The following basic result is in [Goldblatt, 1979].

**THEOREM 62.**  *$F(A(G))$  is isomorphic to  $G$  if and only if  $G$  is descriptive.*

The standard examples of descriptive frames are the general frames derived from *Henkin models* in modal completeness proofs, by taking for  $\mathfrak{W}$  the range of modally definable sets of worlds. It may also be noticed that general frames  $G$  which are themselves of the form  $F(\mathfrak{A})$  are always descriptive. Thus, for certain theoretical purposes, the ‘proper’ bijective correspondence may be said to be that between modal algebras and descriptive frames, which are ‘stable’ under the possible worlds construction described in Section 2.1.

*The categorial connection.* The above connections between modal algebras and Kripke structures run deeper than might appear at first sight. The

general picture is that of two mathematical worlds, or ‘categories’, which turn out to be quite similar in structure:

(Modal algebras, homomorphisms into)  
(General frames, zigzag morphisms into).

The earlier considerations may be summed up in the following two schemata:

$$\begin{array}{ccc}
 G_1 & \xrightarrow{f} & G_2 \\
 \downarrow & & \downarrow \\
 A(G_1) & \xleftarrow{A(f)} & A(G_2)
 \end{array}
 \qquad
 \begin{array}{ccc}
 \mathfrak{A}_1 & \xrightarrow{f} & \mathfrak{A}_2 \\
 \downarrow & & \downarrow \\
 F(\mathfrak{A}_1) & \xleftarrow{F(f)} & F(\mathfrak{A}_2)
 \end{array}$$

So,  $A, F$  are what a category theorist would call ‘contravariant’ functors. Therefore, information concerning the one category may sometimes be transferred to the other. Thus, a ‘categorical transfer’ arises, of which we mention a few phenomena. (The following passage can be skipped by readers unfamiliar with Category Theory or Universal Algebra).

The category of modal algebras has among its internal limit constructions the formation of *terminals* (viz. the degenerate single point algebras) and *pull-backs*. Hence, it is closed under *finite limits* in general. Through  $A, F$ , we may derive that the category of general frames is closed under *finite co-limits*, specifically under *initials* (allowing the *empty* frame) and *push-outs*. (In this connection, the ‘adjointness’ behaviour of  $A, F$  may be investigated.) The preservation behaviour of modal formulas under such limit constructions remains to be studied.

An algebraically well-motivated notion is that of a *free algebra*. What corresponds to these in the realm of general frames? A surprising connection with modal completeness theory appears. The Stone representations of free algebras are essentially *Henkin general frames* (proposition letters correspond to free generators of the algebra). The latter structures were characterised semantically in [Fine, 1975], in terms of certain ‘universal embedding’ properties with respect to zigzag morphisms. This turns out to follow directly, as the dual of the ‘homomorphic extension’ definition of free algebras.

Our final example concerns another algebraic classic, the notion of a *subdirectly irreducible* modal algebra (used with great versatility in [Blok, 1976]). These turn out to correspond almost (not quite) to rooted general frames whose domain consists of one root world together with its  $R$ -successors, their  $R$ -successors, etcetera. The famous Birkhoff Theorem stating that

Every (modal) algebra is a subdirect product of subdirectly irreducibles,

may then be compared with the simple Kripke-semantic observation that

Every general frame is a zigzag-morphic image of the disjoint union of its rooted generated subframes.

These examples will have made it clear how the categorial connection between modal algebra and possible worlds semantics can be a very rewarding perspective.

#### 2.4 From Classical to Modal Logic

Reversing the direction of the earlier correspondence study (Section 2.2), there arises

**DEFINITION 63.**  $\mathbf{P1}$  is the set of all first-order sentences in  $R, =$  for which a modal formula exists defining the same class of Kripke frames.

All earlier examples of formulas in  $\mathbf{M1}$  also provide examples for  $\mathbf{P1}$ , of course. Therefore, here are some more general results straightaway.

Some methods exist for *proving* the existence of modal definitions.

**THEOREM 64.** *Each first-order sentence of the form  $\forall xU\varphi$ , where  $U$  is a (possibly empty) sequence of restricted universal quantifiers, of the form*

$$\forall u(Rvu \rightarrow \quad (\text{with } u, v \text{ distinct}))$$

*followed by a matrix  $\varphi$  of atomic formulas  $u = v, Ruv$  combined through  $\wedge, \vee$ , belongs to  $\mathbf{P1}$ .*

**Proof.** The relevant combinatorial argument is based on the heuristics explained in the introduction. Cf. [van Benthem, 1976]. ■

Some examples of formulas of this type are

$$\text{reflexivity: } \forall xRxx, \text{ transitivity: } \forall x\forall y(Rxy \rightarrow \forall z(Ryz \rightarrow Rxz))$$

and

$$\text{connectedness: } \forall x\forall y(Rxy \rightarrow \forall z(Rxz \rightarrow (Rzy \vee Ryz))).$$

*Disproving* definability proceeds through counter-examples to preservation behaviour.

**EXAMPLE 65.**

1.  $\exists xRxx$  is outside of  $\mathbf{P1}$ .

It holds in  $\langle\{0, 1\}, \{1, 1\}\rangle$ ; but not in its generated subframe  $\langle\{0\}, \emptyset\rangle$ .

2.  $\forall x\forall yRxy$  is outside of  $\mathbf{P1}$ .

It is preserved under generated subframes, but not under disjoint unions. On  $\langle\{0\}, \{0, 0\}\rangle$  and  $\langle\{1\}, \{1, 1\}\rangle$ , the relation is universal; but not on  $\langle\{0, 1\}, \{0, 0\}, \{1, 1\}\rangle$ .

3.  $\forall x\neg Rxx$  is outside of **P1**.

It is preserved under generated subframes and disjoint unions; but not under zigzag-morphic images, witness the Introduction.

4.  $\forall x\exists y(Rxy \wedge Ryy)$  is outside of **P1**.

It is preserved under all three operations mentioned up till now, but not inversely under the formation of ultrafilter extensions. It can be shown to hold in  $ue(\langle N, < \rangle)$ , while failing in  $\langle N, < \rangle$ .

An important general result is casting its shadows here [Goldblatt and Thomason, 1974]:

**THEOREM 66.** *An elementary class of Kripke frames is modally definable if and only if it is closed under the formation of generated subframes, disjoint unions and zigzag-morphic images, while its complement is closed under the formation of ultrafilter extensions.*

**Proof.** This argument is given in heuristic outline here, as it is one of the most elegant applications of algebraic results in modal semantics.

Evidently, modally definable classes of Kripke frames exhibit all the listed closure phenomena: the surprising direction leads from ‘closure’ to ‘definability’.

First, notice that one closure condition can be added for free, by an earlier result. Theorem 30 implies that our class  $\mathfrak{K}$  of frames is itself closed under the formation of ultrafilter extensions: if  $F \in \mathfrak{K}$ , then the relevant elementary equivalent  $F' \in \mathfrak{K}$  ( $\mathfrak{K}$  being elementary), and hence so is its zigzag-morphic image  $ue(F)$ .

Now the obvious strategy is to show that  $\mathfrak{K}$  equals  $\text{MOD}(\text{Th}_{\text{mod}}(\mathfrak{K}))$ , i.e. the class of Kripke frames verifying each modal formula which is valid throughout  $\mathfrak{K}$ . The nontrivial inclusion here requires us to show that

if  $F^* \models \text{Th}_{\text{mod}}(\mathfrak{K})$ , then  $F^* \in \mathfrak{K}$ , for every Kripke frame  $F^*$ .

And here is where an excursion into the realm of modal algebra will help.  $F^*$  verifies  $\text{Th}_{\text{mod}}(\mathfrak{K})$ , and hence  $A(F^*)$  verifies the equational theory of the class  $\{A(G) \mid G \in \mathfrak{K}\}$ . (Recall the earlier correspondence between modal formulas and polynomials.) By Birkhoff’s Theorem, in a suitable version, this implies that  $A(F^*)$  must be constructible as a *homomorphic image* of some *subalgebra* of some *direct product*  $\prod\{A(G_i) \mid i \in I\}$ , with  $G_i \in \mathfrak{K}$ . In a picture,

$$\begin{array}{ccc}
 & \text{surjective} & \\
 A(F^*) & \longleftarrow & \mathfrak{A} \subseteq \prod \{A(G_i) \mid i \in I\}. \\
 & \text{homomorphism} & 
 \end{array}$$

Now the latter algebra is isomorphic to  $A(\oplus\{G_i \mid i \in I\})$ , by the earlier duality. Moreover, the latter disjoint union belongs to  $\mathfrak{K}$  — by the given closure conditions. So, the picture becomes, for some  $G \in \mathfrak{K}$ :

$$\begin{array}{ccc}
 & \text{surjective} & \\
 A(F^*) & \longleftarrow & \mathfrak{A} \subseteq A(G). \\
 & \text{homomorphism} & 
 \end{array}$$

Now, the transformation  $F$  turns this into the corresponding row

$$\begin{array}{ccccc}
 & \text{embedding as} & & \text{surjective} & \\
 FA(F^*) & \longrightarrow & F(\mathfrak{A}) & \longleftarrow & FA(G). \\
 & \text{generated subframe} & & \text{zigzag morphism} & 
 \end{array}$$

But then, finally, the following walk through the diagrams suffices.  $G \in \mathfrak{K} \Rightarrow FA(G) = ue(G) \in \mathfrak{K}$  (by the above observation)  $\Rightarrow F(\mathfrak{A}) \in \mathfrak{K}$  (closure under zigzag images)  $\Rightarrow FA(F^*) \in \mathfrak{K}$  (closure under generated subframes)  $\Rightarrow F^* \in \mathfrak{K}$  ('anti-closure' under ultrafilter extensions). ■

Actually, this result does not yet characterise **P1**, as it talks about modal definability by any set, finite or *infinite*. The additional strengthenings needed for zeroing in on **P1** are hardly enlightening, however.

The result also says a little bit more. Adding closure under ultrafilter extensions, while removing the condition of elementary definability, yields a characterisation of those classes of Kripke frames definable by means of a *canonical* modal logic in the sense of the Introduction (i.e. one which is complete with respect to its Henkin frames). Moreover, the above proof heuristics may also be used to formulate a general closure condition on classes of Kripke frames necessary and sufficient for definability by means of just *any* set of modal formulas ('SA-constructions'; cf. [Goldblatt and Thomason, 1974]).

As with the earlier ultrapower characterisation of **M1**, the above characterisation gives no *effective* information concerning the formulas in **P1**. What is needed are 'preservation theorems' giving the syntactic cash value of the given four closure conditions. Several of these have been given in [van Benthem, 1976], extending earlier results, e.g. of Feferman and Kreisel.

Here is an idea. Preservation under generated subframes allows only formulas constructed from atomic formulas and their negations, using

$\forall, \wedge, \vee$  as well as *restricted* existential quantifiers  $\exists v(Ruv \wedge (u, v \text{ distinct}))$ .

Preservation under disjoint unions admits only one single universal quantifier in front: all others are to be *restricted* to the form  $\forall v(Ruv \rightarrow)$ . Finally, preservation under zigzag images forbids the negations, and we are left with

**THEOREM 67.** *A first-order sentence is preserved under the formation of generated subframes, disjoint unions and zigzag-morphic images if and only if it is equivalent to one of the form  $\forall x\alpha(x)$ , where  $\alpha(x)$  has been constructed from atomic formulas using only conjunction, disjunction and restricted quantifiers.*

**Proof.** By elementary chain constructions, as in [Chang and Keisler, 1973, Chapter 3.1]. ■

For preservation under ultrafilter extensions, only some partial results have been found. (After all, the class of sentences preserved under such a complex operation *need* not even be effectively enumerable.)

As for the total complexity of **P1**, this may well be considerable — as was the case with **M1**. Are the two classes perhaps recursive in each other?

## 2.5 Modal Predicate Logic

As in much technical work in this area, modal *propositional* logic has been studied up till now. Modal *predicate* logic, however important in philosophical applications, is much less understood. (Cf. Chapter 2.5 in this *Handbook*.) Nevertheless, in the case of Correspondence Theory, an excuse for the neglect may be found in Theorem 69 below.

The unfinished state of the art shows already in the fact that no commonly accepted notion of semantic structure, or truth definition exists. Hence, we fix one particular, reasonably motivated choice as a basis for the following sketch of a predicate-logical variant of the earlier theory.

The *language* is the ordinary one of predicate logic, with added modal operators. *Structures* are tuples

$$\mathfrak{M} = \langle W, R, D, V \rangle,$$

where the *skeleton*  $\langle W, R, D \rangle$  is a Kripke frame with a *domain function*  $D$  assigning sets of individuals  $D_w$  to each world  $w \in W$ . The valuation  $V$  supplies the interpretation of the nonlogical vocabulary at each world.

The *truth definition* explicates the notion

$$\text{'}\varphi(x)\text{ is true in } \mathfrak{M} \text{ at } w \text{ for } d\text{'},$$

where the sequence  $d$  assigned to the free individual variables  $x$  comes from  $D_w$ . Its key options are embodied in the clauses for the individual quantifiers: these are to range over  $D_w$ , plus that for the modal operator:



$\Box\varphi(x)$  is true at  $w$  for  $d$  if, for each  $R$ -alternative  $v$  for  $w$  such that  $d$  is in  $D_v$ ,  $\varphi(x)$  is true at  $v$  for  $d$ .

Thus, necessity means ‘truth in all alternatives, where defined’.

As before, truth in a skeleton (at some world, for some sequence of individuals) means truth under all possible valuations. Again, in this way modal axioms start expressing properties of  $R, D$  — and their interplay.

The relevant matching ‘working language’ on the classical side will now be a *two-sorted* one: one sort for worlds, another for individuals. Its basic predicates are the two sortal identities,  $R$  between worlds, as well as the sort-crossing  $Exw$ : ‘ $x$  is in the domain of  $w$ ’, or ‘ $x$  exists at  $w$ ’.

EXAMPLE 68. The Barcan Formula  $\forall x\Box Ax \rightarrow \Box\forall xAx$  defines

$$\forall wv(Rwv \rightarrow \forall x(Exv \rightarrow Exw)).$$

**Proof.** ‘ $\Leftarrow$ ’: Assume  $\forall x\Box Ax$  at  $w$ , and consider any  $R$ -alternative  $v$ . For all  $d \in D_v, d \in D_w$  (by the given condition), whence  $\Box Ad$  holds at  $w$  — and, hence,  $Ad$  holds at  $v$ .

‘ $\Rightarrow$ ’: The Barcan Formula will hold under the following particular assignment:  $V_u(A, d) = 1$  if  $Rwu$  and  $d \in D_w$ .

This  $V$  verifies the antecedent, and hence the consequent. The relational condition follows. ■

Thus, the Barcan Formula expresses an interaction between  $R$  and  $D$ . This is not accidental. For *pure*  $R$ -principles, we have the following *conservation* result.

THEOREM 69. *There exists an effective translation from sentences  $\varphi$  of modal predicate logic to formulas  $p(\varphi)$  of modal propositional logic such that,*

*if  $\varphi$  is equivalent to some pure  $R, =$ -sentence  $\alpha$ , then  $p(\varphi)$  already defines  $\alpha$  in the sense of Section 2.2.*

**Proof.**  $p$  merely crosses out quantifiers in some suitable way. For full details (here and elsewhere) cf. [van Benthem, 1983]. ■

Besides the Barcan Formula, there are three further fundamental ‘de re/de dicto interchanges’. One of these provides a new example of non-first-order definability.

EXAMPLE 70.

1.  $\Box\forall xAx \rightarrow \forall x\Box Ax$  is universally valid,
2.  $\exists x\Box Ax \rightarrow \Box\exists xAx$  defines  $\forall wv(Rwv \rightarrow \forall x(Exw \rightarrow Exv))$ ,

3.  $\Box\exists xAx \rightarrow \exists x\Box Ax$  defines an essentially higher-order condition on  $R, =, E$ .

Despite the superficial resemblance to the McKinsey Axiom of section 2.2., the proof for the latter result is quite different from that of Example 43. Interested readers may notice that the above principle holds in worlds with a *finite* chain of overlapping two-element successors:

$$\{1, 2\}, \{2, 3\}, \{3, 4\}, \dots, \{n-1, n\}, \{n, n+1\}.$$

But, it may fail in the presence of *infinite* such chains, and then compactness lurks.

Further systematic reflection on the above ‘positive’ result yields a method of substitutions again, with an outcome like that of Theorem 48:

**THEOREM 71.** *Formulas of the form  $\varphi \rightarrow \psi$ , with  $\varphi$  constructed from atomic formulas prefixed by a (possibly empty) sequence of  $\forall, \Box$ , using only  $\wedge, \vee, \exists$  and  $\Diamond$ , and  $\psi$  constructed from atomic formulas using  $\wedge, \vee, \exists, \Diamond$  as well as  $\forall, \Box$ , are all uniformly first-order definable.*

The global mathematical characterisation of first-order definability remains essentially the same in this area, whence it is omitted here.

Something which does *not* generalise easily, however, is the algebraic approach of Section 2.3. This is an endemic problem in classical (and intuitionistic) logic already: elegant algebraization stops at the gates of predicate logic. There could be an area of ‘modal cylindric algebra’ of course (cf. [Henkin *et al.*, 1971]), but none exists yet. (For an interesting related area, cf. the extension of modal propositional algebra to the modal program algebra of dynamic logicians (cf. [Kozen, 1979] or the Dynamic Logic chapter in volume 5 of this *Handbook*.) As a consequence, we still lack an elegant characterisation of the modally definable fragment of the present two-sorted first-order language.

What we do have, however, is such a characterisation for that same language with *parametrised predicate constants*  $A(w, -)$  for the predicate constants  $A(-)$  of the modal predicate logic. Thus, this is the appropriate language for the *first-order* transcription of the above truth definition. The Barcan Formula, for example, becomes

$$\begin{aligned} \forall x(Exw \rightarrow \forall v((Ewv \wedge Exv) \rightarrow Avx)) \rightarrow \\ \rightarrow \forall v(Rwv \rightarrow \forall x(Exv \rightarrow Avx)). \end{aligned}$$

As in Theorem 18, two characteristic modal relations suffice for characterising the modal transcriptions among the class of all formulas of this language. In order to end on an optimistic note, here is the relevant result.

First, modal predicate logic knows *generated submodels*, just as in Section 2.1. Moreover, the earlier *zigzag relations* may be enriched so as to incorporate individual back-and-forth choices, as in the Ehrenfeucht–Fraïssé approach to first-order definability.

DEFINITION 72. A *zigzag connection*  $C$  between two models  $M_1, M_2$  relates finite sequences  $(w, x)$  of equal length ( $w$  a world,  $x$  a sequence of individuals in the domain of  $w$ ) in such a way that

1. all such sequences occur: those from  $M_1$  in the domain, those from  $M_2$  in the range of  $C$
2. if  $C(w, x)(v, y)$  and  $w' \in W_1$ , with  $R_1 w w', x \in D_w$ , then  $C(w', x)(v', y)$  for some  $v' \in W_2$  with  $R_2 v v', y \in D_{v'}$ ,  
and analogously in the opposite direction ('world zigzag')
3. if  $C(w, x)(v, y)$  and  $d \in D_w$ ,  
then  $C(w, x \hat{\ } d)(v, y \hat{\ } e)$  for some  $e \in D_v$ ,  
and vice versa. ('individual zigzag')
4. if  $C(w, x)(v, y)$ , then the map  $(x)_i \rightarrow (y)_i$  is a *partial isomorphism* between  $\langle D_w, V_w \rangle$  and  $\langle D_v, V_v \rangle$ .

Now, transcriptions of modal formulas are *invariant* for generated submodels and zigzag connections, in the obvious sense. E.g. the latter have been made precisely in such a way that for modal  $\varphi$ ,

$$\varphi \text{ is true at } w \text{ for } x \quad \text{iff} \quad \varphi \text{ is true at } v \text{ for } y, \quad \text{when } C(w, x)(v, y).$$

THEOREM 73. A formula  $\varphi = \varphi(w, x)$  of the two-sorted world/individual language is (equivalent to the transcription of) a modal formula if and only if it is invariant for generated submodels and zigzag connections.

**Proof.** This follows from the main proof in [van Benthem, 1981b]. ■

On the whole, exciting technical results are yet scarce in modal predicate logic — and Correspondence Theory is no exception.

## 2.6 Higher-Order Correspondence

Modal formulas define second-order ( $\Pi_1^1$ ) conditions on the alternative relation in all cases, and first-order conditions in some. In the perspective of abstract model theory, two possible generalisations arise here.

Instead of the first-order target language, one may consider suitable extensions. For instance, in Theorem 37, the relevant relational condition was definable in  $L_{\omega_1, \omega}$ : first-order logic with countable conjunctions and disjunctions. Not all modal formulas become definable here, however. E.g. Löb's Axiom defined a form of well-foundedness, which is known to be beyond  $L_{\omega_1, \omega}$ , or indeed any language of the  $L_{\infty, \omega}$ -family. On the other hand, this time for instance, the defining condition is already in 'weak second-order logic'  $L^2$ , allowing quantification over *finite* sets of individuals. Thus,

various wider classes of definability could be considered for modal formulas, short of  $\Pi_1^1$ . And, in fact, even the latter case itself is interesting. Which  $\Pi_1^1$ -sentences, for example, admit of modal definitions?

Given the general lack of semantic characterisations for such higher logics, such characterisations for their modal fragments are also difficult to obtain. One observation might be that both  $L_{\omega_1\omega}$  and  $L^2$  have the property of *invariance for partial isomorphism* (cf. van Dalen's chapter in Volume 1 of this *Handbook*). It will be of interest to study this preservation condition on modal formulas. In fact, no counter-examples have been discovered yet; but these do exist in tense logic. (The rationals  $\langle Q, < \rangle$  and the reals  $\langle R, < \rangle$  are a classical example of partially isomorphic structures, but there exists a tense-logical formula expressing Dedekind Completeness, which is valid on the latter, though not on the former frame.)

On the other hand, the modal propositional language could itself be strengthened, notably by the introduction of propositional quantifiers  $\forall p, \exists p$ , which have occurred in various places in the literature (cf. Garson's chapter in Volume 3 of this *Handbook*). Thus, e.g.  $\forall p(\Box\Diamond p \rightarrow \exists q\Diamond\Box q)$  would become an admissible formula, but also  $\Box\exists p\Diamond p \rightarrow \Diamond\forall q\Diamond\Box q$ . Actually, there is a choice here, whether to allow the propositional quantifiers in the scope of modal operators or not. Henceforth, we consider the second, more restricted option.

In the usual manner, a prenex hierarchy arises here, with all propositional quantifiers in front, of which the original modal formulas form the  $\Pi_1^1$ -part (universal prefix). The next simplest cases are  $\Sigma_1^1$  (existential prefix) and  $\Delta_2^1$ . In fact, the latter has a reasonable motivation through the modal 'rules' mentioned in Section 3.2 below.

It has been observed by Gabbay that the following rule defines *irreflexivity* of Kripke frames:

$$\text{'if } F \models (\Box p \wedge \neg p) \rightarrow \varphi[w] \text{ (with } \varphi \text{ } p\text{-free), then } F \models \varphi[w]\text{'}$$

The general pattern here is that of ' $F \models \varphi[w]$  only if  $F \models \psi[w]$ ', i.e. an implication of two  $\Pi_1^1$ -formulas, which is  $\Delta_2^1$ . (It may be written either in the form  $\forall\exists$  or  $\exists\forall$ .)

Actually, the above specific example is already  $\Sigma_1^1$ , as it amounts to  $\forall pq((\Box p \wedge \neg p) \rightarrow q) \rightarrow \forall qq$ , i.e.  $\forall p((\Box p \wedge \neg p) \rightarrow \perp) \rightarrow \forall qq$ , i.e.  $\forall p((\Box p \wedge \neg p) \rightarrow \perp) \rightarrow \perp$ , i.e.  $\exists p(\Box p \wedge \neg p)$ . Another relevant observation is that implications of the above form  $\forall \rightarrow \forall$ , if first-order definable at all, already have a first-order definable consequent. We do not go into these specific matters here, but note a general issue.

As often in higher-order logic, we are interested in *hierarchy* results. For instance, how much power of first-order definability is added at each stage? It is evident that  $\Sigma_1^1$ -definability adds essentially just all negations of the (local) principles in **P1** (cf. Section 2.4), while  $\Delta_2^1$  adds conjunctions and disjunction across **P1** and the latter 'mirror image'.

QUERY. Does the second-order prenex hierarchy induce an *ascending* corresponding hierarchy of modally definable first-order principles about the alternative relation?

This possibly ascending hierarchy cannot exhaust all first-order principles, as higher-order modal formulas do retain one basic preservation property: their local truth is invariant under passing to generated subframes. (The Generation Theorem 15 yields this consequence all the way up, not just for the original modal  $\Pi_1^1$ -formulas.) But then, we know what this semantic constraint means in syntactic terms for first-order formulas (cf. [van Benthem, 1976, Chapter 6]). These will be the ‘almost-restricted’ ones consisting of one universal quantifier followed by a compound of atomic formulas with negation, conjunction and restricted quantifiers  $\exists y(Rxy \wedge)$ .

The other preservation properties of Section 2.1 are lost, however. As was observed earlier, irreflexivity ( $\forall x \neg Rxx$ ) becomes definable and, hence, preservation under zigzag morphisms fails. Anti-preservation under ultrafilter extensions fails, because the earlier example  $\forall x \exists y(Rxy \wedge Ryy)$  becomes definable as well. (A straightforward definition uses a propositional quantifier within a modal scope:  $\diamond \forall p(\Box p \rightarrow p)$ . But there is a nonembedded substitute in the form of  $\exists p(\diamond p \wedge \forall q \Box(p \rightarrow (\Box q \rightarrow q)))$ .)

Thus, we arrive at the following

QUESTION. Can every almost-restricted first-order formula  $\forall x \varphi(x)$  be defined at some level in the modal propositional quantifier hierarchy?

Using ‘simulation’ of restricted first-order quantification by propositional quantifiers, one may indeed handle most obvious cases. Here is one illustration of the procedure

EXAMPLE. Let  $\varphi(x)$  be  $\exists y(Rxy \wedge \forall z(Ryz \rightarrow (Rzz \vee (Rzy \wedge Rzx))))$ . The idea is to define  $\{x\}, \{y\}, \{z\}$ , in a sense, as far as necessary (i.e. on the set consisting of  $x$ , its  $R$ -,  $R^2$ - and  $R^3$ -successors) — and then to express all desired relations between these by means of modal formulas:

$$\begin{aligned} & \exists p_x(p_x \wedge \forall q_x(((p_x \wedge q_x) \vee \diamond(p_x \wedge q_x) \vee \diamond \diamond(p_x \wedge q_x) \vee \diamond \diamond \diamond(p_x \wedge \\ & q_x)) \rightarrow (\Box(p_x \rightarrow q_x) \wedge \Box \Box(p_x \rightarrow q_x) \wedge \Box \Box \Box(p_x \rightarrow q_x))) \text{ [this} \\ & \text{ makes } p_x \text{ unique to the extent indicated]} \wedge \exists p_y(\diamond p_y \wedge \text{ [same} \\ & \text{ uniqueness statement]} \wedge \forall p_z((\diamond(p_y \wedge \diamond p_z) \wedge \text{ [same uniqueness} \\ & \text{ statement]}) \rightarrow (\forall q_z \Box \Box(p_z \rightarrow (\Box q_z \rightarrow q_z)) \text{ [i.e. ‘} Rzz \text{’]} \vee \diamond \diamond(p_z \wedge \\ & \diamond p_x \wedge \diamond p_y) \text{ [i.e. ‘} Rzy \wedge Rzx \text{’]})))). \end{aligned}$$

Accordingly, our conjecture is that the above question has a positive answer.

We conclude with one further

QUESTION. Does the addition of propositional quantifiers within modal scopes add any power of expression?

### 3 OTHER INTENSIONAL NOTIONS

Modal logic is only one branch, be it a paradigmatic one, of intensional logic in general. But also in other intensional areas, a Correspondence Theory is possible. In some cases, the generalisation runs smoothly: existing notions and results may be applied at once, or after only minor modification. A case in point is *tense logic*, to be treated in Section 3.1. More challenging generalisations arise when the relevant intensional semantics exhibit strong peculiarities, diverging from the earlier modal case. Sometimes, these assume the form of pre-conditions on the alternative relation; but maybe the most important hurdle is when a restriction is proclaimed on ‘admissible assignments’. Both phenomena occur in *conditional logic*, the topic of Section 3.2. That, even under such circumstances, an interesting Correspondence Theory may remain, is shown by the example of *intuitionistic logic* in Section 3.3.

These two new features do not exhaust the possible semantic variation. One may also move to the interplay of different kinds of intensional operators, for instance, using correspondence to connect different alternative relations.

EXAMPLE. In *dynamic logic*, two modal operators  $\Box, \Box^*$  figure, which may be provided with two alternative relations  $R, R^*$ . (Recall that  $\Box a$  means ‘after every successful computation of  $a$ ’, while the intuitive meaning of  $\Box^* a$  is to be: ‘after any finite number of runs of  $a$ ’.) Now, from a correspondence point of view, the well-known *Seegerberg Axioms*

$$\begin{aligned} \Box^* p &\rightarrow \Box p \\ \Box^* p &\rightarrow \Box \Box^* p \\ \Box^* (p \rightarrow \Box p) &\rightarrow (\Box p \rightarrow \Box^* p) \end{aligned}$$

define precisely the condition that

$$R^* \text{ coincides with the transitive closure of } R.$$

The very exoticness of this example to many readers may help to show that Correspondence Theory is omnipresent.

No systematic developments will be given in the following sections. Their purpose is to convey an impression of notions and themes, through mainly illustrative examples. Indeed, here is where the reader may wish to carry on the torch herself.

#### 3.1 *Tense Logic*

Traditionally, tense-logical structures have been taken to be temporal orders  $\langle T, < \rangle$ , where  $T$  consists of the points in Time, ordered by precedence  $<$  (‘earlier than’, ‘before’). The simplest formal language to be chosen has

been that of Prior, adding operators  $G$  ('it is always going to be'),  $H$  ('it has always been') to some propositional base. We add  $F$  ('future'),  $P$  ('past') as derived notions. (Cf. the chapter on Basic Tense logic in volume 6 for the necessary background in tense logic.)

Of the amazing diversity of 'ontological' and 'linguistic' questions concerning this temporal semantics, only a few themes will be mentioned here. (Cf. [van Benthem, 1985] for a varied exploration.)

*Explaining philosophical dicta.* In his famous paper 'The Unreality of Time', the philosopher McTaggart enunciated several temporal principles. One of these reads [McTaggart, 1908]:

"If one of the determinations past, present and future can ever be applied to (an event), then one of them has always been and always will be applicable, though of course not always the same one."

When translated into Priorean axioms, this becomes a list:

1.  $Pq \rightarrow H(Fq \vee q \vee Pq)$
2.  $Pq \rightarrow GPq$
3.  $q \rightarrow HFq$
4.  $q \rightarrow GPq$
5.  $Fq \rightarrow HFq$
6.  $Fq \rightarrow G(Fq \vee q \vee Pq)$ .

What do these principles mean? The answer may be obtained through the method of substitutions (fitted to the temporal case — but such generalisations will be presupposed tacitly henceforth).

EXAMPLE 74.

1. defines *left-connectedness*:  $\forall x \forall y < x \forall z < x (y < z \vee z < y \vee y = z)$ ,
2. defines *transitivity*:  $\forall x \forall y < x \forall z > x y < z$ ,
3. defines  $\top$ ,
4. defines  $\top$ .

If  $G, H$  had been interpreted through different relations  $<_G, <_H$ , then (3) and (4) would have expressed that  $<_H$  is the *converse* relation of  $<_G$ .

5. defines *transitivity* again:  $\forall x \forall y > x \forall z < x z < y$ ,

6. defines *right-connectedness*:  $\forall x \forall y > x \forall z > x (y < z \vee z < y \vee y = z)$ .

Thus, the McTaggart temporal picture is one of linear flow.

*An incompleteness theorem.* Simple transfer of earlier modal results establishes the seminal incompleteness result of [Thomason, 1972], in a very simple version.

**THEOREM 75.** *The tense logic axiomatised by*

$$\begin{array}{ll} H(Hp \rightarrow p) \rightarrow Hp & (\text{L\"ob's Axiom}) \\ GFp \rightarrow FGP & (\text{McKinsey Axiom}) \end{array}$$

*is incomplete.*

**Proof.** Specifically, this logic holds in no frame — and yet it is not inconsistent.

First, as to the former statement, recall from Section 2.2 that

1. Löb's Axiom defines *transitivity* of  $>$  and *well-foundedness* of  $<$ .

By the former,  $<$  is transitive as well (transitivity is 'independent of the temporal direction', or *isotropic* (cf. [van Benthem, 1985])). Thus, in this special case, Example 51 applies, and we have

2. McKinsey's Axiom defines atomicity:  $\forall x \exists y > x \forall z > y z = y$ .

A consequence of the latter property is  $\forall x \exists y > x y < y$  (cf. Example 65(4)). So, the temporal order must contain instantaneous loops  $\dots < y < y < y < \dots$ , which contradicts well-foundedness. Therefore, our logic holds in no frame.

Nevertheless, it does hold in a *general* frame, viz. an earlier example from Section 2.1:  $\langle N, <, \mathfrak{W} \rangle$ , with

$$\mathfrak{W} = \{X \subseteq N \mid X \text{ is finite or } N - X \text{ is finite}\}.$$

The reason was that refutations for the McKinsey Axiom are no longer 'admissible', as these involve infinite alterations. (Thomason gives a speculation at this point concerning the Second Law of Thermodynamics: 'event patterns stabilise'.) But then, the logic cannot be inconsistent: its **K**-theorems hold in all general frames where it is valid. ■

*Tense-logical axioms for the temporal order.* In [van Benthem, 1985], the following fundamental axioms are derived for any temporal order induced by a *comparative* (in the linguistic sense) 'earlier than'.

1. *irreflexivity*:  $\forall x \neg x < x$  (‘no vortices in Time’)



2. *transitivity*:  $\forall x \forall y > x \forall z > y \ z > x$  ('flow')
3. *almost-connectedness*: ('arrows are comparative yard sticks')
- $$\forall x \forall y > x \forall z (x < z \vee z < y)$$

A version of the latter principle may also be found as the key axiom in Leibniz' relational theory of Space-Time (cf. [Winnie, 1977]).

Which tense-logical axioms correspond? From Section 2.4, we know that (1) is undefinable, (2) yields  $Gp \rightarrow GGp$ , while (3) just fails to fall under Theorem 67. What the latter result does give is a correspondence between

$$\forall x \forall y > x \forall z > y \forall u > x (y < u \vee u < z)$$

and

$$(F(p \wedge Fq) \wedge Fr) \rightarrow (F(p \wedge Fr) \vee F(r \wedge Fq)).$$

Another example concerns particular temporal orders. One can never hope to fully define such frames categorically by their tense-logical theories. For, by the Generation Theorem, tense-logical formulas cannot distinguish between one single, or several parallel flows of Time — which latter picture is so familiar from contemporary science fiction. Still, if *disjoint unions* of frames are disregarded, we have

THEOREM 76.  $\langle N, < \rangle$  is defined categorically by the axioms

$$\begin{aligned} H(Hp \rightarrow p) &\rightarrow Hp \\ Pp &\rightarrow H(Fp \vee p \vee Pp) \\ Fp &\rightarrow G(Fp \vee p \vee Pp) \\ FT & \\ G(Gp \rightarrow p) &\rightarrow (FGp \rightarrow Gp) \end{aligned}$$

The proof is omitted here.

But, e.g. the integers  $\langle Z, < \rangle$  cannot be thus defined; as the contraction to a single point remains a zigzag morphism preserving their theory. ( $\langle N, < \rangle$  was unafflicted this time: in tense logic, zigzag morphism have *two* backward relational clauses — whence, the earlier contraction fails to quality.)

*Time and modality.* Combined modal-tense logics with two alternative relations  $R, <$  have been repeatedly proposed. For instance, in [White, 1981] we find a logic with characteristic axioms

$$\begin{aligned} Gp \rightarrow GGp, Fp \rightarrow G(Fp \vee p \vee Pp), PT & \quad (\text{D4.3}) \\ Pq \rightarrow \Box Pq & \quad (\text{'irrevocable past'}). \end{aligned}$$

This logic is claimed to be appropriate for an analysis of the famous Diodorean 'Master Argument', identifying *possibility* with *actual or future truth* — a version of what was later to become known as the principle of Plenitude: all metaphysical possibilities are eventually realised in this World.

Our analysis of this claim runs as follows.  $Gp \rightarrow GGp$  defines *transitivity* for  $<$ , the McTaggart Axiom defines right-connectedness; while  $PT$  defines *left-succession*:  $\forall x\exists y y < x$ . The additional ‘mixing postulate’ defines

$$\forall xy(Rxy \rightarrow \forall z(z < x \rightarrow z < y)).$$

CLAIM (1).  $\forall xy(Rxy \rightarrow (y < x \vee y = x \vee x < y))$ .

**Proof.** Assume  $Rxy$ . Let  $z < x$  (by left-succession). Then  $z < y$  (‘mix’). The conclusion follows by right-connectedness. ■

CLAIM (2).  $\forall xy(Rxy \rightarrow (x < y \vee x = y))$ .

**Proof.** If  $Rxy$  and  $y < x$ , then  $y < y$  (‘mix’): *contra* irreflexivity. ■

The outcome is this: *without* ever using transitivity, but *with* irreflexivity (which is presupposed in White’s whole set-up), a relational condition follows which is indeed defined by the Diodorean challenge:

$$\diamond p \rightarrow (Fp \vee p).$$

This is only one of the many possible semantics for temporal modalities, of course. The correspondence aspect of, e.g. the Occamist ‘branching time’ of [Burgess, 1979] remains to be explored.

*Alternative temporal ontologies.* Recently ‘interval structures’ have been proposed as an alternative for the above traditional point ontology. From the manifesto of [Humberstone, 1979], a picture emerges of triples

$$\langle I, \subseteq, < \rangle,$$

where  $\subseteq$  is *inclusion* among intervals, and  $<$  total *precedence*.

Here again, correspondences prove useful in exploring proposed principles. The language has the ordinary tense-logic operators, as well as a modality  $\square$  (‘in all subintervals’). In this notation, Humberstone’s base logic has for its basic axioms

1.  $Fp \rightarrow \square Fp$
2.  $F\diamond p \rightarrow Fp$
3.  $\diamond F\square p \rightarrow (\diamond p \vee Fp)$ .

By the earlier method of substitutions, equivalents may be found illuminating these:

1. defines  $\forall x\forall y > x\forall z \subseteq x y > z$ ,

a property known as *left monotonicity*,

$$2. \text{ defines } \forall x \forall y > x \forall z \subseteq y \ z > z,$$

its dual property of *right monotonicity*. Finally,

$$3. \text{ defines } \forall x \forall y \subseteq x \forall z > y \ (\exists u \subseteq z : u \subseteq x \vee \exists u \subseteq z : u > x),$$

a form of a principle known as *convexity*. ('Stretches of time should be uninterrupted'.)

Starting from the other side, one may impose basic postulates on  $\subseteq, <$ , asking for definitions in this 'interval tense logic'. For  $<$ , these might be the earlier-mentioned ones, for  $\subseteq$ , a minimum seems to be the requirement of *partial order*, while monotonicity (and convexity) take care of minimal connections between  $<, \subseteq$ . This would add only two axioms to the preceding ones, viz. **S4** for inclusion. The further condition of *anti-symmetry* is not definable — as may be seen by noting that the map  $n \mapsto n$  (modulo 2) is a  $\subseteq$ -zigzag morphism sending the anti-symmetric frame  $\langle Z, \leq \rangle$  to the non-antisymmetric one  $\langle \{0, 1\}, \{ \langle 0, 0 \rangle, \langle 0, 1 \rangle, \langle 1, 0 \rangle, \langle 1, 1 \rangle \} \rangle$ .

Many more examples of further correspondences on top of this foundation may be found in Chapter II.3.2 of [van Benthem, 1985].

### 3.2 Conditionals

From among the teeming multitude of 'conditional logics', three specimens have been included here. As no work of the present kind has been done in this area at all, the following considerations are still very much first steps. (Cf. the Conditional Logic chapter in volume 5 for a discussion of conditional logics.)

#### *Constructive implication*

Perhaps the single most effective argument in favour of constructive, as opposed to classical implication is the natural deduction analysis. The *natural* rules for  $\rightarrow$ -introduction and  $\rightarrow$ -elimination give us only a fragment of all classical pure  $\rightarrow$ -tautologies; axiomatised by

$$(A1) \ \varphi \rightarrow (\psi \rightarrow \varphi)$$

$$(A2) \ (\varphi \rightarrow (\psi \rightarrow \chi)) \rightarrow ((\varphi \rightarrow \psi) \rightarrow (\varphi \rightarrow \chi))$$

plus the rule of *modus ponens*. A principle notably outside of this class is Peirce's Law

$$((\varphi \rightarrow \psi) \rightarrow \varphi) \rightarrow \varphi.$$

But really, the same elegance shows up in the Henkin completeness proof. In the usual proof, one starts from a given consistent set — and then has

to extend this arbitrarily to just *any* maximally consistent one, in order to ‘break down’ implications according to the classical truth table. A *canonical* model construction rather uses a unique natural model, viz. that consistent set together with all its consistent extensions, exploiting the evident decomposition rule

$$\Sigma \vdash \varphi \rightarrow \psi \text{ if and only if } \forall \Sigma' \supseteq \Sigma : \text{ if } \Sigma' \vdash \varphi, \text{ then } \Sigma' \vdash \psi.$$

A perfect match arises with the following semantics. Structures are general frames  $F = \langle W, R, \mathfrak{W} \rangle$ , where  $R$  corresponds to the above inclusion relation, and  $\mathfrak{W}$  consists of all *R-hereditary* sets of worlds. (Propositions represent *R-cumulative* knowledge on this view.)

A direct study of the above logic on these frames would yield rather clumsy conditions. One case will be exhibited, as it illustrates a variant concept of correspondence, viz. *correspondence for rules* rather than *axioms*.

EXAMPLE 77. Modus Ponens defines the condition ‘every world belongs to some finite *R*-loop’.

**Proof.** ‘ $\Leftarrow$ ’: Suppose that  $xRx_1R \dots Rx_nRx$ . Let  $V(p), V(q)$  be *R-hereditary* subsets of  $W$ , such that  $p, p \rightarrow q$  hold at  $x$ . Then, successively,  $p, q$  hold at  $x_1, \dots, x_n$ , and finally at  $x$ .

‘ $\Rightarrow$ ’: Suppose that  $x$  belongs to no finite *R*-loop. Set  $V(p) :=$  the smallest *R-hereditary* set containing  $x$ ,  $V(q) =$  the *R-hereditary* closure of  $\{y \mid Rxy\}$ . This verifies  $p, p \rightarrow q$  at  $x$ ; without verifying  $q$ . ■

What will be done instead is to postulate the *partial order* behaviour of  $\subseteq$ : *reflexivity*, *transitivity* and *antisymmetry*. Finer peculiarities of (A1), (A2) remain undetectable below this threshold.

Further restrictions on  $R$  may now be imposed by stronger axioms; e.g. we can see why Peirce’s Law is characteristic for classical logic.

EXAMPLE 78. Peirce’s Law defines the restriction to single points:

$$\forall xy(Rxy \rightarrow y = x).$$

**Proof.** ‘ $\Leftarrow$ ’: A simple calculation suffices.

‘ $\Rightarrow$ ’: Suppose that  $Rxy, x \neq y$ . Set  $V(q) = \emptyset, V(p) = \{z \mid Rxz \wedge x \neq z\}$ . This makes  $(p \rightarrow q) \rightarrow p$  true at  $x$  (notice that  $p \rightarrow q$  is false at  $x$  itself), while falsifying  $p$ . (By the way, that  $V$  is admissible, i.e. that  $V(p)$  is *R-hereditary*, follows from the above general assumption.) ■

But ‘intermediate’ implication axioms exist as well.

EXAMPLE 79. The following principle

$$((p \rightarrow q) \rightarrow p) \rightarrow (((q \rightarrow r) \rightarrow q) \rightarrow p)$$

defines a maximal length 3 for  $R$ -chains:

$$\forall xy(Rxy \rightarrow \forall z(Ryz \rightarrow (x = y \vee y = z \vee \forall u(Rzu \rightarrow z = u)))).$$

**Proof.** Here is the relevant counter-example for the argument in the ‘ $\Rightarrow$ ’-direction. Assume that  $xRyRzRu$ , while  $x \neq y, y \neq z, z \neq u$ . Set  $V(r) = \emptyset, V(q) = \{v \mid Ruv \wedge u \neq v\} \cup \{v \mid Ryv \wedge \neg Rvz\}, V(p) = \{v \mid Ryv \wedge y \neq v\}$ . The principle will be falsified at  $y$ . ■

It has not been possible to find other types of intermediate example. Hence, we conclude with a

CONJECTURE. All principles of pure constructive implication define first-order constraints on  $R$ ; viz. restrictions to some finite chain length.

### *Relevant implication*

Of the various proposed semantics for relevance logic, here is a perspicuous example from [Gabbay, 1976, Chapter 15]. Structures are now tuples  $\langle W, R, V, 0 \rangle$ , where 0 is a special world providing a vantage point from which to compare other worlds through the *ternary* relation  $R$ . Intuitively,  $R_a bc$  is to mean that  $b$  is ‘included’ in  $c$ , at least from the perspective of  $a$ . (One might think of, for example, ‘ $a$ -local inclusion’:  $a \cap b \subseteq a \cap c$ .) No prior conditions are imposed on this relation.

This is not to say that these are not to be found at all. For instance, it may be shown that the mentioned local inclusion relation is characterised by two *betweenness* axioms:

1.  $R_a bc \leftrightarrow R_b ac$  (interchanging boundaries)
2.  $(R_a bc \wedge R_d ae \wedge R_d be) \rightarrow R_d ce$

(I.e. if  $c \in [a, b], a \in [d, e], b \in [d, e]$ , then  $c \in [d, e]$ : a form of convexity.)

The explication of implication reads as follows:

$\varphi \rightarrow \psi$  is true at  $a$  iff, for all  $b, c$  such that  $R_a bc$ , if  $\varphi$  is true at  $b$ , then  $\psi$  is true at  $c$ .

As it stands, this definition makes *no* implication laws universally valid. To obtain at least some indubitable principle, one therefore imposes a restriction on valuations. The most urgent case is that of  $p \rightarrow p$ . On the above bare semantics, it would correspond to  $\forall xyz(R_x yz \rightarrow y = z)$ , collapsing the ternary relation. To avoid this, one again requires ‘cumulation’:

valuations  $V$  are only to assign subsets  $X$  of  $W$  subject to the constraint that  $\forall xy \in W(R_0 xy \rightarrow (x \in X \rightarrow y \in X))$ .

If this constraint is to extend automatically to sets  $X$  defined by complex implicational formulas, then a mild form of *transitivity* is to be imposed on the ternary relation after all:

$$\forall xyzu((R_0xy \wedge R_yzu) \rightarrow R_xzu).$$

Notice how this relates perspectives from different vantage points.

But then, if reasonable forms of transitivity have become respectable, we also add  $(*)\forall xyzu((R_0xy \wedge R_0yz) \rightarrow R_0xz)$ .

Now, at last, some genuine correspondences arise — of a ‘local’ sort (cf. Section 2.2).

EXAMPLE 80.

1. Modus Ponens defines  $R_000$ ,
2. Axiom A1 defines a curious form of ‘transitivity’:  
 $\forall xyzu((R_0xy \wedge R_yzu) \rightarrow R_xu)$ .

**Proof.** (Case (1) only) ‘ $\Leftarrow$ ’: This direction is immediate.

‘ $\Rightarrow$ ’: Let  $V(p) = \{0\} \cup \{x \mid R_00x\}$ ,  $V(q) = \{x \mid R_00x\}$ . By the above principle  $(*)$ , both assignments are admissible. Clearly, both  $p$  and  $p \rightarrow q$  are true at 0, whence also  $q$ : i.e.  $R_000$ . ■

Obviously, the second principle is not very plausible — but then, neither is (A1) for a relevance logician.

A more interesting phenomenon in relevance logic, from the present point of view, is the treatment of *negation*. This formerly inconspicuous notion is now interpreted using a ‘reversal operation’  $+$  on worlds:

$$\neg\varphi \text{ is true at } a \quad \text{iff} \quad \varphi \text{ is true at } a^+.$$

In this light, new combined correspondences appear, such as that between Contraposition and the reversal law

$$\forall xy(R_0xy \rightarrow R_0y^+x^+).$$

Correspondence Theory may be applied to any kind of semantic entity.

### *Counterfactual implication*

Ramsey told us to evaluate conditionals as follows. Make the minimal adjustment of your stock of beliefs needed to accommodate the antecedent: then see if the consequent follows. Various syntactic and semantic implementations of this view exist, of which that of [Lewis, 1973] has deservedly won the greatest favour. A counterfactual  $\varphi \square \rightarrow \psi$  is true in a world, on his

account, if  $\psi$  is true in all worlds most similar to that world *given that*  $\varphi$  holds in them.

As the preceding account has some difficulties in the infinite case, let us consider *finite* models  $\langle W, C, V \rangle$ , where  $C$  is a *ternary* relation of comparative similarity:

$C_x yz$  for: ‘ $y$  is closer to  $x$  than  $z$  is’.

Lewis gives three basic conditions on the relation ‘no closer’:

1. *transitivity*:  $\forall xyz((\neg C_x yz \wedge \neg C_x zu) \rightarrow \neg C_x yu)$ ,
2. *connectedness*:  $\forall xyz(\neg C_x yz \vee \neg C_x zy)$ ,
3. *egocentrism*:  $\forall xy(\neg C_x xy \rightarrow x = y)$ .

Rewriting these for ‘closer’, one finds to one’s surprise that (2) is rather weak, being merely

2'. *asymmetry*:  $\forall xyz(C_x yz \rightarrow \neg C_x zy)$ .

On the other hand, (1) becomes a strong principle

1'.  $\forall xyu(C_x yu \rightarrow \forall z(C_x yz \vee C_x zu))$ ,

which we knew as *almost-connectedness* back in Section 3.1.

From asymmetry and almost-connectedness, one may derive ordinary *transitivity* and *irreflexivity*, whence the three ‘comparative’ axioms of Section 3.1 emerge. These principles justify the appealing picture of ‘similarity spheres’ around the reference world  $x$ .

The tendency has been since 1973 to retain only *transitivity* and *irreflexivity* as fundamental pre-conditions on  $C$ , leaving various forms of connectedness as optional extras. Thus, one finds an axiomatisation of this austere minimal conditional logic in [Burgess, 1981].

The truth definition in this case may be taken to be the following:

$\varphi \Box \rightarrow \psi$  is true at  $w$  if  $w$  holds in all  $\varphi$ -worlds  $C$ -closest to  $w$ .

Indeed, this clause verifies the following list of principles without further ado:

$$\begin{aligned} p \Box \rightarrow p, \\ p \Box \rightarrow q, p \Box \rightarrow r \vdash p \Box \rightarrow q \wedge r, \\ p \wedge q \Box \rightarrow p, \\ p \Box \rightarrow r, q \Box \rightarrow r \vdash p \vee q \Box \rightarrow r. \end{aligned}$$

It is only the last one which requires *transitivity*:

$$p \Box \rightarrow q \wedge r \vdash p \wedge q \Box \rightarrow r.$$

*Egocentrism* is restored by adding the principle of Modus Ponens:

$$p \Box \rightarrow q, p \vdash q$$

But, the original Lewis logic contained even further principles, such as the formidable

$$((p \vee q) \Box \rightarrow p) \vee \neg((p \vee q) \Box \rightarrow r) \vee q \Box \rightarrow r.$$

What does it express? As it happens, it restores *almost-connectedness*.

**Proof.** First, the axiom is valid under this additional assumption — by the above discussion.

Next, suppose almost-connectedness fails; i.e. for some  $xyz u$  we have:  $C_x yz, \neg C_x yu, \neg C_x uz$ . By transitivity, it follows that  $\neg C_x zu$ . Now, set  $V(p) = \{y\}, V(q) = \{z, u\}, V(r) = \{y, u\}$ . Then  $z$  is  $q$ -closest among the worlds falsifying  $r$ . The two  $p \vee q$ -closest worlds  $y, u$  both verify  $r$ . Finally,  $p$  fails in the  $p \vee q$ -closest world  $u$ . Thus, Lewis' axiom has been refuted. ■

Finally, to mention an example outside of Lewis' original logic, there is the Stalnaker principle of 'Conditional Excluded Middle':

$$p \Box \rightarrow q \wedge p \Box \rightarrow \neg q.$$

As was stated in the Introduction, this axiom even requires the similarity order to be a *linear* one. In the present finite case, this means that the above truth definition reduces to:

$$\varphi \Box \rightarrow \psi \text{ is true at } w \text{ if } \psi \text{ holds in the closest } \varphi\text{-alternative to } w.$$

And that was the original Stalnaker explication of conditionals.

The previous examples were all conditional axioms without nestings of  $\Box \rightarrow$ . This is typical for most current logics in this area. Relational conditions matching these have invariably been found to be first-order ones. Hence, in view of Theorem 38, here is our

**CONJECTURE.** All counterfactual axioms without nestings of conditionals are first-order definable.

The reason for this restriction lies in the motivation for the present area. Entailment conditionals such as constructive implication, or modal entailment have often been proposed out of dissatisfaction with classical 'nested principles', such as, say,  $p \rightarrow (q \rightarrow p)$  or Peirce's Law. The non-nested classical fragment was not called into question. Counterfactual conditionals, however, typically disobey classical implicational logic at the level of non-nested inferences, such as the monotonicity rule from  $p \rightarrow q$  to  $p \wedge r \rightarrow q$ .

Nevertheless, there are intrinsic reasons to be found inside the above semantics for considering nested axioms after all. For, one obvious omission



in the above list of semantic conditions was the lack of *index principles* relating the perspectives of different worlds. For instance, when we read  $C$  for a moment as relative proximity in Euclidean space, we find the following *Triangle Inequality*

$$\forall xyz((C_xyz \wedge C_zxy) \rightarrow C_yxz).$$

And there are other elegant principles of this kind.

Now, it is easily seen that such index principles are just what is involved when nested counterfactuals are evaluated: the perspective starts shifting. Thus, it will be rewarding to have correspondences here as well. One, not too exciting example is the following. The Absorption Law

$$p \Box \rightarrow (q \Box \rightarrow r) \vdash (p \wedge q) \Box \rightarrow r$$

defines the index principle

$$\forall xyz(C_xyz \rightarrow \forall u \neg C_yuz).$$

Better examples are still to be found. Indeed, e.g. the counterfactual logic of Euclidean space, the most natural geometric representation of our similarity pictures, is still a mystery.

### 3.3 Intuitionistic Logic

Constructive conditional logic is only a part of the full intuitionistic logic, whose Kripke semantics extends the earlier constructive models. In this section, a sketch will be given of an Intuitionistic Correspondence Theory. (For details on intuitionistic logic, cf. van Dalen's chapter in volume 7 of this *Handbook*.)

*Kripke semantics, intermediate axioms and correspondence.*

DEFINITION 81. An *intuitionistic Kripke model*  $M$  is a tuple  $\langle W, \sqsubseteq, V \rangle$ , where  $\sqsubseteq$  is a partial order ('possible growth') on  $W$  ('stages of knowledge'). The valuation  $V$  assigns  $\sqsubseteq$ -closed subsets of  $W$  to proposition letters ('cumulation of knowledge').

The truth definition has the following familiar pattern,

$$\begin{array}{ll} M \not\models \perp[w] & \text{for all } w \in W, \\ M \models \varphi \rightarrow \psi[w] & \text{if } M \models \psi[v] \text{ for all } v \supseteq w \text{ such that } M \models \varphi[v], \\ M \models \varphi \wedge \psi[w] & \text{if } M \models \varphi[w] \text{ and } M \models \psi[w], \\ M \models \varphi \vee \psi[w] & \text{if } M \models \varphi[w] \text{ or } M \models \psi[w]. \end{array}$$

Negation is defined as usual ( $\neg\varphi$  becoming  $\varphi \rightarrow \perp$ ).

The pre-condition of partial order was motivated earlier on. But, other choices may be defended as well. As is well-known, the above semantics was derived from the modal one, through the *Gödel translation*  $g$ :

$$\begin{aligned} g(p) &= \Box p \\ g(\varphi \rightarrow \psi) &= \Box(g(\varphi) \rightarrow g(\psi)) \\ g(\varphi \wedge \psi) &= g(\varphi) \wedge g(\psi) \\ g(\varphi \vee \psi) &= g(\varphi) \vee g(\psi) \\ g(\perp) &= \perp. \end{aligned}$$

Now, there is a whole range of modal logics whose ‘intuitionistic fragment’ (through  $g$ ) coincides with intuitionistic propositional logics. Amongst others, we have the

**THEOREM 82.** *Let  $X$  be any modal logic in the range from **S4** to **S4.Grz** = **S4** plus the Grzegorzcyk Axiom*

$$\Box(\Box(p \rightarrow \Box p) \rightarrow p) \rightarrow p.$$

*Then, for all intuitionistic formulas  $\varphi$ ,  $\varphi$  is intuitionistically provable in Heyting’s logic if and only if  $g(\varphi)$  is a theorem of  $X$ .*

The earlier modal correspondences yield a corresponding semantic range, between ‘pre-orders’ (*reflexive* and *transitive*) and ‘trees’:

**EXAMPLE 83.** Grzegorzcyk’s Axiom defines the combination of (i) reflexivity, (ii) transitivity, and (iii) well-foundedness in the following sense: ‘from no  $w$  is there an ascending chain  $w = w_1 \subseteq w_2 \subseteq \dots$  with  $w_i \neq w_{i+1}$  ( $i = 1, 2, \dots$ )’.

**Proof.** This goes more or less like the closely related Axiom of Löb. By the way, notice that (iii) implies anti-symmetry. Note also that, semantically, Grzegorzcyk’s axiom alone implies the **S4**-laws: syntactic derivations to match were found around 1979 by W. J. Blok and E. Pledger. ■

Thus, a case may also be made for the *Tree of Knowledge* as a basis for intuitionistic semantics. Nevertheless, we shall stick to partial orders for a start.

Above **S4Grz**, modal logics start producing greater  $g$ -fragments — the so-called *intermediate logics*, ascending to full classical logic. Intermediate axioms impose various restrictions on the pattern of growth for knowledge, classical logic forcing the existence of single (‘complete’) nodes.

**EXAMPLE 84.** (i) Excluded Middle  $p \vee \neg p$  defines  $\forall x \forall y (x \subseteq y \rightarrow x = y)$ .

**Proof.** ‘ $\Leftarrow$ ’ is immediate.

‘ $\Rightarrow$ ’: Suppose  $x \subseteq y, x \neq y$ . (By anti-symmetry then  $y \not\subseteq x$ .) Set  $V(p) = \{z \mid y \subseteq z\}$ . This falsifies both  $p$  and  $\neg p$  at  $x$ . ■

(ii) *Weak Excluded Middle*  $\neg p \vee \neg \neg p$  defines directedness.

**Proof.** ‘ $\Leftarrow$ ’: Suppose that  $\neg p$  fails at  $x$ ; say  $p$  holds at  $y \supseteq x$ . Then consider any  $z \supseteq x$ . As it shares a common successor with  $y$ , and  $V(p)$  is  $\subseteq$ -hereditary, it has a successor verifying  $p$ , whence  $\neg p$  fails at  $z$ . So  $\neg \neg p$  holds at  $x$ .

‘ $\Rightarrow$ ’: Suppose that  $x \subseteq y, z$ , where  $y, z$  share no common successors. Set  $V(p) = \{u \mid z \subseteq u\}$ . (Like above, this is a  $\subseteq$ -closed set.) Notice that  $x, y \notin V(p)$ . It follows that  $\neg p$  fails at  $x$  (consider  $z$ ), but  $\neg \neg p$  fails as well (consider  $y$ ). ■

(iii) *Conditional Choice*  $(p \rightarrow q) \vee (q \rightarrow p)$  defines connectedness.

**Proof.** ‘ $\Leftarrow$ ’: Suppose that  $p \rightarrow q$  fails at  $x$ ; i.e. some  $y \supseteq x$  has  $p$  true, but  $q$  false. Now consider any  $z \supseteq x$  such that  $q$  holds. Either  $z \subseteq x$ , but then, by  $\subseteq$ -heredity,  $q$  is true at  $y$  (*quod non*), or  $y \subseteq z$ , and so, again by  $\subseteq$ -heredity,  $p$  is true at  $z$ , i.e.  $q \rightarrow p$  is true at  $x$ .

‘ $\Rightarrow$ ’: Let  $x \subseteq y, z$  with  $y \not\subseteq z, z \not\subseteq y$ . Set  $V(p) = \{u \mid y \subseteq u\}, V(q) = \{u \mid z \subseteq u\}$ . Then  $p \rightarrow q$  fails at  $x$  (watch  $y$ ), and  $q \rightarrow p$  fails as well (watch  $z$ ). ■

Much more forbidding principles than these have been proposed as intermediate axioms. But surprisingly, these usually turned out to be first-order definable:

EXAMPLE 85. (i) The Stability Principle  $(\neg \neg p \rightarrow p) \rightarrow (p \vee \neg p)$  defines

$$\forall x \neg \exists y z (x \subseteq y \wedge x \subseteq z \wedge \neg \exists u (y \subseteq u \wedge z \subseteq u) \wedge \wedge \forall u (\forall s (u \subseteq s \rightarrow \exists t (s \subseteq t \wedge z \subseteq t)) \rightarrow \neg \exists v (u \subseteq v \wedge y \subseteq v))).$$

(ii) The Kreisel-Putnam Axiom  $(\neg p \rightarrow (q \vee r)) \rightarrow ((\neg p \rightarrow q) \vee (\neg p \rightarrow r))$  defines

$$\forall x \neg \exists y z (x \subseteq y \wedge x \subseteq z \wedge \neg y \subseteq z \wedge \neg z \subseteq y \wedge \wedge \forall u ((x \subseteq u \wedge u \subseteq y \wedge u \subseteq z) \rightarrow \exists v (u \subseteq v \wedge \neg y \subseteq v \wedge \neg z \subseteq v))).$$

No matter how complex such axioms may seem at first sight, proofs of the above assertions are quite simple exercises in ‘imagining what a counterexample would look like’.

This recurrent experience led to the following *conjecture* in [van Benthem, 1976]:

All intermediate axioms express first-order constraints on growth of knowledge.

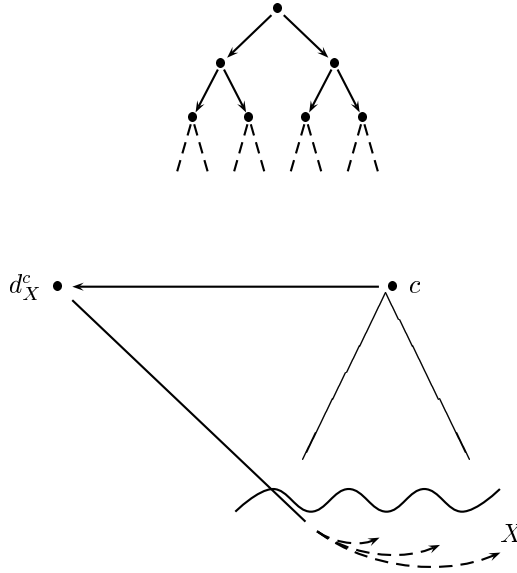
*Two conjectures refuted.* The earlier hope was all but given up in the first version of this chapter; as ‘Scott’s Rule’ turned out to be an essentially

higher-order intermediate inference. The relevant argument was sharpened somewhat by P. Rodenburg:

**THEOREM 86.** *Scott's Axiom  $((\neg\neg p \rightarrow p) \rightarrow (p \vee \neg p)) \rightarrow (\neg p \vee \neg\neg p)$  defines no first-order condition on partial orders.*

**Proof.** An elaborate Löwenheim–Skolem argument works, in the spirit of Example 43. As an illustration of the non-triviality of our present subject matter, it follows here.

*Step 1:* Consider the following Kripke frame  $\langle W, \subseteq \rangle$ :



$W$  consists of the *infinite binary tree*  $T$ , together with, for each node  $c$  in  $T$  and each  $\subseteq$ -hereditary, *cofinal* set  $X$  in  $T_c$  (i.e. the subtree with root  $c$ ), some point  $d_X^c$ .  $\subseteq$  is the usual order on  $T$ , together with

- $c \subseteq d_X^c \subseteq x$ , for all  $x \in X$
- $d_X^c \subseteq d_{X'}^c$ , if  $X' \subseteq X$ .

CLAIM. *Scott's Axiom is true in  $\langle W, \subseteq \rangle$ .*

PROOF. First, let  $c \in T$  be a putative refutation. I.e., for some valuation  $V$ ,

1.  $(\neg\neg p \rightarrow p) \rightarrow p \vee \neg p$  is true at  $c$ ,
2.  $\neg p \vee \neg\neg p$  is false at  $c$ .

Then consider the node  $d_X^c$ , where  $X$  is the cofinal hereditary set

$$T_c \cap (V(p) \cup V(\neg p)).$$

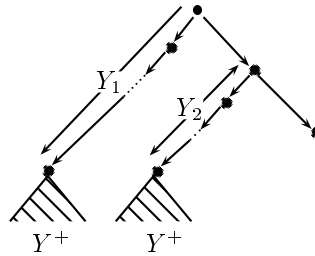
One verifies successively that  $\neg\neg p \rightarrow p$  is true at  $d_X^c$ , whereas both  $p, \neg p$  are false. (E.g. if  $p$  were true at  $d_X^c$ , then  $p$  is true throughout  $X$ , whence  $\neg\neg p$  is true at  $c$  — whereas (2) says the opposite.) Thus, we have a contradiction with (1).

A similar argument works for the case where  $c$  is of the form  $d_X^c$  itself. ■

*Step 2:* A matter of cardinality:

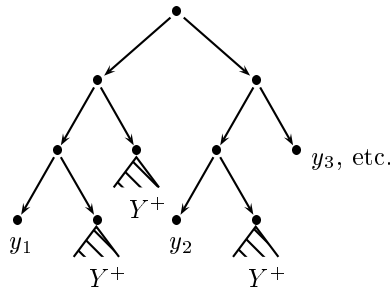
CLAIM. *The above Kripke frame is uncountable.*

PROOF. In particular, there are  $2^{\aleph_0}$  nodes of the form  $d_X^c$ . For, each subset  $Y$  of  $N$  may be coded as follows, using (distinct) hereditary cofinal subsets  $Y^+$  of the infinite binary tree. Let  $Y = \{y_1, y_2, y_3, \dots\}$ .



etc. going down the extreme right branch using the extreme left branches to code  $y_1, y_2, y_3, \dots$

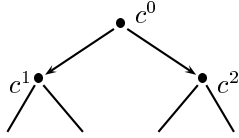
For all nodes not arrived at in this way, one makes  $Y^+$  cofinal by means of the following stipulation:



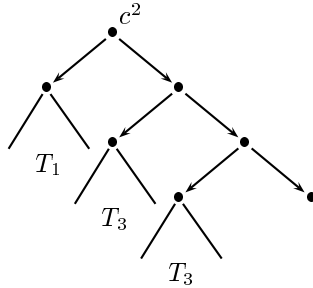
*Step 3:* Take any countable elementary substructure  $F$  of  $\langle W, \subseteq \rangle$  containing the original binary tree. ■

CLAIM. *Scott's Axiom may be falsified in  $F$ .*

PROOF. Consider  $T$  as a double tree



and again  $T_{c^2}$  a countable sequence of 'trees on a string':



Let  $D_{X_1}, D_{X_2}, \dots$  be an enumeration of the points  $d_X^{C_0}$  remaining in  $F$ . Notice that, for each  $i \in N$ ,

1. finite intersections  $T_i \cap X_1 \cap \dots \cap X_n$  are still hereditary cofinal in  $T_i$ ,
2. the total intersection  $T_i \cap \{X_j \mid j = 1, 2, \dots\}$  is empty.

As for the latter observation, it suffices to see that the assertion

$$\forall x \exists d_X^{C_0} \text{ with } d_X^{C_0} \not\subseteq x,$$

which holds in  $\langle W, \subseteq \rangle$ , can be expressed in first-order terms in  $\langle W, \subseteq \rangle$ ; and, hence, it has remained valid in the elementary substructure  $F$ .

Now, define

$$\begin{aligned} X_1^* &= X_1 \\ X_{n+1}^* &= X_1 \cap \dots \cap X_k \text{ for the smallest } k \text{ such that } T_{n+1} \cap X_1 \cap \\ &\quad \dots \cap X_k \subsetneq T_{n+1} \cap X_n^*. \end{aligned}$$

Scott's Axiom may now be falsified at  $c^0$ , by setting

$$X^* = \cup \{T_i \cap X_i^* \mid i = 1, 2, \dots\}, V(p) = \{y \mid \exists x \subseteq y \ x \in X^*\}.$$

to see this, notice, that successively,

1. each point  $d_{X_i}$  has a successor (in  $T_i$ ) outside of  $V(p)$ ,
2.  $(\neg\neg p \rightarrow p) \rightarrow p \vee \neg p$  holds at  $c^0$ ,

3.  $\neg p \vee \neg \neg p$  fails at  $c^0$ . ■

We conclude that Scott's Axiom is not first-order definable — not being preserved under elementary subframes. ■

This complex behaviour disappears on better-behaved structures.

OBSERVATION 87. On *trees*, Scott's Axiom defines the first-order condition

$$\forall x \neg \exists y z u (x \subseteq y \wedge x \subseteq z \wedge z \subseteq u \wedge z \neq u \wedge \neg \exists v (y \subseteq v \wedge z \subseteq v)).$$

This, and other experiences of its kind, led to a revised guess in the first version of this chapter: On trees, all intermediate axioms express first-order constraints on descent. A proof sketch was added, involving *semantic tableaux* as 'patterns of falsification', to be realised in trees.

This conjecture was 'almost' refuted in [Rodenburg, 1982]. The semantic tableau method runs into problems with *disjunctions*, and indeed we have the following counter-example.

EXAMPLE 88. Consider the formula

$$\phi = ((\neg p \wedge \neg q \wedge \neg r) \rightarrow (p \wedge q \wedge r)) \rightarrow (\neg p \wedge \neg q \wedge \neg r)$$

with the simultaneous substitution of:  $p \& q$  for  $p$ ,  $p \& \neg q$  for  $q$ , and  $\neg p \& q$  for  $r$ . This  $\phi$  is not first-order definable on partial orders. On suitably tree-like structures, it expresses the lack of '3-forks' of immediate successors as well as the absence of infinite comb-like structures.

On trees, this negative example probably still works — but there is an instructive difficulty here. The class of trees itself has a higher-order definition;  $\Pi_1^1$ , to be precise. Therefore, current model-theoretic arguments for disproving first-order definability (compactness, Löwenheim–Skolem) run the risk of employing constructions leading outside of this class. Higher-order preconditions are a problem for our Correspondence Theory.

To illustrate this from a purely classical angle, the reader may consider a related problem, showing how soon the familiar methods of model theory fail us. *Finiteness* is first-order undefinable on *partial orders*, even on *trees*. It is thus definable on *linear trees*, however, viz. by 'every non-initial node has an immediate predecessor'. What about the (at most) binary trees? This intermediate case seems to be open.

*The state of the subject.* The progress of science is sometimes startling. Where the first version of this chapter (1981) had some tentative examples, enlightenment reigns in the report [Rodenburg, 1982]. Of its many topics, only a few will be mentioned here.

First, there are several semantic options — as indicated above, ranging from *partial orders* via '*downward linear orders*' to *trees*. But moreover,

there is a legitimate choice of language. Despite appearances, it is the *disjunction* clause which is now strongly constructive in intuitionistic Kripke semantics. ('Choose *now!*' Classical logic would have a more humane clause in this setting:  $\Box\Diamond(\varphi \vee \psi)$ , i.e. ' $\varphi$  or  $\psi$  eventually'.) Thus, it is of interest to consider both the full language and its  $\vee$ -free fragment.

The semantic tableau method mentioned above, in combination with the above counter-examples, has led to the results in the following scheme:

All formulas first-order definable	Partial orders	Downward linear orders	Trees
without $\vee$	YES	YES	YES
with $\vee$	NO	NO	?

But there are also matters of 'fine structure'. For instance, Scott's Axiom had only one proposition letter — and for such intuitionistic formulas we have the beautiful Rieger–Nishimura lattice. Now, Scott's Axiom merely seemed a fit candidate for a counter-example among the intermediate axioms existing in the literature. Rodenburg has proved that it is also *minimal* in the Rieger–Nishimura lattice with respect to non first-order definability. (More precisely, an intuitionistic formula with one proposition letter is first-order definable on the partial orders if and only if it is equivalent to one of  $A_1, \dots, A_9$  in the lattice.)

In the counter-examples needed for the latter result, a uniform method may be seen at work: *compactness*, in the form that sets of formulas which are finitely satisfiable in *finite* models are also simultaneously satisfiable (in some *infinite* model). Now, indeed, intuitionistic truth has a close connection with truth in finite submodels (cf. [Smoryński, 1973]). Our question is whether this may lead to the following improvement in the mathematical characterisation of first-order definability as given in Section 2.2.

CONJECTURE. An intuitionistic formula  $\varphi$  is first-order definable if and only if  $\varphi$  is preserved under *ultraproducts of finite frames*.

*Intuitionistic definability.* As with the direction 'from intensional to classical', the case 'from classical to intuitionistic definability' shows many resemblances with our earlier modal study. For instance, a Goldblatt–Thomason type characterisation was proved in [van Benthem, 1983] (cf. our earlier Theorem 66):

A first-order constraint on the growth pattern is intuitionistically definable if and only if it is preserved under the formation of *generated subframes*, *disjoint unions*, *zigzag-morphic images*, *filter extensions* and '*filter inversions*'.



Merely in order to illustrate this topic, which has a wider semantic significance, here is a sketch of the representation theory in the background.

On the algebraic side, the intuitionistic language may be interpreted in *Heyting Algebras*  $\langle A, 0, 1, +, \cdot, \Rightarrow \rangle$  satisfying suitable postulates. Now, each *Kripke (general) frame* in the above sense induces such a Heyting Algebra, through its  $\subseteq$ -hereditary sets, provided with suitable, obvious operations. But also conversely, a *filter representation* now takes Heyting Algebras to Kripke general frames. Indeed, the earlier categorial duality (cf. Section 2.3) is again forthcoming.

The more general semantic interest of the construction is this. Despite the superficial similarity with structures consisting of the ‘complete’ possible worlds, intuitionistic Kripke models should be regarded as patterns of stages of *partial* information. This comes out quite nicely in the above representation, where ‘worlds’ are no longer complete *ultrafilters*, but merely filters (in the  $\vee$ -free case) or ‘splitting’ filters (for the full language). Filters  $F$  merely satisfy the closure condition that

$$a, b \in F \text{ iff } a \cdot b \in F,$$

a minimal requirement on partial information. Also quite suggestively, the ‘modal’ alternative relation collapses into inclusion (‘growth’):

$$\forall a \Rightarrow b \in F \forall a \in F' : b \in F' \quad \text{iff} \quad F \subseteq F'.$$

The present-day supporters of ‘partial models’ and ‘information semantics’ would do well to study intuitionistic logic.

*Predicate logic.* Again, correspondence phenomena do not stop at the frontier of predicate logic. This will be illustrated by means of some intuitionistic examples.

*Kripke models*  $M = \langle W, \subseteq, D, V \rangle$  will now be of the usual variety; in particular satisfying

1.  $\forall xy(x \subseteq y \rightarrow D_w \subseteq D_v)$  (*monotonicity*)
2.  $\forall xy(x \subseteq y \rightarrow \forall \vec{d} \in D_x (V_x(P, \vec{d}) = 1 \rightarrow V_y(P, \vec{d}) = 1)$  (*heredity*).

But other varieties, say with maps between the domains (cf. [Goldblatt, 1979]) would be suitable as well.

The ‘de re/de dicto’ interchange principles of Section 2.5 now have their obvious counterparts in the following quartetto:

1.  $\neg \exists x Ax \rightarrow \forall x \neg Ax,$
2.  $\forall x \neg Ax \rightarrow \neg \exists x Ax,$
3.  $\exists x \neg Ax \rightarrow \neg \forall x Ax,$

$$4. \neg\forall xAx \rightarrow \exists x\neg Ax.$$

The first three of these are universally valid on the present semantics. That they already hide quite some complexity is shown by the Gödel translation of (3):

$$\Box(\exists x\Box\neg\Box Ax \rightarrow \Box\neg\Box\forall x\Box Ax),$$

or

$$\Box(\exists x\Box\Diamond\neg Ax \rightarrow \Box\Diamond\exists x\Diamond\neg Ax).$$

No wonder that (3), e.g. does not define precisely the above monotonicity constraint on domains — even though its modal cousin  $\exists x\Box Ax \rightarrow \Box\exists xAx$  did.

The first really complex principle in Section 2.5 was the converse implication  $\Box\exists xAx \rightarrow \exists x\Box Ax$ . We shall now investigate its intuitionistic cousin (4) — a rejected classical law.

EXAMPLE 89.

1.  $\neg\forall xAx \rightarrow \exists x\neg Ax$  implies that all domains are equal:

$$\forall xy(x \subseteq y \rightarrow D_x = D_y)$$

2. On frames with constant finite domain,  $\neg\forall xAx \rightarrow \exists x\neg Ax$  expresses the first-order condition that

$$\forall x(\exists!d d \in D_x \vee \forall y(x \subseteq y \rightarrow \forall z(x \subseteq z \rightarrow \exists u(y \subseteq u \wedge z \subseteq u))).$$

**Proof.** Ad 1. Suppose that  $x \subseteq y$ , but  $D_x \not\subseteq D_y$ . Make  $A$  true at  $y$  for all  $d \in D_x$ , and similarly at all  $y' \supseteq y$ . This stipulation defines an admissible assignment verifying  $\neg\forall xAx$  at  $x$ , while falsifying  $\exists x\neg Ax$ .

Ad 2. First, if  $|D_x| = 1$ , then trivially,  $\neg\forall xAx \rightarrow \exists x\neg Ax$  holds at  $x$ . (Recall that all domains are equal.)

Next, if  $|D_x| > 1$ , then one may argue as follows. If  $\subseteq$  is *directed* above  $x$  in the above sense, then the assumption that  $\exists x\neg Ax$  *fails* at  $x$  can be exploited to show that  $\neg\forall xAx$  must fail as well.

For, let  $D_x = \{d_1, \dots, d_k\}$ . By the assumption,  $Ad_i$  will be true at some  $x_i \supseteq x$  ( $1 \leq i \leq k$ ). Then, by successive applications of directedness, there will be found a common successor  $y \supseteq x_1, \dots, y \supseteq x_k$ , where  $\forall xAx$  is true (by heredity). This falsifies  $\neg\forall xAx$  at  $x$ .

If on the other hand, for some  $x$ ,  $|D_x| > 1$  while  $\subseteq$  is not directed above  $x$ , then, say, there exist  $x_1 \supseteq x, x_2 \supseteq x$  without common successors. Then pick any object  $d \in D_x$ , making  $A$  true at  $x_1$  and all its  $\subseteq$ -successors *for all objects except  $d$* ; while making  $A$  true at  $x_2$  and all its  $\subseteq$ -successors *for  $d$  only*. This assignment verifies  $\neg\forall xAx$  at  $x$ , while falsifying  $\exists x\neg Ax$ . ■



*Post-Script: quantum logic.*

Correspondences have not proved uniformly successful in intensional contexts. It seems only fair to finish with a more problematic example.

A possible worlds semantics for *quantum logic* was proposed in [Goldblatt, 1974]. Kripke frames are now regarded as sets of ‘states’ of some physical system, provided with a relation of ‘orthogonality’ ( $\perp$ ). From its physical motivation, two pre-conditions follow for  $\perp$ , viz. *irreflexivity* and *symmetry*. But in addition, there is also a restriction to ‘admissible ranges’ for propositions, in the sense that these sets  $X \subseteq W$  are to be *orthogonally closed*:

$$\forall x \in (W - X) \exists y \in (W - X) (\neg x \perp y \wedge \forall z \in X y \perp z).$$

The key truth clauses are those for conjunction (interpreted as usual), and negation, interpreted as follows:

$$\neg\varphi \text{ is true at } x \quad \text{if } x \text{ is orthogonal to all } \varphi\text{-worlds.}$$

This semantics validates the usual principles for quantum logic, when  $\vee$  is defined in terms of  $\neg, \wedge$  by the De Morgan law. But, one key principle remains invalid, viz. the *ortho-modularity axiom*

$$p \leftrightarrow (p \wedge q) \vee (p \wedge \neg(p \wedge q)).$$

This axiom has a natural motivation in the Hilbert Space semantics for quantum logic — being the key stone in the representation of ortho-modular lattices as subspace algebras of suitable vector spaces. Thus, a minimal expectation would be that an enlightening correspondence is forthcoming with some constraint on the orthogonality relation  $\perp$ .

In reality, no such thing has happened. Quantum logicians pass onto *general* frames, into whose very definition validity of ortho-modularity has been built in. Despite this cover-up, the fact remains that the relational possible worlds perspective fails to do its correspondence duties here. A set-back, or an indication that facile *over-applicability* of Kripke semantics need not be feared for?

#### 4 CONCLUSION

At a purely technical level, Correspondence Theory is an applied subject. Classical tools have been borrowed from model theory and universal algebra. In return to these mother disciplines, the subject offers a good range of (counter-)examples, as well as prospects for generalisability to other suitably chosen fragments of higher-order logic. (Cf. [van Benthem, 1983].)

From a more philosophical point of view, the whole enterprise may be described as finding out what possible worlds semantics really does for us.

It is one thing to make conceptual proposals, and another to really probe their depths. The systematic study of connections between intensional and classical perspectives upon possible world structures is an exploration of the benefits gained by the semantics. This chapter started with the observation that ‘complex’ modal axioms turned out to express ‘simple’ classical requirements (i.e. first-order ones). We have investigated the range and limits of this, and related phenomena. Especially these limits have become quite clear — and, with them, the limits of fruitful application of Kripke semantics. This philosophical conclusion holds for all semantics, of course. But we have earned the moral right to say it, through honest toil.

#### ACKNOWLEDGEMENTS

The classical introduction to a systematic modal model theory remains [Segerberg, 1971]. Some first applications of more sophisticated tools from classical model theory may be found in [Fine, 1975]. The algebraic connection was developed beyond the elementary level by L. Esakia, S. K. Thomason, R. I. Goldblatt and W. J. Blok. Two good surveys are [Blok, 1976] and [Goldblatt, 1979]. The proper perspective upon modal logic as a fragment of second-order logic was given in [Thomason, 1975]. An early appearance of correspondence theory proper is made in [Sahlqvist, 1975], full surveys are found in [van Benthem, 1983] for the case of modal logic and [van Benthem, 1985] for the case of tense logic. Other case studies are still in a preliminary state, with the exception of the intuitionistic treatise [Rodenburg, 1982].

*University of Amsterdam*

#### APPENDIX (1997)

This chapter first appeared in 1984. In the meantime, Modal Logic has evolved, but the basic structure of our original presentation remains valid. Therefore, we have left the old text unchanged, and merely added a short chronicle of further developments, including some answers to open questions. Generally speaking, correspondence methods have become a useful technical tool in pure and applied Modal Logic, without forming a major research area in their own right. A more principled motivation is given in van Benthem [1996a], where correspondence analysis is viewed as a central part in the philosophical quest for logical ‘core theories’ of semantic phenomena in language and computation. In particular, correspondences suggest the introduction of new many-sorted models, inducing decidable geometries of ‘states’ and ‘paths’ in the study of time and computation.

### *Extensions to Other Branches of Intensional Logic*

The first significant extension of correspondence theory concerns *Intuitionistic Logic*. This involves the new feature that all valuations must be restricted to hereditary ones, leading only to formulas whose truth is preserved upward in the relational ordering. Rodenburg [1986] investigates this area in detail. In particular, he shows that the implication-conjunction fragment is totally first-order, whereas disjunctions can lead to non-first-orderness. Moreover, he introduces semantic tableau methods for explicit description of first-order correspondents. A final interesting feature is Rodenburg's analysis of intuitionistic Beth models which employ a second-order truth condition: a disjunction is true when its disjuncts 'bar' all future paths. These also turn out to be amenable to correspondence analysis, over two-sorted frames with both points and paths. Restricted valuations also occur with the ternary relational models of *Relevant Logic*. A full correspondence analysis is given in Kurtonina [1995], which analyses the special effects of working with features like distinguished points (actual worlds), non-standard connectives (including a new product conjunction), as well as the much poorer non-Boolean fragments found in categorial logics for grammatical analysis (cf. [van Benthem, 1991; Moortgat, 1996]). Further extensions have been made to *Epistemic Logic* [van der Hoek, 1992] and *Partial Logics* [Thijssse, 1992; Jaspars, 1994; Huertas, 1994]. Correspondence with restricted valuations for 'convex' propositions has also been proposed in standard *Temporal Logic* (cf. van Benthem [1983; 1986; 1995b]). But also, most axioms for richer interval-based versions have first-order 'Sahlqvist forms' [Venema, 1991]. Zanardo [1994] gives correspondences for modal-temporal models of branching space-time. Finally, correspondence methods have turned out very useful in *Algebraic Logic*. Venema [1991], Marx and Venema [1996] present a systematic study of relational algebra and cylindric algebra along these lines, pointing out the Sahlqvist form of most familiar algebraic axioms, and calculating their frame constraints on algebraic 'atom structures'. This establishes a much wider bridge between algebraic logic and modal logic than our earlier duality.

### *Restricted Frame Classes*

Correspondence behaviour may change on special frame classes. In this chapter, we have looked at some effects of a restriction to transitive frames. But one can also investigate non-first-order frame classes. Van Benthem [1989a] considers *finite frames*, where, amongst others, the McKinsey axiom still defines a non-first-order condition. In this area, standard compactness-based model-theoretic techniques no longer work, and they must be replaced by a more careful combinatorial analysis with Ehrenfeucht-Fraïssé games of model comparison. (More generally, the finite model theory of modal logic

is still undeveloped. Rosen [1995] proves some interesting transfer results, showing better finite model-theoretic behaviour than for first-order logic in general.) Doets [1987] takes up modal Ehrenfeucht games in great depth, investigating, amongst others, correspondence over *countable* and over *well-founded* frames. (For instance, the so-called Fine Axiom turns out to be first-order over countable frames.)

### *Complexity*

This chapter contains some results on the (high) complexity of definability problems for monadic  $\Phi_1^1$ -formulas. It turns out much harder to deal with the modal fragment of these. A lower bound for the complexity of first-orderness of modal formulas has been found in Chagrova [1991]: **M1** is *undecidable*. It seems likely that her methods (involving reductions of Minsky machine computation to correspondence statements) can also be made to yield non-arithmetical complexity. Conversely, undecidability of modal definability for first-order statements has been proved by Wolter [1993]: that is, **P1** is undecidable, too. A more general investigation of time and space complexity for modal logics, and the ‘jumps’ that may occur with different operator vocabularies, may be found in Spaan [1993]. It has improved decidability results for the so-called ‘subframe logics’ defined in Fine [1985], as well as ‘transfer’ of complexity bounds from components to compounds in poly-modal logics (cf. [Kracht and Wolter, 1991]).

### *Correspondence and Completeness*

The main business of modal logic has been the search for completeness theorems over various frame classes. Correspondence theory bypasses this deductive information, focussing on direct semantic definability. Nevertheless, Kracht [1993] shows how the two enterprises can be merged, by a suitably generalized form of modal definability. Perhaps the most powerful result of this kind is the generalized Sahlqvist Theorem in Venema [1991], which shows that over suitably rich modal languages (possessing matched versions for each modality accessing all directions of its alternative relation), and allowing natural additional rules of inference beyond the minimal modal logic, the correspondence and the completeness version of the Sahlqvist Theorem converge in their proofs. The essential observation in the argument is as follows. In standard Henkin models for these richer systems, unlike in the standard case, all definable subsets employed in the correspondence proof (such as singletons or successor sets) are modally definable. Direct frame correspondences for modal rules of inference may be found in van Benthem [1985]. Over frames, the latter correspond to non- $\Phi_1^1$  second-order formulas, but except for a few scattered observations in the literature, correspondence theory for modal rules of inference remains underexplored.

### *Duality with Algebraic Logic*

Algebraic methods have been invaluable in finding key results on correspondence, such as the Goldblatt-Thomason characterization of the modally definable first-order formulas. Nevertheless, a purely model-theoretic re-analysis has been given in van Benthem [1993b], revolving around saturated models instead of descriptive frames. There is no definite preference here, as it is precisely the interplay between algebraic and model-theoretic viewpoints that remains fruitful. For new uses of correspondence methods in algebraic logic, as well as new set-theoretic representations for Boolean algebras with additional modal operators, see Marx [1995], Mikulas [1995]. For instance, Marx has an in-depth study of the duality between algebraic amalgamation and logical interpolation. The latter methods no longer employ simple binary relations as in the Jónsson-Tarski Stone representation, but more complex set-theoretic constructs. (Modal correspondences over finitary relations occur in van Benthem [1992], with a finite neighbourhood semantics for logic programs.) Developing a systematic correspondence theory over such generalized relational structures then becomes the next challenge.

### *Extended Modal Logics*

Perhaps the most striking development in modal logic over the past ten years has been the systematic use of more powerful formalisms, with stronger modal operators over relational frames. A straightforward step is ‘poly-modal logic’, which gives the same expressive power over frames with more alternative relations. Examples of the latter trend are the indexed modalities  $\langle i \rangle$  of propositional dynamic logic (cf. [Harel, 1984; Goldblatt, 1987; Harel *et al.*, 1998]), or  $n$ -ary modalities accessing  $(n + 1)$ -ary alternative relations, as happens in relevant or categorial logics (cf. [Dunn, 2001; Kurtonina, 1995]). The correspondence theory of such extensions is straightforward, whereas there are interesting issues of ‘transfer’ for axiomatic completeness, finite model property, or computational complexity: cf. [Spaan, 1993; Fine and Schurz, 1996]. Transfer may depend very much on the connections between the various modalities. A case in point is modal predicate logic, whose theory has rapidly expanded over the past decade. Van Benthem [1993a] surveys some striking contributions by Ghilardi and Shehtman.

More interesting, from a correspondence point of view, is an increase in expressive power over the original binary relational frames. For temporal logic, the latter research line was initiated by Kamp’s Theorem on functional completeness of the {Since, Until} language over continuous linear orders. In modal logic, the first systematic work emanated from the ‘Sofia School’: cf., e.g., [Gargov and Passy, 1990; Goranko, 1990], Vakarelov [1991; 1996]. These papers study addition of various new operators, such as a universal



modality ranging over all worlds (relationally accessible or not), or various operations on poly-modalities, such as ‘program intersection’. New frame constructions were invented to deal with these, such as ‘duplication’. De Rijke [1992] investigates the ‘difference modality’ (“in at least one different world”), which has turned out to be useful and yet tractable. A more general program for extending modal logic (viewed as a general ‘theory of information’) occurs in van Benthem [1990] but the technical perspective is also clear in the pioneering paper Gabbay [1981]. Finally, de Rijke [1993] is an extensive model-theoretic investigation of definability and correspondence for extended modal languages, producing generalized versions for many results in this chapter (such as frame preservation theorems or effective correspondence algorithms). Still another angle on all this will follow below.

### *Alternatives: Direct Frame Theory*

One may also analyze the frame content of modal logics more directly in terms of mathematical properties of graphs. Fine [1985] is a pioneer of this trend, emphasizing the good behaviour of ‘subframe logics’ which are complete for frame classes that are closed under taking subframes. (Such logics make no ‘existential commitments’.) First-orderness is not a prominent consideration here: e.g., Löb’s Axiom defines a simple subframe logic. Zakharyashev [1992; 1995] is a sophisticated study of modal logic from this viewpoint. Nevertheless, his direct classification of modal logics into three stages of frame preservation behaviour may again be reflected in second-order syntax and hence result in a form of correspondence theory at that higher level. A forthcoming monograph by Chagro and Zakharyashev provides much more background, including references to earlier Russian sources (going back to Jankov in the sixties). Another excellent source, for many of the topics listed here, is the survey chapter [Chagro *et al.*, 1996].

### *Models, Bisimulation and Invariance*

Another noticeable shift of emphasis in the current literature leads away from frames to *models* as the primary objects of semantic interest. This move makes all of basic modal logic first-order, via our standard translation. The main questions then address what makes modal logics special as subspecies of first-order logic. In particular, what is the basic semantic invariance for basic modal logic, which should play a role like Ehrenfeucht games or ‘partial isomorphism’ in first-order model theory? A key result here is the semantic characterization of the modal fragment of first-order logic (modulo logical equivalence) as precisely those formulas in one free variable which are invariant for generated submodels and our ‘zigzag relations’ [van Benthem, 1976]. In modern jargon, this says that these for-

mulas are precisely the ones *invariant for bisimulation*. The latter link was also developed in Hennessy & Milner [1985], which matches modal formalisms in different strengths with coarser or finer process equivalences. For up-to-date expositions of the resulting analogies between modal logics and computational process theories, cf. [van Benthem and Bergstra, 1995; van Benthem *et al.*, 1994], as well as various contributions in the volume [Ponse *et al.*, 1995]. This development has led to a new look at connections between modal formalisms and first-order logic. For instance, there are striking analogies between the meta-theories of both logics, whose precise extent and explanation is explored in de Rijke [1993], and Andr eka, van Benthem & N emeti [1998]. In particular, the latter paper investigates the hierarchy of *finite-variable fragments* for first-order logic as a candidate for a general account of modal logic (cf. [Gabbay, 1981; van Benthem, 1991] for this view). Typically, modal formulas need only two variables over worlds in their standard translation, temporal formulas only three, and so on. Finite-variable fragments are natural, and may be considered as functionally complete modal formalisms (cf. the insightful game-based analysis of Kamp's Theorem in Immerman & Kozen [1987]). Nevertheless, Andr eka, van Benthem & N emeti [1998] also turn up an array of negative properties, and eventually propose another classification for modal languages in terms of restricting atoms for *bounded quantifiers*. The resulting 'guarded fragments' can be analyzed much like the basic modal language, including analogous bisimulation techniques. In particular, these bisimulations now relate finite sequences of objects instead of single worlds, as in many-dimensional modal logics (cf. [Marx and Venema, 1996] for the theory of such formalisms). Their correspondence theory, taken with respect to natural generalized frame conditions for arbitrary first-order relations, still remains to be understood. [van Benthem, 1996b] is a general study of dynamic logics for computation and cognition, pursued via these techniques. One of its central concerns is expressive completeness of modal process logics vis- a-vis process equivalences like bisimulation.

### *Connections with Higher-Order Logic and Set Theory*

From first-order correspondence, forays can be made into higher-order definability. Sometimes, this move is suggested by the modal language itself. E.g., in propositional dynamic logic, program iteration naturally translates into a countable disjunction of finite repetitions. Thus, translation into the *infinitary* standard language  $L_{\omega_1\omega}$  seems the evident route. Infinitary frame correspondences were briefly considered in van Benthem [1983], and their modal model theory is explored in [de Rijke, 1993; van Benthem and Bergstra, 1995]. Of course, one may restore a balance here, and consider an infinitary modal counterpart of  $L_\omega$ , allowing arbitrary set conjunctions and disjunctions, which would be the most natural formalism invariant for

bisimulation. Barwise and Moss [1995] take this line, linking up truth on models and correspondence on frames. (Another perspective on infinitary modal logic is given in [Barwise and van Benthem, 1996].) Among a number of original results, they prove that a modal formula has all its infinitary substitution instances true in a model  $M$  iff it is true (in the usual second-order sense) on the frame collapse of that model taken with respect to the maximal bisimulation over  $M$ . As a direct consequence, frame correspondences for modal formulas imply model correspondences in infinitary modal logic. (The issue of good converses is still open). The original motivation for this type of investigation was that it relates modal logics to (non-well-founded) set theories. Linkages of this kind are further explored in d'Agostino [1995] which also raises the issue of more complex correspondences for modal axioms. For instance, she shows that the second-order Löb Axiom holds in a frame iff that frame is transitive while its collapse with respect to the maximal bisimulation is irreflexive. More generally, then, the interesting point about many correspondences is not that they must always reduce modal axioms to first-order ones, but rather the fact that they reformulate modal principles to any more perspicuous classical formalism. Another natural candidate of the latter kind is second-order monadic  $\Phi_1^1$  logic (cf. [Doets and van Benthem, 2001]). In particular, Doets [1989] shows how modal completeness theorems can sometimes be extended to cover this whole language. Moreover, many effective translation methods (see below) turn out to work for this broader language anyway. Finally, van Benthem [1989b] points out how first-order correspondence theory, suitably restated for second-order  $\Phi_1^1$  formulas, is a natural generalization which handles so-called computable forms of Circumscription in the AI literature (which involves reasoning from a second-order 'predicate-minimal' closure for first-order axioms; cf. [Lifshitz, 1985]).

### *Translations*

Correspondence has become a conspicuous theme in the computational literature on theorem proving with intensional logics. A number of algorithms have been proposed, some of them rediscoveries of the Substitution Method and its ilk (cf. [Simmons, 1994]) and even much older results in second-order logic [Doherty, Lukasiewicz and Szalas, 1994], others working with new 'functional' translations better geared towards complete standard Skolemization and Resolution (cf. Ohlbach [1991; 1993]). One interesting feature of some of these algorithms is that they also produce useful equivalents for second-order modal principles. For instance, the typically non-first-order McKinsey Axiom gets a natural equivalent quantifying over both individual worlds and Skolem functions witnessing its (non-Sahlqvist) antecedent. Finally, we mention the use of set-theoretic interpretations of the standard translation in d'Agostino, van Benthem, Montanari & Policriti [1995], which

read the universal modality as describing a power set. This translation also works with an explicit axiom system for general frames plus one axiom stating that the relational successors of any point in a frame form a set. This shift in perspective reduces theorem proving in modal logics to deduction in weak computational set theories. Many of these translations can also be formulated so as to deal with extended modal formalisms or larger fragments of second-order logic.

### *Designing New Logics*

Finally, correspondence techniques have been used in ‘deconstructing’ standard logics and designing new ones. For instance, one can interpret first-order predicate logic over possible worlds models (‘labelled transition systems’) with assignments replaced by abstract states connected by abstract relations  $R_x$  modelling variable shifts. Then, standard predicate-logical validities turn out to express interesting frame properties, constraining possible computations, e.g., by Church-Rosser confluence properties (which match the first-order axiom  $\exists y \forall x \phi \rightarrow \forall x \exists y \phi$ ). Moreover, one may want to impose certain restrictions on admissible valuations, such as ‘heredity constraints’ for axioms  $\text{Py} \rightarrow \forall x \text{Py}$  or  $\text{Py} \rightarrow [y/x]\text{Px}$  (van Benthem [1997; 1996b] have details). These abstract models reflect certain dependencies between admissible object values that may exist for individual variables. This theme is investigated more explicitly in [Alechina and van Benthem, 1993; Alechina, 1995], which design new generalized quantifier logics over ‘dependence models’, first proposed by Michiel van Lambalgen — where again the force of possible axioms is measured at least initially in terms of (Sahlqvist) frame correspondences. Related modal approaches to first-order logic are found in [Venema, 1991; Marx, 1995].

### ADDED IN PRINT (1999)

Handbooks appear according to their own rhythms. Two years have elapsed since the updates were written for this Appendix. Here are a few further items of interest. D’Agostino [1998] contains new material on definability in infinitary modal logics, a topic also pursued further by Barwise and Moss. Meyer Viol [1995] has examples of correspondence for intuitionistic predicate logic showing how intermediate axioms can be quite surprising in their content. Hollenberg [1998] is an extensive study of definability, invariance and safety in modal process languages. Gerbrandy [1998] has interesting theorems on modal definability and bisimulation invariance in a setting of non-well-founded set theory, with applications to dynamic logic of epistemic updates. Grädel [1999] is an excellent survey of progress made on the program of decidable guarded first-order languages extending modal logic,

including also fixed-point operators. Van Benthem [1998] is an up-to-date survey of the definability/correspondence paradigm, and the corresponding ‘tandem approach’ to modal and classical logics. Finally, two modern texts on modal logic that take correspondence seriously are Blackburn, de Rijke and Venema [1999] and van Benthem [1999].

## BIBLIOGRAPHY

- [d’Agostino *et al.*, 1995] G. d’Agostino, J. van Benthem, A. Montanari, and A. Politicri. Modal deduction in second-order logic and set theory. *Journal of Logic and Computation*, 7:251–265, 1997.
- [d’Agostino, 1995] G. d’Agostino. Model and frame correspondence with bisimulation collapses, 1995. Manuscript, Institute for Logic, Language and Computation, University of Amsterdam.
- [d’Agostino, 1998] G. d’Agostino. *Modal Logic and Non-well-founded Set Theory: Bisimulation, Translation and Interpolation*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1998.
- [Ajtai, 1979] M. Ajtai. Isomorphism and higher-order equivalence. *Ann Math Logic*, 16:181–233, 1979.
- [Alechina and van Benthem, 1998] N. Alechina and J. van Benthem. Modal quantification over structured domains. In M. de Rijke, ed., *Advances in Intensional Logic*, pp. 1–28, Kluwer, Dordrecht.
- [Alechina, 1995] N. Alechina. *Modal Quantifiers*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1995.
- [Andréka *et al.*, 1998] H. Andréka, J. van Benthem, and I. Németi. Back and forth between modal logic and classical logic. *Bulletin of the Interest group in Pure and Applied Logic*, 3:685–720, 1995. Revised version: Modal logics and bounded fragments of predicate logic. *Journal of Philosophical Logic*, 27:217–274, 1998.
- [Barwise and Moss, 1995] J. Barwise and L. Moss. *Vicious Circles*. CSLI Publications, Stanford, 1995.
- [Barwise and van Benthem, 1996] J. Barwise and J. van Benthem. Interpolation, preservation, and pebble games. Report ML-96-12, Institute for Logic, Language and Computation, University of Amsterdam, 1996. To appear in *Journal of Symbolic Logic*, 1999.
- [Blackburn *et al.*, 1999] P. Blackburn, M. de Rijke and Y. Venema. *Modal Logic*, Kluwer, Dordrecht, 1999.
- [Blok, 1976] W. J. Blok. *Varieties of interior algebras*. PhD thesis, Mathematical Institute, University of Amsterdam, 1976.
- [Boolos, 1979] G. Boolos. *The Unprovability of consistency*. Cambridge University Press, Cambridge, 1979.
- [Burgess, 1979] J. P. Burgess. Logic and time. *Journal of Symbolic Logic*, 44:566–582, 1979.
- [Burgess, 1981] J. P. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22:76–84, 1981.
- [Chagrova, 1991] L. Chagrova. An undecidable problem in correspondence theory. *Journal of Symbolic Logic*, 56:1261–1272, 1991.
- [Chagrov *et al.*, 1996] A. Chagrov, F. Wolter and M. Zakharyashev. Advanced modal logic. School of Information Science, JAIST, Japan, 1996. To appear in *this Handbook*.
- [Chang and Keisler, 1973] C. C. Chang and H. J. Keisler. *Model Theory*. North-Holland, Amsterdam, 1973.
- [de Rijke, 1992] M. de Rijke. The modal logic of inequality. *Journal of Symbolic Logic*, 57:566–584, 1992.
- [de Rijke, 1993] M. de Rijke. *Extending Modal Logics*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1993.

- [de Rijke, 1993a] M. de Rijke, ed. *Diamonds and Defaults*, Kluwer Academic Publishers, Dordrecht, 1993.
- [de Rijke, 1997] M. de Rijke. *Advances in Intensional Logic*, Kluwer Academic Publishers, Dordrecht, 1997.
- [Doets and van Benthem, 2001] K. Doets and J. van Benthem. Higher-order logic. In *Handbook of Philosophical Logic*, volume 1, 2nd edition, Kluwer Academic Publishers, 2001. (Volume I of 1st edition, 1983.)
- [Doets, 1987] K. Doets. *Completeness and Definability. Applications of the Ehrenfeucht Game in Second-Order and Intensional Logic*. PhD thesis, Mathematical Institute, University of Amsterdam, 1987.
- [Doets, 1989] K. Doets. Monadic  $\Pi_1^1$  theories of  $\Pi_1^1$  properties. *Notre Dame Journal of Formal Logic*, 30:224–240, 1989.
- [Doherty, Lukasiewicz and Szalas, 1994] P. Doherty, W. Lukasiewicz and A. Szalas. Computing circumscription revisited: a reduction algorithm. Technical report LiTH-IDA-R-94-42, Institutionen för Datavetenskap, University of Linköping, 1994.
- [Dunn, 2001] M. Dunn. Relevant logic. In *Handbook of Philosophical Logic*, volume 8, 2nd edition, Kluwer Academic Publishers, 2001. (Volume III of 1st edition, 1985.)
- [Fine and Schurz, 1996] K. Fine and R. Schurz. Transfer theorems for multimodal logics. In J. Copeland, editor, *Logic and Reality. Essays on the Legacy of Arthur Prior*. Oxford University Press, Oxford, 1996.
- [Fine, 1974] K. Fine. An incomplete logic containing **S4**. *Theoria*, 40:23–29, 1974.
- [Fine, 1975] K. Fine. Some connections between elementary and modal logic. In S. Kanger, editor, *Proceedings of the 3rd Scandinavian Logic symposium*. North-Holland, Amsterdam, 1975.
- [Fine, 1985] K. Fine. Logics containing K4, part II. *Journal of Symbolic Logic*, 50:619–651, 1985.
- [Fitch, 1973] F. B. Fitch. A correlation between modal reduction principles and properties of relations. *Journal of Philosophical Logic*, 2, 97–101, 1973.
- [Gabbay, 1976] D. M. Gabbay. *Investigations in Modal and Tense Logics*. D. Reidel, Dordrecht, 1976.
- [Gabbay, 1981] D. Gabbay. Expressive functional completeness in tense logic. In U. Mönnich, editor, *Aspects of Philosophical Logic*, pages 91–117. Reidel, Dordrecht, 1981.
- [Gargov and Passy, 1990] G. Gargov and S. Passy. A note on Boolean modal logic. In P. Petkov, editor, *Mathematical Logic*, pages 311–321. Plenum Press, New York, 1990.
- [Gerbrandy, 1998] J. Gerbrandy. *Bisimulations on Planet Kripke*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1998.
- [Goldblatt and Thomason, 1974] R. I. Goldblatt and S. K. Thomason. Axiomatic classes in propositional model logic. In J. Crossley, editor, *Algebra and Logic*. Lecture Notes in Mathematics 450, Springer, Berlin, 1974.
- [Goldblatt, 1974] R. I. Goldblatt. Semantic analysis of orthologic. *Journal of Philosophical Logic*, 3:19–35, 1974.
- [Goldblatt, 1979] R. I. Goldblatt. Metamathematics of modal logic. *Reports on Math. Logic*, 6:4–77 and 7:21–52, 1979.
- [Goldblatt, 1987] R. I. Goldblatt. *Logics of Time and Computation*. Vol. 7 of CSLI Lecture Notes, Chicago University Press, 1987.
- [Goranko, 1990] V. Goranko. Modal definability in enriched languages. *Notre Dame Journal of Formal Logic*, 31:81–105, 1990.
- [Grädel, 1999] E. Grädel. Decision procedures for guarded logics. *Mathematische Grundlagen der Informatik*, RWTH Aachen, 1999.
- [Grätzer, 1968] G. Grätzer. *Universal Algebra*. Van Nostrand, Princeton, 1968.
- [Harel et al., 1998] D. Harel, D. Kozen and J. Tiuryn. *Dynamic Logic*. 1998.
- [Harel, 1984] D. Harel. Dynamic logic. In *Vol. II, Handbook of Philosophical Logic*. Reidel, Dordrecht, 1984.
- [Henkin et al., 1971] L. A. Henkin, D. Monk, and A. Tarski. *Cylindric Algebras I*. North-Holland, Amsterdam, 1971.
- [Henkin, 1950] L. Henkin. Completeness in the theory of types. *Journal of Symbolic Logic*, 15:81–91, 1950.

- [Hennessy and Milner, 1985] M. Hennessy and R. Milner. Algebraic laws for indeterminism and concurrency. *Journal of the ACM*, 32:137–162, 1985.
- [Hollenberg, 1998] M. Hollenberg. *Logic and Bisimulation*. PhD Thesis, Institute of Philosophy, Utrecht University, 1998.
- [Huertas, 1994] A. Huertas. *Modal Logics of Predicates and Partial and Heterogeneous Non-Classical Logic*. PhD thesis, Department of Logic, History and Philosophy of Science, Autonomous University of Barcelona, 1994.
- [Humberstone, 1979] I. L. Humberstone. Interval semantics for tense logics. *Journal of Philosophical Logic*, 8:171–196, 1979.
- [Immermann and Kozen, 1987] N. Immermann and D. Kozen. Definability with bounded number of bound variables. In *Proceedings 2nd IEEE Symposium on Logic in Computer Science*, pp. 236–244, 1987.
- [Jaspars, 1994] J. Jaspars. *Calculi for Constructive Communication. The Dynamics of Partial States*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam and Institute for Language and Knowledge Technology, University of Tilburg, 1994.
- [Jennings *et al.*, 1980] R. Jennings, D. Johnstone, and P. Schotch. Universal first-order definability in modal logic. *Zeit. Math. Logik*, 26:327–330, 1980.
- [Kozen, 1979] D. Kozen. On the duality of dynamic algebras and Kripke models. Technical Report RC 7893, IBM, Thomas J. Watson Research Center, New York, 1979.
- [Kracht and Wolter, 1991] M. Kracht and F. Wolter. Properties of independently axiomatizable bimodal logics. *Journal of Symbolic Logic*, 56:1469–1485, 1991.
- [Kracht, 1993] M. Kracht. How completeness and correspondence theory got married. In M. de Rijke, ed. *Diamonds and Defaults*, pages 175–214. Kluwer Academic Publishers, Dordrecht, 1993.
- [Kurtonina, 1995] N. Kurtonina. *Frames and Labels. A Modal Analysis of Categorical Inference*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam and Onderzoeksinstituut voor Taal en Spraak, Universiteit Utrecht, 1995.
- [Lewis, 1973] D. Lewis. Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2:4–18, 1973.
- [Lifshitz, 1985] V. Lifshitz. Computing circumscription. In *Proceedings IJCAI-85*, pages 121–127, 1985.
- [McTaggart, 1908] J. M. E. McTaggart. The unreality of time. *Mind*, 17, 457–474, 1908.
- [Marx and Venema, 1996] M. Marx and Y. Venema. *Multi-Dimensional Modal Logic*. Studies in Pure and Applied Logic. Kluwer Academic Publishers, Dordrecht, 1996.
- [Marx, 1995] M. Marx. *Algebraic Relativization and Arrow Logic*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1995.
- [Meyer Viol, 1995] W. Meyer Viol. *Instantial Logic*. PhD thesis, Utrecht Institute for Linguistics OTS and Institute for Logic, Language and Computation, University of Amsterdam, 1995.
- [Mikulas, 1995] S. Mikulas. *Taming Logics*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1995.
- [Moortgat, 1996] M. Moortgat. Type-logical grammar. In J. van Benthem and A. ter Meulen, editors, *Handbook of Logic and Language*. Elsevier Science Publishers, Amsterdam, 1996.
- [Ohlbach, 1991] H-J Ohlbach. Semantics-based translation methods for modal logics. *Journal of Logic and Computation*, 1:691–746, 1991.
- [Ohlbach, 1993] H. J. Ohlbach. Translation methods for non-classical logics. an overview. *Bulletin of the IGPL*, 1:69–89, 1993.
- [Ponse *et al.*, 1995] A. Ponse, M. de Rijke, and Y. Venema, editors. *Modal Logic and Process Algebra—A Bisimulation Perspective*. Vol. 3 of CSLI Lecture Notes, Cambridge University Press, 1995.
- [Rasiowa and Sikorski, 1970] H. Rasiowa and R. Sikorski. *The Mathematics of Metamathematics*. Polish Scientific Publishers, Warsaw, 1970.
- [Rodenburg, 1982] P. Rodenburg. Intuitionistic correspondence theory. Technical report, Mathematical Institute, University of Amsterdam, 1982.
- [Rodenburg, 1986] P. Rodenburg. *Intuitionistic Correspondence Theory*. PhD thesis, Mathematical Institute, University of Amsterdam, 1986.

- [Rosen, 1995] E. Rosen. Modal logic over finite structures. Technical Report ML-95-08, Institute for Logic, Language and Computation, University of Amsterdam, 1995. To appear in *Journal of Logic, Language and Information*.
- [Sahlqvist, 1975] H. Sahlqvist. Completeness and correspondence in the first- and second- order semantics for modal logic. In S. Kanger, editor, *Proceedings of the 3rd Scandinavian Logic Symposium*, pp. 110–143. North-Holland, Amsterdam, 1975.
- [Seegerberg, 1971] K. Seegerberg. An essay in classical modal logic. *Filosofiska Studier*, 13, 1971.
- [Simmons, 1994] H. Simmons. The monotonous elimination of predicate variables. *Journal of Logic and Computation*, 4:23–68, 1994.
- [Smoryński, 1973] C. S. Smoryński. Applications of Kripke models. In A. S. Troelstra, editor, *Meta-mathematical Investigations of Intuitionistic Arithmetic and Analysis*, pages 329–391. Lecture Notes in Mathematics **344**, Springer, Berlin, 1973.
- [Spaan, 1993] E. Spaan. *Complexity of Modal Logics*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1993.
- [Thijssse, 1992] E. Thijssse. *Partial Logic and Knowledge Representation*. PhD thesis, Institute for Language and Knowledge Technology, University of Tilburg, 1992.
- [Thomason, 1974] S. K. Thomason. An incompleteness theorem in modal logic. *Theoria*, 40:30–34, 1974.
- [Thomason, 1972] S. K. Thomason. Semantic analysis of tense logics. *Journal of Symbolic Logic*, 37:150–158, 1972.
- [Thomason, 1975] S. K. Thomason. Reduction of second-order logic to modal logic I. *Zeit. Math. Logik*, 21:107–114, 1975.
- [Vakarelov, 1991] D. Vakarelov. A modal logic for similarity relations in Pawlak information systems. *Fundamenta Informaticae*, 15:61–79, 1991.
- [Vakarelov, 1996] D. Vakarelov. A modal theory of arrows. To appear in M. de Rijke, ed., 1996.
- [van Benthem and Bergstra, 1995] J. van Benthem and J. Bergstra. Logic of transition systems. *Journal of Logic, Language and Information*, 3:247–283, 1995.
- [van Benthem et al., 1994] J. van Benthem, J. van Eyck, and V. Stebletsova. Modal logic, transition systems and processes. *Logic and Computation*, 4:811–855, 1994.
- [van Benthem, 1976] J. F. A. K. van Benthem. *Modal correspondence theory*. PhD thesis, Instituut voor Grondslagenonderzoek, University of Amsterdam, 1976.
- [van Benthem, 1978] J. F. A. K. van Benthem. Two simple incomplete modal logics. *Theoria*, 44:25–37, 1978.
- [van Benthem, 1979a] J. F. A. K. van Benthem. Canonical modal logics and ultrafilter extensions. *Journal of Symbolic Logic*, 44:1–8, 1979.
- [van Benthem, 1979b] J. F. A. K. van Benthem. Syntactic aspects of modal incompleteness theorems. *Theoria*, 45:67–81, 1979.
- [van Benthem, 1980] J. F. A. K. van Benthem. Some kinds of modal completeness. *Studia Logica*, 39:125–141, 1980.
- [van Benthem, 1981a] J. F. A. K. van Benthem. Intuitionistic definability, 1981. University of Groningen.
- [van Benthem, 1981b] J. F. A. K. van Benthem. Possible worlds semantics for classical logic. Technical Report ZW-8018, Mathematical Institute, University of Groningen, 1981.
- [van Benthem, 1983] J. van Benthem. *The Logic of Time*. Vol. 156 of *Synthese Library*, Reidel, Dordrecht, 1983. Revised edition with Kluwer Academic Publishers, Dordrecht, 1991.
- [van Benthem, 1985] J. van Benthem. *Modal Logic and Classical Logic*. Vol. 3 of *Indices*, Bibliopolis, Napoli, and The Humanities Press, Atlantic Heights, NJ, 1985. Revised edition to appear with Oxford University Press.
- [van Benthem, 1986] J. van Benthem. Tenses in real time. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, 32:61–72, 1986.
- [van Benthem, 1989a] J. van Benthem. Notes on modal definability. *Notre Dame Journal of Formal Logic*, 30:20–35, 1989.



- [van Benthem, 1989b] J. van Benthem. Semantic parallels in natural language and computation. In H.-D. Ebbinghaus *et al.*, editor, *Logic Colloquium. Granada 1987*, pages 331–375. North-Holland, Amsterdam, 1989.
- [van Benthem, 1990] J. van Benthem. Modal logic as a theory of information. In J. Copeland, ed., *Logic and Reality. Essays on the Legacy of Arthur Prior*, pp. 135–186. Oxford University Press, 1995.
- [van Benthem, 1991] J. van Benthem. *Language in Action. Categories, Lambdas and Dynamic Logic*. Vol. 130 of *Studies in Logic*, North-Holland, Amsterdam, 1991.
- [van Benthem, 1992] J. van Benthem. Logic as programming. *Fundamenta Informaticae*, 17:285–317, 1992.
- [van Benthem, 1993a] J. van Benthem. Beyond accessibility: functional models for modal logic. In M. de Rijke, ed. *Diamonds and Defaults*, pp. 1–18. Kluwer, Dordrecht, 1993.
- [van Benthem, 1993b] J. van Benthem. Modal frame classes revisited. *Fundamenta Informaticae*, 18:307–317, 1993.
- [van Benthem, 1995b] J. van Benthem. Temporal logic. In D. Gabbay, C. Hogger, and J. Robinson, editors, *Handbook of Logic in Artificial Intelligence and Logic Programming*, volume 4, pages 241–350. Oxford University Press, 1995.
- [van Benthem, 1996a] J. van Benthem. Contents versus wrappings. In M. Marx, L. Pólos and M. Masuch, eds., *Arrow Logic and Multi-modal Logic*, pp. 203–219. CSLI Publications, Stanford, 1996.
- [van Benthem, 1996b] J. van Benthem. *Exploring Logical Dynamics. Studies in Logic, Language and Information*, CSLI Publications (Stanford) and Cambridge University Press, 1996.
- [van Benthem, 1997] J. van Benthem. Modal foundations for predicate logic. Technical Report LP-95-07, Institute for Logic, Language and Computation, 1995. Appeared in *Bulletin of the IGPL*, 5:259–286, 1997. (R. de Queiroz, ed., Proceedings WoLLIC. Recife 1995) and E. Orłowska, ed., *Logic at Work. Memorial for Elena Rasiowa*, pp. 39–54, Studia Logica Library, Kluwer Academic Publishers, 1999.
- [van Benthem, 1998] J. van Benthem. Modal logic in two gestalts. To appear in M. de Rijke and M. Zakharyashev, eds. *Proceedings AiML-II, Uppsala*, Kluwer, Dordrecht, 1998.
- [van Benthem, 1999] J. van Benthem. Intensional Logic. Electronic lecture notes, Stanford University, 1999. <http://www.turing.wins.uva.nl/~johan/teaching/>
- [van der Hoek, 1992] W. van der Hoek. *Modalities for Reasoning about Knowledge and Quantities*. PhD thesis, Department of Computer Science, Free University, Amsterdam, 1992.
- [Venema, 1991] Y. Venema. *Many-Dimensional Modal Logic*. PhD thesis, Institute for Logic, Language and Information, University of Amsterdam, 1991.
- [White, 1981] M. J. White. Modal-tense logic incompleteness and the Master Argument. University of Arizona, 1981.
- [Winnie, 1977] J. A. Winnie. The causal theory of space-time. In J. S. Earman *et al.*, editor, *Foundations of Space-Time Theories*, pages 134–205. University of Minnesota Press, 1977.
- [Wolter, 1993] F. Wolter. *Lattices of Modal Logics*. PhD thesis, Zweites Mathematisches Institut, Freie Universität, Berlin, 1993.
- [Zanardo, 1994] A. Zanardo. Branching-time logic with quantification over branches. The point of view of modal logic. Reprint 25, Institute for Mathematics, University of Padova, 1994.
- [Zakharyashev, 1992] M. Zakharyashev. Canonical formulae for K4. Part I: Basic results. *Journal of Symbolic Logic*, 57:1377–1402, 1992.
- [Zakharyashev, 1995] M. Zakharyashev. Canonical formulae for modal and superintuitionistic logics: a short outline. In M. de Rijke, ed. 1996.

## INDEX

- $\Box\Diamond$ -formula, 125
- $\Pi_1^1$ -sentences, 328, 350
- $\oplus$ -prime logic, 89
- $\oplus$ -irreducible logic, 89
- $\varkappa$ -complex logic, 115
- $\varkappa$ -generated frame, 93, 160
- $n$ -transitive logic, 87, 160
  
- actual world, 140
- actual world condition, 141
- aggregation, 336
- alternative relation, 339
- amalgamability, 152
- anti-preservation, 345
- atom, 94, 99
- atomic, 326
- axiomatic basis, 89
- axiomatization
  - finite, 89
  - independent, 89
  - problem, 97
  - recursive, 89
  
- Barcan Formula, 368
- Beth Property, 149
- bimodal companion, 224
- Birkhoff's Theorem, 359
- bisimulation, 249, 400
  
- canonical, 334
- canonical formula, 119
  - intuitionistic, 199
  - quasi-normal, 141
- canonicity, 100
- categorical connection, 362
- CDC, 118
- closed domain, 118
- closed domain condition, 118
  
- cluster assignment, 175
- cofinal subframe formula, 125
- cofinal subframe logic, 125
  - quasi-normal, 142
- compact frame, 92
- compactness, 112, 327
- comparative, 375
- complete set of formulas, 89
- completeness, 325
- completeness theorems, 332
- complex variety, 114
- complexity, 367
- complexity function, 243
- conditional logic, 373
- configuration problem, 228
- confluent, 325
- congruential logic, 249
- conservative formula, 155
- correspondence, 325
- counterfactual, 381
- cover, 94
- cycle free frame, 98, 160
  
- d-cyclic set, 94
- deduction theorem, 87
- deductively equivalent formulas, 86
- degree of incompleteness, 108
- depth of a frame, 94
- descriptive, 362
- descriptive frame, 92, 160
- difference operator, 249
- differentiated frame, 92, 160
- direct products, 357
- directed, 325
- disjoint union, 339
- disjunction property, 211
  - modal, 211
- distinguished point, 140

- downward directness, 105
- duality, 331
- Dummett logic, 200
- dynamic logic, 373
- elementary equivalence, 340
- elementary logic, 107
- essentially negative formula, 209
- filter representation, 392
- finite embedding property, 128
- finite model property
  - exponential, 243
  - global, 113
  - polynomial, 243
- first-order definable, 347
- first-order equivalent, 328
- first-order undefinability, 349
- fixed point operator, 250
- focus, 132
- frame formula, 120
- fusion, 162
- Gödel translation, 196
- general frame, 338, 361
- generated subframe, 339
- generation theorem, 341
- global definability, 349
- global derivability, 87
- global Kripke completeness, 113
- graded modality, 249
- Halldén completeness, 156
- Henkin model, 359
- Heyting algebra, 193, 392
- higher-order correspondence, 370
- homomorphic images, 357
- inaccessible world, 159
- incompleteness, 375
- independent set of formulas, 89
- inference rule
  - admissible, 232
  - derivable, 231
- intermediate axioms, 390
- interpolant, 149
  - post-, 156
- interpolation property, 149
  - for a consequence relation, 149
- intersection of logics, 88
- interval, 377
- intuitionistic frame, 193
- intuitionistic logic, 373
- intuitionistic modal frame, 219
- intuitionistic modal logic, 218
- invariance, 371
- isomorphism, 340
- Jankov formula, 120
- Kreisel–Putnam logic, 212
- Kripke frame, 92, 325
- Löb’s Axiom, 116
- Löb’s Axiom, 333
- Löwenheim–Skolem, 327
- Lindenbaum Algebra, 358
- Lindström Theorem, 352
- linear tense logic, 176
- local definability, 349
- local tabularity, 123
- logic of a class of frames, 92
- Łoś Equivalence, 350
- McKinsey Axiom, 326
- Medvedev’s logic, 216
- minimal modal logic, 333
- minimal tense extension, 172
- Minsky machine, 228
- modal algebra, 331, 357
- modal companion, 200
- modal degree, 101
- modal incompleteness, 333
- modal matrix, 140
- modal predicate logic, 367
- modal projection, 339
- modal reduction principle, 354
- models, 327
- negation, 381

- negative formula, 102
- Nishimura formula, 195
- Noetherian frame, 116
- nominal, 249
- non-eliminability, 101
- non-iterative logic, 161
- normal filter, 152
- normal form, 123
  
- open domain, 118, 198
  
- p-morphisms, 331
- p-morphism, 92
- partial order, 325
- persistence, 100
- polymodal frame, 160
- polymodal logic, 159
- polynomial, 358
- polynomially equivalent logics, 243
- positive formula, 102
- possible worlds, 343
- preservation, 342
- pretabularity, 147
- prime filter, 194
- prime formula, 89
- propositional quantifiers, 371
- propositions, 343
- pseudo-Boolean algebra, 193
  
- quasi-normal logic, 139
  
- reduced frame, 101
- reduction, 92, 160
  - weak, 185
- refined frame, 92
- refined refined, 160
- reflexive, 325
- relevance logic, 380
- replacement function, 173
- restricted quantifiers, 364
- Rieger–Nishimura lattice, 195
- root, 92, 160
- rooted frame, 92
  
- Sahlqvist formula, 107, 161
- Sahlqvist Theorem, 354
- saturation, 362
- Scott logic, 212
- second-order equivalent, 327
- semantical consequence, 242
- si-fragment, 200
- si-logic, 193
- similarity, 382
- simulation of a frame, 169
- simulation of a logic, 169
- skeleton, 196
- skeleton lemma, 196
- Smetanich logic, 200
- splitting, 96
  - union-, 97
- splitting pair, 90
- standard translation, 135
- Stone representation, 357
- strict Kripke completeness, 97
- strict sf-completeness, 129
- strong global completeness, 113
- strong Kripke completeness, 112
- strongly positive formula, 103
- structural completeness, 233
- subalgebras, 357
- subdirectly irreducible, 363
- subframe, 116, 145, 160, 198
  - cofinal, 126, 145
  - generated, 92, 160
- subframe formula, 125
- subframe logic, 125, 127
  - quasi-normal, 142
- subreduction, 116
  - cofinal, 116
  - quasi-, 141
  - weak, 185
- substitutions, 354
- sum of logics, 88
- superamalgamability, 152
- superintuitionistic logic, 193
- surrogate, 163
- surrogate frame, 185
- symmetry, 336

t-line logic, 181  
tabularity, 145  
Tarski's criterion, 89  
temporal modalities, 377  
temporal order, 375  
tense frame, 172  
tense logic, 171, 373  
tight frame, 92  
time-line, 181  
topological Boolean algebra, 195  
transfer, 399  
transitive, 325  
translation, 327

ultrafilter extension, 339  
ultrafilters, 344  
ultrapower, 346  
ultraproduct, 328  
undecidable formula, 231  
uniform formula, 124  
uniform interpolation, 156  
universal frame of rank  $n$ , 93  
universal modality, 166  
untied formula, 105  
upward closed set, 92

valuation, 339

weak Kreisel–Putnam formula, 198

zigzag connection, 341, 370  
zigzag morphism, 342  
Zigzag Theorem, 342

# Handbook of Philosophical Logic

2nd Edition

Volume 4

edited by Dov M. Gabbay and F. Guentner



## CONTENTS

Editorial Preface	vii
<b>Dov M. Gabbay</b>	
Conditional Logic	1
<b>D. Nute and C. B. Cross</b>	
Dynamic Logic	99
<b>D. Harel, D. Kozen, and J. Tiuryn</b>	
Logics for Defeasible Argumentation	219
<b>H. Prakken and G. Vreeswijk</b>	
Preference Logic	319
<b>S. O. Hansson</b>	
Diagrammatic Logic	395
<b>E. Hammer</b>	
Index	423





## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good.!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London

<b>Logic</b>	<b>IT</b>			
	<b>Natural language processing</b>	<b>Program control specification, verification, concurrency</b>	<b>Artificial intelligence</b>	<b>Logic programming</b>
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs

<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical frag- ments</b>	Basic back- ground lan- guage	Program syn- thesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely use- ful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calcu- lus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syn- tax	Modules. Combining languages	Logics of space and time	Combining fea- tures
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game seman- tics gaining ground		
<b>Object level/ metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action- revision mod- els</b>			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model





## CONDITIONAL LOGIC

Prior to 1968 several writers had explored the conditions for the truth or assertability of conditionals, but this work did not result in an attempt to provide formal models for the semantical structure of conditionals. It had also been suggested that a proper logic for conditionals might be provided by combining modal operators with material conditionals in some way, but this suggestion never led to any widely accepted formal logic for conditionals.<sup>1</sup> Then Stalnaker [1968] provided both a formal semantics for conditionals and an axiomatic system of conditional logic. This important paper effectively inaugurated that branch of philosophical logic which we today call conditional logic. Nearly all the work on the logic of conditionals for the next ten years, and a great deal of work since then, has either followed Stalnaker's lead in investigating possible worlds semantics for conditionals or posed problems for such an approach. But in 1978, Peter Gärdenfors [1978] initiated a new line of inquiry focused on the use of conditionals to represent policies for belief revision. Thus, two main lines of development appeared, one an ontological approach concerned with truth or assertability conditions for conditionals and the other an epistemological approach focused on conditionals and change of belief.

With these two major lines of development, the material which has appeared on conditionals is prodigious. Consequently, we have had to focus upon certain aspects of conditional logic and to give other aspects less attention. We have followed the trend set in the literature and given the most attention to the analysis of so-called subjunctive conditionals as they are used in ordinary discourse and to triviality results for the Ramsey test. Accordingly, our discussion of conditionals and belief revision will be more heavily technical than our discussion of subjunctive conditionals. Other topics are discussed in less detail. Some of the important papers which it has not been possible to review are included in the accompanying bibliography, but the bibliography itself is far from complete.

## 1 ONTOLOGICAL CONDITIONALS

*1.1 Introduction*

Conditional logic is, in the first place, concerned with the investigation of the logical and semantical properties of a certain class of sentences occurring

<sup>1</sup>Another suggestion which has never been fully developed (but see Hunter [1980; 1982] is that an adequate theory of ordinary conditionals may be derived from relevance logic. We will say no more about this suggestion than it seems to us that conditional logic and relevance logic are concerned with very different problems, and it would be a tremendous coincidence if the correct logic for the conditionals of ordinary usage should turn out to resemble some version of relevance logic at all closely.

in a natural language. We will draw our examples from English, but much of what we have to say can be applied, with due caution, to other natural languages.

Paradigmatically, a conditional declarative sentence in English is one which contains the words ‘if’ and ‘then’. Examples include

1. If it is raining, then we are taking a taxi.

and

2. If I were warm, then I would remove my jacket.

We could delete the occurrences of ‘then’ in (1) and (2) and we would still have perfectly acceptable sentences of English. In the case of (2), we can omit both ‘if’ and ‘then’ if we change the word order. Example (2) surely says the same thing as

3. Were I warm, I would remove my jacket.

Other conditionals in which neither ‘if’ nor ‘then’ occur include

4. When I find a good man, I will praise him.

and

5. You will need my number should you ever wish to call me.

Notice that all of these examples involve two component sentences or clauses, one expressing some sort of condition and another expressing some sort of claim which in some way depends upon the condition. The conditional or ‘if’ part of a conditional sentence is called the antecedent, and the main or ‘then’ part its consequent even when ‘if’ and ‘then’ do not actually occur. Notice that the antecedent precedes the consequent in (1)–(4), but the consequent comes first in (5). These examples should give the reader a fair idea of the types of sentences with which conditional logic is concerned.

While the verbs in (1) are in the indicative mood, those in (2) are in the subjunctive mood. Researchers often rephrase (2), forming a new conditional in which the verbs contained in antecedent and consequent are in the indicative mood. This practice implicitly assumes that (2) has the same content as

6. If it were the case that I am warm, then it would be the case that I remove my jacket.

Even without the rephrasing, it is sometimes said that ‘I am warm’ is the antecedent of both (2) and (6). Thus the mood of the verbs in the grammatical antecedent and consequent of (2) are taken logically to be a component of the conditional construction, while the logical antecedent and consequent

are viewed as containing verbs in the indicative mood. Seen in this way, the conditional constructions in (1) and (2) look quite different and investigators have as a consequence made a distinction between indicative conditionals like (1) and subjunctive conditional like (2). This distinction is important because it appears that these two kinds of conditionals have different logical and semantical properties.

Much of the work done in conditional logic has focused on conditionals having antecedents and consequents which are false. Such conditionals are called counterfactuals. In actual practice, little distinction is made between counterfactuals and subjunctive conditionals which have true antecedents or consequents. Authors frequently refer to conditionals in the subjunctive mood as counterfactuals regardless of whether their antecedents or consequents are true or false. Another special kind of conditional is the so-called counterlegal conditional whose antecedent is incompatible with physical law. An example is

7. If the gravitational constant were to take on a slightly higher value in the immediate vicinity of the earth, then people would suffer bone fractures more frequently.

Also recognized are counteridenticals like

8. If I were the pope, I would support the use of the pill in India.

and countertemporals like

9. If it were 3.00 a.m., it would be dark outside.

Analysis of these special conditionals may involve special difficulties, but we can say very little about these special problems in a paper of this length.

Two other interesting conditional constructions are the even-if construction used in

10. It would rain even if the shaman did not do his dance.

and the might construction used in

11. If you don't take the umbrella, you might get wet.

We might paraphrase (10) using the word 'still' to get

12. It would still rain if the shaman did not do his dance.

even-if and might conditionals have somewhat different properties from those of other conditionals. It is believed by many, though, that these two kinds of conditionals can be analyzed in terms of subjunctive conditionals once we have an acceptable analysis of these. The strategy in this

paper will be to concentrate on the many proposals for subjunctive conditionals, returning later (briefly) to the topics of indicative, even-if and might conditionals.

We will use two different symbols to represent indicative and subjunctive conditionals. For indicative conditionals we will use the double arrow  $\Rightarrow$ , and for the subjunctive conditional we will use the corner  $>$ . (Where context makes our intention clear, we will sometimes use symbols and formulas autonomously to refer to themselves.) With these devices we may represent (1) as

13. It is raining  $\Rightarrow$  I am taking a taxi.

and represent (2) as

14. I am warm  $>$  I remove my jacket.

Frequently we will have no particular antecedent or consequent in mind as we discuss one or the other of these two kinds of conditionals and as we examine forms which arguments involving these conditionals may take. In these cases we will use standard notation for classical first-order logic augmented by our symbols for indicative and subjunctive conditionals to represent the forms of sentences and arguments under discussion. We assume, as have nearly all investigators, that conditionals have truth values and may therefore appear as arguments for truth-functional operators.

Students in introductory symbolic logic courses are normally taught to treat English conditionals as material conditionals. By material conditionals we mean certain truth-functional compounds of simpler sentences. A material condition  $\phi \rightarrow \psi$  is true just in case  $\phi$  is false or  $\psi$  is true. There can be little doubt that neither material implication nor any other truth function can be used by itself to provide an adequate representation of the logical and semantical properties of English conditionals or, presumably, the conditionals of any other language.

Consider the following two examples.

15. If I were seven feet tall, then I would be over two meters tall.

16. If I were seven feet tall, then I would be less than two yards tall.

In fact one of the authors is more than two yards tall but less than two meters tall, so for him the common antecedent and the two consequents of (15) and (16) are all false. Yet surely (15) is true while (16) is false. When both the antecedent and the consequent of an English subjunctive conditional are false, the conditional may be either true or false. Now consider two more examples.

17. If I were eight feet tall, I would be less than seven feet tall.

18. If I were seven feet tall, I would be over six feet tall.

Here we have two conditionals each of which has a false antecedent and a true consequent. but the first of these conditionals is false and the second is true. The moral of these examples is that when the antecedent of an English subjunctive conditional is false, the truth value of the conditional is not determined by the truth values of the antecedent and the consequent of the conditional alone. Some other factors must be involved in determining the truth values of such conditionals.

But what about English conditionals with true antecedents? It is generally accepted that any conditional with a true antecedent and a false consequent is false, but the situation is more controversial where the conditionals with true antecedents and true consequents are concerned. Some researchers have maintained that all such conditionals are true while others have claimed that such conditionals are sometimes false. Later we will consider some of the issues involved in this controversy. For now we simply recognize that there are some very good reasons for rejecting the view that all English conditionals can be represented adequately by material implication or by any other truth function.

### 1.2 *Cotenability theories of conditionals*

Chisholm [1946], Goodman [1955], Sellars [1958], Rescher [1964] and others have proposed accounts of conditionals which share some important features. Borrowing a term from Goodman, we can call these proposals *cotenability* theories of conditionals. The basic idea which these proposals share is that the conditional  $\phi > \psi$  is true in case  $\phi$ , together with some set of laws and true statements, entails  $\psi$ .

A crucial problem for such an analysis is that of determining the appropriate set of true statements to involve in the truth condition for a particular conditional. If the antecedent of the conditional is false, then of course its negation is true. But any proposition together with its negation will entail anything. The set of true statements upon which the truth of the conditional is to depend must at least be logically compatible with the antecedent of the conditional or the conditional will turn out to be trivially true on such an account. But logical compatibility is not enough either. We can have a true proposition  $\chi$  such that  $\phi$  and  $\chi$  are logically compatible but such that  $\chi > \neg\phi$  is also true. Then we should not wish to include  $\chi$  in the set of propositions upon which the evaluation of  $\phi > \psi$  depends. Goodman said of such a  $\chi$  that it is not cotenable with  $\phi$ . So Goodman's ultimate position is that  $\phi > \psi$  is true just in case  $\psi$  is entailed by  $\phi$  together with the set of all physical laws and the set of all true propositions cotenable with  $\phi$ , i.e. with the set of all true propositions such that no member of that set counterfactually implies the negation of  $\phi$  and the negation of no member

of that set is counterfactually implied by  $\phi$ . Such an account is obviously circular since the truth conditions for counterfactuals are given in terms of cotenability, while cotenability is defined in terms of the truth values of various counterfactual conditionals.

Although this is certainly a serious problem, it is not the only problem which theories of this type encounter. As a result of the role which law plays in such a theory, all counterlegal conditionals are counted as trivially true, and this is counterintuitive. Furthermore, even if we could provide a noncircular account of cotenability, another problem arises for conditionals which are *not* counterlegal. Suppose two true propositions  $\chi$  and  $\theta$  are each cotenable with  $\phi$ , but that  $\chi \wedge \theta$  is not. In selecting the set of propositions upon which the evaluation of  $\phi > \psi$  shall rest we must omit either  $\chi$  or  $\theta$  since otherwise our conditional will be trivially true once again. But which of these two propositions shall we omit?

Most recent work in conditional logic is compatible with cotenability theory even though no attempt is made to define and use the notion of cotenability. We might view the resultant theories at least in part as attempts to determine, without ever specifying exactly what cotenability is, the logical and semantical properties which conditionals must have if the cotenability approach is essentially correct for conditionals without counterlegal antecedents. Indeed, the vagueness deliberately built into many of these recent theories suggests that our notion of cotenability, if we have one, varies according to our purposes and the context in which we use a conditional.<sup>2</sup>

### 1.3 *Strict Conditionals*

We have seen that the truth value of a conditional is not always determined by the actual truth values of its antecedent and consequent, but perhaps it is determined by the truth values which its antecedent and consequent take in some other possible worlds. One way such an analysis might be developed is suggested by the role laws play in the cotenability theories. Perhaps we should look not only at the truth values of the antecedent and the consequent in the actual world, but also at their truth values in all possible worlds which have the same laws as does our own. When two worlds obey the same physical laws, we can say that each is a physical alternative of the other. The proposal, then, is that  $\phi > \psi$  is true if  $\psi$  is true at every physical alternative to the actual world at which  $\phi$  is true. Suppose we say a proposition is physically necessary if and only if it is true at every physical alternative to the actual worlds, and suppose we express

---

<sup>2</sup>Bennett [1974] and Loewer [1978] arrive at opposite conclusions concerning the question whether Lewis's semantics is compatible with cotenability theory. Their discussions are instructive for other semantics as well.

the claim that a proposition  $\phi$  is physically necessary by  $\Box\phi$ . Then the proposal we are considering is that the following equivalence always holds:

$$19. (\phi > \psi) \leftrightarrow \Box(\phi \rightarrow \psi).$$

Another way of arriving at (19) is the following. English subjunctive conditionals are not truth-functional because they say more than that the antecedent is false or the consequent is true. The additional content is a claim that there is some sort of connection between the antecedent and the consequent. The kind of connection which seems to occur to people most readily in this context is a physical or causal connection. How can we represent this additional content in our formalization of English subjunctive conditionals? One way is to interpret  $\phi > \psi$  as involving the claim that it is physically impossible that  $\phi$  be true and  $\psi$  false. Once again we come up with (19). A proposal resembling the one we have outlined can be found in [Burks, 1951], although we do not wish to suggest that Burks arrived at his account by exactly the same line of reasoning as we have suggested.

We can generalize the proposal represented by (19). We might suppose that the basic form of (19) is correct but that the sort of necessity involved in English subjunctive conditionals is not pure physical necessity. One reason for suspecting this is that the notion of cotenability has been ignored. It is not simply a consequence of physical law that Jane would develop hives if she were to eat strawberries; it is also in part a consequence of her having a particular physical make-up. In evaluating the claim that Jane would become ill if she were to eat strawberries, we do not count the fact that in some worlds which share the same physical laws as our own but in which Jane has a radically different physical make-up, she is able to eat strawberries with impunity, as a legitimate reason for rejecting this claim. Another reason for seeking a different kind of necessity for the analysis of conditionals is that some conditionals may be true because of connections between their antecedents and consequents which are not physical connections at all. Consider, for example, conditionals such as ‘If you deserted your family you would be a cad’, which seems to be founded on normative rather than physical connections. The general theory we are considering, then, is that English subjunctive conditionals are strict conditionals of some sort, i.e. that their logical form is given by the equivalence (19). There remains the problem of determining which kind of necessity is involved in these conditionals.

Regardless of the kind of necessity we choose in such an analysis of conditionals, we should expect our modal logic to have certain minimal properties. By a modal logic we mean any set  $\bar{L}$  of sentences formed from the symbols of classical sentential logic together with the symbol  $\Box$  in the usual ways, provided that  $\bar{L}$  contains all tautologies and is closed under the rule *modus ponens*. We should expect that for any tautology  $\phi$  our modal logic will contain  $\Box\phi$ . We should also expect our modal logic to contain all substitution



instances of the following thesis:

$$20. \Box(\phi \rightarrow \psi) \rightarrow (\Box\phi \rightarrow \Box\psi).$$

But when we define our conditionals according to (19), our logic will then also contain all substitution instances of the following theses:

$$\text{Transitivity: } [(\phi > \psi) \wedge (\chi > \phi)] \rightarrow (\chi > \psi)$$

$$\text{Contraposition: } (\phi > \neg\psi) \rightarrow (\psi > \neg\phi)$$

$$\text{Strengthening antecedents: } (\phi > \psi) \rightarrow [(\phi \wedge \chi) > \psi].$$

But none of these theses seem to be reliable for English subjunctive conditionals. As a counterexample to Transitivity, consider the following conditionals:

21. If Carter had not lost the election in 1980, Reagan would not have been President in 1981.
22. If Carter had died in 1979, he would not have lost the election in 1980.
23. If Carter had died in 1979, Reagan would not have been President in 1981.

(21) and (22) are true, but it is far from clear that (23) is true. As a counterexample to Contraposition, consider:

24. If it were to rain heavily at noon, the farmer would not irrigate his field at noon.
25. If the farmer were to irrigate his field at noon, it would not rain heavily at noon.

And finally, for Strengthening Antecedents, consider:

26. If the left engine were to fail, the pilot would make an emergency landing.
27. If the left engine were to fail and the right wing were to shear off, the pilot would make an emergency landing.

Since even very weak modal logics will contain all substitution instances of these three theses, and since most speakers of English find counterexamples of the sort we have considered convincing, most investigators are convinced that English conditionals are not a variety of strict conditional.

### 1.4 *Minimal Change Theories*

While treating ordinary conditionals as strict conditionals does not seem too promising, investigators have still found the possible worlds semantics often associated with modal logic very attractive. The basic intuition, that a conditional is true just in case its consequent is true at every member of some set of worlds at which its antecedent is true, may yet be salvageable. We can avoid Transitivity, etc. if we allow that the set of worlds involved in the truth conditions for different conditionals may be different. But we do not wish to allow that this set of worlds be chosen arbitrarily for a given conditional.

Stalnaker [1968] proposes that the conditional  $\phi > \psi$  is true just in case  $\psi$  is true at the world most like the actual world at which  $\phi$  is true. According to Stalnaker, in evaluating a conditional we add the antecedent of the conditional to our set of beliefs and modify our set of beliefs as little as possible in order to accommodate the new belief tentatively adopted. Then we consider whether the consequent of the conditional would be true if this revised set of beliefs were all true. In the ideal case, we would have a belief about every single matter of fact before and after this operation of adding the antecedent of the conditional to our stock of beliefs. Possible worlds correspond to these epistemically ideal situations. Stalnaker's assumption, then, is that at least when the antecedent of a conditional is logically possible, there is always a unique possible world at which the antecedent is true and which is more like the actual world than is any other world at which the antecedent is true. We will call this Stalnaker's Uniqueness Assumption.

On some fairly reasonable assumptions about the notion of similarity of worlds, Stalnaker's truth conditions generate a very interesting logic for conditionals. Essentially these assumptions are that any world is more similar to itself than is any other world, that the  $\phi$ -world closest to world  $i$  (that is, the world at which  $\phi$  is true which is more similar to  $i$  than is any other world at which  $\phi$  is true) is always at least as close as the  $\phi \wedge \psi$ -world closest to  $i$ , and that if the  $\phi$ -world closest to  $i$  is a  $\psi$ -world and the  $\psi$ -world closest to  $i$  is a  $\phi$ -world, then the  $\phi$ -world closest to  $i$  and the  $\psi$ -world closest to  $i$  are the same world.

The model theory Stalnaker develops is complicated by his use of the notion of an absurd world, a world at which every sentence is true. This invention is motivated by the need to provide truth conditions for conditionals with impossible antecedents. Stalnaker's semantics can be simplified by omitting this device and adjusting the rest of the model theory accordingly. When we do this, we produce what could be called simplified Stalnaker models. Such a model is an ordered quadruple  $\langle I, R, s, [ ] \rangle$  where  $I$  is a set of possible worlds,  $R$  is a binary reflexive (accessibility) relation on  $I$ ,  $s$  is a partial world selection function which, when defined, assigns to sentence  $\phi$  and a world  $i$  in  $I$  a world  $s(\phi, i)$  (the  $\phi$ -world closest to  $i$ ), and  $[ ]$  is a

function which assigns to each sentence  $\phi$  a subset  $[\phi]$  of  $I$  (all those worlds in  $I$  at which  $\phi$  is true). Stalnaker's assumptions about the similarity of worlds become a set of restrictions on the items of these models:

- (S1)  $s(\phi, i) \in [\phi]$ ;
- (S2)  $\langle i, s(\phi, i) \rangle \in R$ ;
- (S3) if  $s(\phi, i)$  is not defined then for all  $j \in I$  such that  $\langle i, j \rangle \in R$ ,  $j \notin [\phi]$ ;
- (S4) if  $i \in [\phi]$  then  $s(\phi, i) = i$ ;
- (S5) if  $s(\phi, i) \in [\psi]$  and  $s(\psi, i) \in [\phi]$ , then  $s(\phi, i) = s(\psi, i)$ ;
- (S6)  $i \in [\phi > \psi]$  if and only if  $s(\phi, i) \in [\psi]$  or  $s(\phi, i)$  is undefined.

Until otherwise indicated, we will understand by a conditional logic any set  $\bar{L}$  of sentences which can be constructed from the symbols of classical sentential logic together with the symbol  $>$ , provided that  $\bar{L}$  contains all tautologies and is closed under the inference rule *modus ponens*. The conditional logic determined by Stalnaker's model theory is the smallest conditional logic which is closed under the two inference rules

RCEC: from  $\phi \leftrightarrow \psi$ , to infer  $(\chi > \phi) \leftrightarrow (\chi > \psi)$

RCK: from  $(\phi_1 \wedge \dots \wedge \phi_n) \rightarrow \psi$ , to infer  $[(\chi > \phi_1) \wedge \dots \wedge (\chi > \phi_n)] \rightarrow (\chi > \psi)$ ,  $n \geq 0$

and which contains all substitution instances of the theses

ID:  $\phi > \phi$

MP:  $(\phi > \psi) \rightarrow (\phi \rightarrow \psi)$

MOD:  $(\neg\phi > \phi) \rightarrow (\psi > \phi)$

CSO:  $[(\phi > \psi) \wedge (\psi > \phi)] \rightarrow [(\phi > \chi) \leftrightarrow (\psi > \chi)]$

CV:  $[(\phi > \psi) \wedge \neg(\phi > \neg\chi)] \rightarrow [(\phi \wedge \chi) > \psi]$

CEM:  $(\phi > \psi) \vee (\phi > \neg\psi)$

Together with *modus ponens* and the set of tautologies, these rules and theses can be viewed as an axiomatization of Stalnaker's logic, which he calls **C2**. While Stalnaker supplies a rather different axiomatization for **C2**, these rules and theses enjoy the advantage that they allow easy comparison of **C2** with other conditional logics. Several of these rules and theses are due to Chellas [1975]. It can be shown that a sentence is a member of **C2** if and only if that sentence is true at every world in every simplified

Stalnaker model. Thus we say that the class of simplified Stalnaker models determines or characterizes the conditional logic **C2**. None of Transitivity, Contraposition, and Strengthening Antecedents is contained in **C2**.

A variation of the semantics developed by Stalnaker treats the function  $s$  as taking sets of worlds rather than sentences as arguments and values. In this variation,  $s$  is a function which assigns to each subset  $A$  of  $I$  and each member  $i$  of  $I$  a subset  $s(A, i)$  of  $I$ . Then  $\phi > \psi$  will be true at  $i$  just in case  $s([\phi], i) \subseteq [\psi]$ . By setting our semantics up in this way, we ensure that we can substitute one antecedent for another in a conditional provided that the two antecedents are true at exactly the same worlds, and we can do this without any additional restrictions on the function  $s$ . Since many authors have called sets of worlds propositions, we could call Stalnaker's original semantics a sentential semantics and the present variation on Stalnaker's semantics a propositional semantics to represent this difference in the kind of argument the function  $s$  takes. As we look at alternatives to Stalnaker's semantics we will always consider the sentential forms of these semantics although equivalent propositional forms will often be available. Equivalence of the two versions of a particular semantics is guaranteed so long as the conditional logic characterized by the sentential version is closed under substitution of provable equivalents, i.e. so long as it is closed under both RCEC and

RCEA: from  $\phi \leftrightarrow \psi$  to infer  $(\phi > \chi) \leftrightarrow (\psi > \chi)$ .

**C2** is closed under RCEA as is any conditional logic closed under RCK and containing all substitution instances of CSO. The difference between sentential and propositional formulations of a particular kind of model theory becomes important if we wish to consider conditional logics which are not closed under RCEA. Reasons for considering such 'non-classical' logics are discussed in Section 1.7 below. For parallel development of sentential and propositional versions of certain kinds of model theories for conditional logics, see [Nute, 1980b].

Lewis [1973b; 1973c] questions Stalnaker's assumptions about the similarity of worlds and thus his semantics for conditionals. It is Stalnaker's Uniqueness Assumption which Lewis rejects. Lewis argues that there may be no unique  $\phi$ -world which is closer to  $i$  than is any other  $\phi$ -world. As an example, Lewis asks us to consider a straight line printed in a book and to suppose that this line were longer than it is. No matter what greater length we choose for the line, there is a shorter length which is still greater than the actual length of the line. The conclusion is that worlds which differ from the actual world only in the length of the sample line may be more and more like the actual world as the length of the line in those worlds comes closer to the line's actual length. But none of these worlds is the closest world at which the line is longer. In fact, examples of this sort can also be offered against an assumption about similarity of worlds which is

weaker than Stalnaker's Uniqueness Assumption. This assumption, which Lewis calls the Limit Assumption, is that, at least for a sentence  $\phi$  which is logically possible, there is always at least one  $\phi$ -world which is as much like  $i$  as is any other  $\phi$ -world. Both the Uniqueness Assumption and the weaker Limit Assumption are highly suspect.

If we follow Lewis's advice and drop the Uniqueness Assumption, we must give up Conditional Excluded Middle (CEM). But this is exactly the feature of Stalnaker's logic which is most often cited as objectionable. Both disjuncts in CEM will be true if  $\phi$  is impossible and hence  $s$  is not defined for  $\phi$  and the actual world. On the other hand, if  $\phi$  is possible, then  $\psi$  must be either true or false at the nearest  $\phi$ -world. Lewis ([1973b], p. 80) offers the following as a counterexample to CEM:

28a It is not the case that if Bizet and Verdi were compatriots, Bizet would be Italian; and it is not the case that if Bizet and Verdi were compatriots, Bizet would not be Italian; nevertheless, if Bizet and Verdi were compatriots, Bizet either would or would not be Italian.

Lewis [1973b] admits that (28a) sounds, offhand, like a contradiction, but he insists that the cost of respecting this offhand opinion is too high:

However little there is to choose for closeness between worlds where Bizet and Verdi are compatriots by both being Italian and worlds where they are compatriots by both being French, the selection function still must choose. I do not think it *can* choose—not if it is based entirely on comparative similarity, anyhow. Comparative similarity permits ties, and Stalnaker's selection function does not.<sup>3</sup>

Van Fraassen [1974] has employed the notion of supervaluation in defense of CEM. The suggestion is that in actual practice we do not depend upon a single world selection function  $s$  in evaluating conditionals. Instead we consider a number of different ways in which we might measure the similarity of worlds, each with its appropriate world selection function. Each world selection function provides a way of evaluating conditionals. A sentence can also have the property that it is true regardless of which world selection function we use. We can call such a sentence supertrue. If we accept Stalnaker's semantics together with a multiplicity of world selection functions, it turns out that every instance of CEM is supertrue even though it may be the case that neither disjunct of some instance of CEM is supertrue. In fact, all the members of **C2** are supertrue when we apply Van Fraassen's method of supervaluation, and the method mandates the following reinterpretation of the Bizet-Verdi example:

---

<sup>3</sup>[Lewis, 1973b], p. 80.

28b ‘If Bizet and Verdi were compatriots, Bizet would be Italian’ is not supertrue; and ‘If Bizet and Verdi were compatriots, Bizet would not be Italian’ is not supertrue; nevertheless, ‘If Bizet and Verdi were compatriots, Bizet either would or would not be Italian’ is supertrue. (The relevant instance of CEM is also supertrue: ‘Either Bizet would be Italian if Bizet and Verdi were compatriots, or Bizet would not be Italian if Bizet and Verdi were compatriots.’)

In the Bizet-Verdi example, what Lewis accounts for as a tie in comparative world similarity, the method of supervaluation accounts for as a case of indeterminacy in the choice of a closest compatriot-world.

Lewis [1973b] admits that offhand opinion seems to favor CEM, but, Stalnaker [1981a] shows that there is systematic intuitive evidence for CEM: the apparent absence of scope ambiguities in conditionals where Lewis’ theory predicts we should find them. Consider the following dialogue (see [Stalnaker, 1981a], pp. 93–95):

- X: President Carter has to appoint a woman to the Supreme Court.  
 Y: Who do you think he has to appoint?  
 X: He doesn’t have to appoint any particular woman; he just has to appoint some woman or other.

There is a clear scope ambiguity in X’s statement, and this scope ambiguity explains why X’s response to Y makes sense: Y reads X as having intended ‘a woman’ to have wide scope, and X’s response corrects Y by making it clear that X intended ‘a woman’ to have narrow scope. Now compare this dialogue to another, in which necessity is replaced by the past-tense operator:

- X: President Carter appointed a woman to the Supreme Court.  
 Y: Who do you think he appointed?  
 X: He didn’t appoint any particular woman; he just appointed some woman or other.

In this case X’s response does not make sense. There is no semantically distinct narrow scope reading that X could have had in mind, so there is no room for Y to have misunderstand X’s statement. Finally, consider a dialogue involving a conditional instead of a necessity or past tense statement:

- X: President Carter would have appointed a woman to the Supreme Court last year if there had been a vacancy.  
 Y: Who do you think he would have appointed?

X: He wouldn't have appointed any particular woman; he just would have appointed some woman or other.

If Lewis' analysis of counterfactuals is correct, then in this dialogue, as in the first dialogue, one should perceive an ambiguity in the scope of 'a woman' in X's statement, and X's response should make sense as a correction of Y's misinterpretation. In fact there is no room for Y to have misunderstood X's statement, and X's response simply doesn't make sense. In this respect, the third dialogue parallels the second dialogue, not the first, and the apparent lack of a scope ambiguity in X's statement in the third dialogue is evidence for CEM.<sup>4</sup>

If Stalnaker's example does not convince one to accept CEM, it is quite possible to formulate a logic and a semantics for conditionals which resembles Stalnaker's but which does not include CEM. Lewis [1971; 1973b; 1973c] suggests more than one way of doing this. The first way is to replace the Uniqueness Assumption with the weaker Limit Assumption. Instead of looking at the closest antecedent-world, we look at all closest antecedent-worlds. These functions might better be called class selection functions rather than world selection functions. It is also not necessary to incorporate the accessibility relation into our models for conditionals if we use class selection functions since, if we make a certain reasonable assumption, we can define such a relation in terms of our class selection function. The assumption is that if  $\phi$  is possible at all at  $i$ , then there is at least one closest  $\phi$ -world for our selection function to pick out. Our models are then ordered triples  $\langle I, f, [ ] \rangle$  such that  $I$  and  $[ ]$  are as before and  $f$  is a function which assigns to each sentence  $\phi$  and each world  $i$  in  $I$  a subset of  $I$  (all the  $\phi$ -worlds closest to  $i$ ). By restricting these models appropriately, we can characterize a logic very similar to Stalnaker's **C2**. This logic, which Lewis calls **VC**, is the smallest conditional logic which is closed under the same rules as those listed for **C2** and which contains all those theses used in defining **C2** except that we replace CEM with

CS:  $(\phi \wedge \psi) \rightarrow (\phi > \psi)$ .

CS is contained by **C2** although CEM is not contained by **VC**.<sup>5</sup> A sentence

<sup>4</sup>A different sort of argument for CEM can be found in [Cross, 1985], which adopts Bennett's [1982] analysis of 'even if' conditionals and argues for the validity of CEM based on the intuitive validity of the following formulas:

$$\begin{aligned} &(\text{e } \phi > \psi) \rightarrow (\phi > \psi) \\ &(\psi \wedge \neg(\phi > \neg\psi)) \rightarrow (\text{e } \phi > \psi), \end{aligned}$$

where  $(\text{e } \phi > \psi)$  means 'Even if  $\phi$ ,  $\psi$ '. The argument turns on the fact that in any system of conditional logic that includes classical propositional logic and RCEC, CEM is a theorem iff

$$(\psi \wedge \neg(\phi > \neg\psi)) \rightarrow (\phi > \psi)$$

is a theorem.

<sup>5</sup>This and other independence results cited in this paper are provided in Nute [1979;

is a member of **VC** if and only if it is true at every world in every class selection function model which satisfies the following restrictions:

- (CS1): if  $j \in f(\phi, i)$  then  $j \in [\phi]$ ;
- (CS2): if  $i \in [\phi]$  then  $f(\phi, i) = \{i\}$ ;
- (CS3): if  $f(\phi, i)$  is empty then  $f(\psi, i) \cap [\phi]$  is also empty;
- (CS4): if  $f(\phi, i) \subseteq [\psi]$  and  $f(\psi, i) \subseteq [\phi]$ , then  $f(\phi, i) = f(\psi, i)$ ;
- (CS5): if  $f(\phi, i) \cap [\psi] \neq \emptyset$ , then  $f(\phi \wedge \psi, i) \subseteq f(\phi, i)$ ;
- (CS6):  $i \in [\phi > \psi]$  iff  $f(\phi, i) \subseteq [\psi]$ .

Although Lewis endorses **VC** as the proper logic for subjunctive conditionals, he finds the Limit Assumption and, hence, the version of class selection function semantics we have developed, to be no more satisfactory than the Uniqueness Assumption. Consequently, Lewis proposes an alternative semantics for subjunctive conditionals. This alternative is also based on the similarity of worlds. The difference is in the way Lewis uses similarity in giving the truth conditions for conditionals. A conditional  $\phi > \psi$  with a logically possible antecedent  $\phi$  is true at a world  $i$ , according to Lewis, if there is a  $\phi \wedge \psi$ -world which is closer to  $i$  than is any  $\phi \wedge \neg\psi$ -world. Lewis uses nested systems of spheres in his models to indicate the relative similarity of worlds. A system-of-spheres model is an ordered triple  $\langle I, \$, [ ] \rangle$  such that  $I$  and  $[ ]$  are as before and  $\$$  is a function which assigns to each  $i$  in  $I$  a nested set  $\$_i$  of subsets of  $I$  (the spheres about  $i$ ). If there is some sphere  $S$  about  $i$  such that  $j$  is in  $S$  but  $k$  isn't in  $S$ , then  $j$  is closer to or more similar to  $i$  than is  $k$ . To characterize the logic **VC**, we must adopt the following restrictions of system-of-spheres models:

- (SOS1):  $\{i\} \in \$_i$ ;
- (SOS2):  $i \in [\phi > \psi]$  if and only if  $\$_i \cap [\phi]$  is empty or there is an  $S \in \$_i$  such that  $S \cap [\phi]$  is not empty and  $S \cap [\phi] \subseteq [\psi]$ .

While Lewis rejects the Limit Assumption, it should be noted that in those cases in which there is a closest  $\phi$ -world to  $i$  the conditions for a conditional with antecedent  $\phi$  being true at  $i$  are exactly the same for system-of-spheres models as for the type of class selection function model we examined earlier. For this reason we classify Lewis's semantics as a minimal change semantics to contrast it with other accounts which lack this feature.

Pollock [1976] also develops a minimal change semantics for conditionals. In fact, Pollock's semantics is a type of class selection function semantics. There are two primary reasons why Pollock rejects Lewis's semantics and the

---

1980b].



conditional logic **VC**. First Pollock rejects the thesis **CV**, a thesis which is unavoidable in Lewis's semantics. Second Pollock embraces the Generalized Consequence Principle:

GCP: If  $\Gamma$  is a set of sentences such that  $\phi > \psi$  is true for each  $\psi \in \Gamma$ , and if  $\Gamma$  entails  $\chi$ , then  $\phi > \chi$  is true.

GCP does not hold in all system-of-spheres models, but it does hold in all class selection function models.<sup>6</sup>

The conditional logic **SS** which Pollock favors is the smallest conditional logic closed under the rules listed for **VC** and containing all those theses used in defining **VC** except that we replace **CV** with

CA:  $[(\phi > \psi) \wedge (\chi > \psi)] \rightarrow [(\phi \vee \chi) > \psi]$ .

This again gives us a weaker system since CA is contained by **VC** while **CV** is not contained by **SS**. Obviously, **SS** is not determined by the class of class selection function models which satisfy conditions (CS1)–(CS6) since this class of models characterizes the logic **VC**. Let's replace the condition (CS5) with

(CS5')  $f(\phi \vee \psi, i) \subseteq f(\phi, i) \cup f(\psi, i)$ .

Then **SS** is determined by the class of all class selection function models which satisfy this new set of conditions.

One reason for Pollock's lack of concern for Lewis's counterexamples to the Limit Assumption may be that Pollock conceives of what would count as a minimal change quite differently from the way Lewis does. Pollock [1976] offers a detailed account of the notion of a minimal change, an account

---

<sup>6</sup>To see how GCP might fail in Lewis's semantics, consider the example Lewis uses to show that for a particular  $\phi$  there may be no  $\phi$ -world closest to  $i$ . The example, which we considered earlier, involves a line printed on a page of [Lewis, 1973b]. Lewis invites us to consider worlds in which this line is longer than its actual length, which we will suppose to be exactly one inch. If the only way in which these worlds differ from the actual world is in the length of Lewis's line, then it is plausible that we rank these worlds in their similarity to the actual world according to how close to one inch Lewis's line is in each of these worlds. But no matter how close to one inch the line is, so long as it is longer than one inch there will be another such world in which it is closer to an inch in length. This means that for any length  $m$  greater than one inch, there is a world in which the line is longer than one inch and in which the line does not have length  $m$  which is nearer the actual world than is any world in which the line has length  $m$ . but then Lewis's truth conditions for conditionals dictate that if the line were longer than one inch, its length would not be  $m$ , and this is true for any length  $m$  greater than the actual length of the line. then consider the set  $\Gamma$  of sentences of the form 'Lewis's line is not length  $m$ ' where  $m$  ranges over every length greater than the actual length of Lewis's line. But  $\Gamma$  entails the sentence 'Lewis's line is not greater than one inch in length'. Applying GCP, we conclude that if Lewis's line were greater than one inch in length, then it would not be greater than one inch in length. This conclusion is not intuitively reasonable nor is it true at any world in the system-of-spheres model which Lewis describes in his discussion.

which is later modified in [Pollock, 1981]. The later view, which avoids many problems of the earlier view, will be discussed here.

While Stalnaker, Lewis and others maintain that the notions of similarity of worlds and of minimal change are vague notions which may change given different purposes and contexts, thus accounting for the vagueness we often find in the use of conditionals, Pollock claims that the similarity relation is not vague but quite definite. Pollock's account rests upon his use of two epistemological notions, that of a subjunctive generalisation and that of a simple state of affairs. Subjunctive generalisations are statements of the form 'Any  $F$  would be a  $G$ '; The truth of some subjunctive generalisations like 'Anyone who drank from the Chisholm's bottle would die' depends upon contingent matters of fact, in this case the fact that Chisholm's bottle contains strychnine and the fact that people have a certain physical make-up. Other subjunctive generalisations like 'Any creature with a physical make up like ours who drank strychnine would die' do not depend for their truth on contingent matters of fact in the same way. Pollock calls the former 'weak' subjunctive generalisations and the latter 'strong' subjunctive generalisations. Some subjunctive generalisations are supposed by Pollock to be directly confirmable by their instances, and these he calls basic. The problem of confirmation is discussed in [Pollock, 1984]. The second crucial ingredient in Pollock's analysis is the notion of a simple state of affairs. A state of affairs is simple if it can be known non-inductively to be the case without first coming to know some other state(s) of affairs which entail(s) it.

The actual world is supposed by Pollock to be determined by the set of true basic strong subjunctive generalisations together with the set of true simple states of affairs. The justification conditions for a subjunctive conditional  $\phi > \psi$  are stated in terms of making minimal changes in these two sets in order to accommodate  $\phi$ . The first step is to generate all maximal subsets of the set of true basic strong subjunctive generalisations which are consistent with  $\phi$ . For each such maximally  $\phi$ -consistent set  $N$  of true basic strong subjunctive generalisations, we then generate all sets of true simple states of affairs which are maximally consistent with  $N \cup \{\phi\}$ . Finally, we consider every possible world at which  $\phi$ , every member of some maximally  $\phi$ -consistent set  $N$  of true basic strong subjunctive generalisations, and every member of some set  $S$  of true simple states of affairs maximally consistent with  $N \cup \{\phi\}$  are all true. If  $\psi$  is true at all such worlds, then  $\phi > \psi$  is true at the actual world. The set of worlds determined by this procedure serves as the value of a class selection function.

If we try to define a relative similarity relation for worlds based upon Pollock's analysis of minimal change, we come up with a partial order rather than the 'complete' order assumed by Lewis and, apparently, by Stalnaker. Because we can have two worlds  $j$  and  $k$  such that their similarity to a

third world  $i$  is incomparable, the thesis CV does not hold for Pollock's semantics.<sup>7</sup> A simple model of Pollock's sort which rejects CV as well as another thesis which has been attributed to Pollock's conditional logic **SS** is developed in [Mayer, 1981].

Several authors have proposed theories which resemble Pollock's in important respects. One of these is Blue [1981] who suggests that we think of subjunctive conditionals as metalinguistic statements about a certain semantic relation between an antecedent set of sentences in an object language and another sentence of the object language viewed as a consequent. A theory (set of sentences of the object language) and the set of true basic (atomic and negations of atomic) sentences of the language play roles similar to those played by laws (true basic strong subjunctive generalisations) and simple states of affairs in Pollock's account. One problem with Blue's proposal is that treating conditional metalinguistically as he does prevents iteration of conditionals without climbing a hierarchy of metalanguages. Another problem concerns the role which temporal relations between the basic sentences plays in his theory, a problem for other theories as well. (This problem is discussed in Section 1.8 below.) For a more detailed discussion of Blue's view, see [Nute, 1981c].

The similarity of an account like Pollock's or Blue's to the cotenability theories of conditionals should be obvious. A conditional is true just in case its consequent is entailed (Blue uses a somewhat different relation) by its

---

<sup>7</sup>Pollock has offered various counterexamples to CV, the most recent of which involves a circuit having among its components two light bulbs  $L_1$  and  $L_2$ , three simple switches  $A$ ,  $B$ , and  $C$ , and a power source. These components are supposed to be wired together in such a way that bulb  $L_1$  is lit exactly when switch  $A$  is closed or both switches  $B$  and  $C$  are closed, while bulb  $L_2$  is lit exactly when switch  $A$  is closed or switch  $B$  is closed. At the moment, both bulbs are unlit and all three switches are open. Then the following conditionals are true:

$$(5a) \quad \neg(L_2 > \neg L_1)$$

$$(5b) \quad \neg[(L_2 \wedge L_1) > \neg(B \wedge C)]$$

The justification for (5a) is that one way to bring it about that  $L_2$  (i.e. that bulb  $L_2$  is lit) is to bring it about that  $A$  (i.e. that switch  $A$  is closed), but  $A > L_1$  is true. The justification for (5b) is that one way to make  $L_1$  and  $L_2$  both true is to close both  $B$  and  $C$ . Pollock claims that the following counterfactual is also true:

$$(5c) \quad L_2 > \neg(B \wedge C)$$

If Pollock is correct, then these three counterfactuals comprise a counterexample to CV. Pollock's argument for (5c) is that  $L_2$  requires only  $A$  or  $B$ , and to also make  $C$  the case is a gratuitous change and should therefore not be allowed. But this is an oversimplification. It is not true that only  $A$ ,  $B$ , and  $C$  are involved. Other changes which must be made if  $L_2$  is to be lit include the passage of current through certain lengths of wire where no current is now passing, etc. Which path would the current take if  $L_2$  were lit? We will probably be forced to choose between current passing through a certain piece of wire or switch  $C$  being closed. It is difficult to say exactly what the choices may be without a diagram of the kind of circuit Pollock envisions, but without such a diagram it is also difficult to judge whether closing switch  $C$  is gratuitous in the case of (5c) as Pollock claims.

antecedent together with some subset of the set of laws or theoretical truths and some (cotenable) set of simple states of affairs or basic sentences.

Veltman [1976] and Kratzer [1979; 1981] also propose theories of conditionals which resemble Pollock's in important respects. We will discuss Kratzer's view, although the two are similar. Kratzer suggests what can be called a premise or a partition semantics for subjunctive conditionals. Like Pollock, she associates with each world  $i$  a set  $H_i$  of propositions or states of affairs which uniquely determines that world. The set  $H_i$  is called a partition for  $i$  or a precise set for  $i$ . Kratzer proposes that we evaluate a subjunctive conditional  $\phi > \psi$  by considering  $\phi$ -consistent subsets of  $H_i$ .  $\phi > \psi$  is true at a world  $i$  if and only if each  $\phi$ -consistent subset  $X$  of  $H_i$  is contained in some  $\phi$ -consistent subset  $Y$  of  $H$  such that  $Y \cup \{\phi\}$  entails  $\psi$ . Kratzer points out that if every  $\phi$ -consistent subset of  $H_i$  is contained in some maximally  $\phi$ -consistent subset of  $H_i$ , then this truth condition is equivalent to the condition that  $\phi > \psi$  is true at  $i$  just in case  $\psi$  is entailed by  $X \cup \{\phi\}$  for every maximally  $\phi$ -consistent subset  $X$  of  $H_i$ .

If we assume that every  $\phi$ -consistent subset of  $H_i$  is contained in some maximally  $\phi$ -consistent subset of  $H_i$ , the specialized version of Kratzer's semantics we obtain looks very much like Pollock's. Lewis [1981a] notes that this assumption plays the same role in premise semantics that the Limit Assumption plays in class selection function semantics. In fact, Lewis shows that on this assumption Kratzer's premise semantics is formally equivalent to Pollock's semantics. Given this equivalence, these two semantics will determine exactly the same conditional logic **SS**.

Even if we assume that the required maximal sets always exist and adopt a version of premise semantics which is formally equivalent to Pollock's semantics, Kratzer's position would still differ radically from Pollock's since she does not assign to laws and simple states of affairs a privileged role in her analysis. Nor does she prefer Blue's object language theory and basic sentences for such a role. The set of premises which we associate with a world and use in the evaluation of conditionals varies, according to Kratzer, as the purposes and circumstances of the language users vary. Thus Kratzer reintroduces the vagueness which so many investigators have observed in ordinary usage and which Pollock and Blue would deny or at least eliminate.

Apparently Kratzer does not accept the Limit Assumption, in her case the assumption that the required maximal sets always exist. Yet in [Kratzer, 1981] she describes what she calls the most intuitive analysis of counterfactuals, saying that

The truth of counterfactuals depends on everything which is the case in the world under consideration: in assessing them, we have to consider all the possibilities of adding as many facts to the antecedent as consistency permits.

This certainly suggests maximal antecedent-consistent subsets of a premise set (the Limit Assumption) and a minimal change semantics. But if the Limit Assumption is unacceptable, this initial intuition must be modified. Kratzer's modification takes the form of the truth condition reported earlier. Besides the Limit Assumption, Kratzer's semantics also fails to support the GCP. One principle which does remain, a principle common to all the semantics discussed in this section, is the thesis CS. Beginning with some sort of minimalist intuition, all of these authors claim subjunctive conditionals with true antecedents have the same truth values as their consequents. When the antecedent of the conditional is true, the actual world is the unique closest antecedent world and hence the only world to be considered in evaluating the conditional.

If Lewis's counterexamples to the Limit Assumption are conclusive, we must conclude that all the semantics for subjunctive conditionals which we have discussed in this section must be inadequate except for Lewis' system-of-spheres semantics and the general version of Kratzer's premise semantics. And if the GCP is a principle which we wish to preserve, then Lewis's semantics and Kratzer's semantics are also inadequate. Besides these difficulties, minimal change theories have been criticized because they endorse the thesis CS. As was mentioned in Section 1.1, many researchers claim that some conditionals with true antecedent and consequent are false. For an excellent polemic against the minimal change theorists on this issue, see [Bennett, 1974].

### 1.5 *Small Change Theories*

Åqvist [1973] presents a very interesting analysis of conditionals in which the conditionals in which the conditional operator is defined in terms of material implication and some unusual monadic operators. Simplifying a bit, Åqvist's semantics involves ordered quintuplets  $\langle I, i, R, f, [ ] \rangle$  such that  $I$  and  $[ ]$  are as in other models we have discussed,  $i$  is a member of  $I$ ,  $R$  is an accessibility relation on  $I$ , and  $f$  is a function which assigns to each sentence  $\phi$  a subset  $f(\phi)$  of  $[ \phi ]$  such that for every member  $j$  of  $f(\phi)$ ,  $\langle i, j \rangle \in R$ . A sentence  $*\phi$  whose primary connective is the monadic star operator  $*$  is true at a world  $j$  in  $I$  just in case  $j \in f(\phi)$ . The usual truth conditions are provided for a necessity operator, so that  $\Box\phi$  is true at  $j$  in  $I$  just in case for every world  $k$  such that  $\langle j, k \rangle \in R$ ,  $\phi$  is true at  $k$ , i.e. just in case  $k \in [ \phi ]$ . Finally, a conditional  $\phi > \psi$  is true at a world  $j$  just in case  $\Box(*\phi \rightarrow \psi)$  is true at  $i$ . Åqvist modifies this semantics in an appendix. The modification involves a set of models of the sort described, each with the same set of possible worlds, the same accessibility relation, and the same valuation function, but each with its own designated world and selection function. The resulting semantics turns out once again to be equivalent to a version of class selection semantics.

While the interesting formal details of Åqvist's theory are quite different from those of other investigators, the most significant feature of his account may be his suggestion that a class selection function might properly pick out for a sentence  $\phi$  and a world  $i$  all those  $\phi$ -worlds which are 'sufficiently' similar to  $i$  rather than only those  $\phi$ -worlds which are 'most' similar to  $i$ . By changing the intended interpretation for the class selection function, we avoid the trivialisation of the truth conditions for conditionals in all those cases where the Limit Assumption in either of its forms fails. At the same time, class selection function semantics supports the GCP. Åqvist's suggestion looks very promising.

A similar approach is taken by Nute, [1975a; 1975b; 1980b], but the semantics Nute proposes is explicitly a version of class selection function semantics. This model theory differs from versions of class selection function semantics we examined earlier in two important ways. First the intended interpretation is different, i.e. there is a different informal explanation to be given for the role which the class selection functions play in the models. Second the restriction (CS2) is replaced by the weaker restriction

$$(CS2') \quad \text{if } i \in [\phi] \text{ then } i \in f(\phi, i).$$

The second change is related to the first. Surely any world is more similar to itself than is any other. Thus, if  $f$  picks out for  $\phi$  and  $i$  the  $\phi$ -worlds closest to  $i$ , and if  $i$  is itself a  $\phi$ -world, then  $f$  will pick out  $i$  and nothing else for  $\phi$  and  $i$ . The objection to the thesis CS, though can be thought of as a claim that there may be other worlds sufficiently similar to the actual world so that in some cases we should consider these worlds in evaluating conditionals with true antecedents. When we modify our earlier semantics for Lewis's system **VC** by replacing (CS2) with (CS2'), the resulting class of models characterizes the logic which Lewis [1973b] calls **VW**. **VW** is the smallest conditional logic which is closed under all the rules and contains all the theses listed for **VC** except for the thesis CS. By weakening our semantics further we can characterize a logic which is closed under all the rules and contains all the theses of **VW** except CV. This, of course, would give us a logic for which Pollock's **SS** would be a proper extension.

Although many count it as an advantage of small change class selection function semantics that such theories allow us to avoid CS, it should be noted that such semantics do not commit us to a rejection of CS. As we have seen, both Lewis's **VC** and Pollock's **SS** are characterized by classes of class selection function models. For those who favor CS, it is still possible to avoid the difficulties of the Limit Assumption and embrace the GCP by adopting one of these versions of the class selection function semantics but giving a small change interpretation of the selection functions upon which such a semantics depends.

It is possible to avoid CS within the restrictions of a minimal change semantics. We can do this by 'coarsening' our similarity relation, to use

Lewis's phrase, counting worlds as equally similar to some third world despite fairly large differences in these worlds. For example, we might count some worlds other than  $i$  as being just as similar to  $i$  as is  $i$  itself. When we do this for a minimal change version of class selection function semantics, the formal results are exactly the same as those proposed earlier in this section and the resulting logic is **VW**. Of course, we must still cope with Lewis's objections to the Limit Assumption. But it is even possible to avoid CS within Lewis's system-of-spheres semantics. All we need to do is replace the restriction (SOS1) with the following:

(SOS1')  $i \in \cap s_i$ .

The class of all those system-of-spheres models which satisfy (SOS1') and (SOS2) determines the conditional logic **VW**. While such a concession to the critics of CS is possible within the confines of Lewis's semantics, Lewis does not favor such a move. We should also remember that the resulting semantics still does not support the GCP.

Since we can formulate a kind of minimal change semantics which avoids the controversial thesis CS, the only advantage we have shown for small change theories is that they avoid the problems of the Limit Assumption while giving support for the GCP. But this advantage may be illusory. Loewer [1978] shows that for many versions of class selection function semantics we can always view the selection function as picking out closest worlds. For a model of such a semantics, we can define a relative similarity relation  $R$  between the worlds of the model in terms of the model's selection function  $f$ . It can then be shown that for a sentence  $\phi$  and a world  $i$ ,  $j \in f(\phi, i)$  if and only if  $j$  is a  $\phi$ -world which is at least as close to  $i$  with respect to  $R$  as is any other  $\phi$ -world. Consider such a model and consider a proposition  $\phi$  which is true at world  $j$  just in case there are infinitely many worlds closer to  $i$  with respect to  $R$  than is  $j$ . There will be no  $\phi$ -world closest to  $i$ . Consequently  $f(\phi, i)$  will be empty and any conditional with  $\phi$  as antecedent will be trivially true. How seriously we view this example depends upon our attitude toward the assumption that there exists a proposition which has the properties attributed to  $\phi$ . If we take propositions to be sets of worlds, then the existence of such a proposition is very plausible. We should also note that this argument involves a not so subtle change in our semantics. Until now we have been thinking of our selection functions as taking sentences as arguments rather than propositions. If we restrict ourselves to sentences it is very unlikely that our language for conditional logic will contain a sentence which expresses the troublesome proposition. Nevertheless, it is not entirely clear that every small change version of class selection function semantics will automatically avoid the problems associated with the Limit Assumption.

There is another advantage which can be claimed for small change theories which doesn't involve the logic of conditionals. If, for example, **VW**

is the correct logic for conditionals, we have seen that it is possible to take either the minimal change or the small change approach to semantics for conditionals and still provide a semantics which determines **VW**. But even if agreement is reached about which sentences are valid, these two approaches are still likely to result in different assignments of truth values to contingent conditional sentences. Suppose for example that Fred's lawn is just slightly too short to come into contact with the blades of his lawnmower. Thus his lawnmower will not cut the grass at present. Suppose further that the engine on Fred's lawnmower is so weak that it will only cut about a quarter of an inch of grass. If the height of the grass is more than a quarter of an inch greater than the blade height, the mower will stall. Then is the following sentence true or false?

29. If the grass were higher, Fred's mower would cut it.

On the minimal change approach, whether we use class selection function semantics or system-of-spheres semantics, the answers to this question must be 'yes' for there will be worlds at which the lawn is higher than the blade height but no more than a quarter inch higher than the blade height, which are closer to the actual world than is any world at which the grass is more than a quarter inch higher than the blade height. But the correct answer to the question would seem to be 'no'. If someone were to assert (29) we would likely object, 'Not if the grass were much higher'. This shows that we are inclined to consider changes which are more than minimally small in our evaluations of conditionals. We might avoid particular examples of this sort by 'coarsening' the similarity relation, but it may be possible to generate such examples for any similarity relation no matter how coarse.

All of the small change theories we have considered propose semantics which are at least equivalent to some version of class selection function semantics. There is, however, at least one small change theory which does not share this feature. Warmbrød [1981] presents what he calls a pragmatic theory of conditionals. This theory is based on similarity of worlds but in a radically different way than are any of the theories we have yet examined. According to Warmbrød, the set of worlds we use in evaluating a conditional is determined not by the antecedent of that particular conditional but rather by all the antecedents of conditionals occurring in the piece of discourse containing that particular conditional. Thus a conditional is always evaluated relative to a piece of discourse rather than in isolation. For any piece of discourse  $D$  and world  $i$  we select a set of worlds  $S$  which satisfies the following conditions:

- (W1) if  $\phi > \psi$  occurs in  $D$  and  $\phi$  is logically possible, then some world  $j$  in  $S$  is a  $\phi$ -world;
- (W2) for some  $\phi > \psi$  occurring in  $D$ ,  $j \in S$  if and only if  $j$  is at least as close to  $i$  as are the closest  $\phi$ -worlds to  $i$ .



Condition (W1) ensures that  $S$  is what Warmbröd calls *normal* for  $D$  and (W2) ensures that  $S$  is what Warmbröd calls *standard* for some antecedent occurring in  $D$ . (Warmbröd formulates his theory in terms of an accessibility relation, but the semantics provided here is formally equivalent.) Then a conditional  $\phi > \psi$  is true at  $i$  with respect to  $D$  if and only if  $\phi \rightarrow \psi$  is true at every world in  $S$ . The resulting semantics resembles both class selection function semantics and an analysis of conditionals as strict conditionals, but it differs from each of these approaches in important respects.

Like other proposals which treat subjunctive conditionals as being strict conditionals, Warmbröd's theory validates Transitivity, Contraposition, and Strengthening Antecedents. Warmbröd argues that the evidence against these theses can be explained away. Apparent counterexamples to transitivity, for example, depend according to Warmbröd on the use of different sets  $S$  in the evaluation of the sentences involved in the putative counterexamples. Consider the example (21)–(23) in Section 1.3 above. According to Warmbröd, this example can be a counterexample to Transitivity only if there is some set of worlds  $S$  which contains worlds at which Carter did not lose in 1980, contains some worlds at which Carter died in 1979, which is normal for these two antecedents, and for which the material conditional corresponding to (21) and (22) are true at all members of  $S$  while the material conditional corresponding to (23) is false at some world in  $S$ . But this, Warmbröd claims, is exactly what does not happen. The apparent counterexample depends upon an equivocation, a shift of the set  $S$  during the course of the argument.

Warmbröd's theory has a certain attraction. It is certainly true that Transitivity and other controversial theses are harmless in many contexts, and it is certainly true that these theses are frequently used in ordinary discourse. The problem is to provide an account of the difference between those situations in which the thesis is reliable and those in which it is not. Warmbröd's strategy is to consider the thesis to be always reliable and then to provide a way of falsifying the premises in unhappy cases. An alternative approach is to count these theses as being invalid and then to look for those features of context which sometimes allow us to use them with impunity. We think the second strategy is safer. It is probably better to occasionally overlook a good argument than it is to embrace a bad one. Or to put a bit differently, it is better to force the argument to bear the burden of proof rather than to consider it sound until proven unsound.

Another problem with Warmbröd's theory is that it suggests that we should find apparent counterexamples to certain theses which have until now been considered uncontroversial. For example, we should find apparent counterexamples for CA. (See [Nute, 1981b] for details.)

Warmbröd's semantics also runs into difficulty with the Limit Assumption. The requirement (W2) that  $S$  be standard for some antecedent in  $D$  involves the Limit Assumption explicitly, although Warmbröd's semantics

may tolerate small, non-minimal changes for some of the antecedents in a piece of discourse, it demands that only minimal changes be considered for at least one such antecedent. Of course, we might be able to modify (W2) in such a way as to avoid this problem. There remains, though, the nagging suspicion that none of the small change theories we have considered will in the end be able to escape the Limit Assumption, with all its difficulties, in some form or other.

### 1.6 *Maximal Change Theories*

Both minimal change theories and small change theories of conditionals are based on the premise that a conditional  $\phi > \psi$  is true at  $i$  just in case  $\psi$  is true at some  $\phi$ - world(s) satisfying certain conditions. The difference, of course, is that for the one approach it is sufficient that  $\psi$  be true at all closest  $\phi$ -worlds while the other requires that  $\psi$  be true at all  $\phi$ -worlds which are reasonably or sufficiently close to  $i$ . There is a third type of theory which shares the same basic premise as these two but which does not require that the worlds upon which the evaluation of  $\phi > \psi$  at  $i$  depends be very close or similar to  $i$  at all. According to this way of looking at conditionals, all that is required is that the relevant worlds resemble  $i$  in certain very minimal respects. Otherwise the relevant worlds may differ from  $i$  to any degree whatever. We might even think of this approach as requiring us to consider worlds which differ from  $i$  maximally except for the narrowly defined features which must be shared with  $i$ .

One theory of this sort is developed by Gabbay [1972]. To facilitate comparison, we will simplify Gabbay's account of conditionals rather drastically. When we do this, Gabbay's semantics for conditionals resembles the class selection function semantics we have discussed, but there are some very important differences. A simplified Gabbay model is an ordered triple  $\langle I, g, [ ] \rangle$  such that  $I$  and  $[ ]$  are as in earlier models, and  $g$  is a function which assigns to sentences  $\phi$  and  $\psi$  and world  $i$  in  $I$  a subset  $g(\phi, \psi, i)$  of  $I$ . A conditional  $\phi > \psi$  is true at  $i$  in such a model just in case  $g(\phi, \psi, i) \subseteq [\phi \rightarrow \psi]$ . The difference between this and class selection function semantics of the sort we have seen previously is obvious: the selection function  $g$  takes both antecedent and consequent as argument. This means that quite different sets of worlds might be involved in the truth conditions for two conditionals having exactly the same antecedent. This change in the formal semantics reflects a difference in Gabbay's attitude toward conditionals and toward the way in which we evaluate conditionals. When we evaluate  $\phi > \psi$ , we are not concerned to preserve as much as we can of the actual world in entertaining  $\phi$ ; instead we are concerned to preserve only those features of the actual world which are relevant to the truth of  $\psi$ , or perhaps to the effect  $\phi$  would have on the truth of  $\psi$ . In actual practice the kind of similarity which is required is supposed by Gabbay to be determined by  $\phi$ , by  $\psi$ , and

also by general knowledge and particular circumstances which hold in  $i$  at the time when the conditional is uttered. What this involves is left vague, but it is not more vague than the notions of similarity assumed in earlier theories.

When we modify Gabbay's semantics in this way, we must impose three restrictions on the resulting models:

- (G1)  $i \in g(\phi, \psi, i)$ ;
- (G2) if  $[\phi] = [\psi]$  and  $[\chi] = [\theta]$  then  $g(\phi, \chi, i) = g(\psi, \theta, i)$ ;
- (G3)  $g(\phi, \psi, i) = g(\phi, \neg\psi, i) = g(\neg\phi, \psi, i)$ .

With these restrictions Gabbay's semantics determines the smallest conditional logic which is closed under RCEC and the following two rules:

RCEA: from  $\phi \leftrightarrow \psi$ , to infer  $(\phi > \chi) \leftrightarrow (\psi > \chi)$ .

RCE: from  $\phi \rightarrow \psi$ , to infer  $\phi > \psi$ .

We will call this logic  $\mathbf{G}$ . At the end of [Gabbay, 1972], a conjectured axiomatisation of  $\mathbf{G}$  is presented, but it was later shown to be unsound and incomplete in [Nute, 1977], where the axiomatisation of  $\mathbf{G}$  presented here was conjectured to be sound and complete (see [Nute, 1980b]). Working independently, David Butcher [1978] also disproved Gabbay's conjecture, and proved the soundness and completeness of  $\mathbf{G}$  for the Gabbay semantics (see [Butcher, 1983a]).

It is obvious that  $\mathbf{G}$  is the weakest conditional logic we have yet considered. We can characterize a stronger logic if we place additional restrictions on our Gabbay models, but we may not be able to guarantee a sufficiently strong logic without restricting our models to the point where they become formally equivalent to models we examined earlier. Consider, for example the theses

CC:  $[(\phi > \psi) \wedge (\phi > \chi)] \rightarrow [\phi > (\psi \wedge \chi)]$

CM:  $[\phi > (\psi \wedge \chi)] \rightarrow [(\phi > \psi) \wedge (\phi > \chi)]$ .

To ensure that our conditional logic contains CC and CM, we could impose the following restriction on Gabbay's semantics:

(G4)  $g(\phi, \psi, i) = g(\phi, \chi, i)$ .

Once we do this we have eliminated the most distinctive feature of Gabbay's semantics. According to David Butcher [1983a], it is possible to ensure CC and CM by adopting conditions weaker than (G4). However, Butcher has indicated that these conditions are problematic for other reasons.<sup>8</sup>

<sup>8</sup>Many of these issues are also discussed in Butcher [1978; 1983a].

A rather different and very specialized maximal change theory has been developed in two different forms by Fetzer and Nute [1979; 1980] and by Nute [1981a]. Both forms of this theory are intended not as analyses of ordinary subjunctive conditionals as they are used in ordinary discourse, but rather as analyses of scientific, nomological, or causal conditionals, i.e. of subjunctive conditionals as they are used in the very special circumstances of scientific investigation. Formally the two theories propose class selection function semantics for scientific conditionals, but the intended interpretation is quite different from that of any theory we have yet considered.

In the version of the theory developed by Fetzer and Nute the selection function  $f$  is intended to pick out for a sentence  $\phi$  and a world  $i$  the set of all those  $\phi$ -worlds at which all the individuals mentioned in  $\phi$  possess, in so far as the truth of  $\phi$  will allow, all those dispositional properties which they permanently possess in  $i$ . This forces us to ignore all features of worlds except those assumed by the underlying theory of causality to affect the causal efficacy of the situation, events, etc., described in  $\phi$ . We are forced, in other words, to consider worlds which preserve only these features and otherwise differ maximally from the world at which the scientific conditional is being evaluated. In this way we can ensure that the conditional in question is true if but only if the antecedent and the consequent are related causally or nomologically in an appropriate manner. Physical law statements are then analysed as universal generalisations or sets of universal generalisations of such scientific conditionals.

The view subsequently developed in [Nute, 1981a] departs a bit from the requirement of maximal specificity which we seek in our scientific pronouncements and in doing so comes closer to representing a kind of conditional used in ordinary discourse. Nute suggests that the selection function  $f$  selects for a sentence  $\phi$  and a world  $i$  all those  $\phi$ -worlds at which all those individuals mentioned in  $\phi$  possess, so far as the truth of  $\phi$  allows, not only all those dispositional properties which they permanently possess in  $i$  but also all those dispositional properties which they accidentally or as a matter of particular fact possess in  $i$ . For example, a particular piece of litmus paper permanently possesses the tendency to turn red when dipped in an acidic solution, since it could not lose this tendency and still be litmus paper, but it only accidentally possesses the tendency to reflect blue light, since it could certainly lose this disposition through being dipped in acid and yet still be litmus paper. Where it is impossible to accommodate  $\phi$  without giving up some of the dispositional properties possessed by individuals mentioned in  $\phi$ , preference is given to dispositions which are possessed permanently. On this account, but not on the account developed by Fetzer and Nute conjointly, the following conditional is true where  $x$  is a piece of litmus paper which is in fact blue:

30. If  $x$  were cut in half, it would be blue.

Nute [1981a] suggests that many ordinary conditionals may have such truth conditions, or may be abbreviations of other more explicit conditionals which have such truth conditions.

Each of the theories presented in this section is in fact only a fragment of a more complex theory. It is impossible to discuss the larger theories in any greater detail and the reader is encouraged to consult the original publications. What allows us to consider them under a single category is their departure from the premise that the truth of a conditional depends upon what happens in antecedent- worlds which are very much like the actual world. Each of these theories assumes and even requires that the divergence from the actual world be rather larger than minimal or small change theories would indicate.

### 1.7 *Disjunctive Antecedents*

One thesis in particular has caused considerable controversy among the investigators of conditional logic. This thesis is Simplification of Disjunctive Antecedents:

$$\text{SDA: } [(\phi \vee \psi) > \chi] \rightarrow [(\phi > \chi) \wedge (\psi > \chi)].$$

The intuitive plausibility of SDA has been suggested in [Fine, 1975], in [Nute, 1975b] and in [Ellis *et al.*, 1977]. Unfortunately, any conditional logic which contains SDA and which is also closed under substitution of provable equivalents will also contain the objectionable thesis Strengthening Antecedents. If we add SDA to any of the logics we have discussed, then Transitivity and Contraposition will be contained in the extended logic as well.<sup>9</sup> Ellis *et al.* suggest that the evidence for SDA is so strong and the problems involved in trying to incorporate SDA into any account of conditionals based upon possible worlds semantics is so great that the possibility of an adequate possible worlds semantics for ordinary subjunctive conditionals is quite eliminated. With all the problems which the various theories encounter, the possible worlds approach has still proven to be a powerful tool for the investigation of the logical and semantical properties of conditionals and we should be unwilling to abandon it without first trying to defend it against such a charge.

The first line of defence has been a ‘translation lore’ approach to the problem of disjunctive antecedents. It is first noted that, despite the intuitive appeal of SDA, there are examples from ordinary discourse which show that SDA is not entirely reliable. The following sentences comprise one such example:

---

<sup>9</sup>As further evidence of the problematic character of SDA, David Butcher [1983b] has shown that any logic containing SDA and CS will contain  $\phi \rightarrow \Box\phi$ , where  $\Box\phi$  is defined as  $\neg\phi > \phi$ .

- 31a. If the United States devoted more than half of its national budget to defence or to education, it would devote more than half of its national budget to defence.
- 31b. If the United States devoted more than half of its national budget to education, it would devote more than half of its national budget to defence.

Contrary to what we should expect if SDA were completely reliable, it looks very much as if (31a) is true even though (31b) cannot be true. Fine [1975], Loewer [1976], McKay and Van Inwagen [1977] and others have suggested that those examples which we take to be evidence for SDA actually have a quite different logical form from that which supporters of SDA suppose them to have. While a sentence like (31a) really does have the form  $(\phi \vee \psi) > \chi$ , a sentence like

32. If the world's population were smaller or agricultural productivity were greater, fewer people would starve.

has the quite different logical form  $(\phi > \chi) \wedge (\psi > \chi)$ . According to this suggestion, the word 'or' represents wide scope conjunction rather than narrow scope disjunction in (32). Since we can obviously simplify a conjunction, this confusion about the logical form of sentences like (32) results in the mistaken commitment to a thesis like SDA.

This would be a neat solution to the problem if it would work, but the translation lore approach has a serious flaw. According to the translation theorist, the two sentences (31a) and (32) have different logical forms even though they share the same surface or grammatical structure. We can point out an obvious difference in surface structure since one of the apparent disjuncts in the antecedent of (31a) is also the consequent of (31a), a feature which (32) lacks. But we can easily produce examples where this is not the case. Suppose after asserting (31a) a speaker went on to assert

33. So if the United States devoted over half its national budget to defence or education, my Lockheed stock would be worth much more than it is.

It would be very reasonable to accept this conditional but at the same time to reject the following conditional:

34. So if the United States devoted over half of its national budget to education, my Lockheed stock would be worth much more than it is.

The occurrence of the same component sentence in antecedent and consequent is not a necessary condition for the failure of SDA and cannot be used as a criterion for distinguishing those cases in which English conditional with 'or' in their antecedents are of the logical form  $(\phi \vee \psi) > \chi$  from

those in which they are of the logical form  $(\phi > \chi) \wedge (\psi > \chi)$ . We cannot decide on purely syntactical grounds which of the two possible symbolisations is proper for an English conditional with ‘or’ in its antecedent. Loewer [1976] suggests that this decision may be made on pragmatic grounds, but it is difficult to see what the distinguishing criterion is to be except that English conditionals with disjunctive antecedents are to be symbolized as  $(\phi > \chi) \wedge (\psi > \chi)$  when simplification of their disjunctive antecedents is legitimate and to be symbolized as  $(\phi \vee \psi) > \chi$  when such simplification is not legitimate. Until Loewer’s suggestion concerning the pragmatic pressures which prompt one symbolisation rather than another can be provided with sufficient detail, the translation lore account of disjunctive conditional does not provide us with an adequate solution to our problem.

We find an interesting variation on the translation lore solution in [Humberstone, 1978] and in [Hilpinen, 1981]. Both suggest the use of an antecedent forming operator like Åqvist’s  $*$ . We will discuss Hilpinen’s theory here since it differs the most from Åqvist’s view. Hilpinen’s analysis utilizes two separate operators which we can represent as If and Then. The If operator attaches to a sentence  $\phi$  to produce an antecedent If  $\phi$ . the Then operator connects an antecedent  $\alpha$  and a sentence  $\phi$  to form a conditional  $\alpha$  Then  $\phi$ . The role of the dyadic truth functional connectives is expanded so that  $\vee$ , for example, can connect two antecedents  $\alpha$  and  $\beta$  to form a new antecedent  $\alpha \vee \beta$ . An important difference between Hilpinen’s If operator and Åqvist’s  $*$  is that for Åqvist  $*\phi$  is a sentence or proposition bearing a truth value while for Hilpinen If  $\phi$  is not. Finally Hilpinen proposes that sentences like (31a) be symbolized as If  $(\phi \vee \psi)$  Then  $\chi$  while sentences like (32) be symbolized as (If  $\phi \vee$  If  $\psi$ ) Then  $\chi$ . Hilpinen then accepts a rule similar to SDA for sentences having the latter form but not for sentences having the former. This proposal allows us to incorporate a rule like SDA into our conditional logic while avoiding Strengthening Antecedents, etc., and, unlike other versions of the translation lore approach, Hilpinen’s proposal seems to suggest how it might be possible for sentences like (31a) and (32) to have a legitimate scope ambiguity in their syntactical structure, like the scope ambiguity in ‘President Carter has to appoint a woman’. In fact, however, the ambiguity postulated by Hilpinen’s proposal does not seem simply to be a scope ambiguity. The sentence ‘President Carter has to appoint a woman’ is ambiguous with respect to the scope of the phrase ‘a woman’, but the phrase ‘a woman’ has the same syntactical function and the same semantics on both readings of the sentence. The same cannot be said of the word ‘or’ in Hilpinen’s account of disjunctive antecedents: on one resolution of the ambiguity, what ‘or’ connects in examples like (31a) and (32) are sentences; on the other resolution of the ambiguity, ‘or’ connects phrases that are not sentences. It is difficult to see how the ambiguity in (31a) and (32) can be simply a scope ambiguity if ‘or’ does not have the same syntactical role in both readings of a given sentence.

Another approach to disjunctive antecedents is developed by Nute [1975b; 1978b] and [1980b]. Formally the problem with SDA is that it together with substitution of provable equivalents results in Strengthening Antecedents and other unhappy results. The translation theorist's suggestion is that we abandon SDA. Nute's suggestion, on the other hand, is that we abandon substitution of provable equivalents, at least for antecedents of subjunctive conditionals. One fairly strong logic which does not allow substitution of provably equivalent antecedents is the smallest conditional logic which is closed under RCEC and RCK and contains ID, MP, MOD, CV, and SDA. Logics of this sort have been called 'non-classical' or 'hyperintensional' to contrast them with those intensional logics which are closed under substitution of provable equivalents. Classical logics (those closed under substitution of provable equivalents) are preferred by most investigators.

Besides the fact that non-classical logics are much less elegant than classical logics, Nute's proposal has other very serious difficulties. First, substitution of certain provable equivalents within antecedents appears to be perfectly harmless. For example, we can surely substitute  $\psi \vee \phi$  for  $\phi \vee \psi$  in  $(\phi \vee \psi) > \chi$  with impunity. How are we to decide which substitutions are to be allowed and which are not? Non-classical conditional logics which allow extensive substitutions are developed in Nute [1978b] and [1980b]. But these systems are extremely cumbersome and there still is the extra-formal problem of justifying the particular choice of substitutions which are to be allowed in the logic. Second, we are still left with the apparent counterexamples to SDA like (31a). Nute suggests a pluralist position, maintaining that there are actually several different conditionals in common use. For some of these conditionals SDA is reliable while for others it is not. The conditional involved in (31a), it is claimed, is unusual and should not be represented in the same way as other subjunctive conditionals. While there is good reason to admit a certain pluralism, to admit, for example, the distinction between subjunctive and indicative conditionals, Nute's proposal is little more than a new translation lore in disguise. The translation lore we discussed earlier at least has the virtue that it attempts to explain the perplexities surrounding disjunctive antecedents in terms of a widely accepted set of logical operators without requiring the recognition of any new conditional operators. Non-classical logic appears to be a dead end so far as the problem of disjunctive antecedents is concerned.

A completely different solution is suggested in [Nute, 1980a], a solution based upon the account of conversational score keeping developed in [Lewis, 1979b]. Basically, the proposal concerns the way in which the class selection function (or the system-of-spheres if Lewis-style semantics is employed) becomes more and more definite as a linguistic exchange proceeds. During a conversation, the participants tend to restrict the selection function which they use to interpret conditionals in such a way as to accommodate claims made by their fellow participants. This growing set of restrictions on the se-



lection function forms part of what Lewis calls the score of the conversation at any given stage. Some accommodations, of course, will not be forthcoming since some participant will be unwilling to evaluate conditionals in the way which these accommodations would require. Each restriction on the selection function which the participants implicitly accept will also rule out other restrictions which might otherwise have been allowed. Nute's suggestion is that our inclination is to restrict the selection function in such a way to make SDA reliable, but that this inclination can be overridden in certain circumstances by our desire to accommodate the utterance of another speaker. When we hear the utterance of a sentence like (31a), for example, we restrict our selection function so that SDA becomes unreliable for sentences which have 'the United States devotes more than half its national budget to defence or education' as antecedent. Once (31a) is accommodated in this way, this restriction on the selection function remains in effect so long as the conversational context does not change. Nute completes his account by formulating some 'accommodation' rules for class selection functions. By offering a pragmatic account of the way in which the selection function becomes restricted during the course of a conversation, and by paying attention to the inclination to restrict the selection function in such a way as to make SDA reliable whenever possible, it may be possible to explain the fact that SDA is usually reliable while at the same time avoiding the many difficulties involved in accepting SDA as a thesis of our conditional logic.

This proposal is similar in certain respects to Loewer's [1976]. Like Loewer, Nute is recognising the important role which pragmatic features play in our use of conditionals with disjunctive antecedents. However, Nute's use of Lewis's notion of conversational score keeping results in an account which provides more details about what these pragmatic features might be than does Loewer's account. We also notice that Nute's suggestions might provide the criterion which Loewer needs to distinguish those conditionals which should be symbolized  $s(\phi \vee \psi) > \chi$  from those which should be symbolized as  $(\phi > \chi) \wedge (\psi > \chi)$ . But once the distinction is explained in terms of the evolving restrictions on class selection functions, there is no need to require that these conditionals be symbolized differently. The point of Nute's theory is that all such conditionals have the same logical form, but the reliability of SDA will depend on contextual features.

There is also considerable similarity between Nute's second proposal and Warmbröd's semantics for conditionals which was discussed in Section 1.5. In fact, Warmbröd's semantics is offered at least in part as an alternative to Nute's proposed solution to the problem of disjunctive antecedents. The important similarity between the two approaches is that both recognize that the interpretation of a conditional is a function not of the conditional alone but also of the situation within which the conditional is used. The important difference is that Warmbröd's semantics makes SDA, Transitivity, Contraposition, Strengthening Antecedents, etc. valid and uses pragmatic

considerations to explain and guard us from those cases where it seems to be a mistake to rely upon these principles, while Nute ultimately rejects all of these principles, but uses pragmatic considerations to explain why it is perfectly reasonable to use at least one of these theses, SDA, in many situations.

Warmbrød also offers a translation lore as part of his account. His suggestion about the way in which we should symbolize English conditionals with disjunctive antecedents is essentially that of Fine, Lewis, Loewer, and others, but he offers purely syntactic criteria for determining which symbolisation is appropriate in a particular case. His semantics is offered as a justification for his translation lore in an attempt to make his rules for symbolisation appear less *ad hoc*. Warmbrød points out some difficulties with Nute's rules of accommodation for class selection functions, and his translation rules might be used as a model for improving the formulation of Nute's rules. Nute's theory of disjunctive antecedents in terms of conversational score might also be proposed as an alternative justification for Warmbrød's translation rules.

### 1.8 *The Direction of Time*

We turn now to a problem alluded to in Section 1.4, a problem which concerns the role temporal relations play in the truth conditions for subjunctive conditionals. Actually, there are two different sets of problems to be considered. One of these involves the use of tensed language in conditionals and the other does not depend essentially on the use of tense and conditionals together. We will consider the latter set of problems in this section and save problems concerning tense for the next section.

A particularly thorny problem for logicians working on conditionals has to do with so-called backtracking conditionals, i.e. conditionals having antecedents concerned with events or states of affairs occurring or obtaining at times later than those involved in the consequents of the conditional. It is widely held that such conditionals are rarely true, and that when they are true they usually involve much more complicated antecedents and consequents than do the more usual true non-backtracking conditionals. Consider, for example, the two conditionals:

35. If Hinckley had been a better shot, Reagan would be dead.

36. If Reagan were dead, Hinckley would have been a better shot.

The first of these two conditionals is an ordinary non-backtracking conditional, while the second is a backtracking conditional. the first is very plausible and perhaps true, while the second is surely false. The problem with (36) which makes it so much less plausible than (35) is that Reagan might have died subsequent to the assassination attempt from any number

of causes which would not involve an improvement of Hinckley's aim. The problem for the logician or semanticist is to explain why non-backtracking conditionals are more often true than are backtracking conditionals.

The primary goal of Lewis [1979a] is to explain this phenomenon. Lewis's proposal makes explicit, extensive use of the technical notion of a miracle. In a certain sense miracles do not occur at all in Lewis's analysis: rather a miracle occurs in one world relative to another world. No event ever occurs in any world which violates the physical laws of that world, but events can certainly occur in one world which violate the physical laws of some other world. These are the kinds of miracles Lewis relies upon. Assuming complete determinism, which Lewis does at least for the sake of argument, any world which shares a common history with the actual world up to a certain point in time but which diverges from the actual world after that time cannot obey the same physical laws as does the actual worlds. Basically Lewis proposes that the worlds most similar to the actual world in which some counterfactual sentence  $\phi$  is true are those worlds which share their history with the actual world up until a brief transitional period beginning just prior to the times involved in the truth conditions for  $\phi$ . In the case of (35) this might mean that everything happens exactly as it did except that Hinckley miraculously aimed better than he actually did. This might only require something as small as a neuron firing at a slightly different time than it actually did. This is about as small a miracle as we could hope for. Once this miracle occurs, events are assumed by Lewis to once again follow their lawful course with the result, perhaps, that Reagan is mortally wounded. In the case of (36), on the other hand, Reagan might be dead if the FBI agent miraculously failed to jump in front of Reagan, if Reagan miraculously moved in such a way that the bullet struck him differently, or even if Reagan miraculously had a massive stroke at any time after the assassination attempt. Even if events followed their lawful course after any of these miracles, Hinckley's aim would not be improved.

Lewis notes that the vagueness of conditionals requires that there may be various ways of determining the relative similarity of worlds, different ways being employed on different occasions. There is one way of resolving vagueness which Lewis considers to be standard, and it is this way which provides us with the explanation of (35) and (36) given above. This standard resolution of vagueness is expressed in the following guidelines for determining the relative similarity of worlds:

- (L1) It is of the first importance to avoid big, complicated, varied, widespread violations of law.
- (L2) It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.

- (L3) It is of the third importance to avoid even small, simple, localized violations of law.
- (L4) It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

Lewis would maintain that application of these guidelines together with his system-of-spheres semantics for subjunctive conditionals will have the desired result of making (35) at least plausible while making (36) clearly false.

One major objection to Lewis's account is that once we allow miracles in order to produce a world which diverges from the actual world, there is nothing in Lewis's guidelines to prevent us from allowing another small miracle in order to get the worlds to converge once again. Since Lewis's guidelines place a higher priority on maximising the area of perfect match of particular facts over the avoidance of small, localized violations of law, we should prefer a small convergence miracle to a future which is radically different. Lewis's response to such a suggestion is that divergence miracles tend to be much smaller than convergence miracles or, what amounts to the same thing, that past events are overdetermined to a greater extent than are future events. If correct, then Lewis's guidelines would place greater importance on avoidance of a large convergence miracle than on maximising the area of perfect match of a particular fact. and careful consideration of examples indicates that Lewis's suggestion is at least plausible, although no conclusive argument has been provided. In [Nute, 1980b] examples of very simple worlds are given in which convergence miracles could be quite small and in which Lewis's guidelines would thus dictate that for some counterfactual antecedents the nearest antecedent worlds are those in which such small convergence miracles occur. In these examples, we get the (intuitively) wrong result when we apply Lewis's standard method for resolving the vagueness of conditionals. Lewis [1979a] warns that his guidelines might not work for very simple worlds, though, so the force of Nute's examples is uncertain. Lewis's guidelines may give an adequate explanation for our use of conditionals in the context of a complex world like the actual world, and since our intuitions are developed for such a world they may be unreliable when applied to very simple worlds.

If we consider Lewis's proposal in the context of a probabilistic world, we discover that we no longer need employ the troublesome notion of a miracle. Instead of a miracle, we can accommodate a counterfactual antecedent in a probabilistic world by going back to some underdetermined state of affairs among the causal antecedents of the events or states of affairs which must be eliminated if the antecedent is to be true and change them accordingly. Since these states of affairs were underdetermined to begin with, they could have been otherwise without any violation of the probabilistic laws governing the

universe. But if we do this, Lewis's emphasis on maximising the spatio-temporal area of perfect match of particular fact would require that we always change a more recent rather than an earlier causal antecedent when we have a choice. This consequence is very much like the Requirement of Temporal Priority in [Pollock, 1976], a principle which is superseded by the more complex account to be discussed below. Such a principle is unacceptable. Suppose, for example, that Fred left his coat unattended in a certain room yesterday. Today he returned to the room and found the coat had not been disturbed. Suppose that both yesterday and earlier today a number of people have been in the room who had an opportunity to take the coat. Then a principle like Lewis's L2 or Pollock's RTP will dictate that if the coat had been taken, it would have been taken today rather than yesterday. Other things being equal, the later the coat is taken the greater the area of perfect match of particular fact. But this is counterintuitive. (In fact, experience teaches that unguarded objects tend to disappear earlier rather than later.) While Lewis's theory is intended to explain why many backtracking conditionals are false, a consequence of the theory is that some very unattractive backtracking conditionals turn out to be true. In fact, this particular problem plagues Lewis's analysis whether the world is determined or probabilistic.

As it is presented, Lewis's account does rely upon miracles. As a result, Lewis in effect treats all counterfactual conditionals as also being counter-legals. This is the feature of his account which most writers have found objectionable. Pollock, Blue, and others place a much higher priority on preservation of all law than on preservation of particular fact no matter how large the divergence of particular fact might be. Given such priorities, and given a deterministic world of the sort Lewis supposes, any change in what happens will result in a world which is different at every moment in the past and every moment in the future. If we adopt such a position, how can we hope to explain the asymmetry between normal and backtracking counterfactual conditionals?

Probably the most sophisticated attempt to deal with these problems within the framework of a non-miraculous analysis of counterfactuals is that developed by John Pollock [1976; 1981]. Pollock has refined his account between 1976 and 1981, but we will try to explain what we take to be his latest position on conditionals and temporal relations. Pollock says that a state of affairs  $P$  has historical antecedents if there is a set of true simple states of affairs  $\Gamma$  such that all times of members of  $\Gamma$  are earlier than the time of  $P$  and  $\Gamma$  nominally implies  $P$ .  $\Gamma$  nominally implies  $P$  just in case  $\Gamma$  together with the set of universal generalisations of material implications corresponding to Pollock's true strong subjunctive generalisations entail  $P$  (or entail a sentence  $\phi$  which is true just in case  $P$  obtains). Pollock next defines a nomic pyramid which is supposed to be a set of states of affairs which contains every historical antecedent of each of its members. Then  $P$

undercuts another state of affairs  $Q$  if and only if for every set  $\Gamma$  of true states of affairs such that  $\Gamma$  is a nomic pyramid and  $Q \in \Gamma$ ,  $\Gamma$  nominally implies that  $P$  does not obtain. In revising his set  $S$  of true simple states of affairs to accommodate a particular counterfactual antecedent  $P$ , Pollock tells us that we are to minimize the deletion of members of  $S$  which are not undercut by  $P$ . (We hope the reader will forgive the vacillation here since Pollock talks about entailment and other logical relations holding between states of affairs where most authors prefer to speak of sentences or propositions.)

Perhaps this procedure will give us the correct results for backtracking and non-backtracking conditionals as Pollock suggests it will if the world is deterministic, but problems arise if we allow the possibility that there may be indeterministic states of affairs which lack historical antecedents. Consider a modified version of an example taken from [Pollock, 1981]. Suppose that protons sometimes emit photons when subjected to a strong magnetic field under a set of circumstances  $C$ , but suppose also that protons never emit photons under circumstances  $C$  if they are not also subjected to a strong magnetic field. As a background condition, let us assume that circumstances  $C$  obtain. Now let  $\phi$  be true just in case a certain proton is subjected to a strong magnetic field at time  $t$  and let  $\psi$  be true just in case the same proton emits a photon shortly after  $t$ . Suppose that both  $\phi$  and  $\psi$  are true. Assuming that no other states of affairs nomologically relevant to  $\psi$  obtain, we would intuitively say that  $\neg\phi > \neg\psi$  is true, i.e. if the proton hadn't been subjected to the magnetic field at  $t$ , then it would not have emitted a photon shortly after  $t$ . But Pollock cannot say this. Since  $\psi$  has no historical antecedents in Pollock's sense, it cannot be undercut by  $\neg\phi$ . Because Pollock does not recognize historical antecedents of states of affairs when the nomological connection involved is merely probable, he must say that  $\neg\phi > \psi$  is true.

Pollock's earlier account, which included the Requirement of Temporal Priority, and Lewis's account with its principle L2, in either its original miraculous formulation or the probabilistic, non-miraculous version, both tend to make objectionable backtracking conditionals true when they are intended to explain why they should be false. Blue [1981] includes a feature in his analysis which produces the same result in much the same way. While Pollock's latest theory of counterfactuals avoids examples like that of the unattended coat, it nevertheless encounters new problems with backtracking conditionals in the context of a probabilistic universe. It makes certain backtracking counterfactuals false which our intuitions say are true while making others true which appear to be false. Yet these are the only positive proposals known to the authors at the time of this writing. Other work in the area such as [Nute, 1980b] and [Post, 1981] is essentially critical. An adequate explanation of the role the temporal order plays in the truth conditions for conditionals is still a very live issue.

### 1.9 Tense

There are relatively few papers among the large literature on conditionals which attempt an account of English sentences which involve both tense and conditional constructions. Two of the earliest are [Thomason and Gupta, 1981] and [Van Fraassen, 1981]. Both of these papers attempt the obvious, a fairly straightforward conjunction of tense and conditional operators within a single formal language. Basic items in the semantics for this language are a set of moments, an earlier-than relation on the set of moments which orders moments into tree-like structures, and an equivalence relation which holds between two moments when they are ‘co-present’. A branch on one of these trees plays the role of a possible world in the semantics. Such a branch is called a history, and sentences of the language are interpreted as having truth values at a moment-history pair, i.e. at a moment in a history. Note that a moment is not a clock time but rather a time-slice belonging to each history that passes through it.

The tense operators in the language include two past-tense operators  $P$  and  $H$ , two future-tense operators  $F$  and  $G$ , and a ‘settledness’ or historical necessity operator  $S$ .  $P\phi$  is true at moment  $i$  in history  $h$  just in case  $\phi$  is true at some moment  $j$  in  $h$  where  $j$  is earlier than  $i$ .  $H\phi$  is true at some moment  $i$  in  $h$  if and only if  $\phi$  is true at  $j$  in  $h$  for every moment  $j$  in  $h$  which is earlier than  $i$ .  $F\phi$  is true at  $i$  in  $h$  if  $\phi$  is true at a moment later than  $i$  in  $h$ , and  $G\phi$  is true at  $i$  in  $h$  if  $\phi$  is true at every moment later than  $i$  in  $h$ .  $S\phi$  is true at  $i$  in  $h$  if and only if  $\phi$  is true at  $i$  in every history  $h'$  which contains  $i$ . For a further discussion of semantics for such tense operators, see Burgess [1984] (Chapter 2.2 of this *Handbook*).

In both of these papers, that part of the semantics which is used to interpret conditionals is patterned after the semantics of Stalnaker. A conditional  $\phi > \psi$  is true at a moment  $i$  in a history  $h$  just in case  $\psi$  is true at the pair  $\langle i', h' \rangle$  at which  $\phi$  is true which is closest or most similar to the pair  $\langle i, h \rangle$ . Much of the discussion in the two papers is devoted to the effort to assure that certain theses which the authors favor are valid in their model theories. The measures needed to ensure some of the desired theses within the context of a Stalnakerian semantics are quite complicated, but the set of theses that represents the most important contribution of the account of [Thomason and Gupta, 1981], namely the doctrine of Past Predominance, turns out to be quite tractable model theoretically.

According to Past Predominance, similarities and differences with respect to the present and past have lexical priority over similarities and differences with respect to the future in any evaluation of how close  $\langle i, h \rangle$  is to  $\langle i', h' \rangle$ , where  $i$  and  $i'$  are co-present moments. This doctrine affects the interaction between the settledness operator  $S$  and the conditional. For example, Past Predominance implies the validity of the following thesis:

$$(\neg S\neg\phi \wedge S\psi) \rightarrow (\phi > \psi).$$

This thesis is clearly operative in the reasoning that leads to the two-box solution to Newcomb's Problem: 'If it's not settled that I won't take both boxes but it is settled that there is a million dollars in the opaque box, then if I take both boxes there will (still) be a million dollars in the opaque box.'<sup>10</sup> Cross [1990b] shows that since, concerning the selection of a closest moment-history pair, Past Predominance places no constraints on what is true at past or future moments, Past Predominance can be formalized and axiomatized in terms of settledness and the conditional using ordinary possible worlds models in which relations of temporal priority between moments are not represented.

The issue of how the conditional interacts with tense operators, such as  $P$ ,  $H$ ,  $F$  and  $G$ , is more problematic. The accounts presented by Thomason and Gupta and by Van Fraassen adopt the hypothesis that English sentences involving both tense and conditional constructions can be adequately represented in a formal language containing a conditional operator and the tense operators mentioned above. Nute [1983] argues that this is a mistake. Consider an example discussed in [Thomason and Gupta, 1981]:

37. If Max missed the train he would have taken the bus.

According to Thomason and Gupta, this and other English sentences of similar grammatical form are of the logical form  $P(\phi > F\psi)$ . Nute argues that this is not true. To see why, consider a second example. Suppose we have a computer that upon request will give us a 'random' integer between 1 and 12. Suppose further that what the computer actually does is increment a certain location in memory by a certain amount every time it performs other operations of certain sorts. When asked to return a random number, it consults this memory location and uses the value stored there in its computation. Thus the 'random' number one gets depends upon when one requests it. We just now left the keyboard to roll a pair of dice. If anyone cares, we rolled a 9. Consider the following conditional:

38. If we had used the computer instead of dice, we would have got a 5 instead of a 9.

It is certainly true that there is a time in the past such that if we had used the computer at that time we would have got a 5, so a sentence corresponding to (38) of the form  $P(\phi > F\psi)$  is certainly true. Yet (38) itself is not true. Depending upon when we used the computer and what operations the computer had performed before we used it, we could have obtained any integer from 1 to 12.

Perhaps we are simply using the wrong combination of operators. Instead of  $P(\phi > F\psi)$ , perhaps sentences like (37) and (38) are of the form  $H(\phi > F\psi)$ . A problem with this suggestion is that such conditionals do

<sup>10</sup>See [Gibbard and Harper, 1981].



not normally concern every time prior to the time at which they are uttered but only certain times or periods of time which are determined by context. Suppose in a football game Walker carries the ball into the end zone for a touchdown. During the course of his run, he came very close to the sideline. Consider the conditional

39. If Walker had stepped on the sideline, he would not have scored.

Can this sentence be of the form  $H(\phi > F\psi)$ ? Surely not, for Walker could have stepped on the sideline many times in the past, and probably did, yet he did score on this particular play. Perhaps we can patch things up further by introducing a new tense operator  $H^*$  which has truth conditions similar to  $H$  except that it only concerns times going a certain distance into the past, the distance to be determined by context. Once again, Nute argues, this will not work. Consider the conditional

40. If Fred had received an invitation, he would have gone to the party.

This sentence might very well be accepted even though Fred would not have gone to the party had he received an invitation five minutes before the party began. The period of time involved does not begin with the present moment and extend back to some past moment determined by context. Indeed if this were the case, for (40) to be true it would even have to be true that Fred would have gone to the party if he had received an invitation after the party ended.

It would seem, then, that if a context-dependent operator is to be the solution to the problem Nute describes, then the contextually determined period of time involved in the truth conditions for English sentences of the sort we have been investigating must be some subset of past times, but one that need not be a continuous interval extending back from the present moment. This is the solution suggested by Thomason [1985].<sup>11</sup> Nute [1991] argues for a different approach: the introduction of a new tensed conditional operator, i.e. an operator which involves in its truth conditions both differences in time and differences in world.

Using a class selection function semantics for this task, we could let our selection function  $f$  pick out for a sentence  $\phi$ , a moment or time  $i$ , and a history or world  $h$  a set  $f(\phi, i, h)$  of pairs  $\langle i', h' \rangle$  of times and histories at which  $\phi$  is true and which are otherwise similar enough to  $\langle i, h \rangle$  for our consideration. We would introduce into our formal language a new conditional operator, say  $\rangle PF \rangle$ , and sentences of the form  $\phi \rangle PF \rangle \psi$  would be true in an appropriate model at  $\langle i, h \rangle$  if and only if for every pair  $\langle i', h' \rangle \in$

<sup>11</sup>The following example may be linguistic evidence for this sort of context-dependence in tensed constructions not involving conditionals: a dean, worried about faculty absenteeism, asks a department chair, 'Was Professor X always in his classroom last term?' the correct answer may be 'Yes' even though Professor X was not in his classroom at times last term when his classes were not scheduled to meet.

$f(\phi, i, h)$  such that there is a time  $j$  in  $h'$  which is copresent with  $i$  and later than  $i'$ ,  $\psi$  is true at  $\langle j, h' \rangle$ . It appears that three more operators of this sort will be needed, together with appropriate truth conditions. These operators may be represented as  $\rangle PP \rangle$ ,  $\rangle FF \rangle$ , and  $\rangle FP \rangle$ . These operators would be used to represent sentences like

41. If Fred had gone to the party, he would have had to have received an invitation.
42. If Fred were to receive an invitation, he would go to the party.
43. If Fred were to go to the party, he would have to have received an invitation.

Notice that (41) and (43) are types of backtracking conditionals. Since such conditionals are rarely true, we may use the operators  $\rangle PP \rangle$  and  $\rangle FP \rangle$  infrequently. This may also account for the clumsiness of the English locution which we must use to clearly express what is intended by (41) and (43).

A number of other interesting problems concerning tense and conditionals occur to us. One of these is the way in which the consequent may affect the times included in the pairs picked by a class selection function. Consider the sentences

44. If he had broken his leg, he would have missed the game.
45. If he had broken his leg, the mend would have shown on his X-ray.

The times at which the leg might have been broken varies in the truth conditions for these two conditionals. This suggests that a semantics like Gabbay's which makes both antecedent and consequent arguments for the class selection function might after all be the preferred semantics. Another possibility is that despite its awkwardness we must introduce some sort of context-dependent tense operator like the operator  $H^*$  discussed earlier. When we represent (44) as  $H^*(\phi > F\psi)$ ,  $H^*$  has the whole of the conditional within its scope and can consider the consequent in determining which times are appropriate. A third possibility is that the consequent does not figure as an argument for the selection function but it does figure as part of the context which determines the selection function which is, in fact, used during a particular piece of discourse. This sort of approach utilizes the concept of conversational score discussed in Section 1.7 of this paper. One piece of evidence in favor of this approach is the fact that it would be unusual to assert both (44) and (45) in the same conversation. Whichever of these two sentences was asserted first, the antecedent of the other would likely be modified in some appropriate way to indicate that a change in the times to be considered was required. Besides these interesting puzzles, we need

also to explain the fact that we maintain the distinction between indicative and subjunctive conditionals involving present and past tense much more carefully than we do where the future tense is concerned. These topics are considered in more detail in [Nute, 1982 and 1991] and [Nute, 1991].

### 1.10 *Other Conditionals*

Besides the subjunctive conditionals we have been considering, we also want an analysis for the might conditionals, the even-if conditionals, and the indicative conditionals mentioned in Section 1.1. It is time we took another look at these important classes of conditionals.

Most authors who discuss the might and the even-if conditional constructions propose that their logical structure can be defined by reference to subjunctive conditionals. Lewis [1973b] and Pollock [1976] suggest that English sentences having the form ‘If  $\phi$  were the case, then  $\psi$  might be the case’ should be symbolized as  $\neg(\phi > \neg\psi)$ . Stalnaker [1981a] presents strong linguistic evidence against this suggestion, but the suggestion has achieved wide acceptance nonetheless.

Pollock [1976] also offers a symbolisation of even-if conditionals. English sentences of the form ‘ $\phi$  even if  $\psi$ ’, he suggests, should be symbolized as  $\phi \wedge (\psi > \phi)$ . The adequacy of this suggestion may depend upon our choice of conditional logic and particularly upon whether we accept the thesis CS. If we accept both CS and Pollock’s proposal, then ‘ $\phi$  even if  $\psi$ ’ will be true whenever both  $\phi$  and  $\psi$  are true. An alternative analysis of even-if conditionals is developed in [Gardenförs, 1979]. Gardenförs’s objection to Pollock’s proposal seems to be that a person who knows that both  $\phi$  and  $\psi$  are true might still reject an assertion of the sentence ‘ $\phi$  even if  $\psi$ ’. Normally, says Gardenförs, one does not assert ‘ $\phi$  even if  $\psi$ ’ when one knows that  $\psi$  is true; an assertion of ‘ $\phi$  even if  $\psi$ ’ presupposes that  $\phi$  is true and  $\psi$  is false. Even when the presupposition that  $\psi$  is false turns out to be incorrect, Gardenförs argues that there is a presumption that the falsity of  $\psi$  would not interfere with the truth of  $\phi$ . Consequently, Gardenförs suggests that ‘ $\phi$  even if  $\psi$ ’ has the same truth conditions as  $(\psi > \phi) \wedge (\neg\psi > \phi)$ . Another suggestion comes from Jonathan Bennett [1982]. Bennett gives a comprehensive account of even-if conditionals, fitting them into the context of uses of ‘even’ that don’t involve ‘if’, and uses of ‘if’ that don’t involve ‘even’. That is, Bennett rejects the treatment of ‘even if’ as an idiom with no internal structure.

The first of three proposals we will consider concerning the analysis of indicative conditionals, which can be found in [Lewis, 1973b; Jackson, 1987] and elsewhere, is that indicative conditionals have the same truth conditions as do material conditionals, paradoxes of implication and problems with Transitivity, Contraposition, and Strengthening Antecedents notwithstanding. It is difficult and perhaps impossible to find really persuasive

counterexamples to Transitivity and Strengthening Antecedents using only indicative conditionals, but apparent counterexamples to Contraposition are easy to construct. Consider, for example, the following two sentences:

46. If it is after 3 o'clock, it is not much after 3 o'clock.

47. If it is much after 3 o'clock, it is not after 3 o'clock.

It is easy to imagine situations in which (46) would be true or appropriate, but are there any situations in which (47) would be true or appropriate? Another problem with this analysis concerns denials of indicative conditionals. Stalnaker [1975] offers an interesting example:

48. If the butler didn't do it, then Fred did it.

Being quite sure that Fred didn't do it, we would deny this conditional. At the same time, we may believe that the butler did it, and therefore when we hear someone say what we would express by

49. Either the butler did it or Fred did it.

We might respond, "Yes, one of them did it, but it wasn't Fred". Yet (48) and (49) are equivalent if (48) has the same truth conditions as the corresponding material conditional.

One possible response to these criticisms is that we must distinguish between the truth conditions for an indicative conditional and the assertion conditions for that conditional. It may be that a conditional is true even though certain conventions make it inappropriate to assert the conditional. This might lead us to say that (47) is true even though it would be inappropriate to assert it. We might also attempt to explain away the paradoxes of implication in this way, relying on the assumed convention that it is misleading and therefore inappropriate to assert a weaker sentence  $\phi$  when we are in a position to assert a stronger sentence  $\psi$  which entails  $\phi$ . For example, it is inappropriate to assert  $\phi \vee \psi$  when one knows that  $\phi$  is true. Just so, the argument goes, it is inappropriate to assert  $\phi \Rightarrow \psi$  when one is in a position to assert either  $\neg\phi$  or  $\psi$ . and in general we may reject other putative counterexamples to the proposal that indicative conditionals have the same truth conditions as material conditionals by saying that in these cases not all the assertion conditions are met for some conditional rather than admit that the truth conditions for the conditional are not met. This line of defence is suggested, for example, by [Grice, 1967; Lewis, 1973b; Lewis, 1976] and by [Clark, 1971].

A second proposal is that indicative conditionals are Stalnaker conditionals, i.e. that Stalnaker's world selection function semantics is the correct semantics for indicative conditionals and Stalnaker's conditional logic **C2** is the proper logic for these conditionals. This suggestion is found in [Stalnaker, 1975] and in [Davis, 1979]. While both Stalnaker and Davis propose

the same model theory for indicative and subjunctive conditionals, both also suggest that the properties of the world selection function appropriate to indicative conditionals are different from those of the world selection function appropriate to subjunctive conditionals.

The difference for Stalnaker has to do with the presuppositions involved in the utterance of the conditional. During the course of a conversation, the participants come to share certain presuppositions. In evaluating an indicative conditional  $\phi \Rightarrow \psi$ , Stalnaker says that we look for the closest  $\phi$ -world at which all of these presuppositions are true. In the case of a subjunctive conditional, on the other hand, we may look outside this ‘context set’ for the closest  $\phi$ -world. Of course the overall closest  $\phi$ -world may not be a world at which all of the presuppositions are true since making  $\phi$  true could tend to make one of the presuppositions false. This means that different worlds may be chosen by the selection function used to evaluate indicative conditionals and the selection function used to evaluate subjunctive conditionals.

While accepting Stalnaker’s model theory for both indicative and subjunctive conditionals, Davis offers a different distinction between the world selection function appropriate to indicative conditionals and that the appropriate to subjunctive conditionals. In fact, Davis claims that Stalnaker’s analysis of subjunctive conditionals is actually the correct analysis of indicative conditionals. To evaluate an indicative conditional  $\phi \rightarrow \psi$ , Davis says we look at the  $\phi$ -world which bears the greatest overall similarity to the actual world to see if it is a  $\psi$ -world. For a subjunctive conditional  $\phi > \psi$ , we look at the  $\phi$ -world which most resembles the actual world up until just before what Davis calls the time of reference of  $\phi$ . Apparently, the time of reference of  $\phi$  is the time at which events reported by  $\phi$  occur, or states of affairs described by  $\phi$  obtain, or etc.

A third proposal, due to Adams [1966; 1975b; 1975a; 1981], holds that indicative conditionals lack truth conditions altogether. They do, however, have probabilities and these probabilities are just the corresponding standard conditional probabilities. Thus  $\text{pr}(\phi \Rightarrow \psi) = \text{pr}(\phi \wedge \psi) / \text{pr}(\phi)$ , at least in those cases where  $\text{pr}(\phi)$  is non-zero. We must remember that Adams does not identify the probability of a conditional with the probability that that conditional is true since he rejects the very notion of truth values for conditionals. Adams proposes that an argument involving indicative conditionals is valid just in case its structure makes it possible to ensure that the probability of the conclusion exceeds any arbitrarily chosen value less than 1 by ensuring that the probabilities of each of the premises exceeds some appropriate value less than 1. In other words, we can push the probability of the conclusion arbitrarily high by pushing the probabilities of the premises suitably high. When an argument is valid in this sense, Adams says that the conclusion of the argument is ‘*p*-entailed’ by its premises and the argument itself is ‘*p*-sound’.

Since Adams rejects truth values for conditionals, conditionals can cer-

tainly not occur as arguments for truth functions. Given his identification of the probability of a conditional with the corresponding standard conditional probability, this further entails that conditionals may not occur within the scope of the conditional operator. Adams attempts to justify this consequence of this theory by suggesting that we don't really understand sentences which involve the embedding of one conditional within another in any case. This claim, though is far from obvious. Such sentences as

50. If this glass will break if it is dropped on the carpet, then it will break if it is dropped on the bare wooden floor.

seem absolutely ordinary and at least as comprehensible as most other indicative conditionals. the inability to handle such conditionals must count as a disadvantage of Adams's theory.

In [Adams, 1977] it is shown that  $p$ -soundness is equivalent to soundness in Lewis's system-of-spheres semantics. This implies that the proper logic for indicative conditionals is the 'first-degree fragment' of Lewis's **VC**. By the first degree fragment of **VC** We mean the set of all those sentences in **VC** within which no conditional operator occurs within the scope of any other operator. Since the logic Adams proposes for indicative conditionals can be supported by a semantics which also allows us to interpret sentences involving iterated conditional operators, we will need very strong reasons to accept Adam's account with its restrictions rather than some possible worlds account like Lewis's.

In fact it may be possible to reconcile Lewis's view that the truth conditions for indicative conditionals are the same as those for the corresponding material conditionals with Adams work on the probabilities of conditionals and  $p$ -entailment. Lewis [1973b] suggests that the truth conditions for  $\phi \Rightarrow \psi$  are given by  $\phi \rightarrow \psi$  while the assertion conditions for  $\phi \Rightarrow \psi$  are given by the corresponding standard conditional probability. Jackson [1987] also entertains such a possibility. If we accept this, then we might accept Adams's theory as a basis for an adequate account of the logic of assertion conditions for indicative conditionals. Since we would be assuming that conditionals have truth values as well as probabilities, we could also overcome the restrictions of Adams's theory and assign probabilities to conditionals which have other conditionals embedded in them. One problem with this approach, though, is that it would seem to require that we identify the probability of a conditional with the probability that the conditional is true. When we do this and also take the probability of a conditional to be the corresponding standard conditional probability, serious problems arise as is shown in [Lewis, 1976] and in [Stalnaker, 1976]. These difficulties will be discussed briefly in Section 3.

## 2 EPISTEMIC CONDITIONALS

The idea that there is an important connection between conditionals and belief change seems to have been inspired by this suggestion of Frank Ramsey's:

If two people are arguing “if  $p$  will  $q$ ?” and are both in doubt as to  $p$ , they are adding  $p$  hypothetically to their stock of knowledge and arguing on that basis about  $q$ .<sup>12</sup>

The issue of how, precisely, to formalize Ramsey's suggestion and extend it from the case where  $p$  is in doubt to the general case has received a great deal of attention—too much attention to permit an exhaustive survey here. We will focus here on the Gärdenfors triviality result for the Ramsey test (see [Gärdenfors, 1986]) and related results, and the implications of these results for the project of formalizing the Ramsey test for conditionals. Despite the narrowness of this topic our discussion will not mention all worthy contributions to the subject.

Sections 2.1 and 2.2 provide a general framework for formalizing belief change and the Ramsey test. Section 2.3 makes connections between this framework and the literature on belief change and the Ramsey test. Section 2.4 presents the Ramsey test itself, and Section 2.5 presents versions of several triviality results found in the literature, including a version of Gärdenfors' 1986 result that subsumes several of the definitions of triviality found in the literature. Section 2.6 examines how triviality can be avoided, and section 2.7 examines systems of conditional logic associated with the Ramsey test. We will provide proofs for some of the results stated below and in other cases refer the reader to the literature.

### 2.1 Languages

By a *Boolean language* we will mean any logical language containing at least the propositional constant ' $\perp$ ', the binary operator ' $\wedge$ ', and the unary operator ' $\neg$ '. We will assume that ' $\top$ ' is defined as ' $\neg\perp$ ' and that any other needed Boolean operators are defined. We do not assume anything at this stage about how the operators and propositional constant of a Boolean language are interpreted, but it will turn out that in most cases ' $\wedge$ ', ' $\neg$ ', and ' $\perp$ ' will receive classical truth-functional interpretations. We will use the symbol ' $\vdash$ ' as a variable ranging over logical inference relations. Effective immediately we will cease using quotes when mentioning formulas and logical symbols.

We define a language (whether Boolean or nonBoolean) to be of type  $\mathcal{L}_0$  iff it contains the propositional constants  $\top$  and  $\perp$  but does not contain the

---

<sup>12</sup>[Ramsey, 1990], p. 155.

binary conditional operator  $>$ . We next define two language-types for doing conditional logic. A language, whether Boolean or nonBoolean, is of type  $\mathcal{L}_1$  iff it contains  $\top$  and  $\perp$  and the only  $>$ -conditionals allowed as formulas are first-degree or “flat” conditionals, i.e. conditionals  $\phi > \psi$  where  $\phi, \psi$  are conditional-free. We define a language, whether Boolean or nonBoolean, to be of type  $\mathcal{L}_2$  (a “full” conditional language) iff it contains  $\top$  and  $\perp$  and allows arbitrary nesting of conditionals in formulas.

## 2.2 A general framework for belief change

We will describe belief change using a framework that is related to the AGM (Alchourrón, Gärdenfors, Makinson) framework for belief revision.<sup>13</sup> Our framework extends that of AGM and is adapted (with further enrichment) from the notion of an *enriched belief revision model* introduced in [Cross, 1990a].

For a given language  $L$  containing  $\top, \perp$  as formulas, let  $\text{WFF}_L$  be the set of all formulas of  $L$  and let  $\mathcal{K}_L$  be  $\mathcal{P}(\text{WFF}_L) \sim \{\emptyset\}$  (where  $\mathcal{P}(\text{WFF}_L)$  is the powerset of  $\text{WFF}_L$ ). For a given inference relation  $\vdash$  and set  $\Gamma$  of formulas of  $L$ , define  $\text{Cn}_\vdash(\Gamma)$  (the  $\vdash$ -consequence set for  $\Gamma$ ) to be  $\{\phi : \Gamma \vdash \phi\}$ , and let  $\mathcal{T}_{L,\vdash} = \{\Gamma : \Gamma \subseteq \text{WFF}_L \text{ and } \text{Cn}_\vdash(\Gamma) = \Gamma\}$  be the set of all theories in  $L$  with respect to  $\vdash$ . A set  $\Gamma$  is  $\vdash$ -consistent iff  $\Gamma \not\vdash \perp$ .

We next define the notion of a *belief change model*:

(DefBCM) A *belief change model* on a language  $L$  containing  $\top, \perp$  as formulas is an ordered septuple

$$\langle \mathbf{K}, \mathcal{I}, \vdash, K_\perp, -, *, s \rangle$$

whose components are as follows:

1.  $\mathbf{K} \subseteq \mathcal{K}_L$  and  $\vdash$  is a subset of  $\mathcal{P}(\text{WFF}_L) \times \text{WFF}_L$ ;
2.  $\mathcal{I}$  and  $K_\perp$  are sets of formulas meeting the following requirements:
  - (a)  $K_\perp \in \mathbf{K}$ ;
  - (b)  $\top, \perp \in \mathcal{I}$ ;
  - (c)  $K_\perp$  is the set of all formulas of  $L$  or a fragment of  $L$ ;
  - (d)  $\mathcal{I}$  is the set of all formulas of  $L$  or a fragment of  $L$ , and  $\mathcal{I} \subseteq K_\perp$ .
  - (e)  $K \subseteq K_\perp$  for all  $K \in \mathbf{K}$ .
3.  $-$  and  $*$  are binary functions mapping each  $K \in \mathbf{K}$  and each  $\phi \in \mathcal{I}$  to sets  $K_\phi^-$  and  $K_\phi^*$ , respectively, where  $K_\phi^- \subseteq K_\perp$  and  $K_\phi^* \subseteq K_\perp$ ;

---

<sup>13</sup>See [Alchourrón *et al.*, 1985] and [Gärdenfors, 1988].



4.  $s$  is a function taking values in  $\mathcal{P}(\text{WFF}_L)$ , where  $\mathbf{K} \subseteq \text{dom}(s) \subseteq \mathcal{P}(\text{WFF}_L)$ .

A *classical belief change model* is a belief change model defined on a Boolean language whose logical consequence relation  $\vdash$  includes all classical truth-functional entailments and respects the deduction theorem for the material conditional. A *deductively closed belief change model* is a belief change model for which  $K = \text{Cn}_{\vdash}(K) \cap K_{\perp}$  and  $K_{\phi}^{-} = \text{Cn}_{\vdash}(K_{\phi}^{-}) \cap K_{\perp}$  and  $K_{\phi}^{*} = \text{Cn}_{\vdash}(K_{\phi}^{*}) \cap K_{\perp}$  for all  $K \in \mathbf{K}$  and all  $\phi \in \mathcal{I}$ . Note that in a deductively closed belief change model on language  $L$ , belief sets are theories in the fragment of  $L$  represented by  $K_{\perp}$  and not necessarily theories in  $L$  itself.

Informally, the items in a belief change model can be described as follows.  $\mathbf{K}$  represents the set of all possible belief states recognized by the model; often  $\mathbf{K}$  will be a subset of  $\mathcal{T}_{L,\vdash}$  but not always.  $\mathcal{I}$  represents the set of all formulas eligible to serve as inputs for contraction and revision.  $\vdash$  is an inference relation defined on  $L$  and will in most cases be an extension of truth-functional propositional logic.  $K_{\perp}$  contains all of the formulas of that fragment of  $L$  from which the belief sets in  $\mathbf{K}$  are constructed and represents the absurd belief state; thus every belief set in  $\mathbf{K}$  is a subset of  $K_{\perp}$ . For each  $K \in \mathbf{K}$  and each  $\phi \in \mathcal{I}$ ,  $K_{\phi}^{-}$  represents the result of *contracting*  $K$  to remove  $\phi$  (if possible), whereas  $K_{\phi}^{*}$  represents the result of *revising*  $K$  to include  $\phi$  as a new belief. Revision is normally assumed to involve not only adding the given formula to the given belief set but also resolving any inconsistencies thereby created. For the sake of generality, we have not stipulated that  $K_{\phi}^{-}, K_{\phi}^{*} \in \mathbf{K}$ , though this will usually be the case. Finally,  $s$  is the *support function* for the model, which determines for each belief state  $K$  (and perhaps for other sets, as well) the set of formulas of  $L$  supported by  $K$ . For belief sets  $K$  in belief change models for which the Ramsey test holds,  $s(K)$  will contain Ramsey test conditionals even if  $K$  does not.

### 2.3 Comparisons

With an eye toward our presentation of the basic triviality result for the Ramsey test we will briefly review differing positions about the elements making up a belief change model. The list of authors we mention here is not exhaustive but constitutes a representative sample of the diversity of positions taken with respect to belief change models and their elements in discussions of the Ramsey test.

#### *Belief states: the language of the model and the set $\mathbf{K}$*

Seegerberg places no restrictions on the language in his discussion of the triviality result in [Seegerberg, 1989]. Gärdenfors ([1988] and elsewhere),

Rott [1989], and Cross [1990a] all adopt a type- $\mathcal{L}_2$  language for their Ramsey test belief change models, whereas Makinson [1990], Morreau [1992], Hansson ([1992], section III), Arló-Costa [1995], and Levi ([1996] and elsewhere) restrict themselves to a type- $\mathcal{L}_1$  language. Hansson, Arló-Costa, and Levi allow only the type- $\mathcal{L}_0$  formulas of a type- $\mathcal{L}_1$  language to belong to the sets that individuate belief states. Makinson and Morreau, like Gärdenfors, Rott, Segerberg, and Cross, do not restrict the membership of belief-state-individuating sets to type- $\mathcal{L}_0$  formulas.

Most authors on the Ramsey test follow Gärdenfors in representing the set of all possible belief states as a set of theories. One exception is Hansson [1992], who takes each possible belief state to be represented by a pair consisting of a set of formulas and a revision operator that defines the dynamic properties of the belief state. For Hansson, the set of formulas in question is a *belief base*, a set of conditional-free formulas that need not be deductively closed. The belief base of a belief state is a (not necessarily finite) axiom set for the belief state, the idea being to allow different belief states to be associated with the same deductively closed theory. A belief state in Hansson's model can still be individuated by means of its belief base, however, because the revision operator of a belief state is a function of that belief state's belief base. Morreau [1992] also gives a two-component analysis of belief states, but in Morreau's analysis the two components are a set of "worlds" (truth-value assignments to atomic formulas) and a selection function (of the Stalnaker-Lewis variety) that determines which conditionals are believed in the belief state. A third exception is Rott [1991], who identifies belief states with epistemic entrenchment relations and notes that a nonabsurd belief set can be recovered from an epistemic entrenchment relation that supports at least one strict entrenchment: the belief set will be the set of all formulas strictly more entrenched than  $\perp$ .

Among those authors who take belief states to be deductively closed theories, most follow Gärdenfors in assuming that not every theory corresponds to a possible belief state. On this issue Segerberg and Makinson are exceptions. In their respective extensions of Gärdenfors' basic triviality result Segerberg and Makinson assume that revision is defined on *all* theories in a given language rather than on a nonempty subset of the set of all theories for that language.<sup>14</sup>

We note above that Hansson, Arló-Costa, and Levi allow only conditional-free formulas into the sets that individuate belief states.<sup>15</sup> Why exclude conditionals from these sets? In Levi's view, the formulas eligible for membership in the theories that individuate belief states are precisely those statements about which agents can be concerned to avoid error. Levi argues against including conditionals in the theories that individuate belief

<sup>14</sup>See [Segerberg, 1989] and [Makinson, 1990].

<sup>15</sup>See, for example, [Hansson, 1992], [Arló-Costa, 1995], [Levi, 1996], and [Arló-Costa and Levi, 1996].

states because in his view conditionals do not have truth conditions or truth values and so are not sentences about which agents can be concerned to avoid error.<sup>16</sup> On Levi's view, a conditional  $\phi > \psi$  in a type- $\mathcal{L}_1$  language is acceptable relative to a belief set  $K$  in a type- $\mathcal{L}_0$  language iff  $\neg\psi$  is not epistemically possible relative to the result of revising  $K$  to include  $\phi$ , and the negated conditional  $\neg(\phi > \psi)$  is acceptable relative to  $K$  iff  $\neg\psi$  is epistemically possible relative to the revision of  $K$  to include  $\phi$ . The part of this view governing negated conditionals, the negative Ramsey test, will be discussed later.<sup>17</sup> An important consequence of Levi's view is the thesis that conditionals are "parasitic" on conditional-free statements in the following sense: the set of conditionals supported by a given belief state is determined by the conditional-free formulas accepted in that belief state or a subset thereof. Hansson [1992] shows, however, that it is possible to motivate a parasitic account of conditionals without taking a position on whether conditionals have truth conditions or truth values.

Gärdenfors [1988] criticizes Levi's view of conditionals on the grounds that it fails to account for iterated conditionals, a species of conditional about which Levi has expressed skepticism, but Levi [1996] and Hansson [1992] show that iterated conditionals can be accounted for (if necessary) even if conditionals do not have truth conditions or truth values. Levi [1996] points out, however, that axiom schema (MP) fails to be valid in the sense he favors if iterated conditionals are allowed.<sup>18</sup> In this connection Levi exploits examples like the following, which was described by McGee [1985] as a counterexample to *modus ponens*:

Opinion polls taken just before the 1980 election showed the Republican Ronald Reagan decisively ahead of the Democrat Jimmy Carter, with the other Republican in the race, John Anderson, a distant third. Those apprised of the poll results believed, with good reason:

If a Republican wins the election, then if it's not Reagan who wins it will be Anderson.  
A Republican will win the election.

Yet they will not have good reason to believe

If it's not Reagan who wins, it will be Anderson.<sup>19</sup>

---

<sup>16</sup>See [Levi, 1988] and [Levi, 1996], for example. As Arló-Costa and Levi [1996] point out, Ramsey agreed that conditionals lack truth conditions and truth values: this is clear from the context of the quote from Ramsey with which we began Section 2.

<sup>17</sup>For the most recent account of Levi's views on this topic, see [Levi, 1996].

<sup>18</sup>See [Levi, 1996], pp. 105-112.

<sup>19</sup>[McGee, 1985], p. 462.

Arló-Costa [1998] embraces iterated conditionals and uses McGee’s example to argue, via the Ramsey test, against the following principle of invariance for iterated *supposition*.

(K\*INV) If  $\phi \in K \neq K_{\perp}$ , then  $(K_{\phi}^*)^*_{\psi} = K^*_{\psi}$ .

Supposition, i.e. hypothetical revision of belief “for the sake of argument,” is the notion of revision that Arló-Costa [1998] and Levi [1996] both associate with Ramsey test conditionals. Since (K\*INV) holds in any deductively closed classical belief change model that satisfies (K\*3) and (K\*4), Arló-Costa takes McGee’s example as evidence against (K\*4) as a principle governing supposition.

*Contraction and revision inputs: the set  $\mathcal{I}$*

Gärdenfors does not exclude conditionals from the class of formulas eligible to be inputs for belief change in the models he formulates, but Morreau, Arló-Costa, and Levi do. In Levi’s case this restriction clearly follows from his view that conditionals have neither truth conditions nor truth values, and Arló-Costa appears to agree with this view. Morreau’s exclusion of conditionals as revision inputs appears to be an artifact of the nontriviality theorem he proves for the Ramsey test ([Morreau, 1992], THEOREM 14, p. 48) rather than indicative of a philosophical position about the status of conditionals.

*Logical consequence and support:  $\vdash$  and  $s$*

Most authors on the Ramsey test follow Gärdenfors in assuming a compact background logic  $\vdash$  that includes all truth functional propositional entailments while respecting the deduction theorem for the material conditional, but there has been research on the Ramsey test in frameworks where the background logic is nonclassical or not necessarily classical. For example, Segerberg’s triviality result in [Segerberg, 1989] assumes only the minimal constraints of Reflexiveness, Transitivity, and Monotony for  $\vdash$ ,<sup>20</sup> and in [Gärdenfors, 1987] Gärdenfors credits Peter Lavers with having established in an unpublished note a triviality result for the Ramsey test in which  $\vdash$  is defined to be *minimal logic*<sup>21</sup> instead of an extension of classical truth-functional logic. Also, Cross and Thomason [1987; 1992] investigate a four-valued system of conditional logic that is motivated by an application of the Ramsey test in the context of the nonmonotonic logic of multiple inheritance with exceptions in semantic networks.

<sup>20</sup>See the definition of a Segerberg belief change model near the end of section 2.3.

<sup>21</sup>Minimal logic has modus ponens as its only inference rule and every instance of the following schemata as axioms:

Levi [1988] introduces the function  $RL$ , which maps a conditional-free belief set to a conditional-laden belief set via the Positive and Negative Ramsey tests. Cross [1990a] formulates a version of the triviality result proved in [Gärdenfors, 1986] in a framework where an extension  $\vdash$  of classical logic is coupled with a not-necessarily-monotonic consequence operation  $cl$ . Makinson [1990] does the same, calling his not-necessarily-monotonic consequence operation  $C$ . Hansson [1992] makes use of a function  $s$  which maps each belief state to the set of all formulas the belief state “supports.” Support functions are also adopted by Arló-Costa [1995] and by Arló-Costa and Levi [1996]. Our view is that Levi’s  $RL$ , Cross’  $cl$ , Makinson’s  $C$ , and Hansson’s  $s$  should be regarded as variations on the same theoretical construct, and we will follow Hansson in calling this construct a *support function* and in using  $s$  to represent it. More on this in Section 2.6 below.

The following postulates are examples of requirements that might be imposed on  $s$ . Assume a belief change model on a language  $L$ , and assume that  $\Gamma$  ranges over  $\text{dom}(s)$ , which always includes  $\mathbf{K}$  as a subset:

(Identity over  $\mathbf{K}$ )  $s(K) = K$  for all  $K \in \mathbf{K}$ .

(Monotonicity over  $\mathbf{K}$ ) For all  $H, K \in \mathbf{K}$ , if  $H \subseteq K$  then  $s(H) \subseteq s(K)$ .

(Reflexivity)  $\Gamma \subseteq s(\Gamma)$ .

(Closure)  $\text{Cn}_\vdash[s(\Gamma)] = s(\Gamma)$ .

(Consistency) If  $\Gamma$  is  $\vdash$ -consistent then  $s(\Gamma)$  is  $\vdash$ -consistent.

(Superclassicality)  $\text{Cn}_\vdash(\Gamma) \subseteq s(\Gamma)$ .

(Transitivity)  $s(\Gamma) = s[s(\Gamma)]$ .

(Reasoning by Cases)  $s(\Gamma \cup \{\phi\}) \cap s(\Gamma \cup \{\neg\phi\}) \subseteq s(\Gamma)$  for all  $\Gamma, \phi$  such that  $\Gamma \cup \{\phi\} \in \text{dom}(s)$  and  $\Gamma \cup \{\neg\phi\} \in \text{dom}(s)$ .

(Conservativeness)  $L$  has type- $\mathcal{L}_0$  fragment  $L_0$  and for all  $\phi \in \text{WFF}_{L_0}$ ,  $\phi \in s(\Gamma)$  iff  $\phi \in \text{Cn}_\vdash(\Gamma)$ .

None of Gärdenfors, Morreau, or Segerberg uses the notion of a support function: they assume, in effect, that  $s(K) = K$  for all  $K \in \mathbf{K}$ .

1.  $(\phi \wedge \psi) \rightarrow \phi$ .
2.  $(\phi \wedge \psi) \rightarrow \psi$ .
3.  $\phi \rightarrow (\phi \vee \psi)$ .
4.  $\psi \rightarrow (\phi \vee \psi)$ .
5.  $(\phi \rightarrow \chi) \rightarrow [(\psi \rightarrow \chi) \rightarrow ((\phi \vee \psi) \rightarrow \chi)]$ .
6.  $(\phi \rightarrow \psi) \rightarrow [(\phi \rightarrow \chi) \rightarrow ((\phi \rightarrow (\psi \wedge \chi)))]$ .
7.  $[\phi \rightarrow (\psi \rightarrow \chi)] \rightarrow [(\phi \rightarrow \psi) \rightarrow (\psi \rightarrow \chi)]$ .
8.  $\phi \rightarrow (\psi \rightarrow \phi)$ .

The formula  $\neg\phi$  is defined to be  $\phi \rightarrow \perp$ . This axiomatization is found in [Segerberg, 1968].

*Contraction and revision: postulates for belief change*

Since  $-$  and  $*$  are to represent functions legitimately describable as contraction and revision, respectively, it is appropriate to consider additional conditions on these functions. Which additional conditions should be imposed is a matter of dispute, and some of the additional postulates that will be under consideration are listed below. In the case of postulates (K<sup>+</sup>1), (K<sup>-</sup>1), (K<sup>-</sup>2), (K<sup>-</sup>3), (K<sup>-</sup>4), (K<sup>-</sup>5), (K<sup>-</sup>6), (K<sup>-</sup>7), (K<sup>-</sup>8), (K\*1), (K\*2), (K\*3), (K\*4), (K\*5), (K\*6), (K\*7), (K\*8), (K\*L), (K\*M), and (K\*P) we follow the labeling used in [Gärdenfors, 1988]. Please note that we have not adopted any of the postulates given below in the definition of *belief change model*. In each postulate, the variable  $K$  is understood to range over  $\mathbf{K}$ ; also  $\phi, \psi$  are understood to range over  $\mathcal{I}$ . We begin with a definition of a third important belief change operation: expansion.

*Definition of and postulate for expansion*

$$\text{(Def+)} \quad K_{\phi}^{+} = \text{Cn}_{\vdash}(K \cup \{\phi\}) \cap K_{\perp}.$$

$$\text{(K}^{+}\text{1)} \quad K_{\phi}^{+} \in \mathbf{K}. \text{ (} K_{\phi}^{+} \text{ is a belief set.)}$$

*Postulates for contraction*

$$\text{(K}^{-}\text{1)} \quad K_{\phi}^{-} \in \mathbf{K}. \text{ (} K_{\phi}^{-} \text{ is a belief set.)}$$

$$\text{(K}^{-}\text{2)} \quad K_{\phi}^{-} \subseteq K.$$

$$\text{(K}^{-}\text{3)} \quad \text{If } \phi \notin K, \text{ then } K_{\phi}^{-} = K.$$

$$\text{(K}^{-}\text{4)} \quad \text{If } \not\vdash \phi, \text{ then } \phi \notin K_{\phi}^{-}.$$

$$\text{(K}^{-}\text{4w)} \quad \text{If } \not\vdash \phi \text{ and } K \neq K_{\perp}, \text{ then } \phi \notin K_{\phi}^{-}.$$

$$\text{(K}^{-}\text{5)} \quad \text{If } \phi \in K, \text{ then } K \subseteq (K_{\phi}^{-})_{\phi}^{+}.$$

$$\text{(K}^{-}\text{6)} \quad \text{If } \vdash \phi \leftrightarrow \psi, \text{ then } K_{\phi}^{-} = K_{\psi}^{-}.$$

$$\text{(K}^{-}\text{7)} \quad K_{\phi}^{-} \cap K_{\psi}^{-} \subseteq K_{\phi \wedge \psi}^{-}.$$

$$\text{(K}^{-}\text{8)} \quad \text{If } \phi \notin K_{\phi \wedge \psi}^{-}, \text{ then } K_{\phi \wedge \psi}^{-} \subseteq K_{\phi}^{-}.$$

*Postulates for revision*

$$\text{(K}^{*}\text{1)} \quad K_{\phi}^{*} \in \mathbf{K}. \text{ (} K_{\phi}^{*} \text{ is a belief set.)}$$

$$\text{(K}^{*}\text{2)} \quad \phi \in K_{\phi}^{*}.$$

- (K\*3)  $K_\phi^* \subseteq K_\phi^+$ .
- (K\*4) If  $\neg\phi \notin K$ , then  $K_\phi^+ \subseteq K_\phi^*$ .
- (K\*4s) If  $\neg\phi \notin K$ , then  $K_\phi^+ = K_\phi^*$ .
- (K\*4ss) If  $K_\phi^+ \neq K_\perp$ , then  $K_\phi^+ = K_\phi^*$ .
- (K\*4w) If  $\phi \in K \neq K_\perp$ , then  $K \subseteq K_\phi^*$ .
- (K\*5)  $K_\phi^* = K_\perp$  iff  $\vdash \neg\phi$ .
- (K\*5w) If  $K_\phi^* = K_\perp$ , then  $\vdash \neg\phi$ .
- (K\*5ws) If  $K_\phi^* = K_\perp$ , then  $\text{Cn}_+(\{\phi\}) = K_\perp$ .
- (K\*C) If  $K \neq K_\perp$  and  $K_\phi^* = K_\perp$ , then  $\vdash \neg\phi$ .
- (K\*6) If  $\vdash \phi \leftrightarrow \psi$ , then  $K_\phi^* = K_\psi^*$ .
- (K\*6s) If  $\psi \in K_\phi^*$  and  $\phi \in K_\psi^*$ , then  $K_\phi^* = K_\psi^*$ .
- (K\*7)  $K_{\phi \wedge \psi}^* \subseteq (K_\phi^*)_\psi^+$ .
- (K\*7')  $K_\phi^* \cap K_\psi^* \subseteq K_{\phi \vee \psi}^*$ .
- (K\*8) If  $\neg\psi \notin K_\phi^*$ , then  $(K_\phi^*)_\psi^+ \subseteq K_{\phi \wedge \psi}^*$ .
- (K\*L) If  $\neg(\phi > \neg\psi) \in K$ , then  $(K_\phi^*)_\psi^+ \subseteq K_{\phi \wedge \psi}^*$ .
- (K\*M) If  $s(K) \subseteq s(K')$ , then  $K_\phi^* \subseteq K_\phi'^*$ .
- (K\*IM) If  $K \neq K_\perp \neq K'$  and  $s(K) \subseteq s(K')$ , then  $K_\phi'^* \subseteq K_\phi^*$ .
- (K\*T) If  $K \neq K_\perp$ , then  $K_\top^* = K$ .
- (K\*P) If  $\neg\phi \notin K$ , then  $K \subseteq K_\phi^*$ .
- (K\*P $\mathcal{I}$ ) If  $\neg\phi \notin K$  then  $K \cap \mathcal{I} \subseteq K_\phi^* \cap \mathcal{I}$ .
- (LI)  $K_\phi^* = (K_{\neg\phi}^-)_\phi^+$ .

A few other postulates will be identified as needed.

Our treatment of contraction and revision is not general enough to include every treatment of contraction and revision as a special case. For example, in the formalization of belief revision in [Morreau, 1992], the revision operation is nondeterministic, i.e. its value for a given belief set  $K$  and proposition  $\phi$  is a set of belief sets rather than a belief set. We will not attempt to formalize nondeterministic contraction or revision. Also, for

Levi, contraction and revision are to be evaluated by means of a measure of *informational value*, which we do not explicitly formalize.

The contraction operation, which we include in every belief change model, is not often used in the presentation of triviality results for the Ramsey test. Exceptions to this pattern include Cross [1990a] and Makinson [1990], who involve contraction explicitly in their respective formulations of triviality results for the Ramsey test.

*A catalog of belief change models*

We conclude our discussion of comparisons by defining several categories of belief change model that illustrate how the framework defined above can be made to reflect the differing assumptions of a subset of authors who have written on belief revision and the Ramsey test. In associating a name with a class of belief change models we do not claim that the person named defined this class of models; rather, we claim that the belief change models associated with this name are the appropriate counterpart in our framework of models that the named person did define in the context of work on the Ramsey test. Note that postulates on contraction and revision are not part of these definitions.

1. By a *Gärdenfors belief change model* (see, for example, [Gärdenfors, 1986], [Gärdenfors, 1987], and [Gärdenfors, 1988]) we will mean a deductively closed classical belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  defined on a type- $\mathcal{L}_2$  language  $L$  where  $\mathcal{I} = \text{WFF}_L = K_{\perp}$ , and  $\text{dom}(s) = \mathbf{K}$ , and  $s$  satisfies Identity over  $\mathbf{K}$ .
2. By a *Seegerberg belief change model* (see [Seegerberg, 1989]) we will mean a belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$ , defined on any language, such that the following hold:  $\mathbf{K} = \mathcal{T}_{L, \vdash}$ ;  $\mathcal{I} = \text{WFF}_L = K_{\perp}$ ;  $\text{dom}(s) = \mathbf{K}$ ;  $s$  satisfies identity over  $\mathbf{K}$ ; and  $\text{Cn}_{\vdash}$  meets the following requirements, for all  $\Delta, \Gamma \subseteq \text{WFF}_L$ :
  - (Reflexivity for  $\vdash$ )  $\Gamma \subseteq \text{Cn}_{\vdash}(\Gamma)$ .
  - (Monotonicity for  $\vdash$ ) If  $\Delta \subseteq \Gamma$ , then  $\text{Cn}_{\vdash}(\Delta) \subseteq \text{Cn}_{\vdash}(\Gamma)$ .
  - (Transitivity for  $\vdash$ )  $\text{Cn}_{\vdash}(\Gamma) = \text{Cn}_{\vdash}[\text{Cn}_{\vdash}(\Gamma)]$ .
3. By a *Makinson belief change model* (see [Makinson, 1990]) we will mean a belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  defined on a type- $\mathcal{L}_1$  language  $L$  and satisfying the following:  $\mathbf{K} = \{\Gamma : \Gamma \subseteq \text{WFF}_L \text{ and } s(\Gamma) = \Gamma\}$ ;  $\mathcal{I} = \text{WFF}_L = K_{\perp}$ ,  $\vdash$  is classical propositional consequence;  $\text{dom}(s) = \mathcal{P}(\text{WFF}_L)$ ; and  $s$  satisfies Superclassicality, Transitivity, and Reasoning by Cases.<sup>22</sup>

---

<sup>22</sup>Note that in a Makinson belief change model  $s$  satisfies both Reflexivity and Closure. Closure holds since Superclassicality and Transitivity for  $s$  imply that for each  $\Gamma \subseteq \text{WFF}_L$ , we have  $s(\Gamma) \subseteq \text{Cn}_{\vdash}[s(\Gamma)] \subseteq s[s(\Gamma)] = s(\Gamma)$ .



4. By a *Morreau belief change model* (see [Morreau, 1992]) we will mean a deductively closed classical belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  defined on a type- $\mathcal{L}_1$  language  $L$  whose type- $\mathcal{L}_0$  fragment is  $L_0$  and where the following hold:  $\mathcal{I} = \text{WFF}_{L_0}$ ;  $K_{\perp} = \text{WFF}_L$ ;  $\text{dom}(s) = \mathbf{K}$ ; and  $s$  satisfies Identity over  $\mathbf{K}$ .
5. By a *Hansson belief change model* (see [Hansson, 1992], section 3) we will mean a classical belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  defined on a type- $\mathcal{L}_1$  language  $L$  whose type- $\mathcal{L}_0$  fragment is  $L_0$  and where the following hold:  $\mathbf{K} \subseteq \mathcal{P}(\text{WFF}_{L_0})$ ;  $\mathcal{I} = \text{WFF}_{L_0} = K_{\perp}$ ;  $\text{dom}(s) = \mathbf{K}$ ; and  $s$  satisfies Reflexivity, Conservativeness, and Closure.
6. By an *Arló-Costa/Levi belief change model* (see [Arló-Costa, 1990], [Arló-Costa, 1995], [Arló-Costa and Levi, 1996], and [Levi, 1996]) we will mean a deductively closed classical belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  defined on a type- $\mathcal{L}_1$  language  $L$  whose type- $\mathcal{L}_0$  fragment is  $L_0$  and where the following hold:  $\mathcal{I} = \text{WFF}_{L_0} = K_{\perp}$ ;  $\text{dom}(s) = \mathbf{K}$ ; and  $s$  satisfies Reflexivity, Conservativeness, and Closure.

As we have already noted, our belief change models do not capture every feature of every belief revision model appearing in the literature on the Ramsey test, and the models we associate with the names of authors in some cases omit some of the structure that these authors include in their own respective accounts of what constitutes a belief revision model. On the other hand, we have stipulated more detail for the models we associate with certain authors than do the authors themselves. For example, none of Gärdenfors, Morreau, or Segerberg uses the notion of a support function  $s$  in the sources cited above, and neither Gärdenfors, nor Makinson, nor Segerberg restricts the applicability of contraction and revision to a subset  $\mathcal{I}$  of the set of formulas of the language on which the model is defined. Finally, as was pointed out earlier, the contraction operation, which we include in every belief change model, is not often discussed in connection with the Ramsey test. In general, the stipulation of extra detail will serve to highlight tacit assumptions and make comparisons easier.

#### 2.4 The Ramsey test for conditionals

Ramsey's original suggestion can be put as follows: if an agent's beliefs entail neither  $\phi$  nor  $\neg\phi$ , then the agent's beliefs support  $\phi > \psi$  iff his or her initial beliefs together with  $\phi$  entail  $\psi$ , i.e.

$$\text{(RTR)} \quad \text{For all } K \in \mathbf{K} \text{ and all } \phi \in \mathcal{I} \text{ such that } \phi, \neg\phi \notin K \text{ and all } \psi \in K_{\perp}, \\ \phi > \psi \in s(K) \text{ iff } \psi \in K_{\phi}^+.$$

This suggestion covers only the case in which the epistemic status of  $\phi$  is undetermined. What about the case in which the agent's initial beliefs

entail  $\phi$  and the case in which the agent's initial beliefs entail  $\neg\phi$ ? Stalnaker [1968] suggests the following rule for evaluating a conditional in the general case:

First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is true.<sup>23</sup>

Stalnaker's proposal handles the general case by substituting the operation of revision for that of expansion in Ramsey's original proposal. In our framework Stalnaker's suggestion amounts to the following:

(RT) For all  $K \in \mathbf{K}$  and all  $\phi \in \mathcal{I}$  and all  $\psi \in K_{\perp}$ ,  $\phi > \psi \in s(K)$  iff  $\psi \in K_{\phi}^*$ .

Revision postulates (K\*3) and (K\*4) jointly entail

(K\*4s) If  $\neg\phi \notin K$  then  $K_{\phi}^+ = K_{\phi}^*$ .

Hence, if (K\*3), (K\*4) are assumed, then (RT) agrees with (RTR) in the case where neither  $\phi$  nor  $\neg\phi$  belongs to  $K$ . That is, if (K\*3) and (K\*4) hold, then (RT) can be considered an extension of Ramsey's original proposal. In [Gärdenfors, 1978] and in later writings Gärdenfors adopts Stalnaker's version of the Ramsey test for type- $\mathcal{L}_2$  languages and assumes, in addition, the following: every formula of a type- $\mathcal{L}_2$  language  $L$  is an eligible input for revision and an eligible member of a belief set, i.e.  $\mathcal{I} = \text{WFF}_L = K_{\perp}$ , and a conditional, like any other formula, is accepted with respect to (supported by) a belief set  $K$  iff it belongs to  $K$ , i.e.  $s(K) = K$  for all  $K \in \mathbf{K}$ .

We have already noted that Levi, in contrast to Gärdenfors, excludes conditionals as revision inputs and as members of belief sets. Levi's view is that the conditional  $\phi > \psi$  in a type- $\mathcal{L}_1$  language expresses the attitude of an agent for whom  $\neg\psi$  is not epistemically possible relative to  $K_{\phi}^*$ , and the negated conditional  $\neg(\phi > \psi)$  expresses the attitude of an agent for whom  $\neg\psi$  is epistemically possible relative to  $K_{\phi}^*$ . Assuming a type- $\mathcal{L}_1$  language  $L$  with type- $\mathcal{L}_0$  fragment  $L_0$ , and assuming that  $\mathbf{K} \subseteq \mathcal{T}_{L_0, \vdash}$  and  $\mathcal{I} = \text{WFF}_{L_0} = K_{\perp}$ , Levi's view amounts in our framework to the conjunction of the following:

(PRTL) For all  $\phi \in \mathcal{I}$  and all  $\psi \in K_{\perp}$  and all  $K \in \mathbf{K}$  such that  $K \neq K_{\perp}$ ,  $\phi > \psi \in s(K)$  iff  $\psi \in K_{\phi}^*$ .

(NRTL) For all  $\phi \in \mathcal{I}$  and all  $\psi \in K_{\perp}$  and all  $K \in \mathbf{K}$  such that  $K \neq K_{\perp}$ ,  $\neg(\phi > \psi) \in s(K)$  iff  $\psi \notin K_{\phi}^*$ .

---

<sup>23</sup>[Stalnaker, 1968], p. 44. (The page reference is to [Harper *et al.*, 1981], where [Stalnaker, 1968] is reprinted.)

Note that in both (PRTL) and (NRTL), unlike in (RT),  $K$  is restricted to  $\vdash$ -consistent members of  $\mathbf{K}$ . Note also that the adoption of (RT) (or of (PRTL) without (NRTL)) places no constraints on how negated conditionals are related to belief change.

Other versions of the Ramsey test appearing in the literature include the following, due to Hans Rott, who, like Gärdenfors, assumes a language  $L$  of type  $\mathcal{L}_2$  and no restrictions on which formulas can appear as members of belief sets or as revision inputs (i.e.  $\mathcal{I} = \text{WFF}_L = K_\perp$ ):

- (R1) For all  $K \in \mathbf{K}$  and all  $\phi \in \mathcal{I}$  and all  $\psi \in K_\perp$ ,  $\phi > \psi \in K$  iff  $\psi \in K_\phi^*$  and  $\psi \notin K$ .
- (R2) For all  $K \in \mathbf{K}$  and all  $\phi \in \mathcal{I}$  and all  $\psi \in K_\perp$ ,  $\phi > \psi \in K$  iff  $\psi \in K_\phi^*$  and  $\psi \notin K_{\neg\phi}^*$ .
- (R3) For all  $K \in \mathbf{K}$  and all  $\phi \in \mathcal{I}$  and all  $\psi \in K_\perp$ ,  $\phi > \psi \in K$  iff  $\psi \in (K_\psi^-)_\phi^*$ .

Here we follow the labeling in [Gärdenfors, 1987]. The interest of (R1)-(R3) stems in part from the fact that whereas (RT) can be used with (K\*3) and (K\*4) to derive the following thesis (U), none of (R1)-(R3) can be so used:

- (U) If  $\phi \in K$  and  $\psi \in K$ , then  $\phi > \psi \in K$ .

Thesis (U) is related to the strong centering axiom CS of  $\mathbf{VC}$ , and Rott [1986] suggests that (U) should be rejected. Since none of (R1)-(R3) entails (K\*M), one of the assumptions of Gärdenfors' 1986 triviality result for the Ramsey test, (R1)-(R3) might seem worth investigating as alternatives to (RT), but Gärdenfors [1987] shows that (R1)-(R3) do not avoid the problem faced by (RT). Consider the Weak Ramsey Test:

- (WRT) For all  $K \in \mathbf{K}$  and all  $\phi \in \mathcal{I}$  and all  $\psi \in K_\perp$  such that  $\phi \vee \psi \notin K$ ,  $\phi > \psi \in K$  iff  $\psi \in K_\phi^*$ .

Each of (R1)-(R3) entails (WRT), and Gärdenfors [1987] proves a triviality result that holds for any version of the Ramsey test which entails (WRT), including (R1)-(R3) and (RT).<sup>24</sup>

## 2.5 Triviality results for the Ramsey test

### *The basic result*

Many versions of the basic triviality result for the Ramsey test have appeared in the literature, all of them variations on the result proved by Gärdenfors [1986]. All proofs of the basic triviality result we know of exploit the same maneuver, however, one which Hansson [1992] makes explicit:

<sup>24</sup>See also [Gärdenfors, 1988], Chapter 7, Corollary 7.15

in terms of our framework, the finding of *forking support sets* within a belief change model.

(DefFORK) A belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  will be said to *contain forking support sets* iff there exist  $H, J, K \in \mathbf{K}$ , such that  $H = \text{Cn}_{\vdash}(H) \cap K_{\perp}$ , and  $J = \text{Cn}_{\vdash}(J) \cap K_{\perp}$ , and  $K = \text{Cn}_{\vdash}(K) \cap K_{\perp} \neq K_{\perp}$ , and  $H \cap \mathcal{I} \not\subseteq J$ , and  $J \cap \mathcal{I} \not\subseteq H$ , and  $s(H) \subseteq s(K)$ , and  $s(J) \subseteq s(K)$ .

For Gärdenfors and Segerberg belief change models this condition can be stated in the form in which Hansson originally formulated it:

PROPOSITION 1. *A Gärdenfors or Segerberg belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  contains forking support sets iff there exist  $H, J, K \in \mathbf{K}$ , where  $H, J \subseteq K \neq K_{\perp}$ ,  $H \not\subseteq J$ , and  $J \not\subseteq H$ .*

This proposition follows from the fact that in Gärdenfors and Segerberg belief change models (i)  $s(K) = K = \text{Cn}_{\vdash}(K)$  for all  $K \in \mathbf{K}$ , and (ii)  $\mathcal{I}$  and  $K_{\perp}$  both exhaust the formulas of the language of the model.

Next we present the main lemmas for the basic triviality result:

LEMMA 2. *If (RT) holds in a belief change model, then so does  $(K^*M)$ .*

**Proof.** Trivial; left to reader. ■

Postulate  $(K^*M)$  is a postulate of monotonicity for belief revision. We discuss Gärdenfors' argument against  $(K^*M)$  in Section 2.6 below.

LEMMA 3. *No classical belief change model containing forking support sets satisfies  $(K^*2)$ ,  $(K^*C)$ ,  $(K^*P)$ , and  $(K^*M)$ .*

**Proof.** Assume for *reductio* that  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  is a classical belief change model that contains forking support sets and satisfies  $(K^*2)$ ,  $(K^*C)$ ,  $(K^*P)$ , and  $(K^*M)$ . For clarity, we follow the example of [Rott, 1989] in numbering the steps in the *reductio* argument.

- (1)  $H = \text{Cn}_{\vdash}(H) \cap K_{\perp}$ ,  $J = \text{Cn}_{\vdash}(J) \cap K_{\perp}$ ,  
 $K = \text{Cn}_{\vdash}(K) \cap K_{\perp} \neq K_{\perp}$ ,  $H \cap \mathcal{I} \not\subseteq J$ ,  
 $J \cap \mathcal{I} \not\subseteq H$ , and  $s(H), s(J) \subseteq s(K)$ ,  
 for some  $H, J, K \in \mathbf{K}$  (DefFORK)
- (2)  $\phi \in (H \cap \mathcal{I}) \sim J$ , for some  $\phi$  (1)
- (3)  $\psi \in (J \cap \mathcal{I}) \sim H$ , for some  $\psi$  (1)
- (4)  $\neg(\phi \wedge \psi) \in \mathcal{I}$  (2), (3), (DefBCM)
- (5)  $\neg\neg(\phi \wedge \psi) \notin H$  (3), classicality of  $\vdash$ , fact that  $H = \text{Cn}_{\vdash}(H) \cap K_{\perp}$
- (6)  $H \subseteq H^*_{\neg(\phi \wedge \psi)}$  (4), (5),  $(K^*P)$

- |      |   |  |
|------|---|--|
| (7)  | $\phi \in H_{\neg(\phi \wedge \psi)}^*$   | (2), (6)   |
| (8)  | $\neg\neg(\phi \wedge \psi) \notin J$   | (2), classicality of $\vdash$ , fact that $J = \text{Cn}_{\vdash}(J) \cap K_{\perp}$           |
| (9)  | $J \subseteq J_{\neg(\phi \wedge \psi)}^*$  | (4), (8), (K*P)  |
| (10) | $\psi \in J_{\neg(\phi \wedge \psi)}^*$   | (3), (9)   |
| (11) | $H_{\neg(\phi \wedge \psi)}^*, J_{\neg(\phi \wedge \psi)}^* \subseteq K_{\neg(\phi \wedge \psi)}^*$ | (1), (4), (K*M)  |
| (12) | $\phi, \psi \in K_{\neg(\phi \wedge \psi)}^*$   | (7), (10), (11)  |
| (13) | $\neg(\phi \wedge \psi) \in K_{\neg(\phi \wedge \psi)}^*$   | (4), (K*2)   |
| (14) | $K_{\neg(\phi \wedge \psi)}^*$ is $\vdash$ -inconsistent  | (12), (13), classicality of $\vdash$   |
| (15) | $K$ is $\vdash$ -consistent   | classicality of $\vdash$ , fact that $K = \text{Cn}_{\vdash}(K) \cap K_{\perp} \neq K_{\perp}$ |
| (16) | $\vdash \neg\neg(\phi \wedge \psi)$   | (14), (15), (K*C)  |
| (17) | $\not\vdash \neg\neg(\phi \wedge \psi)$   | (5), classicality of $\vdash$ , fact that $H = \text{Cn}_{\vdash}(H) \cap K_{\perp}$           |

Since (17) contradicts (16), this completes the proof.  $\blacksquare$

Lemmas 2 and 3 suffice to prove the following:

**THEOREM 4.** *No classical belief change model defined on a language of type  $\mathcal{L}_1$  or type  $\mathcal{L}_2$  and containing forking support sets satisfies (K\*2), (K\*C), (K\*P), and (RT).*

Note that we have not assumed that  $\mathbf{K}$  is a set of theories either in the language of the model or in the fragment thereof represented by  $K_{\perp}$ . We have not even assumed (K\*1): that the sets produced by revision always belong to  $\mathbf{K}$ . It is however required that the belief sets  $H$ ,  $J$ , and  $K$  used in the proof be theories in the fragment of the language represented by  $K_{\perp}$ .

Do we have a triviality result? Not yet: we do not yet have a criterion of triviality. The following criteria have appeared in the literature:

1. A belief change model is *Gärdenfors nontrivial* iff there is a  $K' \in \mathbf{K}$  and  $\phi, \psi, \chi \in \mathcal{I}$  such that  $\neg\phi, \neg\psi, \neg\chi \notin \text{Cn}_{\vdash}(K')$  and  $\vdash \neg(\phi \wedge \psi)$  and  $\vdash \neg(\phi \wedge \chi)$  and  $\vdash \neg(\psi \wedge \chi)$ .
2. A belief change model is *Rott nontrivial* iff there is a  $K' \in \mathbf{K}$  and  $\phi, \psi \in \mathcal{I}$  such that  $\phi \not\vdash \psi$  and  $\psi \not\vdash \phi$  and  $\phi \vee \psi, \neg\phi \vee \psi, \phi \vee \neg\psi, \neg\phi \vee \neg\psi \notin \text{Cn}_{\vdash}(K')$ .
3. A belief change model is *Seegerberg nontrivial* iff there exist  $\phi, \psi, \chi \in \mathcal{I}$  such that  $\phi \not\vdash \psi$  and  $\psi \not\vdash \phi$  and  $\text{Cn}_{\vdash}\{\phi, \psi, \chi\} = K_{\perp}$  and  $\text{Cn}_{\vdash}(\{\phi\}), \text{Cn}_{\vdash}(\{\psi\}), \text{Cn}_{\vdash}(\{\phi, \psi\}) \in \mathbf{K}$ .

Recall that a support function  $s$  is *monotone over  $\mathbf{K}$*  iff  $s(H) \subseteq s(K)$  for all  $H, K \in \mathbf{K}$  such that  $H \subseteq K$ . A support function can be monotone

over  $\mathbf{K}$  even if it is a *nonmonotonic* consequence operation provided that  $\mathbf{K}$  does not exhaust  $\text{dom}(s)$ . For example,  $s$  is monotone over  $\mathbf{K}$  (but not necessarily over  $\text{dom}(s)$ ) in all Makinson belief change models, since in a Makinson belief change model  $s(K) = K$  for all  $K \in \mathbf{K}$ . Recall that the operation of expansion is defined by (Def+); it turns out that if  $(K^+1)$  and the monotonicity of  $s$  over  $\mathbf{K}$  are assumed, then nontriviality by any of the above criteria will imply the existence of forking support sets:

**LEMMA 5.** *A classical belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  contains forking support sets if it satisfies  $(K^+1)$  and its support function is monotone over  $\mathbf{K}$  and it is Gärdenfors nontrivial.<sup>25</sup>*

**Proof.** Suppose that the model is Gärdenfors nontrivial; we will show that it contains forking support sets. Let  $K'$ ,  $\phi$ ,  $\psi$ , and  $\chi$  be as in the definition of Gärdenfors nontriviality; also, let  $H = K'_{\phi \vee \psi}^+$ ; let  $J = K'_{\phi \vee \chi}^+$ ; and let  $K = K'_{\phi}^+$ . Then by (Def+) and the classicality of  $\vdash$ ,  $H = \text{Cn}_{\vdash}(H) \cap K_{\perp}$ ,  $J = \text{Cn}_{\vdash}(J) \cap K_{\perp}$ , and  $K = \text{Cn}_{\vdash}(K) \cap K_{\perp} \neq K_{\perp}$ . (Def+) and the classicality of  $\vdash$  also imply that  $H, J \subseteq K$ , hence by the monotonicity of  $s$  we have that  $s(H), s(J) \subseteq s(K)$ .  $H \cap \mathcal{I} \not\subseteq J$  holds because  $\phi \vee \psi \in (H \cap \mathcal{I}) \sim J$ ;  $J \cap \mathcal{I} \not\subseteq H$  holds because  $\phi \vee \chi \in (J \cap \mathcal{I}) \sim H$ . ■

**LEMMA 6.** *A classical belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  contains forking support sets if it satisfies  $(K^+1)$  and its support function is monotone over  $\mathbf{K}$  and it is Rott nontrivial.<sup>26</sup>*

**Proof.** Like the proof of Lemma 5, but let  $H = K'_{\phi \vee \psi}^+$ ; let  $J = K'_{\phi \vee \neg \psi}^+$ , let  $K = K'_{\neg \phi}^+$ , where  $K'$ ,  $\phi$ , and  $\psi$  are as in the definition of Rott nontriviality.  $H \cap \mathcal{I} \not\subseteq J$  holds because  $\phi \vee \psi \in (H \cap \mathcal{I}) \sim J$ ;  $J \cap \mathcal{I} \not\subseteq H$  holds because  $\phi \vee \neg \psi \in (J \cap \mathcal{I}) \sim H$ . ■

**LEMMA 7.** *A classical belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  contains forking support sets if it satisfies  $(K^+1)$  and its support function is monotone over  $\mathbf{K}$  and it is Segerberg nontrivial.*

**Proof.** Like the proof of Lemma 5, but let  $H = \text{Cn}_{\vdash}(\{\phi\})$ ,  $J = \text{Cn}_{\vdash}(\{\psi\})$ ,  $K = \text{Cn}_{\vdash}(\{\phi, \psi\})$ , where  $\phi$ ,  $\psi$ , and  $\chi$  are as in the definition of Segerberg nontriviality. ■

Theorem 4 and Lemmas 5, 6, and 7 immediately imply Theorem 8, the basic triviality result for the Ramsey test:

**THEOREM 8.** *No classical belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  that satisfies  $(K^+1)$ ,  $(K^*2)$ ,  $(K^*C)$ ,  $(K^*P)$ , and  $(RT)$  and*

<sup>25</sup>See [Gärdenfors, 1986].

<sup>26</sup>See [Rott, 1989].

whose support function is monotonic over  $\mathbf{K}$  is Gärdenfors nontrivial or Rott nontrivial or Segerberg nontrivial.

The basic result of [Gärdenfors, 1986] can be derived by applying Theorem 8 to Gärdenfors belief change models.

Gärdenfors [1987; 1988] notes that  $(K^*P)$  and  $(K^*2)$  can be replaced by  $(K^*4)$  in the triviality result he proves there, and this same replacement can be made in Theorem 8, with a corresponding change in Lemma 3 and its proof.<sup>27</sup> As was mentioned in Section 2.3 above, Segerberg has proved a version of the Gärdenfors result in which the constraints on  $\vdash$  are limited to Reflexivity, Transitivity, and Monotonicity. The counterpart in our framework of Segerberg's result is the following:

**THEOREM 9.** *No Segerberg nontrivial Segerberg belief change model satisfies  $(K^*M)$ ,  $(K^*4ss)$ , and  $(K^*5ws)$ .*<sup>28</sup>

If contraction and revision are assumed to be related by the Levi Identity (LI) in a deductively closed belief change model, then triviality results for the Ramsey test can be formulated in terms of contraction rather than in terms of revision. In particular, we have the following as a corollary of Theorem 8:<sup>29</sup>

**THEOREM 10.** *No deductively closed classical belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  that satisfies  $(K^+1)$ ,  $(K^-3)$ ,  $(K^-4w)$ , (LI), and (RT) and whose support function is monotonic over  $\mathbf{K}$  is Gärdenfors nontrivial or Rott nontrivial or Segerberg nontrivial.*

**Proof.** It suffices to note that where  $\vdash$  is classical, we have the following: (Def+) and (LI) jointly imply  $(K^*2)$ ; (LI) and  $(K^-4w)$  jointly imply  $(K^*C)$ ; (Def+),  $(K^-3)$ , and (LI) jointly imply  $(K^*P)$ . ■

Makinson [1990] proves a variant of Theorem 10 for type- $\mathcal{L}_1$  languages that replaces  $(K^-4w)$  and weakens both (RT) and (LI) while making stronger assumptions about  $s$  than merely that it is monotone over  $\mathbf{K}$ :

**THEOREM 11.** *Let  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_\perp, -, *, s \rangle$  be a Makinson belief change model defined on a language  $L$  of type  $\mathcal{L}_1$ . Define postulates (RTM),  $(K^-4c)$ , and (MI) as follows:*

(RTM) *For all  $\phi, \psi \in \text{WFF}_{L_0}$ , where  $L_0$  is the type- $\mathcal{L}_0$  fragment of  $L$ ,*  
 $\phi > \psi \in s(K)$  *iff*  $\psi \in K_\phi^*$ .

<sup>27</sup>Steps (6), (9), and (13) must be differently justified.

<sup>28</sup>See [Segerberg, 1989]. Segerberg's version of the Gärdenfors triviality result makes no assumption about which operators are available in the language, hence  $\chi$  is used in the definition of Segerberg nontriviality to play the role that  $\neg(\phi \wedge \psi)$  plays in the proof of Lemma 3. Also, Segerberg's result does not assume that the language contains both  $\top$  and  $\perp$ , which we assume here in (DefBCM).

<sup>29</sup>A similar result is proved in [Cross, 1990a].

(K<sup>-</sup>4c) If  $\phi \in s(K_\phi^-)$ , then  $\phi \in s(\emptyset)$ .

(MI)  $K_{-\phi}^- \subseteq K_\phi^* \subseteq s(K_{-\phi}^- \cup \{\phi\})$ .

Then we have the following:

(1) *Limiting Case.* If (RTM), (K<sup>-</sup>4c), and (MI) hold for  $K = K_\perp$ , then the model is trivial in the sense that  $s(\emptyset) = K_\perp$ .

(2) *Principal Case.* If (RTM), (K<sup>-</sup>3), (K<sup>-</sup>4c), and (MI) hold for all  $K \in \mathbf{K}$  such that  $K \neq K_\perp$ , then the model is trivial in the sense that there are no conditional-free formulas  $\phi$  and  $\psi$  of  $L$  such that  $\phi \wedge \psi \notin s(\emptyset)$  and  $\phi \notin s(\{\psi\})$  and  $\psi \notin s(\{\phi\})$  and  $s[s(\{\phi\}) \cup s(\{\psi\})] \neq K_\perp$ .

The Limiting Case generalizes Theorem 12 discussed below. It is the Principal Case that more closely corresponds to Theorem 10. (K<sup>-</sup>4c) neither entails nor is entailed by (K<sup>-</sup>4), its AGM counterpart, but (MI), which we will refer to as Makinson’s Inequality, is the result of weakening (LI), the Levi Identity, to say that a revision of  $K$  to include  $\phi$  must lie “between”  $K_{-\phi}^-$  and  $s(K_{-\phi}^- \cup \{\phi\})$ . Since, as we have seen, (LI), (Def+), and (K<sup>-</sup>3) entail (K\*P), one might expect that replacing (LI) with (MI) would leave (K\*P) unsupported, but this is not the case: (MI), (Def+), and (K<sup>-</sup>3) already entail (K\*P). Making up for the fact that (MI) is weaker than (LI) are Makinson’s strengthened assumptions about  $s$ : that it satisfies Superclassicality, Transitivity, and Reasoning by Cases. Makinson [1990] points out that these conditions are known not to imply that  $s$  is monotone over its entire domain ( $\mathcal{P}(\text{WFF}_L)$ ), but since contraction and revision in a Makinson belief change model are defined only on  $K$  such that  $s(K) = K$ ,  $s$  is nevertheless monotone “where it counts”, namely over the set  $\mathbf{K}$  of belief sets on which contraction and revision are defined. Several authors (e.g. Grahne [1991], Hansson [1992], and Morreau [1992]) have concluded from Makinson’s result that nonmonotonic consequence does not provide a way out of the Gärdenfors triviality result. In fact, adopting a nonmonotonic consequence operation *does* provide a way out, provided that this consequence relation plays the role of a support function  $s$  that is nonmonotonic over the belief sets to which contraction and revision are applied. Indeed, it is by adopting such support functions that Hansson, Arló-Costa and Levi are able to make the Ramsey test nontrivial, though these authors do not describe the support function as a consequence operation. (See also Section 2.6 below.)

Theorem 8 and its variants pose a dilemma: which of an inconsistent set of constraints on belief change models should be rejected? We return to this later in Section 2.6 below.



*The problem with (K\*5w)*

In the version of Theorem 8 that Gärdenfors proves in [Gärdenfors, 1988], postulate (K\*C) is replaced by the stronger (K\*5w), but Arló-Costa [1990] proves that (K\*5w) faces problems that have nothing to do with (K\*P). Arló-Costa's result, which is not so much a triviality result as an impossibility result, can be formulated as follows in our framework:

**THEOREM 12.** *There is no Gärdenfors belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  in which  $\not\vdash \perp$ , (K\*5w), and (RT) hold.*

Whereas Gärdenfors' results against the Ramsey test exploit the fact that (RT) entails (K\*M), Arló-Costa's result exploits the fact that (RT) entails the following, which Arló-Costa calls "Unsuccess":

(US) If  $K = K_\perp$ , then  $K_\phi^* = K_\perp$ .

In other words, the Ramsey test requires revision into inconsistency if the initial belief state is already inconsistent, regardless whether the revision input is a consistent proposition. Contrary to this, (K\*5w) prohibits revision into inconsistency when the revision input is a consistent proposition, regardless whether the initial belief state is consistent. The labeling of (US) as a postulate of "unsuccess" is appropriate since (K\*5w), which (US) contradicts, follows from (Def+), (LI), and the Postulate of Success for *contraction* (K<sup>-</sup>4).

Arló-Costa's result can be strengthened to include belief change models in which  $s(K) = K$  does not always hold:

**THEOREM 13.** *There is no deductively closed classical belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  whose support function satisfies Reflexivity and Closure and in which  $\not\vdash \perp$ , (K\*5w), and (RT) hold.*

**Proof.** Let a deductively closed classical belief change model on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  be given and suppose for reductio that  $\not\vdash \perp$ , that  $s$  satisfies Reflexivity and Closure, and that the model satisfies (K\*5w) and (RT).

First we prove (US). Let  $K = K_\perp$ ; then we have

$$\begin{aligned} K &\subseteq s(K) && \text{Reflexivity of } s \\ &= \text{Cn}_+[s(K)] && \text{Closure of } s \end{aligned}$$

Since  $\perp \in K_\perp = K$  we have  $s(K) = \text{WFF}_L$ , by the classicality of  $\vdash$ . Next, let  $\psi \in K_\perp$ , and let  $\phi \in \mathcal{I}$ . Since  $s(K) = \text{WFF}_L$ , we have  $\phi > \psi \in s(K)$ . Hence by (RT) we have  $B \in K_\phi^*$ . Thus  $K_\perp \subseteq K_\phi^*$ ; the converse inclusion holds by (DefBCM), so  $K_\phi^* = K_\perp$ , as required to show (US).

By (DefBCM)  $K_\perp \in \mathbf{K}$  and  $\neg\perp \in \mathcal{I}$ ; by (US) we have  $(K_\perp)_{\neg\perp}^* = K_\perp$ . By hypothesis  $\not\vdash \perp$ , hence by the classicality of  $\vdash$  we have not only  $(K_\perp)_{\neg\perp}^* = K_\perp$  but also  $\not\vdash \neg\neg\perp$ , which contradicts (K\*5w). ■

The Limiting Case of Theorem 11, like Theorem 13, is a strengthening of Theorem 12. The problem posed by Theorems 12 and 13 can be solved by retreating from  $(K^*5w)$  to something weaker, such as the postulate  $(K^*C)$  mentioned in Theorem 8, or by restricting the applicability of the Ramsey test, as Aró-Costa and Levi both do by adopting (PRTL) (see Section 2.4 above), which eliminates  $K_{\perp}$  from the domain of belief sets to which the Ramsey test can be applied.

*The negative Ramsey test*

Levi has argued (see, for example, [Levi, 1988] and [Levi, 1996]) that a negated conditional  $\neg(\phi > \psi)$  expresses the propositional attitude of an agent for whom  $\neg\psi$  is a serious (i.e. epistemic) possibility relative to  $K_{\phi}^*$ . Abstracting from Levi’s requirements on what is allowed to be a revision input, the result is this thesis, the negative Ramsey test:

(NRTL) For all  $\phi \in \mathcal{I}$  and all  $\psi \in K_{\perp}$  and all  $K \in \mathbf{K}$  such that  $K \neq K_{\perp}$ ,  
 $\neg(\phi > \psi) \in s(K)$  iff  $\psi \notin K_{\phi}^*$ .

Rott [1989] takes the view that adopting both the negative Ramsey test and the Ramsey test amounts to an assumption of autoepistemic omniscience. Given the view of Gärdenfors, Rott, and others that conditionals and negated conditionals belong in belief sets along with other beliefs (so that  $s$  satisfies Identity over  $\mathbf{K}$ ), the conjunction of (RT) and (NRTL) does amount to a kind of epistemic omniscience. That is, if  $s$  satisfies Identity over  $\mathbf{K}$ , then “closing” each belief set under (RT) and (NRTL) amounts to an idealization that parallels the idealization represented by “closing” each belief set under  $\vdash$ . On Levi’s view, conditionals do not express propositions and so are not objects of belief, thus on Levi’s view the positive and negative Ramsey tests cannot be said to represent an idealization concerning what *beliefs* an agent holds. For Levi, what the positive and negative Ramsey tests represent is not a pair of closure conditions on the unary propositional attitude of belief but rather a definition of a binary propositional attitude toward the antecedent and consequent of a conditional that an agent is said to ‘accept’.

Regardless how the issue of autoepistemic omniscience is resolved, the adoption of (NRTL) has consequences. Gärdenfors, Lindström, Morreau, and Rabinowicz [1991] prove what they consider to be a triviality result for (NRTL) with assumptions weaker than those needed for Gärdenfors’ 1986 triviality result for (RT); in particular,  $(K^*P)$  is not needed. In our framework their result is equivalent to the following:

**THEOREM 14.** *If  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  is a belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  for which both (NRTL) and  $(K^*T)$  hold and for which  $s$  satisfies Identity over  $\mathbf{K}$ , then there are no  $K, K' \in \mathbf{K}$  such that  $K \neq K_{\perp} \neq K'$  and  $K' \neq K \subseteq K'$ .*

The latter can be derived as a corollary of the following stronger result:

**THEOREM 15.** *If  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  is a belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  for which both (NRTL) and (K\*T) hold and for which  $s$  is monotone over  $\mathbf{K}$ , then there are no  $K, K' \in \mathbf{K}$  such that  $K \neq K_{\perp} \neq K'$  and  $K' \neq K \subseteq K'$ .*

**Proof.** Suppose that  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  is a belief change model defined on a language of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  such that  $s$  is monotone over  $\mathbf{K}$ .

First we prove that (NRTL) implies (K\*IM): assume (NRTL) and suppose that  $K, K' \in \mathbf{K}$  and  $\phi \in \mathcal{I}$  and  $K \neq K_{\perp} \neq K'$  and  $s(K) \subseteq s(K')$ , and let  $\psi \notin K_{\phi}^*$ . Then by (NRTL)  $\neg(\phi > \psi) \in s(K)$ , hence  $\neg(\phi > \psi) \in s(K')$ . By (NRTL) it follows that  $\psi \notin K'_{\phi}*$ , as required to establish (K\*IM).

Now suppose for reductio that  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  satisfies both (NRTL) and (K\*T), that  $s$  is monotone over  $\mathbf{K}$ , and there are  $K, K' \in \mathbf{K}$  such that  $K \neq K_{\perp} \neq K'$  and  $K' \neq K \subseteq K'$ . Since  $K \subseteq K'$  we have  $s(K) \subseteq s(K')$  by the monotonicity of  $s$ . By (DefBCM) we know that  $\top \in \mathcal{I}$ , so we have  $K'_{\top} \subseteq K_{\top}$  by (K\*IM). But  $K_{\top} = K$  and  $K'_{\top} = K'$  by (K\*T), hence  $K' \subseteq K$ , which contradicts  $K' \neq K \subseteq K'$ , completing the reductio.  $\blacksquare$

Note that neither theorem assumes that all members of  $\mathbf{K}$  must be deductively closed, nor does either result include any assumption about  $\vdash$ . In [Gärdenfors *et al.*, 1991] Theorem 14 is presented as a triviality result because the authors maintain that a model of belief change is trivial if it contains no consistent, conditional-laden belief sets  $K, K'$  such that  $K$  is a proper subset of  $K'$ . As Rott [1989], Morreau [1992], and Hansson [1992] point out, however, it is a substantive (and, they argue, mistaken) assumption to hold that principles of belief revision that are justified in the context of conditional-free belief sets (e.g. the closure of  $\mathbf{K}$  under expansions) can be carried over without modification to conditional-laden belief sets. More on this in Section 2.6 below. One might therefore respond to Theorem 14 by questioning the criterion of triviality: perhaps a model of belief change whose belief sets are conditional-laden should not be classified as trivial simply because it contains no consistent  $K, K'$  such that  $K$  is a proper subset of  $K'$ , even though a belief change model with conditional-free belief sets would be trivial in that case. But if the criterion of triviality espoused by Gärdenfors, *et al* [1991] is appropriate for conditional-free belief sets, then what about Theorem 15, which *does* cover belief change models with conditional-free belief sets? Our discussion in Section 2.6 below may appear to suggest that the problem raised by Theorems 14 and 15 might ultimately

be solved by giving up the monotonicity of  $s$  over  $\mathbf{K}$ , but even that is not guaranteed to be enough. Consider this result:<sup>30</sup>

**THEOREM 16.** *No classical belief change model defined on a language of type- $\mathcal{L}_1$  or type- $\mathcal{L}_2$  that satisfies  $(K^*T)$ ,  $(PRTL)$ , and  $(NRTL)$ , and whose support function satisfies Conservativeness and Closure, is Gärdenfors nontrivial or Rott nontrivial or Segerberg nontrivial.*

**Proof.** By Lemmas 5, 6, and 7, it suffices to show that no classical belief change model defined on a language of type- $\mathcal{L}_1$  or type- $\mathcal{L}_2$  that satisfies  $(K^*T)$ ,  $(PRTL)$ , and  $(NRTL)$ , and whose support function satisfies Conservativeness and Closure, contains forking support sets.

Consider a classical belief change model  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_\perp, -, *, s \rangle$  defined on a language of type- $\mathcal{L}_1$  or type- $\mathcal{L}_2$  that satisfies  $(K^*T)$ ,  $(PRTL)$ , and  $(NRTL)$ , and whose support function satisfies Conservativeness and Closure. Note first that since  $s$  satisfies Conservativeness and Closure,  $s$  must also satisfy Consistency.

Suppose the model contains forking support sets. Then there exist  $H, J, K \in \mathbf{K}$  such that  $H = \text{Cn}_\vdash(H) \cap K_\perp$ , and  $J = \text{Cn}_\vdash(J) \cap K_\perp$ , and  $K = \text{Cn}_\vdash(K) \cap K_\perp \neq K_\perp$ , and  $H \cap \mathcal{I} \not\subseteq J$ , and  $J \cap \mathcal{I} \not\subseteq H$ , and  $s(H) \subseteq s(K)$ , and  $s(J) \subseteq s(K)$ .

Since  $J \cap \mathcal{I} \not\subseteq H$  we have  $\phi \in J$  but  $\phi \notin H$  for some conditional-free  $\phi$ . By  $(K^*T)$  we have  $H = H_\top^*$  and  $J = J_\top^*$  and  $K = K_\top^*$ , hence  $\phi \in J_\top^*$  and  $\phi \notin H_\top^*$ . By  $(PRTL)$  we have  $\top > \phi \in s(J)$ , and by  $(NRTL)$  we have  $\neg(\top > \phi) \in s(H)$ . We also have  $\top > \phi \in s(K)$ , since  $s(J) \subseteq s(K)$ ; hence  $s(K)$  is not  $\vdash$ -consistent. This contradicts the  $\vdash$ -consistency of  $K$ , since  $s$  satisfies Consistency.  $\blacksquare$

As Theorem 16 shows,  $(NRTL)$ ,  $(K^*T)$  and  $(PRTL)$  cannot be nontrivially combined, even in a broad category of models where  $s$  is not monotone over  $\mathbf{K}$ , unless we abandon the Rott, Gärdenfors, and Segerberg criteria of nontriviality.

## 2.6 Resolving the conflict

### *On giving up (RT)*

Gärdenfors interprets Theorem 8 as forcing a choice between  $(K^*P)$  and the Ramsey test (RT), and he has argued (see, e.g., [Gärdenfors, 1986], pp. 86-87 and [Gärdenfors, 1988], p. 59 and p. 159) that  $(K^*M)$  and with it (RT) should be rejected. In this connection he offers the following example:

Let us assume that Miss Julie, in her present state of belief  $K$ , believes that her own blood group is O and that Johan is her

---

<sup>30</sup>The authors thank Horacio Arló-Costa for showing us the proof of this result in correspondence.

father, but she does not know anything about Johan's blood group. Let  $A$  be the proposition that Johan's blood group is AB and  $C$  the proposition that Johan is Miss Julie's father. If she were to revise her beliefs by adding the proposition  $A$ , she would still believe that  $C$ , that is,  $C \in K_A^*$ . But in fact she now learns that a person with blood group AB can never have a child with blood group O. This information, which entails  $C \rightarrow \neg A$ , is consistent with her present state of belief  $K$ , and thus her new state of belief, call it  $K'$ , is an expansion of  $K$ . If she then revises  $K'$  by adding the information that Johan's blood group is AB, she will no longer believe that Johan is her father, that is  $C \notin K'_A$ . Thus (K\*M) is violated. ([Gärdenfors, 1986], pp. 86-87)

The example assumes that  $s$  satisfies Identity over  $\mathbf{K}$ , so let us assume that as well. In reply to Gärdenfors one might say that if (RT) and Identity over  $\mathbf{K}$  are assumed, then the presence of conditionals in belief sets prevents this example from being a counterexample to (K\*M): if (RT) and Identity over  $\mathbf{K}$  are assumed, then since we have  $C \in K_A^*$  and  $C \notin K'_A$  it follows that  $A > C \in K$  and  $A > C \notin K'$ , in which case  $K \not\subseteq K'$ , i.e.  $K'$  is not an expansion of  $K$ . But to accept this, Gärdenfors argues, would violate certain intuitions:

[I]f we assume (RT) and not only (K\*M), then Miss Julie would have believed  $A > C$  in  $K$ . But then the information that a person with blood group AB can never have a child with blood group O, would *contradict* her beliefs in  $K$ , which violates our intuitions that this information is indeed consistent with her beliefs in  $K$ . ([Gärdenfors, 1986], p. 87)

Let  $B$  stand for the statement that a person with blood group AB can never have a child with blood group O. Gärdenfors has claimed in a context where  $s$  satisfies Identity over  $\mathbf{K}$  that if (RT) holds, then Miss Julie's beliefs in  $K$  contradict  $B$ , but this claim requires further justification: how exactly does  $K$  contradict  $B$ ? We might suppose that  $B$  entails  $\mathbf{L}(C \rightarrow \neg A)$ , where  $\mathbf{L}$  is an alethic nomological necessity operator expressing the modal force of  $B$ . Assuming (RT) and Identity over  $\mathbf{K}$ , the question whether  $K$  contradicts  $B$  depends on whether the set  $\{C, B, A > C\}$  is consistent, and this can be assumed to depend on whether the set  $\{C, \mathbf{L}(C \rightarrow \neg A), A > C\}$  is consistent. But the latter set *is* consistent if the semantics for  $>$  is not tied to nomological necessity. For example, given a selection function semantics for  $>$  and an accessibility relation semantics for  $\mathbf{L}$ , all of  $C$ ,  $A > C$ , and  $\mathbf{L}(C \rightarrow \neg A)$  can be true at possible world  $w$  if none of the  $A$ -worlds selected relative to  $w$  happen to be nomologically accessible at  $w$ .<sup>31</sup> So in

<sup>31</sup>For a less abstract version of essentially this point, see [Cross, 1990a], pp. 229-232.

order to sustain Gärdenfors' claim in the passage cited above, the claim that  $K$  contradicts  $B$  if (RT) holds (and if  $s$  satisfies Identity), we would have to assume the right sort of semantic connection between Ramsey test conditionals and the nomological modality in  $B$ , but the case for assuming that connection is not at all obvious: Ramsey test conditionals, after all, are *epistemic*. And if  $K$  indeed does not contradict  $B$ , then the same conditional that prevents the example from being a counterexample to (K\*M) makes the example a counterexample to (K\*P): since  $K' = K_B^*$ , and since  $C \in K_A^*$  but  $C \notin K_A'^*$ , it follows that if (RT) holds, then  $A > C \in K \not\subseteq K_B^* \not\Rightarrow A > C$  even though  $\neg B \notin K$ .

So it might be argued that the presence of Ramsey test conditionals in belief sets will render (K\*M) intuitively innocuous while providing perfectly reasonable counterexamples to (K\*P). Still, the arguments in favor of (K\*P) seem strong. One argument appeals to the Bayesian model of rationality. Suppose that an agent's belief state is represented as a probability function  $P$ . According to Bayesian doctrine, upon becoming certain of  $\phi$  a rational agent in belief state  $P$  revises her belief state by conditionalizing on  $\phi$ , assuming  $P(\phi) > 0$ . If this doctrine is correct and if an agent's belief set consists of those statements to which she assigns unit probability, then (K\*P) reduces to a theorem of probability theory: if  $P(\neg\phi) \neq 1$  and  $P(\psi) = 1$ , then  $P(\psi|\phi) = 1$ . A second argument appeals to the doctrine that revision can be defined in terms of contraction and expansion via the Levi Identity (LI), which prescribes the following: to revise with  $\phi$ , first contract relative to  $\neg\phi$  and then expand with  $\phi$ . If  $\vdash$  is classical and if (LI), (Def+), and (K-3) hold, then (K\*P) follows, the role of (K-3) being to require any contraction of  $K$  to be vacuous if the proposition contracted does not belong to  $K$ : if one does not believe a given proposition then no prior belief need be discarded when one contracts one's beliefs to exclude that proposition—it is already excluded.

As Theorem 10 shows, we can incorporate the second of these arguments for (K\*P) directly into the triviality result by recasting Theorem 8 in terms of an inconsistency between (RT), (LI), and postulates (K+1), (K-3), and (K-4w). (K-3) deals with what is in some sense the degenerate case of contraction: contraction with respect to an absent proposition. Postulate (K-4w) is similarly weak: it requires only that a contraction should really *be* a contraction in any case where a logically contingent proposition is contracted from a logically consistent belief set. Both postulates are very weak constraints on contraction, and their weakness makes the case against (RT) seem strong, as long as we assume that  $s$  satisfies Identity over  $\mathbf{K}$  or at least Monotonicity over  $\mathbf{K}$ .

Is there a weakened version of (RT) that is compatible with the other postulates mentioned in Theorem 8? Lindström and Rabinowicz [1992] show that there is. They suggest replacing (RT) with a condition whose counterpart in our framework is the following:

(SRT) For all  $K \in \mathbf{K}$  and all  $\phi \in \mathcal{I}$  and all  $\psi \in K_{\perp}$ ,  $\phi > \psi \in s(K)$  iff  $\psi \in K'_{\phi}$  for all  $K' \in \mathbf{K}$  such that  $K \subseteq K'$ .

Like Gärdenfors, Lindström and Rabinowicz do not distinguish between  $K$  and  $s(K)$ , and in a context where this distinction is not made (i.e. where  $s$  satisfies Identity over  $\mathbf{K}$ ), replacing (RT) with (SRT) has the effect of excluding from belief sets many of the conditionals that must be present in them if (RT) and Identity over  $\mathbf{K}$  are assumed. For example, in Gärdenfors' Miss Julie case, (RT) and Identity over  $\mathbf{K}$  force the conclusion that  $A > C \in K$  and  $A > C \notin K'_A$ , since  $C \in K_A^*$  and  $C \notin K'_A$ , giving us a counterexample to (K\*P) since  $\neg B \notin K$  and  $K' = K_B^*$ . If (SRT) and Identity over  $\mathbf{K}$  are assumed instead, then the falsity of (K\*P) no longer follows. The question whether  $A > C$  belongs to  $K$  depends not simply on  $K_A^*$  but on the revision behaviour of all belief sets that include  $K$  as a subset, and similarly for the question whether  $A > C$  belongs to  $K'$ .

*On giving up (K\*P)*

Should the Ramsey test be preserved at the expense of (K\*P)? The answer is certainly *yes* if the Ramsey test is applied to the notion of theory change to which Katsuno and Mendelzon in [Katsuno and Mendelzon, 1992] attach the label *update*. They write:

... [*U*]update, consists of bringing the knowledge base up to date when the world described by it changes. For example, most database updates are of this variety, e.g. "increase Joe's salary by 5%". Another example is the incorporation into the knowledge base of changes caused in the world by the actions of a robot.<sup>32</sup>

Update, according to Katsuno and Mendelzon, contrasts with *revision*:<sup>33</sup>

... [*R*]evision, is used when we are obtaining new information about a static world. For example, we may be trying to diagnose a faulty circuit and want to incorporate into the knowledge base the results of successive tests, where newer results may contradict old ones. We claim the AGM postulates describe only revision.<sup>34</sup>

Katsuno and Mendelzon represent knowledge bases as formulas  $\chi$  and introduce a binary modal connective to represent the update operation. Following [Grahne, 1991] we will use the symbol ' $\circ$ ' for this operation; then the

<sup>32</sup>[Katsuno and Mendelzon, 1992], p. 183.

<sup>33</sup>See also [Winslett, 1990].

<sup>34</sup>*Ibid.*

formula  $\chi \circ \phi$  is the knowledge base that results from updating knowledge base  $\chi$  with new information  $\phi$ .

Grahne [1991] provides an interpretation of the Ramsey test in terms of update. Given a type- $\mathcal{L}_2$  language that includes the binary connective ‘ $\circ$ ’ Grahne simply adds to Lewis’ system **VCU** the following (validity preserving) rule of inference:

RR: From  $\chi \rightarrow (\phi > \psi)$  infer  $(\chi \circ \phi) \rightarrow \psi$ , and from  $(\chi \circ \phi) \rightarrow \psi$  infer  $\chi \rightarrow (\phi > \psi)$ .

Grahne calls the resulting logical system **VCU<sup>2</sup>**. In Grahne’s framework, the formula  $\chi \circ \phi$  is true in possible world  $w$  iff  $w$  belongs to the set of closest worlds to  $w'$  in which  $\phi$  is true for at least one world  $w'$  in which  $\chi$  is true.

Grahne proves soundness, completeness, decidability, and nontriviality results for **VCU<sup>2</sup>**, and he notes that **VCU<sup>2</sup>** fails to satisfy the following principle:

(U\*4s) If  $\not\vdash \neg(\chi \wedge \phi)$ , then  $\vdash (\chi \circ \phi) \leftrightarrow (\chi \wedge \phi)$ .

(U\*4s) states that if  $\phi$  is consistent with knowledge base  $\chi$ , then the result of updating  $\chi$  with  $\phi$  is a formula logically equivalent to  $\chi \wedge \phi$ . Grahne cites the following example to illustrate the failure of (U\*4s), which is the update counterpart of revision postulate (K\*4s):

A room has two objects in it, a book and a magazine. Suppose  $p_1$  means that the book is on the floor, and  $p_2$  means that the magazine is on the floor. Let the knowledge base be  $(p_1 \vee p_2) \wedge \neg(p_1 \wedge p_2)$ , i.e. either the book or the magazine is on the floor, but not both. Now we order a robot to put the book on the floor, that is, our new piece of knowledge is  $p_1$ . If this change is taken as a revision [so that (K\*4s) is assumed], then we find that since the knowledge base is consistent with  $p_1$ , our new knowledge base will be equivalent to  $p_1 \wedge \neg p_2$ , i.e. the book is on the floor and the magazine is not.

But the above change is inadequate. After the robot moves the book to the floor, all we know is that the book is on the floor; why should we conclude that the magazine is not on the floor?<sup>35</sup>

That is, upon updating to include  $p_1$ , we should *give up* something we believed in our initial epistemic state, namely  $\neg(p_1 \wedge p_2)$ , even though the new information  $p_1$  is consistent with our initial epistemic state. Apparently, then, we have made a belief change using a method that does not satisfy an appropriate counterpart of (K\*P). Isaac Levi disagrees. The mechanism which underlies update is *imaging*: the “image” of a set  $S$  of possible worlds

---

<sup>35</sup>[Grahne, 1991], pp. 274–275.



under  $\phi$  is the set of worlds each of which is one of the closest  $\phi$ -worlds to some world belonging to  $S$ . Levi [Levi, 1996] argues that while imaging may be useful for describing how changes over time in the state of a system (such as the room in Grahne's example) are regulated, such changes are not an example of *belief* change. We may, of course, have *beliefs* about how changes in a system over time are regulated, but an analysis of Grahne's example along the lines recommended by Levi would show it to be a straightforward case in which belief revision took place via *expansion*: if  $t$  is a time before the book was moved and  $t'$  is a time just after the book is moved and if propositional variables  $p_1$  and  $p_2$  are replaced by formulas containing predicates  $P_1$  and  $P_2$ , where  $P_i u$  means that  $p_i$  is true at time  $u$ , then our initial epistemic state can be represented as  $(P_1 t \vee P_2 t) \wedge \neg(P_1 t \wedge P_2 t)$ , and upon learning of the change in the position of the book our new epistemic state is  $(P_1 t \vee P_2 t) \wedge \neg(P_1 t \wedge P_2 t) \wedge P_1 t'$ .

### *On giving up (K<sup>+</sup>1)*

Gärdenfors interprets Theorem 8 as forcing a choice between (RT) and (K\*P), but Rott [1989], Morreau [1992], and Hansson [1992] have argued that (K<sup>+</sup>1) is the real culprit.

Postulates (K\*3) and (K\*4) entail (K\*4s): if  $\phi$  is consistent with  $K$ , then a revision to accept  $\phi$  should be the result of expanding  $K$  with  $\phi$ . Rott [1989] argues that (K\*4s), while fine for belief revision in a type- $\mathcal{L}_0$  language, is an inappropriate requirement on belief revision in a language with Ramsey test conditionals. Once (K\*4s) is rejected in the context of Ramsey test conditionals, Rott argues, (K<sup>+</sup>1) is robbed of any intuitive basis: the only reason for thinking that belief change models should be closed under expansion would be the assumption that expansion is a species of revision. Why think that expansion is a species of revision in the first place? One could justify (K\*4s) as the qualitative analog of the Bayesian doctrine that upon becoming certain of  $\phi$  a rational agent whose belief state is represented by probability function  $P$  revises her belief state by conditionalizing on  $\phi$  if  $P(\phi) > 0$ . This doctrine supports (K\*4s) because if  $P(\phi) > 0$ , then the set  $\{\psi : P(\psi|\phi) = 1\}$  is precisely the result of expanding the set  $\{\psi : P(\psi) = 1\}$  with  $\phi$ . But, Morreau [1992] counters, Bayesian doctrine supports (K\*4s) in this way only for belief sets over a type- $\mathcal{L}_0$  language.

Still, one might argue, regardless whether revision ever leads from a belief set to one of its expansions, should not the expansion of every belief set in a belief change model be available in the model as a possible *starting point* for revision? Not so, argues Morreau [1992]: a belief change model over which (RT) holds and in which belief sets contain conditionals incorporates the idealizing assumption that the conditionals an agent believes form a complete and correct record of how the agent would revise his or her beliefs.

Not just any collection of theories in a conditional language can be the belief sets of a Ramsey test respecting belief change model because not just any theory will conform to the idealization. Morreau interprets the Gärdenfors triviality result as showing in particular that the idealization required by the Ramsey test cannot be achieved in a nontrivial belief change model that respects  $(K^+1)$  while incorporating conditionals in belief sets.

But are there in fact nontrivial belief change models containing conditional-laden belief sets in which (RT) holds but  $(K^+1)$  does not? Morreau's Example 6 ([Morreau, 1992], p. 41), which we adapt to our framework, confirms that there are. Let  $L$  be a type- $\mathcal{L}_1$  language and let  $L_0$  be its type- $\mathcal{L}_0$  fragment. Assume that  $L_0$  contains at least two distinct atomic formulas. Let  $\vdash_0$  be truth-functional consequence, and assume (Def+), and let  $\mathbf{K}_0$  be the set of all  $\vdash_0$ -theories in  $L_0$ . Define a belief revision operation  $\star$  as follows for all  $\mu \in \mathbf{K}_0$  and all formulas  $\phi$  of  $L_0$ :

$$\mu_\phi^\star = \begin{cases} \mu & \text{if } \perp \in \mu; \\ \mu_\phi^+ & \text{if } \perp \notin \mu \text{ and } \neg\phi \notin \mu; \\ \text{Cn}_{\vdash_0}(\{\phi\}) & \text{otherwise.} \end{cases}$$

Let  $\mathcal{I}_0 = \text{WFF}_{L_0} = K_{\perp_0}$ ; let  $\text{dom}(s_0) = \mathbf{K}_0$ ; and let  $s_0(K) = K$  for all  $K \in \mathbf{K}_0$ . Letting the contraction operation  $(-_0)$  be arbitrary, note that  $\langle \mathbf{K}_0, \mathcal{I}_0, \vdash_0, K_{\perp_0}, -_0, \star, s_0 \rangle$  satisfies  $(K^+1)$ ,  $(K^*2)$ ,  $(K^*C)$ , and  $(K^*P)$ , but not (RT). Using  $\mathbf{K}_0$  and  $\star$  we construct a second, Ramsey test supporting belief change model with conditional-laden belief sets as follows: for each  $\mu \in \mathbf{K}_0$ , let  $K_\mu = \text{Cn}_{\vdash_0}(\mu \cup \{\phi > \psi : \psi \in \mu_\phi^\star\})$ . Let  $\mathbf{K} = \{K_\mu : \mu \in \mathbf{K}_0\}$ ; let  $(K_\mu)_\phi^* = K_{\mu_\phi^\star}$ ; let  $\mathcal{I} = \text{WFF}_{L_0}$  (as before); let  $K_\perp = \text{WFF}_L$ ; let  $\text{dom}(s) = \mathbf{K}$ ; and let  $s(K) = K$  for all  $K \in \mathbf{K}$ . Letting contraction  $(-)$  again be arbitrary,  $\langle \mathbf{K}, \mathcal{I}, \vdash_0, K_\perp, -, *, s \rangle$  satisfies  $(K^*2)$ ,  $(K^*C)$ , and (RT), but this model, unlike the first, does not satisfy  $(K^+1)$  or  $(K^*P)$ .<sup>36</sup> For example, let  $A, B$  be distinct atomic formulas of  $L_0$ , and let  $\mu' = \text{Cn}_{\vdash_0}(\{B\})$ ; thus  $\mu' \in \mathbf{K}_0$ . Since  $\neg A \notin \mu'$ , we have that  $\mu'_A{}^\star = \mu'_A{}^+ = \text{Cn}_{\vdash_0}(\{A, B\})$ . Accordingly,  $A > B \in K_{\mu'} \in \mathbf{K}$ , but note that  $(K_{\mu'})_{\neg A}^+$  does not belong to  $\mathbf{K}$ , for there is no  $\mu \in \mathbf{K}_0$  such that both  $\neg A$  and  $A > B$  belong to  $K_\mu$ .

The second belief change model constructed above is the result of closing the first model under the Ramsey test (restricted to non-nested conditionals), and both models are Gärdenfors nontrivial. The Gärdenfors nontriviality of the second model is established by  $K_{\text{Cn}_{\vdash_0}(\emptyset)}$ , which belongs to  $\mathbf{K}$ , and  $A \wedge \neg B$ ,  $B \wedge \neg A$ , and  $A \wedge B$  which belong to  $L$ . In addition to this example Morreau provides a general recipe for constructing nontrivial models of belief revision in a type- $\mathcal{L}_1$  language  $L$  whose type- $\mathcal{L}_0$  fragment is  $L_0$  and where  $(K^*1)$ ,  $(K^*2)$ ,  $(K^*C)$ , (RT), and  $(K^*P_{\mathcal{I}})$  hold and  $\mathcal{I} = \text{WFF}_{L_0}$ .

<sup>36</sup>The model does satisfy a weakened version of  $(K^*P)$ , however, as Morreau points out:

$(K^*P_{\mathcal{I}})$  For all  $\phi \in \mathcal{I}$ , if  $\neg\phi \notin K$  then  $K \cap \mathcal{I} \subseteq K_\phi^* \cap \mathcal{I}$ .

Where does the triviality proof break down when applied to Morreau's example? Hansson [1992] proves a theorem that provides the answer: Morreau's example is one of a set of belief change models in which forking support sets cannot be constructed. The counterpart of Hansson's theorem in our framework is the following:

**THEOREM 17.** *Suppose that  $\langle \mathbf{K}, \mathcal{I}, \vdash, K_{\perp}, -, *, s \rangle$  is a belief change model defined on a language  $L$  of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  and that  $\vdash$  includes all truth-functional entailments and respects the deduction theorem for the material conditional. Suppose also that  $\text{dom}(s) = \mathbf{K}$ , that  $s(K) \subseteq \text{WFF}_L$  for all  $K \in \mathbf{K}$ , and that  $s$  satisfies the following for all  $K \in \mathbf{K}$ , all  $\phi \in \mathcal{I}$  and all  $\psi, \chi \in K_{\perp}$ :*

1. *If  $\Gamma \subseteq s(K)$  and  $\Gamma \vdash \psi$  then  $\psi \in s(K)$ .*
2. *If  $K$  is  $\vdash$ -consistent and  $\psi \in s(K)$ , then  $\neg\psi \notin s(K)$ .*
3. *If  $K$  is  $\vdash$ -consistent and  $\not\vdash \neg\phi$  and  $\phi > \psi$ ,  $\phi > \chi \in s(K)$ , then  $\not\vdash \neg(\psi \wedge \chi)$ .*
4. *If  $\phi > (\psi \wedge \phi) \in \text{WFF}_L$  and  $\psi \in s(K)$  and  $\phi \notin s(K)$  and  $\neg\phi \notin s(K)$ , then  $\phi > (\psi \wedge \phi) \in s(K)$ .*

*Suppose that  $K_1, K_2, K \in \mathbf{K} \sim \{K_{\perp}\}$  and that  $s(K_1)$  and  $s(K_2)$  are both subsets of  $s(K)$ . Then either  $s(K_1) \subseteq s(K_2)$  or  $s(K_2) \subseteq s(K_1)$ .*

Conditions 1 and 2 are equivalent to Closure and Consistency for  $s$ , respectively. Note that the Ramsey test itself is not assumed: the point is that simply having conditionals in a belief change model that meets these four conditions ensures that forking support sets cannot be constructed.

#### *On giving up the monotonicity of $s$ over $\mathbf{K}$*

Rott [1989; 1991] suggests that nonmonotonic reasoning may provide a solution to the dilemma posed by the Gärdenfors triviality result, and Cross [1990a] argues that Gärdenfors' triviality result should be interpreted as showing not that the Ramsey test should be abandoned but that, given the Ramsey test,  $s$  must be nonmonotonic over  $\mathbf{K}$ , i.e. for some  $H, K \in \mathbf{K}$   $H \subseteq K$  but  $s(H) \not\subseteq s(K)$ .<sup>37</sup> Makinson counters in [Makinson, 1990] with a triviality result for models in which  $s$  is permitted to be nonmonotonic, but Makinson's result does not bear on the suggestion endorsed by Cross and by Rott. More on this below.

Other authors have brought nonmonotonic reasoning into the discussion of the Ramsey test without advertising it as such. For example, in [Hansson, 1992] Hansson writes:

---

<sup>37</sup>Since the monotonicity of  $s$  is not assumed in Theorem 13, however, it is clear that the problem for (RT) posed by (K\*5) cannot be solved by making  $s$  nonmonotonic.

... the addition of an indicative sentence that is compatible with all previously supported indicative sentences typically withdraws the support of conditional sentences that were previously supported.<sup>38</sup>

The type- $\mathcal{L}_1$  statements that are in Hansson's sense *supported by* a given "indicative" (i.e. conditional-free) belief base  $K$  represent what Cross (and possibly Rott) would classify as the nonmonotonic consequences of  $K$ . Hansson and Cross both think of the sets that individuate belief states as belief bases and define contraction and revision as operations on these sets, but Cross' belief bases differ from Hansson's in two respects: first, whereas Hansson's belief bases are conditional-free, Cross' are not; secondly, whereas Hansson's belief bases are not closed under  $\vdash$  or closed under  $s$ , in Cross' enriched belief revision models belief bases are closed under  $\vdash$ , though not under  $s$ . That is, for Hansson, belief states are individuated in terms of sets that function as belief bases with respect to both  $\vdash$  and  $s$ , whereas for Cross, belief states are individuated in terms of sets that function as belief bases only with respect to  $s$ . For Hansson, belief bases need not be closed under  $\vdash$  and are never closed under  $s$ . Makinson [1990], like Cross [1990a], supplements the classical  $\vdash$  with a not-necessarily-monotonic  $s$ ,<sup>39</sup> and like Cross, Makinson explicitly advertises  $s$  as a consequence operation. But in Makinson's discussion revision and contraction are defined only on  $K$  that are closed under  $s$ , and the proof of Makinson's triviality theorem, whose counterpart here is Theorem 11 above, requires a belief change model containing three belief sets closed under  $s$ . Makinson's triviality result does not apply to belief change models in which  $\mathbf{K}$  contains *no*  $K$  such that  $s(K) = K$ , and such authors as Arló-Costa and Levi (see [Levi, 1988], [Arló-Costa, 1995], and [Arló-Costa and Levi, 1996]) avoid Makinson's triviality result precisely by requiring  $s(K) \neq K$  for all  $K \in \mathbf{K}$ .

As we noted above, Hansson does not explicitly speak of the support function as a consequence operation, nor does Arló-Costa or Levi. Yet, if one looks at the conditions that Hansson, Arló-Costa, and Levi place on the support function, mirrored here in the definitions of Hansson and Arló-Costa/Levi belief change models as the requirements of Reflexivity, Conservativeness, and Closure, it seems natural to think of  $s$  as a nonmonotonic consequence operation. But if we do think of  $s$  in a Hansson or Arló-Costa/Levi belief change model as a nonmonotonic consequence operation, what sort of nonmonotonic reasoning does it represent? In [Moore, 1983] Robert Moore distinguishes two types of nonmonotonic reasoning:

By default reasoning, we mean drawing plausible inferences from less than conclusive evidence in the absence of any information

---

<sup>38</sup>[Hansson, 1992], p. 526.

<sup>39</sup>Makinson uses the symbol  $C$  and Cross the symbol  $cl$  for  $s$ .

to the contrary. The examples about birds being able to fly are of this type.<sup>40</sup>

He continues:

Default reasoning is nonmonotonic because, to use a term from philosophy, it is *defeasible*. Its conclusions are tentative, so, given better information, they may be withdrawn.<sup>41</sup>

Default reasoning, according to Moore, contrasts with *autoepistemic reasoning*, or reasoning about one's state of belief. Moore writes:

Autoepistemic reasoning is nonmonotonic because the meaning of an autoepistemic statement is context-sensitive; it depends on the theory in which the statement is embedded.<sup>42</sup>

For example, if  $\diamond\phi$  is defined as being accepted in belief state  $K$  just in case  $\neg\phi$  is not accepted in  $K$ , then  $\diamond\phi$  is an autoepistemic statement in Moore's sense. If the support function  $s$  in a belief change model is thought of as a nonmonotonic consequence operation, then how should  $s$  be classified with respect to Moore's distinction? It depends on the properties  $s$  is assumed to have.

If a belief change model satisfies some version of the Ramsey test (e.g. (RT), (PRTL), or (NRTL)), then the support function of that model is *at least* a form of autoepistemic reasoning. This is clear since the acceptability of a Ramsey test conditional for a given agent is in part a function of the agent's current epistemic state, and this holds true regardless whether conditionals themselves are objects of belief.<sup>43</sup> Moreover, the context sensitivity to which Moore refers in the passage just quoted is clearly present in the support function of any belief change model that satisfies (RT), and indeed this context sensitivity was exploited by Morreau [1992] in his construction of a nontrivial Ramsey test, by Lindström and Rabinowicz [1995] and Lindström [1996] in a proposed indexical interpretation of conditionals,<sup>44</sup> by Hansson [1992] in his accounts of type- $\mathcal{L}_1$  conditionals and iterated conditionals, respectively, and by Boutilier and Goldszmidt [1995] in their account of the revision of conditional belief sets.

Given a support function  $s$  for a belief change model that satisfies a version of the Ramsey test, can  $s$  be not only a mechanism for autoepistemic reasoning but a mechanism for default reasoning, too? This depends on whether  $s$  can be used to make ampliative inferences to conclusions that

---

<sup>40</sup>[Moore, 1983], p. 273.

<sup>41</sup>[Moore, 1983], p. 274.

<sup>42</sup>[Moore, 1983], p. 274.

<sup>43</sup>Rott [1989] and Morreau [1992] explicitly adopt the view that Ramsey test conditionals are autoepistemic.

<sup>44</sup>See also [Döring, 1997].

are not epistemically context-sensitive from premises that are not epistemically context-sensitive. Since Hansson, Arló-Costa, and Levi assume that  $s$  satisfies Conservativeness, it is clear that for them  $s$  is an operation of autoepistemic reasoning but *not* an operation of default reasoning:  $s(K)$  will contain conditionals, i.e. autoepistemic statements, that are not logical consequences of  $K$ , but no conditional-free formula gets into  $s(K)$  without being a logical consequence of  $K$ , which is itself conditional-free for any  $K \in \text{dom}(s)$  according to Hansson, Arló-Costa, and Levi. Cross and Makinson, on the other hand, do not require the support function to satisfy Conservativeness; accordingly, they allow belief change models in which  $s$  supports default reasoning. No distinction between  $s(K)$  and  $K$  exists for Morreau, Gärdenfors, and Segerberg, hence the issue of the status of  $s$  does not arise in their respective cases.

### 2.7 Logics for Ramsey test conditionals

Gärdenfors [1978] proves the soundness and completeness of David Lewis' system of conditional logic **VC** with respect to an epistemic, Ramsey test semantics for the conditional. Several other authors have proposed variants of Gärdenfors' Ramsey test semantics, including variants that generalize Gärdenfors' semantics, but it will be convenient for our purposes to adopt formalisms similar to those of [Arló-Costa, 1995] and [Arló-Costa and Levi, 1996].

#### *Primitive belief revision models*

Since the conditional is to be given a *semantics* in terms of belief revision, the notion of a belief set must be defined in terms that do not assume a logic for the conditional. To this end we define *primitive belief sets*, *primitive expansion*, and *primitive belief revision models*.

For a Boolean language  $L$  of type  $\mathcal{L}_1$  or  $\mathcal{L}_2$  let a *primitive belief set* defined on this language be any set  $K$  of formulas of  $L$  meeting three requirements:

(pBS1)  $K \neq \emptyset$ ;

(pBS2) if  $\phi \in K$  and  $\psi \in K$ , then  $\phi \wedge \psi \in K$ ;

(pBS3) if  $\phi \in K$  and  $\phi \rightarrow \psi$  is a truth-functional tautology, then  $\psi \in K$ .

If  $K \subseteq K'$  and  $K, K'$  are both primitive belief sets, then  $K'$  is a *primitive expansion* of  $K$ . The *operation* of primitive expansion is defined as follows:

(DEF+)  $K_\phi^+ = \{\psi : \phi \rightarrow \psi \in K\}$ .

It is easy to see that if  $K$  is a primitive belief set, then so is  $K_\phi^+$ . Finally, let us define the notion of a *primitive belief revision model* on  $L$ :

(DefpBRM) A *primitive belief revision model* (or pBRM) on a Boolean language  $L$  is an ordered quadruple  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  whose components are as follows:

1.  $K_{\perp} = \text{WFF}_{L'}$ , where  $L'$  is  $L$  or a fragment of  $L$ ;
2.  $\mathbf{K}$  is a nonempty set of primitive belief sets defined on  $L'$ , and if  $K \in \mathbf{K}$  then  $\mathbf{K}$  contains every primitive expansion of  $K$  on  $L'$ ;
3.  $*$  is a function mapping each  $K \in \mathbf{K}$  and each formula  $\phi \in K_{\perp}$  to a primitive belief set  $K_{\phi}^*$  belonging to  $\mathbf{K}$ ;
4.  $s$  is a function mapping each  $K \in \mathbf{K}$  to a primitive belief set  $s(K)$  of formulas of  $L$ , where  $s$  satisfies the following:
  - (a) if  $\phi \in K_{\perp}$  and  $\phi \in s(K)$  then  $\phi \in K$ ;
  - (b)  $K \subseteq s(K)$  if  $K_{\perp} \neq K \in \mathbf{K}$ .

Note that while  $K$  and  $s(K)$  must both be primitive belief sets, they need not be primitive belief sets of the same language. When referring to the belief revision postulates (K\*1), (K\*2), etc., in the context of primitive belief revision models we will assume that  $\phi$  and  $\psi$  range over  $K_{\perp}$ .

A primitive belief revision model  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  defined on a Boolean language  $L$  is a *Gärdenfors pBRM* iff  $L$  is of type  $\mathcal{L}_2$  and  $K_{\perp} = \text{WFF}_L$  and  $s$  is the identity function on  $K$  and  $s$  satisfies the following unrestricted version of the positive Ramsey test:

(pRTG) For all  $K \in \mathbf{K}$ , if  $\phi, \psi \in K_{\perp}$ , then  $(\phi > \psi) \in s(K)$  iff  $\psi \in K_{\phi}^*$ .

A primitive belief revision model  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  defined on a Boolean language  $L$  is an *Arló-Costa/Levi pBRM* iff  $L$  is of type  $\mathcal{L}_1$  and  $K_{\perp}$  is the set of all formulas of the largest conditional-free fragment of  $L$  and  $s$  satisfies the following versions of both the positive and negative Ramsey tests:

(pPRT) For all  $K \in \mathbf{K}$  such that  $K \neq K_{\perp}$ , if  $\phi, \psi \in K_{\perp}$ , then  $(\phi > \psi) \in s(K)$  iff  $\psi \in K_{\phi}^*$ .

(pNRT) For all  $K \in \mathbf{K}$  such that  $K \neq K_{\perp}$ , if  $\phi, \psi \in K_{\perp}$ , then  $\neg(\phi > \psi) \in s(K)$  iff  $\psi \notin K_{\phi}^*$ .

#### *Positive and negative validity*

In [Arló-Costa, 1995] (and in [Arló-Costa and Levi, 1996], with Isaac Levi) Arló-Costa distinguishes between positive and negative concepts of validity. The concepts are distinct in Arló-Costa/Levi pBRMs, though not in Gärdenfors pBRMs.

A formula  $\phi$  is *positively valid (PV) relative to*  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$ , where the latter is a primitive belief revision model, iff  $\phi \in s(K)$  for all  $K$  such that

$K_{\perp} \neq K \in \mathbf{K}$ ; and  $\phi$  is *positively valid relative to a set of belief revision models* iff  $\phi$  is positively valid relative to each member of the set.  $\phi$  is *negatively valid (NV) relative to  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$*  iff  $\neg\phi \notin s(K)$  for each  $K$  such that  $K_{\perp} \neq K \in \mathbf{K}$ ; and  $\phi$  is *negatively valid relative to a set of belief revision models* iff  $\phi$  is negatively valid relative to each member of the set.

Notions of entailment can be associated with positive and negative validity, respectively. Given a set  $\Gamma$  of formulas of a type- $\mathcal{L}_1$  or type- $\mathcal{L}_2$  language  $L$ , and a formula  $\phi$  of  $L$ ,  $\Gamma$  *positively entails  $\phi$*  ( $\Gamma \models_+ \phi$ ) with respect to a primitive belief revision model  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  iff  $\phi \in s(K)$  for every  $K$  such that  $\Gamma \subseteq s(K)$  and  $K_{\perp} \neq K \in \mathbf{K}$ . By contrast,  $\Gamma$  *negatively entails  $\phi$*  ( $\Gamma \models_- \phi$ ) with respect to a primitive belief revision model  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  iff there is no  $K$  such that  $K_{\perp} \neq K \in \mathbf{K}$  and  $\Gamma \cup \{\neg\phi\} \subseteq s(K)$ .

In a Gärdenfors pBRM, positive and negative validity coincide:

**PROPOSITION 18.** *Relative to any Gärdenfors pBRM defined on a language  $L$  of type  $\mathcal{L}_2$ , a formula  $\phi$  of  $L$  is positively valid iff  $\phi$  is negatively valid.*

**Proof.** Let  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  be a Gärdenfors pBRM defined on a language  $L$  of type  $\mathcal{L}_2$ , and let  $\phi$  be a formula of  $L$ . First, suppose that  $\phi$  is positively valid relative to  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  and choose an arbitrary  $K$  such that  $K_{\perp} \neq K \in \mathbf{K}$ . Then  $\phi \in s(K)$ , but since  $s(K) = K \neq K_{\perp}$  we have  $\neg\phi \notin s(K)$ , as required. Conversely, assume that  $\phi$  is negatively valid relative to  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  and choose an arbitrary  $K$  such that  $K_{\perp} \neq K \in \mathbf{K}$ . Assume for *reductio* that  $\phi \notin s(K)$ . Since  $s$  is the identity function, we have that  $\phi \notin K$ , in which case  $K_{-\phi}^+ \neq K_{\perp}$ . Since  $\mathbf{K}$  is closed under primitive expansions, we have in addition that  $K_{-\phi}^+ \in \mathbf{K}$ . Thus,  $\neg\phi \in s(K_{-\phi}^+)$  and  $K_{\perp} \neq K_{-\phi}^+ \in \mathbf{K}$ , which is contrary to the negative validity of  $\phi$ . ■

Positive and negative validity do not coincide in Arló-Costa/Levi pBRMs, however. The negative Ramsey test prevents it. Consider the following pair of lemmas regarding the thesis (CS):

**LEMMA 19.** *For any type- $\mathcal{L}_1$  language  $L$ , if  $\phi, \psi$  are conditional-free, then  $(\phi \wedge \psi) \rightarrow (\phi > \psi)$  is negatively valid in an Arló-Costa/Levi pBRM defined on  $L$  iff the model satisfies  $(K^* \not\vdash w)$ .*

The latter is equivalent to Observation 4.7 of [Arló-Costa and Levi, 1996].

**LEMMA 20.** *For any type- $\mathcal{L}_1$  language  $L$  containing at least one atomic formula  $\chi$  other than  $\top$  and  $\perp$ , there are conditional-free  $\phi$  and  $\psi$  such that  $(\phi \wedge \psi) \rightarrow (\phi > \psi)$  is not positively valid in any Arló-Costa/Levi pBRM defined on  $L$  that satisfies  $(K^* 3)$  and contains a primitive belief set  $K$  where  $\neg\chi, \chi \notin K$ .*

**Proof.** Suppose  $L$  is a type- $\mathcal{L}_1$  language containing at least one atomic formula  $\chi$  different from  $\top$  and  $\perp$ , and consider an Arló-Costa/Levi pBRM



$\langle \mathbf{K}, *, K_{\perp}, s \rangle$  defined on  $L$  that satisfies  $(\mathbf{K}^*3)$  and contains a primitive belief set  $K$  such that  $\neg\chi, \chi \notin K$ . Suppose for *reductio* that  $(\phi \wedge \psi) \rightarrow (\phi > \psi)$  is positively valid for all conditional-free  $\phi$  and  $\psi$ . Then, in particular  $(\top \wedge \chi) \rightarrow (\top > \chi)$  is positively valid relative to  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$ . By  $(\mathbf{K}^*3)$  we have  $K_{\top}^* \subseteq K_{\top}^+ = K$ . Since, in addition,  $\neg\chi, \chi \notin K$  we have that  $\neg\chi, \chi \notin K_{\top}^*$ . Since  $\chi \notin K_{\top}^*$ , it follows by the Negative Ramsey test that  $\neg(\top > \chi) \in s(K)$ , hence by the positive validity of  $(\top \wedge \chi) \rightarrow (\top > \chi)$  relative to  $\langle \mathbf{K}, *, K_{\perp}, s \rangle$  we have that  $\neg(\top \wedge \chi) \in s(K)$ . Since  $\neg(\top \wedge \chi)$  is conditional-free, it follows that  $\neg(\top \wedge \chi) \in K$ . Since primitive belief sets are deductively closed, we have  $\neg\chi \in K$ , contrary to assumption.  $\blacksquare$

The proof just given is derived from that given by Arló-Costa for Observation 3.14 in [Arló-Costa, 1995]. Finally, we state the following obvious but necessary lemma:

LEMMA 21. *For some type- $\mathcal{L}_1$  language  $L$ , there is an Arló-Costa/Levi pBRM defined on  $L$  that satisfies  $(\mathbf{K}^*3)$  and  $(\mathbf{K}^*4w)$  and also contains a primitive belief set  $K$  where  $\neg\phi, \phi \notin K$  for some conditional-free formula  $\phi$  of  $L$ .*

These three lemmas suffice to show the following:

THEOREM 22. *There are Arló-Costa/Levi pBRMs relative to which at least some formulas of the form  $(\phi \wedge \psi) \rightarrow (\phi > \psi)$  are negatively valid but not positively valid.*

Interestingly, despite Theorem 22,  $\{\phi \wedge \psi\} \models_+ \phi > \psi$  holds relative to every Arló-Costa/Levi pBRM that satisfies  $(\mathbf{K}^*4w)$ .<sup>45</sup>

### *Belief revision models for VC*

Gärdenfors provides an epistemic semantics for **VC** based on negative validity. He begins with a minimal conditional logic **CM** defined as follows:

#### *Axiom schemata*

Taut: All truth-functional tautologies;

CC:  $[(\phi > \psi) \wedge (\phi > \chi)] \rightarrow [\phi > (\psi \wedge \chi)];$

CN:  $\phi > \top.$

#### *Rules of inference*

*Modus Ponens* From  $\phi$  and  $\phi \rightarrow \psi$  to infer  $\psi$ ;

RCM: From  $\psi \rightarrow \chi$  to infer  $(\phi > \psi) \rightarrow (\phi > \chi).$

<sup>45</sup>See OBSERVATION 3.15 in [Arló-Costa, 1995].

Gärdenfors [1978] proves a soundness/completeness theorem for **CM** that is equivalent to the following:

**THEOREM 23.** *A formula of any type  $\mathcal{L}_2$  language is a theorem of **CM** iff it is negatively valid in every Gärdenfors pBRM.*

Gärdenfors then proves the following:

**THEOREM 24.** *A formula is a theorem of **VC** iff it is derivable from **CM** together with (ID), (CSO'), (CS), (MP), (CA), and (CV) as additional axiom schemata:*

$$\begin{aligned}
 \text{ID:} & \quad \phi > \phi \\
 \text{CSO':} & \quad [(\phi > \psi) \wedge (\psi > \phi)] \rightarrow [(\phi > \chi) \rightarrow (\psi > \chi)] \\
 \text{CS:} & \quad (\phi \wedge \psi) \rightarrow (\phi > \psi) \\
 \text{MP:} & \quad (\phi > \psi) \rightarrow (\phi \rightarrow \psi) \\
 \text{CA:} & \quad [(\phi > \chi) \wedge (\psi > \chi)] \rightarrow [(\phi \vee \psi) > \chi] \\
 \text{CV:} & \quad [(\phi > \psi) \wedge \neg(\phi > \neg\chi)] \rightarrow [(\phi \wedge \chi) > \psi]
 \end{aligned}$$

An epistemic semantics for **VC** is obtained by restricting attention to Gärdenfors pBRMs that satisfy constraints corresponding to axioms ID, CSO', CS, MP, CA, and CV. Gärdenfors [1978] proves lemmas equivalent to the following:

**LEMMA 25.** *Where  $\mathcal{M}$  is any Gärdenfors pBRM,*

1. *all instances of ID are negatively valid in  $\mathcal{M}$  iff  $\mathcal{M}$  satisfies (K\*2);*
2. *all instances of CSO' are negatively valid in  $\mathcal{M}$  iff  $\mathcal{M}$  satisfies (K\*6s);*
3. *all instances of CS are negatively valid in  $\mathcal{M}$  iff  $\mathcal{M}$  satisfies (K\*4w);*
4. *all instances of MP are negatively valid in  $\mathcal{M}$  iff  $\mathcal{M}$  satisfies (K\*3);*
5. *if  $\mathcal{M}$  satisfies (K\*2), (K\*6s), (K\*4w), and (K\*3), then all instances of CA are negatively valid in  $\mathcal{M}$  if  $\mathcal{M}$  satisfies (K\*7);*
6. *if all instances of ID, CSO', CS, and MP are negatively valid in  $\mathcal{M}$ , then  $\mathcal{M}$  satisfies (K\*7) if all instances of CA are negatively valid in  $\mathcal{M}$ ;*
7. *if  $\mathcal{M}$  satisfies (K\*2), (K\*6s), (K\*4w), and (K\*3), then all instances of CV are negatively valid in  $\mathcal{M}$  if  $\mathcal{M}$  satisfies (K\*L);*
8. *if all instances of ID, CSO', CS, and MP are negatively valid in  $\mathcal{M}$ , then  $\mathcal{M}$  satisfies (K\*L) if all instances of CV are negatively valid in  $\mathcal{M}$ .*

Theorem 24 and Lemma 25 allow the soundness and completeness result of Theorem 23 to be extended to yield the following:

**THEOREM 26.** *A formula is a theorem of  $\mathbf{VC}$  iff it is negatively valid in all Gärdenfors pBRMs that satisfy  $(K^*2)$ ,  $(K^*3)$ ,  $(K^*4w)$ ,  $(K^*6s)$ ,  $(K^*7)$ , and  $(K^*L)$ .*

This theorem shows that if  $\mathbf{VC}$  is translated into a theory of belief revision on Gärdenfors pBRMs using that version of the Ramsey test which is built into the notion of a Gärdenfors pBRM, then the resulting theory of belief revision is defined by  $(K^*1)$  (which is built into the definition of a pBRM),  $(K^*2)$ ,  $(K^*3)$ ,  $(K^*4w)$ ,  $(K^*6s)$ ,  $(K^*7)$ , and  $(K^*L)$ . The absence of  $(K^*5w)$  should not be surprising, given Theorem 13. Since in a Gärdenfors pBRM  $(K^*3)$ ,  $(K^*4w)$ , and  $(\text{DEF}+)$  imply  $(K^*T)$ , and since  $(K^*T)$  together with  $(\text{DEF}+)$ ,  $(K^*6s)$  and  $(K^*8)$  imply  $(K^*P)$ , and given Theorem 8, the absence of  $(K^*8)$  should not be surprising.

#### *A conditional logic that approximates AGM belief revision*

Whereas Gärdenfors [1978] sets out to find epistemic models for Lewis's system  $\mathbf{VC}$  of conditional logic, Arló-Costa [1995] sets out to find a system of conditional logic whose primitive belief revision models are defined at least approximately by the AGM belief revision postulates for transitive relational partial meet contraction (see [Gärdenfors, 1988], Chapters 3–4). The result is the system  $\mathcal{EF}$ , which is defined only on languages of type  $\mathcal{L}_1$  (languages of *flat* conditionals).  $\mathcal{EF}$  has the following axioms and rules, where  $\phi$ ,  $\psi$ , and  $\chi$  are conditional free:

#### *Axiom schemata*

Taut: All truth-functional tautologies

ID:  $\phi > \phi$

MP:  $(\phi > \psi) \rightarrow (\phi \rightarrow \psi)$

CC:  $[(\phi > \psi) \wedge (\phi > \chi)] \rightarrow [\phi > (\psi \wedge \chi)]$

CA:  $[(\phi > \chi) \wedge (\psi > \chi)] \rightarrow [(\phi \vee \psi) > \chi]$

CV:  $[(\phi > \psi) \wedge \neg(\phi > \neg\chi)] \rightarrow [(\phi \wedge \chi) > \psi]$

CN:  $\phi > \top$

CD:  $\neg(\phi > \perp)$  for all non-tautologous  $\phi$ .

*Rules of inference*

Modus Ponens: From  $\phi$  and  $\phi \rightarrow \psi$  to infer  $\psi$ .

RCM: From  $\psi \rightarrow \chi$  to infer  $(\phi > \psi) \rightarrow (\phi > \chi)$ .

RCEA: From  $\phi \leftrightarrow \psi$  to infer  $(\phi > \chi) \leftrightarrow (\psi > \chi)$ .

One obvious difference between **VC** and Arló-Costa's  $\mathcal{EF}$  is that  $\mathcal{EF}$  is defined for type  $\mathcal{L}_1$  languages only whereas **VC** is defined for type  $\mathcal{L}_2$  languages. Another difference is that **CS**, an axiom of **VC**, is not a theorem of  $\mathcal{EF}$ . A third difference is that **CD**, an axiom of  $\mathcal{EF}$ , is not a theorem of **VC**.

Arló-Costa's epistemic semantics for  $\mathcal{EF}$  is crucially different from Gärdenfors' epistemic semantics for **VC** in that the semantics of  $\mathcal{EF}$  is defined in terms of positive validity over Arló-Costa/Levi pBRMs rather than in terms of negative validity over Gärdenfors pBRMs. Positive and negative validity coincide in Gärdenfors pBRMs (see Proposition 18) but not in Arló-Costa/Levi pBRMs (see Theorem 22). Which notion of validity should then be adopted? Arló-Costa and Levi argue that positive validity should be adopted rather than negative validity both because positive validity is more intuitive and because in Arló-Costa/Levi models, which satisfy the Negative Ramsey Test favored by Arló-Costa and Levi, the inference rule *modus ponens* does not preserve negative validity.<sup>46</sup>

Consider a type- $\mathcal{L}_1$  language  $L$ ; relative to  $L$  the logical system **Flat CM** is the smallest set of formulas of  $L$  that contains all instances of the axiom schemata of Gärdenfors' **CM** and is closed under the rules of **CM**. Note that  $\mathcal{EF}$  is an extension of **Flat CM**. Arló-Costa [1995] proves the completeness of  $\mathcal{EF}$  with respect to an epistemic semantics (based on positive validity) by proving a result equivalent to Theorem 31 below. We begin with a series of results to be used as lemmas for Theorem 31:<sup>47</sup>

**THEOREM 27.** *A formula of any type  $\mathcal{L}_1$  language  $L$  is a theorem of **Flat CM** iff it is positively valid in every Arló-Costa/Levi pBRM defined on  $L$ .*

**THEOREM 28.** *Let  $\mathcal{CM}^+$  be the result of extending **Flat CM** by adding the rule (RCEA) (restricted to the conditionals of a type- $\mathcal{L}_1$  language). A formula is derivable in  $\mathcal{CM}^+$  iff it is positively valid in the class of all Arló-Costa/Levi pBRMs that satisfy  $(K^*6)$ .*

**THEOREM 29.** *Let  $\mathcal{CMU}^+$  be the result of extending  $\mathcal{CM}^+$  by adding  $\neg(\phi > \perp)$  for every non-tautologous conditional-free  $\phi$ . A formula is derivable in  $\mathcal{CMU}^+$  iff it is positively valid in the class of all Arló-Costa/Levi pBRMs that satisfy  $(K^*6)$  and  $(K^*C)$ .*

<sup>46</sup>See [Arló-Costa and Levi, 1996], pp. 239-240.

<sup>47</sup>Our formulation of these results reflects the organization found in [Arló-Costa and Levi, 1996].

LEMMA 30. *Where  $\mathcal{M}$  is any Arló-Costa/Levi pBRM,*

1. *all instances of ID are positively valid in  $\mathcal{M}$  iff  $\mathcal{M}$  satisfies (K\*2);*
2. *all instances of MP are positively valid in  $\mathcal{M}$  iff  $\mathcal{M}$  satisfies (K\*3);*
3. *all instances of CA are positively valid in  $\mathcal{M}$  iff  $\mathcal{M}$  satisfies (K\*7');*
4. *if all instances of ID are positively valid in  $\mathcal{M}$ , then all instances of CV are positively valid in  $\mathcal{M}$  if  $\mathcal{M}$  satisfies (K\*8);*

Theorems 27, 28, and 29, together with Lemma 30 and the fact that (K\*7) and (K\*7') are equivalent in any pBRM that satisfies (K\*2) and (K\*6), yield the following completeness theorem for  $\mathcal{EF}$ :<sup>48</sup>

THEOREM 31. *A formula of any type  $\mathcal{L}_1$  language  $L$  is a theorem of  $\mathcal{EF}$  iff it is positively valid in every Arló-Costa/Levi pBRM defined on  $L$  satisfying (K\*2), (K\*3), (K\*C), (K\*6), (K\*7), and (K\*8).*

Postulates (K\*1), (K\*2), (K\*3), (K\*4), (K\*5), (K\*6), (K\*7), and (K\*8) jointly capture that notion of revision that is derivable via the Levi Identity (LI) from the AGM notion of transitively relational partial meet contraction (*AGM Revision*, for short).<sup>49</sup> Since (K\*1) holds in all pBRMs,  $\mathcal{EF}$  comes very close to capturing AGM Revision, but (K\*1) and the postulates mentioned in Theorem 31 define a notion of revision ( $\mathcal{EF}$  Revision, for short) that is strictly weaker than AGM revision in two respects.

First, whereas AGM revision includes (K\*4),  $\mathcal{EF}$  revision does not. It turns out that (K\*4) does not correspond to the positive validity of any type- $\mathcal{L}_1$  formula. Still, (K\*4) does correspond to a certain positive entailment, as Arló-Costa [1995] shows:

PROPOSITION 32. *If  $\mathcal{M}$  is an Arló-Costa/Levi pBRM defined on a type- $\mathcal{L}_1$  language  $L$ , then  $\mathcal{M}$  satisfies (K\*4) iff*

$$\{\phi \rightarrow \psi, \neg(\top > \neg\phi)\} \models_+ \phi > \psi$$

*holds in  $\mathcal{M}$  for all conditional-free formulas  $\phi$  and  $\psi$  of  $L$ .*

This result is equivalent to OBSERVATION 3.16 of [Arló-Costa, 1995]. Note that Proposition 32 does *not* establish conditions for the positive validity of

$$(\phi \rightarrow \psi) \rightarrow [\neg(\top > \neg\phi) \rightarrow (\phi > \psi)].$$

But if nesting of conditionals is allowed, then (K\*4) can be associated with the positive validity of nested conditionals of the form  $[(\phi \rightarrow \psi) \wedge \neg(\top >$

<sup>48</sup>Arló-Costa [1995] notes that Theorems 27 and 29 and Lemma 30 suffice to yield completeness theorem for the type- $\mathcal{L}_1$  fragment of David Lewis' system VW.

<sup>49</sup>See, for example, [Gärdenfors, 1988], Chapters 3 and 4.

$\neg\phi$ )  $>$  ( $\phi > \psi$ ) (see THEOREM 8.1 and OBSERVATION 8.3 in [Arló-Costa, 1995]). In general,  $\Gamma \cup \{\phi\} \models_+ \psi$  is not equivalent to  $\Gamma \models_+ \phi > \psi$  in an Arló-Costa/Levi pBRM, but this equivalence does hold for certain  $\phi$  and  $\psi$  when  $\Gamma = \emptyset$  (see OBSERVATION 3.17 of [Arló-Costa, 1995]).

A second difference between  $\mathcal{EF}$  revision and AGM revision is this: whereas AGM Revision includes (K\*5), which entails (K\*5w),  $\mathcal{EF}$  revision includes neither (K\*5) nor (K\*5w) but instead includes (K\*C). The only difference between (K\*5w) and (K\*C) is that (K\*5w) places a constraint on the revision of all belief sets that (K\*C) places just on the revision of consistent belief sets. In particular, where  $\phi$  is nontautologous, (K\*5w) requires  $(K_\perp)_\phi^*$  to be distinct from  $K_\perp$  (and therefore, actually, a contraction of  $K_\perp$ ), whereas (K\*C) implies no such requirement. Theorem 29 reveals that (K\*C) is secured in Arló-Costa/Levi pBRMs via (pNRT) and the positive validity of negated conditionals of the form  $\neg(\phi > \perp)$ , where  $\phi$  is nontautologous. These negated conditionals also belong to  $K_\perp$ , of course, but allowing  $K$  to take  $K_\perp$  as a value in (pNRT) is not an option. Allowing  $K$  to take  $K_\perp$  as a value in (pPRT) also does not help: Theorem 13 shows that (K\*5w) and a consistent underlying logic cannot be combined with the positive Ramsey test in that case. Still, leaving aside the revision of  $K_\perp$ , it is true, as Arló-Costa [1995] has shown, that AGM revision of *nonabsurd* belief sets can be specified in terms of positive validity in a type- $\mathcal{L}_2$  language or in terms of positive validity and positive entailment in a type- $\mathcal{L}_1$  language.

### 3 OTHER TOPICS

Our discussion of the major kinds of conditionals is far from exhaustive. We have looked at several different approaches to the problem of providing an adequate formal semantics and logic for various kinds of conditionals without being able to demonstrate that one approach is clearly superior to all the others. Furthermore, there are many problems involved in the analysis of conditionals which we either have not discussed at all or have only just mentioned in passing. In this section we will look at several of these, giving each the very briefest attention.

One issue which has received much attention is the relationship between conditionals and probability. Stalnaker [1970] proposed that the probability that a conditional is true should be identical with the standard conditional probability. Lewis demonstrates in [Lewis, 1976], however, that this assumption can only be true if we restrict our probability functions to those which assign only a small finite number of distinct values to propositions. Stalnaker [1976] provides a different proof for a similar result, a proof which does not depend upon certain assumptions which Lewis used and which some investigators have questioned. Van Fraassen [1976] avoids

these Triviality Results for a weakened, non-classical version of Stalnaker's conditional logic **C2**. Lewis [1976] shows, however, that we can embrace a result which resembles Stalnaker's while avoiding the Triviality Result. Lewis's suggestion depends upon a technique which he calls imaging. This technique, which provides an alternative method for determining conditional probabilities, requires that in conditionalising a probability assignment with respect to  $\phi$ , i.e. in modifying the assignment in a way which produces a new assignment which assigns probability 1 to  $\phi$ , all the probability which was originally assigned to each  $\neg\phi$ -world  $i$  would be transferred to the  $\phi$ -world closest to  $i$ . Lewis demonstrates that if we accept Stalnaker's semantics and if we assign conditional probabilities in this new, non-standard way, then the probability that a conditional is true turns out to be identical with the conditional probability even when the probabilities of truth for conditionals take on infinitely many different values. Lewis's imaging techniques can be adapted to semantics other than Stalnaker's. Nute [1980b] adapts Lewis's imaging technique to class selection function semantics, producing a notion of subjunctive probability which differs from both the standard conditional probability and the probability that the corresponding conditional is true. While promising in some ways, Nute's account is extremely cumbersome. Gärdenfors [1982] presents a generalized form of imaging and shows that conditional probability cannot be described even in terms of generalized imaging. Other papers on conditionals and probability include [Döring, 1994; Fetzer and Nute, 1979; Fetzer and Nute, 1980; Hájek, 1994; Hall, 1994; Lance, 1991; Lewis, 1981b; Lewis, 1986; McGee, 1989; Nute, 1981a; Stalnaker and Jeffrey, 1994]. For a careful and comprehensive survey of results relating the probabilities of conditionals to conditional probabilities see [Hájek and Hall, 1994].

The relationship between causation and conditionals has certainly not been overlooked either. Many authors like Jackson [1977] and Kwart [1980; 1986] assign a special role to causation in their analyses of counterfactual conditionals. Others like Lewis [1973a] and Swain [1978] attempt to provide analyses of causation in terms of counterfactual dependence. Still others like Fetzer and Nute [1979; 1980] have tried to develop a semantics for a special kind of causal conditional. These special causal conditionals have then been employed in the formulation of a single-case propensity interpretation of law statements.

Conditional logic also has applications in deontic logic (see, for example, [Hilpinen, 1981]), in decision theory (see, for example, [Gibbard and Harper, 1981; Stalnaker, 1981a]), and in nonmonotonic logic (for a summary of some of this work, see [Nute, 1994]). In addition, there has been significant attention in recent years to the issue of whether so-called future indicative conditionals (e.g. 'If Oswald doesn't shoot President Kennedy, then someone else will') should be classified as indicative or as subjunctive (see, for example, [Bennett, 1988; Bennett, 1995; Dudman, 1984; Dudman, 1989;

Dudman, 1994; Jackson, 1990]). For a careful and comprehensive review of this and other recent topics of discussion see [Edgington, 1995]. It is not possible in this essay to discuss or even to list all of the material that can be found in the literature on conditional logic and its applications.

#### 4 LIST OF SOME IMPORTANT RULES, THESES, AND LOGICS

In this section we collect some of the most important rules and theses of conditional logic together with definitions for a few of the better known conditional logics.

##### *Rules*

RCEC: from  $\phi \leftrightarrow \psi$ , to infer  $(\chi > \phi) \leftrightarrow (\chi > \psi)$ .

RCK: from  $(\phi_1 \wedge \dots \wedge \phi_n) \rightarrow \psi$ , to infer  $[(\chi > \phi_1)(\wedge \dots (\chi > \phi_n))] \rightarrow (\chi > \psi)$ ,  $n \geq 0$ .

RCEA: from  $\phi \leftrightarrow \psi$ , to infer  $(\phi > \chi) \leftrightarrow (\psi > \chi)$ .

RCE: from  $\phi \rightarrow \psi$ , to infer  $\phi > \psi$ .

RCM: from  $\psi \rightarrow \chi$ , to infer  $(\phi > \psi) \rightarrow (\phi > \chi)$ .

RR: from  $\chi \rightarrow (\phi > \psi)$  infer  $(\chi \circ \phi) \rightarrow \psi$ , and from  $(\chi \circ \phi) \rightarrow \psi$  infer  $\chi \rightarrow (\phi > \psi)$ .

##### *Theses*

Transitivity:  $[(\phi > \psi) \wedge (\chi > \phi)] \rightarrow (\chi > \psi)$

Contraposition:  $(\phi > \neg\psi) \rightarrow (\psi > \neg\phi)$

Strengthening Antecedents:  $(\phi > \psi) \rightarrow [(\phi \wedge \chi) > \psi]$

ID:  $\phi > \phi$

MP:  $(\phi > \psi) \rightarrow (\phi \rightarrow \psi)$

MOD:  $(\neg\phi > \phi) \rightarrow (\psi > \phi)$

CSO:  $[(\phi > \psi) \wedge (\psi > \phi)] \rightarrow [(\phi > \chi) \leftrightarrow (\psi > \chi)]$

CSO':  $[(\phi > \psi) \wedge (\psi > \phi)] \rightarrow [(\phi > \chi) \rightarrow (\psi > \chi)]$

CV:  $[(\phi > \psi) \wedge \neg(\phi > \neg\chi)] \rightarrow [(\phi \wedge \chi) > \psi]$

CEM:  $(\phi > \psi) \vee (\phi > \neg\psi)$

CS:  $(\phi \wedge \psi) \rightarrow (\phi > \psi)$



- CC:  $[(\phi > \psi) \wedge (\phi > \chi)] \rightarrow [\phi > \psi \wedge \chi]$   
 CM:  $[\phi > (\psi \wedge \chi)] \rightarrow [(\phi > \psi) \wedge (\phi > \chi)]$   
 CA:  $[(\phi > \psi) \wedge (\chi > \psi)] \rightarrow [(\phi \vee \chi) > \psi]$   
 SDA:  $[(\phi \vee \psi) > \chi] \rightarrow [(\phi > \chi) \wedge (\psi > \chi)]$   
 CN:  $\phi > \top$   
 CT:  $(\neg\phi > \perp) \rightarrow \phi$   
 CU:  $\neg(\phi > \perp) \rightarrow ((\phi > \perp) > \perp)$   
 CD:  $\neg(\phi > \perp)$  for all non-tautologous  $\phi$ .

Recall that in Section 1 we defined a conditional logic as any collection  $\overline{L}$  of sentences formed in the usual way from the symbols of classical sentential logic together with a conditional operator  $>$ , such that  $\overline{L}$  is closed under *modus ponens* and  $\overline{L}$  contains every tautology. We now modify this definition as follows, adopting the terminology of Section 2, to take different language types for conditional logic into account: let a conditional logic on a Boolean language  $L$  of type  $\mathcal{L}_1$  or type  $\mathcal{L}_2$  be any collection  $\overline{L}$  of sentences of  $L$  such that  $\overline{L}$  is closed under *modus ponens* and  $\overline{L}$  contains every tautology.

*Logics for full conditional languages*

For a given Boolean language  $L$  of type  $\mathcal{L}_2$ , each of the following is the smallest conditional logic on  $L$  closed under all the rules and containing all the theses associated with it below.

- CM:** RCM, CC, CN  
**VW:** RCEC, RCK; ID, MOD, CSO, MP, CV  
**SS:** RCEC, RCK; ID, MOD, CSO, MP, CA, CS  
**VC:** RCEC, RCK; ID, MOD, CSO, MP, CV, CS  
**VCU:** RCEC, RCK; ID, MOD, CSO, MP, CV, CS, CT, CU  
**VCU<sup>2</sup>:** RCEC, RCK, RR; ID, MOD, CSO, MP, CV, CS, CT, CU (with  $\circ$  as an additional binary operator)  
**C2:** RCEC, RCK; ID, MOD, CSO, MP, CV, CEM

Neither of **VW** and **SS** is an extension of the other, and neither of **VCU** and **C2** is an extension of the other. **VCU<sup>2</sup>** is an extension of **VCU**, and **C2** and **VCU** are both extensions of **VC**, which is an extension of both **VW** and **SS**. **VW** and **SS** are both extensions of **CM**. For the definitions

of several weaker conditional logics, see [Lewis, 1973b; Chellas, 1975; Nute, 1980b].

*Logics for languages of “flat” conditionals*

If  $L$  is a Boolean language of type  $\mathcal{L}_1$ , then each of the following logics is the smallest conditional logic on  $L$  closed under all the rules and containing all the theses associated with it below.

**Flat CM:** RCM, CC, CN

**Flat VW:** RCEC, RCK; ID, MOD, CSO, MP, CV

**Flat VC:** RCEC, RCK; ID, MOD, CSO, MP, CV, CS

$\mathcal{EF}$ : RCM, RCEA, ID, MP, CC, CA, CV, CN, CD

**Flat CM** is contained in **Flat VW**, which is contained in both **Flat VC** and  $\mathcal{EF}$ , but neither of **Flat VC** and  $\mathcal{EF}$  is contained in the other. For a discussion of the logic of flat conditionals aimed at being as true as possible to Ramsey’s ideas, see [Levi, 1996], Chapter 4.

*Acknowledgements*

Sections 1, 3 and 4 are primarily the work of the first author, but revised from the first edition of this *Handbook* with input from the second author. Section 2 is primarily the work of the second author.

We are grateful to Lennart Åqvist, Horacio Arló-Costa, Ermanno Ben-civenga, John Burgess, David Butcher, Michael Dunn, Dov Gabbay, Christopher Gauker, Franz Guenther, Hans Kamp, David Lewis and Christian Rohrer for their helpful comments and suggestions on material contained in this paper. We are also grateful to Richmond Thomason for help in assembling our list of references. Finally, we thank Kluwer and the editor of the *Journal of Philosophical Logic* for permission to use material from [Nute, 1981b] in this article.

Donald Nute  
*University of Georgia, USA.*

Charles B. Cross  
*University of Georgia, USA.*

BIBLIOGRAPHY

- [Adams, 1966] E. Adams. Probability and the logic of conditionals. In J. Hintikka and P. Suppes, editors, *Aspects of Inductive Logic*. North Holland, Amsterdam, 1966.  
[Adams, 1975a] E. Adams. Counterfactual conditionals and prior probabilities. In A. Hooker and W. Harper, editors, *Proceedings of International Congress on the Foundations of Statistics*. Reidel, Dordrecht, 1975.

- [Adams, 1975b] E. Adams. *The Logic of Conditionals; An Application of Probability to Deductive Logic*. Reidel, Dordrecht, 1975.
- [Adams, 1977] E. Adams. A note on comparing probabilistic and modal logics of conditionals. *Theoria*, 43:186–194, 1977.
- [Adams, 1981] E. Adams. Transmissible improbabilities and marginal essentialness of premises in inferences involving indicative conditionals. *Journal of Philosophical Logic*, 10:149–177, 1981.
- [Adams, 1995] E. Adams. Remarks on a theorem of McGee. *Journal of Philosophical Logic*, 24:343–348, 1995.
- [Adams, 1996] E. Adams. Four probability preserving properties of inferences. *Journal of Philosophical Logic*, 25:1–24, 1996.
- [Adams, 1997] E. Adams. *A Primer of Probability Logic*. Cambridge University Press, Cambridge, England, 1997.
- [Alchourrón *et al.*, 1985] C. Alchourrón, P. Gärdenfors, and D. Makinson. On the logic of theory change: partial meet contraction and revision functions. *The Journal of Symbolic Logic*, 50:510–530, 1985.
- [Alchourrón, 1994] C. Alchourrón. Philosophical foundations of deontic logic and the logic of defeasible conditionals. In J.J. Meyer and R.J. Wieringa, editors, *Deontic Logic in Computer Science: Normative System Specification*, pages 43–84. John Wiley and Sons, New York, 1994.
- [Appiah, 1984] A. Appiah. Generalizing the probabilistic semantics of conditionals. *Journal of Philosophical Logic*, 13:351–372, 1984.
- [Appiah, 1985] A. Appiah. *Assertion and Conditionals*. Cambridge University Press, Cambridge, England, 1985.
- [Åqvist, 1973] Lennart Åqvist. Modal logic with subjunctive conditionals and dispositional predicates. *Journal of Philosophical Logic*, 2:1–76, 1973.
- [Arló-Costa and Levi, 1996] H. Arló-Costa and I. Levi. Two notions of epistemic validity: epistemic models for Ramsey’s conditionals. *Synthese*, 109:217–262, 1996.
- [Arló-Costa and Segerberg, 1998] H. Arló-Costa and K. Segerberg. Conditionals and hypothetical belief revision (abstract). *Theoria*, 1998. Forthcoming.
- [Arló-Costa, 1990] H. Arló-Costa. Conditionals and monotonic belief revisions: the success postulate. *Studia Logica*, 49:557–566, 1990.
- [Arló-Costa, 1995] H. Arló-Costa. Epistemic conditionals, snakes, and stars. In G. Crocco, L. Fariñas del Cerro, and A. Herzig, editors, *Conditionals: From Philosophy to Computer Science*. Oxford University Press, Oxford, 1995.
- [Arló-Costa, 1998] H. Arló-Costa. Belief revision conditionals: Basic iterated systems. *Annals of Pure and Applied Logic*, 1998. Forthcoming.
- [Asher and Morreau, 1991] N. Asher and M. Morreau. Commonsense entailment: a modal theory of nonmonotonic reasoning. In J. Mylopoulos and R. Reiter, editors, *Proceedings of the Twelfth International Joint Conference on Artificial Intelligence*, Los Altos, California, 1991. Morgan Kaufmann.
- [Asher and Morreau, 1995] N. Asher and M. Morreau. What some generic sentences mean. In Gregory Carlson and Francis Jeffrey Pelletier, editors, *The Generic Book*. Chicago University Press, Chicago, IL, 1995.
- [Balke and Pearl, 1994a] A. Balke and J. Pearl. Counterfactual probabilities: Computational methods, bounds, and applications. In R. Lopez de Mantaras and D. Poole, editors, *Uncertainty in Artificial Intelligence 10*. Morgan Kaufmann, San Mateo, California, 1994.
- [Balke and Pearl, 1994b] A. Balke and J. Pearl. Counterfactuals and policy analysis in structural models. In P. Besnard and S. Hanks, editors, *Uncertainty in Artificial Intelligence 11*. Morgan Kaufmann, San Francisco, 1994.
- [Balke and Pearl, 1994c] A. Balke and J. Pearl. Probabilistic evaluation of counterfactual queries. In B. Hayes-Roth and R. Korf, editors, *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Menlo Park, California, 1994. American Association for Artificial Intelligence, AAAI Press.
- [Barwise, 1986] J. Barwise. Conditionals and conditional information. In E. Traugott, A. ter Meulen, J. Reilly, and C. Ferguson, editors, *On Conditionals*. Cambridge University Press, Cambridge, England, 1986.

- [Bell, 1988] J. Bell. *Predictive Conditionals, Nonmonotonicity, and Reasoning About the Future*. Ph.D. dissertation, University of Essex, Colchester, 1988.
- [Benferat *et al.*, 1997] S. Benferat, D. Dubois, and H. Prade. Nonmonotonic reasoning, conditional objects, and possibility theory. *Artificial Intelligence*, 92:259–276, 1997.
- [Bennett, 1974] J. Bennett. Counterfactuals and possible worlds. *Canadian Journal of Philosophy*, 4:381–402, 1974.
- [Bennett, 1982] J. Bennett. Even if. *Linguistics and Philosophy*, 5:403–418, 1982.
- [Bennett, 1984] J. Bennett. Counterfactuals and temporal direction. *Philosophical Review*, 43:57–91, 1984.
- [Bennett, 1988] J. Bennett. Farewell to the phlogiston theory of conditionals. *Mind*, 97:509–527, 1988.
- [Bennett, 1995] J. Bennett. Classifying conditionals: the traditional way is right. *Mind*, 104:331–354, 1995.
- [Bigelow, 1976] J. C. Bigelow. If-then meets the possible worlds. *Philosophia*, 6:215–236, 1976.
- [Bigelow, 1980] J. Bigelow. Review of [Pollock, 1976]. *Linguistics and Philosophy*, 4:129–139, 1980.
- [Blue, 1981] N. A. Blue. A metalinguistic interpretation of counterfactual conditionals. *Journal of Philosophical Logic*, 10:179–200, 1981.
- [Boutilier and Goldszmidt, 1995] C. Boutilier and M. Goldszmidt. On the revision of conditional belief sets. In G. Crocco, L. Fariñas del Cerro, and A. Herzig, editors, *Conditionals: From Philosophy to Computer Science*. Oxford University Press, Oxford, 1995.
- [Boutilier, 1990] C. Boutilier. Conditional logics of normality as modal systems. In T. Dietterich and W. Swartout, editors, *Proceedings of the Eighth National Conference on Artificial Intelligence*, Menlo Park, California, 1990. American Association for Artificial Intelligence, AAAI Press.
- [Boutilier, 1992] C. Boutilier. Conditional logics for default reasoning and belief revision. Technical Report KRR-TR-92-1, Computer Science Department, University of Toronto, Toronto, Ontario, 1992.
- [Boutilier, 1993a] C. Boutilier. Belief revision and nested conditionals. In R. Bajcsy, editor, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, San Mateo, California, 1993. Morgan Kaufmann.
- [Boutilier, 1993b] C. Boutilier. A modal characterization of defeasible deontic conditionals and conditional goals. In *Working Notes of the AAAI Spring Symposium on Reasoning about Mental States*, Menlo Park, California, 1993. American Association for Artificial Intelligence.
- [Boutilier, 1993c] C. Boutilier. Revision by conditional beliefs. In R. Fikes and W. Lehnert, editors, *Proceedings of the Eleventh National Conference on Artificial Intelligence*, Menlo Park, California, 1993. American Association for Artificial Intelligence, AAAI Press.
- [Boutilier, 1996] C. Boutilier. Iterated revision and minimal change of conditional beliefs. *Journal of Philosophical Logic*, 25:263–305, 1996.
- [Bowie, 1979] G. Lee Bowie. The similarity approach to counterfactuals: some problems. *Noûs*, 13:477–497, 1979.
- [Burgess, 1979] J. P. Burgess. Quick completeness proofs for some logics of conditionals. *Notre Dame Journal of Formal Logic*, 22:76–84, 1979.
- [Burgess, 1984] J. P. Burgess. Chapter II.2: Basic tense logic. In *Handbook of Philosophical Logic*. Reidel, Dordrecht, 1984.
- [Burks, 1951] A. W. Burks. The logic of causal propositions. *Mind*, 60:363–382, 1951.
- [Butcher, 1978] D. Butcher. *Subjunctive conditional modal logic*. Ph.D. dissertation, Stanford, 1978.
- [Butcher, 1983a] D. Butcher. Consequent-relative subjunctive implication, 1983. Unpublished.
- [Butcher, 1983b] D. Butcher. An incompatible pair of subjunctive conditional modal axioms. *Philosophical Studies*, 44:71–110, 1983.
- [Chellas, 1975] B. F. Chellas. Basic conditional logic. *Journal of Philosophical Logic*, 4:133–153, 1975.

- [Chisholm, 1946] R. Chisholm. The contrary-to-fact conditional. *Mind*, 55:289–307, 1946.
- [Clark, 1971] M. Clark. Ifs and hooks. *Analysis*, 32:33–39, 1971.
- [Costello, 1996] T. Costello. Modeling belief change using counterfactuals. In L. Carlucci Aiello, J. Doyle, and S. Shapiro, editors, *KR'96: Principles of Knowledge Representation and Reasoning*. Morgan Kaufmann, San Francisco, California, 1996.
- [Creary and Hill, 1975] L. G. Creary and C. S. Hill. Review of [Lewis, 1973b]. *Philosophy of Science*, 43:431–344, 1975.
- [Crocco and Fariñas del Cerro, 1996] G. Crocco and L. Fariñas del Cerro. Counterfactuals: Foundations for nonmonotonic inferences. In A. Fuhrmann and H. Rott, editors, *Logic, Action, and Information: Essays on Logic in Philosophy and Artificial Intelligence*. Walter de Gruyter, Berlin, 1996.
- [Cross and Thomason, 1987] C. Cross and R. Thomason. Update and conditionals. In Z. Ras and M. Zemankova, editors, *Methodologies for Intelligent Systems*. North-Holland, Amsterdam, 1987.
- [Cross and Thomason, 1992] C. Cross and R. Thomason. Conditionals and knowledge-base update. In P. Gärdenfors, editor, *Cambridge Tracts in Theoretical Computer Science: Belief Revision*, volume 29. Cambridge University Press, Cambridge, England, 1992.
- [Cross, 1985] C. Cross. Jonathan Bennett on ‘even if’. *Linguistics and Philosophy*, 8:353–357, 1985.
- [Cross, 1990a] C. Cross. Belief revision, nonmonotonic reasoning, and the Ramsey test. In H. Kyburg and R. Loui, editors, *Knowledge Representation and Defeasible Reasoning*. Kluwer, Boston, 1990.
- [Cross, 1990b] C. Cross. Temporal necessity and the conditional. *Studia Logica*, 49:345–363, 1990.
- [Daniels and Freeman, 1980] B. Daniels and J. B. Freeman. An analysis of the subjunctive conditional. *Notre Dame Journal of Formal Logic*, 21:639–655, 1980.
- [Darwiche and Pearl, 1994] A. Darwiche and J. Pearl. On the logic of iterated belief revision. In Ronald Fagin, editor, *Theoretical Aspects of Reasoning About Knowledge: Proceedings of the Fifth Conference*, San Francisco, 1994. Morgan Kaufmann.
- [Davis, 1979] W. Davis. Indicative and subjunctive conditionals. *Philosophical Review*, 88:544–564, 1979.
- [Decew, 1981] J. Decew. Conditional obligation and counterfactuals. *Journal of Philosophical Logic*, 10(1):55–72, 1981.
- [Delgrande, 1988] J. Delgrande. An approach to default reasoning based on a first-order conditional logic: Revised report. *Artificial Intelligence*, 36:63–90, 1988.
- [Delgrande, 1995] J. Delgrande. Syntactic conditional closures for defeasible reasoning. In C. Mellish, editor, *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence*, San Francisco, 1995. Morgan Kaufmann.
- [Döring, 1994] F. Döring. On the probabilities of conditionals. *Philosophical Review*, 103:689–699, 1994. See *Philosophical Review*, 105: 231, 1996, for corrections.
- [Döring, 1997] F. Döring. The Ramsey test and conditional semantics. *Journal of Philosophical Logic*, 26:359–376, 1997.
- [Dudman, 1983] V. Dudman. Tense and time in English verb clusters of the primary pattern. *Australasian Journal of Linguistics*, 3:25–44, 1983.
- [Dudman, 1984] V. Dudman. Parsing ‘if’-sentences. *Analysis*, 44:145–153, 1984.
- [Dudman, 1984a] V. Dudman. Conditional interpretations of if-sentences. *Australian Journal of Linguistics*, 4, 143–204, 1984.
- [Dudman, 1986] V. Dudman. Antecedents and consequents. *Theoria*, 52, 168–199, 1986.
- [Dudman, 1989] V. Dudman. Vive la revolution! *Mind*, 98:591–603, 1989.
- [Dudman, 1991] V. Dudman. V. Dudman. Jackson classifying conditionals. *Analysis*, 51:131–136, 1991.
- [Dudman, 1994] V. Dudman. On conditionals. *Journal of Philosophy*, 91:113–128, 1994.
- [Dudman, 1994a] V. Dudman. Against the indicative. *Australasian Journal of Philosophy*, 72:17–26, 1994.
- [Dudman, 2000] V. Dudman. Classifying ‘conditionals’: the traditional way is wrong, *Analysis*, 60:147, 2000.

- [Edgington, 1995] D. Edgington. On conditionals. *Mind*, 104:235–329, 1995.
- [Eells and Skyrms, 1994] E. Eells and B. Skyrms, editors. *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge University Press, Cambridge, England, 1994.
- [Eiter and Gottlob, 1993] T. Eiter and G. Gottlob. The complexity of nested counterfactuals and iterated knowledge base revision. In R. Bajcsy, editor, *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, San Mateo, California, 1993. Morgan Kaufmann.
- [Ellis *et al.*, 1977] B. Ellis, F. Jackson, and R. Pargetter. An objection to possible worlds semantics for counterfactual logics. *Journal of Philosophical Logic*, 6:355–357, 1977.
- [Ellis, 1978] Brian Ellis. A unified theory of conditionals. *Journal of Philosophical Logic*, 7:107–124, 1978.
- [Farkas and Sugioka, 1987] D. Farkas and Y. Sugioka. Restrictive if/when clauses. *Linguistics and Philosophy*, 6:225–258, 1987.
- [Fetzer and Nute, 1979] J. H. Fetzer and D. Nute. Syntax, semantics and ontology: a probabilistic causal calculus. *Synthese*, 40:453–495, 1979.
- [Fetzer and Nute, 1980] J. H. Fetzer and D. Nute. A probabilistic causal calculus: conflicting conceptions. *Synthese*, 44:241–246, 1980.
- [Fine, 1975] K. Fine. Review of [Lewis, 1973b]. *Mind*, 84:451–458, 1975.
- [Friedman and Halpern, 1994] N. Friedman and J. Halpern. Conditional logics for belief change. In B. Hayes-Roth and R. Korf, editors, *Proceedings of the Twelfth National Conference on Artificial Intelligence*, Menlo Park, California, 1994. American Association for Artificial Intelligence, AAAI Press.
- [Fuhrmann and Levi, 1994] A. Fuhrmann and I. Levi. Undercutting and the Ramsey test for conditionals. *Synthese*, 101:157–169, 1994.
- [Gabbay, 1972] D. M. Gabbay. A general theory of the conditional in terms of a ternary operator. *Theoria*, 38:97–104, 1972.
- [Galles and Pearl, 1997] D. Galles and J. Pearl. An axiomatic characterization of causal counterfactuals. Technical report, Computer Science Department, UCLA, Los Angeles, California, 1997.
- [Gärdenfors *et al.*, 1991] P. Gärdenfors, S. Lindström, M. Morreau, and R. Rabinowicz. The negative Ramsey test: another triviality result. In A. Fuhrmann and M. Morreau, editors, *The Logic of Theory Change*. Cambridge University Press, Cambridge, England, 1991.
- [Gärdenfors, 1978] P. Gärdenfors. Conditionals and changes of belief. In I. Niiniluoto and R. Tuomela, editors, *The Logic and Epistemology of Scientific Change*. North Holland, Amsterdam, 1978.
- [Gärdenfors, 1979] P. Gärdenfors. Even if. In F. V. Jensen, B. H. Mayoh, and K. K. Moller, editors, *Proceedings from 5th Scandinavian Logic Symposium*. Aalborg University Press, Aalborg, 1979.
- [Gärdenfors, 1982] P. Gärdenfors. Imaging and conditionalization. *Journal of Philosophy*, 79:747–760, 1982.
- [Gärdenfors, 1986] P. Gärdenfors. Belief revisions and the Ramsey test for conditionals. *Philosophical Review*, 95:81–93, 1986.
- [Gärdenfors, 1987] P. Gärdenfors. Variations on the Ramsey test: more triviality results. *Studia Logica*, 46:321–327, 1987.
- [Gärdenfors, 1988] P. Gärdenfors. *Knowledge in Flux*. MIT Press, Cambridge, MA, 1988.
- [Gibbard and Harper, 1981] A. Gibbard and W. Harper. Counterfactuals and two kinds of expected utility. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*. Reidel, Dordrecht, 1981.
- [Ginsberg, 1985] M. Ginsberg. Counterfactuals. In A. Joshi, editor, *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, Los Altos, California, 1985. Morgan Kaufmann.
- [Goodman, 1955] N. Goodman. *Fact, Fiction and Forecast*. Harvard, Cambridge, MA, 1955.
- [Grahne and Mendelzon, 1994] G. Grahne and A. Mendelzon. Updates and subjunctive queries. *Information and Computation*, 116:241–252, 1994.

- [Grahne, 1991] G. Grahne. Updates and counterfactuals. In J. Allen, R. Fikes, and E. Sandewall, editors, *Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference*. Morgan-Kaufmann, Los Altos, 1991.
- [Grice, 1967] P. Grice. Logic and conversation, 1967. The William James Lectures, given at Harvard University.
- [Hájek and Hall, 1994] A. Hájek and N. Hall. The hypothesis of the conditional construal of conditional probability. In E. Eells and B. Skyrms, editors, *Probability and Conditionals*. Cambridge University Press, Cambridge, England, 1994.
- [Hájek, 1989] A. Hájek. Probabilities of conditionals—revisited. *Journal of Philosophical Logic*, 18:423–428, 1989.
- [Hájek, 1994] A. Hájek. Triviality on the cheap. In E. Eells and B. Skyrms, editors, *Probability and Conditionals*. Cambridge University Press, Cambridge, England, 1994.
- [Hall, 1994] N. Hall. Back in the CCCP. In E. Eells and B. Skyrms, editors, *Probability and Conditionals*. Cambridge University Press, Cambridge, England, 1994.
- [Hansson, 1992] S. Hansson. In defense of the Ramsey test. *Journal of Philosophy*, 89:499–521, 1992.
- [Harper *et al.*, 1981] W. Harper, R. Stalnaker, and G. Pearce, editors. *Ifs: conditionals, belief, decision, chance, and time*. D. Reidel, Dordrecht, 1981.
- [Harper, 1975] W. Harper. Rational belief change, Popper functions and the counterfactuals. *Synthese*, 30:221–262, 1975.
- [Hausman, 1996] D. Hausman. Causation and counterfactual dependence reconsidered. *Noûs*, 30:55–74, 1996.
- [Hawthorne, 1996] J. Hawthorne. On the logic of nonmonotonic conditionals and conditional probabilities. *Journal of Philosophical Logic*, 25:185–218, 1996.
- [Herzberger, 1979] H. Herzberger. Counterfactuals and consistency. *Journal of Philosophy*, 76:83–88, 1979.
- [Hilpinen, 1981] R. Hilpinen. Conditionals and possible worlds. In G. Floistad, editor, *Contemporary Philosophy: A New Survey*, volume I: Philosophy of Language/Philosophical Logic. Martinus Nijhoff, The Hague, 1981.
- [Hilpinen, 1982] R. Hilpinen. Disjunctive permissions and conditionals with disjunctive antecedents. In I. Niiniluoto and Esa Saarinen, editors, *Proceedings of the Second Soviet–Finnish Logic Conference, Moscow, December 1979*. Acta Philosophica Fennica, 1982.
- [Horgan, 1981] T. Horgan. Counterfactuals and Newcomb’s problem. *Journal of Philosophy*, 78:331–356, 1981.
- [Horty and Thomason, 1991] J. Horty and R. Thomason. Conditionals and artificial intelligence. *Fundamenta Informaticae*, 15:301–324, 1991.
- [Humberstone, 1978] I. L. Humberstone. Two merits of the circumstantial operator language for conditional logic. *Australasian Journal of Philosophy*, 56:21–24, 1978.
- [Hunter, 1980] G. Hunter. Conditionals, indicative and subjunctive. In J. Dancey, editor, *Papers on Language and Logic*. Keele University Library, 1980.
- [Hunter, 1982] G. Hunter. Review of [Nute, 1980b]. *Mind*, 91:136–138, 1982.
- [Jackson, 1977] F. Jackson. A causal theory of counterfactuals. *Australasian Journal of Philosophy*, 55:3–21, 1977.
- [Jackson, 1979] F. Jackson. On assertion and indicative conditionals. *Philosophical Review*, 88:565–589, 1979.
- [Jackson, 1987] F. Jackson. *Conditionals*. Blackwell, Oxford, 1987.
- [Jackson, 1990] F. Jackson. Classifying conditionals. *Analysis*, 50:134–147, 1990.
- [Jackson, 1991] F. Jackson, editor. *Conditionals*. Oxford University Press, Oxford, 1991.
- [Jackson, 1991a] F. Jackson, Classifying conditionals II. *Analysis*, 51:137–143, 1991.
- [Katsuno and Mendelzon, 1992] H. Katsuno and A. Mendelzon. Updates and counterfactuals. In P. Gärdenfors, editor, *Cambridge Tracts in Theoretical Computer Science: Belief Revision*, volume 29. Cambridge University Press, Cambridge, England, 1992.
- [Kim, 1973] J. Kim. Causes and counterfactuals. *Journal of Philosophy*, 70:570–572, 1973.
- [Kratzer, 1979] A. Kratzer. Conditional necessity and possibility. In R. Bauerle, U. Egli, and A. von Stechow, editors, *Semantics from Different Points of View*. Springer-Verlag, Berlin, 1979.

- [Kratzer, 1981] A. Kratzer. Partition and revision: the semantics of counterfactuals. *Journal of Philosophical Logic*, 10:201–216, 1981.
- [Kremer, 1987] M. Kremer. ‘if’ is unambiguous. *Noûs*, 21:199–217, 1987.
- [Kvart, 1980] I. Kvart. Formal semantics for temporal logic and counterfactuals. *Logique et analyse*, 23:35–62, 1980.
- [Kvart, 1986] I. Kvart. *A Theory of Counterfactuals*. Hackett, Indianapolis, 1986.
- [Kvart, 1987] I. Kvart. Putnam’s counterexample to ‘A theory of counterfactuals’. *Philosophical Papers*, 16:235–239, 1987.
- [Kvart, 1991] I. Kvart. Counterfactuals and causal relevance. *Pacific Philosophical Quarterly*, 72:314–337, 1991.
- [Kvart, 1992] I. Kvart. Counterfactuals. *Erkenntnis*, 36:139–179, 1992.
- [Kvart, 1994] I. Kvart. Counterfactual ambiguities, true premises and knowledge. *Synthese*, 100:133–164, 1994.
- [Lance, 1991] M. Lance. Probabilistic dependence among conditionals. *Philosophical Review*, 100:269–276, 1991.
- [Lehmann and Magidor, 1992] D. Lehmann and M. Magidor. What does a conditional knowledge base entail? *Artificial intelligence*, 55:1–60, 1992.
- [Levi, 1977] I. Levi. Subjunctives, dispositions and chances. *Synthese*, 34:423–455, 1977.
- [Levi, 1988] I. Levi. Iteration of conditionals and the Ramsey test. *Synthese*, 76:49–81, 1988.
- [Levi, 1996] I. Levi. *For the Sake of the Argument: Ramsey Test Conditionals, Inductive Inference, and Nonmonotonic Reasoning*. Cambridge University Press, Cambridge, England, 1996.
- [Lewis, 1971] D. Lewis. Completeness and decidability of three logics of counterfactual conditionals. *Theoria*, 37:74–85, 1971.
- [Lewis, 1973a] D. Lewis. Causation. *Journal of Philosophy*, 70:556–567, 1973.
- [Lewis, 1973b] D. Lewis. *Counterfactuals*. Harvard, Cambridge, MA, 1973.
- [Lewis, 1973c] D. Lewis. Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2:418–446, 1973.
- [Lewis, 1976] D. Lewis. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85:297–315, 1976.
- [Lewis, 1977] D. Lewis. Possible world semantics for counterfactuals logics: a rejoinder. *Journal of Philosophical Logic*, 6:359–363, 1977.
- [Lewis, 1979a] D. Lewis. Counterfactual dependence and time’s arrow. *Noûs*, 13:455–476, 1979.
- [Lewis, 1979b] D. Lewis. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359, 1979.
- [Lewis, 1981a] D. Lewis. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, 10:217–234, 1981.
- [Lewis, 1981b] D. Lewis. A subjectivist’s guide to objective change. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*. Reidel, Dordrecht, 1981.
- [Lewis, 1986] D. Lewis. Probabilities of conditionals and conditional probabilities II. *Philosophical Review*, 95:581–589, 1986.
- [Lindström and Rabinowicz, 1992] S. Lindström and W. Rabinowicz. Belief revision, epistemic conditionals, and the Ramsey test. *Synthese*, 91:195–237, 1992.
- [Lindström and Rabinowicz, 1995] S. Lindström and W. Rabinowicz. The Ramsey test revisited. In G. Crocco, L. Fariñas del Cerro, and A. Herzig, editors, *Conditionals: From Philosophy to Computer Science*. Oxford University Press, Oxford, 1995.
- [Lindström, 1996] S. Lindström. The Ramsey test and the indexicality of conditionals: A proposed resolution of Gärdenfors’ paradox. In A. Fuhrmann and H. Rott, editors, *Logic, Action, and Information: Essays on Logic in Philosophy and Artificial Intelligence*. Walter de Gruyter, Berlin, 1996.
- [Loewer, 1976] B. Loewer. Counterfactuals with disjunctive antecedents. *Journal of Philosophy*, 73:531–536, 1976.
- [Loewer, 1978] B. Loewer. Cotenability and counterfactual logics. *Journal of Philosophical Logic*, 8:99–116, 1978.
- [Lowe, 1991] E.J. Lowe. Jackson on classifying conditionals. *Analysis*, 51:126–130, 1991.



- [Makinson, 1989] D. Makinson. General theory of cumulative inference. In M. Reinfrank, J. de Kleer, and M. Ginsberg, editors, *Lecture Notes in Artificial Intelligence: Non-Monotonic Reasoning*, volume 346. Springer-Verlag, Berlin, 1989.
- [Makinson, 1990] D. Makinson. The Gärdenfors impossibility theorem in nonmonotonic contexts. *Studia Logica*, 49:1–6, 1990.
- [Mayer, 1981] J. C. Mayer. A misplaced thesis of conditional logic. *Journal of Philosophical Logic*, 10:235–238, 1981.
- [McDermott, 1996] M. McDermott. On the truth conditions of certain ‘if’-sentences. *The Philosophical Review*, 105:1–37, 1996.
- [Mcgee, 1981] V. Mcgee. Finite matrices and the logic of conditionals. *Journal of Philosophical Logic*, 10:349–351, 1981.
- [McGee, 1985] V. McGee. A counterexample to modus ponens. *Journal of Philosophy*, 82:462–471, 1985.
- [McGee, 1989] V. McGee. Conditional probabilities and compounds of conditionals. *Philosophical Review*, 98:485–541, 1989.
- [McGee, 2000] V. McGee. To tell the truth about conditionals. *Analysis*, 60:107–111, 2000.
- [McKay and Inwagen, 1977] T. McKay and P. Van Inwagen. Counterfactuals with disjunctive antecedents. *Philosophical Studies*, 31:353–356, 1977.
- [Mellor, 1993] D.H. Mellor. How to believe a conditional. *Journal of Philosophy*, 90:233–248, 1993.
- [Moore, 1983] R. Moore. Semantical considerations on nonmonotonic logic. In *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, volume 1. Morgan Kaufman, San Mateo, 1983.
- [Morreau, 1992] M. Morreau. Epistemic semantics for counterfactuals. *Journal of Philosophical Logic*, 21:33–62, 1992.
- [Morreau, 1997] M. Morreau. Fainthearted conditionals. *The Journal of Philosophy*, 94:187–211, 1997.
- [Nute and Mitcheltree, 1982] D. Nute and W. Mitcheltree. Review of [Adams, 1975b]. *Noûs*, 15:432–436, 1982.
- [Nute, 1975a] D. Nute. Counterfactuals. *Notre Dame Journal of Formal Logic*, 16:476–482, 1975.
- [Nute, 1975b] D. Nute. Counterfactuals and the similarity of worlds. *Journal of Philosophy*, 72:73–778, 1975.
- [Nute, 1977] D. Nute. Scientific law and nomological conditionals. Technical report, National Science Foundation, 1977.
- [Nute, 1978a] D. Nute. An incompleteness theorem for conditional logic. *Notre Dame Journal of Formal Logic*, 19:634–636, 1978.
- [Nute, 1978b] D. Nute. Simplification and substitution of counterfactual antecedents. *Philosophia*, 7:317–326, 1978.
- [Nute, 1979] D. Nute. Algebraic semantics for conditional logics. *Reports on Mathematical Logic*, 10:79–101, 1979.
- [Nute, 1980a] D. Nute. Conversational scorekeeping and conditionals. *Journal of Philosophical Logic*, 9:153–166, 1980.
- [Nute, 1980b] D. Nute. *Topics in Conditional Logic*. Reidel, Dordrecht, 1980.
- [Nute, 1981a] D. Nute. Causes, laws and law statements. *Synthese*, 48:347–370, 1981.
- [Nute, 1981b] D. Nute. Introduction. *Journal of Philosophical Logic*, 10:127–147, 1981.
- [Nute, 1981c] D. Nute. Review of [Pollock, 1976]. *Noûs*, 15:212–219, 1981.
- [Nute, 1982 and 1991] D. Nute. Tense and conditionals. Technical report, Deutsche Forschungsgemeinschaft and the University of Georgia, 1982 and 1991.
- [Nute, 1983] D. Nute. Review of [Harper *et al.*, 1981]. *Philosophy of Science*, 50:518–520, 1983.
- [Nute, 1991] D. Nute. Historical necessity and conditionals. *Noûs*, 25:161–175, 1991.
- [Nute, 1994] D. Nute. Defeasible logic. In D. Gabbay and C. Hogger, editors, *Handbook of Logic for Artificial Intelligence and Logic Programming*, volume III. Oxford University Press, Oxford, 1994.

- [Pearl, 1994] J. Pearl. From Adams' conditionals to default expressions, causal conditionals, and counterfactuals. In E. Eells and B. Skyrms, editors, *Probability and Conditionals: Belief Revision and Rational Decision*. Cambridge University Press, Cambridge, England, 1994.
- [Pearl, 1995] J. Pearl. Causation, action, and counterfactuals. In A. Gammerman, editor, *Computational Learning and Probabilistic Learning*. John Wiley and Sons, New York, 1995.
- [Pollock, 1976] J. Pollock. *Subjunctive Reasoning*. Reidel, Dordrecht, 1976.
- [Pollock, 1981] J. Pollock. A refined theory of counterfactuals. *Journal of Philosophical Logic*, 10:239–266, 1981.
- [Pollock, 1984] J. Pollock. *Knowledge and Justification*. Princeton University Press, Princeton, 1984.
- [Posch, 1980] G. Posch. *Zur Semantik der Kontrafaktischen Konditionale*. Narr, Tübingen, 1980.
- [Post, 1981] J. Post. Review of [Pollock, 1976]. *Philosophia*, 9:405–420, 1981.
- [Ramsey, 1990] F. Ramsey. *Philosophical Papers*. Cambridge University Press, Cambridge, England, 1990.
- [Rescher, 1964] N. Rescher. *Hypothetical Reasoning*. Reidel, Dordrecht, 1964.
- [Rott, 1986] H. Rott. Ifs, though, and because. *Erkenntnis*, 25:345–370, 1986.
- [Rott, 1989] H. Rott. Conditionals and theory change: revisions, expansions, and additions. *Synthese*, 81:91–113, 1989.
- [Rott, 1991] H. Rott. A nonmonotonic conditional logic for belief revision. In A. Fuhrmann and M. Morreau, editors, *The Logic of Theory Change*. Cambridge University Press, Cambridge, England, 1991.
- [Sanford, 1992] D. Sanford. *If P then Q*. Routledge, London, 1992.
- [Schlechta and Makinson, 1994] K. Schlechta and D. Makinson. Local and global metrics for the semantics of counterfactual conditionals. *Journal of applied Non-Classical Logics*, 4:129–140, 1994.
- [Segerberg, 1968] K. Segerberg. Propositional logics related to Heyting's and Johanson's. *Theoria*, 34:26–61, 1968.
- [Segerberg, 1989] K. Segerberg. A note on an impossibility theorem of Gärdenfors. *Noûs*, 23:351–354, 1989.
- [Sellars, 1958] W. S. Sellars. Counterfactuals, dispositions and the causal modalities. In Feigl, Scriven, and Maxwell, editors, *Minnesota Studies in the Philosophy of Science*, volume 2. University of Minnesota, Minneapolis, 1958.
- [Slote, 1978] M. A. Slote. Time and counterfactuals. *Philosophical Review*, 87:3–27, 1978.
- [Stalnaker and Jeffrey, 1994] R. Stalnaker and R. Jeffrey. Conditionals as random variables. In E. Eells and B. Skyrms, editors, *Probability and Conditionals*. Cambridge University Press, Cambridge, England, 1994.
- [Stalnaker and Thomason, 1970] R. Stalnaker and R. Thomason. A semantical analysis of conditional logic. *Theoria*, 36:23–42, 1970.
- [Stalnaker, 1968] R. Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in Logical Theory*. American Philosophical quarterly Monograph Series, No. 2, Blackwell, Oxford, 1968. Reprinted in [Harper *et al.*, 1981].
- [Stalnaker, 1970] R. Stalnaker. Probabilities and conditionals. *Philosophy of Science*, 28:64–80, 1970.
- [Stalnaker, 1975] R. Stalnaker. Indicative conditionals. *Philosophia*, 5:269–286, 1975.
- [Stalnaker, 1976] R. Stalnaker. Stalnaker to Van Fraassen. In C. Hooker and W. Harper, editors, *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*. Reidel, Dordrecht, 1976.
- [Stalnaker, 1981a] R. Stalnaker. A defense of conditional excluded middle. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*. Reidel, Dordrecht, 1981.
- [Stalnaker, 1981b] R. Stalnaker. Letter to David Lewis. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*. Reidel, Dordrecht, 1981.
- [Stalnaker, 1984] R. Stalnaker. *Inquiry*. MIT Press, Cambridge, MA, 1984.
- [Swain, 1978] M. Swain. A counterfactual analysis of event causation. *Philosophical Studies*, 34:1–19, 1978.

- [Thomason and Gupta, 1981] R. Thomason and A. Gupta. A theory of conditionals in the context of branching time. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*. Reidel, Dordrecht, 1981.
- [Thomason, 1985] R. Thomason. Note on tense and subjunctive conditionals. *Philosophy of Science*, pages 151–153, 1985.
- [Traugott *et al.*, 1986] E. Traugott, A. ter Meulen, J. Reilly, and C. Ferguson, editors. *On Conditionals*. Cambridge University Press, Cambridge, England, 1986.
- [Turner, 1981] R. Turner. Counterfactuals without possible worlds. *Journal of Philosophical Logic*, 10:453–493, 1981.
- [van Benthem, 1984] J. van Benthem. Foundations of conditional logic. *Journal of Philosophical Logic*, 13:303–349, 1984.
- [Van Fraassen, 1974] B. C. Van Fraassen. Hidden variables in conditional logic. *Theoria*, 40:176–190, 1974.
- [Van Fraassen, 1976] B. C. Van Fraassen. Probabilities of conditionals. In C. Hooker and W. Harper, editors, *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*. Reidel, Dordrecht, 1976.
- [Van Fraassen, 1981] B. C. Van Fraassen. A temporal framework for conditionals and chance. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*. Reidel, Dordrecht, 1981.
- [Veltman, 1976] F. Veltman. Prejudices, presuppositions and the theory of conditionals. In J. Groenendijk and M. Stokhof, editors, *Amsterdam Papers in Formal Grammar*. Vol. 1, Centrale Interfaculteit, Universiteit van Amsterdam, 1976.
- [Veltman, 1985] Frank Veltman. *Logics for Conditionals*. Ph.D. dissertation, University of Amsterdam, Amsterdam, 1985.
- [Warmbröd, 1981] Warmbröd. Counterfactuals and substitution of equivalent antecedents. *Journal of Philosophical Logic*, 10:267–289, 1981.
- [Winslett, 1990] M. Winslett. *Updating Logical Databases*. Cambridge University Press, Cambridge, England, 1990.
- [Woods, 1997] M. Woods. *Conditionals*. Oxford University Press, Oxford, 1997. Published posthumously. Edited by D. Wiggins, with a commentary by D. Edgington.

## DYNAMIC LOGIC

## PREFACE

Dynamic Logic (DL) is a formal system for reasoning about programs. Traditionally, this has meant formalizing correctness specifications and proving rigorously that those specifications are met by a particular program. Other activities fall into this category as well: determining the equivalence of programs, comparing the expressive power of various programming constructs, synthesizing programs from specifications, etc. Formal systems too numerous to mention have been proposed for these purposes, each with its own peculiarities.

DL can be described as a blend of three complementary classical ingredients: first-order predicate logic, modal logic, and the algebra of regular events. These components merge to form a system of remarkable unity that is theoretically rich as well as practical.

The name *Dynamic Logic* emphasizes the principal feature distinguishing it from classical predicate logic. In the latter, truth is *static*: the truth value of a formula  $\varphi$  is determined by a valuation of its free variables over some structure. The valuation and the truth value of  $\varphi$  it induces are regarded as immutable; there is no formalism relating them to any other valuations or truth values. In Dynamic Logic, there are explicit syntactic constructs called *programs* whose main role is to change the values of variables, thereby changing the truth values of formulas. For example, the program  $x := x + 1$  over the natural numbers changes the truth value of the formula “ $x$  is even”.

Such changes occur on a metalogical level in classical predicate logic. For example, in Tarski’s definition of truth of a formula, if  $u : \{x, y, \dots\} \rightarrow \mathbb{N}$  is a valuation of variables over the natural numbers  $\mathbb{N}$ , then the formula  $\exists x \ x^2 = y$  is defined to be true under the valuation  $u$  iff there exists an  $a \in \mathbb{N}$  such that the formula  $x^2 = y$  is true under the valuation  $u[x/a]$ , where  $u[x/a]$  agrees with  $u$  everywhere except  $x$ , on which it takes the value  $a$ . This definition involves a metalogical operation that produces  $u[x/a]$  from  $u$  for all possible values  $a \in \mathbb{N}$ . This operation becomes explicit in DL in the form of the program  $x := ?$ , called a *nondeterministic* or *wildcard assignment*. This is a rather unconventional program, since it is not effective; however, it is quite useful as a descriptive tool. A more conventional way to obtain a square root of  $y$ , if it exists, would be the program

(1)  $x := 0$ ; **while**  $x^2 < y$  **do**  $x := x + 1$ .

In DL, such programs are first-class objects on a par with formulas, complete with a collection of operators for forming compound programs inductively

from a basis of primitive programs. To discuss the effect of the execution of a program  $\alpha$  on the truth of a formula  $\varphi$ , DL uses a modal construct  $\langle \alpha \rangle \varphi$ , which intuitively states, “It is possible to execute  $\alpha$  starting from the current state and halt in a state satisfying  $\varphi$ .” There is also the dual construct  $[\alpha] \varphi$ , which intuitively states, “If  $\alpha$  halts when started in the current state, then it does so in a state satisfying  $\varphi$ .” For example, the first-order formula  $\exists x x^2 = y$  is equivalent to the DL formula  $\langle x := ? \rangle x^2 = y$ . In order to instantiate the quantifier effectively, we might replace the nondeterministic assignment inside the  $\langle \rangle$  with the **while** program (1); over  $\mathbb{N}$ , the two formulas would be equivalent.

Apart from the obvious heavy reliance on classical logic, computability theory and programming, the subject has its roots in the work of [Thiele, 1966] and [Engeler, 1967] in the late 1960’s, who were the first to advance the idea of formulating and investigating formal systems dealing with properties of programs in an abstract setting. Research in program verification flourished thereafter with the work of many researchers, notably [Floyd, 1967], [Hoare, 1969], [Manna, 1974], and [Salwicki, 1970]. The first precise development of a DL-like system was carried out by [Salwicki, 1970], following [Engeler, 1967]. This system was called Algorithmic Logic. A similar system, called Monadic Programming Logic, was developed by [Constable, 1977]. Dynamic Logic, which emphasizes the modal nature of the program/assertion interaction, was introduced by [Pratt, 1976].

Background material on mathematical logic, computability, formal languages and automata, and program verification can be found in [Shoenfield, 1967] (logic), [Rogers, 1967] (recursion theory), [Kozen, 1997a] (formal languages, automata, and computability), [Keisler, 1971] (infinitary logic), [Manna, 1974] (program verification), and [Harel, 1992; Lewis and Papadimitriou, 1981; Davis *et al.*, 1994] (computability and complexity). Much of this introductory material as it pertains to DL can be found in the authors’ text [Harel *et al.*, 2000].

There are by now a number of books and survey papers treating logics of programs, program verification, and Dynamic Logic [Apt and Olderog, 1991; Backhouse, 1986; Harel, 1979; Harel, 1984; Parikh, 1981; Goldblatt, 1982; Goldblatt, 1987; Knijnenburg, 1988; Cousot, 1990; Emerson, 1990; Kozen and Tiuryn, 1990]. In particular, much of this chapter is an abbreviated summary of material from the authors’ text [Harel *et al.*, 2000], to which we refer the reader for a more complete treatment. Full proofs of many of the theorems cited in this chapter can be found there, as well as extensive introductory material on logic and complexity along with numerous examples and exercises.

## 1 REASONING ABOUT PROGRAMS

### 1.1 Programs

For us, a *program* is a recipe written in a formal language for computing desired output data from given input data.

EXAMPLE 1. The following program implements the Euclidean algorithm for calculating the greatest common divisor (gcd) of two integers. It takes as input a pair of integers in variables  $x$  and  $y$  and outputs their gcd in variable  $x$ :

```
while  $y \neq 0$  do  
  begin  
     $z := x \bmod y$ ;  
     $x := y$ ;  
     $y := z$   
  end
```

The value of the expression  $x \bmod y$  is the (nonnegative) remainder obtained when dividing  $x$  by  $y$  using ordinary integer division.

Programs normally use *variables* to hold input and output values and intermediate results. Each variable can assume values from a specific *domain of computation*, which is a structure consisting of a set of data values along with certain distinguished constants, basic operations, and tests that can be performed on those values, as in classical first-order logic. In the program above, the domain of  $x$ ,  $y$ , and  $z$  might be the integers  $\mathbb{Z}$  along with basic operations including integer division with remainder and tests including  $\neq$ . In contrast with the usual use of variables in mathematics, a variable in a program normally assumes different values during the course of the computation. The value of a variable  $x$  may change whenever an assignment  $x := t$  is performed with  $x$  on the left-hand side.

In order to make these notions precise, we will have to specify the programming language and its semantics in a mathematically rigorous way. In this section we give a brief introduction to some of these languages and the role they play in program verification.

### 1.2 States and Executions

As mentioned above, a program can change the values of variables as it runs. However, if we could freeze time at some instant during the execution of the program, we could presumably read the values of the variables at that instant, and that would give us an instantaneous snapshot of all information that we would need to determine how the computation would proceed from that point. This leads to the concept of a *state*—intuitively, an instantaneous description of reality.

Formally, we will define a *state* to be a function that assigns a value to each program variable. The value for variable  $x$  must belong to the domain associated with  $x$ . In logic, such a function is called a *valuation*. At any given instant in time during its execution, the program is thought to be “in” some state, determined by the instantaneous values of all its variables. If an assignment statement is executed, say  $x := 2$ , then the state changes to a new state in which the new value of  $x$  is 2 and the values of all other variables are the same as they were before. We assume that this change takes place instantaneously; note that this is a mathematical abstraction, since in reality basic operations take some time to execute.

A typical state for the gcd program above is  $(15, 27, 0, \dots)$ , where (say) the first, second, and third components of the sequence denote the values assigned to  $x$ ,  $y$ , and  $z$  respectively. The ellipsis “ $\dots$ ” refers to the values of the other variables, which we do not care about, since they do not occur in the program.

A program can be viewed as a transformation on states. Given an initial (input) state, the program will go through a series of intermediate states, perhaps eventually halting in a final (output) state. A sequence of states that can occur from the execution of a program  $\alpha$  starting from a particular input state is called a *trace*. As a typical example of a trace for the program above, consider the initial state  $(15, 27, 0)$  (we suppress the ellipsis). The program goes through the following sequence of states:

$$(15, 27, 0), (15, 27, 15), (27, 27, 15), (27, 15, 15), (27, 15, 12), (15, 15, 12), \\ (15, 12, 12), (15, 12, 3), (12, 12, 3), (12, 3, 3), (12, 3, 0), (3, 3, 0), (3, 0, 0).$$

The value of  $x$  in the last (output) state is 3, the gcd of 15 and 27.

The binary relation consisting of the set of all pairs of the form (input state, output state) that can occur from the execution of a program  $\alpha$ , or in other words, the set of all first and last states of traces of  $\alpha$ , is called the *input/output relation* of  $\alpha$ . For example, the pair  $((15, 27, 0), (3, 0, 0))$  is a member of the input/output relation of the gcd program above, as is the pair  $((-6, -4, 303), (2, 0, 0))$ . The values of other variables besides  $x$ ,  $y$ , and  $z$  are not changed by the program. These values are therefore the same in the output state as in the input state. In this example, we may think of the variables  $x$  and  $y$  as the *input variables*,  $x$  as the *output variable*, and  $z$  as a *work variable*, although formally there is no distinction between any of the variables, including the ones not occurring in the program.

### 1.3 Programming Constructs

In subsequent sections we will consider a number of programming constructs. In this section we introduce some of these constructs and define a few general classes of languages built on them.

In general, programs are built inductively from *atomic programs* and *tests* using various *program operators*.

### While Programs

A popular choice of programming language in the literature on DL is the family of deterministic **while** programs. This language is a natural abstraction of familiar imperative programming languages such as Pascal or C. Different versions can be defined depending on the choice of tests allowed and whether or not nondeterminism is permitted.

The language of **while** programs is defined inductively. There are atomic programs and atomic tests, as well as program constructs for forming compound programs from simpler ones.

In the propositional version of Dynamic Logic (PDL), atomic programs are simply letters  $a, b, \dots$  from some alphabet. Thus PDL abstracts away from the nature of the domain of computation and studies the pure interaction between programs and propositions. For the first-order versions of DL, atomic programs are *simple assignments*  $x := t$ , where  $x$  is a variable and  $t$  is a term. In addition, a *nondeterministic* or *wildcard assignment*  $x := ?$  or *nondeterministic choice* construct may be allowed.

Tests can be *atomic tests*, which for propositional versions are simply propositional letters  $p$ , and for first-order versions are atomic formulas  $p(t_1, \dots, t_n)$ , where  $t_1, \dots, t_n$  are terms and  $p$  is an  $n$ -ary relation symbol in the vocabulary of the domain of computation. In addition, we include the *constant tests* **1** and **0**. Boolean combinations of atomic tests are often allowed, although this adds no expressive power. These versions of DL are called *poor test*.

More complicated tests can also be included. These versions of DL are sometimes called *rich test*. In rich test versions, the families of programs and tests are defined by mutual induction.

Compound programs are formed from the atomic programs and tests by induction, using the *composition*, *conditional*, and *while* operators. Formally, if  $\varphi$  is a test and  $\alpha$  and  $\beta$  are programs, then the following are programs:

- $\alpha ; \beta$
- **if**  $\varphi$  **then**  $\alpha$  **else**  $\beta$
- **while**  $\varphi$  **do**  $\alpha$ .

We can also parenthesize with **begin** ... **end** where necessary. The gcd program of Example 1 above is an example of a **while** program.

The semantics of these constructs is defined to correspond to the ordinary operational semantics familiar from common programming languages.



### *Regular Programs*

*Regular programs* are more general than **while** programs, but not by much. The advantage of regular programs is that they reduce the relatively more complicated **while** program operators to much simpler constructs. The deductive system becomes comparatively simpler too. They also incorporate a simple form of nondeterminism.

For a given set of atomic programs and tests, the set of *regular programs* is defined as follows:

- (i) any atomic program is a program
- (ii) if  $\varphi$  is a test, then  $\varphi?$  is a program
- (iii) if  $\alpha$  and  $\beta$  are programs, then  $\alpha ; \beta$  is a program;
- (iv) if  $\alpha$  and  $\beta$  are programs, then  $\alpha \cup \beta$  is a program;
- (v) if  $\alpha$  is a program, then  $\alpha^*$  is a program.

These constructs have the following intuitive meaning:

- (i) Atomic programs are basic and indivisible; they execute in a single step. They are called *atomic* because they cannot be decomposed further.
- (ii) The program  $\varphi?$  tests whether the property  $\varphi$  holds in the current state. If so, it continues without changing state. If not, it blocks without halting.
- (iii) The operator  $;$  is the *sequential composition* operator. The program  $\alpha ; \beta$  means, “Do  $\alpha$ , then do  $\beta$ .”
- (iv) The operator  $\cup$  is the *nondeterministic choice* operator. The program  $\alpha \cup \beta$  means, “Nondeterministically choose one of  $\alpha$  or  $\beta$  and execute it.”
- (v) The operator  $*$  is the *iteration* operator. The program  $\alpha$  means, “Execute  $\alpha$  some nondeterministically chosen finite number of times.”

Keep in mind that these descriptions are meant only as intuitive aids. A formal semantics will be given in Section 2.2, in which programs will be interpreted as binary input/output relations and the programming constructs above as operators on binary relations.

The operators  $\cup, ;, *$  may be familiar from automata and formal language theory (see [Kozen, 1997a]), where they are interpreted as operators on sets of strings over a finite alphabet. The language-theoretic and relation-theoretic semantics share much in common; in fact, they have the same equational theory, as shown in [Kozen, 1994a].

The operators of deterministic **while** programs can be defined in terms of the regular operators:

- (2) **if**  $\varphi$  **then**  $\alpha$  **else**  $\beta$   $\stackrel{\text{def}}{=} \varphi? ; \alpha \cup \neg\varphi? ; \beta$   
 (3) **while**  $\varphi$  **do**  $\alpha$   $\stackrel{\text{def}}{=} (\varphi? ; \alpha)^* ; \neg\varphi?$

The class of **while** programs is equivalent to the subclass of the regular programs in which the program operators  $\cup$ ,  $?$ , and  $*$  are constrained to appear only in these forms.

### *Recursion*

*Recursion* can appear in programming languages in several forms. Two such manifestations are *recursive calls* and *stacks*. Under certain very general conditions, the two constructs can simulate each other. It can also be shown that recursive programs and **while** programs are equally expressive over the natural numbers, whereas over arbitrary domains, **while** programs are strictly weaker. **While** programs correspond to what is often called *tail recursion* or *iteration*.

### *R.E. Programs*

A *finite computation sequence* of a program  $\alpha$ , or *seq* for short, is a finite-length string of atomic programs and tests representing a possible sequence of atomic steps that can occur in a halting execution of  $\alpha$ . Seqs are denoted  $\sigma, \tau, \dots$ . The set of all seqs of a program  $\alpha$  is denoted  $CS(\alpha)$ . We use the word “possible” loosely— $CS(\alpha)$  is determined by the syntax of  $\alpha$  alone. Because of tests that evaluate to false,  $CS(\alpha)$  may contain seqs that are never executed under any interpretation.

The set  $CS(\alpha)$  is a subset of  $A^*$ , where  $A$  is the set of atomic programs and tests occurring in  $\alpha$ . For **while** programs, regular programs, or recursive programs, we can define the set  $CS(\alpha)$  formally by induction on syntax. For example, for regular programs,

$$\begin{aligned} CS(a) &\stackrel{\text{def}}{=} \{a\}, \quad a \text{ an atomic program or test} \\ CS(\mathbf{skip}) &\stackrel{\text{def}}{=} \{\varepsilon\} \\ CS(\mathbf{fail}) &\stackrel{\text{def}}{=} \emptyset \\ CS(\alpha ; \beta) &\stackrel{\text{def}}{=} \{\sigma ; \tau \mid \sigma \in CS(\alpha), \tau \in CS(\beta)\} \\ CS(\alpha \cup \beta) &\stackrel{\text{def}}{=} CS(\alpha) \cup CS(\beta) \\ CS(\alpha^*) &\stackrel{\text{def}}{=} CS(\alpha)^* \\ &= \bigcup_{n \geq 0} CS(\alpha^n), \end{aligned}$$

where

$$\begin{aligned}\alpha^0 &\stackrel{\text{def}}{=} \mathbf{skip} \\ \alpha^{n+1} &\stackrel{\text{def}}{=} \alpha^n ; \alpha.\end{aligned}$$

For example, if  $a$  is an atomic program and  $p$  an atomic formula, then the program

$$\mathbf{while } p \mathbf{ do } a = (p? ; a)^* ; \neg p?$$

has as seqs all strings of the form

$$(p? ; a)^n ; \neg p? = \underbrace{p? ; a ; p? ; a ; \cdots ; p? ; a}_n ; \neg p?$$

for all  $n \geq 0$ . Note that each seq  $\sigma$  of a program  $\alpha$  is itself a program, and

$$CS(\sigma) = \{\sigma\}.$$

**While** programs and regular programs give rise to regular sets of seqs, and recursive programs give rise to context-free sets of seqs. Taking this a step further, we can define an *r.e. program* to be simply a recursively enumerable set of seqs. This is the most general programming language we will consider in the context of DL; it subsumes all the others in expressive power.

### *Nondeterminism*

We should say a few words about the concept of *nondeterminism* and its role in the study of logics and languages, since this concept often presents difficulty the first time it is encountered.

In some programming languages we will consider, the traces of a program need not be uniquely determined by their start states. When this is possible, we say that the program is *nondeterministic*. A nondeterministic program can have both divergent and convergent traces starting from the same input state, and for such programs it does not make sense to say that the program halts on a certain input state or that it loops on a certain input state; there may be different computations starting from the same input state that do each.

There are several concrete ways nondeterminism can enter into programs. One construct is the *nondeterministic* or *wildcard assignment*  $x := ?$ . Intuitively, this operation assigns an arbitrary element of the domain to the variable  $x$ , but it is not determined which one.<sup>1</sup> Another source of nondeterminism is the unconstrained use of the choice operator  $\cup$  in regular

<sup>1</sup>This construct is often called *random assignment* in the literature. This terminology is misleading, because it has nothing at all to do with probability.

programs. A third source is the iteration operator  $*$  in regular programs. A fourth source is r.e. programs, which are just r.e. sets of seqs; initially, the seq to execute is chosen nondeterministically. For example, over  $\mathbb{N}$ , the r.e. program

$$\{x := n \mid n \geq 0\}$$

is equivalent to the regular program

$$x := 0; (x := x + 1)^*.$$

Nondeterministic programs provide no explicit mechanism for resolving the nondeterminism. That is, there is no way to determine which of many possible next steps will be taken from a given state. This is hardly realistic. So why study nondeterminism at all if it does not correspond to anything operational? One good answer is that nondeterminism is a valuable tool that helps us understand the expressiveness of programming language constructs. It is useful in situations in which we cannot necessarily predict the outcome of a particular choice, but we may know the range of possibilities. In reality, computations may depend on information that is out of the programmer's control, such as input from the user or actions of other processes in the system. Nondeterminism is useful in modeling such situations.

The importance of nondeterminism is not limited to logics of programs. Indeed, the most important open problem in the field of computational complexity theory, the  $P=NP$  problem, is formulated in terms of nondeterminism.

#### 1.4 Program Verification

Dynamic Logic and other program logics are meant to be useful tools for facilitating the process of producing correct programs. One need only look at the miasma of buggy software to understand the dire need for such tools. But before we can produce correct software, we need to know what it means for it to be correct. It is not good enough to have some vague idea of what is supposed to happen when a program is run or to observe it running on some collection of inputs. In order to apply formal verification tools, we must have a formal specification of correctness for the verification tools to work with.

In general, a *correctness specification* is a formal description of how the program is supposed to behave. A given program is *correct* with respect to a correctness specification if its behavior fulfills that specification. For the gcd program of Example 1, the correctness might be specified informally by the assertion

If the input values of  $x$  and  $y$  are positive integers  $c$  and  $d$ , respectively, then

- (i) the output value of  $x$  is the gcd of  $c$  and  $d$ , and
- (ii) the program halts.

Of course, in order to work with a formal verification system, these properties must be expressed formally in a language such as first-order logic.

The assertion (ii) is part of the correctness specification because programs do not necessarily halt, but may produce infinite traces for certain inputs. A finite trace, as for example the one produced by the gcd program above on input state  $(15,27,0)$ , is called *halting*, *terminating*, or *convergent*. Infinite traces are called *looping* or *divergent*. For example, the program

**while**  $x > 7$  **do**  $x := x + 3$

loops on input state  $(8, \dots)$ , producing the infinite trace

$(8, \dots), (11, \dots), (14, \dots), \dots$

Dynamic Logic can reason about the behavior of a program that is manifested in its input/output relation. It is not well suited to reasoning about program behavior manifested in intermediate states of a computation (although there are close relatives, such as Process Logic and Temporal Logic, that are). This is not to say that all interesting program behavior is captured by the input/output relation, and that other types of behavior are irrelevant or uninteresting. Indeed, the restriction to input/output relations is reasonable only when programs are supposed to halt after a finite time and yield output results. This approach will not be adequate for dealing with programs that normally are not supposed to halt, such as operating systems.

For programs that are supposed to halt, correctness criteria are traditionally given in the form of an *input/output specification* consisting of a formal relation between the input and output states that the program is supposed to maintain, along with a description of the set of input states on which the program is supposed to halt. The input/output relation of a program carries all the information necessary to determine whether the program is correct relative to such a specification. Dynamic Logic is well suited to this type of verification.

It is not always obvious what the correctness specification ought to be. Sometimes, producing a formal specification of correctness is as difficult as producing the program itself, since both must be written in a formal language. Moreover, specifications are as prone to bugs as programs. Why bother then? Why not just implement the program with some vague specification in mind?

There are several good reasons for taking the effort to produce formal specifications:

1. Often when implementing a large program from scratch, the programmer may have been given only a vague idea of what the finished product is supposed to do. This is especially true when producing software for a less technically inclined employer. There may be a rough informal description available, but the minor details are often left to the programmer. It is very often the case that a large part of the programming process consists of taking a vaguely specified problem and making it precise. The process of formulating the problem precisely can be considered a *definition* of what the program is supposed to do. And it is just good programming practice to have a very clear idea of what we want to do before we start doing it.
2. In the process of formulating the specification, several unforeseen cases may become apparent, for which it is not clear what the appropriate action of the program should be. This is especially true with error handling and other exceptional situations. Formulating a specification can define the action of the program in such situations and thereby tie up loose ends.
3. The process of formulating a rigorous specification can sometimes suggest ideas for implementation, because it forces us to isolate the issues that drive design decisions. When we know all the ways our data are going to be accessed, we are in a better position to choose the right data structures that optimize the tradeoffs between efficiency and generality.
4. The specification is often expressed in a language quite different from the programming language. The specification is *functional*—it tells *what* the program is supposed to do—as opposed to *imperative*—*how* to do it. It is often easier to specify the desired functionality independent of the details of how it will be implemented. For example, we can quite easily express what it means for a number  $x$  to be the gcd of  $y$  and  $z$  in first-order logic without even knowing how to compute it.
5. Verifying that a program meets its specification is a kind of sanity check. It allows us to give two solutions to the problem—once as a functional specification, and once as an algorithmic implementation—and lets us verify that the two are compatible. Any incompatibilities between the program and the specification are either bugs in the program, bugs in the specification, or both. The cycle of refining the specification, modifying the program to meet the specification, and re-verifying until the process converges can lead to software in which we have much more confidence.

### *Partial and Total Correctness*

Typically, a program is designed to implement some functionality. As mentioned above, that functionality can often be expressed formally in the form of an input/output specification. Concretely, such a specification consists of an *input condition* or *precondition*  $\varphi$  and an *output condition* or *post-condition*  $\psi$ . These are properties of the input state and the output state, respectively, expressed in some formal language such as the first-order language of the domain of computation. The program is supposed to halt in a state satisfying the output condition whenever the input state satisfies the input condition. We say that a program is *partially correct* with respect to a given input/output specification  $\varphi, \psi$  if, whenever the program is started in a state satisfying the input condition  $\varphi$ , then if and when it ever halts, it does so in a state satisfying the output condition  $\psi$ . The definition of partial correctness does not stipulate that the program halts; this is what we mean by *partial*.

A program is *totally correct* with respect to an input/output specification  $\varphi, \psi$  if

- it is partially correct with respect to that specification; and
- it halts whenever it is started in a state satisfying the input condition  $\varphi$ .

The input/output specification imposes no requirements when the input state does not satisfy the input condition  $\varphi$ —the program might as well loop infinitely or erase memory. This is the “garbage in, garbage out” philosophy. If we really do care what the program does on some of those input states, then we had better rewrite the input condition to include them and say formally what we want to happen in those cases.

For example, in the gcd program of Example 1, the output condition  $\psi$  might be the condition (i) stating that the output value of  $x$  is the gcd of the input values of  $x$  and  $y$ . We can express this completely formally in the language of first-order number theory. We may try to start off with the input specification  $\varphi_0 = \mathbf{1}$  (*true*); that is, no restrictions on the input state at all. Unfortunately, if the initial value of  $y$  is 0 and  $x$  is negative, the final value of  $x$  will be the same as the initial value, thus negative. If we expect all gcds to be positive, this would be wrong. Another problematic situation arises when the initial values of  $x$  and  $y$  are both 0; in this case the gcd is not defined. Therefore, the program as written is not partially correct with respect to the specification  $\varphi_0, \psi$ .

We can remedy the situation by providing an input specification that rules out these troublesome input values. We can limit the input states to those in which  $x$  and  $y$  are both nonnegative and not both zero by taking

the input specification

$$\varphi_1 = (x \geq 0 \wedge y > 0) \vee (x > 0 \wedge y \geq 0).$$

The gcd program of Example 1 above would be partially correct with respect to the specification  $\varphi_1, \psi$ . It is also totally correct, since the program halts on all inputs satisfying  $\varphi_1$ .

Perhaps we want to allow any input in which not both  $x$  and  $y$  are zero. In that case, we should use the input specification  $\varphi_2 = \neg(x = 0 \wedge y = 0)$ . But then the program of Example 1 is not partially correct with respect to  $\varphi_2, \psi$ ; we must amend the program to produce the correct (positive) gcd on negative inputs.

### 1.5 Exogenous and Endogenous Logics

There are two main approaches to modal logics of programs: the *exogenous* approach, exemplified by Dynamic Logic and its precursor Hoare Logic [Hoare, 1969], and the *endogenous* approach, exemplified by Temporal Logic and its precursor, the invariant assertions method of [Floyd, 1967]. A logic is *exogenous* if its programs are explicit in the language. Syntactically, a Dynamic Logic program is a well-formed expression built inductively from primitive programs using a small set of program operators. Semantically, a program is interpreted as its input/output relation. The relation denoted by a compound program is determined by the relations denoted by its parts. This aspect of *compositionality* allows analysis by structural induction.

The importance of compositionality is discussed in [van Emde Boas, 1978]. In Temporal Logic, the program is fixed and is considered part of the structure over which the logic is interpreted. The current location in the program during execution is stored in a special variable for that purpose, called the *program counter*, and is part of the state along with the values of the program variables. Instead of program operators, there are temporal operators that describe how the program variables, including the program counter, change with time. Thus Temporal Logic sacrifices compositionality for a less restricted formalism. We discuss Temporal Logic further in Section 14.2.

## 2 PROPOSITIONAL DYNAMIC LOGIC (PDL)

Propositional Dynamic Logic (PDL) plays the same role in Dynamic Logic that classical propositional logic plays in classical predicate logic. It describes the properties of the interaction between programs and propositions that are independent of the domain of computation. Since PDL is a subsystem of first-order DL, we can be sure that all properties of PDL that we discuss in this section will also be valid in first-order DL.



Since there is no domain of computation in PDL, there can be no notion of assignment to a variable. Instead, primitive programs are interpreted as arbitrary binary relations on an abstract set of states  $K$ . Likewise, primitive assertions are just atomic propositions and are interpreted as arbitrary subsets of  $K$ . Other than this, no special structure is imposed.

This level of abstraction may at first appear too general to say anything of interest. On the contrary, it is a very natural level of abstraction at which many fundamental relationships between programs and propositions can be observed.

For example, consider the PDL formula

$$(4) \quad [\alpha](\varphi \wedge \psi) \leftrightarrow [\alpha]\varphi \wedge [\alpha]\psi.$$

The left-hand side asserts that the formula  $\varphi \wedge \psi$  must hold after the execution of program  $\alpha$ , and the right-hand side asserts that  $\varphi$  must hold after execution of  $\alpha$  and so must  $\psi$ . The formula (4) asserts that these two statements are equivalent. This implies that to verify a conjunction of two postconditions, it suffices to verify each of them separately. The assertion (4) holds universally, regardless of the domain of computation and the nature of the particular  $\alpha$ ,  $\varphi$ , and  $\psi$ .

As another example, consider

$$(5) \quad [\alpha; \beta]\varphi \leftrightarrow [\alpha][\beta]\varphi.$$

The left-hand side asserts that after execution of the composite program  $\alpha; \beta$ ,  $\varphi$  must hold. The right-hand side asserts that after execution of the program  $\alpha$ ,  $[\beta]\varphi$  must hold, which in turn says that after execution of  $\beta$ ,  $\varphi$  must hold. The formula (5) asserts the logical equivalence of these two statements. It holds regardless of the nature of  $\alpha$ ,  $\beta$ , and  $\varphi$ . Like (4), (5) can be used to simplify the verification of complicated programs.

As a final example, consider the assertion

$$(6) \quad [\alpha]p \leftrightarrow [\beta]p$$

where  $p$  is a primitive proposition symbol and  $\alpha$  and  $\beta$  are programs. If this formula is true under all interpretations, then  $\alpha$  and  $\beta$  are *equivalent* in the sense that they behave identically with respect to any property expressible in PDL or any formal system containing PDL as a subsystem. This is because the assertion will hold for any substitution instance of (6). For example, the two programs

$$\begin{aligned} \alpha &= \mathbf{if} \ \varphi \ \mathbf{then} \ \gamma \ \mathbf{else} \ \delta \\ \beta &= \mathbf{if} \ \neg\varphi \ \mathbf{then} \ \delta \ \mathbf{else} \ \gamma \end{aligned}$$

are equivalent in the sense of (6).

## 2.1 Syntax

Syntactically, PDL is a blend of three classical ingredients: propositional logic, modal logic, and the algebra of regular expressions. There are several versions of PDL, depending on the choice of program operators allowed. In this section we will introduce the basic version, called *regular PDL*. Variations of this basic version will be considered in later sections.

The language of regular PDL has expressions of two sorts: *propositions* or *formulas*  $\varphi, \psi, \dots$  and *programs*  $\alpha, \beta, \gamma, \dots$ . There are countably many *atomic symbols* of each sort. Atomic programs are denoted  $a, b, c, \dots$  and the set of all atomic programs is denoted  $\Pi_0$ . Atomic propositions are denoted  $p, q, r, \dots$  and the set of all atomic propositions is denoted  $\Phi_0$ . The set of all programs is denoted  $\Pi$  and the set of all propositions is denoted  $\Phi$ . Programs and propositions are built inductively from the atomic ones using the following operators:

Propositional operators:

$\rightarrow$	implication
$\mathbf{0}$	falsity

Program operators:

$;$	composition
$\cup$	choice
$*$	iteration

Mixed operators:

$[ ]$	necessity
$?$	test

The definition of programs and propositions is by mutual induction. All atomic programs are programs and all atomic propositions are propositions. If  $\varphi, \psi$  are propositions and  $\alpha, \beta$  are programs, then

$\varphi \rightarrow \psi$	propositional implication
$\mathbf{0}$	propositional falsity
$[\alpha]\varphi$	program necessity

are propositions and

$\alpha ; \beta$	sequential composition
$\alpha \cup \beta$	nondeterministic choice
$\alpha^*$	iteration
$\varphi?$	test

are programs. In more formal terms, we define the set  $\Pi$  of all programs and the set  $\Phi$  of all propositions to be the smallest sets such that

- $\Phi_0 \subseteq \Phi$
- $\Pi_0 \subseteq \Pi$
- if  $\varphi, \psi \in \Phi$ , then  $\varphi \rightarrow \psi \in \Phi$  and  $\mathbf{0} \in \Phi$
- if  $\alpha, \beta \in \Pi$ , then  $\alpha; \beta$ ,  $\alpha \cup \beta$ , and  $\alpha^* \in \Pi$
- if  $\alpha \in \Pi$  and  $\varphi \in \Phi$ , then  $[\alpha]\varphi \in \Phi$
- if  $\varphi \in \Phi$  then  $\varphi? \in \Pi$ .

Note that the inductive definitions of programs  $\Pi$  and propositions  $\Phi$  are intertwined and cannot be separated. The definition of propositions depends on the definition of programs because of the construct  $[\alpha]\varphi$ , and the definition of programs depends on the definition of propositions because of the construct  $\varphi?$ . Note also that we have allowed all formulas as tests. This is the *rich test* version of PDL.

Compound programs and propositions have the following intuitive meanings:

$[\alpha]\varphi$  “It is necessary that after executing  $\alpha$ ,  $\varphi$  is true.”

$\alpha; \beta$  “Execute  $\alpha$ , then execute  $\beta$ .”

$\alpha \cup \beta$  “Choose either  $\alpha$  or  $\beta$  nondeterministically and execute it.”

$\alpha^*$  “Execute  $\alpha$  a nondeterministically chosen finite number of times (zero or more).”

$\varphi?$  “Test  $\varphi$ ; proceed if true, fail if false.”

We avoid parentheses by assigning precedence to the operators: unary operators, including  $[\alpha]$ , bind tighter than binary ones, and  $;$  binds tighter than  $\cup$ . Thus the expression

$$[\alpha; \beta^* \cup \gamma^*]\varphi \vee \psi$$

should be read

$$([\alpha; (\beta^*)] \cup (\gamma^*))\varphi \vee \psi.$$

Of course, parentheses can always be used to enforce a particular parse of an expression or to enhance readability. Also, under the semantics to be given in the next section, the operators  $;$  and  $\cup$  will turn out to be associative, so we may write  $\alpha; \beta; \gamma$  and  $\alpha \cup \beta \cup \gamma$  without ambiguity. We often omit the symbol  $;$  and write the composition  $\alpha; \beta$  as  $\alpha\beta$ .

The propositional operators  $\wedge$ ,  $\vee$ ,  $\neg$ ,  $\leftrightarrow$ , and  $\mathbf{1}$  can be defined from  $\rightarrow$  and  $\mathbf{0}$  in the usual way.

The possibility operator  $\langle \alpha \rangle$  is the modal dual of the necessity operator  $[\ ]$ . It is defined by

$$\langle \alpha \rangle \varphi \stackrel{\text{def}}{=} \neg [\alpha] \neg \varphi.$$

The propositions  $[\alpha] \varphi$  and  $\langle \alpha \rangle \varphi$  are read “box  $\alpha \varphi$ ” and “diamond  $\alpha \varphi$ ,” respectively. The latter has the intuitive meaning, “There is a computation of  $\alpha$  that terminates in a state satisfying  $\varphi$ .”

One important difference between  $\langle \alpha \rangle$  and  $[\ ]$  is that  $\langle \alpha \rangle \varphi$  implies that  $\alpha$  terminates, whereas  $[\alpha] \varphi$  does not. Indeed, the formula  $[\alpha] \mathbf{0}$  asserts that no computation of  $\alpha$  terminates, and the formula  $[\alpha] \mathbf{1}$  is always true, regardless of  $\alpha$ .

In addition, we define

$$\begin{aligned} \mathbf{skip} &\stackrel{\text{def}}{=} \mathbf{1?} \\ \mathbf{fail} &\stackrel{\text{def}}{=} \mathbf{0?} \\ \mathbf{if } \varphi_1 \rightarrow \alpha_1 \mid \cdots \mid \varphi_n \rightarrow \alpha_n \mathbf{ fi} &\stackrel{\text{def}}{=} \varphi_1?; \alpha_1 \cup \cdots \cup \varphi_n?; \alpha_n \\ \mathbf{do } \varphi_1 \rightarrow \alpha_1 \mid \cdots \mid \varphi_n \rightarrow \alpha_n \mathbf{ od} &\stackrel{\text{def}}{=} \left( \bigcup_{i=1}^n \varphi_i?; \alpha_i \right)^*; \left( \bigwedge_{i=1}^n \neg \varphi_i \right)? \\ \mathbf{if } \varphi \mathbf{ then } \alpha \mathbf{ else } \beta &\stackrel{\text{def}}{=} \mathbf{if } \varphi \rightarrow \alpha \mid \neg \varphi \rightarrow \beta \mathbf{ fi} \\ &= \varphi?; \alpha \cup \neg \varphi?; \beta \\ \mathbf{while } \varphi \mathbf{ do } \alpha &\stackrel{\text{def}}{=} \mathbf{do } \varphi \rightarrow \alpha \mathbf{ od} \\ &= (\varphi?; \alpha)^*; \neg \varphi? \\ \mathbf{repeat } \alpha \mathbf{ until } \varphi &\stackrel{\text{def}}{=} \alpha; \mathbf{while } \neg \varphi \mathbf{ do } \alpha \\ &= \alpha; (\neg \varphi?; \alpha)^*; \varphi? \\ \{ \varphi \} \alpha \{ \psi \} &\stackrel{\text{def}}{=} \varphi \rightarrow [\alpha] \psi. \end{aligned}$$

The programs **skip** and **fail** are the program that does nothing (no-op) and the failing program, respectively. The ternary **if-then-else** operator and the binary **while-do** operator are the usual *conditional* and *while loop* constructs found in conventional programming languages. The constructs **if-|fi** and **do-|od** are the *alternative guarded command* and *iterative guarded command* constructs, respectively. The construct  $\{ \varphi \} \alpha \{ \psi \}$  is the Hoare partial correctness assertion. We will argue later that the formal definitions of these operators given above correctly model their intuitive behavior.

## 2.2 Semantics

The semantics of PDL comes from the semantics for modal logic. The structures over which programs and propositions of PDL are interpreted

are called *Kripke frames* in honor of Saul Kripke, the inventor of the formal semantics of modal logic. A *Kripke frame* is a pair

$$\mathfrak{K} = (K, \mathbf{m}_{\mathfrak{K}}),$$

where  $K$  is a set of elements  $u, v, w, \dots$  called *states* and  $\mathbf{m}_{\mathfrak{K}}$  is a *meaning function* assigning a subset of  $K$  to each atomic proposition and a binary relation on  $K$  to each atomic program. That is,

$$\begin{aligned} \mathbf{m}_{\mathfrak{K}}(p) &\subseteq K, & p \in \Phi_0 \\ \mathbf{m}_{\mathfrak{K}}(a) &\subseteq K \times K, & a \in \Pi_0. \end{aligned}$$

We will extend the definition of the function  $\mathbf{m}_{\mathfrak{K}}$  by induction below to give a meaning to all elements of  $\Pi$  and  $\Phi$  such that

$$\begin{aligned} \mathbf{m}_{\mathfrak{K}}(\varphi) &\subseteq K, & \varphi \in \Phi \\ \mathbf{m}_{\mathfrak{K}}(\alpha) &\subseteq K \times K, & \alpha \in \Pi. \end{aligned}$$

Intuitively, we can think of the set  $\mathbf{m}_{\mathfrak{K}}(\varphi)$  as the set of states *satisfying* the proposition  $\varphi$  in the model  $\mathfrak{K}$ , and we can think of the binary relation  $\mathbf{m}_{\mathfrak{K}}(\alpha)$  as the set of input/output pairs of states of the program  $\alpha$ .

Formally, the meanings  $\mathbf{m}_{\mathfrak{K}}(\varphi)$  of  $\varphi \in \Phi$  and  $\mathbf{m}_{\mathfrak{K}}(\alpha)$  of  $\alpha \in \Pi$  are defined by mutual induction on the structure of  $\varphi$  and  $\alpha$ . The basis of the induction, which specifies the meanings of the atomic symbols  $p \in \Phi_0$  and  $a \in \Pi_0$ , is already given in the specification of  $\mathfrak{K}$ . The meanings of compound propositions and programs are defined as follows.

$$\begin{aligned} \mathbf{m}_{\mathfrak{K}}(\varphi \rightarrow \psi) &\stackrel{\text{def}}{=} (K - \mathbf{m}_{\mathfrak{K}}(\varphi)) \cup \mathbf{m}_{\mathfrak{K}}(\psi) \\ \mathbf{m}_{\mathfrak{K}}(\mathbf{0}) &\stackrel{\text{def}}{=} \emptyset \\ \mathbf{m}_{\mathfrak{K}}([\alpha]\varphi) &\stackrel{\text{def}}{=} K - (\mathbf{m}_{\mathfrak{K}}(\alpha) \circ (K - \mathbf{m}_{\mathfrak{K}}(\varphi))) \\ &= \{u \mid \forall v \in K \text{ if } (u, v) \in \mathbf{m}_{\mathfrak{K}}(\alpha) \text{ then } v \in \mathbf{m}_{\mathfrak{K}}(\varphi)\} \\ (7) \quad \mathbf{m}_{\mathfrak{K}}(\alpha; \beta) &\stackrel{\text{def}}{=} \mathbf{m}_{\mathfrak{K}}(\alpha) \circ \mathbf{m}_{\mathfrak{K}}(\beta) \\ &= \{(u, v) \mid \exists w \in K (u, w) \in \mathbf{m}_{\mathfrak{K}}(\alpha) \text{ and } (w, v) \in \mathbf{m}_{\mathfrak{K}}(\beta)\} \\ \mathbf{m}_{\mathfrak{K}}(\alpha \cup \beta) &\stackrel{\text{def}}{=} \mathbf{m}_{\mathfrak{K}}(\alpha) \cup \mathbf{m}_{\mathfrak{K}}(\beta) \\ (8) \quad \mathbf{m}_{\mathfrak{K}}(\alpha^*) &\stackrel{\text{def}}{=} \mathbf{m}_{\mathfrak{K}}(\alpha)^* = \bigcup_{n \geq 0} \mathbf{m}_{\mathfrak{K}}(\alpha)^n \\ \mathbf{m}_{\mathfrak{K}}(\varphi?) &\stackrel{\text{def}}{=} \{(u, u) \mid u \in \mathbf{m}_{\mathfrak{K}}(\varphi)\}. \end{aligned}$$

The operator  $\circ$  in (7) is relational composition. In (8), the first occurrence of  $*$  is the iteration symbol of PDL, and the second is the reflexive transitive closure operator on binary relations. Thus (8) says that the program  $\alpha^*$  is interpreted as the reflexive transitive closure of  $\mathbf{m}_{\mathfrak{K}}(\alpha)$ .

We write  $\mathfrak{K}, u \models \varphi$  and  $u \in \mathbf{m}_{\mathfrak{K}}(\varphi)$  interchangeably, and say that  $u$  *satisfies*  $\varphi$  in  $\mathfrak{K}$ , or that  $\varphi$  is *true* at state  $u$  in  $\mathfrak{K}$ . We may omit the  $\mathfrak{K}$  and write  $u \models \varphi$

when  $\mathfrak{R}$  is understood. The notation  $u \not\models \varphi$  means that  $u$  does not satisfy  $\varphi$ , or in other words that  $u \notin \mathfrak{m}_{\mathfrak{R}}(\varphi)$ . In this notation, we can restate the definition above equivalently as follows:

$$\begin{aligned}
u \models \varphi \rightarrow \psi &\stackrel{\text{def}}{\iff} u \models \varphi \text{ implies } u \models \psi \\
&u \not\models \mathbf{0} \\
u \models [\alpha]\varphi &\stackrel{\text{def}}{\iff} \forall v \text{ if } (u, v) \in \mathfrak{m}_{\mathfrak{R}}(\alpha) \text{ then } v \models \varphi \\
(u, v) \in \mathfrak{m}_{\mathfrak{R}}(\alpha\beta) &\stackrel{\text{def}}{\iff} \exists w (u, w) \in \mathfrak{m}_{\mathfrak{R}}(\alpha) \text{ and } (w, v) \in \mathfrak{m}_{\mathfrak{R}}(\beta) \\
(u, v) \in \mathfrak{m}_{\mathfrak{R}}(\alpha \cup \beta) &\stackrel{\text{def}}{\iff} (u, v) \in \mathfrak{m}_{\mathfrak{R}}(\alpha) \text{ or } (u, v) \in \mathfrak{m}_{\mathfrak{R}}(\beta) \\
(u, v) \in \mathfrak{m}_{\mathfrak{R}}(\alpha^*) &\stackrel{\text{def}}{\iff} \exists n \geq 0 \exists u_0, \dots, u_n \ u = u_0, \ v = u_n, \\
&\text{and } (u_i, u_{i+1}) \in \mathfrak{m}_{\mathfrak{R}}(\alpha), \ 0 \leq i \leq n-1 \\
(u, v) \in \mathfrak{m}_{\mathfrak{R}}(\varphi?) &\stackrel{\text{def}}{\iff} u = v \text{ and } u \models \varphi.
\end{aligned}$$

The defined operators inherit their meanings from these definitions:

$$\begin{aligned}
\mathfrak{m}_{\mathfrak{R}}(\varphi \vee \psi) &\stackrel{\text{def}}{=} \mathfrak{m}_{\mathfrak{R}}(\varphi) \cup \mathfrak{m}_{\mathfrak{R}}(\psi) \\
\mathfrak{m}_{\mathfrak{R}}(\varphi \wedge \psi) &\stackrel{\text{def}}{=} \mathfrak{m}_{\mathfrak{R}}(\varphi) \cap \mathfrak{m}_{\mathfrak{R}}(\psi) \\
\mathfrak{m}_{\mathfrak{R}}(\neg\varphi) &\stackrel{\text{def}}{=} K - \mathfrak{m}_{\mathfrak{R}}(\varphi) \\
\mathfrak{m}_{\mathfrak{R}}(\langle\alpha\rangle\varphi) &\stackrel{\text{def}}{=} \{u \mid \exists v \in K \ (u, v) \in \mathfrak{m}_{\mathfrak{R}}(\alpha) \text{ and } v \in \mathfrak{m}_{\mathfrak{R}}(\varphi)\} \\
&= \mathfrak{m}_{\mathfrak{R}}(\alpha) \circ \mathfrak{m}_{\mathfrak{R}}(\varphi) \\
\mathfrak{m}_{\mathfrak{R}}(\mathbf{1}) &\stackrel{\text{def}}{=} K \\
\mathfrak{m}_{\mathfrak{R}}(\mathbf{skip}) &\stackrel{\text{def}}{=} \mathfrak{m}_{\mathfrak{R}}(\mathbf{1}?) = \iota, \text{ the identity relation} \\
\mathfrak{m}_{\mathfrak{R}}(\mathbf{fail}) &\stackrel{\text{def}}{=} \mathfrak{m}_{\mathfrak{R}}(\mathbf{0}?) = \emptyset.
\end{aligned}$$

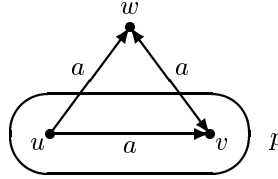
In addition, the **if-then-else**, **while-do**, and guarded commands inherit their semantics from the above definitions, and the input/output relations given by the formal semantics capture their intuitive operational meanings. For example, the relation associated with the program **while**  $\varphi$  **do**  $\alpha$  is the set of pairs  $(u, v)$  for which there exist states  $u_0, u_1, \dots, u_n$ ,  $n \geq 0$ , such that  $u = u_0$ ,  $v = u_n$ ,  $u_i \in \mathfrak{m}_{\mathfrak{R}}(\varphi)$  and  $(u_i, u_{i+1}) \in \mathfrak{m}_{\mathfrak{R}}(\alpha)$  for  $0 \leq i < n$ , and  $u_n \notin \mathfrak{m}_{\mathfrak{R}}(\varphi)$ .

This version of PDL is usually called *regular PDL* and the elements of  $\Pi$  are called *regular programs* because of the primitive operators  $\cup$ ,  $;$ , and  $*$ , which are familiar from regular expressions. Programs can be viewed as regular expressions over the atomic programs and tests. In fact, it can be shown that if  $p$  is an atomic proposition symbol, then any two test-free programs  $\alpha, \beta$  are equivalent as regular expressions—that is, they represent the same regular set—if and only if the formula  $\langle\alpha\rangle p \leftrightarrow \langle\beta\rangle p$  is valid.

EXAMPLE 2. Let  $p$  be an atomic proposition, let  $a$  be an atomic program, and let  $\mathfrak{K} = (K, \mathfrak{m}_{\mathfrak{K}})$  be a Kripke frame with

$$\begin{aligned} K &= \{u, v, w\} \\ \mathfrak{m}_{\mathfrak{K}}(p) &= \{u, v\} \\ \mathfrak{m}_{\mathfrak{K}}(a) &= \{(u, v), (u, w), (v, w), (w, v)\}. \end{aligned}$$

The following diagram illustrates  $\mathfrak{K}$ .



In this structure,  $u \models \langle a \rangle \neg p \wedge \langle a \rangle p$ , but  $v \models [a] \neg p$  and  $w \models [a] p$ . Moreover, every state of  $\mathfrak{K}$  satisfies the formula

$$\langle a^* \rangle [(aa)^*] p \wedge \langle a^* \rangle [(aa)^*] \neg p.$$

### 2.3 Computation Sequences

Let  $\alpha$  be a program. Recall from Section 1.3 that a *finite computation sequence* of  $\alpha$  is a finite-length string of atomic programs and tests representing a possible sequence of atomic steps that can occur in a halting execution of  $\alpha$ . These strings are called *seqs* and are denoted  $\sigma, \tau, \dots$ . The set of all such sequences is denoted  $CS(\alpha)$ . We use the word “possible” here loosely— $CS(\alpha)$  is determined by the syntax of  $\alpha$  alone, and may contain strings that are never executed in any interpretation. The formal definition of  $CS(\alpha)$  was given in Section 1.3.

Note that each finite computation sequence  $\beta$  of a program  $\alpha$  is itself a program, and  $CS(\beta) = \{\beta\}$ . Moreover, the following proposition is not difficult to prove by induction on the structure of  $\alpha$ :

PROPOSITION 3.

$$\mathfrak{m}_{\mathfrak{K}}(\alpha) = \bigcup_{\sigma \in CS(\alpha)} \mathfrak{m}_{\mathfrak{K}}(\sigma).$$

### 2.4 Satisfiability and Validity

The definitions of satisfiability and validity of propositions come from modal logic. Let  $\mathfrak{K} = (K, \mathfrak{m}_{\mathfrak{K}})$  be a Kripke frame and let  $\varphi$  be a proposition. We have defined in Section 2.2 what it means for  $\mathfrak{K}, u \models \varphi$ . If  $\mathfrak{K}, u \models \varphi$  for some

$u \in K$ , we say that  $\varphi$  is *satisfiable* in  $\mathfrak{K}$ . If  $\varphi$  is satisfiable in some  $\mathfrak{K}$ , we say that  $\varphi$  is *satisfiable*.

If  $\mathfrak{K}, u \models \varphi$  for all  $u \in K$ , we write  $\mathfrak{K} \models \varphi$  and say that  $\varphi$  is *valid* in  $\mathfrak{K}$ . If  $\mathfrak{K} \models \varphi$  for all Kripke frames  $\mathfrak{K}$ , we write  $\models \varphi$  and say that  $\varphi$  is *valid*.

If  $\Sigma$  is a set of propositions, we write  $\mathfrak{K} \models \Sigma$  if  $\mathfrak{K} \models \varphi$  for all  $\varphi \in \Sigma$ . A proposition  $\psi$  is said to be a *logical consequence* of  $\Sigma$  if  $\mathfrak{K} \models \psi$  whenever  $\mathfrak{K} \models \Sigma$ , in which case we write  $\Sigma \models \psi$ . (Note that this is *not* the same as saying that  $\mathfrak{K}, u \models \psi$  whenever  $\mathfrak{K}, u \models \Sigma$ .) We say that an inference rule

$$\frac{\varphi_1, \dots, \varphi_n}{\varphi}$$

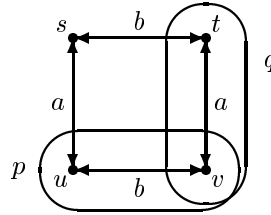
is *sound* if  $\varphi$  is a logical consequence of  $\{\varphi_1, \dots, \varphi_n\}$ .

Satisfiability and validity are dual in the same sense that  $\exists$  and  $\forall$  are dual and  $\langle \rangle$  and  $[ ]$  are dual: a proposition is valid (in  $\mathfrak{K}$ ) if and only if its negation is not satisfiable (in  $\mathfrak{K}$ ).

EXAMPLE 4. Let  $p, q$  be atomic propositions, let  $a, b$  be atomic programs, and let  $\mathfrak{K} = (K, m_{\mathfrak{K}})$  be a Kripke frame with

$$\begin{aligned} K &= \{s, t, u, v\} \\ m_{\mathfrak{K}}(p) &= \{u, v\} \\ m_{\mathfrak{K}}(q) &= \{t, v\} \\ m_{\mathfrak{K}}(a) &= \{(t, v), (v, t), (s, u), (u, s)\} \\ m_{\mathfrak{K}}(b) &= \{(u, v), (v, u), (s, t), (t, s)\}. \end{aligned}$$

The following figure illustrates  $\mathfrak{K}$ .



The following formulas are valid in  $\mathfrak{K}$ .

$$\begin{aligned} p &\leftrightarrow [(ab^*a)^*]p \\ q &\leftrightarrow [(ba^*b)^*]q. \end{aligned}$$

Also, let  $\alpha$  be the program

$$\alpha = (aa \cup bb \cup (ab \cup ba)(aa \cup bb)^*(ab \cup ba))^*.$$

Thinking of  $\alpha$  as a regular expression,  $\alpha$  generates all words over the alphabet  $\{a, b\}$  with an even number of occurrences of each of  $a$  and  $b$ . It can be shown that for any proposition  $\varphi$ , the proposition  $\varphi \leftrightarrow [\alpha]\varphi$  is valid in  $\mathfrak{K}$ .



EXAMPLE 5. The formula

$$p \wedge [a^*]((p \rightarrow [a]\neg p) \wedge (\neg p \rightarrow [a]p)) \leftrightarrow [(aa)^*]p \wedge [a(aa)^*]\neg p$$

is valid. Both sides assert in different ways that  $p$  is alternately true and false along paths of execution of the atomic program  $a$ .

## 2.5 Basic Properties

THEOREM 6. *The following are valid formulas of PDL:*

- (i)  $\langle \alpha \rangle (\varphi \vee \psi) \leftrightarrow \langle \alpha \rangle \varphi \vee \langle \alpha \rangle \psi$
- (ii)  $[\alpha] (\varphi \wedge \psi) \leftrightarrow [\alpha] \varphi \wedge [\alpha] \psi$
- (iii)  $\langle \alpha \rangle \varphi \wedge [\alpha] \psi \rightarrow \langle \alpha \rangle (\varphi \wedge \psi)$
- (iv)  $[\alpha] (\varphi \rightarrow \psi) \rightarrow ([\alpha] \varphi \rightarrow [\alpha] \psi)$
- (v)  $\langle \alpha \rangle (\varphi \wedge \psi) \rightarrow \langle \alpha \rangle \varphi \wedge \langle \alpha \rangle \psi$
- (vi)  $[\alpha] \varphi \vee [\alpha] \psi \rightarrow [\alpha] (\varphi \vee \psi)$
- (vii)  $\langle \alpha \rangle \mathbf{0} \leftrightarrow \mathbf{0}$
- (viii)  $[\alpha] \varphi \leftrightarrow \neg \langle \alpha \rangle \neg \varphi$ .
- (ix)  $\langle \alpha \cup \beta \rangle \varphi \leftrightarrow \langle \alpha \rangle \varphi \vee \langle \beta \rangle \varphi$
- (x)  $[\alpha \cup \beta] \varphi \leftrightarrow [\alpha] \varphi \wedge [\beta] \varphi$
- (xi)  $\langle \alpha ; \beta \rangle \varphi \leftrightarrow \langle \alpha \rangle \langle \beta \rangle \varphi$
- (xii)  $[\alpha ; \beta] \varphi \leftrightarrow [\alpha] [\beta] \varphi$
- (xiii)  $\langle \varphi ? \rangle \psi \leftrightarrow (\varphi \wedge \psi)$
- (xiv)  $[\varphi ?] \psi \leftrightarrow (\varphi \rightarrow \psi)$ .

THEOREM 7. *The following are sound rules of inference of PDL:*

- (i) *Modal generalization (GEN):*

$$\frac{\varphi}{[\alpha] \varphi}$$

- (ii) *Monotonicity of  $\langle \alpha \rangle$ :*

$$\frac{\varphi \rightarrow \psi}{\langle \alpha \rangle \varphi \rightarrow \langle \alpha \rangle \psi}$$

(iii) *Monotonicity of  $[\alpha]$* :

$$\frac{\varphi \rightarrow \psi}{[\alpha]\varphi \rightarrow [\alpha]\psi}$$

The *converse operator*  $^-$  is a program operator with semantics

$$\mathfrak{m}_{\mathfrak{R}}(\alpha^-) = \mathfrak{m}_{\mathfrak{R}}(\alpha)^- = \{(v, u) \mid (u, v) \in \mathfrak{m}_{\mathfrak{R}}(\alpha)\}.$$

Intuitively, the converse operator allows us to “run a program backwards;” semantically, the input/output relation of the program  $\alpha^-$  is the output/input relation of  $\alpha$ . Although this is not always possible to realize in practice, it is nevertheless a useful expressive tool. For example, it gives us a convenient way to talk about *backtracking*, or rolling back a computation to a previous state.

**THEOREM 8.** *For any programs  $\alpha$  and  $\beta$ ,*

- (i)  $\mathfrak{m}_{\mathfrak{R}}((\alpha \cup \beta)^-) = \mathfrak{m}_{\mathfrak{R}}(\alpha^- \cup \beta^-)$
- (ii)  $\mathfrak{m}_{\mathfrak{R}}((\alpha ; \beta)^-) = \mathfrak{m}_{\mathfrak{R}}(\beta^- ; \alpha^-)$
- (iii)  $\mathfrak{m}_{\mathfrak{R}}(\varphi?^-) = \mathfrak{m}_{\mathfrak{R}}(\varphi?)$
- (iv)  $\mathfrak{m}_{\mathfrak{R}}(\alpha^{*-}) = \mathfrak{m}_{\mathfrak{R}}(\alpha^{-*})$
- (v)  $\mathfrak{m}_{\mathfrak{R}}(\alpha^{--}) = \mathfrak{m}_{\mathfrak{R}}(\alpha)$ .

**THEOREM 9.** *The following are valid formulas of PDL:*

- (i)  $\varphi \rightarrow [\alpha]\langle\alpha^-\rangle\varphi$
- (ii)  $\varphi \rightarrow [\alpha^-]\langle\alpha\rangle\varphi$
- (iii)  $\langle\alpha\rangle[\alpha^-]\varphi \rightarrow \varphi$
- (iv)  $\langle\alpha^-\rangle[\alpha]\varphi \rightarrow \varphi$ .

The iteration operator  $^*$  is interpreted as the reflexive transitive closure operator on binary relations. It is the means by which iteration is coded in PDL. This operator differs from the other operators in that it is infinitary in nature, as reflected by its semantics:

$$\mathfrak{m}_{\mathfrak{R}}(\alpha^*) = \mathfrak{m}_{\mathfrak{R}}(\alpha)^* = \bigcup_{n < \omega} \mathfrak{m}_{\mathfrak{R}}(\alpha)^n$$

(see Section 2.2). This introduces a level of complexity to PDL beyond the other operators. Because of it, PDL is not compact: the set

$$(9) \quad \{\langle\alpha^*\rangle\varphi\} \cup \{\neg\varphi, \neg\langle\alpha\rangle\varphi, \neg\langle\alpha^2\rangle\varphi, \dots\}$$

is finitely satisfiable but not satisfiable. Because of this infinitary behavior, it is rather surprising that PDL should be decidable and that there should be a finitary complete axiomatization.

The properties of the  $*$  operator of PDL come directly from the properties of the reflexive transitive closure operator  $*$  on binary relations. In a nutshell, for any binary relation  $R$ ,  $R^*$  is the  $\subseteq$ -least reflexive and transitive relation containing  $R$ .

**THEOREM 10.** *The following are valid formulas of PDL:*

- (i)  $[\alpha^*]\varphi \rightarrow \varphi$
- (ii)  $\varphi \rightarrow \langle \alpha^* \rangle \varphi$
- (iii)  $[\alpha^*]\varphi \rightarrow [\alpha]\varphi$
- (iv)  $\langle \alpha \rangle \varphi \rightarrow \langle \alpha^* \rangle \varphi$
- (v)  $[\alpha^*]\varphi \leftrightarrow [\alpha^*\alpha^*]\varphi$
- (vi)  $\langle \alpha^* \rangle \varphi \leftrightarrow \langle \alpha^*\alpha^* \rangle \varphi$
- (vii)  $[\alpha^*]\varphi \leftrightarrow [\alpha^{**}]\varphi$
- (viii)  $\langle \alpha^* \rangle \varphi \leftrightarrow \langle \alpha^{**} \rangle \varphi$
- (ix)  $[\alpha^*]\varphi \leftrightarrow \varphi \wedge [\alpha][\alpha^*]\varphi.$
- (x)  $\langle \alpha^* \rangle \varphi \leftrightarrow \varphi \vee \langle \alpha \rangle \langle \alpha^* \rangle \varphi.$
- (xi)  $[\alpha^*]\varphi \leftrightarrow \varphi \wedge [\alpha^*](\varphi \rightarrow [\alpha]\varphi).$
- (xii)  $\langle \alpha^* \rangle \varphi \leftrightarrow \varphi \vee \langle \alpha^* \rangle (\neg \varphi \wedge \langle \alpha \rangle \varphi).$

Semantically,  $\alpha^*$  is a reflexive and transitive relation containing  $\alpha$ , and Theorem 10 captures this. That  $\alpha^*$  is reflexive is captured in (ii); that it is transitive is captured in (vi); and that it contains  $\alpha$  is captured in (iv). These three properties are captured by the single property (x).

### *Reflexive Transitive Closure and Induction*

To prove properties of iteration, it is not enough to know that  $\alpha^*$  is a reflexive and transitive relation containing  $\alpha$ . So is the universal relation  $K \times K$ , and that is not very interesting. We also need some way of capturing the idea that  $\alpha^*$  is the *least* reflexive and transitive relation containing  $\alpha$ . There are several equivalent ways this can be done:

**(RTC)** The *reflexive transitive closure rule*:

$$\frac{(\varphi \vee \langle \alpha \rangle \psi) \rightarrow \psi}{\langle \alpha^* \rangle \varphi \rightarrow \psi}$$

**(LI)** The *loop invariance rule*:

$$\frac{\psi \rightarrow [\alpha] \psi}{\psi \rightarrow [\alpha^*] \psi}$$

**(IND)** The *induction axiom* (box form):

$$\varphi \wedge [\alpha^*](\varphi \rightarrow [\alpha] \varphi) \rightarrow [\alpha^*] \varphi$$

**(IND)** The *induction axiom* (diamond form):

$$\langle \alpha^* \rangle \varphi \rightarrow \varphi \vee \langle \alpha^* \rangle (\neg \varphi \wedge \langle \alpha \rangle \varphi)$$

The rule (RTC) is called the *reflexive transitive closure rule*. Its importance is best described in terms of its relationship to the valid PDL formula of Theorem 10(x). Observe that the right-to-left implication of this formula is obtained by substituting  $\langle \alpha^* \rangle \varphi$  for  $R$  in the expression

$$(10) \quad \varphi \vee \langle \alpha \rangle R \rightarrow R.$$

Theorem 10(x) implies that  $\langle \alpha^* \rangle \varphi$  is a solution of (10); that is, (10) is valid when  $\langle \alpha^* \rangle \varphi$  is substituted for  $R$ . The rule (RTC) says that  $\langle \alpha^* \rangle \varphi$  is the *least* such solution with respect to logical implication. That is, it is the least PDL-definable set of states that when substituted for  $R$  in (10) results in a valid formula.

The dual propositions labeled (IND) are jointly called the PDL *induction axiom*. Intuitively, the box form of (IND) says, “If  $\varphi$  is true initially, and if, after any number of iterations of the program  $\alpha$ , the truth of  $\varphi$  is preserved by one more iteration of  $\alpha$ , then  $\varphi$  will be true after any number of iterations of  $\alpha$ .” The diamond form of (IND) says, “If it is possible to reach a state satisfying  $\varphi$  in some number of iterations of  $\alpha$ , then either  $\varphi$  is true now, or it is possible to reach a state in which  $\varphi$  is false but becomes true after one more iteration of  $\alpha$ .”

Note that the box form of (IND) bears a strong resemblance to the induction axiom of Peano arithmetic:

$$\varphi(0) \wedge \forall n (\varphi(n) \rightarrow \varphi(n+1)) \rightarrow \forall n \varphi(n).$$

Here  $\varphi(0)$  is the basis of the induction and  $\forall n (\varphi(n) \rightarrow \varphi(n+1))$  is the induction step, from which the conclusion  $\forall n \varphi(n)$  can be drawn. In the PDL axiom (IND), the basis is  $\varphi$  and the induction step is  $[\alpha^*](\varphi \rightarrow [\alpha]\varphi)$ , from which the conclusion  $[\alpha^*]\varphi$  can be drawn.

## 2.6 Encoding Hoare Logic

The Hoare partial correctness assertion  $\{\varphi\} \alpha \{\psi\}$  is encoded as  $\varphi \rightarrow [\alpha]\psi$  in PDL. The following theorem says that under this encoding, Dynamic Logic subsumes Hoare Logic.

**THEOREM 11.** *The following rules of Hoare Logic are derivable in PDL:*

(i) *Composition rule:*

$$\frac{\{\varphi\} \alpha \{\sigma\}, \{\sigma\} \beta \{\psi\}}{\{\varphi\} \alpha; \beta \{\psi\}}$$

(ii) *Conditional rule:*

$$\frac{\{\varphi \wedge \sigma\} \alpha \{\psi\}, \{\neg\varphi \wedge \sigma\} \beta \{\psi\}}{\{\sigma\} \text{ if } \varphi \text{ then } \alpha \text{ else } \beta \{\psi\}}$$

(iii) **While rule:**

$$\frac{\{\varphi \wedge \psi\} \alpha \{\psi\}}{\{\psi\} \text{ while } \varphi \text{ do } \alpha \{\neg\varphi \wedge \psi\}}$$

(iv) *Weakening rule:*

$$\frac{\varphi' \rightarrow \varphi, \{\varphi\} \alpha \{\psi\}, \psi \rightarrow \psi'}{\{\varphi'\} \alpha \{\psi'\}}$$

### 3 FILTRATION AND DECIDABILITY

The *small model property* for PDL says that if  $\varphi$  is satisfiable, then it is satisfied at a state in a Kripke frame with no more than  $2^{|\varphi|}$  states, where  $|\varphi|$  is the number of symbols of  $\varphi$ . This result and the technique used to prove it, called *filtration*, come directly from modal logic. This immediately gives a naive decision procedure for the satisfiability problem for PDL: to determine whether  $\varphi$  is satisfiable, construct all Kripke frames with at most  $2^{|\varphi|}$  states and check whether  $\varphi$  is satisfied at some state in one of them. Considering only interpretations of the primitive formulas and primitive programs appearing in  $\varphi$ , there are roughly  $2^{2^{|\varphi|}}$  such models, so this algorithm is too inefficient to be practical. A more efficient algorithm will be described in Section 5.

#### 3.1 The Fischer–Ladner Closure

Many proofs in simpler modal systems use induction on the well-founded subformula relation. In PDL, the situation is complicated by the simultaneous inductive definitions of programs and propositions and by the behavior of the  $*$  operator, which make the induction proofs somewhat tricky. Nevertheless, we can still use the well-founded subexpression relation in inductive proofs. Here an *expression* can be either a program or a proposition. Either one can be a subexpression of the other because of the mixed operators  $[ ]$  and  $?$ .

We start by defining two functions

$$\begin{aligned} FL & : \Phi \rightarrow 2^\Phi \\ FL^\square & : \{[\alpha]\varphi \mid \alpha \in \Psi, \varphi \in \Phi\} \rightarrow 2^\Phi \end{aligned}$$

by simultaneous induction. The set  $FL(\varphi)$  is called the *Fischer–Ladner closure* of  $\varphi$ . The filtration construction for PDL uses the Fischer–Ladner closure of a given formula where the corresponding proof for propositional modal logic would use the set of subformulas.

The functions  $FL$  and  $FL^\square$  are defined inductively as follows:

- (a)  $FL(p) \stackrel{\text{def}}{=} \{p\}$ ,  $p$  an atomic proposition
- (b)  $FL(\varphi \rightarrow \psi) \stackrel{\text{def}}{=} \{\varphi \rightarrow \psi\} \cup FL(\varphi) \cup FL(\psi)$
- (c)  $FL(\mathbf{0}) \stackrel{\text{def}}{=} \{\mathbf{0}\}$
- (d)  $FL([\alpha]\varphi) \stackrel{\text{def}}{=} FL^\square([\alpha]\varphi) \cup FL(\varphi)$
- (e)  $FL^\square([a]\varphi) \stackrel{\text{def}}{=} \{[a]\varphi\}$ ,  $a$  an atomic program

- (f)  $FL^\square([\alpha \cup \beta]\varphi) \stackrel{\text{def}}{=} \{[\alpha \cup \beta]\varphi\} \cup FL^\square([\alpha]\varphi) \cup FL^\square([\beta]\varphi)$
- (g)  $FL^\square([\alpha ; \beta]\varphi) \stackrel{\text{def}}{=} \{[\alpha ; \beta]\varphi\} \cup FL^\square([\alpha][\beta]\varphi) \cup FL^\square([\beta]\varphi)$
- (h)  $FL^\square([\alpha^*]\varphi) \stackrel{\text{def}}{=} \{[\alpha^*]\varphi\} \cup FL^\square([\alpha][\alpha^*]\varphi)$
- (i)  $FL^\square([\psi?]\varphi) \stackrel{\text{def}}{=} \{[\psi?]\varphi\} \cup FL(\psi).$

This definition is apparently quite a bit more involved than for mere subexpressions. In fact, at first glance it may appear circular because of the rule (h). The auxiliary function  $FL^\square$  is introduced for the express purpose of avoiding any such circularity. It is defined only for formulas of the form  $[\alpha]\varphi$  and intuitively produces those elements of  $FL([\alpha]\varphi)$  obtained by breaking down  $\alpha$  and ignoring  $\varphi$ .

LEMMA 12.

- (i) *If  $[\alpha]\psi \in FL(\varphi)$ , then  $\psi \in FL(\varphi)$ .*
- (ii) *If  $[\rho?]\psi \in FL(\varphi)$ , then  $\rho \in FL(\varphi)$ .*
- (iii) *If  $[\alpha \cup \beta]\psi \in FL(\varphi)$ , then  $[\alpha]\psi \in FL(\varphi)$  and  $[\beta]\psi \in FL(\varphi)$ .*
- (iv) *If  $[\alpha ; \beta]\psi \in FL(\varphi)$ , then  $[\alpha][\beta]\psi \in FL(\varphi)$  and  $[\beta]\psi \in FL(\varphi)$ .*
- (v) *If  $[\alpha^*]\psi \in FL(\varphi)$ , then  $[\alpha][\alpha^*]\psi \in FL(\varphi)$ .*

Even after convincing ourselves that the definition is noncircular, it may not be clear how the size of  $FL(\varphi)$  depends on the length of  $\varphi$ . Indeed, the right-hand side of rule (h) involves a formula that is larger than the formula on the left-hand side. However, it can be shown by induction on subformulas that the relationship is linear:

LEMMA 13.

- (i) *For any formula  $\varphi$ ,  $\#FL(\varphi) \leq |\varphi|$ .*
- (ii) *For any formula  $[\alpha]\varphi$ ,  $\#FL^\square([\alpha]\varphi) \leq |\alpha|$ .*

### 3.2 Filtration

Given a PDL proposition  $\varphi$  and a Kripke frame  $\mathfrak{K} = (K, \mathfrak{m}_{\mathfrak{K}})$ , we define a new frame  $\mathfrak{K}/FL(\varphi) = (K/FL(\varphi), \mathfrak{m}_{\mathfrak{K}/FL(\varphi)})$ , called the *filtration of  $\mathfrak{K}$  by  $FL(\varphi)$* , as follows. Define a binary relation  $\equiv$  on states of  $\mathfrak{K}$  by:

$$u \equiv v \stackrel{\text{def}}{\iff} \forall \psi \in FL(\varphi) (u \in \mathfrak{m}_{\mathfrak{K}}(\psi) \iff v \in \mathfrak{m}_{\mathfrak{K}}(\psi)).$$

In other words, we collapse states  $u$  and  $v$  if they are not distinguishable by any formula of  $FL(\varphi)$ . Let

$$\begin{aligned} [u] &\stackrel{\text{def}}{=} \{v \mid v \equiv u\} \\ K/FL(\varphi) &\stackrel{\text{def}}{=} \{[u] \mid u \in K\} \\ \mathfrak{m}_{\mathfrak{K}/FL(\varphi)}(p) &\stackrel{\text{def}}{=} \{[u] \mid u \in \mathfrak{m}_{\mathfrak{K}}(p)\}, \quad p \text{ an atomic proposition} \\ \mathfrak{m}_{\mathfrak{K}/FL(\varphi)}(a) &\stackrel{\text{def}}{=} \{([u], [v]) \mid (u, v) \in \mathfrak{m}_{\mathfrak{K}}(a)\}, \quad a \text{ an atomic program.} \end{aligned}$$

The map  $\mathfrak{m}_{\mathfrak{K}/FL(\varphi)}$  is extended inductively to compound propositions and programs as described in Section 2.2.

The following key lemma relates  $\mathfrak{K}$  and  $\mathfrak{K}/FL(\varphi)$ . Most of the difficulty in the following lemma is in the correct formulation of the induction hypotheses in the statement of the lemma. Once this is done, the proof is a fairly straightforward induction on the well-founded subexpression relation.

**LEMMA 14 (Filtration Lemma).** *Let  $\mathfrak{K}$  be a Kripke frame and let  $u, v$  be states of  $\mathfrak{K}$ .*

- (i) *For all  $\psi \in FL(\varphi)$ ,  $u \in \mathfrak{m}_{\mathfrak{K}}(\psi)$  iff  $[u] \in \mathfrak{m}_{\mathfrak{K}/FL(\varphi)}(\psi)$ .*
- (ii) *For all  $[\alpha]\psi \in FL(\varphi)$ ,*
  - (a) *if  $(u, v) \in \mathfrak{m}_{\mathfrak{K}}(\alpha)$  then  $([u], [v]) \in \mathfrak{m}_{\mathfrak{K}/FL(\varphi)}(\alpha)$ ;*
  - (b) *if  $([u], [v]) \in \mathfrak{m}_{\mathfrak{K}/FL(\varphi)}(\alpha)$  and  $u \in \mathfrak{m}_{\mathfrak{K}}([\alpha]\psi)$ , then  $v \in \mathfrak{m}_{\mathfrak{K}}(\psi)$ .*

Using the filtration lemma, we can prove the small model theorem easily.

**THEOREM 15 (Small Model Theorem).** *Let  $\varphi$  be a satisfiable formula of PDL. Then  $\varphi$  is satisfied in a Kripke frame with no more than  $2^{|\varphi|}$  states.*

**Proof.** If  $\varphi$  is satisfiable, then there is a Kripke frame  $\mathfrak{K}$  and state  $u \in \mathfrak{K}$  with  $u \in \mathfrak{m}_{\mathfrak{K}}(\varphi)$ . Let  $FL(\varphi)$  be the Fischer-Ladner closure of  $\varphi$ . By the filtration lemma (Lemma 14),  $[u] \in \mathfrak{m}_{\mathfrak{K}/FL(\varphi)}(\varphi)$ . Moreover,  $\mathfrak{K}/FL(\varphi)$  has no more states than the number of truth assignments to formulas in  $FL(\varphi)$ , which by Lemma 13(i) is at most  $2^{|\varphi|}$ . ■

It follows immediately that the satisfiability problem for PDL is decidable, since there are only finitely many possible Kripke frames of size at most  $2^{|\varphi|}$  to check, and there is a polynomial-time algorithm to check whether a given formula is satisfied at a given state in a given Kripke frame. A more efficient algorithm exists (see Section 5).

The completeness proof for PDL also makes use of the filtration lemma (Lemma 14), but in a somewhat stronger form. We need to know that it also holds for *nonstandard Kripke frames* as well as the standard Kripke frames defined in Section 2.2.



A *nonstandard Kripke frame* is any structure  $\mathfrak{N} = (N, m_{\mathfrak{N}})$  that is a Kripke frame in the sense of Section 2.2 in every respect, except that  $m_{\mathfrak{N}}(\alpha^*)$  need not be the reflexive transitive closure of  $m_{\mathfrak{N}}(\alpha)$ , but only a reflexive, transitive binary relation containing  $m_{\mathfrak{N}}(\alpha)$  satisfying the PDL axioms for \* (Axioms 17(vii) and (viii) of Section 4.1).

**LEMMA 16** (Filtration for Nonstandard Models). *Let  $\mathfrak{N}$  be a nonstandard Kripke frame and let  $u, v$  be states of  $\mathfrak{N}$ .*

- (i) *For all  $\psi \in FL(\varphi)$ ,  $u \in m_{\mathfrak{N}}(\psi)$  iff  $[u] \in m_{\mathfrak{N}/FL(\varphi)}(\psi)$ .*
- (ii) *For all  $[\alpha]\psi \in FL(\varphi)$ ,*
  - (a) *if  $(u, v) \in m_{\mathfrak{N}}(\alpha)$  then  $([u], [v]) \in m_{\mathfrak{N}/FL(\varphi)}(\alpha)$ ;*
  - (b) *if  $([u], [v]) \in m_{\mathfrak{N}/FL(\varphi)}(\alpha)$  and  $u \in m_{\mathfrak{N}}([\alpha]\psi)$ , then  $v \in m_{\mathfrak{N}}(\psi)$ .*

## 4 DEDUCTIVE COMPLETENESS OF PDL

### 4.1 A Deductive System

The following list of axioms and rules constitutes a sound and complete Hilbert-style deductive system for PDL.

**Axiom System 17.**

- (i) *Axioms for propositional logic*
- (ii)  $[\alpha](\varphi \rightarrow \psi) \rightarrow ([\alpha]\varphi \rightarrow [\alpha]\psi)$
- (iii)  $[\alpha](\varphi \wedge \psi) \leftrightarrow [\alpha]\varphi \wedge [\alpha]\psi$
- (iv)  $[\alpha \cup \beta]\varphi \leftrightarrow [\alpha]\varphi \wedge [\beta]\varphi$
- (v)  $[\alpha; \beta]\varphi \leftrightarrow [\alpha][\beta]\varphi$
- (vi)  $[\psi?]\varphi \leftrightarrow (\psi \rightarrow \varphi)$
- (vii)  $\varphi \wedge [\alpha][\alpha^*]\varphi \leftrightarrow [\alpha^*]\varphi$
- (viii)  $\varphi \wedge [\alpha^*](\varphi \rightarrow [\alpha]\varphi) \rightarrow [\alpha^*]\varphi$

*In PDL with converse  $^-$ , we also include*

- (ix)  $\varphi \rightarrow [\alpha]\langle\alpha^-\rangle\varphi$
- (x)  $\varphi \rightarrow [\alpha^-\rangle\langle\alpha\rangle\varphi$

*Rules of Inference*

$$(MP) \quad \frac{\varphi, \varphi \rightarrow \psi}{\psi}$$

$$(GEN) \quad \frac{\varphi}{[\alpha]\varphi}$$

□

The axioms (ii) and (iii) and the two rules of inference are not particular to PDL, but come from modal logic. The rules (MP) and (GEN) are called *modus ponens* and (*modal*) *generalization*, respectively.

Axiom (viii) is called the *PDL induction axiom*. Intuitively, (viii) says: “Suppose  $\varphi$  is true in the current state, and suppose that after any number of iterations of  $\alpha$ , if  $\varphi$  is still true, then it will be true after one more iteration of  $\alpha$ . Then  $\varphi$  will be true after any number of iterations of  $\alpha$ .” In other words, if  $\varphi$  is true initially, and if the truth of  $\varphi$  is preserved by the program  $\alpha$ , then  $\varphi$  will be true after any number of iterations of  $\alpha$ .

We write  $\vdash \varphi$  if the proposition  $\varphi$  is a theorem of this system, and say that  $\varphi$  is *consistent* if  $\not\vdash \neg\varphi$ ; that is, if it is not the case that  $\vdash \neg\varphi$ . A set  $\Sigma$  of propositions is *consistent* if all finite conjunctions of elements of  $\Sigma$  are consistent.

The soundness of these axioms and rules over Kripke frames can be established by elementary arguments in relational algebra using the semantics of Section 2.2.

We write  $\vdash \varphi$  if the formula  $\varphi$  is provable in this deductive system. A formula  $\varphi$  is *consistent* if  $\not\vdash \neg\varphi$ , that is, if it is not the case that  $\vdash \neg\varphi$ ; that a finite set  $\Sigma$  of formulas is *consistent* if its conjunction  $\bigwedge \Sigma$  is consistent; and that an infinite set of formulas is *consistent* if every finite subset is consistent.

Axiom System 17 is complete: all valid formulas of PDL are theorems. This fact can be proved by constructing a nonstandard Kripke frame from maximal consistent sets of formulas, then using the filtration lemma for nonstandard models (Lemma 16) to collapse this nonstandard model to a finite standard model.

**THEOREM 18 (Completeness of PDL).** *If  $\models \varphi$  then  $\vdash \varphi$ .*

In classical logics, a completeness theorem of the form of Theorem 18 can be adapted to handle the relation of logical consequence  $\varphi \models \psi$  between formulas because of the deduction theorem, which says

$$\varphi \vdash \psi \Leftrightarrow \vdash \varphi \rightarrow \psi.$$

Unfortunately, the deduction theorem fails in PDL, as can be seen by taking  $\psi = [a]p$  and  $\varphi = p$ . However, the following result allows Theorem

18, as well as the deterministic exponential-time satisfiability algorithm described in the next section, to be extended to handle the logical consequence relation:

**THEOREM 19.** *Let  $\varphi$  and  $\psi$  be any PDL formulas. Then*

$$\varphi \models \psi \Leftrightarrow \models [(a_1 \cup \dots \cup a_n)^*] \varphi \rightarrow \psi,$$

where  $a_1, \dots, a_n$  are all atomic programs appearing in  $\varphi$  or  $\psi$ . Allowing infinitary conjunctions, if  $\Sigma$  is a set of formulas in which only finitely many atomic programs appear, then

$$\Sigma \models \psi \Leftrightarrow \models \bigwedge \{[(a_1 \cup \dots \cup a_n)^*] \varphi \mid \varphi \in \Sigma\} \rightarrow \psi,$$

where  $a_1, \dots, a_n$  are all atomic programs appearing in  $\Sigma$  or  $\psi$ .

## 5 COMPLEXITY OF PDL

The small model theorem (Theorem 15) gives a naive deterministic algorithm for the satisfiability problem: construct all Kripke frames of at most  $2^{|\varphi|}$  states and check whether  $\varphi$  is satisfied at any state in any of them. Although checking whether a given formula is satisfied in a given state of a given Kripke frame can be done quite efficiently, the naive satisfiability algorithm is highly inefficient. For one thing, the models constructed are of exponential size in the length of the given formula; for another, there are  $2^{2^{O(|\varphi|)}}$  of them. Thus the naive satisfiability algorithm takes double exponential time in the worst case.

There is a more efficient algorithm [Pratt, 1979b] that runs in deterministic single-exponential time. One cannot expect to improve this significantly due to a corresponding lower bound.

**THEOREM 20.** *There is an exponential-time algorithm for deciding whether a given formula of PDL is satisfiable.*

**THEOREM 21.** *The satisfiability problem for PDL is EXPTIME-complete.*

**COROLLARY 22.** *There is a constant  $c > 1$  such that the satisfiability problem for PDL is not solvable in deterministic time  $c^{n/\log n}$ , where  $n$  is the size of the input formula.*

EXPTIME-hardness can be established by constructing a formula of PDL whose models encode the computation of a given linear-space-bounded one-tape alternating Turing machine  $M$  on a given input  $x$  of length  $n$  over  $M$ 's input alphabet. Since the membership problem for alternating polynomial-space machines is EXPTIME-hard [Chandra *et al.*, 1981], so is the satisfiability problem for PDL.

It is interesting to compare the complexity of satisfiability in PDL with the complexity of satisfiability in propositional logic. In the latter, satisfiability is *NP*-complete; but at present it is not known whether the two complexity classes *EXPTIME* and *NP* differ. Thus, as far as current knowledge goes, the satisfiability problem is no easier in the worst case for propositional logic than for its far richer superset PDL.

As we have seen, current knowledge does not permit a significant difference to be observed between the complexity of satisfiability in propositional logic and in PDL. However, there is one easily verified and important behavioral difference: propositional logic is *compact*, whereas PDL is not.

Compactness has significant implications regarding the relation of logical consequence. If a propositional formula  $\varphi$  is a consequence of a set  $\Gamma$  of propositional formulas, then it is already a consequence of some finite subset of  $\Gamma$ ; but this is not true in PDL.

Recall that we write  $\Gamma \models \varphi$  and say that  $\varphi$  is a *logical consequence* of  $\Gamma$  if  $\varphi$  is satisfied in any state of any Kripke frame  $\mathfrak{K}$  all of whose states satisfy all the formulas of  $\Gamma$ . That is, if  $\mathfrak{K} \models \Gamma$ , then  $\mathfrak{K} \models \varphi$ .

An alternative interpretation of logical consequence, not equivalent to the above, is that in any Kripke frame, the formula  $\varphi$  holds in any state satisfying all formulas in  $\Gamma$ . Allowing infinite conjunctions, we might write this as  $\models \bigwedge \Gamma \rightarrow \varphi$ . This is not the same as  $\Gamma \models \varphi$ , since  $\models \bigwedge \Gamma \rightarrow \varphi$  implies  $\Gamma \models \varphi$ , but not necessarily vice versa. A counterexample is provided by  $\Gamma = \{p\}$  and  $\varphi = [a]p$ . However, if  $\Gamma$  contains only finitely many atomic programs, we can reduce the problem  $\Gamma \models \varphi$  to the problem  $\models \bigwedge \Gamma' \rightarrow \varphi$  for a related  $\Gamma'$ , as shown in Theorem 19.

Under either interpretation, compactness fails:

**THEOREM 23.** *There is an infinite set of formulas  $\Gamma$  and a formula  $\varphi$  such that  $\models \bigwedge \Gamma \rightarrow \varphi$  (hence  $\Gamma \models \varphi$ ), but for no proper subset  $\Gamma' \subsetneq \Gamma$  is it the case that  $\Gamma' \models \varphi$  (hence neither is it the case that  $\models \bigwedge \Gamma' \rightarrow \varphi$ ).*

As shown in Theorem 19, logical consequences  $\Gamma \models \varphi$  for finite  $\Gamma$  are no more difficult to decide than validity of single formulas. But what if  $\Gamma$  is infinite? Here compactness is the key factor. If  $\Gamma$  is an r.e. set and the logic is compact, then the consequence problem is r.e.: to check whether  $\Gamma \models \varphi$ , the finite subsets of  $\Gamma$  can be effectively enumerated, and checking  $\Gamma \models \varphi$  for finite  $\Gamma$  is a decidable problem.

Since compactness fails in PDL, this observation does us no good, even when  $\Gamma$  is known to be recursively enumerable. However, the following result shows that the situation is much worse than we might expect: even if  $\Gamma$  is taken to be the set of substitution instances of a single formula of PDL, the consequence problem becomes very highly undecidable. This is a rather striking manifestation of PDL's lack of compactness.

Let  $\varphi$  be a given formula. The set  $S_\varphi$  of *substitution instances* of  $\varphi$  is the set of all formulas obtained by substituting a formula for each atomic proposition appearing in  $\varphi$ .

**THEOREM 24.** *The problem of deciding whether  $S_\varphi \models \psi$  is  $\Pi_1^1$ -complete. The problem is  $\Pi_1^1$ -hard even for a particular fixed  $\varphi$ .*

## 6 NONREGULAR PDL

In this section we enrich the class of regular programs in PDL by introducing programs whose control structure requires more than a finite automaton. For example, the class of *context-free programs* requires a pushdown automaton (PDA), and moving up from regular to context-free programs is really going from iterative programs to ones with parameterless recursive procedures. Several questions arise when enriching the class of programs of PDL, such as whether the expressive power of the logic grows, and if so whether the resulting logics are still decidable. It turns out that *any* nonregular program increases PDL's expressive power and that the validity problem for PDL with context-free programs is undecidable. The bulk of the section is then devoted to the difficult problem of trying to characterize the borderline between decidable and undecidable extensions. On the one hand, validity for PDL with the addition of even a single extremely simple nonregular program is already  $\Pi_1^1$ -complete; but on the other hand, when we add another equally simple program, the problem remains decidable. Besides these results, which pertain to very specific extensions, we discuss some broad decidability results that cover many languages, including some that are not even context-free. Since no similarly general undecidability results are known, we also address the weaker issue of whether nonregular extensions admit the finite model property and present a negative result that covers many cases.

### 6.1 Nonregular Programs

Consider the following self-explanatory program:

(11) **while**  $p$  **do**  $a$ ; **now do**  $b$  **the same number of times**

This program is meant to represent the following set of computation sequences:

$$\{(p? ; a)^i ; \neg p? ; b^i \mid i \geq 0\}.$$

Viewed as a language over the alphabet  $\{a, b, p, \neg p\}$ , this set is not regular, thus cannot be programmed in PDL. However, it can be represented by the following parameterless recursive procedure:

```

proc  $V$  {
  if  $p$  then {  $a$ ; call  $V$ ;  $b$  }
  else return
}

```

The set of computation sequences of this program is captured by the context-free grammar

$$V \rightarrow \neg p? \mid p?aVb.$$

We are thus led to the idea of allowing context-free programs inside the boxes and diamonds of PDL. From a pragmatic point of view, this amounts to extending the logic with the ability to reason about parameterless recursive procedures. The particular representation of the context-free programs is unimportant; we can use pushdown automata, context-free grammars, recursive procedures, or any other formalism that can be effectively translated into these.

In the rest of this section, a number of specific programs will be of interest, and we employ special abbreviations for them. For example, we define:

$$\begin{aligned} a^\Delta b a^\Delta &\stackrel{\text{def}}{=} \{a^i b a^i \mid i \geq 0\} \\ a^\Delta b^\Delta &\stackrel{\text{def}}{=} \{a^i b^i \mid i \geq 0\} \\ b^\Delta a^\Delta &\stackrel{\text{def}}{=} \{b^i a^i \mid i \geq 0\}. \end{aligned}$$

Note that  $a^\Delta b^\Delta$  is really just a nondeterministic version of the program (11) in which there is simply no  $p$  to control the iteration. In fact, (11) could have been written in this notation as  $(p?a)^\Delta \neg p?b^\Delta$ .<sup>2</sup> In programming terms, we can compare the regular program  $(ab)^*$  with the nonregular one  $a^\Delta b^\Delta$  by observing that if  $a$  is “purchase a loaf of bread” and  $b$  is “pay \$1.00,” then the former program captures the process of paying for each loaf when purchased, while the latter one captures the process of paying for them all at the end of the month.

It turns out that enriching PDL with even a single arbitrary nonregular program increases expressive power.

If  $L$  is any language over atomic programs and tests, then  $\text{PDL} + L$  is defined exactly as PDL, but with the additional syntax rule stating that for any formula  $\varphi$ , the expression  $\langle L \rangle \varphi$  is a new formula. The semantics of  $\text{PDL} + L$  is like that of PDL with the addition of the clause

$$m_{\bar{\kappa}}(L) \stackrel{\text{def}}{=} \bigcup_{\beta \in L} m_{\bar{\kappa}}(\beta).$$

---

<sup>2</sup>It is noteworthy that the results of this section do not depend on nondeterminism. For example, the negative Theorem 28 holds for the deterministic version (11) too. Also, most of the results in this section involve nonregular programs over atomic programs only, but can be generalized to allow tests as well.

Note that  $\text{PDL} + L$  does not allow  $L$  to be used as a formation rule for new programs or to be combined with other programs. It is added to the programming language as a single new stand-alone program only.

If  $\text{PDL}_1$  and  $\text{PDL}_2$  are two extensions of  $\text{PDL}$ , we say that  $\text{PDL}_1$  is *as expressive as*  $\text{PDL}_2$  if for each formula  $\varphi$  of  $\text{PDL}_2$  there is a formula  $\psi$  of  $\text{PDL}_1$  such that  $\models \varphi \leftrightarrow \psi$ . If  $\text{PDL}_1$  is as expressive as  $\text{PDL}_2$  but  $\text{PDL}_2$  is not as expressive as  $\text{PDL}_1$ , we say that  $\text{PDL}_1$  is *strictly more expressive than*  $\text{PDL}_2$ .

Thus, one version of  $\text{PDL}$  is strictly more expressive than another if anything the latter can express the former can too, but there is something the former can express that the latter cannot.

A language is *test-free* if it is a subset of  $\Pi_0^*$ ; that is, if its seqs contain no tests.

**THEOREM 25.** *If  $L$  is any nonregular test-free language, then  $\text{PDL} + L$  is strictly more expressive than  $\text{PDL}$ .*

We can view the decidability of regular  $\text{PDL}$  as showing that propositional-level reasoning about iterative programs is computable. We now wish to know if the same is true for recursive procedures. We define *context-free*  $\text{PDL}$  to be  $\text{PDL}$  extended with context-free programs, where a *context-free program* is one whose seqs form a context-free language. The precise syntax is unimportant, but for definiteness we might take as programs the set of context-free grammars  $G$  over atomic programs and tests and define

$$\mathfrak{m}_{\mathfrak{R}}(G) \stackrel{\text{def}}{=} \bigcup_{\beta \in CS(G)} \mathfrak{m}_{\mathfrak{R}}(\beta),$$

where  $CS(G)$  is the set of computation sequences generated by  $G$  as described in Section 1.3.

**THEOREM 26.** *The validity problem for context-free  $\text{PDL}$  is undecidable.*

Theorem 26 leaves several interesting questions unanswered. What is the level of undecidability of context-free  $\text{PDL}$ ? What happens if we want to add only a small number of specific nonregular programs? The first of these questions arises from the fact that the equivalence problem for context-free languages is co-r.e.-complete, or complete for  $\Pi_1^0$  in the arithmetic hierarchy. Hence, all Theorem 26 shows is that the validity problem for context-free  $\text{PDL}$  is  $\Pi_1^0$ -hard, while it might in fact be worse. The second question is far more general. We might be interested in reasoning only about deterministic or linear context-free programs,<sup>3</sup> or we might be interested only in a few special context-free programs such as  $a^\Delta b a^\Delta$  or  $a^\Delta b^\Delta$ . Perhaps  $\text{PDL}$

---

<sup>3</sup>A *linear program* is one whose seqs are generated by a context-free grammar in which there is at most one nonterminal symbol on the right-hand side of each rule. This corresponds to a family of recursive procedures in which there is at most one recursive call in each procedure.

remains decidable when these programs are added. The general question is to determine the borderline between the decidable and the undecidable when it comes to enriching the class of programs allowed in PDL.

Interestingly, if we wish to consider such simple nonregular extensions as  $\text{PDL} + a^\Delta b a^\Delta$  or  $\text{PDL} + a^\Delta b^\Delta$ , we will not be able to prove undecidability by the technique used for context-free PDL in Theorem 26, since standard problems that are undecidable for context-free languages, such as equivalence and inclusion, are decidable for classes containing the regular languages and the likes of  $a^\Delta b a^\Delta$  and  $a^\Delta b^\Delta$ . Moreover, we cannot prove decidability by the technique used for PDL in Section 3.2, since logics like  $\text{PDL} + a^\Delta b a^\Delta$  and  $\text{PDL} + a^\Delta b^\Delta$  do not enjoy the finite model property. Thus, if we want to determine the decidability status of such extensions, we will have to work harder.

**THEOREM 27.** *There is a satisfiable formula in  $\text{PDL} + a^\Delta b^\Delta$  that is not satisfied in any finite structure.*

For  $\text{PDL} + a^\Delta b a^\Delta$ , the news is worse than mere undecidability:

**THEOREM 28.** *The validity problem for  $\text{PDL} + a^\Delta b a^\Delta$  is  $\Pi_1^1$ -complete.*

The  $\Pi_1^1$  result holds also for PDL extended with the two programs  $a^\Delta b^\Delta$  and  $b^\Delta a^\Delta$ .

It is easy to show that the validity problem for context-free PDL in its entirety remains in  $\Pi_1^1$ . Together with the fact that  $a^\Delta b a^\Delta$  is a context-free language, this yields an answer to the first question mentioned earlier: context-free PDL is  $\Pi_1^1$ -complete. As to the second question, Theorem 28 shows that the high undecidability phenomenon starts occurring even with the addition of one very simple nonregular program.

We now turn to nonregular programs over a single letter. Consider the language of powers of 2:

$$a^{2^*} \stackrel{\text{def}}{=} \{a^{2^i} \mid i \geq 0\}.$$

Here we have:

**THEOREM 29.** *The validity problem for  $\text{PDL} + a^{2^*}$  is undecidable.*

It is actually possible to prove this result for powers of any fixed  $k \geq 2$ . Thus PDL with the addition of any language of the form  $\{a^{k^i} \mid i \geq 0\}$  for fixed  $k \geq 2$  is undecidable. Another class of one-letter extensions that has been proven to be undecidable consists of Fibonacci-like sequences:

**THEOREM 30.** *Let  $f_0, f_1$  be arbitrary elements of  $\mathbb{N}$  with  $f_0 < f_1$ , and let  $F$  be the sequence  $f_0, f_1, f_2, \dots$  generated by the recurrence  $f_i = f_{i-1} + f_{i-2}$  for  $i \geq 2$ . Let  $a^F \stackrel{\text{def}}{=} \{a^{f_i} \mid i \geq 0\}$ . Then the validity problem for  $\text{PDL} + a^F$  is undecidable.*

In both these theorems, the fact that the sequences of  $a$ 's in the programs grow exponentially is crucial to the proofs. Indeed, we know of no



undecidability results for any one-letter extension in which the lengths of the sequences of  $a$ 's grow subexponentially. Particularly intriguing are the cases of squares and cubes:

$$\begin{aligned} a^{*2} &\stackrel{\text{def}}{=} \{a^{i^2} \mid i \geq 0\}, \\ a^{*3} &\stackrel{\text{def}}{=} \{a^{i^3} \mid i \geq 0\}. \end{aligned}$$

Are  $\text{PDL} + a^{*2}$  and  $\text{PDL} + a^{*3}$  undecidable?

There is a decidability result for a slightly restricted version of the squares extension, which seems to indicate that the full unrestricted version  $\text{PDL} + a^{*2}$  is decidable too. However, we conjecture that for cubes the problem is undecidable. Interestingly, several classical open problems in number theory reduce to instances of the validity problem for  $\text{PDL} + a^{*3}$ . For example, while no one knows whether every integer greater than 10000 is the sum of five cubes, the following formula is valid if and only if the answer is yes:

$$[(a^{*3})^5]p \rightarrow [a^{10001}a^*]p.$$

(The 5-fold and 10001-fold iterations have to be written out in full, of course.) If  $\text{PDL} + a^{*3}$  were decidable, then we could compute the answer in a simple manner, at least in principle.

## 6.2 Decidable Extensions

We now turn to positive results. Theorem 27 states that  $\text{PDL} + a^\Delta b^\Delta$  does not have the finite model property. Nevertheless, we have the following:

**THEOREM 31.** *The validity problem for  $\text{PDL} + a^\Delta b^\Delta$  is decidable.*

When contrasted with Theorem 28, the decidability of  $\text{PDL} + a^\Delta b^\Delta$  is very surprising. We have two of the simplest nonregular languages— $a^\Delta b a^\Delta$  and  $a^\Delta b^\Delta$ —which are extremely similar, yet the addition of one to  $\text{PDL}$  yields high undecidability while the other leaves the logic decidable.

Theorem 31 was proved originally by showing that, although  $\text{PDL} + a^\Delta b^\Delta$  does not always admit finite models, it does admit finite *pushdown models*, in which transitions are labeled not only with atomic programs but also with push and pop instructions for a particular kind of stack. A close study of the proof (which relies heavily on the idiosyncrasies of the language  $a^\Delta b^\Delta$ ) suggests that the decidability or undecidability has to do with the manner in which an automaton accepts the languages involved. For example, in the usual way of accepting  $a^\Delta b a^\Delta$ , a pushdown automaton (PDA) reading an  $a$  will carry out a push or a pop, depending upon its location in the input word. However, in the standard way of accepting  $a^\Delta b^\Delta$ , the  $a$ 's are always pushed and the  $b$ 's are always popped, regardless of the location; the input symbol alone determines what the automaton does. More recent work, which we now set out to describe, has yielded a general decidability result

that confirms this intuition. It is of special interest due to its generality, since it does not depend on specific programs.

Let  $M = (Q, \Sigma, \Gamma, q_0, z_0, \delta)$  be a PDA that accepts by empty stack. We say that  $M$  is *simple-minded* if, whenever  $\delta(q, \sigma, \gamma) = (p, b)$ , then for each  $q'$  and  $\gamma'$ , either  $\delta(q', \sigma, \gamma') = (p, b)$  or  $\delta(q', \sigma, \gamma')$  is undefined. A context-free language is said to be *simple-minded* (a simple-minded CFL) if there exists a simple-minded PDA that accepts it.

In other words, the action of a simple-minded automaton is determined uniquely by the input symbol; the state and stack symbol are only used to help determine whether the machine halts (rejecting the input) or continues. Note that such an automaton is necessarily deterministic.

It is noteworthy that simple-minded PDAs accept a large fragment of the context-free languages, including  $a^\Delta b^\Delta$  and  $b^\Delta a^\Delta$ , as well as all balanced parenthesis languages (Dyck sets) and many of their intersections with regular languages.

**THEOREM 32.** *If  $L$  is accepted by a simple-minded PDA, then  $\text{PDL} + L$  is decidable.*

We can obtain another general decidability result involving languages accepted by deterministic stack automata. A stack automaton is a one-way PDA whose head can travel up and down the stack reading its contents, but can make changes only at the top of the stack. Stack automata can accept non-context-free languages such as  $a^\Delta b^\Delta c^\Delta$  and its generalizations  $a_1^\Delta a_2^\Delta \dots a_n^\Delta$  for any  $n$ , as well as many variants thereof. It would be nice to be able to prove decidability of PDL when augmented by any language accepted by such a machine, but this is not known. What has been proven, however, is that if each word in such a language is preceded by a new symbol to mark its beginning, then the enriched PDL is decidable:

**THEOREM 33.** *Let  $e \notin \Pi_0$ , and let  $L$  be a language over  $\Pi_0$  that is accepted by a deterministic stack automaton. If we let  $eL$  denote the language  $\{eu \mid u \in L\}$ , then  $\text{PDL} + eL$  is decidable.*

While Theorems 32 and 33 are general and cover many languages, they do not prove decidability of  $\text{PDL} + a^\Delta b^\Delta c^\Delta$ , which may be considered the simplest non-context-free extension of PDL. Nevertheless, the constructions used in the proofs of the two general results have been combined to yield:

**THEOREM 34.**  *$\text{PDL} + a^\Delta b^\Delta c^\Delta$  is decidable.*

As explained, we know of no undecidable extension of PDL with a polynomially growing language, although we conjecture that the cubes extension is undecidable. Since the decidability status of such extensions seems hard to determine, we now address a weaker notion: the presence or absence of a finite model property. The technique used in Theorem 27 to show that  $\text{PDL} + a^\Delta b^\Delta$  violates the finite model property does not work for one-letter alphabets. Nevertheless, we now state a general result leading to many one-

letter extensions that violate the finite model property. In particular, the theorem will yield the following:

**PROPOSITION 35** (squares and cubes). *The logics  $\text{PDL} + a^{*2}$  and  $\text{PDL} + a^{*3}$  do not have the finite model property.*

**PROPOSITION 36** (polynomials). *For every polynomial of the form*

$$p(n) = c_i n^i + c_{i-1} n^{i-1} + \cdots + c_0 \in \mathbb{Z}[n]$$

*with  $i \geq 2$  and positive leading coefficient  $c_i > 0$ , let  $S_p = \{p(m) \mid m \in \mathbb{N}\} \cap \mathbb{N}$ . Then  $\text{PDL} + a^{S_p}$  does not have the finite model property.*

**PROPOSITION 37** (sums of primes). *Let  $p_i$  be the  $i^{\text{th}}$  prime (with  $p_1 = 2$ ), and define*

$$S_{\text{sop}} \stackrel{\text{def}}{=} \left\{ \sum_{i=1}^n p_i \mid n \geq 1 \right\}.$$

*Then  $\text{PDL} + a^{S_{\text{sop}}}$  does not have the finite model property.*

**PROPOSITION 38** (factorials). *Let  $S_{\text{fac}} \stackrel{\text{def}}{=} \{n! \mid n \in \mathbb{N}\}$ . Then  $\text{PDL} + a^{S_{\text{fac}}}$  does not have the finite model property.*

The finite model property fails for any sufficiently fast-growing integer linear recurrence, not just the Fibonacci sequence, although we do not know whether these extensions also render PDL undecidable. A  $k^{\text{th}}$ -order integer linear recurrence is an inductively defined sequence

$$(12) \quad \ell_n \stackrel{\text{def}}{=} c_1 \ell_{n-1} + \cdots + c_k \ell_{n-k} + c_0, \quad n \geq k,$$

where  $k \geq 1$ ,  $c_0, \dots, c_k \in \mathbb{N}$ ,  $c_k \neq 0$ , and  $\ell_0, \dots, \ell_{k-1} \in \mathbb{N}$  are given.

**PROPOSITION 39** (linear recurrences). *Let  $S_{\text{lr}} = \{\ell_n \mid n \geq 0\}$  be the set defined inductively by (12). The following conditions are equivalent:*

- (i)  $a^{S_{\text{lr}}}$  is nonregular;
- (ii)  $\text{PDL} + a^{S_{\text{lr}}}$  does not have the finite model property;
- (iii) not all  $\ell_0, \dots, \ell_{k-1}$  are zero and  $\sum_{i=1}^k c_i > 1$ .

## 7 OTHER VARIANTS OF PDL

### 7.1 Deterministic Programs

Nondeterminism arises in PDL in two ways:

- atomic programs can be interpreted in a structure as (not necessarily single-valued) binary relations on states; and

- the programming constructs  $\alpha \cup \beta$  and  $\alpha^*$  involve nondeterministic choice.

Many modern programming languages have facilities for concurrency and distributed computation, certain aspects of which can be modeled by non-determinism. Nevertheless, the majority of programs written in practice are still deterministic. Here we investigate the effect of eliminating either one or both of these sources of nondeterminism from PDL.

A program  $\alpha$  is said to be (*semantically*) *deterministic* in a Kripke frame  $\mathfrak{K}$  if its traces are uniquely determined by their first states. If  $\alpha$  is an atomic program  $a$ , this is equivalent to the requirement that  $\mathfrak{m}_{\mathfrak{K}}(a)$  be a partial function; that is, if both  $(s, t)$  and  $(s, t') \in \mathfrak{m}_{\mathfrak{K}}(a)$ , then  $t = t'$ . A *deterministic Kripke frame*  $\mathfrak{K} = (K, \mathfrak{m}_{\mathfrak{K}})$  is one in which all atomic  $a$  are semantically deterministic.

The class of *deterministic while programs*, denoted DWP, is the class of programs in which

- the operators  $\cup$ ,  $?$ , and  $*$  may appear only in the context of the conditional test, **while** loop, **skip**, or **fail**;
- tests in the conditional test and **while** loop are purely propositional; that is, there is no occurrence of the  $\langle \rangle$  or  $[ ]$  operators.

The class of *nondeterministic while programs*, denoted WP, is the same, except unconstrained use of the nondeterministic choice construct  $\cup$  is allowed. It is easily shown that if  $\alpha$  and  $\beta$  are semantically deterministic in  $\mathfrak{K}$ , then so are **if**  $\varphi$  **then**  $\alpha$  **else**  $\beta$  and **while**  $\varphi$  **do**  $\alpha$ .

By restricting either the syntax or the semantics or both, we obtain the following logics:

- DPDL (deterministic PDL), which is syntactically identical to PDL, but interpreted over deterministic structures only;
- SPDL (strict PDL), in which only deterministic **while** programs are allowed; and
- SDPDL (strict deterministic PDL), in which both restrictions are in force.

Validity and satisfiability in DPDL and SDPDL are defined just as in PDL, but with respect to deterministic structures only. If  $\varphi$  is valid in PDL, then  $\varphi$  is also valid in DPDL, but not conversely: the formula

$$(13) \langle a \rangle \varphi \rightarrow [a] \varphi$$

is valid in DPDL but not in PDL. Also, SPDL and SDPDL are strictly less expressive than PDL or DPDL, since the formula

$$(14) \langle (a \cup b)^* \rangle \varphi$$

is not expressible in SPDL, as shown in [Halpern and Reif, 1983].

**THEOREM 40.** *If the axiom scheme*

$$(15) \langle a \rangle \varphi \rightarrow [a] \varphi, \quad a \in \Pi_0$$

*is added to Axiom System 17, then the resulting system is sound and complete for DPDL.*

**THEOREM 41.** *Validity in DPDL is deterministic exponential-time complete.*

Now we turn to SPDL, in which atomic programs can be nondeterministic but can be composed into larger programs only with deterministic constructs.

**THEOREM 42.** *Validity in SPDL is deterministic exponential-time complete.*

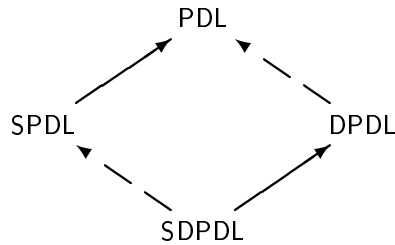
The final version of interest is SDPDL, in which both the syntactic restrictions of SPDL and the semantic ones of DPDL are adopted. The exponential-time lower bound fails here, and we have:

**THEOREM 43.** *The validity problem for SDPDL is complete in polynomial space.*

The question of relative power of expression is of interest here. Is DPDL < PDL? Is SDPDL < DPDL? The first of these questions is inappropriate, since the syntax of both languages is the same but they are interpreted over different classes of structures. Considering the second, we have:

**THEOREM 44.** SDPDL < DPDL *and* SPDL < PDL.

In summary, we have the following diagram describing the relations of expressiveness between these logics. The solid arrows indicate added expressive power and broken ones a difference in semantics. The validity problem is exponential-time complete for all but SDPDL, for which it is PSPACE-complete. Straightforward variants of Axiom System 17 are complete for all versions.



## 7.2 Representation by Automata

A PDL program represents a regular set of computation sequences. This same regular set could possibly be represented exponentially more succinctly

by a finite automaton. The difference between these two representations corresponds roughly to the difference between **while** programs and flowcharts.

Since finite automata are exponentially more succinct in general, the upper bound of Section 5 could conceivably fail if finite automata were allowed as programs. Moreover, we must also rework the deductive system of Section 4.1.

However, it turns out that the completeness and exponential-time decidability results of PDL are not sensitive to the representation and still go through in the presence of finite automata as programs, provided the deductive system of Section 4.1 and the techniques of Sections 4 and 5 are suitably modified, as shown in [Pratt, 1979b; Pratt, 1981b] and [Harel and Sherman, 1985].

In recent years, the automata-theoretic approach to logics of programs has yielded significant insight into propositional logics more powerful than PDL, as well as substantial reductions in the complexity of their decision procedures. Especially enlightening are the connections with automata on infinite strings and infinite trees. By viewing a formula as an automaton and a treelike model as an input to that automaton, the satisfiability problem for a given formula becomes the emptiness problem for a given automaton. Logical questions are thereby transformed into purely automata-theoretic questions.

We assume that nondeterministic finite automata are given in the form

$$(16) \quad M = (n, i, j, \delta),$$

where  $\bar{n} = \{0, \dots, n-1\}$  is the set of states,  $i, j \in \bar{n}$  are the start and final states respectively, and  $\delta$  assigns a subset of  $\Pi_0 \cup \{\varphi? \mid \varphi \in \Phi\}$  to each pair of states. Intuitively, when visiting state  $\ell$  and seeing symbol  $a$ , the automaton may move to state  $k$  if  $a \in \delta(\ell, k)$ .

The fact that the automata (16) have only one accept state is without loss of generality. If  $M$  is an arbitrary nondeterministic finite automaton with accept states  $F$ , then the set accepted by  $M$  is the union of the sets accepted by  $M_k$  for  $k \in F$ , where  $M_k$  is identical to  $M$  except that it has unique accept state  $k$ . A desired formula  $[M]\varphi$  can be written as a conjunction

$$\bigwedge_{k \in F} [M_k]\varphi$$

with at most quadratic growth.

We now obtain a new logic APDL (*automata PDL*) by defining  $\Phi$  and  $\Pi$  inductively using the clauses for  $\Phi$  from Section 2.1 and letting  $\Pi = \Pi_0 \cup \{\varphi? \mid \varphi \in \Phi\} \cup F$ , where  $F$  is the set of automata of the form (16).

Axioms 17(iv), (v), and (vii) are replaced by:

$$(17) \quad [n, i, j, \delta]\varphi \leftrightarrow \bigwedge_{\substack{k \in \bar{n} \\ \alpha \in \delta(i, k)}} [\alpha][n, k, j, \delta]\varphi, \quad i \neq j$$

$$(18) \quad [n, i, i, \delta]\varphi \leftrightarrow \varphi \wedge \bigwedge_{\substack{k \in \bar{n} \\ \alpha \in \delta(i, k)}} [\alpha][n, k, i, \delta]\varphi.$$

The induction axiom 17(viii) becomes

$$(19) \quad \left( \bigwedge_{k \in \bar{n}} [n, i, k, \delta](\varphi_k \rightarrow \bigwedge_{\substack{m \in \bar{n} \\ \alpha \in \delta(k, m)}} [\alpha]\varphi_m) \right) \rightarrow (\varphi_i \rightarrow [n, i, j, \delta]\varphi_j).$$

These and other similar changes can be used to prove:

**THEOREM 45.** *Validity in APDL is decidable in exponential time.*

**THEOREM 46.** *The axiom system described above is complete for APDL.*

### 7.3 Converse

The *converse operator*  $-$  is a program operator that allows a program to be “run backwards”:

$$\mathbf{m}_{\bar{R}}(\alpha^-) \stackrel{\text{def}}{=} \{(s, t) \mid (t, s) \in \mathbf{m}_{\bar{R}}(\alpha)\}.$$

PDL with converse is called CPDL.

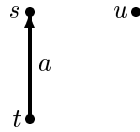
The following identities allow us to assume without loss of generality that the converse operator is applied to atomic programs only.

$$\begin{aligned} (\alpha; \beta)^- &\leftrightarrow \beta^-; \alpha^- \\ (\alpha \cup \beta)^- &\leftrightarrow \alpha^- \cup \beta^- \\ \alpha^{*-} &\leftrightarrow \alpha^{-*}. \end{aligned}$$

The converse operator strictly increases the expressive power of PDL, since the formula  $\langle \alpha^- \rangle \mathbf{1}$  is not expressible without it.

**THEOREM 47.**  $\text{PDL} < \text{CPDL}$ .

**Proof.** Consider the structure described in the following figure:



In this structure,  $s \models \langle a^- \rangle \mathbf{1}$  but  $u \not\models \langle a^- \rangle \mathbf{1}$ . On the other hand, it can be shown by induction on the structure of formulas that if  $s$  and  $u$  agree on all atomic formulas, then no formula of PDL can distinguish between the two. ■

More interestingly, the presence of the converse operator implies that the operator  $\langle \alpha \rangle$  is *continuous* in the sense that if  $A$  is any (possibly infinite) family of formulas possessing a join  $\bigvee A$ , then  $\bigvee \langle \alpha \rangle A$  exists and is logically equivalent to  $\langle \alpha \rangle \bigvee A$ . In the absence of the converse operator, one can construct nonstandard models for which this fails.

The completeness and exponential time decidability results of Sections 4 and 5 can be extended to CPDL provided the following two axioms are added:

$$\begin{aligned} \varphi &\rightarrow [\alpha] \langle \alpha^- \rangle \varphi \\ \varphi &\rightarrow [\alpha^-] \langle \alpha \rangle \varphi. \end{aligned}$$

The filtration lemma (Lemma 14) still holds in the presence of  $\bar{\phantom{x}}$ , as does the finite model property.

#### 7.4 Well-foundedness

If  $\alpha$  is a deterministic program, the formula  $\varphi \rightarrow \langle \alpha \rangle \psi$  asserts the total correctness of  $\alpha$  with respect to pre- and postconditions  $\varphi$  and  $\psi$ , respectively. For *nondeterministic* programs, however, this formula does not express the right notion of total correctness. It asserts that  $\varphi$  implies that *there exists* a halting computation sequence of  $\alpha$  yielding  $\psi$ , whereas we would really like to assert that  $\varphi$  implies that *all* computation sequences of  $\alpha$  terminate and yield  $\psi$ . Let us denote the latter property by

$$TC(\varphi, \alpha, \psi).$$

Unfortunately, this is not expressible in PDL.

The problem is intimately connected with the notion of *well-foundedness*. A program  $\alpha$  is said to be *well-founded* at a state  $u_0$  if there exists no infinite sequence of states  $u_0, u_1, u_2, \dots$  with  $(u_i, u_{i+1}) \in \mathfrak{m}_{\mathcal{R}}(\alpha)$  for all  $i \geq 0$ . This property is not expressible in PDL either, as we will see.

Several very powerful logics have been proposed to deal with this situation. The most powerful is perhaps the propositional  $\mu$ -calculus, which is essentially propositional modal logic augmented with a least fixpoint operator  $\mu$ . Using this operator, one can express any property that can be formulated as the least fixpoint of a monotone transformation on sets of states defined by the PDL operators. For example, the well-foundedness of a program  $\alpha$  is expressed



$$(20) \quad \mu X. [\alpha] X$$

in this logic.

Two somewhat weaker ways of capturing well-foundedness without resorting to the full  $\mu$ -calculus have been studied. One is to add to PDL an explicit predicate **wf** for well-foundedness:

$$\mathbf{m}_{\mathcal{R}}(\mathbf{wf} \alpha) \stackrel{\text{def}}{=} \{s_0 \mid \neg \exists s_1, s_2, \dots \forall i \geq 0 (s_i, s_{i+1}) \in \mathbf{m}_{\mathcal{R}}(\alpha)\}.$$

Another is to add an explicit predicate **halt**, which asserts that all computations of its argument  $\alpha$  terminate. The predicate **halt** can be defined inductively from **wf** as follows:

$$(21) \quad \mathbf{halt} a \stackrel{\text{def}}{\iff} \mathbf{1}, \quad a \text{ an atomic program or test,}$$

$$(22) \quad \mathbf{halt} \alpha; \beta \stackrel{\text{def}}{\iff} \mathbf{halt} \alpha \wedge [\alpha] \mathbf{halt} \beta,$$

$$(23) \quad \mathbf{halt} \alpha \cup \beta \stackrel{\text{def}}{\iff} \mathbf{halt} \alpha \wedge \mathbf{halt} \beta,$$

$$(24) \quad \mathbf{halt} \alpha^* \stackrel{\text{def}}{\iff} \mathbf{wf} \alpha \wedge [\alpha^*] \mathbf{halt} \alpha.$$

These constructs have been investigated under the various names **loop**, **repeat**, and  $\Delta$ . The predicates **loop** and **repeat** are just the complements of **halt** and **wf**, respectively:

$$\mathbf{loop} \alpha \stackrel{\text{def}}{\iff} \neg \mathbf{halt} \alpha$$

$$\mathbf{repeat} \alpha \stackrel{\text{def}}{\iff} \neg \mathbf{wf} \alpha.$$

Clause (24) is equivalent to the assertion

$$\mathbf{loop} \alpha^* \stackrel{\text{def}}{\iff} \mathbf{repeat} \alpha \vee \langle \alpha^* \rangle \mathbf{loop} \alpha.$$

It asserts that a nonhalting computation of  $\alpha^*$  consists of either an infinite sequence of halting computations of  $\alpha$  or a finite sequence of halting computations of  $\alpha$  followed by a nonhalting computation of  $\alpha$ .

Let RPDL and LPDL denote the logics obtained by augmenting PDL with the **wf** and **halt** predicates, respectively.<sup>4</sup> It follows from the preceding discussion that

$$\text{PDL} \leq \text{LPDL} \leq \text{RPDL} \leq \text{the propositional } \mu\text{-calculus}.$$

Moreover, all these inclusions are known to be strict.

The logic LPDL is powerful enough to express the total correctness of nondeterministic programs. The total correctness of  $\alpha$  with respect to precondition  $\varphi$  and postcondition  $\psi$  is expressed

$$\mathbf{TC}(\varphi, \alpha, \psi) \stackrel{\text{def}}{\iff} \varphi \rightarrow \mathbf{halt} \alpha \wedge [\alpha] \psi.$$

<sup>4</sup>The L in LPDL stands for “loop” and the R in RPDL stands for “repeat.” We retain these names for historical reasons.

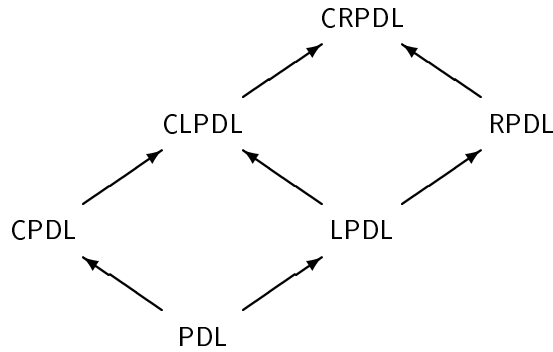
Conversely, **halt** can be expressed in terms of  $TC$ :

$$\mathbf{halt} \alpha \Leftrightarrow TC(\mathbf{1}, \alpha, \mathbf{1}).$$

THEOREM 48.  $PDL < LPDL$ .

THEOREM 49.  $LPDL < RPD$ .

It is possible to extend Theorem 49 to versions CRPDL and CLPDL in which converse is allowed in addition to **wf** or **halt**. Also, the proof of Theorem 47 goes through for LPDL and RPD, so that  $\langle a^- \rangle \mathbf{1}$  is not expressible in either. Theorem 48 goes through for the converse versions too. We obtain the situation illustrated in the following figure, in which the arrows indicate  $<$  and the absence of a path between two logics means that each can express properties that the other cannot.



The filtration lemma fails for all **halt** and **wf** versions as in Theorem 48. However, satisfiable formulas of the  $\mu$ -calculus (hence of RPD and LPDL) do have finite models. This finite model property is not shared by CLPDL or CRPDL.

THEOREM 50. *The CLPDL formula*

$$\neg \mathbf{halt} a^* \wedge [a^*] \mathbf{halt} a^{*-}$$

*is satisfiable but has no finite model.*

As it turns out, Theorem 50 does not prevent CRPDL from being decidable.

**THEOREM 51.** *The validity problems for CRPDL, CLPDL, RPD, LPDL, and the propositional  $\mu$ -calculus are all decidable in deterministic exponential time.*

Obviously, the simpler the logic, the simpler the arguments needed to show exponential time decidability. Over the years all these logics have been gradually shown to be decidable in exponential time by various authors using various techniques. Here we point to the exponential time decidability of the propositional  $\mu$ -calculus with forward and backward modalities, proved in [Vardi, 1998b], from which all these can be seen easily to follow. The proof in [Vardi, 1998b] is carried out by exhibiting an exponential time decision procedure for two-way alternating automata on infinite trees.

As mentioned above, RPD possesses the finite (but not necessarily the small and not the collapsed) model property.

**THEOREM 52.** *Every satisfiable formula of RPD, LPDL, and the propositional  $\mu$ -calculus has a finite model.*

CRPDL and CLPDL are extensions of PDL that, like  $\text{PDL} + a^\Delta b^\Delta$  (Theorems 27 and 31), are decidable despite lacking a finite model property.

Complete axiomatizations for RPD and LPDL can be obtained by embedding them into the  $\mu$ -calculus (see Section 14.4).

## 7.5 Concurrency

Another interesting extension of PDL concerns concurrent programs. One can define an intersection operator  $\cap$  such that the binary relation on states corresponding to the program  $\alpha \cap \beta$  is the intersection of the binary relations corresponding to  $\alpha$  and  $\beta$ . This can be viewed as a kind of concurrency operator that admits transitions to those states that both  $\alpha$  and  $\beta$  would have admitted.

Here we consider a different and perhaps more natural notion of concurrency. The interpretation of a program will not be a binary relation on states, which relates initial states to possible final states, but rather a relation between a states and *sets* of states. Thus  $\mathfrak{m}_{\mathfrak{R}}(\alpha)$  will relate a start state  $u$  to a collection of sets of states  $U$ . The intuition is that starting in state  $u$ , the (concurrent) program  $\alpha$  can be run with its concurrent execution threads ending in the set of final states  $U$ . The basic concurrency operator will be denoted here by  $\wedge$ , although in the original work on concurrent Dynamic Logic [Peleg, 1987b; Peleg, 1987c; Peleg, 1987a] the notation  $\cap$  is used.

The syntax of *concurrent* PDL is the same as PDL, with the addition of the clause:

- if  $\alpha, \beta \in \Pi$ , then  $\alpha \wedge \beta \in \Pi$ .

The program  $\alpha \wedge \beta$  means intuitively, “Execute  $\alpha$  and  $\beta$  in parallel.”

The semantics of concurrent PDL is defined on Kripke frames  $\mathfrak{K} = (K, m_{\mathfrak{K}})$  as with PDL, except that for programs  $\alpha$ ,

$$m_{\mathfrak{K}}(\alpha) \subseteq K \times 2^K.$$

Thus the meaning of  $\alpha$  is a collection of *reachability pairs* of the form  $(u, U)$ , where  $u \in K$  and  $U \subseteq K$ . In this brief description of concurrent PDL, we require that structures assign to atomic programs *sequential*, non-parallel, meaning; that is, for each  $a \in \Pi_0$ , we require that if  $(u, U) \in m_{\mathfrak{K}}(a)$ , then  $\#U = 1$ . The true parallelism will stem from applying the concurrency operator to build larger sets  $U$  in the reachability pairs of compound programs. For details, see [Peleg, 1987b; Peleg, 1987c].

The relevant results for this logic are the following:

**THEOREM 53.** PDL  $<$  concurrent PDL.

**THEOREM 54.** *The validity problem for concurrent PDL is decidable in deterministic exponential time.*

Axiom System 17, augmented with the following axiom, can be shown to be complete for concurrent PDL:

$$\langle \alpha \wedge \beta \rangle \varphi \leftrightarrow \langle \alpha \rangle \varphi \wedge \langle \beta \rangle \varphi.$$

## 8 FIRST-ORDER DYNAMIC LOGIC (DL)

In this section we begin the study of first-order Dynamic Logic. The main difference between first-order DL and the propositional version PDL discussed in previous sections is the presence of a first-order structure  $\mathfrak{A}$ , called the *domain of computation*, over which first-order quantification is allowed. States are no longer abstract points, but *valuations* of a set of variables over  $A$ , the *carrier* of  $\mathfrak{A}$ . Atomic programs in DL are no longer abstract binary relations, but *assignment statements* of various forms, all based on assigning values to variables during the computation. The most basic example of such an assignment is the *simple assignment*  $x := t$ , where  $x$  is a variable and  $t$  is a term. The atomic formulas of DL are generally taken to be atomic first-order formulas.

In addition to the constructs of PDL, the basic DL syntax contains individual variables ranging over  $A$ , function and predicate symbols for distinguished functions and predicates of  $\mathfrak{A}$ , and quantifiers ranging over  $A$ , exactly as in classical first-order logic. More powerful versions of the logic contain array and stack variables and other constructs, as well as primitive operations for manipulating them, and assignments for changing their values. Sometimes the introduction of a new construct increases expressive power and sometimes not; sometimes it has an effect on the complexity of

deciding satisfiability and sometimes not. Indeed, one of the central goals of research has been to classify these constructs in terms of their relative expressive power and complexity.

In this section we lay the groundwork for this by defining the various logical and programming constructs we shall need.

### 8.1 Basic Syntax

The language of first-order Dynamic Logic is built upon classical first-order logic. There is always an underlying first-order vocabulary  $\Sigma$ , which involves a vocabulary of function symbols and predicate (or relation) symbols. On top of this vocabulary, we define a set of *programs* and a set of *formulas*. These two sets interact by means of the modal construct  $[ ]$  exactly as in the propositional case. Programs and formulas are usually defined by mutual induction.

Let  $\Sigma = \{f, g, \dots, p, r, \dots\}$  be a finite first-order vocabulary. Here  $f$  and  $g$  denote typical function symbols of  $\Sigma$ , and  $p$  and  $r$  denote typical relation symbols. Associated with each function and relation symbol of  $\Sigma$  is a fixed *arity* (number of arguments), although we do not represent the arity explicitly. We assume that  $\Sigma$  always contains the equality symbol  $=$ , whose arity is 2. Functions and relations of arity 0, 1, 2, 3 and  $n$  are called *nullary*, *unary*, *binary*, *ternary*, and *n-ary*, respectively. Nullary functions are also called *constants*. We shall be using a countable set of *individual variables*  $V = \{x_0, x_1, \dots\}$ .

We always assume that  $\Sigma$  contains at least one function symbol of positive arity. A vocabulary  $\Sigma$  is *polyadic* if it contains a function symbol of arity greater than one. Vocabularies whose function symbols are all unary are called *monadic*.

A vocabulary  $\Sigma$  is *rich* if either it contains at least one predicate symbol besides the equality symbol or the sum of arities of the function symbols is at least two. Examples of rich vocabularies are: two unary function symbols, or one binary function symbol, or one unary function symbol and one unary predicate symbol. A vocabulary that is not rich is *poor*. Hence a poor vocabulary has just one unary function symbol and possibly some constants, but no relation symbols other than equality. The main difference between rich and poor vocabularies is that the former admit exponentially many pairwise non-isomorphic structures of a given finite cardinality, whereas the latter admit only polynomially many.

We say that the vocabulary  $\Sigma$  is *mono-unary* if it contains no function symbols other than a single unary one. It may contain constants and predicate symbols.

The definitions of DL programs and formulas below depend on the vocabulary  $\Sigma$ , but in general we shall not make this dependence explicit unless we have some specific reason for doing so.

### *Atomic Formulas and Programs*

In all versions of DL that we will consider, atomic formulas are atomic formulas of the first-order vocabulary  $\Sigma$ ; that is, formulas of the form  $r(t_1, \dots, t_n)$ , where  $r$  is an  $n$ -ary relation symbol of  $\Sigma$  and  $t_1, \dots, t_n$  are terms of  $\Sigma$ .

As in PDL, programs are defined inductively from atomic programs using various programming constructs. The meaning of a compound program is given inductively in terms of the meanings of its constituent parts. Different classes of programs are obtained by choosing different classes of atomic programs and programming constructs.

In the basic version of DL, an atomic program is a *simple assignment*  $x := t$ , where  $x \in V$  and  $t$  is a term of  $\Sigma$ . Intuitively, this program assigns the value of  $t$  to the variable  $x$ . This is the same form of assignment found in most conventional programming languages.

More powerful forms of assignment such as stack and array assignments and nondeterministic “wildcard” assignments will be discussed later. The precise choice of atomic programs will be made explicit when needed, but for now, we use the term *atomic program* to cover all of these possibilities.

### *Tests*

As in PDL, DL contains a test operator  $?$ , which turns a formula into a program. In most versions of DL that we shall discuss, we allow only quantifier-free first-order formulas as tests. We sometimes call these versions *poor test*. Alternatively, we might allow any first-order formula as a test. Most generally, we might place no restrictions on the form of tests, allowing any DL formula whatsoever, including those that contain other programs, perhaps containing other tests, etc. These versions of DL are labeled *rich test* as in Section 2.1. Whereas programs can be defined independently from formulas in poor test versions, rich test versions require a mutually inductive definition of programs and formulas.

As with atomic programs, the precise logic we consider at any given time depends on the choice of tests we allow. We will make this explicit when needed, but for now, we use the term *test* to cover all possibilities.

### *Regular Programs*

For a given set of atomic programs and tests, the set of *regular programs* is defined as in PDL (see Section 2.1):

- any atomic program or test is a program;
- if  $\alpha$  and  $\beta$  are programs, then  $\alpha ; \beta$  is a program;
- if  $\alpha$  and  $\beta$  are programs, then  $\alpha \cup \beta$  is a program;
- if  $\alpha$  is a program then  $\alpha^*$  is a program.

### While Programs

Much of the literature on DL is concerned with the class of **while programs** (see Section 2.1). Formally, *deterministic while programs* form the subclass of the regular programs in which the program operators  $\cup$ ,  $?$ , and  $*$  are constrained to appear only in the forms

$$\begin{aligned}
 \text{skip} &\stackrel{\text{def}}{=} 1? \\
 \text{fail} &\stackrel{\text{def}}{=} 0? \\
 (25) \text{ if } \varphi \text{ then } \alpha \text{ else } \beta &\stackrel{\text{def}}{=} (\varphi?; \alpha) \cup (\neg\varphi?; \beta) \\
 (26) \text{ while } \varphi \text{ do } \alpha &\stackrel{\text{def}}{=} (\varphi?; \alpha)^*; \neg\varphi?
 \end{aligned}$$

The class of *nondeterministic while programs* is the same, except that we allow unrestricted use of the nondeterministic choice construct  $\cup$ . Of course, unrestricted use of the sequential composition operator is allowed in both languages.

Restrictions on the form of atomic programs and tests apply as with regular programs. For example, if we are allowing only poor tests, then the  $\varphi$  occurring in the programs (25) and (26) must be a quantifier-free first-order formula.

The class of deterministic **while programs** is important because it captures the basic programming constructs common to many real-life imperative programming languages. Over the standard structure of the natural numbers  $\mathbb{N}$ , deterministic **while programs** are powerful enough to define all partial recursive functions, and thus over  $\mathbb{N}$  they are as expressive as regular programs. A similar result holds for a wide class of models similar to  $\mathbb{N}$ , for a suitable definition of “partial recursive functions” in these models. However, it is not true in general that **while programs**, even nondeterministic ones, are universally expressive. We discuss these results in Section 12.

### Formulas

A *formula* of DL is defined in way similar to that of PDL, with the addition of a rule for quantification. Equivalently, we might say that a formula of DL is defined in a way similar to that of first-order logic, with the addition of a rule for modality. The basic version of DL is defined with regular programs:

- the false formula  $\mathbf{0}$  is a formula;
- any atomic formula is a formula;
- if  $\varphi$  and  $\psi$  are formulas, then  $\varphi \rightarrow \psi$  is a formula;
- if  $\varphi$  is a formula and  $x \in V$ , then  $\forall x \varphi$  is a formula;

- if  $\varphi$  is a formula and  $\alpha$  is a program, then  $[\alpha]\varphi$  is a formula.

The only missing rule in the definition of the syntax of DL are the tests. In our basic version we would have:

- if  $\varphi$  is a quantifier-free first-order formula, then  $\varphi?$  is a test.

For the rich test version, the definitions of programs and formulas are mutually dependent, and the rule defining tests is simply:

- if  $\varphi$  is a formula, then  $\varphi?$  is a test.

We will use the same notation as in propositional logic that  $\neg\varphi$  stands for  $\varphi \rightarrow \mathbf{0}$ . As in first-order logic, the first-order existential quantifier  $\exists$  is considered a defined construct:  $\exists x \varphi$  abbreviates  $\neg\forall x \neg\varphi$ . Similarly, the modal construct  $\langle \rangle$  is considered a defined construct as in Section 2.1, since it is the modal dual of  $[\ ]$ . The other propositional constructs  $\wedge$ ,  $\vee$ ,  $\leftrightarrow$  are defined as in Section 2.1. Of course, we use parentheses where necessary to ensure unique readability.

Note that the individual variables in  $V$  serve a dual purpose: they are both program variables and logical variables.

## 8.2 Richer Programs

### *Seqs and R.E. Programs*

Some classes of programs are most conveniently defined as certain sets of seqs. Recall from Section 2.3 that a *seq* is a program of the form  $\sigma_1; \dots; \sigma_k$ , where each  $\sigma_i$  is an assignment statement or a quantifier-free first-order test. Each regular program  $\alpha$  is associated with a unique set of seqs  $CS(\alpha)$  (Section 2.3). These definitions were made in the propositional context, but they apply equally well to the first-order case; the only difference is in the form of atomic programs and tests.

Construing the word in the broadest possible sense, we can consider a *program* to be an arbitrary set of seqs. Although this makes sense semantically—we can assign an input/output relation to such a set in a meaningful way—such programs can hardly be called executable. At the very least we should require that the set of seqs be recursively enumerable, so that there will be some effective procedure that can list all possible executions of a given program. However, there is a subtle issue that arises with this notion. Consider the set of seqs

$$\{x_i := f^i(c) \mid i \in \mathbb{N}\}.$$

This set satisfies the above restriction, yet it can hardly be called a program. It uses infinitely many variables, and as a consequence it might change a



valuation at infinitely many places. Another pathological example is the set of seqs

$$\{x_{i+1} := f(x_i) \mid i \in \mathbb{N}\},$$

which not only could change a valuation at infinitely many locations, but also depends on infinitely many locations of the input valuation.

In order to avoid such pathologies, we will require that each program use only finitely many variables. This gives rise to the following definition of *r.e. programs*, which is the most general family of programs we will consider. Specifically, an r.e. program  $\alpha$  is a Turing machine that enumerates a set of seqs over a finite set of variables. The set of seqs enumerated will be called  $CS(\alpha)$ . By  $FV(\alpha)$  we will denote the finite set of variables that occur in seqs of  $CS(\alpha)$ .

An important issue connected with r.e. programs is that of *bounded memory*. The assignment statements or tests in an r.e. program may have infinitely many terms with increasingly deep nesting of function symbols (although, as discussed, these terms only use finitely many variables), and these could require an unbounded amount of memory to compute. We define a set of seqs to be *bounded memory* if the depth of terms appearing in it is bounded. In fact, without sacrificing computational power, we could require that all terms be of the form  $f(x_1, \dots, x_n)$  in a bounded-memory set of seqs.

### *Arrays and Stacks*

Interesting variants of the programming language we use in DL arise from allowing auxiliary data structures. We shall define versions with *arrays* and *stacks*, as well as a version with a nondeterministic assignment statement called *wildcard assignment*.

Besides these, one can imagine augmenting **while** programs with many other kinds of constructs such as blocks with declarations, recursive procedures with various parameter passing mechanisms, higher-order procedures, concurrent processes, etc. It is easy to arrive at a family consisting of thousands of programming languages, giving rise to thousands of logics. Obviously, we have had to restrict ourselves. It is worth mentioning, however, that certain kinds of recursive procedures are captured by our stack operations, as explained below.

### *Arrays*

To handle arrays, we include a countable set of *array variables*

$$V_{\text{array}} = \{F_0, F_1, \dots\}.$$

Each array variable has an associated *arity*, or number of arguments, which we do not represent explicitly. We assume that there are countably many

variables of each arity  $n \geq 0$ . In the presence of array variables, we equate the set  $V$  of individual variables with the set of nullary array variables; thus  $V \subseteq V_{\text{array}}$ .

The variables in  $V_{\text{array}}$  of arity  $n$  will range over  $n$ -ary functions with arguments and values in the domain of computation. In our exposition, elements of the domain of computation play two roles: they are used both as *indices* into an array and as *values* that can be stored in an array. One might equally well introduce a separate sort for array indices; although conceptually simple, this would complicate the notation and would give no new insight.

We extend the set of first-order terms to allow the unrestricted occurrence of array variables, provided arities are respected.

The classes of *regular programs with arrays* and *deterministic and non-deterministic while programs with arrays* are defined similarly to the classes without, except that we allow *array assignments* in addition to simple assignments. Array assignments are similar to simple assignments, but on the left-hand side we allow a term in which the outermost symbol is an array variable:

$$F(t_1, \dots, t_n) := t.$$

Here  $F$  is an  $n$ -ary array variable and  $t_1, \dots, t_n, t$  are terms, possibly involving other array variables. Note that when  $n = 0$ , this reduces to the ordinary simple assignment.

#### *Recursion via an Algebraic Stack*

We now consider DL in which the programs can manipulate a stack. The literature in automata theory and formal languages often distinguishes a stack from a pushdown store. In the former, the automaton is allowed to inspect the contents of the stack but to make changes only at the top. We shall use the term stack to denote the more common pushdown store, where the only inspection allowed is at the top of the stack.

The motivation for this extension is to be able to capture recursion. It is well known that recursive procedures can be modeled using a stack, and for various technical reasons we prefer to extend the data-manipulation capabilities of our programs than to introduce new control constructs. When it encounters a recursive call, the stack simulation of recursion will push the return location and values of local variables and parameters on the stack. It will pop them upon completion of the call. The LIFO (last-in-first-out) nature of stack storage fits the order in which control executes recursive calls.

To handle the stack in our stack version of DL, we add two new atomic programs

$$\mathbf{push}(t) \quad \text{and} \quad \mathbf{pop}(y),$$

where  $t$  is a term and  $y \in V$ . Intuitively, **push**( $t$ ) pushes the current value of  $t$  onto the top of the stack, and **pop**( $y$ ) pops the top value off the top of the stack and assigns that value to the variable  $y$ . If the stack is empty, the pop operation does not change anything. We could have added a test for stack emptiness, but it can be shown to be redundant. Formally, the stack is simply a finite string of elements of the domain of computation.

The classes of *regular programs with stack* and *deterministic and non-deterministic while programs with stack* are obtained by augmenting the respective classes of programs with the **push** and **pop** operations as atomic programs in addition to simple assignments.

In contrast to the case of arrays, here there is only a single stack. In fact, expressiveness changes dramatically when two or more stacks are allowed. Also, in order to be able to simulate recursion, the domain must have at least two distinct elements so that return addresses can be properly encoded in the stack. One way of doing this is to store the return address itself in unary using one element of the domain, then store one occurrence of the second element as a delimiter symbol, followed by domain elements constituting the current values of parameters and local variables.

The kind of stack described here is often termed *algebraic*, since it contains elements from the domain of computation. It should be contrasted with the Boolean stack described next.

#### *Parameterless Recursion via a Boolean Stack*

An interesting special case is when the stack can contain only two distinct elements. This version of our programming language can be shown to capture recursive procedures without parameters or local variables. This is because we only need to store return addresses, but no actual data items from the domain of computation. This can be achieved using two values, as described above. We thus arrive at the idea of a Boolean stack.

To handle such a stack in this version of DL, we add three new kinds of atomic programs and one new test. The atomic programs are

**push-1**      **push-0**      **pop**,

and the test is simply **top?**. Intuitively, **push-1** and **push-0** push the corresponding distinct Boolean values on the stack, **pop** removes the top element, and the test **top?** evaluates to true iff the top element of the stack is **1**, but with no side effect.

With the test **top?** only, there is no explicit operator that distinguishes a stack with top element **0** from the empty stack. We might have defined such an operator, and in a more realistic language we would certainly do so. However, it is mathematically redundant, since it can be simulated with the operators we already have.

### *Wildcard Assignment*

The nondeterministic assignment  $x := ?$  is a device that arises in the study of fairness; see [Apt and Plotkin, 1986]. It has often been called *random assignment* in the literature, although it has nothing to do with randomness or probability. We shall call it *wildcard assignment*. Intuitively, it operates by assigning a nondeterministically chosen element of the domain of computation to the variable  $x$ . This construct together with the  $[ ]$  modality is similar to the first-order universal quantifier, since it will follow from the semantics that the two formulas  $[x := ?]\varphi$  and  $\forall x \varphi$  are equivalent. However, wildcard assignment may appear in programs and can therefore be iterated.

### 8.3 *Semantics*

In this section we assign meanings to the syntactic constructs described in the previous sections. We interpret programs and formulas over a first-order structure  $\mathfrak{A}$ . Variables range over the carrier of this structure. We take an *operational* view of program semantics: programs change the values of variables by sequences of simple assignments  $x := t$  or other assignments, and flow of control is determined by the truth values of tests performed at various times during the computation.

#### *States as Valuations*

An instantaneous snapshot of all relevant information at any moment during the computation is determined by the values of the program variables. Thus our *states* will be *valuations*  $u, v, \dots$  of the variables  $V$  over the carrier of the structure  $\mathfrak{A}$ . Our formal definition will associate the pair  $(u, v)$  of such valuations with the program  $\alpha$  if it is possible to start in valuation  $u$ , execute the program  $\alpha$ , and halt in valuation  $v$ . In this case, we will call  $(u, v)$  an *input/output pair* of  $\alpha$  and write  $(u, v) \in \mathfrak{m}_{\mathfrak{A}}(\alpha)$ . This will result in a Kripke frame exactly as in Section 2.

Let  $\mathfrak{A} = (A, \mathfrak{m}_{\mathfrak{A}})$  be a first-order structure for the vocabulary  $\Sigma$ . We call  $\mathfrak{A}$  the *domain of computation*. Here  $A$  is a set, called the *carrier* of  $\mathfrak{A}$ , and  $\mathfrak{m}_{\mathfrak{A}}$  is a *meaning function* such that  $\mathfrak{m}_{\mathfrak{A}}(f)$  is an  $n$ -ary function  $\mathfrak{m}_{\mathfrak{A}}(f) : A^n \rightarrow A$  interpreting the  $n$ -ary function symbol  $f$  of  $\Sigma$ , and  $\mathfrak{m}_{\mathfrak{A}}(r)$  is an  $n$ -ary relation  $\mathfrak{m}_{\mathfrak{A}}(r) \subseteq A^n$  interpreting the  $n$ -ary relation symbol  $r$  of  $\Sigma$ . The equality symbol  $=$  is always interpreted as the identity relation.

For  $n \geq 0$ , let  $A^n \rightarrow A$  denote the set of all  $n$ -ary functions in  $A$ . By convention, we take  $A^0 \rightarrow A = A$ . Let  $A^*$  denote the set of all finite-length strings over  $A$ .

The structure  $\mathfrak{A}$  determines a Kripke frame, which we will also denote by  $\mathfrak{A}$ , as follows. A *valuation* over  $\mathfrak{A}$  is a function  $u$  assigning an  $n$ -ary function over  $A$  to each  $n$ -ary array variable. It also assigns meanings to

the stacks as follows. We shall use the two unique variable names  $STK$  and  $BSTK$  to denote the algebraic stack and the Boolean stack, respectively. The valuation  $u$  assigns a finite-length string of elements of  $A$  to  $STK$  and a finite-length string of Boolean values  $\mathbf{1}$  and  $\mathbf{0}$  to  $BSTK$ . Formally:

$$\begin{aligned} u(F) &\in A^n \rightarrow A, \quad \text{if } F \text{ is an } n\text{-ary array variable,} \\ u(STK) &\in A^*, \\ u(BSTK) &\in \{\mathbf{1}, \mathbf{0}\}^*. \end{aligned}$$

By our convention  $A^0 \rightarrow A = A$ , and assuming that  $V \subseteq V_{\text{array}}$ , the individual variables (that is, the nullary array variables) are assigned elements of  $A$  under this definition:

$$u(x) \in A \text{ if } x \in V.$$

The valuation  $u$  extends uniquely to terms  $t$  by induction. For an  $n$ -ary function symbol  $f$  and an  $n$ -ary array variable  $F$ ,

$$\begin{aligned} u(f(t_1, \dots, t_n)) &\stackrel{\text{def}}{=} \mathfrak{m}_{\mathfrak{A}}(f)(u(t_1), \dots, u(t_n)) \\ u(F(t_1, \dots, t_n)) &\stackrel{\text{def}}{=} u(F)(u(t_1), \dots, u(t_n)). \end{aligned}$$

The *function-patching* operator is defined as follows: if  $X$  and  $D$  are sets,  $f : X \rightarrow D$  is any function,  $x \in X$ , and  $d \in D$ , then  $f[x/d] : X \rightarrow D$  is the function defined by

$$f[x/d](y) \stackrel{\text{def}}{=} \begin{cases} d, & \text{if } x = y \\ f(y), & \text{otherwise.} \end{cases}$$

We will be using this notation in several ways, both at the logical and metalogical levels. For example:

- If  $u$  is a valuation,  $x$  is an individual variable, and  $a \in A$ , then  $u[x/a]$  is the new valuation obtained from  $u$  by changing the value of  $x$  to  $a$  and leaving the values of all other variables intact.
- If  $F$  is an  $n$ -ary array variable and  $f : A^n \rightarrow A$ , then  $u[F/f]$  is the new valuation that assigns the same value as  $u$  to the stack variables and to all array variables other than  $F$ , and

$$u[F/f](F) = f.$$

- If  $f : A^n \rightarrow A$  is an  $n$ -ary function and  $\bar{a} = a_1, \dots, a_n \in A^n$  and  $a \in A$ , then the expression  $f[\bar{a}/a]$  denotes the  $n$ -ary function that agrees with  $f$  everywhere except for input  $\bar{a}$ , on which it takes the value  $a$ . More precisely,

$$f[\bar{a}/a](\bar{b}) = \begin{cases} a, & \text{if } \bar{b} = \bar{a} \\ f(\bar{b}), & \text{otherwise.} \end{cases}$$

We call valuations  $u$  and  $v$  *finite variants* of each other if

$$u(F)(a_1, \dots, a_n) = v(F)(a_1, \dots, a_n)$$

for all but finitely many array variables  $F$  and  $n$ -tuples  $a_1, \dots, a_n \in A^n$ . In other words,  $u$  and  $v$  differ on at most finitely many array variables, and for those  $F$  on which they do differ, the functions  $u(F)$  and  $v(F)$  differ on at most finitely many values.

The relation “is a finite variant of” is an equivalence relation on valuations. Since a halting computation can run for only a finite amount of time, it can execute only finitely many assignments. It will therefore not be able to cross equivalence class boundaries; that is, in the binary relation semantics given below, if the pair  $(u, v)$  is an input/output pair of the program  $\alpha$ , then  $v$  is a finite variant of  $u$ .

We are now ready to define the *states* of our Kripke frame. For  $a \in A$ , let  $w_a$  be the valuation in which the stacks are empty and all array and individual variables are interpreted as constant functions taking the value  $a$  everywhere. A *state* of  $\mathfrak{A}$  is any finite variant of a valuation  $w_a$ . The set of states of  $\mathfrak{A}$  is denoted  $S^{\mathfrak{A}}$ .

Call a state *initial* if it differs from some  $w_a$  only at the values of individual variables.

It is meaningful, and indeed useful in some contexts, to take as states the set of *all* valuations. Our purpose in restricting our attention to states as defined above is to prevent arrays from being initialized with highly complex oracles that would compromise the value of the relative expressiveness results of Section 12.

### Assignment Statements

As in Section 2.2, with every program  $\alpha$  we associate a binary relation

$$\mathfrak{m}_{\mathfrak{A}}(\alpha) \subseteq S^{\mathfrak{A}} \times S^{\mathfrak{A}}$$

(called the *input/output relation* of  $\alpha$ ), and with every formula  $\varphi$  we associate a set

$$\mathfrak{m}_{\mathfrak{A}}(\varphi) \subseteq S^{\mathfrak{A}}.$$

The sets  $\mathfrak{m}_{\mathfrak{A}}(\alpha)$  and  $\mathfrak{m}_{\mathfrak{A}}(\varphi)$  are defined by mutual induction on the structure of  $\alpha$  and  $\varphi$ .

For the basis of this inductive definition, we first give the semantics of all the assignment statements discussed earlier.

- The array assignment  $F(t_1, \dots, t_n) := t$  is interpreted as the binary relation

$$\begin{aligned} \mathfrak{m}_{\mathfrak{A}}(F(t_1, \dots, t_n) := t) \\ \stackrel{\text{def}}{=} \{ (u, u[F/u(F)][u(t_1), \dots, u(t_n)/u(t)]) \mid u \in S^{\mathfrak{A}} \}. \end{aligned}$$

In other words, starting in state  $u$ , the array assignment has the effect of changing the value of  $F$  on input  $u(t_1), \dots, u(t_n)$  to  $u(t)$ , and leaving the value of  $F$  on all other inputs and the values of all other variables intact. For  $n = 0$ , this definition reduces to the following definition of simple assignment:

$$\mathfrak{m}_{\mathfrak{A}}(x := t) \stackrel{\text{def}}{=} \{(u, u[x/u(t)]) \mid u \in S^{\mathfrak{A}}\}.$$

- The push operations, **push**( $t$ ) for the algebraic stack and **push-1** and **push-0** for the Boolean stack, are interpreted as the binary relations

$$\begin{aligned} \mathfrak{m}_{\mathfrak{A}}(\mathbf{push}(t)) &\stackrel{\text{def}}{=} \{(u, u[STK/(u(t) \cdot u(STK))]) \mid u \in S^{\mathfrak{A}}\} \\ \mathfrak{m}_{\mathfrak{A}}(\mathbf{push-1}) &\stackrel{\text{def}}{=} \{(u, u[BSTK/(1 \cdot u(BSTK))]) \mid u \in S^{\mathfrak{A}}\} \\ \mathfrak{m}_{\mathfrak{A}}(\mathbf{push-0}) &\stackrel{\text{def}}{=} \{(u, u[BSTK/(0 \cdot u(BSTK))]) \mid u \in S^{\mathfrak{A}}\}, \end{aligned}$$

respectively. In other words, **push**( $t$ ) changes the value of the algebraic stack variable  $STK$  from  $u(STK)$  to the string  $u(t) \cdot u(STK)$ , the concatenation of the value  $u(t)$  with the string  $u(STK)$ , and everything else is left intact. The effects of **push-1** and **push-0** are similar, except that the special constants **1** and **0** are concatenated with  $u(BSTK)$  instead of  $u(t)$ .

- The pop operations, **pop**( $y$ ) for the algebraic stack and **pop** for the Boolean stack, are interpreted as the binary relations

$$\begin{aligned} \mathfrak{m}_{\mathfrak{A}}(\mathbf{pop}(y)) &\stackrel{\text{def}}{=} \{(u, u[STK/\mathbf{tail}(u(STK))][y/\mathbf{head}(u(STK), \\ &\quad u(y))]) \mid u \in S^{\mathfrak{A}}\} \\ \mathfrak{m}_{\mathfrak{A}}(\mathbf{pop}) &\stackrel{\text{def}}{=} \{(u, u[BSTK/\mathbf{tail}(u(BSTK))]) \mid u \in S^{\mathfrak{A}}\}, \end{aligned}$$

respectively, where

$$\begin{aligned} \mathbf{tail}(a \cdot \sigma) &\stackrel{\text{def}}{=} \sigma \\ \mathbf{tail}(\varepsilon) &\stackrel{\text{def}}{=} \varepsilon \\ \mathbf{head}(a \cdot \sigma, b) &\stackrel{\text{def}}{=} a \\ \mathbf{head}(\varepsilon, b) &\stackrel{\text{def}}{=} b \end{aligned}$$

and  $\varepsilon$  is the empty string. In other words, if  $u(STK) \neq \varepsilon$ , this operation changes the value of  $STK$  from  $u(STK)$  to the string obtained by

deleting the first element of  $u(STK)$  and assigns that element to the variable  $y$ . If  $u(STK) = \varepsilon$ , then nothing is changed. Everything else is left intact. The Boolean stack operation **pop** changes the value of  $BSTK$  only, with no additional changes. We do not include explicit constructs to test whether the stacks are empty, since these can be simulated. However, we do need to be able to refer to the value of the top element of the Boolean stack, hence we include the **top?** test.

- The Boolean test program **top?** is interpreted as the binary relation

$$\mathfrak{m}_{\mathfrak{A}}(\mathbf{top?}) \stackrel{\text{def}}{=} \{(u, u) \mid u \in S^{\mathfrak{A}}, \mathbf{head}(u(BSTK)) = \mathbf{1}\}.$$

In other words, this test changes nothing at all, but allows control to proceed iff the top of the Boolean stack contains **1**.

- The wildcard assignment  $x := ?$  for  $x \in V$  is interpreted as the relation

$$\mathfrak{m}_{\mathfrak{A}}(x := ?) \stackrel{\text{def}}{=} \{(u, u[x/a]) \mid u \in S^{\mathfrak{A}}, a \in A\}.$$

As a result of executing this statement,  $x$  will be assigned some arbitrary value of the carrier set  $A$ , and the values of all other variables will remain unchanged.

### *Programs and Formulas*

The meanings of compound programs and formulas are defined by mutual induction on the structure of  $\alpha$  and  $\varphi$  exactly as in the propositional case (see Section 2.2).

### *Seqs and R.E. Programs*

Recall that an r.e. program is a Turing machine enumerating a set  $CS(\alpha)$  of seqs. If  $\alpha$  is an r.e. program, we define

$$\mathfrak{m}_{\mathfrak{A}}(\alpha) \stackrel{\text{def}}{=} \bigcup_{\sigma \in CS(\alpha)} \mathfrak{m}_{\mathfrak{A}}(\sigma).$$

Thus, the meaning of  $\alpha$  is defined to be the union of the meanings of the seqs in  $CS(\alpha)$ . The meaning  $\mathfrak{m}_{\mathfrak{A}}(\sigma)$  of a seq  $\sigma$  is determined by the meanings of atomic programs and tests and the sequential composition operator.

There is an interesting point here regarding the translation of programs using other programming constructs into r.e. programs. This can be done for arrays and stacks (for Booleans stacks, even into r.e. programs with bounded memory), but not for wildcard assignment. Since later in the book we shall be referring to the r.e. set of seqs associated with such programs, it



is important to be able to carry out this translation. To see how this is done for the case of arrays, for example, consider an algorithm for simulating the execution of a program by generating only ordinary assignments and tests. It does not generate an array assignment of the form  $F(t_1, \dots, t_n) := t$ , but rather “remembers” it and when it reaches an assignment of the form  $x := F(t_1, \dots, t_n)$  it will aim at generating  $x := t$  instead. This requires care, since we must keep track of changes in the variables inside  $t$  and  $t_1, \dots, t_n$  and incorporate them into the generated assignments.

### Formulas

Here are the semantic definitions for the constructs of formulas of DL. The semantics of atomic first-order formulas is the standard semantics of classical first-order logic.

$$(27) \quad \mathfrak{m}_{\mathfrak{A}}(\mathbf{0}) \stackrel{\text{def}}{=} \emptyset$$

$$(28) \quad \mathfrak{m}_{\mathfrak{A}}(\varphi \rightarrow \psi) \stackrel{\text{def}}{=} \{u \mid \text{if } u \in \mathfrak{m}_{\mathfrak{A}}(\varphi) \text{ then } u \in \mathfrak{m}_{\mathfrak{A}}(\psi)\}$$

$$(29) \quad \mathfrak{m}_{\mathfrak{A}}(\forall x \varphi) \stackrel{\text{def}}{=} \{u \mid \forall a \in A \ u[x/a] \in \mathfrak{m}_{\mathfrak{A}}(\varphi)\}$$

$$(30) \quad \mathfrak{m}_{\mathfrak{A}}([\alpha]\varphi) \stackrel{\text{def}}{=} \{u \mid \forall v \text{ if } (u, v) \in \mathfrak{m}_{\mathfrak{A}}(\alpha) \text{ then } v \in \mathfrak{m}_{\mathfrak{A}}(\varphi)\}.$$

Equivalently, we could define the first-order quantifiers  $\forall$  and  $\exists$  in terms of the wildcard assignment:

$$(31) \quad \forall x \varphi \leftrightarrow [x := ?]\varphi$$

$$(32) \quad \exists x \varphi \leftrightarrow \langle x := ? \rangle \varphi.$$

Note that for *deterministic* programs  $\alpha$  (for example, those obtained by using the **while** programming language instead of regular programs and disallowing wildcard assignments),  $\mathfrak{m}_{\mathfrak{A}}(\alpha)$  is a partial function from states to states; that is, for every state  $u$ , there is at most one  $v$  such that  $(u, v) \in \mathfrak{m}_{\mathfrak{A}}(\alpha)$ . The partiality of the function arises from the possibility that  $\alpha$  may not halt when started in certain states. For example,  $\mathfrak{m}_{\mathfrak{A}}(\mathbf{while\ 1\ do\ skip})$  is the empty relation. In general, the relation  $\mathfrak{m}_{\mathfrak{A}}(\alpha)$  need not be single-valued.

If  $K$  is a given set of syntactic constructs, we refer to the version of Dynamic Logic with programs built from these constructs as *Dynamic Logic with  $K$*  or simply as  $\text{DL}(K)$ . Thus, we have  $\text{DL}(\text{r.e.})$ ,  $\text{DL}(\text{array})$ ,  $\text{DL}(\text{stk})$ ,  $\text{DL}(\text{bstk})$ ,  $\text{DL}(\text{wild})$ , and so on. As a default, these logics are the poor-test versions, in which only quantifier-free first-order formulas may appear as tests. The unadorned DL is used to abbreviate  $\text{DL}(\text{reg})$ , and we use  $\text{DL}(\text{dreg})$  to denote DL with **while** programs, which are really deterministic regular programs. Again, **while** programs use only poor tests. Combinations such as  $\text{DL}(\text{dreg}+\text{wild})$  are also allowed.

### 8.4 Satisfiability and Validity

The concepts of satisfiability, validity, etc. are defined as for PDL in Section 2 or as for first-order logic under the standard semantics.

Let  $\mathfrak{A} = (A, \mathfrak{m}_{\mathfrak{A}})$  be a structure, and let  $u$  be a state in  $S^{\mathfrak{A}}$ . For a formula  $\varphi$ , we write  $\mathfrak{A}, u \models \varphi$  if  $u \in \mathfrak{m}_{\mathfrak{A}}(\varphi)$  and say that  $u$  *satisfies*  $\varphi$  in  $\mathfrak{A}$ . We sometimes write  $u \models \varphi$  when  $\mathfrak{A}$  is understood. We say that  $\varphi$  is  $\mathfrak{A}$ -*valid* and write  $\mathfrak{A} \models \varphi$  if  $\mathfrak{A}, u \models \varphi$  for all  $u$  in  $\mathfrak{A}$ . We say that  $\varphi$  is *valid* and write  $\models \varphi$  if  $\mathfrak{A} \models \varphi$  for all  $\mathfrak{A}$ . We say that  $\varphi$  is *satisfiable* if  $\mathfrak{A}, u \models \varphi$  for some  $\mathfrak{A}, u$ .

For a set of formulas  $\Delta$ , we write  $\mathfrak{A} \models \Delta$  if  $\mathfrak{A} \models \varphi$  for all  $\varphi \in \Delta$ .

Informally,  $\mathfrak{A}, u \models [\alpha]\varphi$  iff every terminating computation of  $\alpha$  starting in state  $u$  terminates in a state satisfying  $\varphi$ , and  $\mathfrak{A}, u \models \langle \alpha \rangle \varphi$  iff there exists a computation of  $\alpha$  starting in state  $u$  and terminating in a state satisfying  $\varphi$ . For a pure first-order formula  $\varphi$ , the metastatement  $\mathfrak{A}, u \models \varphi$  has the same meaning as in first-order logic.

## 9 RELATIONSHIPS WITH STATIC LOGICS

### 9.1 Uninterpreted Reasoning

In contrast to the propositional version PDL discussed in Sections 2–7, DL formulas involve variables, functions, predicates, and quantifiers, a state is a mapping from variables to values in some domain, and atomic programs are assignment statements. To give semantic meaning to these constructs requires a first-order structure  $\mathfrak{A}$  over which to interpret the function and predicate symbols. Nevertheless, we are not obliged to assume anything special about  $\mathfrak{A}$  or the nature of the interpretations of the function and predicate symbols, except as dictated by first-order semantics. Any conclusions we draw from this level of reasoning will be valid under all possible interpretations. *Uninterpreted reasoning* refers to this style of reasoning.

For example, the formula

$$p(f(x), g(y, f(x))) \rightarrow \langle z := f(x) \rangle p(z, g(y, z))$$

is true over any domain, irrespective of the interpretations of  $p$ ,  $f$ , and  $g$ .

Another example of a valid formula is

$$\begin{aligned} z = y \wedge \forall x f(g(x)) = x \\ \rightarrow [\mathbf{while} \ p(y) \ \mathbf{do} \ y := g(y)] \langle \mathbf{while} \ y \neq z \ \mathbf{do} \ y := f(y) \rangle \mathbf{1}. \end{aligned}$$

Note the use of  $[ ]$  applied to  $\langle \rangle$ . This formula asserts that under the assumption that  $f$  “undoes”  $g$ , any computation consisting of applying  $g$  some number of times to  $z$  can be backtracked to the original  $z$  by applying  $f$  some number of times to the result.

We now observe that three basic properties of classical (uninterpreted) first-order logic, the *Löwenheim–Skolem theorem*, *completeness*, and *compactness*, fail for even fairly weak versions of DL.

The Löwenheim–Skolem theorem for classical first-order logic states that if a formula  $\varphi$  has an infinite model then it has models of all infinite cardinalities. Because of this theorem, classical first-order logic cannot define the structure of elementary arithmetic

$$\mathbb{N} = (\omega, +, \cdot, 0, 1, =)$$

up to isomorphism. That is, there is no first-order sentence that is true in a structure  $\mathfrak{A}$  if and only if  $\mathfrak{A}$  is isomorphic to  $\mathbb{N}$ . However, this can be done in DL.

**PROPOSITION 55.** *There exists a formula  $\Theta_{\mathbb{N}}$  of DL(dreg) that defines  $\mathbb{N}$  up to isomorphism.*

The Löwenheim–Skolem theorem does not hold for DL, because  $\Theta_{\mathbb{N}}$  has an infinite model (namely  $\mathbb{N}$ ), but all models are isomorphic to  $\mathbb{N}$  and are therefore countable.

Besides the Löwenheim–Skolem Theorem, compactness fails in DL as well. Consider the following countable set  $\Gamma$  of formulas:

$$\{\langle \text{while } p(x) \text{ do } x := f(x) \rangle \mathbf{1}\} \cup \{p(f^n(x)) \mid n \geq 0\}.$$

It is easy to see that  $\Gamma$  is not satisfiable, but it is finitely satisfiable, i.e. each finite subset of it is satisfiable.

Worst of all, completeness cannot hold for any deductive system as we normally think of it (a finite effective system of axioms schemes and finitary inference rules). The set of theorems of such a system would be r.e., since they could be enumerated by writing down the axioms and systematically applying the rules of inference in all possible ways. However, the set of valid statements of DL is not recursively enumerable. In fact, we will describe in Section 10 exactly how bad the situation is.

This is not to say that we cannot say anything meaningful about proofs and deduction in DL. On the contrary, there is a wealth of interesting and practical results on axiom systems for DL that we will cover in Section 11.

In this section we investigate the power of DL relative to classical static logics on the uninterpreted level. In particular, *rich test DL of r.e. programs* is equivalent to the infinitary language  $L_{\omega_1^{\text{ck}}, \omega}$ . Some consequences of this fact are drawn in later sections.

First we introduce a definition that allows to compare different variants of DL. Let us recall from Section 8.3 that a state is *initial* if it differs from a constant state  $w_a$  only at the values of individual variables. If  $\text{DL}_1$  and  $\text{DL}_2$  are two variants of DL over the same vocabulary, we say that  $\text{DL}_2$  is *as expressive as*  $\text{DL}_1$  and write  $\text{DL}_1 \leq \text{DL}_2$  if for each formula  $\varphi$  in  $\text{DL}_1$  there is a formula  $\psi$  in  $\text{DL}_2$  such that  $\mathfrak{A}, u \models \varphi \leftrightarrow \psi$  for all structures  $\mathfrak{A}$  and initial

states  $u$ . If  $DL_2$  is as expressive as  $DL_1$  but  $DL_1$  is not as expressive as  $DL_2$ , we say that  $DL_2$  is *strictly more expressive than*  $DL_1$ , and write  $DL_1 < DL_2$ . If  $DL_2$  is as expressive as  $DL_1$  and  $DL_1$  is as expressive as  $DL_2$ , we say that  $DL_1$  and  $DL_2$  are of *equal expressive power*, or are simply *equivalent*, and write  $DL_1 \equiv DL_2$ . We will also use these notions for comparing versions of DL with static logics such as  $L_{\omega\omega}$ .

There is a technical reason for the restriction to initial states in the above definition. If  $DL_1$  and  $DL_2$  have access to different sets of data types, then they may be trivially incomparable for uninteresting reasons, unless we are careful to limit the states on which they are compared. We shall see examples of this in Section 12.

Also, in the definition of  $DL(K)$  given in Section 8.4, the programming language  $K$  is an explicit parameter. Actually, the particular first-order vocabulary  $\Sigma$  over which  $DL(K)$  and  $K$  are considered should be treated as a parameter too. It turns out that the relative expressiveness of versions of DL is sensitive not only to  $K$ , but also to  $\Sigma$ . This second parameter is often ignored in the literature, creating a source of potential misinterpretation of the results. For now, we assume a fixed first-order vocabulary  $\Sigma$ .

### *Rich Test Dynamic Logic of R.E. Programs*

We are about to introduce the most general version of DL we will ever consider. This logic is called *rich test Dynamic Logic of r.e. programs*, and it will be denoted  $DL(\text{rich-test r.e.})$ . Programs of  $DL(\text{rich-test r.e.})$  are r.e. sets of seqs as defined in Section 8.2, except that the seqs may contain tests  $\varphi?$  for any previously constructed formula  $\varphi$ .

The formal definition is inductive. All atomic programs are programs and all atomic formulas are formulas. If  $\varphi, \psi$  are formulas,  $\alpha, \beta$  are programs,  $\{\alpha_n \mid n \in \omega\}$  is an r.e. set of programs over a finite set of variables (free or bound), and  $x$  is a variable, then

- $\mathbf{0}$
- $\varphi \rightarrow \psi$
- $[\alpha]\varphi$
- $\forall x \varphi$

are formulas and

- $\alpha ; \beta$
- $\{\alpha_n \mid n \in \omega\}$
- $\varphi?$

are programs. The set  $CS(\alpha)$  of computation sequences of a rich test r.e. program  $\alpha$  is defined as usual.

The language  $L_{\omega_1\omega}$  is the language with the formation rules of the first-order language  $L_{\omega\omega}$ , but in which countably infinite conjunctions and disjunctions  $\bigwedge_{i \in I} \varphi_i$  and  $\bigvee_{i \in I} \varphi_i$  are also allowed. In addition, if  $\{\varphi_i \mid i \in I\}$  is recursively enumerable, then the resulting language is denoted  $L_{\omega_1^{\text{ck}}\omega}$  and is sometimes called *constructive*  $L_{\omega_1\omega}$ .

PROPOSITION 56.  $\text{DL}(\text{rich-test r.e.}) \equiv L_{\omega_1^{\text{ck}}\omega}$ .

Since r.e. programs as defined in Section 8.2 are clearly a special case of general rich-test r.e. programs, it follows that  $\text{DL}(\text{rich-test r.e.})$  is as expressive as  $\text{DL}(\text{r.e.})$ . In fact they are not of the same expressive power.

THEOREM 57.  $\text{DL}(\text{r.e.}) < \text{DL}(\text{rich-test r.e.})$ .

Henceforth, we shall assume that the first-order vocabulary  $\Sigma$  contains at least one function symbol of positive arity. Under this assumption, DL can easily be shown to be strictly more expressive than  $L_{\omega\omega}$ :

THEOREM 58.  $L_{\omega\omega} < \text{DL}$ .

COROLLARY 59.

$$L_{\omega\omega} < \text{DL} < \text{DL}(\text{r.e.}) < \text{DL}(\text{rich-test r.e.}) \equiv L_{\omega_1^{\text{ck}}\omega}.$$

The situation with the intermediate versions of DL, e.g.  $\text{DL}(\text{stk})$ ,  $\text{DL}(\text{bstk})$ ,  $\text{DL}(\text{wild})$ , etc., is of interest. We deal with the relative expressive power of these in Section 12.

## 9.2 Interpreted Reasoning

### *Arithmetical Structures*

This is the most detailed level we will consider. It is the closest to the actual process of reasoning about concrete, fully specified programs. Syntactically, the programs and formulas are as on the uninterpreted level, but here we assume a fixed structure or class of structures.

In this framework, we can study programs whose computational behavior depends on (sometimes deep) properties of the particular structures over which they are interpreted. In fact, almost any task of verifying the correctness of an actual program falls under the heading of interpreted reasoning.

One specific structure we will look at carefully is the natural numbers with the usual arithmetic operations:

$$\mathbb{N} = (\omega, 0, 1, +, \cdot, =).$$

Let  $-$  denote the (first-order-definable) operation of subtraction and  $\text{gcd}(x, y)$  denote the first-order-definable operation giving the greatest common divisor of  $x$  and  $y$ . The following formula of DL is  $\mathbb{N}$ -valid, i.e., true in

all states of  $\mathbb{N}$ :

$$(33) \quad x = x' \wedge y = y' \wedge xy \geq 1 \rightarrow \langle \alpha \rangle (x = \text{gcd}(x', y'))$$

where  $\alpha$  is the **while** program of Example 1 or the regular program

$$(x \neq y?; ((x > y?; x := x - y) \cup (x < y?; y := y - x)))^* x = y?.$$

Formula (33) states the correctness and termination of an actual program over  $\mathbb{N}$  computing the greatest common divisor.

As another example, consider the following formula over  $\mathbb{N}$ :

$$\forall x \geq 1 \langle (\text{if even}(x) \text{ then } x := x/2 \text{ else } x := 3x + 1)^* \rangle (x = 1).$$

Here  $/$  denotes integer division, and **even**( $\cdot$ ) is the relation that tests if its argument is even. Both of these are first-order definable. This innocent-looking formula asserts that starting with an arbitrary positive integer and repeating the following two operations, we will eventually reach 1:

- if the number is even, divide it by 2;
- if the number is odd, triple it and add 1.

The truth of this formula is as yet unknown, and it constitutes a problem in number theory (dubbed “the  $3x + 1$  problem”) that has been open for over 60 years. The formula  $\forall x \geq 1 \langle \alpha \rangle \mathbf{1}$ , where  $\alpha$  is

$$\text{while } x \neq 1 \text{ do if even}(x) \text{ then } x := x/2 \text{ else } x := 3x + 1,$$

says this in a slightly different way.

The specific structure  $\mathbb{N}$  can be generalized, resulting in the class of *arithmetical structures*. Briefly, a structure  $\mathfrak{A}$  is *arithmetical* if it contains a first-order-definable copy of  $\mathbb{N}$  and has first-order definable functions for coding finite sequences of elements of  $\mathfrak{A}$  into single elements and for the corresponding decoding.

Arithmetical structures are important because (i) most structures arising naturally in computer science (e.g., discrete structures with recursively defined data types) are arithmetical, and (ii) any structure can be extended to an arithmetical one by adding appropriate encoding and decoding capabilities. While most of the results we present for the interpreted level are given in terms of  $\mathbb{N}$  alone, many of them hold for any arithmetical structure, so their significance is greater.

### *Expressive Power over $\mathbb{N}$*

The results of Corollary 59 establishing that

$$L_{\omega\omega} < \text{DL} < \text{DL(r.e.)} < \text{DL(rich-test r.e.)}$$

were on the uninterpreted level, where all structures are taken into account. Thus first-order logic, regular DL, and DL(rich-test r.e.) form a sequence of increasingly more powerful logics when interpreted uniformly over all structures.

What happens if one fixes a structure, say  $\mathbb{N}$ ? Do these differences in expressive power still hold? We now address these questions.

First, we introduce notation for comparing expressive power over  $\mathbb{N}$ . If  $DL_1$  and  $DL_2$  are variants of DL (or static logics, such as  $L_{\omega\omega}$ ) and are defined over the vocabulary of  $\mathbb{N}$ , we write  $DL_1 \leq_{\mathbb{N}} DL_2$  if for each  $\varphi \in DL_1$  there is  $\psi \in DL_2$  such that  $\mathbb{N} \models \varphi \leftrightarrow \psi$ . We define  $<_{\mathbb{N}}$  and  $\equiv_{\mathbb{N}}$  from  $\leq_{\mathbb{N}}$  in a way analogous to the definition of  $<$  and  $\equiv$  from  $\leq$ .

It turns out that over  $\mathbb{N}$ , DL is no more expressive than first-order logic  $L_{\omega\omega}$ . This is true even for finite-test DL. The result is stated for  $\mathbb{N}$ , but is actually true for any arithmetical structure.

**THEOREM 60.**  $L_{\omega\omega} \equiv_{\mathbb{N}} DL \equiv_{\mathbb{N}} DL(\text{r.e.})$ .

The significance of this result is that in principle, one can carry out all reasoning about programs interpreted over  $\mathbb{N}$  in the first-order logic  $L_{\omega\omega}$  by translating each DL formula into an equivalent first-order formula. The translation is effective. Moreover, Theorem 60 holds for any arithmetical structure containing the requisite coding power. As mentioned earlier, every structure can be extended to an arithmetical one.

However, the translation of Theorem 60 produces unwieldy formulas having little resemblance to the original ones. This mechanism is thus somewhat unnatural and does not correspond closely to the type of arguments one would find in practical program verification. In Section 11, a remedy is provided that makes the process more orderly.

We now observe that over  $\mathbb{N}$ , DL(rich-test r.e.) has considerably more power than the equivalent logics of Theorem 60. This too is true for any arithmetical structure.

**THEOREM 61.** *Over  $\mathbb{N}$ , DL(rich-test r.e.) defines precisely the  $\Delta_1^1$  (hyperarithmetical) sets.*

Theorems 60 and 61 say that over  $\mathbb{N}$ , the languages DL and DL(r.e.) define the arithmetic (first-order definable) sets and DL(rich-test r.e.) defines the hyperarithmetical or  $\Delta_1^1$  sets. Since the inclusion between these classes is strict—for example, first-order number theory is hyperarithmetical but not arithmetic—we have

**COROLLARY 62.**  $DL(\text{r.e.}) <_{\mathbb{N}} DL(\text{rich-test r.e.})$ .

## 10 COMPLEXITY OF DL

This section addresses the complexity of first-order Dynamic Logic.

Since all versions of DL subsume first-order logic, the truth, satisfiability, or validity of a given formula can be no easier to establish than in  $L_{\omega\omega}$ . Also, since DL(r.e.) is subsumed by  $L_{\omega_1^{\text{ck}}\omega}$ , these questions are no harder to establish than in  $L_{\omega_1^{\text{ck}}\omega}$ . These bounds hold for both uninterpreted and interpreted levels of reasoning.

### 10.1 The Uninterpreted Level

In this section we discuss the complexity of the validity problem for DL. By the remarks above, this problem is between  $\Sigma_1^0$  and  $\Pi_1^1$ . That is, as a lower bound it is undecidable and can be no better than recursively enumerable, and as an upper bound it is in  $\Pi_1^1$ . This is a rather large gap, so we are still interested in determining more precise complexity bounds for DL and its variants. An interesting related question is whether there is some nontrivial<sup>5</sup> fragment of DL that is in  $\Sigma_1^0$ , since this would allow a complete axiomatization.

In the following, we consider these questions for full DL(reg), but we also consider two important subclasses of formulas for which better upper bounds are derivable:

- partial correctness assertions of the form  $\psi \rightarrow [\alpha]\varphi$ , and
- termination or total correctness assertions of the form  $\psi \rightarrow \langle \alpha \rangle \varphi$ ,

where  $\varphi$  and  $\psi$  are first-order formulas. The results are stated for regular programs, but they remain true for the more powerful programming languages too. They also hold for deterministic **while** programs.

We state the results without mentioning the underlying first-order vocabulary  $\Sigma$ . For the upper bounds this is irrelevant. For the lower bounds, we assume the  $\Sigma$  contains a unary function symbol and ternary predicate symbols.

**THEOREM 63.** *The validity problem for DL is  $\Pi_1^1$ -hard, even for formulas of the form  $\exists x [\alpha]\varphi$ , where  $\alpha$  is a regular program and  $\varphi$  is first-order.*

**THEOREM 64.** *The validity problem for DL and DL(rich-test r.e.), as well as all intermediate versions, is  $\Pi_1^1$ -complete.*

To soften the negative flavor of these results, we now observe that the special cases of unquantified one-program DL(r.e.) formulas have easier validity problems (though, as mentioned, they are still undecidable).

<sup>5</sup>*Nontrivial* here means containing  $L_{\omega\omega}$  and allowing programs with iteration. The reason for this requirement is that loop-free programs add no expressive power over first-order logic.



**THEOREM 65.** *The validity problem for the sublanguage of DL (r.e.) consisting of formulas of the form  $\langle \alpha \rangle \varphi$ , where  $\varphi$  is first-order and  $\alpha$  is an r.e. program, is  $\Sigma_1^0$ -complete.*

It is easy to see that the result holds for formulas of the form  $\psi \rightarrow \langle \alpha \rangle \varphi$ , where  $\psi$  is also first-order. Thus, termination assertions for nondeterministic programs with first-order tests (or total correctness assertions for deterministic programs), on the uninterpreted level of reasoning, are recursively enumerable and therefore axiomatizable. We shall give an explicit axiomatization in Section 11.

We now turn to partial correctness.

**THEOREM 66.** *The validity problem for the sublanguage of DL (r.e.) consisting of formulas of the form  $[\alpha] \varphi$ , where  $\varphi$  is first-order and  $\alpha$  is an r.e. program, is  $\Pi_2^0$ -complete. The  $\Pi_2^0$ -completeness property holds even if we restrict  $\alpha$  to range over deterministic **while** programs.*

Theorem 66 extends easily to partial correctness assertions; that is, to formulas of the form  $\psi \rightarrow [\alpha] \varphi$ , where  $\psi$  is also first-order. Thus, while  $\Pi_2^0$  is obviously better than  $\Pi_1^1$ , it is noteworthy that on the uninterpreted level of reasoning, the truth of even simple correctness assertions for simple programs is not r.e., so that no finitary complete axiomatization for such validities can be given.

## 10.2 The Interpreted Level

The characterizations of the various versions of DL in terms of classical static logics established in Section 9.2 provide us with the precise complexity of the validity problem over  $\mathbb{N}$ .

**THEOREM 67.** *The  $\mathbb{N}$ -validity problem for DL (dreg) and DL (rich-test r.e.), as well as all intermediate versions, when defined over the vocabulary of  $\mathbb{N}$ , is hyperarithmetical ( $\Delta_1^1$ ) but not arithmetic.*

## 10.3 Spectral Complexity

We now introduce the *spectral complexity* of a programming language. As mentioned, this notion provides a measure of the complexity of the halting problem for programs over finite interpretations.

Recall that a *state* is a finite variant of a constant valuation  $w_a$  for some  $a \in A$  (see Section 8.3), and a state  $w$  is *initial* if it differs from  $w_a$  for individual variables only. Thus, an initial state can be uniquely defined by specifying its relevant portion of values on individual variables. For  $m \in \mathbb{N}$ , we call an initial state  $w$  an *m-state* if for some  $a \in A$  and for all  $i \geq m$ ,  $w(x_i) = a$ . An *m-state* can be specified by an  $(m + 1)$ -tuple of values  $(a_0, \dots, a_m)$  that represent values of  $w$  for the first  $m + 1$  individual variables  $x_0, \dots, x_m$ . Call an *m-state*  $w = (a_0, \dots, a_m)$  *Herbrand-like* if the

set  $\{a_0, \dots, a_m\}$  generates  $A$ ; that is, if every element of  $A$  can be obtained as a value of a term in the state  $w$ .

We are now ready to define the notion of a *spectrum* of a programming language. Let  $K$  be a programming language and let  $\alpha \in K$  and  $m \geq 0$ . The  $m^{\text{th}}$  *spectrum* of  $\alpha$  is the set

$$SP_m(\alpha) \stackrel{\text{def}}{=} \{ \ulcorner \mathfrak{A}_w \urcorner \mid \mathfrak{A} \text{ is a finite } \Sigma\text{-structure, } w \text{ is an } m\text{-state in } \mathfrak{A}, \text{ and } \mathfrak{A}, w \models \langle \alpha \rangle \mathbf{1} \}.$$

The *spectrum* of  $K$  is the set

$$SP(K) \stackrel{\text{def}}{=} \{ SP_m(\alpha) \mid \alpha \in K, m \in \mathbb{N} \}.$$

Given  $m \geq 0$ , observe that structures in  $S_n^{\Sigma \cup \{c_0, \dots, c_m\}}$  can be viewed as structures of the form  $\mathfrak{A}_w$  for a certain  $\Sigma$ -structure  $\mathfrak{A}$  and an  $m$ -state  $w$  in  $\mathfrak{A}$ . This representation is unique.

In this section we establish the complexity of spectra; that is, the complexity of the halting problem in finite interpretations. Let us fix  $m \geq 0$ , a rich vocabulary  $\Sigma$ , and new constants  $c_0, \dots, c_m$ . Since not every binary string is of the form  $\ulcorner \mathfrak{A} \urcorner$  for some  $\Sigma$ -structure  $\mathfrak{A}$  and  $m$ -state  $w$  in  $\mathfrak{A}$ , we will restrict our attention to strings that are of this form. Let

$$H_m^\Sigma \stackrel{\text{def}}{=} \{ \ulcorner \mathfrak{A} \urcorner \mid \mathfrak{A} \in S_n^{\Sigma \cup \{c_0, \dots, c_m\}} \text{ for some } n \geq 1 \}.$$

It is easy to show that the language  $H_m^\Sigma$  is in *LOGSPACE* for every vocabulary  $\Sigma$  and  $m \geq 0$ .

We are now ready to connect complexity classes with spectra. Let  $K$  be any programming language and let  $C \subseteq 2^{\{0,1\}^*}$  be a family of sets. We say that  $SP(K)$  *captures*  $C$ , denoted  $SP(K) \approx C$ , if

- $SP(K) \subseteq C$ , and
- for every  $X \in C$  and  $m \geq 0$ , if  $X \subseteq H_m^\Sigma$ , then there is a program  $\alpha \in K$  such that  $SP_m(\alpha) = X$ .

For example, if  $C$  is the class of all sets recognizable in polynomial time, then  $SP(K) \approx P$  means that

- the halting problem over finite interpretations for programs from  $K$  is decidable in polynomial time, and
- every polynomial-time-recognizable set of codes of finite interpretations is the spectrum of some program from  $K$ .

We conclude this section by characterizing the spectral complexity of some of the programming languages introduced in Section 8.

THEOREM 68. *Let  $\Sigma$  be a rich vocabulary. Then*

(i)  $SP(\text{dreg}) \subseteq LOGSPACE$ .

(ii)  $SP(\text{reg}) \subseteq NLOGSPACE$ .

*Moreover, if  $\Sigma$  is mono-unary, then  $SP(\text{dreg})$  captures  $LOGSPACE$  and  $SP(\text{reg})$  captures  $NLOGSPACE$ .*

THEOREM 69. *Over a rich vocabulary  $\Sigma$ ,  $SP(\text{dstk})$  and  $SP(\text{stk})$  capture  $P$ .*

THEOREM 70. *If  $\Sigma$  is a rich vocabulary, then  $SP(\text{darray})$  and  $SP(\text{array})$  capture  $PSPACE$ .*

## 11 AXIOMATIZATION OF DL

### 11.1 Uninterpreted Reasoning

Recall from Section 10.1 that validity in DL is  $\Pi_1^1$ -complete, but only r.e. when restricted to simple termination assertions. This means that termination (or total correctness when the programs are deterministic) can be fully axiomatized in the standard sense. This we do first, and we then turn to the problem of axiomatizing full DL.

Since the validity problem for such termination assertions is r.e., it is of interest to find a nicely-structured complete axiom system. We propose the following.

#### Axiom System S1

##### Axiom Schemes

- all instances of valid first-order formulas;
- all instances of valid formulas of PDL;
- $\varphi[x/t] \rightarrow \langle x := t \rangle \varphi$ , where  $\varphi$  is a first-order formula.

##### Inference Rules

- modus ponens:

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi}$$

We denote provability in Axiom System S1 by  $\vdash_{S1}$ .

**THEOREM 71.** *For any DL formula of the form  $\varphi \rightarrow \langle \alpha \rangle \psi$ , for first-order  $\varphi$  and  $\psi$  and program  $\alpha$  containing first-order tests only,*

$$\models \varphi \rightarrow \langle \alpha \rangle \psi \quad \Leftrightarrow \quad \vdash_{S1} \varphi \rightarrow \langle \alpha \rangle \psi.$$

Given the high undecidability of validity in DL, we cannot hope for a complete axiom system in the usual sense. Nevertheless, we do want to provide an orderly axiomatization of valid DL formulas, even if this means that we have to give up the finitary nature of standard axiom systems.

Below we present a complete infinitary axiomatization S2 of DL that includes an inference rule with infinitely many premises. Before doing so, however, we must get a certain technical complication out of the way. We would like to be able to consider valid first-order formulas as axiom schemes, but instantiated by general formulas of DL. In order to make formulas amenable to first-order manipulation, we must be able to make sense of such notions as “a free occurrence of  $x$  in  $\varphi$ ” and the substitution  $\varphi[x/t]$ . For example, we would like to be able to use the axiom scheme of the predicate calculus  $\forall x \varphi \rightarrow \varphi[x/t]$ , even if  $\varphi$  contains programs.

The problem arises because the dynamic nature of the semantics of DL may cause a single occurrence of a variable in a DL formula to act as both a free and bound occurrence. For example, in the formula  $\langle \mathbf{while} \ x \leq 99 \ \mathbf{do} \ x := x + 1 \rangle$ , the occurrence of  $x$  in the expression  $x + 1$  acts as both a free occurrence (for the first assignment) and as a bound occurrence (for subsequent assignments).

There are several reasonable ways to deal with this, and we present one for definiteness. Without loss of generality, we assume that whenever required, all programs appear in the special form

$$(34) \quad \langle \bar{z} := \bar{x}; \alpha; \bar{x} := \bar{z} \rangle \varphi$$

where  $\bar{x} = (x_1, \dots, x_n)$  and  $\bar{z} = (z_1, \dots, z_n)$  are tuples of variables,  $\bar{z} := \bar{x}$  stands for

$$z_1 := x_1; \dots; z_n := x_n$$

(and similarly for  $\bar{x} := \bar{z}$ ), the  $x_i$  do not appear in  $\alpha$ , and the  $z_i$  are new variables appearing nowhere in the relevant context outside of the program  $\alpha$ . The idea is to make programs act on the “local” variables  $z_i$  by first copying the values of the  $x_i$  into the  $z_i$ , thus freezing the  $x_i$ , executing the program with the  $z_i$ , and then restoring the  $x_i$ . This form can be easily obtained from any DL formula by consistently changing all variables of any program to new ones and adding the appropriate assignments that copy and then restore the values. Clearly, the new formula is equivalent to the old. Given a DL formula in this form, the following are bound occurrences of variables:

- all occurrences of  $x$  in a subformula of the form  $\exists x \varphi$ ;
- all occurrences of  $z_i$  in a subformula of the form (34) (note, though, that  $z_i$  does not occur in  $\varphi$  at all);
- all occurrences of  $x_i$  in a subformula of the form (34) except for its occurrence in the assignment  $z_i := x_i$ .

Every occurrence of a variable that is not bound is free. Our axiom system will have an axiom that enables free translation into the special form discussed, and in the sequel we assume that the special form is used whenever required (for example, in the assignment axiom scheme below).

As an example, consider the formula:

$$\begin{aligned} \forall x (\langle y := f(x); x := g(y, x) \rangle p(x, y)) &\rightarrow \\ \langle z_1 := h(z); z_2 := y; z_2 := f(z_1); z_1 := g(z_2, z_1); x := z_1; \\ y := z_2 \rangle p(x, y). \end{aligned}$$

Denoting  $\langle y := f(x); x := g(y, x) \rangle p(x, y)$  by  $\varphi$ , the conclusion of the implication is just  $\varphi[x/h(z)]$  according to the convention above; that is, the result of replacing all free occurrences of  $x$  in  $\varphi$  by  $h(z)$  after  $\varphi$  has been transformed into special form. We want the above formula to be considered a legal instance of the assignment axiom scheme below.

## Axiom System S2

### Axiom Schemes

- all instances of valid first-order formulas;
- all instances of valid formulas of PDL;
- $\langle x := t \rangle \varphi \leftrightarrow \varphi[x/t]$ ;
- $\varphi \leftrightarrow \widehat{\varphi}$ , where  $\widehat{\varphi}$  is  $\varphi$  in which some occurrence of a program  $\alpha$  has been replaced by the program  $z := x$ ;  $\alpha'$ ;  $x := z$  for  $z$  not appearing in  $\varphi$ , and where  $\alpha'$  is  $\alpha$  with all occurrences of  $x$  replaced by  $z$ .

### Inference Rules

- modus ponens:

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi}$$

- generalization:

$$\frac{\varphi}{[\alpha]\varphi} \quad \text{and} \quad \frac{\varphi}{\forall x \varphi}$$

- infinitary convergence:

$$\frac{\varphi \rightarrow [\alpha^n]\psi, n \in \omega}{\varphi \rightarrow [\alpha^*]\psi}$$

Provability in Axiom System S2, denoted by  $\vdash_{S2}$ , is the usual concept for systems with infinitary rules of inference; that is, deriving a formula using the infinitary rule requires infinitely many premises to have been previously derived.

Axiom System S2 consists of an axiom for assignment, facilities for propositional reasoning about programs and first-order reasoning with no programs (but with programs possibly appearing in instantiated first-order formulas), and an infinitary rule for  $[\alpha^*]$ . The dual construct,  $\langle \alpha^* \rangle$ , is taken care of by the “unfolding” validity of PDL:

$$\langle \alpha^* \rangle \varphi \leftrightarrow (\varphi \vee \langle \alpha; \alpha^* \rangle \varphi).$$

**THEOREM 72.** *For any formula  $\varphi$  of DL,*

$$\models \varphi \Leftrightarrow \vdash_{S2} \varphi.$$

## 11.2 Interpreted Reasoning

Proving properties of real programs very often involves reasoning on the interpreted level, where one is interested in  $\mathfrak{A}$ -validity for a particular structure  $\mathfrak{A}$ . A typical proof might use induction on the length of the computation to establish an invariant for partial correctness or to exhibit a decreasing value in some well-founded set for termination. In each case, the problem is reduced to the problem of verifying some domain-dependent facts, sometimes called *verification conditions*. Mathematically speaking, this kind of activity is really an effective transformation of assertions about programs into ones about the underlying structure.

For DL, this transformation can be guided by a direct induction on program structure using an axiom system that is complete *relative to* any given arithmetical structure  $\mathfrak{A}$ . The essential idea is to exploit the existence, for any given DL formula, of a first-order equivalent in  $\mathfrak{A}$ , as guaranteed by Theorem 60. In the axiom systems we construct, instead of dealing with the  $\Pi_1^1$ -hardness of the validity problem by an infinitary rule, we take all  $\mathfrak{A}$ -valid first-order formulas as additional axioms. Relative to this set of axioms, proofs are finite and effective.

For partial correctness assertions of the form  $\varphi \rightarrow [\alpha]\psi$  with  $\varphi$  and  $\psi$  first-order and  $\alpha$  containing first-order tests, it suffices to show that DL reduces to the first-order logic  $L_{\omega\omega}$ , and there is no need for the natural numbers to be present. Thus, Axiom System S3 below works for finite structures too. Axiom System S4 is an *arithmetically complete* system for full DL that does make explicit use of natural numbers.

It follows from Theorem 66 that for partial correctness formulas we cannot hope to obtain a completeness result similar to the one proved in Theorem 71 for termination formulas. A way around this difficulty is to consider only *expressive* structures.

A structure  $\mathfrak{A}$  for the first-order vocabulary  $\Sigma$  is said to be *expressive* for a programming language  $K$  if for every  $\alpha \in K$  and for every first-order formula  $\varphi$ , there exists a first-order formula  $\psi_L$  such that  $\mathfrak{A} \models \psi_L \leftrightarrow [\alpha]\varphi$ . Examples of structures that are expressive for most programming languages are finite structures and arithmetical structures.

### Axiom System S3

#### Axiom Schemes

- all instances of valid formulas of PDL;
- $\langle x := t \rangle \varphi \leftrightarrow \varphi[x/t]$  for first-order  $\varphi$ .

#### Inference Rules

- modus ponens:

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi}$$

- generalization:

$$\frac{\varphi}{[\alpha]\varphi}.$$

Note that Axiom System S3 is really the axiom system for PDL from Section 4 with the addition of the assignment axiom. Given a DL formula  $\varphi$  and a structure  $\mathfrak{A}$ , denote by  $\mathfrak{A} \vdash_{S3} \varphi$  provability of  $\varphi$  in the system obtained from Axiom System S3 by adding the following set of axioms:

- all  $\mathfrak{A}$ -valid first-order sentences.

**THEOREM 73.** *For every expressive structure  $\mathfrak{A}$  and for every formula  $\xi$  of DL of the form  $\varphi \rightarrow [\alpha]\psi$ , where  $\varphi$  and  $\psi$  are first-order and  $\alpha$  involves only first-order tests, we have*

$$\mathfrak{A} \models \xi \iff \mathfrak{A} \vdash_{S_3} \xi.$$

Now we present an axiom system S4 for full DL. It is similar in spirit to S3 in that it is complete relative to the formulas valid in the structure under consideration. However, this system works for arithmetical structures only. It is not tailored to deal with other expressive structures, notably finite ones, since it requires the use of the natural numbers. The kind of completeness result stated here is thus termed *arithmetical*.

As in Section 9.2, we state the results for the special structure  $\mathbb{N}$ , omitting the technicalities needed to deal with general arithmetical structures. The main difference is that in  $\mathbb{N}$  we can use variables  $n$ ,  $m$ , etc., knowing that their values will be natural numbers. We can thus write  $n + 1$ , for example, assuming the standard interpretation. When working in an unspecified arithmetical structure, we have to precede such usage with appropriate predicates that guarantee that we are indeed talking about that part of the domain that is isomorphic to the natural numbers. For example, we would often have to use the first-order formula, call it  $\text{nat}(n)$ , which is true precisely for the elements representing natural numbers, and which exists by the definition of an arithmetical structure.

### Axiom System S4

#### Axiom Schemes

- all instances of valid first-order formulas;
- all instances of valid formulas of PDL;
- $\langle x := t \rangle \varphi \leftrightarrow \varphi[x/t]$  for first-order  $\varphi$ .

#### Inference Rules

- modus ponens:

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi}$$

- generalization:

$$\frac{\varphi}{[\alpha]\varphi} \quad \text{and} \quad \frac{\varphi}{\forall x \varphi}$$



- convergence:

$$\frac{\varphi(n+1) \rightarrow \langle \alpha \rangle \varphi(n)}{\varphi(n) \rightarrow \langle \alpha^* \rangle \varphi(0)}$$

for first order  $\varphi$  and variable  $n$  not appearing in  $\alpha$ .

REMARK 74. For general arithmetical structures, the  $+1$  and  $0$  in the rule of convergence denote suitable first-order definitions.

As in Axiom System S3, denote by  $\mathfrak{A} \vdash_{S4} \varphi$  provability of  $\varphi$  in the system obtained from Axiom System S4 by adding all  $\mathfrak{A}$ -valid first-order sentences as axioms.

THEOREM 75. *For every formula  $\xi$  of DL,*

$$\mathbb{N} \models \xi \Leftrightarrow \mathbb{N} \vdash_{S4} \xi.$$

The use of the natural numbers as a device for counting down to 0 in the convergence rule of Axiom System S4 can be relaxed. In fact, any well-founded set suitably expressible in any given arithmetical structure suffices. Also, it is not necessary to require that an execution of  $\alpha$  causes the truth of the parameterized  $\varphi(n)$  in that rule to decrease exactly by 1; it suffices that the decrease is positive at each iteration.

In closing, we note that appropriately restricted versions of all axiom systems of this section are complete for DL (dreg). In particular, as pointed out in Section 2.6, the Hoare **while**-rule

$$\frac{\varphi \wedge \xi \rightarrow [\alpha] \varphi}{\varphi \rightarrow [\mathbf{while} \ \xi \ \mathbf{do} \ \alpha] (\varphi \wedge \neg \xi)}$$

results from combining the generalization rule with the induction and test axioms of PDL, when  $*$  is restricted to appear only in the context of a **while** statement; that is, only in the form  $(\xi?; p)^*$ ;  $(\neg \xi)?$ .

## 12 EXPRESSIVENESS OF DL

The subject of study in this section is the relative expressive power of languages. We will be primarily interested in comparing, on the uninterpreted level, the expressive power of various versions of DL. That is, for programming languages  $P_1$  and  $P_2$  we will study whether  $\text{DL}(P_1) \leq \text{DL}(P_2)$  holds. Recall from Section 9 that the latter relation means that for each formula  $\varphi$  in  $\text{DL}(P_1)$ , there is a formula  $\psi$  in  $\text{DL}(P_2)$  such that  $\mathfrak{A}, u \models \varphi \leftrightarrow \psi$  for all structures  $\mathfrak{A}$  and initial states  $u$ .

Studying the expressive power of logics rather than the computational power of programs allows us to compare, for example, deterministic and

nondeterministic programming languages. Also, we will see that the answer to the fundamental question “ $DL(P_1) \leq DL(P_2)$ ?” may depend crucially on the vocabulary over which we consider logics and programs. For this reason we always make clear in the theorems of this section our assumptions on the vocabulary.

**THEOREM 76.** *Let  $\Sigma$  be a rich vocabulary. Then*

- (i)  $DL(\text{stk}) \leq DL(\text{array})$ .
- (ii)  $DL(\text{stk}) \equiv DL(\text{array})$  iff  $P = PSPACE$ .

*Moreover, the same holds for deterministic regular programs with an algebraic stack and deterministic regular programs with arrays.*

**THEOREM 77.** *Over a monadic vocabulary, nondeterministic regular programs with a Boolean stack have the same computational power as nondeterministic regular programs with an algebraic stack.*

Now we investigate the role that nondeterminism plays in the expressive power of logics of programs. As we shall see, the general conclusion is that for a programming language of sufficient computational power, nondeterminism does not increase the expressive power of the logic.

We start our discussion of the role of nondeterminism with the basic case of regular programs. Recall that  $DL$  and  $DDL$  denote the logics of nondeterministic and deterministic regular programs, respectively.

We can now state the main result that separates the expressive power of deterministic and nondeterministic **while** programs.

**THEOREM 78.** *For every vocabulary containing at least two unary function symbols or at least one function symbol of arity greater than one,  $DDL$  is strictly less expressive than  $DL$ ; that is,  $DDL < DL$ .*

It turns out that Theorem 78 cannot be extended to vocabularies containing just one unary function symbol without solving a well known open problem in complexity theory.

**THEOREM 79.** *For every rich mono-unary vocabulary, the statement “ $DDL$  is strictly less expressive than  $DL$ ” is equivalent to  $LOGSPACE \neq NLOGSPACE$ .*

We now turn our attention to the discussion of the role nondeterminism plays in the expressive power of regular programs with a Boolean stack. For a vocabulary containing at least two unary function symbols, nondeterminism increases the expressive power of  $DL$  over regular programs ‘ with a Boolean stack.

For the rest of this section, we let the vocabulary contain two unary function symbols.

**THEOREM 80.** *For a vocabulary containing at least two unary function symbols or a function symbol of arity greater than two,  $DL(\text{dbstk}) < DL(\text{bstk})$ .*

It turns out that for programming languages that use sufficiently strong data types, nondeterminism does not increase the expressive power of Dynamic Logic.

**THEOREM 81.** *For every vocabulary,*

- (i)  $\text{DL}(\text{dstk}) \equiv \text{DL}(\text{stk});$
- (ii)  $\text{DL}(\text{darray}) \equiv \text{DL}(\text{array}).$

We will discuss the role of unbounded memory of programs for the expressive power of the corresponding logic. However, this result depends on assumptions about the vocabulary  $\Sigma$ .

Recall from Section 8.2 that an r.e. program  $\alpha$  has *bounded memory* if the set  $CS(\alpha)$  contains only finitely many distinct variables from  $V$ , and if in addition the nesting of function symbols in terms that occur in seqs of  $CS(\alpha)$  is bounded. This restriction implies that such a program can be simulated in all interpretations by a device that uses a fixed finite number of registers, say  $x_1, \dots, x_n$ , and all its elementary steps consist of either performing a test of the form

$$r(x_{i_1}, \dots, x_{i_m})?,$$

where  $r$  is an  $m$ -ary relation symbol of  $\Sigma$ , or executing a simple assignment of either of the following two forms:

$$x_i := f(x_{i_1}, \dots, x_{i_k}) \qquad x_i := x_j.$$

In general, however, such a device may need a very powerful control (that of a Turing machine) to decide which elementary step to take next.

An example of a programming language with bounded memory is the class of regular programs with a Boolean stack. Indeed, the Boolean stack strengthens the control structure of a regular program without introducing extra registers for storing algebraic elements. It can be shown without much difficulty that regular programs with a Boolean stack have bounded memory. On the other hand, regular programs with an algebraic stack or with arrays are programming languages with unbounded memory.

For monadic vocabularies, the class of nondeterministic regular programs with a Boolean stack is computationally equivalent to the class of nondeterministic regular programs with an algebraic stack. For deterministic programs, the situation is slightly different.

**THEOREM 82.**

- (i) *For every vocabulary containing a function symbol of arity greater than one,  $\text{DL}(\text{dbstk}) < \text{DL}(\text{dstk})$  and  $\text{DL}(\text{bstk}) < \text{DL}(\text{stk}).$*
- (ii) *For all monadic vocabularies,  $\text{DL}(\text{bstk}) \equiv \text{DL}(\text{stk}).$*

- (iii) For all mono-unary vocabularies,  $DL(\text{dbstk}) \equiv DL(\text{dstk})$ .
- (iv) For all monadic vocabularies containing at least two function symbols,  $DL(\text{dbstk}) < DL(\text{dstk})$ .

Regular programs with a Boolean stack are situated between pure regular programs and regular programs with an algebraic stack. We start our discussion by comparing the expressive power of regular programs with and without a Boolean stack. The only known definite answer to this problem is given in the following result, which covers the case of deterministic programs only.

**THEOREM 83.**

- (i) Let the vocabulary be rich and mono-unary. Then

$$DL(\text{dreg}) \equiv DL(\text{dstk}) \Leftrightarrow LOGSPACE = P.$$

- (ii) If the vocabulary contains at least one function symbol of arity greater than one or at least two unary function symbols, then  $DL(\text{dreg}) < DL(\text{dbstk})$ .

It is not known whether Theorem 83(ii) holds for nondeterministic programs, and neither is its statement known to be equivalent to any of the well known open problems in complexity theory. In contrast, it follows from Theorems 83(i) and 82(iii) that for rich mono-unary vocabularies,  $DL(\text{dreg}) \equiv DL(\text{dbstk})$  if and only if  $LOGSPACE = P$ . Hence, this problem cannot be solved without solving one of the major open problems in complexity theory.

The wildcard assignment statement  $x := ?$  discussed in Section 8.2 chooses an element of the domain of computation nondeterministically and assigns it to  $x$ . It is a device that represents *unbounded nondeterminism* as opposed to the binary nondeterminism of the nondeterministic choice construct  $\cup$ . The programming language of regular programs augmented with wildcard assignment is not an acceptable programming language, since a wildcard assignment can produce values that are outside the substructure generated by the input.

Our first result shows that wildcard assignment increases the expressive power in quite a substantial way; it cannot be simulated even by r.e. programs.

**THEOREM 84.** *Let the vocabulary  $\Sigma$  contain two constants  $c_1, c_2$ , a binary predicate symbol  $p$ , the symbol  $=$  for equality, and no other function or predicate symbols. There is a formula of  $DL(\text{wild})$  that is equivalent to no formula of  $DL(\text{r.e.})$ , thus  $DL(\text{wild}) \not\leq DL(\text{r.e.})$ .*

It is not known whether any of the logics with unbounded memory are reducible to DL(wild).

When both wildcard and array assignments are allowed, it is possible to define the finiteness of (the domain of) a structure, but not in the logics with either of the additions removed. Thus, having both memory and non-determinism unbounded provides more power than having either of them bounded.

**THEOREM 85.** *Let vocabulary  $\Sigma$  contain only the symbol of equality. There is a formula of DL(array+wild) equivalent to no formula of either DL(array) or DL(wild).*

### 13 VARIANTS OF DL

In this section we consider some restrictions and extensions of DL. We are interested mainly in questions of comparative expressive power on the uninterpreted level. In arithmetical structures these questions usually become trivial, since it is difficult to go beyond the power of first-order arithmetic without allowing infinitely many distinct tests in programs (see Theorems 60 and 61). In regular DL this luxury is not present.

#### 13.1 Algorithmic Logic

Algorithmic Logic (AL) is the predecessor of Dynamic Logic. The basic system was defined by [Salwicki, 1970] and generated an extensive amount of subsequent research carried out by a group of mathematicians working in Warsaw. Two surveys of the first few years of their work can be found in [Banachowski *et al.*, 1977] and [Salwicki, 1977].

The original version of AL allowed deterministic **while** programs and formulas built from the constructs

$$\alpha\varphi \quad \cup \alpha\varphi \quad \cap \alpha\varphi$$

corresponding in our terminology to

$$\langle \alpha \rangle \varphi \quad \langle \alpha^* \rangle \varphi \quad \bigwedge_{n \in \omega} \langle \alpha^n \rangle \varphi,$$

respectively, where  $\alpha$  is a deterministic **while** program and  $\varphi$  is a quantifier-free first-order formula.

In [Mirkowska, 1980; Mirkowska, 1981a; Mirkowska, 1981b], AL was extended to allow nondeterministic **while** programs and the constructs

$$\nabla \alpha\varphi \quad \Delta \alpha\varphi$$

corresponding in our terminology to

$$\langle \alpha \rangle \varphi \quad \mathbf{halt}(\alpha) \wedge [\alpha]\varphi \wedge \langle \alpha \rangle \varphi,$$

respectively. The latter asserts that all traces of  $\alpha$  are finite and terminate in a state satisfying  $\varphi$ .

A feature present in AL but not in DL is the set of “dynamic terms” in addition to dynamic formulas. For a first-order term  $t$  and a deterministic **while** program  $\alpha$ , the meaning of the expression  $\alpha t$  is the value of  $t$  after executing program  $\alpha$ . If  $\alpha$  does not halt, the meaning is undefined. Such terms can be systematically eliminated; for example,  $P(x, \alpha t)$  is replaced by  $\exists z (\langle \alpha \rangle (z = t) \wedge P(x, z))$ .

The emphasis in the early research on AL was in obtaining infinitary completeness results, developing normal forms for programs, investigating recursive procedures with parameters, and axiomatizing certain aspects of programming using formulas of AL. As an example of the latter, the algorithmic formula

$$(\mathbf{while} \ s \neq \varepsilon \ \mathbf{do} \ s := \mathbf{pop}(s))\mathbf{1}$$

can be viewed as an axiom connected with the data structure *stack*. One can then investigate the consequences of such axioms within AL, regarding them as properties of the corresponding data structures.

Complete infinitary deductive systems for first-order and propositional versions are given in [Mirkowska, 1980; Mirkowska, 1981a; Mirkowska, 1981b]. The infinitary completeness results for AL are usually proved by the algebraic methods of [Rasiowa and Sikorski, 1963].

[Constable, 1977], [Constable and O’Donnell, 1978] and [Goldblatt, 1982] present logics similar to AL and DL for reasoning about deterministic **while** programs.

### 13.2 Well-Foundedness

As in Section 7 for PDL, we consider adding to DL assertions to the effect that programs can enter infinite computations. Here too, we shall be interested both in LDL and in RDL versions; i.e., those in which **halt**  $\alpha$  and **wf**  $\alpha$ , respectively, have been added inductively as new formulas for any program  $\alpha$ . As mentioned there, the connection with the more common notation **repeat**  $\alpha$  and **loop**  $\alpha$  (from which the L and R in the names LDL and RDL derive) is by:

$$\begin{aligned} \mathbf{loop} \ \alpha &\stackrel{\text{def}}{\iff} \neg \mathbf{halt} \ \alpha \\ \mathbf{repeat} \ \alpha &\stackrel{\text{def}}{\iff} \neg \mathbf{wf} \ \alpha. \end{aligned}$$

We now state some of the relevant results. The first concerns the addition of **halt**  $\alpha$ :

**THEOREM 86.** LDL  $\equiv$  DL.

In contrast to this, we have:

THEOREM 87.  $\text{LDL} < \text{RDL}$ .

Turning to the validity problem for these extensions, clearly they cannot be any harder to decide than that of  $\text{DL}$ , which is  $\Pi_1^1$ -complete. However, the following result shows that detecting the absence of infinite computations of even simple uninterpreted programs is extremely hard.

THEOREM 88. *The validity problems for formulas of the form  $\varphi \rightarrow \mathbf{wf}\alpha$  and formulas of the form  $\varphi \rightarrow \mathbf{halt}\alpha$ , for first-order  $\varphi$  and regular  $\alpha$ , are both  $\Pi_1^1$ -complete. If  $\alpha$  is constrained to have only first-order tests then the  $\varphi \rightarrow \mathbf{wf}\alpha$  case remains  $\Pi_1^1$ -complete but the  $\varphi \rightarrow \mathbf{halt}\alpha$  case is r.e.; that is, it is  $\Sigma_1^0$ -complete.*

We just mention here that the additions to Axiom System S4 of Section 11 that are used to obtain an arithmetically complete system for  $\text{RDL}$  are the axiom

$$[\alpha^*](\varphi \rightarrow \langle \alpha \rangle \varphi) \rightarrow (\varphi \rightarrow \neg \mathbf{wf} \alpha)$$

and the inference rule

$$\frac{\varphi(n+1) \rightarrow [\alpha]\varphi(n), \neg\varphi(0)}{\varphi(n) \rightarrow \mathbf{wf} \alpha}$$

for first-order  $\varphi$  and  $n$  not occurring in  $\alpha$ .

### 13.3 Probabilistic Programs

There is wide interest recently in programs that employ probabilistic moves such as coin tossing or random number draws and whose behavior is described probabilistically (for example,  $\alpha$  is “correct” if it does what it is meant to do with probability 1). To give one well known example taken from [Miller, 1976] and [Rabin, 1980], there are fast probabilistic algorithms for checking primality of numbers but no known fast nonprobabilistic ones. Many synchronization problems including digital contract signing, guaranteeing mutual exclusion, etc. are often solved by probabilistic means.

This interest has prompted research into formal and informal methods for reasoning about probabilistic programs. It should be noted that such methods are also applicable for reasoning probabilistically about ordinary programs, for example, in average-case complexity analysis of a program, where inputs are regarded as coming from some set with a probability distribution.

[Kozen, 1981d] provided a formal semantics for probabilistic first-order **while** programs with a random assignment statement  $x := ?$ . Here the term “random” is quite appropriate (contrast with Section 8.2) as the statement essentially picks an element out of some fixed distribution over the domain  $D$ . This domain is assumed to be given with an appropriate set of measurable subsets. Programs are then interpreted as measurable functions on a certain measurable product space of copies of  $D$ .

In [Feldman and Harel, 1984] a probabilistic version of first-order Dynamic Logic,  $Pr(DL)$ , was investigated on the interpreted level. Kozen's semantics is extended as described below to a semantics for formulas that are closed under Boolean connectives and quantification over reals and integers and that employ terms of the form  $Fr(\varphi)$  for first-order  $\varphi$ . In addition, if  $\alpha$  is a **while** program with nondeterministic assignments and  $\varphi$  is a formula, then  $\{\alpha\}\varphi$  is a new formula.

The semantics assumes a domain  $D$ , say the reals, with a measure space consisting of an appropriate family of *measurable subsets* of  $D$ . The states  $\mu, \nu, \dots$  are then taken to be the positive measures on this measure space. Terms are interpreted as functions from states to real numbers, with  $Fr(\varphi)$  in  $\mu$  being the *frequency* (or simply, the *measure*) of  $\varphi$  in  $\mu$ . Frequency is to positive measures as probability is to probability measures. The formula  $\{\alpha\}\varphi$  is true in  $\mu$  if  $\varphi$  is true in  $\nu$ , the state (i.e., measure) that is the result of applying  $\alpha$  to  $\mu$  in Kozen's semantics. Thus  $\{\alpha\}\varphi$  means "after  $\alpha$ ,  $\varphi$ " and is the construct analogous to  $\langle \alpha \rangle \varphi$  of DL.

For example, in  $Pr(DL)$  one can write

$$Fr(1) = 1 \rightarrow \{\alpha\}Fr(1) \geq p$$

to mean, " $\alpha$  halts with probability at least  $p$ ." The formula

$$\begin{aligned} Fr(1) = 1 \rightarrow & [i := 1; x := ?; \mathbf{while} \ x > 1/2 \ \mathbf{do} \ (x := ?; i := i + 1)] \\ & \forall n \ ((n \geq 1 \rightarrow Fr(i = n) = 2^{-n}) \wedge \\ & (n < 1 \rightarrow Fr(i = n) = 0)) \end{aligned}$$

is valid in all structures in which the distribution of the random variable used in  $x := ?$  is a uniform distribution on the real interval  $[0, 1]$ .

An axiom system for  $Pr(DL)$  was proved in [Feldman and Harel, 1984] to be complete relative to an extension of first-order analysis with integer variables, and for discrete probabilities first-order analysis with integer variables was shown to suffice.

## 14 OTHER APPROACHES

Here we discuss briefly some topics closely related to Dynamic Logic.

### 14.1 Logic of Effective Definitions

The Logic of Effective Definitions (LED), introduced by [Tiuryn, 1981a], was intended to study notions of computability over abstract models and to provide a universal framework for the study of logics of programs over such models. It consists of first-order logic augmented with new atomic formulas of the form  $\alpha = \beta$ , where  $\alpha$  and  $\beta$  are *effective definitional schemes* (the latter notion is due to [Friedman, 1971]):



```

if  $\varphi_1$  then  $t_1$ 
  else if  $\varphi_2$  then  $t_2$ 
    else if  $\varphi_3$  then  $t_3$ 
      else if ...

```

where the  $\varphi_i$  are quantifier-free formulas and  $t_i$  are terms over a bounded set of variables, and the function  $i \mapsto (\varphi_i, t_i)$  is recursive. The formula  $\alpha = \beta$  is defined to be true in a state if both  $\alpha$  and  $\beta$  terminate and yield the same value, or neither terminates.

Model theory and infinitary completeness of LED are treated in [Tiuryn, 1981a].

Effective definitional schemes in the definition of LED can be replaced by any programming language  $K$ , giving rise to various logical formalisms. The following result, which relates LED to other logics discussed here, is proved in [Meyer and Tiuryn, 1981; Meyer and Tiuryn, 1984].

**THEOREM 89.** *For every vocabulary  $L$ ,  $\text{LED} \equiv \text{DL}(\text{r.e.})$ .*

## 14.2 Temporal Logic

Temporal Logic (TL) is an alternative application of modal logic to program specification and verification. It was first proposed as a useful tool in program verification by [Pnueli, 1977] and has since been developed by many authors in various forms. This topic is surveyed in depth in [Emerson, 1990] and [Gabbay *et al.*, 1994].

TL differs from DL chiefly in that it is *endogenous*; that is, programs are not explicit in the language. Every application has a single program associated with it, and the language may contain program-specific statements such as **at**  $L$ , meaning “execution is currently at location  $L$  in the program.” There are two competing semantics, giving rise to two different theories called *linear-time* and *branching-time* TL. In the former, a model is a linear sequence of program states representing an execution sequence of a deterministic program or a possible execution sequence of a nondeterministic or concurrent program. In the latter, a model is a tree of program states representing the space of all possible traces of a nondeterministic or concurrent program. Depending on the application and the semantics, different syntactic constructs can be chosen. The relative advantages of linear and branching time semantics are discussed in [Lamport, 1980; Emerson and Halpern, 1986; Emerson and Lei, 1987; Vardi, 1998a].

Modal constructs used in TL include

$\Box\varphi$	“ $\varphi$ holds in all future states”
$\Diamond\varphi$	“ $\varphi$ holds in some future state”
$\bigcirc\varphi$	“ $\varphi$ holds in the next state”
$\varphi \text{ until } \psi$	“there exists some strictly future point $t$ at which $\psi$ will be satisfied and all points strictly between the current state and $t$ satisfy $\varphi$ ”

for linear-time logic, as well as constructs for expressing

“for all traces starting from the present state . . . ”  
 “for some trace starting from the present state . . . ”

for branching-time logic.

Temporal logic is useful in situations where programs are not normally supposed to halt, such as operating systems, and is particularly well suited to the study of concurrency. Many classical program verification methods such as the *intermittent assertions method* are treated quite elegantly in this framework.

Temporal logic has been most successful in providing tools for proving properties of concurrent *finite state* protocols, such as solutions to the *dining philosophers* and *mutual exclusion* problems, which are popular abstract versions of synchronization and resource management problems in distributed systems.

The induction principle of TL takes the form:

$$(35) \quad \varphi \wedge \Box(\varphi \rightarrow \bigcirc\varphi) \rightarrow \Box\varphi.$$

Note the similarity to the PDL induction axiom (Axiom 17(viii)):

$$\varphi \wedge [\alpha^*](\varphi \rightarrow [\alpha]\varphi) \rightarrow [\alpha^*]\varphi.$$

This is a classical program verification method known as *inductive* or *invariant assertions*.

The operators  $\bigcirc$ ,  $\Diamond$ , and  $\Box$  can all be defined in terms of **until**:

$$\begin{aligned} \bigcirc\varphi &\Leftrightarrow \neg(\mathbf{0} \text{ until } \neg\varphi) \\ \Diamond\varphi &\Leftrightarrow \varphi \vee (\mathbf{1} \text{ until } \varphi) \\ \Box\varphi &\Leftrightarrow \varphi \wedge \neg(\mathbf{1} \text{ until } \neg\varphi), \end{aligned}$$

but not vice-versa. It has been shown in [Kamp, 1968] and [Gabbay *et al.*, 1980] that the **until** operator is powerful enough to express anything that can be expressed in the first-order theory of  $(\omega, <)$ . It has also been shown in [Wolper, 1981; Wolper, 1983] that there are very simple predicates that cannot be expressed by **until**; for example, “ $\varphi$  is true at every multiple of 4.”

The **until** operator has been shown to be very useful in expressing properties of programs that are not properties of the input/output relation, such as: “If process  $p$  requests a resource before  $q$  does, then it will receive it before  $q$  does.” Indeed, much of the research in TL has concentrated on providing useful methods for proving these and other kinds of properties (see [Manna and Pnueli, 1981; Gabbay *et al.*, 1980]).

### *Concurrency and Nondeterminism*

Unlike DL, TL can be applied to programs that are not normally supposed to halt, such as operating systems, because programs are interpreted as *traces* instead of pairs of states.

Up to now we have only considered deterministic, single-process programs. There is no reason however not to apply TL to *nondeterministic* and *concurrent (multiprocessor)* systems, in which next states are not unique. The computation is no longer a single trace, but many different traces are possible. We can assemble them all together to get a *computation tree* in which each node represents a state accessible from the start state.

As above, an *invariance property* is a property of the form  $\Box\varphi$ . However, the dual  $\Diamond$  of the operator  $\Box$  defined in this way does not really capture what we mean by *eventuality* or *liveness* properties. We would like to be able to say that *every* possible trace in the computation tree has a state satisfying  $\varphi$ . For instance, a nondeterministic program is *total* if there is no chance of an infinite trace out of the start state  $s$ ; that is, every trace out of  $s$  satisfies  $\Diamond\mathbf{halt}$ . The dual  $\Diamond$  of  $\Box$  as defined by  $\Diamond\varphi = \neg\Box\neg\varphi$  does not really express this. It says instead

$$s \models \Diamond\varphi \Leftrightarrow \text{there is some node } t \text{ in the tree below } s \text{ such that } t \models \varphi.$$

This is not a very useful statement.

One way to fix this is to introduce the branching time operator **A** that says, “For all traces in the tree ...,” and then use  $\Box$ ,  $\Diamond$  in the sense of linear TL applied to the trace quantified by **A**. The dual of **A** is **E**, which says, “There exists a trace in the tree ... .” Thus, in order to say that the computation tree starting from the current state satisfies a safety or invariance property, we would write

$$\mathbf{A}\Box\varphi,$$

which says, “For all traces  $\pi$  out of the current state,  $\pi$  satisfies  $\Box\varphi$ ,” and to say that the tree satisfies an eventuality property, we would write

$$\mathbf{A}\Diamond\varphi,$$

which says, “For all traces  $\pi$  out of the current state,  $\pi$  satisfies  $\Diamond\varphi$ ; that is,  $\varphi$  occurs somewhere along the trace  $\pi$ .” The logic with the linear temporal

operators augmented with the trace quantifiers **A** and **E** is known as CTL; see [Emerson, 1990; Emerson and Halpern, 1986; Emerson and Halpern, 1985; Emerson and Lei, 1987; Emerson and Sistla, 1984].

### *Complexity and Deductive Completeness*

A useful axiomatization of linear-time TL without the until operator is given by the axioms

$$\begin{aligned}
\Box(\varphi \rightarrow \psi) &\rightarrow (\Box\varphi \rightarrow \Box\psi) \\
\Box(\varphi \wedge \psi) &\leftrightarrow \Box\varphi \wedge \Box\psi \\
\Diamond\varphi &\leftrightarrow \varphi \vee \Box\Diamond\varphi \\
\bigcirc(\varphi \vee \psi) &\leftrightarrow \bigcirc\varphi \vee \bigcirc\psi \\
\bigcirc(\varphi \wedge \psi) &\leftrightarrow \bigcirc\varphi \wedge \bigcirc\psi \\
\varphi \wedge \Box(\varphi \rightarrow \bigcirc\varphi) &\rightarrow \Box\varphi \\
\forall x \varphi(x) &\rightarrow \varphi(t) \quad (t \text{ is free for } x \text{ in } \varphi) \\
\forall x \Box\varphi &\rightarrow \Box\forall x \varphi
\end{aligned}$$

and rules

$$\frac{\varphi, \varphi \rightarrow \psi}{\psi} \quad \frac{\varphi}{\Box\varphi} \quad \frac{\varphi}{\forall x \varphi}.$$

Compare the axioms of PDL (Axioms 17). The propositional fragment of this deductive system is complete for linear-time propositional TL, as shown in [Gabbay *et al.*, 1980].

[Sistla and Clarke, 1982] and [Emerson and Halpern, 1985] have shown that the validity problem for most versions of propositional TL is *PSPACE*-complete for linear structures and *EXPTIME*-complete for branching structures.

### *Embedding TL in DL*

TL is subsumed by DL. To embed propositional TL into PDL, take an atomic program  $a$  to mean “one step of program  $p$ .” In the linear model, the TL constructs  $\bigcirc\varphi$ ,  $\Box\varphi$ ,  $\Diamond\varphi$ , and  $\varphi$  **until**  $\psi$  are then expressed by  $[a]\varphi$ ,  $[a^*]\varphi$ ,  $\langle a^* \rangle\varphi$ , and  $\langle (a; \varphi?)^* \rangle; a \rangle\psi$ , respectively.

### 14.3 Process Logic

Dynamic Logic and Temporal Logic embody markedly different approaches to reasoning about programs. This dichotomy has prompted researchers to search for an appropriate process logic that combines the best features of both. An appropriate candidate should combine the ability to reason

about programs compositionally with the ability to reason directly about the intermediate states encountered during the course of a computation.

[Pratt, 1979c], [Parikh, 1978b], [Nishimura, 1980], and [Harel *et al.*, 1982b] all suggested increasingly more powerful propositional-level formalisms in which the basic idea is to interpret formulas in *traces* rather than in states. In particular, [Harel *et al.*, 1982b] present a system called *Process Logic* (PL), which is essentially a union of TL and test-free regular PDL. That paper proves that the satisfiability problem is decidable and gives a complete finitary axiomatization.

Syntactically, we have programs  $\alpha, \beta, \dots$  and propositions  $\varphi, \psi, \dots$  as in PDL. We have atomic symbols of each type and compound expressions built up from the operators  $\rightarrow, \mathbf{0}, ;, \cup, *, ?$  (applied to Boolean combinations of atomic formulas only),  $\omega$ , and  $[ ]$ . In addition we have the temporal operators **first** and **until**. The temporal operators are available for expressing and reasoning about trace properties, but programs are constructed compositionally as in PDL. Other operators are defined as in PDL (see Section 2.1) except for **skip**, which is handled specially.

Semantically, both programs and propositions are interpreted as sets of traces. We start with a Kripke frame  $\mathfrak{K} = (K, \mathfrak{m}_{\mathfrak{K}})$  as in Section 2.2, where  $K$  is a set of *states*  $s, t, \dots$  and the function  $\mathfrak{m}_{\mathfrak{K}}$  interprets atomic formulas  $p$  as subsets of  $K$  and atomic programs  $a$  as binary relations on  $K$ . The temporal operators are defined as in TL.

Trace models satisfy (most of) the PDL axioms. As in Section 14.2, define

$$\begin{aligned} \mathbf{halt} &\stackrel{\text{def}}{\iff} \mathbf{0} \\ \mathbf{fin} &\stackrel{\text{def}}{\iff} \diamond \mathbf{halt} \\ \mathbf{inf} &\stackrel{\text{def}}{\iff} \neg \mathbf{fin}, \end{aligned}$$

which say that the trace is of length 0, of finite length, or of infinite length, respectively. Define two new operators  $\llbracket \ ]$  and  $\langle \! \langle \ \rangle \! \rangle$ :

$$\begin{aligned} \llbracket \alpha \rrbracket \varphi &\stackrel{\text{def}}{\iff} \mathbf{fin} \rightarrow [\alpha] \varphi \\ \langle \! \langle \alpha \rangle \! \rangle \varphi &\stackrel{\text{def}}{\iff} \neg \llbracket \alpha \rrbracket \neg \varphi \iff \mathbf{fin} \wedge \langle \alpha \rangle \varphi. \end{aligned}$$

The  $*$  operator is the same as in PDL. It can be shown that the two PDL axioms

$$\begin{aligned} \varphi \wedge [\alpha] [\alpha^*] \varphi &\leftrightarrow [\alpha^*] \varphi \\ \varphi \wedge [\alpha^*] (\varphi \rightarrow [\alpha] \varphi) &\rightarrow [\alpha^*] \varphi \end{aligned}$$

hold by establishing that

$$\begin{aligned} \bigcup_{n \geq 0} \mathfrak{m}_{\mathfrak{K}}(\alpha^n) &= \mathfrak{m}_{\mathfrak{K}}(\alpha^0) \cup (\mathfrak{m}_{\mathfrak{K}}(\alpha) \circ \bigcup_{n \geq 0} \mathfrak{m}_{\mathfrak{K}}(\alpha^n)) \\ &= \mathfrak{m}_{\mathfrak{K}}(\alpha^0) \cup ((\bigcup_{n \geq 0} \mathfrak{m}_{\mathfrak{K}}(\alpha^n)) \circ \mathfrak{m}_{\mathfrak{K}}(\alpha)). \end{aligned}$$

As mentioned, the version of PL of [Harel *et al.*, 1982b] is decidable (but, it seems, in nonelementary time only) and complete. It has also been shown that if we restrict the semantics to include only finite traces (not a necessary restriction for obtaining the results above), then PL is no more expressive than PDL. Translations of PL structures into PDL structures have also been investigated, making possible an elementary time decision procedure for deterministic PL; see [Halpern, 1982; Halpern, 1983]. An extension of PL in which **first** and **until** are replaced by regular operators on formulas has been shown to be decidable but nonelementary in [Harel *et al.*, 1982b]. This logic perhaps comes closer to the desired objective of a powerful decidable logic of traces with natural syntactic operators that is closed under attachment of regular programs to formulas.

#### 14.4 The $\mu$ -Calculus

The  $\mu$ -calculus was suggested as a formalism for reasoning about programs in [Scott and de Bakker, 1969] and was further developed in [Hitchcock and Park, 1972], [Park, 1976], and [de Bakker, 1980].

The heart of the approach is  $\mu$ , the *least fixpoint* operator, which captures the notions of iteration and recursion. The calculus was originally defined as a first-order-level formalism, but propositional versions have become popular.

The  $\mu$  operator binds relation variables. If  $\varphi(X)$  is a logical expression with a free relation variable  $X$ , then the expression  $\mu X.\varphi(X)$  represents the least  $X$  such that  $\varphi(X) = X$ , if such an  $X$  exists. For example, the reflexive transitive closure  $R^*$  of a binary relation  $R$  is the least binary relation containing  $R$  and closed under reflexivity and transitivity; this would be expressed in the first-order  $\mu$ -calculus as

$$(36) \quad R^* \stackrel{\text{def}}{=} \mu X(x, y).(x = y \vee \exists z (R(x, z) \wedge X(z, y))).$$

This should be read as, “the least binary relation  $X(x, y)$  such that either  $x = y$  or  $x$  is related by  $R$  to some  $z$  such that  $z$  and  $y$  are already related by  $X$ .” This captures the usual fixpoint formulation of reflexive transitive closure. The formula (36) can be regarded either as a recursive program computing  $R^*$  or as an inductively defined assertion that is true of a pair  $(x, y)$  iff that pair is in the reflexive transitive closure of  $R$ .

The existence of a least fixpoint is not guaranteed except under certain restrictions. Indeed, the formula  $\neg X$  has no fixpoint, therefore  $\mu X.\neg X$  does not exist. Typically, one restricts the application of the binding operator  $\mu X$  to formulas that are *positive* or *syntactically monotone* in  $X$ ; that is, those formulas in which every free occurrence of  $X$  occurs in the scope of an even number of negations. This implies that the relation operator  $X \mapsto \varphi(X)$  is (semantically) monotone, which by the Knaster–Tarski theorem ensures the existence of a least fixpoint.

The first-order  $\mu$ -calculus can define all sets definable by first-order induction and more. In particular, it can capture the input/output relation of any program built from any of the DL programming constructs we have discussed. Since the first-order  $\mu$ -calculus also admits first-order quantification, it is easily seen to be as powerful as DL.

It was shown by [Park, 1976] that finiteness is not definable in the first-order  $\mu$ -calculus with the monotonicity restriction, but well-foundedness is. Thus this version of the  $\mu$ -calculus is independent of  $L_{\omega_1^{\text{ck}}\omega}$  (and hence of DL (r.e.)) in expressive power. Well-foundedness of a binary relation  $R$  can be written

$$\forall x (\mu X(x). \forall y (R(y, x) \rightarrow X(y))).$$

A more severe syntactic restriction on the binding operator  $\mu X$  is to allow its application only to formulas that are *syntactically continuous* in  $X$ ; that is, those formulas in which  $X$  does not occur free in the scope of any negation or any universal quantifier. It can be shown that this syntactic restriction implies semantic continuity, so the least fixpoint is the union of  $\emptyset$ ,  $\varphi(\emptyset)$ ,  $\varphi(\varphi(\emptyset))$ ,  $\dots$ . As shown in [Park, 1976], this version is strictly weaker than  $L_{\omega_1^{\text{ck}}\omega}$ .

In [Pratt, 1981a] and [Kozen, 1982; Kozen, 1983], propositional versions of the  $\mu$ -calculus were introduced. The latter version consists of propositional modal logic with a least fixpoint operator. It is the most powerful logic of its type, subsuming all known variants of PDL, game logic of [Parikh, 1983], various forms of temporal logic (see Section 14.2), and other seemingly stronger forms of the  $\mu$ -calculus ([Vardi and Wolper, 1986b]). In the following presentation we focus on this version, since it has gained fairly widespread acceptance; see [Kozen, 1984; Kozen and Parikh, 1983; Streett, 1985b; Streett and Emerson, 1984; Vardi and Wolper, 1986b; Walukiewicz, 1993; Walukiewicz, 1995; Walukiewicz, 2000; Stirling, 1992; Mader, 1997; Kaivola, 1997].

The language of the propositional  $\mu$ -calculus, also called the *modal  $\mu$ -calculus*, is syntactically simpler than PDL. It consists of the usual propositional constructs  $\rightarrow$  and  $\mathbf{0}$ , atomic modalities  $[a]$ , and the least fixpoint operator  $\mu$ . A greatest fixpoint operator dual to  $\mu$  can be defined:

$$\nu X. \varphi(X) \stackrel{\text{def}}{\iff} \neg \mu X. \neg \varphi(\neg X).$$

Variables are monadic, and the  $\mu$  operator may be applied only to syntactically monotone formulas. As discussed above, this ensures monotonicity of the corresponding set operator. The language is interpreted over Kripke frames in which atomic propositions are interpreted as sets of states and atomic programs are interpreted as binary relations on states.

The propositional  $\mu$ -calculus subsumes PDL. For example, the PDL formula  $\langle a^* \rangle \varphi$  for atomic  $a$  can be written  $\mu X. (\varphi \vee \langle a \rangle X)$ . The formula

$\mu X.\langle a \rangle[a]X$ , which expresses the existence of a forced win for the first player in a two-player game, and the formula  $\mu X.[a]X$ , which expresses well-foundedness and is equivalent to  $\mathbf{wf} a$  (see Section 7), are both inexpressible in PDL, as shown in [Streett, 1981; Kozen, 1981c]. [Niwinski, 1984] has shown that even with the addition of the **halt** construct, PDL is strictly less expressive than the  $\mu$ -calculus.

The propositional  $\mu$ -calculus satisfies a finite model theorem, as first shown in [Kozen, 1988]. Progressively better decidability results were obtained in [Kozen and Parikh, 1983; Vardi and Stockmeyer, 1985; Vardi, 1985b], culminating in a deterministic exponential-time algorithm of [Emerson and Jutla, 1988] based on an automata-theoretic lemma of [Safra, 1988]. Since the  $\mu$ -calculus subsumes PDL, it is *EXPTIME*-complete.

In [Kozen, 1982; Kozen, 1983], an axiomatization of the propositional  $\mu$ -calculus was proposed and conjectured to be complete. The axiomatization consists of the axioms and rules of propositional modal logic, plus the axiom

$$\varphi[X/\mu X.\varphi] \rightarrow \mu X.\varphi$$

and rule

$$\frac{\varphi[X/\psi] \rightarrow \psi}{\mu X.\varphi \rightarrow \psi}$$

for  $\mu$ . Completeness of this deductive system for a syntactically restricted subset of formulas was shown in [Kozen, 1982; Kozen, 1983]. Completeness for the full language was proved by [Walukiewicz, 1995; Walukiewicz, 2000]. This was quickly followed by simpler alternative proofs by [Ambler *et al.*, 1995; Bonsangue and Kwiatkowska, 1995; Hartonas, 1998]. [Bradfield, 1996] showed that the alternating  $\mu/\nu$  hierarchy (least/greatest fixpoints) is strict. An interesting open question is the complexity of *model checking*: does a given formula of the propositional  $\mu$ -calculus hold in a given state of a given Kripke frame? Although some progress has been made (see [Bhat and Cleaveland, 1996; Cleaveland, 1996; Emerson and Lei, 1986; Sokolsky and Smolka, 1994; Stirling and Walker, 1989]), it is still unknown whether this problem has a polynomial-time algorithm.

The propositional  $\mu$ -calculus has become a popular system for the specification and verification of properties of transition systems, where it has had some practical impact ([Steffen *et al.*, 1996]). Several recent papers on model checking work in this context; see [Bhat and Cleaveland, 1996; Cleaveland, 1996; Emerson and Lei, 1986; Sokolsky and Smolka, 1994; Stirling and Walker, 1989]. A comprehensive introduction can be found in [Stirling, 1992].

### 14.5 Kleene Algebra

*Kleene algebra* (KA) is the algebra of regular expressions. It is named for the mathematician S. C. Kleene (1909–1994), who among his many other



achievements invented regular expressions and proved their equivalence to finite automata in [Kleene, 1956].

Kleene algebra has appeared in various guises and under many names in relational algebra [Ng, 1984; Ng and Tarski, 1977], semantics and logics of programs [Kozen, 1981b; Pratt, 1988], automata and formal language theory [Kuich, 1987; Kuich and Salomaa, 1986], and the design and analysis of algorithms [Aho *et al.*, 1975; Tarjan, 1981; Mehlhorn, 1984; Iwano and Steiglitz, 1990; Kozen, 1991b]. As discussed in Section 13, Kleene algebra plays a prominent role in dynamic algebra as an algebraic model of program behavior.

Beginning with the monograph of [Conway, 1971], many authors have contributed over the years to the development of the algebraic theory; see [Backhouse, 1975; Krob, 1991; Kleene, 1956; Kuich and Salomaa, 1986; Sakarovitch, 1987; Kozen, 1990; Bloom and Ésik, 1992; Hopkins and Kozen, 1999]. See also [Kozen, 1996] for further references.

A *Kleene algebra* is an algebraic structure  $(K, +, \cdot, *, 0, 1)$  satisfying the axioms

$$\begin{aligned}
 \alpha + (\beta + \gamma) &= (\alpha + \beta) + \gamma \\
 \alpha + \beta &= \beta + \alpha \\
 \alpha + 0 &= \alpha + \alpha = \alpha \\
 \alpha(\beta\gamma) &= (\alpha\beta)\gamma \\
 1\alpha &= \alpha 1 = \alpha \\
 \alpha(\beta + \gamma) &= \alpha\beta + \alpha\gamma \\
 (\alpha + \beta)\gamma &= \alpha\gamma + \beta\gamma \\
 0\alpha &= \alpha 0 = 0 \\
 (37) \quad 1 + \alpha\alpha^* &= 1 + \alpha^*\alpha = \alpha^* \\
 (38) \quad \beta + \alpha\gamma \leq \gamma &\rightarrow \alpha^*\beta \leq \gamma \\
 (39) \quad \beta + \gamma\alpha \leq \gamma &\rightarrow \beta\alpha^* \leq \gamma
 \end{aligned}$$

where  $\leq$  refers to the natural partial order on  $K$ :

$$\alpha \leq \beta \stackrel{\text{def}}{\iff} \alpha + \beta = \beta.$$

In short, a KA is an idempotent semiring under  $+$ ,  $\cdot$ ,  $0$ ,  $1$  such that  $\alpha^*\beta$  is the least solution to  $\beta + \alpha x \leq x$  and  $\beta\alpha^*$  is the least solution to  $\beta + x\alpha \leq x$ . The axioms (37)–(39) say essentially that  $*$  behaves like the asterate operator on sets of strings or reflexive transitive closure on binary relations. This particular axiomatization is from [Kozen, 1991a; Kozen, 1994a], but there are other competing ones.

The axioms (38) and (39) correspond to the reflexive transitive closure rule (RTC) of PDL (Section 2.5). Instead, we might postulate the equivalent

axioms

$$(40) \quad \alpha\gamma \leq \gamma \rightarrow \alpha^*\gamma \leq \gamma$$

$$(41) \quad \gamma\alpha \leq \gamma \rightarrow \gamma\alpha^* \leq \gamma,$$

which correspond to the loop invariance rule (LI). The induction axiom (IND) is inexpressible in KA, since there is no negation.

A Kleene algebra is *\*-continuous* if it satisfies the infinitary condition

$$(42) \quad \alpha\beta^*\gamma = \sup_{n \geq 0} \alpha\beta^n\gamma$$

where

$$\beta^0 \stackrel{\text{def}}{=} 1 \quad \beta^{n+1} \stackrel{\text{def}}{=} \beta\beta^n$$

and where the supremum is with respect to the natural order  $\leq$ . We can think of (42) as a conjunction of the infinitely many axioms  $\alpha\beta^n\gamma \leq \alpha\beta^*\gamma$ ,  $n \geq 0$ , and the infinitary Horn formula

$$\left( \bigwedge_{n \geq 0} \alpha\beta^n\gamma \leq \delta \right) \rightarrow \alpha\beta^*\gamma \leq \delta.$$

In the presence of the other axioms, the *\*-continuity* condition (42) implies (38)–(41) and is strictly stronger in the sense that there exist Kleene algebras that are not *\*-continuous* [Kozen, 1990].

The fundamental motivating example of a Kleene algebra is the family of regular sets of strings over a finite alphabet, but other classes of structures share the same equational theory, notably the binary relations on a set. In fact it is the latter interpretation that makes Kleene algebra a suitable choice for modeling programs in dynamic algebras. Other more unusual interpretations are the min, + algebra used in shortest path algorithms (see [Aho *et al.*, 1975; Tarjan, 1981; Mehlhorn, 1984; Kozen, 1991b]) and KAs of convex polyhedra used in computational geometry as described in [Iwano and Steiglitz, 1990].

Axiomatization of the equational theory of the regular sets is a central question going back to the original paper of [Kleene, 1956]. A completeness theorem for relational algebras was given in an extended language by [Ng, 1984; Ng and Tarski, 1977]. Axiomatization is a central focus of the monograph of [Conway, 1971], but the bulk of his treatment is infinitary. [Redko, 1964] proved that there is no finite equational axiomatization. Schematic equational axiomatizations for the algebra of regular sets, necessarily representing infinitely many equations, have been given by [Krob, 1991] and [Bloom and Ésik, 1993]. [Salomaa, 1966] gave two finitary complete axiomatizations that are sound for the regular sets but not sound in general over other standard interpretations, including relational interpretations. The axiomatization given above is a finitary universal Horn axiomatization that

is sound and complete for the equational theory of standard relational and language-theoretic models, including the regular sets [Kozen, 1991a; Kozen, 1994a]. Other work on completeness appears in [Krob, 1991; Boffa, 1990; Boffa, 1995; Archangelsky, 1992].

The literature contains a bewildering array of inequivalent definitions of Kleene algebras and related algebraic structures; see [Conway, 1971; Pratt, 1988; Pratt, 1990; Kozen, 1981b; Kozen, 1991a; Aho *et al.*, 1975; Mehlhorn, 1984; Kuich, 1987; Kozen, 1994b]. As demonstrated in [Kozen, 1990], many of these are strongly related. One important property shared by most of them is closure under the formation of  $n \times n$  matrices. This was proved for the axiomatization above in [Kozen, 1991a; Kozen, 1994a], but the idea essentially goes back to [Kleene, 1956; Conway, 1971; Backhouse, 1975]. This result gives rise to an algebraic treatment of finite automata in which the automata are represented by their transition matrices.

The equational theory of Kleene algebra is *PSPACE*-complete [Stockmeyer and Meyer, 1973]; thus it is apparently less complex than PDL, which is *EXPTIME*-complete (Theorem 21), although the strict separation of the two complexity classes is still open.

### *Kleene Algebra with Tests*

From a practical standpoint, many simple program manipulations such as loop unwinding and basic safety analysis do not require the full power of PDL, but can be carried out in a purely equational subsystem using the axioms of Kleene algebra. However, *tests* are an essential ingredient, since they are needed to model conventional programming constructs such as conditionals and **while** loops and to handle assertions. This motivates the definition of the following variant of KA introduced in [Kozen, 1996; Kozen, 1997b].

A *Kleene algebra with tests* (KAT) is a Kleene algebra with an embedded Boolean subalgebra. Formally, it is a two-sorted algebra

$$(K, B, +, \cdot, *, \bar{\phantom{x}}, 0, 1)$$

such that

- $(K, +, \cdot, *, 0, 1)$  is a Kleene algebra
- $(B, +, \cdot, \bar{\phantom{x}}, 0, 1)$  is a Boolean algebra
- $B \subseteq K$ .

The unary negation operator  $\bar{\phantom{x}}$  is defined only on  $B$ . Elements of  $B$  are called *tests* and are written  $\varphi, \psi, \dots$ . Elements of  $K$  (including elements of  $B$ ) are written  $\alpha, \beta, \dots$ . In PDL, a test would be written  $\varphi?$ , but in KAT we dispense with the symbol  $?$ .

This deceptively concise definition actually carries a lot of information. The operators  $+$ ,  $\cdot$ ,  $0$ ,  $1$  each play two roles: applied to arbitrary elements of  $K$ , they refer to nondeterministic choice, composition, fail, and skip, respectively; and applied to tests, they take on the additional meaning of Boolean disjunction, conjunction, falsity, and truth, respectively. These two usages do not conflict—for example, sequential testing of two tests is the same as testing their conjunction—and their coexistence admits considerable economy of expression.

For applications in program verification, the standard interpretation would be a Kleene algebra of binary relations on a set and the Boolean algebra of subsets of the identity relation. One could also consider trace models, in which the Kleene elements are sets of traces (sequences of states) and the Boolean elements are sets of states (traces of length 0). As with KA, one can form the algebra  $n \times n$  matrices over a KAT  $(K, B)$ ; the Boolean elements of this structure are the diagonal matrices over  $B$ .

KAT can express conventional imperative programming constructs such as conditionals and while loops as in PDL. It can perform elementary program manipulation such as loop unwinding, constant propagation, and basic safety analysis in a purely equational manner. The applicability of KAT and related equational systems in practical program verification has been explored in [Cohen, 1994a; Cohen, 1994b; Cohen, 1994c; Kozen, 1996; Kozen and Patron, 2000].

There is a language-theoretic model that plays the same role in KAT that the regular sets play in KA, namely the algebra of regular sets of *guarded strings*, and a corresponding completeness result was obtained by [Kozen and Smith, 1996]. Moreover, KAT is complete for the equational theory of relational models, as shown in [Kozen and Smith, 1996]. Although less expressive than PDL, KAT is also apparently less difficult to decide: it is *PSPACE*-complete, the same as KA, as shown in [Cohen *et al.*, 1996].

In [Kozen, 1999a], it is shown that KAT subsumes propositional Hoare Logic in the following sense. The partial correctness assertion  $\{\varphi\} \alpha \{\psi\}$  is encoded in KAT as the equation  $\varphi \alpha \bar{\psi} = 0$ , or equivalently  $\varphi \alpha = \varphi \alpha \psi$ . If a rule

$$\frac{\{\varphi_1\} \alpha_1 \{\psi_1\}, \dots, \{\varphi_n\} \alpha_n \{\psi_n\}}{\{\varphi\} \alpha \{\psi\}}$$

is derivable in propositional Hoare Logic, then its translation, the universal Horn formula

$$\varphi_1 \alpha_1 \bar{\psi}_1 = 0 \wedge \dots \wedge \varphi_n \alpha_n \bar{\psi}_n = 0 \rightarrow \varphi \alpha \bar{\psi} = 0,$$

is a theorem of KAT. For example, the **while** rule of Hoare logic (see Section 2.6) becomes

$$\sigma \varphi \alpha \bar{\varphi} = 0 \rightarrow \varphi (\sigma \alpha)^* \bar{\sigma} \bar{\sigma} \varphi = 0.$$

More generally, all relationally valid Horn formulas of the form

$$\gamma_1 = 0 \wedge \cdots \wedge \gamma_n = 0 \quad \rightarrow \quad \alpha = \beta$$

are theorems of KAT [Kozen, 1999a].

Horn formulas are important from a practical standpoint. For example, commutativity conditions are used to model the idea that the execution of certain instructions does not affect the result of certain tests. In light of this, the complexity of the universal Horn theory of KA and KAT are of interest. There are both positive and negative results. It is shown in [Kozen, 1997c] that for a Horn formula  $\Phi \rightarrow \varphi$  over  $*$ -continuous Kleene algebras,

- if  $\Phi$  contains only commutativity conditions  $\alpha\beta = \beta\alpha$ , the universal Horn theory is  $\Pi_1^0$ -complete;
- if  $\Phi$  contains only monoid equations, the problem is  $\Pi_2^0$ -complete;
- for arbitrary finite sets of equations  $\Phi$ , the problem is  $\Pi_1^1$ -complete.

On the other hand, commutativity assumptions of the form  $\alpha\varphi = \varphi\alpha$ , where  $\varphi$  is a test, and assumptions of the form  $\gamma = 0$  can be eliminated without loss of efficiency, as shown in [Cohen, 1994a; Kozen and Smith, 1996]. Note that assumptions of this form are all we need to encode Hoare Logic as described above.

In typed Kleene algebra introduced in [Kozen, 1998; Kozen, 1999b], elements have types  $s \rightarrow t$ . This allows Kleene algebras of nonsquare matrices, among other applications. It is shown in [Kozen, 1999b] that Hoare Logic is subsumed by the type calculus of typed KA augmented with a typecast or coercion rule for tests. Thus Hoare-style reasoning with partial correctness assertions reduces to typechecking in a relatively simple type system.

### 14.6 Dynamic Algebra

Dynamic algebra provides an abstract algebraic framework that relates to PDL as Boolean algebra relates to propositional logic. A *dynamic algebra* is defined to be any two-sorted algebraic structure  $(K, B, \cdot)$ , where  $B = (B, \rightarrow, 0)$  is a Boolean algebra,  $K = (K, +, \cdot, *, 0, 1)$  is a Kleene algebra (see Section 14.5), and  $\cdot : K \times B \rightarrow B$  is a scalar multiplication satisfying algebraic constraints corresponding to the dual forms of the PDL axioms (Axioms 17). For example, all dynamic algebras satisfy the equations

$$\begin{aligned} (\alpha\beta) \cdot \varphi &= \alpha \cdot (\beta \cdot \varphi) \\ \alpha \cdot 0 &= 0 \\ 0 \cdot \varphi &= 0 \\ \alpha \cdot (\varphi \vee \psi) &= \alpha \cdot \varphi \vee \alpha \cdot \psi, \end{aligned}$$

which correspond to the PDL validities

$$\begin{aligned} \langle \alpha ; \beta \rangle \varphi &\leftrightarrow \langle \alpha \rangle \langle \beta \rangle \varphi \\ \langle \alpha \rangle \mathbf{0} &\leftrightarrow \mathbf{0} \\ \langle \mathbf{0} ? \rangle \varphi &\leftrightarrow \mathbf{0} \\ \langle \alpha \rangle (\varphi \vee \psi) &\leftrightarrow \langle \alpha \rangle \varphi \vee \langle \alpha \rangle \psi, \end{aligned}$$

respectively. The Boolean algebra  $B$  is an abstraction of the formulas of PDL and the Kleene algebra  $K$  is an abstraction of the programs.

The interaction of scalar multiplication with iteration can be axiomatized in a finitary or infinitary way. One can postulate

$$(43) \quad \alpha^* \cdot \varphi \leq \varphi \vee (\alpha^* \cdot (\neg \varphi \wedge (\alpha \cdot \varphi)))$$

corresponding to the diamond form of the PDL induction axiom (Axiom 17(viii)). Here  $\varphi \leq \psi$  in  $B$  iff  $\varphi \vee \psi = \psi$ . Alternatively, one can postulate the stronger axiom of *\*-continuity*:

$$(44) \quad \alpha^* \cdot \varphi = \sup_n (\alpha^n \cdot \varphi).$$

We can think of (44) as a conjunction of infinitely many axioms  $\alpha^n \cdot \varphi \leq \alpha^* \cdot \varphi$ ,  $n \geq 0$ , and the infinitary Horn formula

$$\left( \bigwedge_{n \geq 0} \alpha^n \cdot \varphi \leq \psi \right) \rightarrow \alpha^* \cdot \varphi \leq \psi.$$

In the presence of the other axioms, (44) implies (43) [Kozen, 1980b], and is strictly stronger in the sense that there are dynamic algebras that are not *\*-continuous* [Pratt, 1979a].

A standard Kripke frame  $\mathfrak{K} = (U, \mathfrak{m}_{\mathfrak{K}})$  of PDL gives rise to a *\*-continuous* dynamic algebra consisting of a Boolean algebra of subsets of  $U$  and a Kleene algebra of binary relations on  $U$ . Operators are interpreted as in PDL, including  $0$  as  $\mathbf{0}?$  (the empty program),  $1$  as  $\mathbf{1}?$  (the identity program), and  $\alpha \cdot \varphi$  as  $\langle \alpha \rangle \varphi$ . Nonstandard Kripke frames (see Section 3.2) also give rise to dynamic algebras, but not necessarily *\*-continuous* ones. A dynamic algebra is *separable* if any pair of distinct Kleene elements can be distinguished by some Boolean element; that is, if  $\alpha \neq \beta$ , then there exists  $\varphi \in B$  with  $\alpha \cdot \varphi \neq \beta \cdot \varphi$ .

Research directions in this area include the following.

- *Representation theory.* It is known that any separable dynamic algebra is isomorphic to some possibly nonstandard Kripke frame. Under certain conditions, “possibly nonstandard” can be replaced by “standard,” but not in general, even for *\*-continuous* algebras [Kozen, 1980b; Kozen, 1979c; Kozen, 1980a].

- *Algebraic methods in PDL.* The small model property (Theorem 15) and completeness (Theorem 18) for PDL can be established by purely algebraic considerations [Pratt, 1980a].
- *Comparative study of alternative axiomatizations of  $*$ .* For example, it is known that separable dynamic algebras can be distinguished from standard Kripke frames by a first-order formula, but even  $L_{\omega_1\omega}$  cannot distinguish the latter from  $*$ -continuous separable dynamic algebras [Kozen, 1981b].
- *Equational theory of dynamic algebras.* Many seemingly unrelated models of computation share the same equational theory, namely that of dynamic algebras [Pratt, 1979b; Pratt, 1979a].

In addition, many interesting questions arise from the algebraic viewpoint, and interesting connections with topology, classical algebra, and model theory have been made [Kozen, 1979b; Németi, 1980].

## 15 BIBLIOGRAPHICAL NOTES

Systematic program verification originated with the work of [Floyd, 1967] and [Hoare, 1969]. Hoare Logic was introduced in [Hoare, 1969]; see [Cousot, 1990; Apt, 1981; Apt and Olderog, 1991] for surveys.

The *digital abstraction*, the view of computers as state transformers that operate by performing a sequence of discrete and instantaneous primitive steps, can be attributed to [Turing, 1936]. Finite-state transition systems were defined formally by [McCulloch and Pitts, 1943]. State-transition semantics is based on this idea and is quite prevalent in early work on program semantics and verification; see [Hennessy and Plotkin, 1979]. The relational-algebraic approach taken here, in which programs are interpreted as binary input/output relations, was introduced in the context of DL by [Pratt, 1976].

The notions of partial and total correctness were present in the early work of [Hoare, 1969]. Regular programs were introduced by [Fischer and Ladner, 1979] in the context of PDL. The concept of nondeterminism was introduced in the original paper of [Turing, 1936], although he did not develop the idea. Nondeterminism was further developed by [Rabin and Scott, 1959] in the context of finite automata.

[Burstall, 1974] suggested using modal logic for reasoning about programs, but it was not until the work of [Pratt, 1976], prompted by a suggestion of R. Moore, that it was actually shown how to extend modal logic in a useful way by considering a separate modality for every program. The first research devoted to propositional reasoning about programs seems to be that of [Fischer and Ladner, 1977; Fischer and Ladner, 1979] on PDL. As

mentioned in the Preface, the general use of logical systems for reasoning about programs was suggested by [Engeler, 1967].

Other semantics besides Kripke semantics have been studied; see [Berman, 1979; Nishimura, 1979; Kozen, 1979b; Trnkova and Reiterman, 1980; Kozen, 1980b; Pratt, 1979b]. Modal logic has many applications and a vast literature; good introductions can be found in [Hughes and Cresswell, 1968; Chellas, 1980]. Alternative and iterative guarded commands were studied in [Gries, 1981]. Partial correctness assertions and the Hoare rules given in Section 2.6 were first formulated by [Hoare, 1969]. Regular expressions, on which the regular program operators are based, were introduced by [Kleene, 1956]. Their algebraic theory was further investigated by [Conway, 1971]. They were first applied in the context of DL by [Fischer and Ladner, 1977; Fischer and Ladner, 1979]. The axiomatization of PDL given in Axioms 17 was formulated by [Seegerberg, 1977]. Tests and converse were investigated by various authors; see [Peterson, 1978; Berman, 1978; Berman and Paterson, 1981; Streett, 1981; Streett, 1982; Vardi, 1985b]. The continuity of the diamond operator in the presence of reverse is due to [Trnkova and Reiterman, 1980].

The filtration argument and the small model property for PDL are due to [Fischer and Ladner, 1977; Fischer and Ladner, 1979]. Nonstandard Kripke frames for PDL were studied by [Berman, 1979; Berman, 1982], [Parikh, 1978a], [Pratt, 1979a; Pratt, 1980a], and [Kozen, 1979c; Kozen, 1979b; Kozen, 1980a; Kozen, 1980b; Kozen, 1981b].

The axiomatization of PDL used here (Axiom System 17) was introduced by [Seegerberg, 1977]. Completeness was shown independently by [Gabbay, 1977] and [Parikh, 1978a]. A short and easy-to-follow proof is given in [Kozen and Parikh, 1981]. Completeness is also treated in [Pratt, 1978; Pratt, 1980a; Berman, 1979; Nishimura, 1979; Kozen, 1981a].

The exponential-time lower bound for PDL was established by [Fischer and Ladner, 1977; Fischer and Ladner, 1979] by showing how PDL formulas can encode computations of linear-space-bounded alternating Turing machines.

Deterministic exponential-time algorithms were first given in [Pratt, 1978; Pratt, 1979b; Pratt, 1980b].

Theorem 24 showing that the problem of deciding whether  $\Gamma \models \psi$ , where  $\Gamma$  is a fixed r.e. set of PDL formulas, is  $\Pi_1^1$ -complete is due to [Meyer *et al.*, 1981].

The computational difficulty of the validity problem for nonregular PDL and the borderline between the decidable and undecidable were discussed in [Harel *et al.*, 1983]. The fact that any nonregular program adds expressive power to PDL, Theorem 25, first appeared explicitly in [Harel and Singerman, 1996]. Theorem 26 on the undecidability of context-free PDL was observed by [Ladner, 1977].



Theorems 27 and 28 are from [Harel *et al.*, 1983]. An alternative proof of Theorem 28 using tiling is supplied in [Harel, 1985]; see [Harel *et al.*, 2000]. The existence of a primitive recursive one-letter extension of PDL that is undecidable was shown already in [Harel *et al.*, 1983], but undecidability for the particular case of  $a^{2^*}$ , Theorem 29, is from [Harel and Paterson, 1984]. Theorem 30 is from [Harel and Singerman, 1996].

As to decidable extensions, Theorem 31 was proved in [Koren and Pnueli, 1983]. The more general results of Section 6.2, namely Theorems 32, 33, and 34, are from [Harel and Raz, 1993], as is the notion of a simple-minded PDA. The decidability of emptiness for pushdown and stack automata on trees that is needed for the proofs of these is from [Harel and Raz, 1994]. A better bound on the complexity of the emptiness results can be found in [Peng and Iyer, 1995].

A sufficient condition for PDL with the addition of a program over a single letter alphabet not to have the finite model property is given in [Harel and Singerman, 1996].

Completeness and exponential time decidability for DPDL, Theorem 40 and the upper bound of Theorem 41, are proved in [Ben-Ari *et al.*, 1982] and [Valiev, 1980]. The lower bound of Theorem 41 is from [Parikh, 1981]. Theorems 43 and 44 on SDPDL are from [Halpern and Reif, 1981; Halpern and Reif, 1983].

That tests add to the power of PDL is proved in [Berman and Paterson, 1981]. It is also known that the test-depth hierarchy is strict [Berman, 1978; Peterson, 1978] and that rich-test PDL is strictly more expressive than poor-test PDL [Peterson, 1978; Berman, 1978; Berman and Paterson, 1981]. These results also hold for SDPDL.

The results on programs as automata (Theorems 45 and 46) appear in [Pratt, 1981b]. Alternative proofs are given in [Harel and Sherman, 1985]; see [Harel *et al.*, 2000]. In recent years, the development of the automata-theoretic approach to logics of programs has prompted renewed inquiry into the complexity of automata on infinite objects, with considerable success. See [Courcoubetis and Yannakakis, 1988; Emerson, 1985; Emerson and Jutla, 1988; Emerson and Sistla, 1984; Manna and Pnueli, 1987; Muller *et al.*, 1988; Pecuchet, 1986; Safra, 1988; Sistla *et al.*, 1987; Streett, 1982; Vardi, 1985a; Vardi, 1985b; Vardi, 1987; Vardi and Stockmeyer, 1985; Vardi and Wolper, 1986b; Vardi and Wolper, 1986a; Arnold, 1997a; Arnold, 1997b]; and [Thomas, 1997]. Especially noteworthy in this area is the result of [Safra, 1988] involving the complexity of converting a nondeterministic automaton on infinite strings into an equivalent deterministic one. This result has already had a significant impact on the complexity of decision procedures for several logics of programs; see [Courcoubetis and Yannakakis, 1988; Emerson and Jutla, 1988; Emerson and Jutla, 1989]; and [Safra, 1988].

Intersection of programs was studied in [Harel *et al.*, 1982a]. That the axioms for converse yield completeness for CPDL is proved in [Parikh, 1978a].

The complexity of PDL with converse and various forms of well-foundedness constructs is studied in [Vardi, 1985b]. Many authors have studied logics with a least-fixpoint operator, both on the propositional and first-order levels ([Scott and de Bakker, 1969; Hitchcock and Park, 1972; Park, 1976; Pratt, 1981a; Kozen, 1982; Kozen, 1983; Kozen, 1988; Kozen and Parikh, 1983; Niwinski, 1984; Streett, 1985a; Vardi and Stockmeyer, 1985]). The version of the propositional  $\mu$ -calculus presented here was introduced in [Kozen, 1982; Kozen, 1983].

That the propositional  $\mu$ -calculus is strictly more expressive than PDL with **wf** was shown in [Niwinski, 1984] and [Streett, 1985a]. That this logic is strictly more expressive than PDL with **halt** was shown in [Harel and Sherman, 1982]. That this logic is strictly more expressive than PDL was shown in [Streett, 1981].

The **wf** construct (actually its complement, **repeat**) is investigated in [Streett, 1981; Streett, 1982], in which Theorems 48 (which is actually due to Pratt) and 50–52 are proved. The **halt** construct (actually its complement, **loop**) was introduced in [Harel and Pratt, 1978] and Theorem 49 is from [Harel and Sherman, 1982]. Finite model properties for the logics LPDL, RPD, CLPDL, CRPDL, and the propositional  $\mu$ -calculus were established in [Streett, 1981; Streett, 1982] and [Kozen, 1988]. Decidability results were obtained in [Streett, 1981; Streett, 1982; Kozen and Parikh, 1983; Vardi and Stockmeyer, 1985]; and [Vardi, 1985b]. Deterministic exponential-time completeness was established in [Emerson and Jutla, 1988] and [Safra, 1988]. For the strongest variant, CRPDL, exponential-time decidability follows from [Vardi, 1998b].

Concurrent PDL is defined and studied in [Peleg, 1987b]. Additional versions of this logic, which employ various mechanisms for communication among the concurrent parts of a program, are considered in [Peleg, 1987c; Peleg, 1987a]. These papers contain many results concerning expressive power, decidability and undecidability for concurrent PDL with communication.

Other work on PDL not described here includes work on nonstandard models, studied in [Berman, 1979; Berman, 1982] and [Parikh, 1981]; PDL with Boolean assignments, studied in [Abrahamson, 1980]; and restricted forms of the consequence problem, studied in [Parikh, 1981].

First-order DL was defined in [Harel *et al.*, 1977], where it was also first named Dynamic Logic. That paper was carried out as a direct continuation of the original work of [Pratt, 1976].

Many variants of DL were defined in [Harel, 1979]. In particular, DL(bst)k is very close to the context-free Dynamic Logic investigated there.

Uninterpreted reasoning in the form of program schematology has been a common activity ever since the work of [Iarov, 1960]. It was given con-

siderable impetus by the work of [Luckham *et al.*, 1970] and [Paterson and Hewitt, 1970]; see also [Greibach, 1975]. The study of the correctness of interpreted programs goes back to the work of Turing and von Neumann, but seems to have become a well-defined area of research following [Floyd, 1967], [Hoare, 1969] and [Manna, 1974].

Embedding logics of programs in  $L_{\omega_1\omega}$  is based on observations of [Engeler, 1967]. Theorem 57 is from [Meyer and Parikh, 1981]. Theorem 60 is from [Harel, 1979] (see also [Harel, 1984] and [Harel and Kozen, 1984]); it is similar to the expressiveness result of [Cook, 1978]. Theorem 61 and Corollary 62 are from [Harel and Kozen, 1984].

Arithmetical structures were first defined by [Moschovakis, 1974] under the name *acceptable structures*. In the context of logics of programs, they were reintroduced and studied in [Harel, 1979].

The  $\Pi_1^1$ -completeness of DL was first proved by Meyer, and Theorem 63 appears in [Harel *et al.*, 1977]. An alternative proof is given in [Harel, 1985]; see [Harel *et al.*, 2000]. Theorem 65 is from [Meyer and Halpern, 1982]. That the fragment of DL considered in Theorem 66 is not r.e., was proved by [Pratt, 1976]. Theorem 67 follows from [Harel and Kozen, 1984].

The name “spectral complexity” was proposed by [Tiuryn, 1986], although the main ideas and many results concerning this notion were already present in [Tiuryn and Urzyczyn, 1983] (see [Tiuryn and Urzyczyn, 1988] for the full version). This notion is an instance of the so-called *second-order spectrum* of a formula. First-order spectra were investigated by [Sholz, 1952], from which originates the well known *Spectralproblem*. The reader can find more about this problem and related results in the survey paper by [Börger, 1984]. The notion of a natural chain is from [Urzyczyn, 1983]. The results presented here are from [Tiuryn and Urzyczyn, 1983; Tiuryn and Urzyczyn, 1988]. A result similar to Theorem 69 in the area of finite model theory was obtained by [Sazonov, 1980] and independently by [Gurevich, 1983]. Higher-order stacks were introduced in [Engelfriet, 1983] to study complexity classes. Higher-order arrays and stacks in DL were considered by [Tiuryn, 1986], where a strict hierarchy within the class of elementary recursive sets was established. The main tool used in the proof of the strictness of this hierarchy is a generalization of Cook’s auxiliary pushdown automata theorem for higher-order stacks, which is due to [Kowalczyk *et al.*, 1987].

[Meyer and Halpern, 1982] showed completeness for termination assertions (Theorem 71). Infinitary completeness for DL (Theorem 72) is based upon a similar result for Algorithmic Logic (see Section 13.1) by [Mirkowska, 1971]. The proof sketch presented in [Harel *et al.*, 2000] is an adaptation of Henkin’s proof for  $L_{\omega_1\omega}$  appearing in [Keisler, 1971].

The notion of relative completeness and Theorem 73 are due to [Cook, 1978]. The notion of arithmetical completeness and Theorem 75 is from [Harel, 1979].

The use of invariants to prove partial correctness and of well-founded sets to prove termination are due to [Floyd, 1967]. An excellent survey of such methods and the corresponding completeness results appears in [Apt, 1981].

Some contrasting negative results are contained in [Clarke, 1979], [Lipton, 1977], and [Wand, 1978].

Many of the results on relative expressiveness presented herein answer questions posed in [Harel, 1979]. Similar uninterpreted research, comparing the expressive power of classes of programs (but detached from any surrounding logic) has taken place under the name *comparative schematology* quite extensively ever since [Ianov, 1960]; see [Greibach, 1975] and [Manna, 1974].

Theorems 76, 79 and 83(i) result as an application of the so-called *spectral theorem*, which connects expressive power of logics with complexity classes. This theorem was obtained by [Tiuryn and Urzyczyn, 1983; Tiuryn and Urzyczyn, 1984; Tiuryn and Urzyczyn, 1988]. A simplified framework for this approach and a statement of this theorem together with a proof is given in [Harel *et al.*, 2000].

Theorem 78 appears in [Berman *et al.*, 1982] and was proved independently in [Stolboushkin and Taitlin, 1983]. An alternative proof is given in [Tiuryn, 1989]. These results extend in a substantial way an earlier and much simpler result for the case of regular programs without equality in the vocabulary, which appears in [Halpern, 1981]. A simpler proof of the special case of the quantifier-free fragment of the logic of regular programs appears in [Meyer and Winklmann, 1982]. Theorem 79 is from [Tiuryn and Urzyczyn, 1984].

Theorem 80 is from [Stolboushkin, 1983]. The proof, as in the case of regular programs (see [Stolboushkin and Taitlin, 1983]), uses Adian's result from group theory ([Adian, 1979]). Results on the expressive power of DL with deterministic **while** programs and a Boolean stack can be found in [Stolboushkin, 1983; Kfoury, 1985]. Theorem 81 is from [Tiuryn and Urzyczyn, 1983; Tiuryn and Urzyczyn, 1988].

[Erimbetov, 1981; Tiuryn, 1981b; Tiuryn, 1984; Kfoury, 1983; Kfoury and Stolboushkin, 1997] contain results on the expressive power of DL over programming languages with bounded memory. [Erimbetov, 1981] shows that  $DL(dreg) < DL(dstk)$ . The main proof technique is pebble games on finite trees.

Theorem 83 is from [Urzyczyn, 1987]. There is a different proof of this result, using Adian structures, which appears in [Stolboushkin, 1989]. Theorem 77 is from [Urzyczyn, 1988], which also studies programs with Boolean arrays.

Wildcard assignments were considered in [Harel *et al.*, 1977] under the name *nondeterministic assignments*. Theorem 84 is from [Meyer and Winklmann, 1982]. Theorem 85 is from [Meyer and Parikh, 1981].

In our exposition of the comparison of the expressive power of logics, we have made the assumption that programs use only quantifier-free first-order tests. It follows from the results of [Urzyczyn, 1986] that allowing full first-order tests in many cases results in increased expressive power. [Urzyczyn, 1986] also proves that adding array assignments to nondeterministic r.e. programs increases the expressive power of the logic. This should be contrasted with the result of [Meyer and Tiuryn, 1981; Meyer and Tiuryn, 1984] to the effect that for deterministic r.e. programs, array assignments do not increase expressive power.

[Makowsky, 1980] considers a weaker notion of equivalence between logics common in investigations in abstract model theory, whereby models are extended with interpretations for additional predicate symbols. With this notion it is shown in [Makowsky, 1980] that most of the versions of logics of programs treated here become equivalent.

Algorithmic logic was introduced by [Salwicki, 1970]. [Mirkowska, 1980; Mirkowska, 1981a; Mirkowska, 1981b] extended **AL** to allow nondeterministic **while** programs and studied the operators  $\nabla$  and  $\Delta$ . Complete infinitary deductive systems for propositional and first-order versions were given by [Mirkowska, 1980; Mirkowska, 1981a; Mirkowska, 1981b] using the algebraic methods of [Rasiowa and Sikorski, 1963]. Surveys of early work in **AL** can be found in [Banachowski *et al.*, 1977; Salwicki, 1977]. [Constable, 1977; Constable and O'Donnell, 1978; Goldblatt, 1982] presented logics similar to **AL** and **DL** for reasoning about deterministic **while** programs.

Nonstandard Dynamic Logic was introduced by [Németi, 1981] and [Andréka *et al.*, 1982a; Andréka *et al.*, 1982b] and studied in [Csirmaz, 1985]. See [Makowsky and Sain, 1986] for more information and further references.

The **halt** construct (actually its complement, **loop**) was introduced in [Harel and Pratt, 1978], and the **wf** construct (actually its complement, **repeat**) was investigated for PDL in [Streett, 1981; Streett, 1982]. Theorem 86 is from [Meyer and Winklmann, 1982], Theorem 87 is from [Harel and Peleg, 1985], Theorem 88 is from [Harel, 1984], and the axiomatizations of LDL and PDL are discussed in [Harel, 1979; Harel, 1984].

Dynamic algebra was introduced in [Kozen, 1980b] and [Pratt, 1979b] and studied by numerous authors; see [Kozen, 1979c; Kozen, 1979b; Kozen, 1980a; Kozen, 1981b; Pratt, 1979a; Pratt, 1980a; Pratt, 1988; Németi, 1980; Trnkova and Reiterman, 1980]. A survey of the main results appears in [Kozen, 1979a].

The PhD thesis [Ramshaw, 1981] contains an engaging introduction to the subject of probabilistic semantics and verification. [Kozen, 1981d] provided a formal semantics for probabilistic programs. The logic  $Pr(\text{DL})$  was presented in [Feldman and Harel, 1984], along with a deductive system that is complete for Kozen's semantics relative to an extension of first-order analysis. Various propositional versions of probabilistic DL have been proposed

in [Reif, 1980; Makowsky and Tiomkin, 1980; Feldman, 1984; Parikh and Mahoney, 1983; Kozen, 1985]. The temporal approach to probabilistic verification has been studied in [Lehmann and Shelah, 1982; Hart *et al.*, 1982; Courcoubetis and Yannakakis, 1988; Vardi, 1985a]. Interest in the subject of probabilistic verification has undergone a recent revival; see [Morgan *et al.*, 1999; Segala and Lynch, 1994; Hansson and Jonsson, 1994; Jou and Smolka, 1990; Baier and Kwiatkowska, 1998; Huth and Kwiatkowska, 1997; Blute *et al.*, 1997].

Concurrent DL is defined and studied in [Peleg, 1987b]. Additional versions of this logic, which employ various mechanisms for communication among the concurrent parts of a program, are also considered in [Peleg, 1987c; Peleg, 1987a].

David Harel

*The Weizmann Institute of Science, Rehovot, Israel*

Dexter Kozen

*Cornell University, Ithaca, New York*

Jerzy Tiuryn

*The University of Warsaw, Warsaw, Poland*

## BIBLIOGRAPHY

- [Abrahamson, 1980] K. Abrahamson. *Decidability and expressiveness of logics of processes*. PhD thesis, Univ. of Washington, 1980.
- [Adian, 1979] S. I. Adian. *The Burnside Problem and Identities in Groups*. Springer-Verlag, 1979.
- [Aho *et al.*, 1975] A. V. Aho, J. E. Hopcroft, and J. D. Ullman. *The Design and Analysis of Computer Algorithms*. Addison-Wesley, Reading, Mass., 1975.
- [Ambler *et al.*, 1995] S. Ambler, M. Kwiatkowska, and N. Measor. Duality and the completeness of the modal  $\mu$ -calculus. *Theor. Comput. Sci.*, 151(1):3–27, November 1995.
- [Andréka *et al.*, 1982a] H. Andréka, I. Németi, and I. Sain. A complete logic for reasoning about programs via nonstandard model theory, part I. *Theor. Comput. Sci.*, 17:193–212, 1982.
- [Andréka *et al.*, 1982b] H. Andréka, I. Németi, and I. Sain. A complete logic for reasoning about programs via nonstandard model theory, part II. *Theor. Comput. Sci.*, 17:259–278, 1982.
- [Apt and Olderog, 1991] K. R. Apt and E.-R. Olderog. *Verification of Sequential and Concurrent Programs*. Springer-Verlag, 1991.
- [Apt and Plotkin, 1986] K. R. Apt and G. Plotkin. Countable nondeterminism and random assignment. *J. Assoc. Comput. Mach.*, 33:724–767, 1986.
- [Apt, 1981] K. R. Apt. Ten years of Hoare's logic: a survey—part I. *ACM Trans. Programming Languages and Systems*, 3:431–483, 1981.
- [Archangelsky, 1992] K. V. Archangelsky. A new finite complete solvable quasiequational calculus for algebra of regular languages. Manuscript, Kiev State University, 1992.
- [Arnold, 1997a] A. Arnold. An initial semantics for the  $\mu$ -calculus on trees and Rabin's complementation lemma. Technical report, University of Bordeaux, 1997.
- [Arnold, 1997b] A. Arnold. The  $\mu$ -calculus on trees and Rabin's complementation theorem. Technical report, University of Bordeaux, 1997.

- [Backhouse, 1975] R. C. Backhouse. *Closure Algorithms and the Star-Height Problem of Regular Languages*. PhD thesis, Imperial College, London, U.K., 1975.
- [Backhouse, 1986] R. C. Backhouse. *Program Construction and Verification*. Prentice-Hall, 1986.
- [Baier and Kwiatkowska, 1998] C. Baier and M. Kwiatkowska. On the verification of qualitative properties of probabilistic processes under fairness constraints. *Information Processing Letters*, 66(2):71–79, April 1998.
- [Banachowski et al., 1977] L. Banachowski, A. Kreczmar, G. Mirkowska, H. Rasiowa, and A. Salwicki. An introduction to algorithmic logic: metamathematical investigations in the theory of programs. In Mazurkiewicz and Pawlak, editors, *Math. Found. Comput. Sci.*, pages 7–99. Banach Center, Warsaw, 1977.
- [Ben-Ari et al., 1982] M. Ben-Ari, J. Y. Halpern, and A. Pnueli. Deterministic propositional dynamic logic: finite models, complexity and completeness. *J. Comput. Syst. Sci.*, 25:402–417, 1982.
- [Berman and Paterson, 1981] F. Berman and M. Paterson. Propositional dynamic logic is weaker without tests. *Theor. Comput. Sci.*, 16:321–328, 1981.
- [Berman et al., 1982] P. Berman, J. Y. Halpern, and J. Tiuryn. On the power of non-determinism in dynamic logic. In Nielsen and Schmidt, editors, *Proc 9th Colloq. Automata Lang. Prog.*, volume 140 of *Lect. Notes in Comput. Sci.*, pages 48–60. Springer-Verlag, 1982.
- [Berman, 1978] F. Berman. Expressiveness hierarchy for PDL with rich tests. Technical Report 78-11-01, Comput. Sci. Dept., Univ. of Washington, 1978.
- [Berman, 1979] F. Berman. A completeness technique for  $D$ -axiomatizable semantics. In *Proc. 11th Symp. Theory of Comput.*, pages 160–166. ACM, 1979.
- [Berman, 1982] F. Berman. Semantics of looping programs in propositional dynamic logic. *Math. Syst. Theory*, 15:285–294, 1982.
- [Bhat and Cleaveland, 1996] G. Bhat and R. Cleaveland. Efficient local model checking for fragments of the modal  $\mu$ -calculus. In T. Margaria and B. Steffen, editors, *Proc. Second Int. Workshop Tools and Algorithms for the Construction and Analysis of Systems (TACAS'96)*, volume 1055 of *Lect. Notes in Comput. Sci.*, pages 107–112. Springer-Verlag, March 1996.
- [Bloom and Ésik, 1992] S. L. Bloom and Z. Ésik. Program correctness and matricial iteration theories. In *Proc. 7th Int. Conf. Mathematical Foundations of Programming Semantics*, volume 598 of *Lecture Notes in Computer Science*, pages 457–476. Springer-Verlag, 1992.
- [Bloom and Ésik, 1993] S. L. Bloom and Z. Ésik. Equational axioms for regular sets. *Math. Struct. Comput. Sci.*, 3:1–24, 1993.
- [Blute et al., 1997] R. Blute, J. Desharnais, A. Edalat, and P. Panangaden. Bisimulation for labeled Markov processes. In *Proc. 12th Symp. Logic in Comput. Sci.*, pages 149–158. IEEE, 1997.
- [Boffa, 1990] M. Boffa. Une remarque sur les systèmes complets d'identités rationnelles. *Informatique Théorique et Applications/Theoretical Informatics and Applications*, 24(4):419–423, 1990.
- [Boffa, 1995] Maurice Boffa. Une condition impliquant toutes les identités rationnelles. *Informatique Théorique et Applications/Theoretical Informatics and Applications*, 29(6):515–518, 1995.
- [Bonsangue and Kwiatkowska, 1995] M. Bonsangue and M. Kwiatkowska. Reinterpreting the modal  $\mu$ -calculus. In A. Ponse, M. de Rijke, and Y. Venema, editors, *Modal Logic and Process Algebra*, pages 65–83. CSLI Lecture Notes, August 1995.
- [Börger, 1984] E. Börger. Spectral problem and completeness of logical decision problems. In G. Hasenjaeger E. Börger and D. Rödding, editors, *Logic and Machines: Decision Problems and Complexity, Proceedings*, volume 171 of *Lect. Notes in Comput. Sci.*, pages 333–356. Springer-Verlag, 1984.
- [Bradfield, 1996] J. C. Bradfield. The modal  $\mu$ -calculus alternation hierarchy is strict. In U. Montanari and V. Sassone, editors, *Proc. CONCUR'96*, volume 1119 of *Lect. Notes in Comput. Sci.*, pages 233–246. Springer, 1996.
- [Burstall, 1974] R. M. Burstall. Program proving as hand simulation with a little induction. *Information Processing*, pages 308–312, 1974.

- [Chandra *et al.*, 1981] A. Chandra, D. Kozen, and L. Stockmeyer. Alternation. *J. Assoc. Comput. Mach.*, 28(1):114–133, 1981.
- [Chellas, 1980] B. F. Chellas. *Modal Logic: An Introduction*. Cambridge University Press, 1980.
- [Clarke, 1979] E. M. Clarke. Programming language constructs for which it is impossible to obtain good Hoare axiom systems. *J. Assoc. Comput. Mach.*, 26:129–147, 1979.
- [Cleaveland, 1996] R. Cleaveland. Efficient model checking via the equational  $\mu$ -calculus. In *Proc. 11th Symp. Logic in Comput. Sci.*, pages 304–312. IEEE, July 1996.
- [Cohen *et al.*, 1996] Ernie Cohen, Dexter Kozen, and Frederick Smith. The complexity of Kleene algebra with tests. Technical Report 96-1598, Computer Science Department, Cornell University, July 1996.
- [Cohen, 1994a] E. Cohen. Hypotheses in Kleene algebra. Available as <ftp://ftp.telcordia.com/pub/ernie/research/homepage.html>, April 1994.
- [Cohen, 1994b] E. Cohen. Lazy caching. Available as <ftp://ftp.telcordia.com/pub/ernie/research/homepage.html>, 1994.
- [Cohen, 1994c] E. Cohen. Using Kleene algebra to reason about concurrency control. Available as <ftp://ftp.telcordia.com/pub/ernie/research/homepage.html>, 1994.
- [Constable and O'Donnell, 1978] R. L. Constable and M. O'Donnell. *A Programming Logic*. Winthrop, 1978.
- [Constable, 1977] R. L. Constable. On the theory of programming logics. In *Proc. 9th Symp. Theory of Comput.*, pages 269–285. ACM, May 1977.
- [Conway, 1971] J. H. Conway. *Regular Algebra and Finite Machines*. Chapman and Hall, London, 1971.
- [Cook, 1978] S. A. Cook. Soundness and completeness of an axiom system for program verification. *SIAM J. Comput.*, 7:70–80, 1978.
- [Courcoubetis and Yannakakis, 1988] C. Courcoubetis and M. Yannakakis. Verifying temporal properties of finite-state probabilistic programs. In *Proc. 29th Symp. Foundations of Comput. Sci.*, pages 338–345. IEEE, October 1988.
- [Cousot, 1990] P. Cousot. Methods and logics for proving programs. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 841–993. Elsevier, Amsterdam, 1990.
- [Csirmaz, 1985] L. Csirmaz. A completeness theorem for dynamic logic. *Notre Dame J. Formal Logic*, 26:51–60, 1985.
- [Davis *et al.*, 1994] M. D. Davis, R. Sigal, and E. J. Weyuker. *Computability, Complexity, and Languages: Fundamentals of Theoretical Computer Science*. Academic Press, 1994.
- [de Bakker, 1980] J. de Bakker. *Mathematical Theory of Program Correctness*. Prentice-Hall, 1980.
- [Emerson and Halpern, 1985] E. A. Emerson and J. Y. Halpern. Decision procedures and expressiveness in the temporal logic of branching time. *J. Comput. Syst. Sci.*, 30(1):1–24, 1985.
- [Emerson and Halpern, 1986] E. A. Emerson and J. Y. Halpern. “Sometimes” and “not never” revisited: on branching vs. linear time temporal logic. *J. ACM*, 33(1):151–178, 1986.
- [Emerson and Jutla, 1988] E. A. Emerson and C. Jutla. The complexity of tree automata and logics of programs. In *Proc. 29th Symp. Foundations of Comput. Sci.*, pages 328–337. IEEE, October 1988.
- [Emerson and Jutla, 1989] E. A. Emerson and C. Jutla. On simultaneously determinizing and complementing  $\omega$ -automata. In *Proc. 4th Symp. Logic in Comput. Sci.* IEEE, June 1989.
- [Emerson and Lei, 1986] E. A. Emerson and C.-L. Lei. Efficient model checking in fragments of the propositional  $\mu$ -calculus. In *Proc. 1st Symp. Logic in Comput. Sci.*, pages 267–278. IEEE, June 1986.
- [Emerson and Lei, 1987] E. A. Emerson and C. L. Lei. Modalities for model checking: branching time strikes back. *Sci. Comput. Programming*, 8:275–306, 1987.
- [Emerson and Sistla, 1984] E. A. Emerson and P. A. Sistla. Deciding full branching-time logic. *Infor. and Control*, 61:175–201, 1984.



- [Emerson, 1985] E. A. Emerson. Automata, tableaux, and temporal logics. In R. Parikh, editor, *Proc. Workshop on Logics of Programs*, volume 193 of *Lect. Notes in Comput. Sci.*, pages 79–88. Springer-Verlag, 1985.
- [Emerson, 1990] E. A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of theoretical computer science*, volume B: formal models and semantics, pages 995–1072. Elsevier, 1990.
- [Engeler, 1967] E. Engeler. Algorithmic properties of structures. *Math. Syst. Theory*, 1:183–195, 1967.
- [Engelfriet, 1983] J. Engelfriet. Iterated pushdown automata and complexity classes. In *Proceedings of the Fifteenth Annual ACM Symposium on Theory of Computing*, pages 365–373, Boston, Massachusetts, 1983.
- [Erimbetov, 1981] M. M. Erimbetov. On the expressive power of programming logics. In *Proc. Alma-Ata Conf. Research in Theoretical Programming*, pages 49–68, 1981. In Russian.
- [Feldman and Harel, 1984] Y. A. Feldman and D. Harel. A probabilistic dynamic logic. *J. Comput. Syst. Sci.*, 28:193–215, 1984.
- [Feldman, 1984] Y. A. Feldman. A decidable propositional dynamic logic with explicit probabilities. *Infor. and Control*, 63:11–38, 1984.
- [Fischer and Ladner, 1977] M. J. Fischer and R. E. Ladner. Propositional modal logic of programs. In *Proc. 9th Symp. Theory of Comput.*, pages 286–294. ACM, 1977.
- [Fischer and Ladner, 1979] M. J. Fischer and R. E. Ladner. Propositional dynamic logic of regular programs. *J. Comput. Syst. Sci.*, 18(2):194–211, 1979.
- [Floyd, 1967] R. W. Floyd. Assigning meanings to programs. In *Proc. Symp. Appl. Math.*, volume 19, pages 19–31. AMS, 1967.
- [Friedman, 1971] H. Friedman. Algorithmic procedures, generalized Turing algorithms, and elementary recursion theory. In Gandy and Yates, editors, *Logic Colloq. 1969*, pages 361–390. North-Holland, 1971.
- [Gabbay *et al.*, 1980] D. Gabbay, A. Pnueli, S. Shelah, and J. Stavi. On the temporal analysis of fairness. In *Proc. 7th Symp. Princip. Prog. Lang.*, pages 163–173. ACM, 1980.
- [Gabbay *et al.*, 1994] D. Gabbay, I. Hodkinson, and M. Reynolds. *Temporal Logic: Mathematical Foundations and Computational Aspects*. Oxford University Press, 1994.
- [Gabbay, 1977] D. Gabbay. Axiomatizations of logics of programs. Unpublished, 1977.
- [Goldblatt, 1982] R. Goldblatt. *Axiomatizing the Logic of Computer Programming*, volume 130 of *Lect. Notes in Comput. Sci.* Springer-Verlag, 1982.
- [Goldblatt, 1987] R. Goldblatt. Logics of time and computation. Technical Report Lect. Notes 7, Center for the Study of Language and Information, Stanford Univ., 1987.
- [Greibach, 1975] S. Greibach. *Theory of Program Structures: Schemes, Semantics, Verification*, volume 36 of *Lecture Notes in Computer Science*. Springer Verlag, 1975.
- [Gries, 1981] D. Gries. *The Science of Programming*. Springer-Verlag, 1981.
- [Gurevich, 1983] Yu. Gurevich. Algebras of feasible functions. In *24-th IEEE Annual Symposium on Foundations of Computer Science*, pages 210–214, 1983.
- [Halpern and Reif, 1981] J. Y. Halpern and J. H. Reif. The propositional dynamic logic of deterministic, well-structured programs. In *Proc. 22nd Symp. Found. Comput. Sci.*, pages 322–334. IEEE, 1981.
- [Halpern and Reif, 1983] J. Y. Halpern and J. H. Reif. The propositional dynamic logic of deterministic, well-structured programs. *Theor. Comput. Sci.*, 27:127–165, 1983.
- [Halpern, 1981] J. Y. Halpern. On the expressive power of dynamic logic II. Technical Report TM-204, MIT/LCS, 1981.
- [Halpern, 1982] J. Y. Halpern. Deterministic process logic is elementary. In *Proc. 23rd Symp. Found. Comput. Sci.*, pages 204–216. IEEE, 1982.
- [Halpern, 1983] J. Y. Halpern. Deterministic process logic is elementary. *Infor. and Control*, 57(1):56–89, 1983.
- [Hansson and Jonsson, 1994] H. Hansson and B. Jonsson. A logic for reasoning about time and probability. *Formal Aspects of Computing*, 6:512–535, 1994.
- [Harel and Kozen, 1984] D. Harel and D. Kozen. A programming language for the inductive sets, and applications. *Information and Control*, 63(1–2):118–139, 1984.

- [Harel and Paterson, 1984] D. Harel and M. S. Paterson. Undecidability of PDL with  $L = \{a^{2^i} \mid i \geq 0\}$ . *J. Comput. Syst. Sci.*, 29:359–365, 1984.
- [Harel and Peleg, 1985] D. Harel and D. Peleg. More on looping vs. repeating in dynamic logic. *Information Processing Letters*, 20:87–90, 1985.
- [Harel and Pratt, 1978] D. Harel and V. R. Pratt. Nondeterminism in logics of programs. In *Proc. 5th Symp. Princip. Prog. Lang.*, pages 203–213. ACM, 1978.
- [Harel and Raz, 1993] D. Harel and D. Raz. Deciding properties of nonregular programs. *SIAM J. Comput.*, 22:857–874, 1993.
- [Harel and Raz, 1994] D. Harel and D. Raz. Deciding emptiness for stack automata on infinite trees. *Information and Computation*, 113:278–299, 1994.
- [Harel and Sherman, 1982] D. Harel and R. Sherman. Looping vs. repeating in dynamic logic. *Infor. and Control*, 55:175–192, 1982.
- [Harel and Sherman, 1985] D. Harel and R. Sherman. Propositional dynamic logic of flowcharts. *Infor. and Control*, 64:119–135, 1985.
- [Harel and Singerman, 1996] D. Harel and E. Singerman. More on nonregular PDL: Finite models and Fibonacci-like programs. *Information and Computation*, 128:109–118, 1996.
- [Harel *et al.*, 1977] D. Harel, A. R. Meyer, and V. R. Pratt. Computability and completeness in logics of programs. In *Proc. 9th Symp. Theory of Comput.*, pages 261–268. ACM, 1977.
- [Harel *et al.*, 1982a] D. Harel, A. Pnueli, and M. Vardi. Two dimensional temporal logic and PDL with intersection. Unpublished, 1982.
- [Harel *et al.*, 1982b] D. Harel, D. Kozen, and R. Parikh. Process logic: Expressiveness, decidability, completeness. *J. Comput. Syst. Sci.*, 25(2):144–170, 1982.
- [Harel *et al.*, 1983] D. Harel, A. Pnueli, and J. Stavi. Propositional dynamic logic of nonregular programs. *J. Comput. Syst. Sci.*, 26:222–243, 1983.
- [Harel *et al.*, 2000] D. Harel, D. Kozen, and J. Tiuryn. *Dynamic Logic*. MIT Press, Cambridge, MA, 2000.
- [Harel, 1979] D. Harel. *First-Order Dynamic Logic*, volume 68 of *Lect. Notes in Comput. Sci.* Springer-Verlag, 1979.
- [Harel, 1984] D. Harel. Dynamic logic. In Gabbay and Guenther, editors, *Handbook of Philosophical Logic*, volume II: Extensions of Classical Logic, pages 497–604. Reidel, 1984.
- [Harel, 1985] D. Harel. Recurring dominoes: Making the highly undecidable highly understandable. *Annals of Discrete Mathematics*, 24:51–72, 1985.
- [Harel, 1992] D. Harel. *Algorithmics: The Spirit of Computing*. Addison-Wesley, second edition, 1992.
- [Hart *et al.*, 1982] S. Hart, M. Sharir, and A. Pnueli. Termination of probabilistic concurrent programs. In *Proc. 9th Symp. Princip. Prog. Lang.*, pages 1–6. ACM, 1982.
- [Hartonas, 1998] C. Hartonas. Duality for modal  $\mu$ -logics. *Theor. Comput. Sci.*, 202(1–2):193–222, 1998.
- [Hennessy and Plotkin, 1979] M. C. B. Hennessy and G. D. Plotkin. Full abstraction for a simple programming language. In *Proc. Symp. Semantics of Algorithmic Languages*, volume 74 of *Lecture Notes in Computer Science*, pages 108–120. Springer-Verlag, 1979.
- [Hitchcock and Park, 1972] P. Hitchcock and D. Park. Induction rules and termination proofs. In M. Nivat, editor, *Int. Colloq. Automata Lang. Prog.*, pages 225–251. North-Holland, 1972.
- [Hoare, 1969] C. A. R. Hoare. An axiomatic basis for computer programming. *Comm. Assoc. Comput. Mach.*, 12:576–580, 583, 1969.
- [Hopkins and Kozen, 1999] M. Hopkins and D. Kozen. Parikh's theorem in commutative Kleene algebra. In *Proc. Conf. Logic in Computer Science (LICS'99)*, pages 394–401. IEEE, July 1999.
- [Hughes and Cresswell, 1968] G. E. Hughes and M. J. Cresswell. *An Introduction to Modal Logic*. Methuen, 1968.
- [Huth and Kwiatkowska, 1997] M. Huth and M. Kwiatkowska. Quantitative analysis and model checking. In *Proc. 12th Symp. Logic in Comput. Sci.*, pages 111–122. IEEE, 1997.

- [Ianov, 1960] Y. I. Ianov. The logical schemes of algorithms. In *Problems of Cybernetics*, volume 1, pages 82–140. Pergamon Press, 1960.
- [Iwano and Steiglitz, 1990] K. Iwano and K. Steiglitz. A semiring on convex polygons and zero-sum cycle problems. *SIAM J. Comput.*, 19(5):883–901, 1990.
- [Jou and Smolka, 1990] C. Jou and S. Smolka. Equivalences, congruences and complete axiomatizations for probabilistic processes. In *Proc. CONCUR'90*, volume 458 of *Lecture Notes in Comput. Sci.*, pages 367–383. Springer-Verlag, 1990.
- [Kaivola, 1997] R. Kaivola. *Using Automata to Characterise Fixed Point Temporal Logics*. PhD thesis, University of Edinburgh, April 1997. Report CST-135-97.
- [Kamp, 1968] H. W. Kamp. *Tense logics and the theory of linear order*. PhD thesis, UCLA, 1968.
- [Keisler, 1971] J. Keisler. *Model Theory for Infinitary Logic*. North Holland, 1971.
- [Kfoury and Stolboushkin, 1997] A.J. Kfoury and A.P. Stolboushkin. An infinite pebble game and applications. *Information and Computation*, 136:53–66, 1997.
- [Kfoury, 1983] A.J. Kfoury. Definability by programs in first-order structures. *Theoretical Computer Science*, 25:1–66, 1983.
- [Kfoury, 1985] A. J. Kfoury. Definability by deterministic and nondeterministic programs with applications to first-order dynamic logic. *Infor. and Control*, 65(2–3):98–121, 1985.
- [Kleene, 1956] S. C. Kleene. Representation of events in nerve nets and finite automata. In C. E. Shannon and J. McCarthy, editors, *Automata Studies*, pages 3–41. Princeton University Press, Princeton, N.J., 1956.
- [Knijnenburg, 1988] P. M. W. Knijnenburg. On axiomatizations for propositional logics of programs. Technical Report RUU-CS-88-34, Rijksuniversiteit Utrecht, November 1988.
- [Koren and Pnueli, 1983] T. Koren and A. Pnueli. There exist decidable context-free propositional dynamic logics. In *Proc. Symp. on Logics of Programs*, volume 164 of *Lecture Notes in Computer Science*, pages 290–312. Springer-Verlag, 1983.
- [Kowalczyk et al., 1987] W. Kowalczyk, D. Niwiński, and J. Tiuryn. A generalization of Cook's auxiliary-pushdown-automata theorem. *Fundamenta Informaticae*, XII:497–506, 1987.
- [Kozen and Parikh, 1981] D. Kozen and R. Parikh. An elementary proof of the completeness of *PDL*. *Theor. Comput. Sci.*, 14(1):113–118, 1981.
- [Kozen and Parikh, 1983] D. Kozen and R. Parikh. A decision procedure for the propositional  $\mu$ -calculus. In Clarke and Kozen, editors, *Proc. Workshop on Logics of Programs*, volume 164 of *Lecture Notes in Computer Science*, pages 313–325. Springer-Verlag, 1983.
- [Kozen and Patron, 2000] D. Kozen and M.-C. Patron. Certification of compiler optimizations using Kleene algebra with tests. In J. Lloyd, V. Dahl, U. Furbach, M. Kerber, K.-K. Lau, C. Palamidessi, L. M. Pereira, Y. Sagiv, and P. J. Stuckey, editors, *Proc. 1st Int. Conf. Computational Logic (CL2000)*, volume 1861 of *Lecture Notes in Artificial Intelligence*, pages 568–582, London, July 2000. Springer-Verlag.
- [Kozen and Smith, 1996] D. Kozen and F. Smith. Kleene algebra with tests: Completeness and decidability. In D. van Dalen and M. Bezem, editors, *Proc. 10th Int. Workshop Computer Science Logic (CSL'96)*, volume 1258 of *Lecture Notes in Computer Science*, pages 244–259, Utrecht, The Netherlands, September 1996. Springer-Verlag.
- [Kozen and Tiuryn, 1990] D. Kozen and J. Tiuryn. Logics of programs. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B, pages 789–840. North Holland, Amsterdam, 1990.
- [Kozen, 1979a] D. Kozen. Dynamic algebra. In E. Engeler, editor, *Proc. Workshop on Logic of Programs*, volume 125 of *Lecture Notes in Computer Science*, pages 102–144. Springer-Verlag, 1979. chapter of *Propositional dynamic logics of programs: A survey* by Rohit Parikh.
- [Kozen, 1979b] D. Kozen. On the duality of dynamic algebras and Kripke models. In E. Engeler, editor, *Proc. Workshop on Logic of Programs*, volume 125 of *Lecture Notes in Computer Science*, pages 1–11. Springer-Verlag, 1979.
- [Kozen, 1979c] D. Kozen. On the representation of dynamic algebras. Technical Report RC7898, IBM Thomas J. Watson Research Center, October 1979.

- [Kozen, 1980a] D. Kozen. On the representation of dynamic algebras II. Technical Report RC8290, IBM Thomas J. Watson Research Center, May 1980.
- [Kozen, 1980b] D. Kozen. A representation theorem for models of  $*$ -free *PDL*. In *Proc. 7th Colloq. Automata, Languages, and Programming*, pages 351–362. EATCS, July 1980.
- [Kozen, 1981a] D. Kozen. Logics of programs. Lecture notes, Aarhus University, Denmark, 1981.
- [Kozen, 1981b] D. Kozen. On induction vs.  $*$ -continuity. In Kozen, editor, *Proc. Workshop on Logic of Programs*, volume 131 of *Lecture Notes in Computer Science*, pages 167–176, New York, 1981. Springer-Verlag.
- [Kozen, 1981c] D. Kozen. On the expressiveness of  $\mu$ . Manuscript, 1981.
- [Kozen, 1981d] D. Kozen. Semantics of probabilistic programs. *J. Comput. Syst. Sci.*, 22:328–350, 1981.
- [Kozen, 1982] D. Kozen. Results on the propositional  $\mu$ -calculus. In *Proc. 9th Int. Colloq. Automata, Languages, and Programming*, pages 348–359, Aarhus, Denmark, July 1982. EATCS.
- [Kozen, 1983] D. Kozen. Results on the propositional  $\mu$ -calculus. *Theor. Comput. Sci.*, 27:333–354, 1983.
- [Kozen, 1984] D. Kozen. A Ramsey theorem with infinitely many colors. In Lenstra, Lenstra, and van Emde Boas, editors, *Dopo Le Parole*, pages 71–72. University of Amsterdam, Amsterdam, May 1984.
- [Kozen, 1985] D. Kozen. A probabilistic *PDL*. *J. Comput. Syst. Sci.*, 30(2):162–178, April 1985.
- [Kozen, 1988] D. Kozen. A finite model theorem for the propositional  $\mu$ -calculus. *Studia Logica*, 47(3):233–241, 1988.
- [Kozen, 1990] D. Kozen. On Kleene algebras and closed semirings. In Rován, editor, *Proc. Math. Found. Comput. Sci.*, volume 452 of *Lecture Notes in Computer Science*, pages 26–47, Banská-Bystrica, Slovakia, 1990. Springer-Verlag.
- [Kozen, 1991a] D. Kozen. A completeness theorem for Kleene algebras and the algebra of regular events. In *Proc. 6th Symp. Logic in Comput. Sci.*, pages 214–225, Amsterdam, July 1991. IEEE.
- [Kozen, 1991b] D. Kozen. *The Design and Analysis of Algorithms*. Springer-Verlag, New York, 1991.
- [Kozen, 1994a] D. Kozen. A completeness theorem for Kleene algebras and the algebra of regular events. *Infor. and Comput.*, 110(2):366–390, May 1994.
- [Kozen, 1994b] D. Kozen. On action algebras. In J. van Eijck and A. Visser, editors, *Logic and Information Flow*, pages 78–88. MIT Press, 1994.
- [Kozen, 1996] D. Kozen. Kleene algebra with tests and commutativity conditions. In T. Margaria and B. Steffen, editors, *Proc. Second Int. Workshop Tools and Algorithms for the Construction and Analysis of Systems (TACAS'96)*, volume 1055 of *Lecture Notes in Computer Science*, pages 14–33, Passau, Germany, March 1996. Springer-Verlag.
- [Kozen, 1997a] D. Kozen. *Automata and Computability*. Springer-Verlag, New York, 1997.
- [Kozen, 1997b] D. Kozen. Kleene algebra with tests. *Transactions on Programming Languages and Systems*, 19(3):427–443, May 1997.
- [Kozen, 1997c] D. Kozen. On the complexity of reasoning in Kleene algebra. In *Proc. 12th Symp. Logic in Comput. Sci.*, pages 195–202, Los Alamitos, Ca., June 1997. IEEE.
- [Kozen, 1998] D. Kozen. Typed Kleene algebra. Technical Report 98-1669, Computer Science Department, Cornell University, March 1998.
- [Kozen, 1999a] D. Kozen. On Hoare logic and Kleene algebra with tests. In *Proc. Conf. Logic in Computer Science (LICS'99)*, pages 167–172. IEEE, July 1999.
- [Kozen, 1999b] D. Kozen. On Hoare logic, Kleene algebra, and types. Technical Report 99-1760, Computer Science Department, Cornell University, July 1999. Abstract in: Abstracts of 11th Int. Congress Logic, Methodology and Philosophy of Science, Ed. J. Cachro and K. Kijania-Placek, Krakow, Poland, August 1999, p. 15. To appear

- in: Proc. 11th Int. Congress Logic, Methodology and Philosophy of Science, ed. P. Gardenfors, K. Kijania-Placek and J. Wolenski, Kluwer.
- [Krob, 1991] Daniel Krob. A complete system of  $B$ -rational identities. *Theoretical Computer Science*, 89(2):207–343, October 1991.
- [Kuich and Salomaa, 1986] W. Kuich and A. Salomaa. *Semirings, Automata, and Languages*. Springer-Verlag, Berlin, 1986.
- [Kuich, 1987] W. Kuich. The Kleene and Parikh theorem in complete semirings. In T. Ottmann, editor, *Proc. 14th Colloq. Automata, Languages, and Programming*, volume 267 of *Lecture Notes in Computer Science*, pages 212–225, New York, 1987. EATCS, Springer-Verlag.
- [Ladner, 1977] R. E. Ladner. Unpublished, 1977.
- [Lamport, 1980] L. Lamport. “Sometime” is sometimes “not never”. *Proc. 7th Symp. Princip. Prog. Lang.*, pages 174–185, 1980.
- [Lehmann and Shelah, 1982] D. Lehmann and S. Shelah. Reasoning with time and chance. *Infor. and Control*, 53(3):165–198, 1982.
- [Lewis and Papadimitriou, 1981] H. R. Lewis and C. H. Papadimitriou. *Elements of the Theory of Computation*. Prentice Hall, 1981.
- [Lipton, 1977] R. J. Lipton. A necessary and sufficient condition for the existence of Hoare logics. In *Proc. 18th Symp. Found. Comput. Sci.*, pages 1–6. IEEE, 1977.
- [Luckham *et al.*, 1970] D. C. Luckham, D. Park, and M. Paterson. On formalized computer programs. *J. Comput. Syst. Sci.*, 4:220–249, 1970.
- [Mader, 1997] A. Mader. *Verification of Modal Properties Using Boolean Equation Systems*. PhD thesis, Fakultt fr Informatik, Technische Universitt Mnchen, September 1997.
- [Makowsky and Sain, 1986] J. A. Makowsky and I. Sain. On the equivalence of weak second-order and nonstandard time semantics for various program verification systems. In *Proc. 1st Symp. Logic in Comput. Sci.*, pages 293–300. IEEE, 1986.
- [Makowsky, 1980] J. A. Makowsky. Measuring the expressive power of dynamic logics: an application of abstract model theory. In *Proc. 7th Int. Colloq. Automata Lang. Prog.*, volume 80 of *Lect. Notes in Comput. Sci.*, pages 409–421. Springer-Verlag, 1980.
- [Makowsky and Tiomkin, 1980] J. A. Makowsky and M. L. Tiomkin. Probabilistic propositional dynamic logic. Manuscript, 1980.
- [Manna and Pnueli, 1981] Z. Manna and A. Pnueli. Verification of concurrent programs: temporal proof principles. In D. Kozen, editor, *Proc. Workshop on Logics of Programs*, volume 131 of *Lect. Notes in Comput. Sci.*, pages 200–252. Springer-Verlag, 1981.
- [Manna and Pnueli, 1987] Z. Manna and A. Pnueli. Specification and verification of concurrent programs by  $\forall$ -automata. In *Proc. 14th Symp. Principles of Programming Languages*, pages 1–12. ACM, January 1987.
- [Manna, 1974] Z. Manna. *Mathematical Theory of Computation*. McGraw-Hill, 1974.
- [McCulloch and Pitts, 1943] W. S. McCulloch and W. Pitts. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophysics*, 5:115–143, 1943.
- [Mehlhorn, 1984] K. Mehlhorn. *Graph Algorithms and NP-Completeness*, volume II of *Data Structures and Algorithms*. Springer-Verlag, 1984.
- [Meyer and Halpern, 1982] A. R. Meyer and J. Y. Halpern. Axiomatic definitions of programming languages: a theoretical assessment. *J. Assoc. Comput. Mach.*, 29:555–576, 1982.
- [Meyer and Parikh, 1981] A. R. Meyer and R. Parikh. Definability in dynamic logic. *J. Comput. Syst. Sci.*, 23:279–298, 1981.
- [Meyer and Tiuryn, 1981] A. R. Meyer and J. Tiuryn. A note on equivalences among logics of programs. In D. Kozen, editor, *Proc. Workshop on Logics of Programs*, volume 131 of *Lect. Notes in Comput. Sci.*, pages 282–299. Springer-Verlag, 1981.
- [Meyer and Tiuryn, 1984] A. R. Meyer and J. Tiuryn. Equivalences among logics of programs. *Journal of Computer and Systems Science*, 29:160–170, 1984.
- [Meyer and Winklmann, 1982] A. R. Meyer and K. Winklmann. Expressing program looping in regular dynamic logic. *Theor. Comput. Sci.*, 18:301–323, 1982.

- [Meyer *et al.*, 1981] A. R. Meyer, R. S. Streett, and G. Mirkowska. The deducibility problem in propositional dynamic logic. In E. Engeler, editor, *Proc. Workshop Logic of Programs*, volume 125 of *Lect. Notes in Comput. Sci.*, pages 12–22. Springer-Verlag, 1981.
- [Miller, 1976] G. L. Miller. Riemann's hypothesis and tests for primality. *J. Comput. Syst. Sci.*, 13:300–317, 1976.
- [Mirkowska, 1971] G. Mirkowska. On formalized systems of algorithmic logic. *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys.*, 19:421–428, 1971.
- [Mirkowska, 1980] G. Mirkowska. Algorithmic logic with nondeterministic programs. *Fund. Informaticae*, III:45–64, 1980.
- [Mirkowska, 1981a] G. Mirkowska. PAL—propositional algorithmic logic. In E. Engeler, editor, *Proc. Workshop Logic of Programs*, volume 125 of *Lect. Notes in Comput. Sci.*, pages 23–101. Springer-Verlag, 1981.
- [Mirkowska, 1981b] G. Mirkowska. PAL—propositional algorithmic logic. *Fund. Informaticae*, IV:675–760, 1981.
- [Morgan *et al.*, 1999] C. Morgan, A. McIver, and K. Seidel. Probabilistic predicate transformers. *ACM Trans. Programming Languages and Systems*, 8(1):1–30, 1999.
- [Moschovakis, 1974] Y. N. Moschovakis. *Elementary Induction on Abstract Structures*. North-Holland, 1974.
- [Muller *et al.*, 1988] D. E. Muller, A. Saoudi, and P. E. Schupp. Weak alternating automata give a simple explanation of why most temporal and dynamic logics are decidable in exponential time. In *Proc. 3rd Symp. Logic in Computer Science*, pages 422–427. IEEE, July 1988.
- [Németi, 1980] I. Németi. Every free algebra in the variety generated by the representable dynamic algebras is separable and representable. Manuscript, 1980.
- [Németi, 1981] I. Németi. Nonstandard dynamic logic. In D. Kozen, editor, *Proc. Workshop on Logics of Programs*, volume 131 of *Lect. Notes in Comput. Sci.*, pages 311–348. Springer-Verlag, 1981.
- [Ng and Tarski, 1977] K. C. Ng and A. Tarski. Relation algebras with transitive closure, abstract 742-02-09. *Notices Amer. Math. Soc.*, 24:A29–A30, 1977.
- [Ng, 1984] K. C. Ng. *Relation Algebras with Transitive Closure*. PhD thesis, University of California, Berkeley, 1984.
- [Nishimura, 1979] H. Nishimura. Sequential method in propositional dynamic logic. *Acta Informatica*, 12:377–400, 1979.
- [Nishimura, 1980] H. Nishimura. Descriptively complete process logic. *Acta Informatica*, 14:359–369, 1980.
- [Niwinski, 1984] D. Niwinski. The propositional  $\mu$ -calculus is more expressive than the propositional dynamic logic of looping. University of Warsaw, 1984.
- [Parikh and Mahoney, 1983] R. Parikh and A. Mahoney. A theory of probabilistic programs. In E. Clarke and D. Kozen, editors, *Proc. Workshop on Logics of Programs*, volume 164 of *Lect. Notes in Comput. Sci.*, pages 396–402. Springer-Verlag, 1983.
- [Parikh, 1978a] R. Parikh. The completeness of propositional dynamic logic. In *Proc. 7th Symp. on Math. Found. of Comput. Sci.*, volume 64 of *Lect. Notes in Comput. Sci.*, pages 403–415. Springer-Verlag, 1978.
- [Parikh, 1978b] R. Parikh. A decidability result for second order process logic. In *Proc. 19th Symp. Found. Comput. Sci.*, pages 177–183. IEEE, 1978.
- [Parikh, 1981] R. Parikh. Propositional dynamic logics of programs: a survey. In E. Engeler, editor, *Proc. Workshop on Logics of Programs*, volume 125 of *Lect. Notes in Comput. Sci.*, pages 102–144. Springer-Verlag, 1981.
- [Parikh, 1983] R. Parikh. Propositional game logic. In *Proc. 23rd IEEE Symp. Foundations of Computer Science*, 1983.
- [Park, 1976] D. Park. Finiteness is  $\mu$ -ineffable. *Theor. Comput. Sci.*, 3:173–181, 1976.
- [Paterson and Hewitt, 1970] M. S. Paterson and C. E. Hewitt. Comparative schematology. In *Record Project MAC Conf. on Concurrent Systems and Parallel Computation*, pages 119–128. ACM, 1970.
- [Pecuchet, 1986] J. P. Pecuchet. On the complementation of Büchi automata. *Theor. Comput. Sci.*, 47:95–98, 1986.

- [Peleg, 1987a] D. Peleg. Communication in concurrent dynamic logic. *J. Comput. Sys. Sci.*, 35:23–58, 1987.
- [Peleg, 1987b] D. Peleg. Concurrent dynamic logic. *J. Assoc. Comput. Mach.*, 34(2):450–479, 1987.
- [Peleg, 1987c] D. Peleg. Concurrent program schemes and their logics. *Theor. Comput. Sci.*, 55:1–45, 1987.
- [Peng and Iyer, 1995] W. Peng and S. Purushothaman Iyer. A new type of pushdown-tree automata on infinite trees. *Int. J. of Found. of Comput. Sci.*, 6(2):169–186, 1995.
- [Peterson, 1978] G. L. Peterson. The power of tests in propositional dynamic logic. Technical Report 47, Comput. Sci. Dept., Univ. of Rochester, 1978.
- [Pnueli, 1977] A. Pnueli. The temporal logic of programs. In *Proc. 18th Symp. Found. Comput. Sci.*, pages 46–57. IEEE, 1977.
- [Pratt, 1976] V. R. Pratt. Semantical considerations on Floyd-Hoare logic. In *Proc. 17th Symp. Found. Comput. Sci.*, pages 109–121. IEEE, 1976.
- [Pratt, 1978] V. R. Pratt. A practical decision method for propositional dynamic logic. In *Proc. 10th Symp. Theory of Comput.*, pages 326–337. ACM, 1978.
- [Pratt, 1979a] V. R. Pratt. Dynamic algebras: examples, constructions, applications. Technical Report TM-138, MIT/LCS, July 1979.
- [Pratt, 1979b] V. R. Pratt. Models of program logics. In *Proc. 20th Symp. Found. Comput. Sci.*, pages 115–122. IEEE, 1979.
- [Pratt, 1979c] V. R. Pratt. Process logic. In *Proc. 6th Symp. Princip. Prog. Lang.*, pages 93–100. ACM, 1979.
- [Pratt, 1980a] V. R. Pratt. Dynamic algebras and the nature of induction. In *Proc. 12th Symp. Theory of Comput.*, pages 22–28. ACM, 1980.
- [Pratt, 1980b] V. R. Pratt. A near-optimal method for reasoning about actions. *J. Comput. Syst. Sci.*, 20(2):231–254, 1980.
- [Pratt, 1981a] V. R. Pratt. A decidable  $\mu$ -calculus: preliminary report. In *Proc. 22nd Symp. Found. Comput. Sci.*, pages 421–427. IEEE, 1981.
- [Pratt, 1981b] V. R. Pratt. Using graphs to understand PDL. In D. Kozen, editor, *Proc. Workshop on Logics of Programs*, volume 131 of *Lect. Notes in Comput. Sci.*, pages 387–396. Springer-Verlag, 1981.
- [Pratt, 1988] V. Pratt. Dynamic algebras as a well-behaved fragment of relation algebras. In D. Pigozzi, editor, *Proc. Conf. on Algebra and Computer Science*, volume 425 of *Lecture Notes in Computer Science*, pages 77–110, Ames, Iowa, June 1988. Springer-Verlag.
- [Pratt, 1990] V. Pratt. Action logic and pure induction. In J. van Eijck, editor, *Proc. Logics in AI: European Workshop JELIA '90*, volume 478 of *Lecture Notes in Computer Science*, pages 97–120, New York, September 1990. Springer-Verlag.
- [Rabin and Scott, 1959] M. O. Rabin and D. S. Scott. Finite automata and their decision problems. *IBM J. Res. Develop.*, 3(2):115–125, 1959.
- [Rabin, 1980] M. O. Rabin. Probabilistic algorithms for testing primality. *J. Number Theory*, 12:128–138, 1980.
- [Ramshaw, 1981] L. H. Ramshaw. *Formalizing the analysis of algorithms*. PhD thesis, Stanford Univ., 1981.
- [Rasiowa and Sikorski, 1963] H. Rasiowa and R. Sikorski. *Mathematics of Metamathematics*. Polish Scientific Publishers, PWN, 1963.
- [Redko, 1964] V. N. Redko. On defining relations for the algebra of regular events. *Ukrain. Mat. Z.*, 16:120–126, 1964. In Russian.
- [Reif, 1980] J. Reif. Logics for probabilistic programming. In *Proc. 12th Symp. Theory of Comput.*, pages 8–13. ACM, 1980.
- [Rogers, 1967] H. Rogers. *Theory of Recursive Functions and Effective Computability*. McGraw-Hill, 1967.
- [Safra, 1988] S. Safra. On the complexity of  $\omega$ -automata. In *Proc. 29th Symp. Foundations of Comput. Sci.*, pages 319–327. IEEE, October 1988.

- [Sakarovitch, 1987] J. Sakarovitch. Kleene's theorem revisited: A formal path from Kleene to Chomsky. In A. Kelemenova and J. Keleman, editors, *Trends, Techniques, and Problems in Theoretical Computer Science*, volume 281 of *Lecture Notes in Computer Science*, pages 39–50, New York, 1987. Springer-Verlag.
- [Salomaa, 1966] A. Salomaa. Two complete axiom systems for the algebra of regular events. *J. Assoc. Comput. Mach.*, 13(1):158–169, January 1966.
- [Salwicki, 1970] A. Salwicki. Formalized algorithmic languages. *Bull. Acad. Polon. Sci. Ser. Sci. Math. Astron. Phys.*, 18:227–232, 1970.
- [Salwicki, 1977] A. Salwicki. Algorithmic logic: a tool for investigations of programs. In Butts and Hintikka, editors, *Logic Foundations of Mathematics and Computability Theory*, pages 281–295. Reidel, 1977.
- [Sazonov, 1980] V.Y. Sazonov. Polynomial computability and recursivity in finite domains. *Elektronische Informationsverarbeitung und Kibernetik*, 16:319–323, 1980.
- [Scott and de Bakker, 1969] D. S. Scott and J. W. de Bakker. A theory of programs. IBM Vienna, 1969.
- [Segala and Lynch, 1994] R. Segala and N. Lynch. Probabilistic simulations for probabilistic processes. In *Proc. CONCUR'94*, volume 836 of *Lecture Notes in Computer Sci.*, pages 481–496. Springer-Verlag, 1994.
- [Segerberg, 1977] K. Segerberg. A completeness theorem in the modal logic of programs (preliminary report). *Not. Amer. Math. Soc.*, 24(6):A-552, 1977.
- [Shoenfield, 1967] J. R. Shoenfield. *Mathematical Logic*. Addison-Wesley, 1967.
- [Sholz, 1952] H. Sholz. Ein ungelöstes Problem in der symbolischen Logik. *The Journal of Symbolic Logic*, 17:160, 1952.
- [Sistla and Clarke, 1982] A. P. Sistla and E. M. Clarke. The complexity of propositional linear temporal logics. In *Proc. 14th Symp. Theory of Comput.*, pages 159–168. ACM, 1982.
- [Sistla et al., 1987] A. P. Sistla, M. Y. Vardi, and P. Wolper. The complementation problem for Büchi automata with application to temporal logic. *Theor. Comput. Sci.*, 49:217–237, 1987.
- [Sokolsky and Smolka, 1994] O. Sokolsky and S. Smolka. Incremental model checking in the modal  $\mu$ -calculus. In D. Dill, editor, *Proc. Conf. Computer Aided Verification*, volume 818 of *Lect. Notes in Comput. Sci.*, pages 352–363. Springer, June 1994.
- [Steffen et al., 1996] B. Steffen, T. Margaria, A. Classen, V. Braun, R. Nisius, and M. Reitenspiess. A constraint oriented service environment. In T. Margaria and B. Steffen, editors, *Proc. Second Int. Workshop Tools and Algorithms for the Construction and Analysis of Systems (TACAS'96)*, volume 1055 of *Lect. Notes in Comput. Sci.*, pages 418–421. Springer, March 1996.
- [Stirling and Walker, 1989] C. Stirling and D. Walker. Local model checking in the modal  $\mu$ -calculus. In *Proc. Int. Joint Conf. Theory and Practice of Software Develop. (TAPSOFT89)*, volume 352 of *Lect. Notes in Comput. Sci.*, pages 369–383. Springer, March 1989.
- [Stirling, 1992] C. Stirling. Modal and temporal logics. In S. Abramsky, D. Gabbay, and T. Maibaum, editors, *Handbook of Logic in Computer Science*, pages 477–563. Clarendon Press, 1992.
- [Stockmeyer and Meyer, 1973] L. J. Stockmeyer and A. R. Meyer. Word problems requiring exponential time. In *Proc. 5th Symp. Theory of Computing*, pages 1–9, New York, 1973. ACM.
- [Stolboushkin and Taitslin, 1983] A. P. Stolboushkin and M. A. Taitslin. Deterministic dynamic logic is strictly weaker than dynamic logic. *Infor. and Control*, 57:48–55, 1983.
- [Stolboushkin, 1983] A.P. Stolboushkin. Regular dynamic logic is not interpretable in deterministic context-free dynamic logic. *Information and Computation*, 59:94–107, 1983.
- [Stolboushkin, 1989] A.P. Stolboushkin. Some complexity bounds for dynamic logic. In *Proc. 4th Symp. Logic in Comput. Sci.*, pages 324–332. IEEE, June 1989.
- [Street and Emerson, 1984] R. Street and E. A. Emerson. The propositional  $\mu$ -calculus is elementary. In *Proc. 11th Int. Colloq. on Automata Languages and Programming*, pages 465–472. Springer, 1984. *Lect. Notes in Comput. Sci.* 172.



- [Streett, 1981] R. S. Streett. Propositional dynamic logic of looping and converse. In *Proc. 13th Symp. Theory of Comput.*, pages 375–381. ACM, 1981.
- [Streett, 1982] R. S. Streett. Propositional dynamic logic of looping and converse is elementarily decidable. *Infor. and Control*, 54:121–141, 1982.
- [Streett, 1985a] R. S. Streett. Fixpoints and program looping: reductions from the propositional  $\mu$ -calculus into propositional dynamic logics of looping. In R. Parikh, editor, *Proc. Workshop on Logics of Programs*, volume 193 of *Lect. Notes in Comput. Sci.*, pages 359–372. Springer-Verlag, 1985.
- [Streett, 1985b] R. Streett. Fixpoints and program looping: reductions from the propositional  $\mu$ -calculus into propositional dynamic logics of looping. In Parikh, editor, *Proc. Workshop on Logics of Programs 1985*, pages 359–372. Springer, 1985. *Lect. Notes in Comput. Sci.* 193.
- [Tarjan, 1981] R. E. Tarjan. A unified approach to path problems. *J. Assoc. Comput. Mach.*, pages 577–593, 1981.
- [Thiele, 1966] H. Thiele. Wissenschaftstheoretische untersuchungen in algorithmischen sprachen. In *Theorie der Graphschemata-Kalkale Veb Deutscher Verlag der Wissenschaften*. Berlin, 1966.
- [Thomas, 1997] W. Thomas. Languages, automata, and logic. Technical Report 9607, Christian-Albrechts-Universität Kiel, May 1997.
- [Tiuryn and Urzyczyn, 1983] J. Tiuryn and P. Urzyczyn. Some relationships between logics of programs and complexity theory. In *Proc. 24th Symp. Found. Comput. Sci.*, pages 180–184. IEEE, 1983.
- [Tiuryn and Urzyczyn, 1984] J. Tiuryn and P. Urzyczyn. Remarks on comparing expressive power of logics of programs. In Chytil and Koubek, editors, *Proc. Math. Found. Comput. Sci.*, volume 176 of *Lect. Notes in Comput. Sci.*, pages 535–543. Springer-Verlag, 1984.
- [Tiuryn and Urzyczyn, 1988] J. Tiuryn and P. Urzyczyn. Some relationships between logics of programs and complexity theory. *Theor. Comput. Sci.*, 60:83–108, 1988.
- [Tiuryn, 1981a] J. Tiuryn. A survey of the logic of effective definitions. In E. Engeler, editor, *Proc. Workshop on Logics of Programs*, volume 125 of *Lect. Notes in Comput. Sci.*, pages 198–245. Springer-Verlag, 1981.
- [Tiuryn, 1981b] J. Tiuryn. Unbounded program memory adds to the expressive power of first-order programming logics. In *Proc. 22nd Symp. Found. Comput. Sci.*, pages 335–339. IEEE, 1981.
- [Tiuryn, 1984] J. Tiuryn. Unbounded program memory adds to the expressive power of first-order programming logics. *Infor. and Control*, 60:12–35, 1984.
- [Tiuryn, 1986] J. Tiuryn. Higher-order arrays and stacks in programming: an application of complexity theory to logics of programs. In Gruska and Rován, editors, *Proc. Math. Found. Comput. Sci.*, volume 233 of *Lect. Notes in Comput. Sci.*, pages 177–198. Springer-Verlag, 1986.
- [Tiuryn, 1989] J. Tiuryn. A simplified proof of  $DDL < DL$ . *Information and Computation*, 81:1–12, 1989.
- [Trnkova and Reiterman, 1980] V. Trnkova and J. Reiterman. Dynamic algebras which are not Kripke structures. In *Proc. 9th Symp. on Math. Found. Comput. Sci.*, pages 528–538, 1980.
- [Turing, 1936] A. M. Turing. On computable numbers with an application to the Entscheidungsproblem. *Proc. London Math. Soc.*, 42:230–265, 1936. Erratum: *Ibid.*, 43 (1937), pp. 544–546.
- [Urzyczyn, 1983] P. Urzyczyn. A necessary and sufficient condition in order that a Herbrand interpretation be expressive relative to recursive programs. *Information and Control*, 56:212–219, 1983.
- [Urzyczyn, 1986] P. Urzyczyn. “During” cannot be expressed by “after”. *Journal of Computer and System Sciences*, 32:97–104, 1986.
- [Urzyczyn, 1987] P. Urzyczyn. Deterministic context-free dynamic logic is more expressive than deterministic dynamic logic of regular programs. *Fundamenta Informaticae*, 10:123–142, 1987.
- [Urzyczyn, 1988] P. Urzyczyn. Logics of programs with Boolean memory. *Fundamenta Informaticae*, XI:21–40, 1988.

- [Valiev, 1980] M. K. Valiev. Decision complexity of variants of propositional dynamic logic. In *Proc. 9th Symp. Math. Found. Comput. Sci.*, volume 88 of *Lect. Notes in Comput. Sci.*, pages 656–664. Springer-Verlag, 1980.
- [van Emde Boas, 1978] P. van Emde Boas. The connection between modal logic and algorithmic logics. In *Symp. on Math. Found. of Comp. Sci.*, pages 1–15, 1978.
- [Vardi and Stockmeyer, 1985] M. Y. Vardi and L. Stockmeyer. Improved upper and lower bounds for modal logics of programs: preliminary report. In *Proc. 17th Symp. Theory of Comput.*, pages 240–251. ACM, May 1985.
- [Vardi and Wolper, 1986a] M. Y. Vardi and P. Wolper. An automata-theoretic approach to automatic program verification. In *Proc. 1st Symp. Logic in Computer Science*, pages 332–344. IEEE, June 1986.
- [Vardi and Wolper, 1986b] M. Y. Vardi and P. Wolper. Automata-theoretic techniques for modal logics of programs. *J. Comput. Syst. Sci.*, 32:183–221, 1986.
- [Vardi, 1985a] M. Y. Vardi. Automatic verification of probabilistic concurrent finite-state programs. In *Proc. 26th Symp. Found. Comput. Sci.*, pages 327–338. IEEE, October 1985.
- [Vardi, 1985b] M. Y. Vardi. The taming of the converse: reasoning about two-way computations. In R. Parikh, editor, *Proc. Workshop on Logics of Programs*, volume 193 of *Lect. Notes in Comput. Sci.*, pages 413–424. Springer-Verlag, 1985.
- [Vardi, 1987] M. Y. Vardi. Verification of concurrent programs: the automata-theoretic framework. In *Proc. 2nd Symp. Logic in Comput. Sci.*, pages 167–176. IEEE, June 1987.
- [Vardi, 1998a] M. Vardi. Linear vs. branching time: a complexity-theoretic perspective. In *Proc. 13th Symp. Logic in Comput. Sci.*, pages 394–405. IEEE, 1998.
- [Vardi, 1998b] M. Y. Vardi. Reasoning about the past with two-way automata. In *Proc. 25th Int. Colloq. Automata Lang. Prog.*, volume 1443 of *Lect. Notes in Comput. Sci.*, pages 628–641. Springer-Verlag, July 1998.
- [Walukiewicz, 1993] I. Walukiewicz. Completeness result for the propositional  $\mu$ -calculus. In *Proc. 8th IEEE Symp. Logic in Comput. Sci.*, June 1993.
- [Walukiewicz, 1995] I. Walukiewicz. Completeness of Kozen’s axiomatisation of the propositional  $\mu$ -calculus. In *Proc. 10th Symp. Logic in Comput. Sci.*, pages 14–24. IEEE, June 1995.
- [Walukiewicz, 2000] I. Walukiewicz. Completeness of Kozen’s axiomatisation of the propositional  $\mu$ -calculus. *Infor. and Comput.*, 157(1–2):142–182, February–March 2000.
- [Wand, 1978] M. Wand. A new incompleteness result for Hoare’s system. *J. Assoc. Comput. Mach.*, 25:168–175, 1978.
- [Wolper, 1981] P. Wolper. Temporal logic can be more expressive. In *Proc. 22nd Symp. Foundations of Computer Science*, pages 340–348. IEEE, 1981.
- [Wolper, 1983] P. Wolper. Temporal logic can be more expressive. *Infor. and Control*, 56:72–99, 1983.



## LOGICS FOR DEFEASIBLE ARGUMENTATION

## 1 INTRODUCTION

Logic is the science that deals with the formal principles and criteria of validity of patterns of inference. This chapter surveys logics for a particular group of patterns of inference, namely those where arguments for and against a certain claim are produced and evaluated, to test the tenability of the claim. Such reasoning processes are usually analysed under the common term ‘defeasible argumentation’. We shall illustrate this form of reasoning with a dispute between two persons, *A* and *B*. They disagree on whether it is morally acceptable for a newspaper to publish a certain piece of information concerning a politician’s private life.<sup>1</sup> Let us assume that the two parties have reached agreement on the following points.

- (1) The piece of information *I* concerns the health of person *P*;
- (2) *P* does not agree with publication of *I*;
- (3) Information concerning a person’s health is information concerning that person’s private life

*A* now states the moral principle that

- (4) Information concerning a person’s private life may not be published if that person does not agree with publication.

and *A* says “So the newspapers may not publish *I*” (Fig. 1, page 220). Although *B* accepts principle (4) and is therefore now committed to (1-4), *B* still refuses to accept the conclusion that the newspapers may not publish *I*. *B* motivates his refusal by replying that:

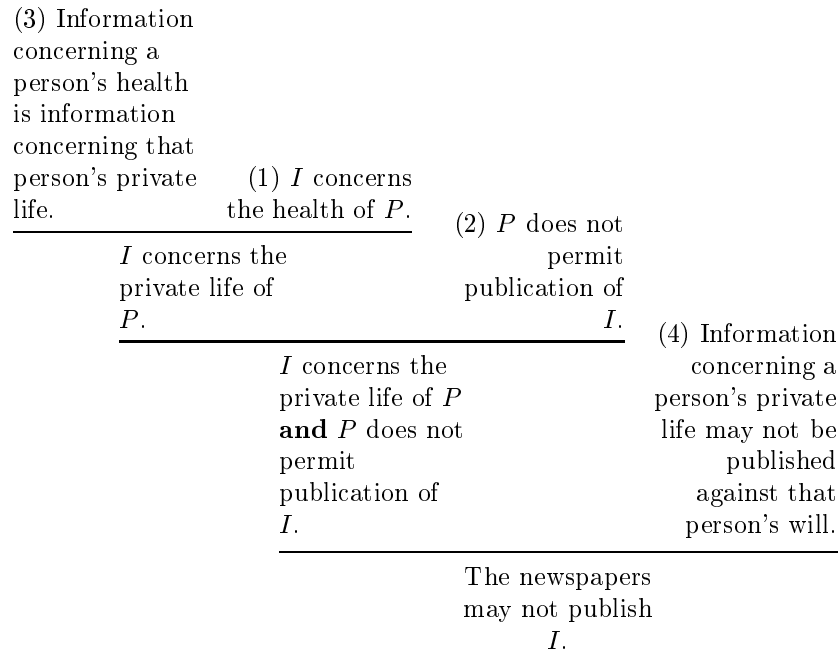
- (5) *P* is a cabinet minister
- (6) *I* is about a disease that might affect *P*’s political functioning
- (7) Information about things that might affect a cabinet minister’s political functioning has public significance

Furthermore, *B* maintains that there is also the moral principle that

- (8) Newspapers may publish any information that has public significance

---

<sup>1</sup>Adapted from [Sartor, 1994].

Figure 1. *A*'s argument.

*B* concludes by saying that therefore the newspapers may write about *P*'s disease (Fig. 2, page 221). *A* agrees with (5–7) and even accepts (8) as a moral principle, but *A* does not give up his initial claim. (It is assumed that *A* and *B* are both male.) Instead he tries to defend it by arguing that he has the stronger argument: he does so by arguing that in this case

- (9) The likelihood that the disease mentioned in *I* affects *P*'s functioning is small.
- (10) If the likelihood that the disease mentioned in *I* affects *P*'s functioning is small, then principle (4) has priority over principle (8).

Thus it can be derived that the principle used in *A*'s first argument is stronger than the principle used by *B* (Fig. 3, page 222), which makes *A*'s first argument stronger than *B*'s, so that it follows after all that the newspapers should be silent about *P*'s disease.

Let us examine the various stages of this dispute in some detail. Intuitively, it seems obvious that the accepted basis for discussion after *A* has stated (4) and *B* has accepted it, viz. (1,2,3,4), warrants the conclusion that the piece of information *I* may not be published. However, after *B*'s counterargument and *A*'s acceptance of its premises (5–8) things have changed.

(5) <i>P</i> is a cabinet minister.	(6) <i>I</i> is about a disease that might affect <i>P</i> 's political functioning.	(7) Information about things that might affect a cabinet minister's political functioning has public significance.	(8) Newspapers may publish any information that has public significance.
<div style="display: flex; justify-content: space-between;"> <div style="width: 45%;"> <i>I</i> is about a disease that might affect a cabinet minister's political functioning. </div> <div style="width: 45%;"> <i>I</i> has public significance. </div> </div>		The newspapers may publish <i>I</i> .	

Figure 2. *B*'s argument.

At this stage the joint basis for discussion is (1-8), which gives rise to two conflicting arguments. Moreover, (1-8) does not yield reasons to prefer one argument over the other: so at this point *A*'s conclusion has ceased to be warranted. But then *A*'s second argument, which states a preference between the two conflicting moral principles, tips the balance in favour of his first argument: so after the basis for discussion has been extended to (1-10), we must again accept *A*'s moral claim as warranted.

This chapter is about logical systems that formalise this kind of reasoning. We shall call them 'logics for defeasible argumentation', or 'argumentation systems'. As the example shows, these systems lack one feature of 'standard', deductive logic (say, first-order predicate logic, FOL). The notion of 'warrant' that we used in explaining the example is clearly not the same as first-order logical consequence, which has the property of monotonicity: in FOL any conclusion that can be drawn from a given set of premises, remains valid if we add new premises to this set. So according to FOL, if *A*'s claim is implied by (1-4), it is surely also implied by (1-8). From the point of view of FOL it is pointless for *B* to accept (1-4) and yet state a counterargument; *B* should also have refused to accept one of the premises, for instance, (4).

(9) The likelihood that the disease mentioned in <i>I</i> affects <i>P</i> 's functioning is small.	(10) If the likelihood that the disease mentioned in <i>I</i> affects <i>P</i> 's functioning is small, then principle (4) has priority over principle (8).
---	---

---

Principle (4) has priority  
over principle (8).

Figure 3. *A*'s priority argument.

Does this mean that our informal account of the example is misleading, that it conceals a subtle change in the interpretation of, say, (4) as the dispute progresses? This is not so easy to answer in general. Although in some cases it might indeed be best to analyse an argument move like *B*'s as a reinterpretation of a premise, in other cases this is different. In actual reasoning, rules are not always neatly labelled with an exhaustive list of possible exceptions; rather, people are often forced to apply 'rules of thumb' or 'default rules', in the absence of evidence to the contrary, and it seems natural to analyse an argument like *B*'s as an attempt to provide such evidence to the contrary. When the example is thus analysed, the force of the conclusions drawn in it can only be captured by a consequence notion that is nonmonotonic: although *A*'s claim is warranted on the basis of (1–4), it is not warranted on the basis of (1–8).

Such nonmonotonic consequence notions have been studied over the last twenty years in an area of artificial intelligence called 'nonmonotonic reasoning' (recently the term 'defeasible reasoning' has also become popular), and logics for defeasible argumentation are largely a result of this development. Some might say that the lack of the property of monotonicity disqualifies these notions from being notions of logical consequence: isn't the very idea of calling an inference 'logical' that it is (given the premises) beyond any doubt? We are not so sure. Our view on logic is that it studies criteria of warrant, that is, criteria that determine the degree according to which it is reasonable to accept logical conclusions, even though some of these conclusions are established non-deductively: sometimes it is reasonable to accept a conclusion of an argument even though this argument is not strong enough to establish its conclusion with absolute certainty.

Several ways to formalise nonmonotonic, or defeasible reasoning have been studied. This chapter is not meant to survey all of them but only discusses the argument-based approach, which defines notions like argument, counterargument, attack and defeat, and defines consequence notions in terms of the interaction of arguments for and against certain conclusions. This approach was initiated by the philosopher John Pollock [1987], based

on his earlier work in epistemology, e.g. [1974], and the computer scientist Ronald Loui [1987]. As we shall see, argumentation systems are able to incorporate the traditional, monotonic notions of logical consequence as a special case, for instance, in their definition of what an argument is.

The field of defeasible argumentation is relatively young, and researchers disagree on many issues, while the formal meta-theory is still in its early stages. Yet we think that the field has sufficiently matured to devote a handbook survey to it.<sup>2</sup> We aim to show that there are also many similarities and connections between the various systems, and that many differences are variations on a few basic notions, or are caused by different focus or different levels of abstraction. Moreover, we shall show that some recent developments pave the way for a more elaborate meta-theory of defeasible argumentation.

Although when discussing individual systems we aim to be as formal as possible, when comparing them we shall mostly use conceptual or quasi-formal terms. We shall also report on some formal results on this comparison, but it is not our aim to present new technical results; this we regard as a task for further research in the field.

The structure of this chapter is as follows. In Section 2 we give an overview of the main approaches in nonmonotonic reasoning, and argue why the study of this kind of reasoning is relevant not only for artificial intelligence but also for philosophy. In Section 3 we give a brief conceptual sketch of logics for defeasible argumentation, and we argue that it is not obvious that they need a model-theoretic semantics. In Section 4 we become formal, studying how semantic consequence notions for argumentation systems can be defined given a set of arguments ordered by a defeat relation. This discussion is still abstract, leaving the structure of arguments and the origin of the defeat relation largely unspecified. In Section 5 we become more concrete, in discussing particular logics for defeasible argumentation. Then in Section 6 we discuss one way in which argumentation systems can be formulated, viz. in the form of rules for dispute. We end this chapter in Section 7 with some concluding remarks, and with a list of the main open issues in the field.

## 2 NONMONOTONIC LOGICS: OVERVIEW AND PHILOSOPHICAL RELEVANCE

Before discussing argumentation systems, we place them in the context of the study of nonmonotonic reasoning, and discuss why this study deserves a place in philosophical logic.

---

<sup>2</sup>For a survey of this topic from a computer science perspective, see [Chesñevar *et al.*, 1999].



### 2.1 *Research in nonmonotonic reasoning*

Although this chapter is not about nonmonotonic logics in general, it is still useful to give a brief impression of this field, to put systems for defeasible argumentation in context. Several styles of nonmonotonic logics exist. Most of them take as the basic ‘nonstandard’ unit the notion of a default, or defeasible conditional or rule: this is a conditional that can be qualified with phrases like ‘typically’, ‘normally’ or ‘unless shown otherwise’ (the two principles in our example may be regarded as defaults). Defaults do not guarantee that their consequent holds whenever their antecedent holds; instead they allow us in such cases to defeasibly derive their consequent, i.e., if nothing is known about exceptional circumstances. Most nonmonotonic logics aim to formalise this phenomenon of ‘default reasoning’, but they do so in different ways.

Firstly, they differ in whether the above qualifications are regarded as extra conditions in the antecedent of a default, as aspects of the use of a default, or as inherent in the meaning of a defeasible conditional operator. In addition, within each of these views on defaults, nonmonotonic logics differ in the technical means by which they formalise it. Let us briefly review the main approaches. (More detailed overviews can be found in e.g. [Brewka, 1991] and [Gabbay *et al.*, 1994].)

#### *Preferential entailment*

Preferential entailment, e.g. [Shoham, 1988], is a model-theoretic approach based on standard first-order logic, which weakens the standard notion of entailment. The idea is that instead of checking *all* models of the premises to see if the conclusion holds, only some of the models are checked, viz. those in which as few exceptions to the defaults hold as possible. This technique is usually combined with the ‘extra condition’ view on defaults, by adding a special ‘normality condition’ to their antecedent, as in

$$(1) \quad \forall x. \text{Birds}(x) \wedge \neg \text{ab}_1(x) \supset \text{Canfly}(x)$$

Informally, this reads as ‘Birds can fly, unless they are abnormal with respect to flying’. Let us now also assume that Tweety is a bird:

$$(2) \quad \text{Bird}(\text{Tweety})$$

We want to infer from (1) and (2) that  $\text{Canfly}(\text{Tweety})$ , since there is no reason to believe that  $\text{ab}_1(\text{Tweety})$ . This inference is formalised by only looking at those models of (1,2) where the extension of the  $\text{ab}_i$  predicates are minimal (with respect to set inclusion). Thus, since on the basis of (1) and (2) nothing is known about whether Tweety is an abnormal bird, there are both FOL-models of these premises where  $\text{ab}_1(\text{Tweety})$  is satisfied and FOL-models where this is not satisfied. The idea is then that we can

disregard the models satisfying  $\text{ab}_1(\textit{Tweety})$ , and only look at the models satisfying  $\neg \text{ab}_1(\textit{Tweety})$ ; clearly in all those models  $\text{Canfly}(\textit{Tweety})$  holds.

The defeasibility of this inference can be shown by adding  $\text{ab}_1(\textit{Tweety})$  to the premises. Then all models of the premises satisfy  $\text{ab}_1(\textit{Tweety})$ , and the preferred models are now those in which the extension of  $\text{ab}_1$  is  $\{\textit{Tweety}\}$ . Some of those models satisfy  $\text{Canfly}(\textit{Tweety})$  but others satisfy  $\neg \text{Canfly}(\textit{Tweety})$ , so we cannot any more draw the conclusion  $\text{Canfly}(\textit{Tweety})$ .

A variant of this approach is Poole's [1988] 'abductive framework for default reasoning'. Poole also represents defaults with normality conditions, but he does not define a new semantics. Instead, he recommends a new way of using first-order logic, viz. for constructing 'extensions' of a theory. Essentially, extensions can be formed by adding as many normality statements to a theory as is consistently possible. The standard first-order models of a theory extension correspond to the preferred models of the original theory.

### *Intensional semantics for defaults*

There are also intensional approaches to the semantics of defaults, e.g. [Delgrande, 1988; Asher & Morreau, 1990]. The idea is to interpret defaults in a possible-worlds semantics, and to evaluate their truth in a model by focusing on a subset of the set of possible worlds within a model. This is similar to the focusing on certain models of a theory in preferential entailment. On the other hand, intensional semantics capture the defeasibility of defaults not with extra normality conditions, but in the meaning of the conditional operator. This development draws its inspiration from the similarity semantics for counterfactuals in conditional logics, e.g. [Lewis, 1973]. In these logics a counterfactual conditional is interpreted as follows:  $\varphi \Rightarrow \psi$  is true just in case  $\psi$  is true in a subset of the possible worlds in which  $\varphi$  is true, viz. in the possible worlds which resemble the actual world as much as possible, given that in them  $\varphi$  holds. Now with respect to defeasible conditionals the idea is to define in a similar way a possible-worlds semantics for defeasible conditionals. A defeasible conditional  $\varphi \Rightarrow \psi$  is roughly interpreted as 'in all most normal worlds in which  $\varphi$  holds,  $\psi$  holds as well'. Obviously, if read in this way, then modus ponens is not valid for such conditionals, since even if  $\varphi$  holds in the actual world, the actual world need not be a normal world. This is different for counterfactual conditionals, where the actual world is always among the worlds most similar to itself. This difference makes that intensional defeasible logics need a component that is absent in counterfactual logics, and which is similar to the selection of the 'most normal' models in preferential entailment: in order to derive default conclusions from defeasible conditionals, the actual world is assumed to be as normal as possible given the premises. It is this assumption that makes the resulting conclusions defeasible: it validates modus ponens for those

defaults for which there is no evidence of exceptions.

### *Consistency and non-provability statements*

Yet another approach is to somehow make the expression possible of consistency or non-provability statements. This is, for instance, the idea behind Reiter's [1980] *default logic*, which extends first-order logic with constructs that technically play the role of inference rules, but that express domain-specific generalisations instead of logical inference principles. In default logic, the Tweety default can be written as follows.

$$\text{Bird}(x) : \text{Canfly}(x) / \text{Canfly}(x)$$

The middle part of this 'default' can be used to express consistency statements. Informally the default reads as 'If it is provable that Tweety is a bird, and it is not provable that Tweety cannot fly, then we may infer that Tweety can fly'. To see how this works, assume that in addition to this default we have a first-order theory

$$W = \{\text{Bird}(\text{Tweety}), \forall x. \text{Penguin}(x) \supset \neg \text{Canfly}(x)\}$$

Then (informally) since  $\text{Canfly}(\text{Tweety})$  is consistent with what is known, we can apply the default to *Tweety* and defeasibly derive  $\text{Canfly}(\text{Tweety})$  from  $W$ . That this inference is indeed defeasible becomes apparent if  $\text{Penguin}(\text{Tweety})$  is also added to  $W$ : then  $\neg \text{Canfly}(\text{Tweety})$  is classically entailed by what is known and the consistency check for applying the default fails, for which reason  $\text{Canfly}(\text{Tweety})$  cannot be derived from  $W \cup \{\text{Penguin}(\text{Tweety})\}$ .

This example seems straightforward but the formal definition of default-logical consequence is tricky: in this approach, what is provable is determined by what is not provable, so the problem is how to avoid a circular definition. In default logic (as in related logics) this is solved by giving the definition a fixed-point appearance; see below in Section 5.4. Similar equilibrium-like definitions for argumentation systems will be discussed throughout this chapter.

### *Inconsistency handling*

It has also been proposed to formalise defeasible reasoning as strategies for dealing with inconsistent information, e.g. by Brewka [1989]. In this approach defaults are formalised with ordinary material implications and without normality conditions, and their defeasible nature is captured in how they are used by the consistency handling strategies. In particular, in case of inconsistency, alternative consistent subsets (subtheories) of the premises give rise to alternative default conclusions, after which a choice can be made for the subtheory containing the exceptional rule.

In our birds example this works out as follows.

- (1)  $\text{bird} \supset \text{canfly}$
- (2)  $\text{penguin} \supset \neg \text{canfly}$
- (3)  $\text{bird}$
- (4)  $\text{penguin}$

The set  $\{(1), (3)\}$  is a subtheory supporting the conclusion  $\text{canfly}$ , while  $\{(2), (4)\}$  is a subtheory supporting the opposite. The exceptional nature of (2) over (1) can be captured by preferring the latter subtheory.

### *Systems for defeasible argumentation*

Argumentation systems are yet another way to formalise nonmonotonic reasoning, viz. as the construction and comparison of arguments for and against certain conclusions. In these systems the basic notion is not that of a defeasible conditional but that of a defeasible argument. The idea is that the construction of arguments is monotonic, i.e., an argument stays an argument if more premises are added. Nonmonotonicity, or defeasibility, is not explained in terms of the interpretation of a defeasible conditional, but in terms of the interactions between conflicting arguments: in argumentation systems nonmonotonicity arises from the fact that new premises may give rise to stronger counterarguments, which defeat the original argument. So in case of Tweety we may construct one argument that Tweety flies because it is a bird, and another argument that Tweety does not fly because it is a penguin, and then we may prefer the latter argument because it is about a specific class of birds, and is therefore an exception to the general rule.

Argumentation systems can be combined with each of the above-discussed views on defaults. The ‘normality condition’ view can be formalised by regarding an argument as a standard derivation from a set of premises augmented with normality statements. Thus a counterargument is an attack on such a normality statement. A variant of this method can be applied to the use of consistency and nonprovability expressions. The ‘pragmatic’ view on defaults (as in inconsistency handling) can be formalised by regarding arguments as a standard derivation from a consistent subset of the premises. Here a counterargument attacks a premise of an argument. Finally, the ‘semantic’ view on defaults could be formalised by allowing the construction of arguments with inference rules (such as modus ponens) that are invalid in the semantics. In that case a counterargument attacks the use of such an inference rule.

It is important to note, however, that argumentation systems have wider scope than just reasoning with defaults. Firstly, argumentation systems can be applied to any form of reasoning with contradictory information, whether the contradictions have to do with rules and exceptions or not. For instance, the contradictions may arise from reasoning with several sources

of information, or they may be caused by disagreement about beliefs or about moral, ethical or political claims. Moreover, it is important that several argumentation systems allow the construction and attack of arguments that are traditionally called ‘ampliative’, such as inductive, analogical and abductive arguments; these reasoning forms fall outside the scope of most other nonmonotonic logics.

Most argumentation systems have been developed in artificial intelligence research on nonmonotonic reasoning, although Pollock’s work, which was the first logical formalisation of defeasible argumentation, was initially applied to the philosophy of knowledge and justification (epistemology). The first artificial intelligence paper on argumentation systems was [Loui, 1987]. One domain in which argumentation systems have become popular is legal reasoning [Loui *et al.*, 1993; Prakken, 1993; Sartor, 1994; Gordon, 1995; Loui & Norman, 1995; Prakken & Sartor, 1996; Freeman & Farley, 1996; Prakken & Sartor, 1997a; Prakken, 1997; Gordon & Karacapilidis, 1997]. This is not surprising, since legal reasoning often takes place in an adversarial context, where notions like argument, counterargument, rebuttal and defeat are very common. However, argumentation systems have also been applied to such domains as medical reasoning [Das *et al.*, 1996], negotiation [Parsons *et al.*, 1998] and risk assessment in oil exploration [Clark, 1990].

## 2.2 *Nonmonotonic reasoning: artificial intelligence or logic?*

Usually, nonmonotonic logics are studied as a branch of artificial intelligence. However, it is more than justified to regard these logics as also part of philosophical logic. In fact, several issues in nonmonotonic logic have come up earlier in philosophy. For instance, in the context of moral reasoning, Ross [1930] has studied the notion of *prima facie* obligations. According to Ross an act is *prima facie* obligatory if it has a characteristic that makes the act (by virtue of an underlying moral principle) *tend* to be a ‘duty proper’. Fulfilling a promise is a *prima facie* duty because it is the fulfillment of a promise, i.e., because of the moral principle that one should do what one has promised to do. But the act may also have other characteristics which make the act tend to be forbidden. For instance, if John has promised a friend to visit him for a cup of tea, and then John’s mother suddenly falls ill, then he also has a *prima facie* duty to do his mother’s shopping, based, say, on the principle that we ought to help our parents when they need it. To find out what one’s duty proper is, one should ‘consider all things’, i.e., compare all *prima facie* duties that can be based on any aspect of the factual circumstances and find which one is ‘more incumbent’ than any conflicting one. If we qualify the all-things-considered clause as ‘consider all things that you know’, then the reasoning involved is clearly nonmonotonic: if we are first only told that John has promised his friend to visit him, then we conclude that John’s duty proper is to visit his friend. But if we next also

hear that John's mother has become ill, we conclude instead that John's duty proper is to help his mother.

The term 'defeasibility' was first introduced not in logic but in legal philosophy, viz. by Hart [1949] (see the historical discussion in [Loui, 1995]). Hart observed that legal concepts are defeasible in the sense that the conditions for when a fact situation classifies as an instance of a legal concept (such as 'contract'), are only ordinarily, or presumptively, sufficient. If a party in a law suit succeeds in proving these conditions, this does not have the effect that the case is settled; instead, legal procedure is such that the burden of proof shifts to the opponent, whose turn it then is to prove additional facts which, despite the facts proven by the proponent, nevertheless prevent the claim from being granted (for instance, insanity of one of the contracting parties). Hart's discussion of this phenomenon stays within legal-procedural terms, but it is obvious that it provides a challenge for standard logic: an explanation is needed of how proving new facts without rejecting what was proven by the other party can reverse the outcome of a case.

Toulmin [1958], who criticised the logicians of his days for neglecting many features of ordinary reasoning, was aware of the implications of this phenomenon for logic. In his well-known pictorial scheme for arguments he leaves room for rebuttals of an argument. He also urges logicians to take the *procedural* aspect (in the legal sense) of argumentation seriously. In particular, Toulmin argues that (outside mathematics) an argument is valid if it can stand against criticism in a properly conducted dispute, and the task of logicians is to find criteria for when a dispute has been conducted properly.

The notion of burden of proof, and its role in dialectical inquiry, has also been studied by Rescher [1977], in the context of epistemology. Among other things, Rescher claims that a dialectical model of scientific reasoning can explain the rational force of inductive arguments: they must be accepted if they cannot be successfully challenged in a properly conducted scientific dispute. Rescher thereby assumes that the standards for *constructing* inductive arguments are somehow given by generally accepted practices of scientific reasoning; he only focuses on the dialectical interaction between conflicting inductive arguments.

Another philosopher who has studied defeasible reasoning is John Pollock. Although his work, to be presented below, is also well-known in the field of artificial intelligence, it was initially a contribution to epistemology, with, like Rescher, much attention for induction as a form of defeasible reasoning.

As this overview shows, a logical study of nonmonotonic, or defeasible reasoning fully deserves a place in philosophical logic. Let us now turn to the discussion of logics for defeasible argumentation.

### 3 SYSTEMS FOR DEFEASIBLE ARGUMENTATION: A CONCEPTUAL SKETCH

In this section we give a conceptual sketch of the general ideas behind logics for defeasible argumentation. These systems contain the following five elements (although sometimes implicitly): an underlying logical language, definitions of an argument, of conflicts between arguments and of defeat among arguments and, finally, a definition of the status of arguments, which can be used to define a notion of defeasible logical consequence.

Argumentation systems are built around an underlying logical language and an associated notion of logical consequence, defining the notion of an argument. As noted above, the idea is that this consequence notion is monotonic: new premises cannot invalidate arguments as arguments but only give rise to counterarguments. Some argumentation systems assume a particular logic, while other systems leave the underlying logic partly or wholly unspecified; thus these systems can be instantiated with various alternative logics, which makes them frameworks rather than systems. The notion of an argument corresponds to a proof (or the existence of a proof) in the underlying logic. As for the layout of arguments, in the literature on argumentation systems three basic formats can be distinguished, all familiar from the logic literature. Sometimes arguments are defined as a tree of inferences grounded in the premises, and sometimes as a sequence of such inferences, i.e., as a deduction. Finally, some systems simply define an argument as a premises - conclusion pair, leaving implicit that the underlying logic validates a proof of the conclusion from the premises. One argumentation system, viz. Dung [1995], leaves the internal structure of an argument completely unspecified. Dung treats the notion of an argument as a primitive, and exclusively focuses on the ways arguments interact. Thus Dung's framework is of the most abstract kind.

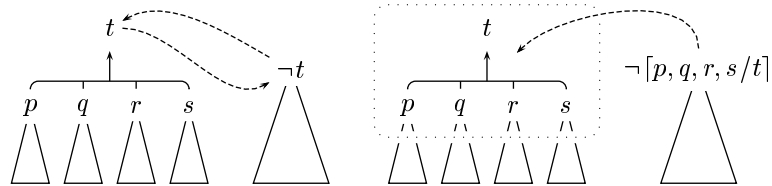


Figure 4. Rebutting attack (left) vs. undercutting attack (right).

The notions of an underlying logic and an argument still fit with the standard picture of what a logical system is. The remaining three elements are what makes an argumentation system a framework for defeasible argumentation.

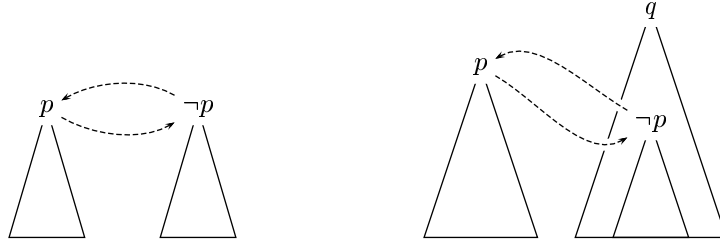


Figure 5. Direct attack (left) vs. indirect attack (right).

The first is the notion of a *conflict* between arguments (also used are the terms ‘attack’ and ‘counterargument’). In the literature, three types of conflicts are discussed. The first type is when arguments have contradictory conclusions, as in ‘Tweety flies, because it is a bird’ and ‘Tweety does not fly because it is a penguin’ (cf. the left part of Fig. 4). Clearly, this form of attack, which is often called *rebutting* an argument, is symmetric. The other two types of conflict are not symmetric. One is where one argument makes a non-provability assumption (as in default logic) and another argument proves what was assumed unprovable by the first. For example, an argument ‘Tweety flies because it is a bird, and it is not provable that Tweety is a penguin’, is attacked by any argument with conclusion ‘Tweety is a penguin’. We shall call this *assumption attack*. The final type of conflict (first discussed by Pollock [1970]) is when one argument challenges, not a proposition, but a rule of inference of another argument (cf. the right part of Fig. 4). After Pollock, this is usually called *undercutting* an inference. Obviously, a rule of inference can only be undercut if it is not deductive. Non-deductive rules of inference occur in argumentation systems that allow inductive, abductive or analogical arguments. To consider an example, the inductive argument ‘Raven<sub>101</sub> is black since the observed ravens raven<sub>1</sub> ... raven<sub>100</sub> were black’ is undercut by an argument ‘I saw raven<sub>102</sub>, which was white’. In order to formalise this type of conflict, the rule of inference that is to be undercut (in Fig. 4: the rule that is enclosed in the dotted box, in flat text written as  $p, q, r, s/t$ ) must be expressed in the object language:  $[p, q, r, s/t]$  and denied:  $\neg[p, q, r, s/t]$ .<sup>3</sup>

Note that all these senses of attack have a direct and an indirect version; indirect attack is directed against a subconclusion or a substep of an argument, as illustrated by Figure 5 for indirect rebutting.

The notion of conflicting, or attacking arguments does not embody any form of evaluation; evaluating conflicting pairs of arguments, or in other

<sup>3</sup>Ceiling brackets around a meta-level formula denote a conversion of that formula to the object language, provided that the object language is expressive enough to enable such a conversion.



words, determining whether an attack is successful, is another element of argumentation systems. It has the form of a binary relation between arguments, standing for ‘attacking and not weaker’ (in a weak form) or ‘attacking and stronger’ (in a strong form). The terminology varies: some terms that have been used are ‘defeat’ [Prakken & Sartor, 1997b], ‘attack’ [Dung, 1995; Bondarenko *et al.*, 1997] and ‘interference’ [Loui, 1998]. Other systems do not explicitly name this notion but leave it implicit in the definitions. In this chapter we shall use ‘defeat’ for the weak notion and ‘strict defeat’ for the strong, asymmetric notion. Note that the several forms of attack, rebutting vs. assumption vs. undercutting and direct vs. indirect, have their counterparts for defeat.

Argumentation systems vary in their grounds for the evaluation of arguments. In artificial intelligence the specificity principle, which prefers arguments based on the most specific defaults, is by many regarded as very important, but several researchers, e.g. Vreeswijk [1989], Pollock [1995] and Prakken & Sartor [1996], have argued that specificity is not a general principle of common-sense reasoning but just one of the many standards that might or might not be used. Moreover, some have claimed that general, domain-independent principles of defeat do not exist or are very weak, and that information from the semantics of the domain will be the most important way of deciding among competing arguments [Konolige, 1988; Vreeswijk, 1989]. For these reasons several argumentation systems are parametrised by user-provided criteria. Some, e.g. Prakken & Sartor, even argue that the evaluation criteria are debatable, just as the rest of the domain theory is, and that argumentation systems should therefore allow for defeasible arguments on these criteria. (Our example in the introduction contains such an argument, viz. *A*’s use of a priority rule (10) based on the expected consequences of certain events. This argument might, for instance, be attacked by an argument that in case of important officials even a small likelihood that the disease affects the official’s functioning justifies publication, or by an argument that the negative consequences of publication for the official are small.)

The notion of defeat is a binary relation on the set of arguments. It is important to note that this relation does not yet tell us with what arguments a dispute can be won; it only tells us something about the relative strength of two individual conflicting arguments. The ultimate status of an argument depends on the interaction between all available arguments: it may very well be that argument *B* defeats argument *A*, but that *B* is in turn defeated by a third argument *C*; in that case *C* ‘reinstates’ *A* (see Figure 6).<sup>4</sup> Suppose, for instance, that the argument *A* that Tweety flies because it is a bird is regarded as being defeated by the argument *B* that

---

<sup>4</sup>While in figures 4 and 5 the arrows stood for attack relations, from now on they will depict defeat relations.

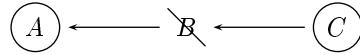


Figure 6. Argument  $C$  reinstates argument  $A$ .

Tweety does not fly because it is a penguin (for instance, because conflicting arguments are compared with respect to specificity). And suppose that  $B$  is in turn defeated by an argument  $C$ , attacking  $B$ 's intermediate conclusion that Tweety is a penguin.  $C$  might, for instance, say that the penguin observation was done with faulty instruments. In that case  $C$  reinstates argument  $A$ .

Therefore, what is also needed is a definition of the status of arguments on the basis of all the ways in which they interact. Besides reinstatement, this definition must also capture the principle that an argument cannot be justified unless all its subarguments are justified (by Vreeswijk [1997] called the 'compositionality principle'). There is a close relation between these two notions, since reinstatement often proceeds by indirect attack, i.e., attacking a subargument of the attacking argument. (Cf. Fig. 5 on page 231.) It is this definition of the status of arguments that produces the output of an argumentation system: it typically divides arguments in at least two classes: arguments with which a dispute can be 'won' and arguments with which a dispute should be 'lost'. Sometimes a third, intermediate category is also distinguished, of arguments that leave the dispute undecided. The terminology varies here also: terms that have been used are justified vs. defensible vs. defeated (or overruled), defeated vs. undefeated, in force vs. not in force, preferred vs. not preferred, etcetera. Unless indicated otherwise, this chapter shall use the terms 'justified', 'defensible' and 'overruled' arguments.

These notions can be defined both in a 'declarative' and in a 'procedural' form. The declarative form, usually with fixed-point definitions, just declares certain sets of arguments as acceptable, (given a set of premises and evaluation criteria) without defining a procedure for testing whether an argument is a member of this set; the procedural form amounts to defining just such a procedure. Thus the declarative form of an argumentation system can be regarded as its (argumentation-theoretic) semantics, and the procedural form as its proof theory. Note that it is very well possible that, while an argumentation system has an argumentation-theoretic semantics, at the same time its underlying logic for constructing arguments has a model-theoretic semantics in the usual sense, for instance, the semantics of standard first-order logic, or a possible-worlds semantics of some modal logic.

In fact, this point is not universally accepted, and therefore we devote a separate subsection to it.

*Semantics: model-theoretic or not?*

A much-discussed issue is whether logics for nonmonotonic reasoning should have a model-theoretic semantics or not. In the early days of this field it was usual to criticise several systems (such as default logic) for the lack of a model-theoretic semantics. However, when such semantics were provided, this was not always felt to be a major step forward, unlike when, for instance, possible-worlds semantics for modal logic was introduced. In addition, several researchers argued that nonmonotonic reasoning needs a different kind of semantics than a model theory, viz. an argumentation-theoretic semantics. It is here not the place to decide the discussion. Instead we confine ourselves to presenting some main arguments for this view that have been put forward.

Traditionally, model theory has been used in logic to define the meaning of logical languages. Formulas of such languages were regarded as telling us something about reality (however defined). Model-theoretic semantics defines the meaning of logical symbols by defining how the world looks like if an expression with these symbols is true, and it defines logical consequence, entailment, by looking at what else must be true if the premises are true. For defaults this means that their semantics should be in terms of how the world normally, or typically looks like when defaults are true; logical consequence should, in this approach, be determined by looking at the most normal worlds, models or situations that satisfy the premises.

However, others, e.g. Pollock [1991, p. 40], Vreeswijk [1993a, pp. 88–9] and Loui [1998], have argued that the meaning of defaults should not be found in a correspondence with reality, but in their role in dialectical inquiry. That a relation between premises and conclusion is defeasible means that a certain burden of proof is induced. In this approach, the central notions of defeasible reasoning are notions like attack, rebuttal and defeat among arguments, and these notions are not ‘propositional’, for which reason their meaning is not naturally captured in terms of correspondence between a proposition and the world. This approach instead defines ‘argumentation-theoretic’ semantics for such notions. The basic idea of such a semantics is to capture sets of arguments that are as large as possible, and adequately defend themselves against attacks on their members.

It should be noted that this approach does not deny the usefulness of model theory but only wants to define its proper place. Model theory should not be applied for things for which it is not suitable, but should be reserved for the initial components of an argumentation system, the notions of a logical language and a consequence relation defining what an argument is.

It should also be noted, however, that some have proposed argumentation systems as proof theories for model-theoretic semantics of preferential entailment (in particular Geffner & Pearl [1992]). In our opinion, one criterion for success of such model-theoretic semantics of argumentation systems

is whether *natural* criteria for model preference can be defined. For certain restricted cases this seems possible, but whether this approach is extendable to more general argumentation systems, for instance, those allowing inductive, analogical or abductive arguments, remains to be investigated.

#### 4 GENERAL FEATURES OF ARGUMENT-BASED SEMANTICS

Let us now, before looking at some systems in detail, become more formal about some of the notions that these systems have in common. We shall focus in particular on the semantics of argumentation systems, i.e., on the conditions that sets of justified arguments should satisfy. In line with the discussion at the end of Section 3, we can say that argumentation systems are not concerned with truth of propositions, but with justification of accepting a proposition as true. In particular, one is justified in accepting a proposition as true if there is an argument for the proposition that one is justified in accepting. Let us concentrate on the task of defining the notion of a justified argument. Which properties should such a definition have?

Let us assume as background a set of arguments, with a binary relation of ‘defeat’ defined over it. Recall that we read ‘ $A$  defeats  $B$ ’ in the weak sense of ‘ $A$  conflicts with  $B$  and is not weaker than  $B$ ’; so in some cases it may happen that  $A$  defeats  $B$  and  $B$  defeats  $A$ . For the moment we leave the internal structure of an argument unspecified, as well as the precise definition of defeat.<sup>5</sup> Then a simple definition of the status of an argument is the following.

DEFINITION 1. Arguments are either justified or not justified.

1. An argument is *justified* if all arguments defeating it (if any) are not justified.
2. An argument is *not justified* if it is defeated by an argument that is justified.

This definition works well in simple cases, in which it is clear which arguments should emerge victorious, as in the following example.

EXAMPLE 2. Consider three arguments  $A$ ,  $B$  and  $C$  such that  $B$  defeats  $A$  and  $C$  defeats  $B$ :

$$A \longleftarrow B \longleftarrow C$$

---

<sup>5</sup>This style of discussion is inspired by Dung [1995]; see further Subsection 5.1 below.

A concrete version of this example is

$A =$  ‘Tweety flies because it is a bird’

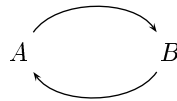
$B =$  ‘Tweety does not fly because it is a penguin’

$C =$  ‘The observation that Tweety is a penguin is unreliable’

$C$  is justified since it is not defeated by any other argument. This makes  $B$  not justified, since  $B$  is defeated by  $C$ . This in turn makes  $A$  justified: although  $A$  is defeated by  $B$ ,  $A$  is reinstated by  $C$ , since  $C$  makes  $B$  not justified.

In other cases, however, Definition 1 is circular or ambiguous. Especially when arguments of equal strength interfere with each other, it is not clear which argument should remain undefeated.

EXAMPLE 3. (Even cycle.) Consider the arguments  $A$  and  $B$  such that  $A$  defeats  $B$  and  $B$  defeats  $A$ .



A concrete example is

$A =$  ‘Nixon was a pacifist because he was a quaker’

$B =$  ‘Nixon was not a pacifist because he was a republican’

Can we regard  $A$  as justified? Yes, we can, if  $B$  is not justified. Can we regard  $B$  as not justified? Yes, we can, if  $A$  is justified. So, if we regard  $A$  as justified and  $B$  as not justified, Definition 1 is satisfied. However, it is obvious that by a completely symmetrical line of reasoning we can also regard  $B$  as justified and  $A$  as not justified. So there are two possible ‘status assignments’ to  $A$  and  $B$  that satisfy Definition 1: one in which  $A$  is justified at the expense of  $B$ , and one in which  $B$  is justified at the expense of  $A$ . Yet intuitively, we are not justified in accepting either of them.

In the literature, two approaches to the solution of this problem can be found. The first approach consists of changing Definition 1 in such a way that there is always precisely one possible way to assign a status to arguments, and which is such that with ‘undecided conflicts’ as in our example both of the conflicting arguments receive the status ‘not justified’. The second approach instead regards the existence of multiple status assignments not as a problem but as a feature: it allows for multiple assignments and defines an argument as ‘genuinely’ justified if and only if it receives this status in all possible assignments. The following two subsections discuss the details of both approaches.

First, however, another problem with Definition 1 must be explained, having to do with self-defeating arguments.

EXAMPLE 4. (Self-defeat.) Consider an argument  $L$ , such that  $L$  defeats  $L$ . Suppose  $L$  is not justified. Then all arguments defeating  $L$  are not



Figure 7. A self-defeating argument.

justified, so by clause 1 of Definition 1  $L$  is justified. Contradiction. Suppose now  $L$  is justified. Then  $L$  is defeated by a justified argument, so by clause 2 of Definition 1  $L$  is not justified. Contradiction.

Thus, Definition 1 implies that there are no self-defeating arguments. Yet the notion of self-defeating arguments seems intuitively plausible, as is illustrated by the following example.

EXAMPLE 5. (The Liar.) An elementary self-defeating argument can be fabricated on the basis of the so-called *paradox of the Liar*. There are many versions of this paradox. The one we use here, runs as follows:

Dutch people can be divided into two classes: people who always tell the truth, and people who always lie. Hendrik is a Dutch monk, and of Dutch monks we know that they tend to be consistent truth-tellers. Therefore, it is reasonable to assume that Hendrik is a consistent truth-teller. However, Hendrik *says* he is a liar. Is Hendrik a truth-teller or a liar?

The Liar-paradox is a paradox, because either answer leads to a contradiction.

1. Suppose that Hendrik tells the truth. Then what Hendrik says must be true. So, Hendrik is a liar. Contradiction.
2. Suppose that Hendrik lies. Then what Hendrik says must be false. So, Hendrik is not a liar. Because Dutch people are either consistent truth-tellers or consistent liars, it follows that Hendrik always tells the truth. Contradiction.

From this paradox, a self-defeating argument  $L$  can be made out of (1):

	Dutch monks tend to be consistent truth-tellers	Hendrik is a Dutch monk
	<hr/>	Hendrik is a consistent truth-teller
Hendrik says: “I lie”	<hr/>	
	Hendrik lies	
	<hr/>	
	Hendrik is <b>not</b> a consistent truth-teller	

If the argument for “Hendrik is *not* a consistent truth-teller” is as strong as its subargument for “Hendrik is a consistent truth-teller,” then  $L$  defeats one of its own sub-arguments, and thus is a self-defeating argument.

In conclusion, it seems that Definition 1 needs another revision, to leave room for the existence of self-defeating arguments. Below we shall not discuss this in general terms since, perhaps surprisingly, in the literature it is hard to find generally applicable solutions to this problem. Instead we shall discuss for each particular system how it deals with self-defeat.

#### 4.1 *The unique-status-assignment approach*

The idea to enforce unique status assignments basically comes in two variants. The first defines status assignments in terms of some fixed-point operator, and the second involves a recursive definition of a justified argument, by introducing the notion of a subargument of an argument. We first discuss the fixed-point approach.

##### *Fixed-point definitions*

This approach, followed by e.g. Pollock [1987; 1992], Simari & Loui [1992] and Prakken & Sartor [1997b], can best be explained with the notion of ‘reinstatement’ (see above, Section 3). The key observation is that an argument that is defeated by another argument can only be justified if it is reinstated by a third argument, viz. by a justified argument that defeats its defeater. This idea is captured by Dung’s [1995] notion of *acceptability*.

**DEFINITION 6.** An argument  $A$  is *acceptable* with respect to a set  $S$  of arguments iff each argument defeating  $A$  is defeated by an argument in  $S$ .

The arguments in  $S$  can be seen as the arguments capable of reinstating  $A$  in case  $A$  is defeated.

However, the notion of acceptability is not sufficient. Consider in Example 3 the set  $S = \{A\}$ . It is easy to see that  $A$  is acceptable with respect to  $S$ , since all arguments defeating  $A$  (viz.  $B$ ) are defeated by an argument in  $S$ , viz.  $A$  itself. Clearly, we do not want that an argument can re-instate itself, and this is the reason why a fixed-point operator must be used. Consider the following operator from [Dung, 1995], which for each set of arguments returns the set of all arguments that are acceptable to it.

DEFINITION 7. (Dung's [1995] grounded semantics.) Let  $Args$  be a set of arguments ordered by a binary relation of defeat,<sup>6</sup> and let  $S \subseteq Args$ . Then the operator  $F$  is defined as follows:

- $F(S) = \{A \in Args \mid A \text{ is acceptable with respect to } S\}$

Dung proves that the operator  $F$  has a least fixed point. (The basic idea is that if an argument is acceptable with respect to  $S$ , it is also acceptable with respect to any superset of  $S$ , so that  $F$  is monotonic.) Self-reinstatement can then be avoided by defining the set of justified arguments as that least fixed point. Note that in Example 3 the sets  $\{A\}$  and  $\{B\}$  are fixed points of  $F$  but not its least fixed point, which is the empty set. In general we have that if no argument is undefeated, then  $F(\emptyset) = \emptyset$ .

These observations allow the following definition of a justified argument.

DEFINITION 8. An argument is *justified* iff it is a member of the least fixed point of  $F$ .

It is possible to reformulate Definition 7 in various ways, which are either equivalent to, or approximations of the least fixed point of  $F$ . To start with, Dung shows that it can be approximated from below, and when each argument has at most finitely many defeaters even be obtained, by iterative application of  $F$  to the empty set.

PROPOSITION 9. Consider the following sequence of arguments.

- $F^0 = \emptyset$
- $F^{i+1} = \{A \in Args \mid A \text{ is acceptable with respect to } F^i\}$ .

The following observations hold [Dung, 1995].

1. All arguments in  $\cup_{i=0}^{\infty} (F^i)$  are justified.
2. If each argument is defeated by at most a finite number of arguments, then an argument is justified iff it is in  $\cup_{i=0}^{\infty} (F^i)$

---

<sup>6</sup>As remarked above, Dung uses the term 'attack' instead of 'defeat'.



In the iterative construction first all arguments that are not defeated by any argument are added, and at each further application of  $F$  all arguments that are reinstated by arguments that are already in the set are added. This is achieved through the notion of acceptability. To see this, suppose we apply  $F$  for the  $i$ th time: then for any argument  $A$ , if all arguments that defeat  $A$  are themselves defeated by an argument in  $F^{i-1}$ , then  $A$  is in  $F^i$ .

It is instructive to see how this works in Example 2. We have that

$$\begin{aligned} F^1 &= F(\emptyset) = \{C\} \\ F^2 &= F(F^1) = \{A, C\} \\ F^3 &= F(F^2) = F^2 \end{aligned}$$

Dung [1995] also shows that  $F$  is equivalent to double application of a simpler operator  $G$ , i.e.  $F = G \circ G$ . The operator  $G$  returns for each set of arguments all arguments that are not defeated by any argument in that set.

DEFINITION 10. Let  $Args$  be a set of arguments ordered by a binary relation of defeat. Then the operator  $G$  is defined as follows:

- $G(S) = \{A \in Args \mid A \text{ is not defeated by any argument in } S\}$

The  $G$  operator is in turn very similar to the one used by Pollock [1987; 1992]. To see this, we reformulate  $G$  in Pollock's style, by considering the sequence obtained by iterative application of  $G$  to the empty set, and defining an argument  $A$  to be justified if and only if at some point (or "level")  $m$  in the sequence  $A$  remains in  $G_n$  for all  $n \geq m$ .

DEFINITION 11. (Levels in justification.)

- All arguments are *in at level 0*.
- An argument is *in at level  $n + 1$*  iff it is not defeated by any argument in at level  $n$ .
- An argument is *justified* iff there is an  $m$  such that for every  $n \geq m$ , the argument is in at level  $n$ .

As shown by Dung [1995], this definition stands to Definition 10 as the construction of Proposition 9 stands to Definition 7. Dung also remarks that Definition 11 is equivalent to Pollock's [1987; 1992] definition, but as we shall see below, this is not completely accurate.

In Example 2, Definition 11 works out as follows.

level	in
0	$A, B, C$
1	$C$
2	$A, C$
3	$A, C$
.	...

$C$  is in at all levels, while  $A$  becomes in at 2 and stays in at all subsequent levels.

And in Example 3 both  $A$  and  $B$  are in at all even levels and out at all odd levels.

level	in
0	$A, B$
1	
2	$A, B$
3	
4	$A, B$
.	...

The following example, with an infinite chain of defeat relations, gives another illustration of Definitions 7 and 11.

EXAMPLE 12. (Infinite defeat chain.) Consider an infinite chain of arguments  $A_1, \dots, A_n, \dots$  such that  $A_1$  is defeated by  $A_2$ ,  $A_2$  is defeated by  $A_3$ , and so on.

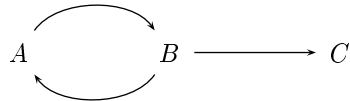
$$A_1 \longleftarrow A_2 \longleftarrow A_3 \longleftarrow A_4 \longleftarrow A_5 \longleftarrow \dots$$

The least fixed point of this chain is empty, since no argument is undefeated. Consequently,  $F(\emptyset) = \emptyset$ . Note that this example has two other fixed points, which also satisfy Definition 1, viz. the set of all  $A_i$  where  $i$  is odd, and the set of all  $A_i$  where  $i$  is even.

### *Defensible arguments*

A final peculiarity of the definitions is that they allow a distinction between two types of arguments that are not justified. Consider first again Example 2 and observe that, although  $B$  defeats  $A$ ,  $A$  is still justified since it is reinstated by  $C$ . Consider next the following extension of Example 3.

EXAMPLE 13. (Zombie arguments.) Consider three arguments  $A$ ,  $B$  and  $C$  such that  $A$  defeats  $B$ ,  $B$  defeats  $A$ , and  $B$  defeats  $C$ .



A concrete example is

- $A =$  ‘Dixon is no pacifist because he is a republican’  
 $B =$  ‘Dixon is a pacifist because he is a quaker, and he has no gun  
because he is a pacifist’  
 $C =$  ‘Dixon has a gun because he lives in Chicago’

According to Definitions 8 and 11, neither of the three arguments are justified. For  $A$  and  $B$  this is since their relation is the same as in Example 3, and for  $C$  this is since it is defeated by  $B$ . Here a crucial distinction between the two examples becomes apparent: unlike in Example 2,  $B$  is, although not justified, not defeated by any justified argument and therefore  $B$  retains the potential to prevent  $C$  from becoming justified: there is no justified argument that reinstates  $C$  by defeating  $B$ . Makinson & Schlechta [1991] call arguments like  $B$  ‘zombie arguments’:<sup>7</sup>  $B$  is not ‘alive’, (i.e., not justified) but it is not fully dead either; it has an intermediate status, in which it can still influence the status of other arguments. Following Prakken & Sartor [1997b], we shall call this intermediate status ‘defensible’. In the unique-status-assignment approach it can be defined as follows.

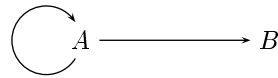
DEFINITION 14. (Overruled and defensible arguments.)

- An argument is *overruled* iff it is not justified, and defeated by a justified argument.
- An argument is *defensible* iff it is not justified and not overruled.

#### *Self-defeating arguments*

Finally, we must come back to the problem of self-defeating arguments. How does Definition 7 deal with them? Consider the following extension of Example 4.

EXAMPLE 15. Consider two arguments  $A$  and  $B$  such that  $A$  defeats  $A$  and  $A$  defeats  $B$ .



Intuitively, we want that  $B$  is justified, since the only argument defeating it is self-defeating. However, we have that  $F(\emptyset) = \emptyset$ , so neither  $A$  nor  $B$  are justified. Moreover, they are both defensible, since they are not defeated by any justified argument.

How can Definitions 7 and 11 be modified to obtain the intuitive result that  $A$  is overruled and  $B$  is justified? Here is where Pollock’s deviation from the latter definition becomes relevant. His version is as follows.

<sup>7</sup>Actually, they talk about ‘zombie paths’, since their article is about inheritance systems.

DEFINITION 16. (Pollock, [1992])

- An argument is *in at level 0* iff it is not self-defeating.
- An argument is *in at level  $n + 1$*  iff it is in at level 0 and it is not defeated by any argument in at level  $n$ .
- An argument is *justified* iff there is an  $m$  such that for every  $n \geq m$ , the argument is in at level  $n$ .

The additions *iff it is not self-defeating* in the first condition and *iff it is in at level 0* in the second make the difference: they render all self-defeating arguments out at every level, and incapable of preventing other arguments from being out.

Another solution is provided by Prakken & Sartor [1997b] and Vreeswijk [1997], who distinguish a special ‘empty’ argument, which is not defeated by any other argument and which by definition defeats any self-defeating argument. Other solutions are possible, but we shall not pursue them here.

### *Recursive definitions*

Sometimes a second approach to the enforcement of unique status assignments is employed, e.g. by Prakken [1993] and Nute [1994]. The idea is to make explicit that arguments are usually constructed step-by-step, proceeding from intermediate to final conclusions (as in Example 13, where  $A$  has an intermediate conclusion ‘Dixon is a pacifist’ and a final conclusion ‘Dixon has no gun’). This approach results in an explicitly recursive definition of justified arguments, reflecting the basic intuition that an argument cannot be justified if not all its subarguments are justified. At first sight, this recursive style is very natural, particularly for implementing the definition in a computer program. However, the approach is not so straightforward as it seems, as the following discussion aims to show.

To formalise the recursive approach, we must make a first assumption on the structure of arguments, viz. that they have subarguments (which are ‘proper’ iff they are not identical to the entire argument). Justified arguments are then defined as follows. (We already add how self-defeating arguments can be dealt with, so that our discussion can be confined to the issue of avoiding multiple status assignments. Note that the explicit notion of a subargument makes it possible to regard an argument as self-defeating if it defeats one of its subarguments, as in Example 5.)

DEFINITION 17. (Recursively justified arguments.) An argument  $A$  is *justified* iff

1.  $A$  is not self-defeating; and
2. All proper subarguments of  $A$  are justified; and

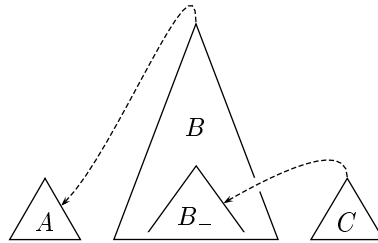
3. All arguments defeating  $A$  are self-defeating, or have at least one proper subargument that is not justified.

How does this definition avoid multiple status assignments in Example 3? The ‘trick’ is that for an argument to be justified, clause (2) requires that it have no (non self-defeating) defeaters of which all proper subarguments are justified. This is different in Definition 1, which leaves room for such defeaters, and instead requires that these themselves are not justified; thus this definition implies in Example 3 that  $A$  is justified if and only if  $B$  is not justified, inducing two status assignments. With Definition 17, on the other hand,  $A$  is prevented from being justified by the existence of a (non-selfdefeating) defeater with justified subarguments, viz.  $B$  (and likewise for  $B$ ).

The reader might wonder whether this solution is not too drastic, since it would seem to give up the property of reinstatement. For instance, when applied to Example 2, Definition 17 says that argument  $A$  is not justified, since it is defeated by  $B$ , which is not self-defeating. That  $B$  is in turn defeated by  $C$  is irrelevant, even though  $C$  is justified.

However, here it is important that Definition 17 allows us to distinguish between two kinds of reinstatement. Intuitively, the reason why  $C$  defeats  $B$  in Example 2, is that it defeats  $B$ ’s proper subargument that Tweety is a penguin. And if the subarguments in the example are made explicit as follows, Definition 17 yields the intuitive result. (As for notation, for any pair of arguments  $X$  and  $X^-$ , the latter is a proper subargument of the first.)

EXAMPLE 18. Consider four arguments  $A$ ,  $B$ ,  $B^-$  and  $C$  such that  $B$  defeats  $A$  and  $C$  defeats  $B^-$ .



According to Definition 17,  $A$  and  $C$  are justified and  $B$  and  $B^-$  are not justified. Note that  $B$  is not justified by Clause 2. So  $C$  reinstates  $A$  not by directly defeating  $B$  but by defeating  $B$ ’s subargument  $B^-$ .

The crucial difference between the Examples 2 and 3 is that in the latter example the defeat relation is of a different kind, in that  $A$  and  $B$  are in conflict on their final conclusions (respectively that Nixon is, or is not a pacifist). The only way to reinstate, say, the argument  $A$  that Nixon was a

pacifist is by finding a defeater of  $B$ 's proper subargument that Nixon was a republican (while making the subargument relations explicit).

So the only case in which Definition 17 does not capture reinstatement is when all relevant defeat relations concern the final conclusions of the arguments involved. This might even be regarded as a virtue of the definition, as is illustrated by the following modification of Example 2 (taken from [Nute, 1994]).

EXAMPLE 19. Consider three arguments  $A$ ,  $B$  and  $C$  such that  $B$  defeats  $A$  and  $C$  defeats  $B$ . Read the arguments as follows.

$A$  = 'Tweety flies because it is a bird'

$B$  = 'Tweety does not fly because it is a penguin'

$C$  = 'Tweety might fly because it is a genetically altered penguin'

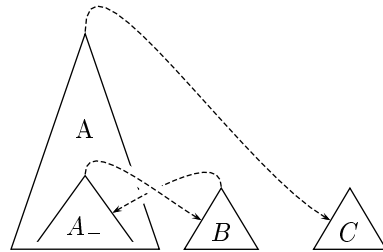
Note that, unlike in Example 2, these three arguments are in conflict on the same issue, viz. on whether Tweety can fly. According to Definitions 7 and 11 both  $A$  and  $C$  are justified; in particular,  $A$  is justified since it is reinstated by  $C$ . However, according to Definition 17 only  $C$  is justified, since  $A$  has a non-self-defeating defeater, viz.  $B$ . The latter outcome might be regarded as the intuitively correct one, since we still accept that Tweety is a penguin, which blocks the 'birds fly' default, and  $C$  allows us at most to conclude that Tweety *might* fly.

So does this example show that Definitions 7 and 11 must be modified? We think not, since it is possible to represent the arguments in such a way that these definitions give the intuitive outcome. However, this solution requires a particular logical language, for which reason its discussion must be postponed (see Section 5.2, p. 269).

Nevertheless, we can at least conclude that while the indirect form of reinstatement (by defeating a subargument) clearly seems a basic principle of argumentation, Example 19 shows that with direct reinstatement this is not so clear.

Unfortunately, Definition 17 is not yet fully adequate, as can be shown with the following extension of Example 3. It is a version of Example 13 with the subarguments made explicit.

EXAMPLE 20. (Zombie arguments 2.) Consider the arguments  $A^-$ ,  $A$ ,  $B$  and  $C$  such that  $A^-$  and  $B$  defeat each other and  $A$  defeats  $C$ .



A concrete example is

- $A^-$  = ‘Dixon is a pacifist because he is a quaker’
- $B$  = ‘Dixon is no pacifist because he is a republican’
- $A$  = ‘Dixon has no gun because he is a pacifist’
- $C$  = ‘Dixon has a gun because he lives in Chicago’

According to Definition 17,  $C$  is justified since its only defeater,  $A$ , has a proper subargument that is not justified, viz.  $A^-$ . Yet, as we explained above with Example 13, intuitively  $A$  should retain its capacity to prevent  $C$  from being justified, since the defeater of its subargument is not justified.

There is an obvious way to repair Definition 17: it must be made explicitly ‘three-valued’ by changing the phrase ‘not justified’ in Clause 3 into ‘overruled’,<sup>8</sup> where the latter term is defined as follows.

DEFINITION 21. (Defensible and overruled arguments 2.)

- An argument is *overruled* iff it is not justified and either it is self-defeating, or it or one of its proper subarguments is defeated by a justified argument.
- An argument is *defensible* iff it is not justified and not overruled.

This results in the following definition of justified arguments.

DEFINITION 22. (Recursively justified arguments—revised.) An argument  $A$  is *justified* iff

1.  $A$  is not self-defeating; and
2. All proper subarguments of  $A$  are justified; and
3. All arguments defeating  $A$  are self-defeating, or have at least one proper subargument that is overruled.

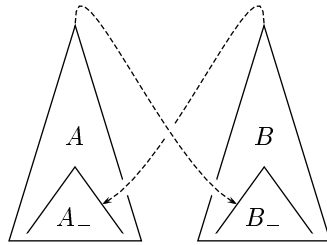
In Example 20 this has the following result. Note first that none of the arguments are self-defeating. Then to determine whether  $C$  is justified, we must determine the status of  $A$ .  $A$  defeats  $C$ , so  $C$  is only justified if  $A$  is overruled. Since  $A$  is not defeated,  $A$  can only be overruled if its proper subargument  $A^-$  is overruled. No proper subargument of  $A^-$  is defeated, but  $A^-$  is defeated by  $B$ . So if  $B$  is justified,  $A^-$  is overruled. Is  $B$  justified? No, since it is defeated by  $A^-$ , and  $A^-$  is not self-defeating and has no overruled proper subarguments. But then  $A$  is not overruled, which means that  $C$  is not justified. In fact, all arguments in the example are defensible, as can be easily verified.

<sup>8</sup>Makinson & Schlechta [1991] criticise this possibility and recommend the approach with multiple status assignments.

*Comparing fixed-point and recursive definitions*

Comparing the fixed-point and recursive definitions, we have seen that in the main example where their outcomes differ (Example 19), the intuitions seem to favour the outcome of the recursive definitions (but see below, p. 269). We have also seen that the recursive definition, if made ‘three-valued’, can deal with zombie arguments just as well as the fixed-point definitions. So must we favour the recursive form? The answer is negative, since it also has a problem: Definitions 17 and 22 do not always enforce a unique status assignment. Consider the following example.

EXAMPLE 23. (Crossover defeat.)<sup>9</sup> Consider four arguments  $A^-$ ,  $A$ ,  $B^-$ ,  $B$  such that  $A$  defeats  $B^-$  while  $B$  defeats  $A^-$ .



Definition 17 allows for two status assignments, viz. one in which only  $A^-$  and  $A$  are justified, and one in which only  $B^-$  and  $B$  are justified. In addition, Definition 22 also allows for the status assignment which makes all arguments defensible. Clearly, the latter status assignment is the intuitively intended one. However, without fixed-point constructions it seems hard to enforce it as the unique one.

Note, finally, that in our discussion of the non-recursive approach we implicitly assumed that when a proper subargument of an argument is defeated, thereby the argument itself is also defeated (see e.g. Example 2). In fact, any particular argumentation system that has no explicitly recursive definition of justified arguments should satisfy this assumption. By contrast, systems that have a recursive definition, can leave defeat of an argument independent from defeat of its proper subarguments. Furthermore, if a system has no recursive definition of justified arguments, but still distinguishes arguments and subarguments for other reasons (as e.g. [Simari & Loui, 1992] and [Prakken & Sartor, 1997b]), then a proof is required that Clause 2 of Definition 17 holds. Further illustration of this point must be postponed to the discussion of concrete systems in Section 5.

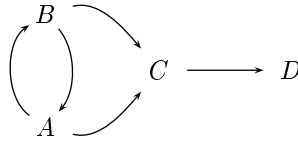
<sup>9</sup>The name ‘crossover’ is taken from Hunter [1993].



*Unique status assignments: evaluation*

Evaluating the unique-status-assignment approach, we have seen that it can be formalised in an elegant way if fixed-point definitions are used, while the, perhaps more natural attempt with a recursive definition has some problems. However, regardless of its precise formalisation, this approach has inherent problems with certain types of examples, such as the following.

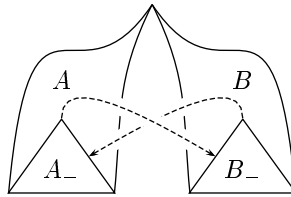
EXAMPLE 24. (Floating arguments.) Consider the arguments  $A, B, C$  and  $D$  such that  $A$  defeats  $B$ ,  $B$  defeats  $A$ ,  $A$  defeats  $C$ ,  $B$  defeats  $C$  and  $C$  defeats  $D$ .



Since no argument is undefeated, Definition 8 tells us that all of them are defensible. However, it might be argued that for  $C$  and  $D$  this should be otherwise: since  $C$  is defeated by both  $A$  and  $B$ ,  $C$  should be overruled. The reason is that as far as the status of  $C$  is concerned, there is no need to resolve the conflict between  $A$  and  $B$ : the status of  $C$  ‘floats’ on that of  $A$  and  $B$ . And if  $C$  should be overruled, then  $D$  should be justified, since  $C$  is its only defeater.

A variant of this example is the following piece of default reasoning. To analyse this example, we must again make an assumption on the structure of arguments, viz. that they have a conclusion.

EXAMPLE 25. (Floating conclusions.)<sup>10</sup> Consider the arguments  $A^-$ ,  $A$ ,  $B^-$  and  $B$  such that  $A^-$  and  $B^-$  defeat each other and  $A$  and  $B$  have the same conclusion.



An intuitive reading is

<sup>10</sup>The term ‘floating conclusions’ was coined by Makinson & Schlechta [1991].

- $A^-$  = Brygt Rykkje is Dutch because he was born in Holland
- $B^-$  = Brygt Rykkje is Norwegian because he has a Norwegian name
- $A$  = Brygt Rykkje likes ice skating because he is Dutch
- $B$  = Brygt Rykkje likes ice skating because he is Norwegian

The point is that whichever way the conflict between  $A^-$  and  $B^-$  is decided, we always end up with an argument for the conclusion that Brygt Rykkje likes ice skating, so it seems that it is justified to accept this conclusion as true, even though it is not supported by a justified argument. In other words, the status of this conclusion floats on the status of the arguments  $A^-$  and  $B^-$ .

While the unique-assignment approach is inherently unable to capture floating arguments and conclusions, there is a way to capture them, viz. by working with multiple status assignments. To this approach we now turn.

#### 4.2 *The multiple-status-assignments approach*

A second way to deal with competing arguments of equal strength is to let them induce two alternative status assignments, in both of which one is justified at the expense of the other. Note that both these assignments will satisfy Definition 1. In this approach, an argument is ‘genuinely’ justified iff it receives this status in all status assignments. To prevent terminological confusion, we now slightly reformulate the notion of a status assignment.

**DEFINITION 26.** A *status assignment* to a set  $X$  of arguments ordered by a binary defeat relation is an assignment to each argument of either the status ‘in’ or the status ‘out’ (but not both), satisfying the following conditions:

1. An argument is *in* if all arguments defeating it (if any) are out.
2. An argument is *out* if it is defeated by an argument that is in.

Note that the conditions (1) and (2) are just the conditions of Definition 1. In Example 3 there are precisely two possible status assignments:



Recall that an argumentation system is supposed to define when it is justified to accept an argument. What can we say in case of  $A$  and  $B$ ? Since both of them are ‘in’ in one status assignment but ‘out’ in the other, we must conclude that neither of them is justified. This is captured by redefining the notion of a justified argument as follows:

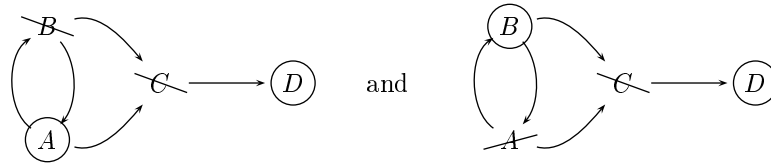
DEFINITION 27. Given a set  $X$  of arguments and a relation of defeat on  $X$ , an argument is *justified* iff it is ‘in’ in all status assignments to  $X$ .

However, this is not all; just as in the unique-status-assignment approach, it is possible to distinguish between two different categories of arguments that are not justified. Some of those arguments are in no extension, but others are at least in some extensions. The first category can be called the *overruled*, and the latter category the *defensible* arguments.

DEFINITION 28. Given a set  $X$  of arguments and a relation of defeat on  $X$

- An argument is *overruled* iff it is ‘out’ in all status assignments to  $X$ ;
- An argument is *defensible* iff it is ‘in’ in some and ‘out’ in some status assignments to  $X$ .

It is easy to see that the unique-assignment and multiple-assignments approaches are not equivalent. Consider again Example 24. Argument  $A$  and  $B$  form an even loop, thus, according to the multiple-assignments approach, either  $A$  and  $B$  can be assigned ‘in’ but not both. So the above defeat relation induces two status assignments:



While in the unique-assignment approach all arguments are defensible, we now have that  $D$  is justified and  $C$  is overruled.

Multiple status assignments also make it possible to capture floating conclusions. This can be done by defining the status of formulas as follows.

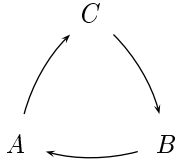
DEFINITION 29. (The status of conclusions.)

- $\varphi$  is a *justified conclusion* iff every status assignment assigns ‘in’ to an argument with conclusion  $\varphi$ ;
- $\varphi$  is a *defensible conclusion* iff  $\varphi$  is not justified, and a conclusion of a defensible argument.
- $\varphi$  is an *overruled conclusion* iff  $\varphi$  is not justified or defensible, and a conclusion of an overruled argument.

Changing the first clause into ‘ $\varphi$  is a justified conclusion iff  $\varphi$  is the conclusion of a justified argument’ would express a stronger notion, not recognising floating conclusions as justified.

There is reason to distinguish several variants of the multiple-status-assignments approach. Consider the following example, with an ‘odd loop’ of defeat relations.

EXAMPLE 30. (Odd loop.) Let  $A, B$  and  $C$  be three arguments, represented in a triangle, such that  $A$  defeats  $C$ ,  $B$  defeats  $A$ , and  $C$  defeats  $B$ .



In this situation, Definition 27 has some problems, since this example has no status assignments.

1. Assume that  $A$  is ‘in’. Then, since  $A$  defeats  $C$ ,  $C$  is ‘out’. Since  $C$  is ‘out’,  $B$  is ‘in’, but then, since  $B$  defeats  $A$ ,  $A$  is ‘out’. Contradiction.
2. Assume next that  $A$  is ‘out’. Then, since  $A$  is the only defeater of  $C$ ,  $C$  is ‘in’. Then, since  $C$  defeats  $B$ ,  $B$  is ‘out’. But then, since  $B$  is the only defeater of  $A$ ,  $A$  is ‘in’. Contradiction.

Note that a self-defeating argument is a special case of Example 30, viz. the case where  $B$  and  $C$  are identical to  $A$ . This means that sets of arguments containing a self-defeating argument might have no status assignment.

To deal with the problem of odd defeat cycles, several alternatives to Definition 26 have been studied in the literature. They will be discussed in Section 5, in particular in 5.1 and 5.2.

### 4.3 Comparing the two approaches

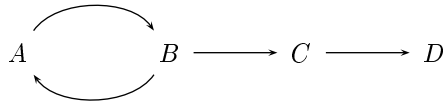
How do the unique- and multiple-assignment approaches compare to each other? It is sometimes said that their difference reflects a difference between a ‘sceptical’ and ‘credulous’ attitude towards drawing defeasible conclusions: when faced with an unresolvable conflict between two arguments, a sceptic would refrain from drawing any conclusion, while a credulous reasoner would choose one conclusion at random (or both alternatively) and further explore its consequences. The sceptical approach is often defended by saying that since in an unresolvable conflict no argument is stronger than the other, neither of them can be accepted as justified, while the credulous approach

has sometimes been defended by saying that the practical circumstances often require a person to act, whether or not s/he has conclusive reasons to decide which act to perform.

In our opinion this interpretation of the two approaches is incorrect. When deciding what to accept as a justified belief, what is important is not whether one or more possible status assignments are considered, but how the arguments are evaluated given these assignments. And this evaluation is captured by the qualifications ‘justified’ and ‘defensible’, which thus capture the distinction between ‘sceptical’ and ‘credulous’ reasoning. And since, as we have seen, the distinction justified vs. defensible arguments can be made in both the unique-assignment and the multiple-assignments approach, these approaches are independent of the distinction ‘sceptical’ vs. ‘credulous’ reasoning.

Although both approaches can capture the notion of a defensible argument, they do so with one important difference. The multiple-assignments approach is more convenient for identifying *sets* of arguments that are compatible with each other. The reason is that while with unique assignments the defensible arguments are defensible on an individual basis, with multiple assignments they are defensible because they belong to a set of arguments that are ‘in’ and thus can be defended simultaneously. Even if two defensible arguments do not defeat each other, they might be incompatible in the sense that no status assignment makes them both ‘in’, as in the following example.

EXAMPLE 31. *A* and *B* defeat each other, *B* defeats *C*, *C* defeats *D*.



This example has two status assignments, viz.  $\{A, C\}$  and  $\{B, D\}$ . Accordingly, all four arguments are defensible. Note that, although *A* and *D* do not defeat each other, *A* is in iff *D* is out. So *A* and *D* are in some sense incompatible. In the unique-assignment approach this notion of incompatibility seems harder to capture.

As we have seen, the unique-assignment approach has no inherent difficulty to recognise ‘zombie arguments’; this problem only occurs if this approach uses a recursive two-valued definition of the status of arguments.

As for their outcomes, the approaches mainly differ in their treatment of floating arguments and conclusions. With respect to these examples, the question easily arises whether one approach is the right one. However, we prefer a different attitude: instead of speaking about the ‘right’ or ‘wrong’ definition, we prefer to speak of ‘senses’ in which an argument or conclusion

can be justified. For instance, the sense in which the conclusion that Brygt Rykkje likes ice skating in Example 25 is justified is different from the sense in which, for instance, the conclusion that Tweety flies in Example 2 is justified: only in the second case is the conclusion supported by a justified argument. And the status of  $D$  in Example 24 is not quite the same as the status of, for instance,  $A$  in Example 2. Although both arguments need the help of other arguments to be justified, the argument helping  $A$  is itself justified, while the arguments helping  $D$  are merely defensible. In the concluding section we come back to this point, and generalise it to other differences between the various systems.

#### 4.4 General properties of consequence notions

We conclude this section with a much-discussed issue, viz. whether any nonmonotonic consequence notion, although lacking the property of monotonicity, should still satisfy other criteria. Many argue that this is the case, and much research has been devoted to formulating such criteria and designing systems that satisfy them; see e.g. [Gabbay, 1985; Makinson, 1989; Kraus *et al.*, 1990]. We, however, do not follow this approach, since we think that it is hard to find any criterion that should really hold for any argumentation system, or nonmonotonic consequence notion, for that matter. We shall illustrate this with the condition that is perhaps most often defended, called *cumulativity*. In terms of argumentation systems this principle says that if a formula  $\varphi$  is justified on the basis of a set of premises  $T$ , then any formula  $\psi$  is justified on the basis of  $T$  if and only if  $\psi$  is also justified on the basis of  $T \cup \{\varphi\}$ . We shall in particular give counterexamples to the ‘if’ part of the biconditional, which is often called *cautious monotony*. This condition in fact says that adding justified conclusions to the premises cannot make other justified conclusions unjustified.

At first sight, this principle would seem uncontroversial. However, we shall now (quasi-formally) discuss reasonably behaving argumentation systems, with plausible criteria for defeat, and show by example that they do not satisfy cautious monotony and are therefore not cumulative. These examples illustrate two points. First they illustrate Makinson & Schlechta’s [1991] remark that systems that do not satisfy cumulativity assign facts a special status. Second, since the examples are quite natural, they illustrate that argumentation systems *should* assign facts a special status and therefore should not be cumulative.

Below, the  $\longrightarrow$  symbols stand for unspecified reasoning steps in an argument, and the formulas stand for the conclusion drawn in such a step.

EXAMPLE 32. Consider two (schematic) arguments

$$\begin{aligned} A : p &\longrightarrow q \longrightarrow r \longrightarrow \neg q \longrightarrow s \\ B : &\longrightarrow \neg s \end{aligned}$$

Suppose we have a system in which self-defeating arguments have no capacity to prevent other arguments from being justified. Assume also that  $A$  is self-defeating, since a subconclusion,  $\neg q$ , is based on a subargument for a conclusion  $q$ . Assume, finally, that the system makes  $A$ 's subargument for  $r$  justified (since it has no non-selfdefeating counterarguments). Then  $B$  is justified. However, if  $r$  is now added to the 'facts', the following argument can be constructed:

$$A' : r \longrightarrow \neg q \longrightarrow s$$

This argument is not self-defeating, and therefore it might have the capacity to prevent  $B$  from being justified.

EXAMPLE 33. Consider next the following arguments.

$$A \text{ is a two-step argument } p \longrightarrow q \longrightarrow r$$

$$B \text{ is a three-step argument } s \longrightarrow t \longrightarrow u \longrightarrow \neg r$$

And assume that conflicting arguments are compared on their length (the shorter, the better). Then  $A$  strictly defeats  $B$ , so  $A$  is justified. Assume, however, also that  $B$ 's subargument

$$s \longrightarrow t \longrightarrow u$$

is justified, since it has no counterarguments, and assume that  $u$  is added to the facts. Then we have a new argument for  $\neg r$ , viz.

$$B' : u \longrightarrow \neg r$$

which is shorter than  $A$  and therefore strictly defeats  $A$ .

Yet another type of example uses numerical assessments of arguments.

EXAMPLE 34. Consider the arguments

$$A : p \longrightarrow q \longrightarrow r$$

$$B : s \longrightarrow \neg r$$

Assume that in  $A$  the strength of the derivation of  $q$  from  $p$  is 0.7 and that the strength of the derivation of  $r$  from  $q$  is 0.85, while in  $B$  the strength of the derivation of  $\neg r$  from  $s$  is 0.8. Consider now an argumentation system where arguments are compared with respect to their weakest links. Then  $B$  strictly defeats  $A$ , since  $B$ 's weakest link is 0.8 while  $A$ 's weakest link is 0.7. However, assume once more that  $A$ 's subargument for  $q$  is justified because it has no counterargument, and then assume that  $q$  is added as a fact. Then a new argument

$$A' : q \longrightarrow r$$

can be constructed, with as weakest link 0.85, so that it strictly defeats  $B$ .

The point of these examples is that reasonable argumentation systems with plausible criteria for defeat are conceivable which do not satisfy cumulativity, so that cumulativity cannot be required as a minimum requirement for justified belief. Vreeswijk [1993a, pp. 82–8] has shown that other

properties of nonmonotonic consequence relations also turn out to be counterintuitive in a number of realistic logical scenario's.

## 5 SOME ARGUMENTATION SYSTEMS

Let us, after our general discussions, now turn to individual argumentation systems and frameworks. We shall present them according to the conceptual sketch of Section 3, and also evaluate them in the light of Section 4.

### 5.1 *The abstract approach of Bondarenko, Dung, Kowalski and Toni*

#### *Introductory remarks*

We first discuss an abstract approach to nonmonotonic logic developed in several articles by Bondarenko, Dung, Toni and Kowalski (below called the 'BDKT approach'). Historically, this work came after the development by others of a number of argumentation systems (to be discussed below). The major innovation of the BDKT approach is that it provides a framework and vocabulary for investigating the general features of these other systems, and also of nonmonotonic logics that are not argument-based.

The latest and most comprehensive account of the BDKT approach is Bondarenko *et al.* [1997]. In this account, the basic notion is that of a set of "assumptions". In their approach the premises come in two kinds: 'ordinary' premises, comprising a *theory*, and *assumptions*, which are formulas (of whatever form) that are designated (on whatever ground) as having default status. Inspired by Poole [1988], Bondarenko *et al.* [1997] regard nonmonotonic reasoning as adding sets of assumptions to theories formulated in an underlying monotonic logic, provided that the contrary of the assumptions cannot be shown. What in their view makes the theory argumentation-theoretic is that this provision is formalised in terms of sets of assumptions attacking each other. In other words, according to Bondarenko *et al.* [1997] an argument is a set of assumptions. This approach has especially proven successful in capturing existing nonmonotonic logics.

Another version of the BDKT approach, presented by Dung [1995], completely abstracts from both the internal structure of an argument and the origin of the set of arguments; all that is assumed is the existence of a set of arguments, ordered by a binary relation of 'defeat'.<sup>11</sup> This more abstract point of view seems more in line with the aims of this chapter, and therefore we shall below mainly discuss Dung's version of the BDKT approach. As remarked above, it inspired much of our discussion in Section 4. The

---

<sup>11</sup>BDKT use the term 'attack', but to maintain uniformity we shall use 'defeat'.



assumption-based version of Bondarenko *et al.* [1997] will be briefly outlined at the end of this subsection.

### *Basic notions*

As just remarked, Dung's [1995] primitive notion is a set of arguments ordered by a binary relation of defeat. Dung then defines various notions of so-called argument extensions, which are intended to capture various types of defeasible consequence. These notions are declarative, just declaring sets of arguments as having a certain status. Finally, Dung shows that many existing nonmonotonic logics can be reformulated as instances of the abstract framework.

Dung's basic formal notions are as follows.

DEFINITION 35. An *argumentation framework* (AF) is a pair  $(Args, \text{defeat})$ , where  $Args$  is a set of arguments, and  $\text{defeat}$  a binary relation on  $Args$ .

- An AF is *finitary* iff each argument in  $Args$  is defeated by at most a finite number of arguments in  $Args$ .
- A set of arguments is *conflict-free* iff no argument in the set is defeated by an argument in the set.

One might think of the set  $Args$  as all arguments that can be constructed in a given logic from a given set of premises (although this is not always the case; see the discussions below of 'partial computation'). Unless stated otherwise, we shall below implicitly assume an arbitrary but fixed AF.

Dung interprets *defeat*, like us, in the weak sense of 'conflicting and not being weaker'. Thus in Dung's approach two arguments can defeat each other. Dung does not explicitly use the stronger (and asymmetric) notion of strict defeat, but we shall sometimes use it below.

A central notion of Dung's framework is acceptability, already defined above in Definition 6. We repeat it here. It captures how an argument that cannot defend itself, can be protected from attacks by a set of arguments.

DEFINITION 36. An argument  $A$  is *acceptable* with respect to a set  $S$  of arguments iff each argument defeating  $A$  is defeated by an argument in  $S$ .

As remarked above, the arguments in  $S$  can be seen as the arguments capable of reinstating  $A$  in case  $A$  is defeated. To illustrate acceptability, consider again Example 2, which in terms of Dung has an AF (called 'TT' for 'Tweety Triangle') with  $Args = \{A, B, C\}$  and  $\text{defeat} = \{(B, A), (C, B)\}$  ( $B$  strictly defeats  $A$  and  $C$  strictly defeats  $B$ ).  $A$  is acceptable with respect to  $\{C\}$ ,  $\{A, C\}$ ,  $\{B, C\}$  and  $\{A, B, C\}$ , but not with respect to  $\emptyset$  and  $\{B\}$ .

Another central notion is that of an admissible set.

DEFINITION 37. A conflict-free set of arguments  $S$  is *admissible* iff each argument in  $S$  is acceptable with respect to  $S$ .

Intuitively, an admissible set represents an admissible, or defensible, point of view. In Example 2 the sets  $\emptyset$ ,  $\{C\}$  and  $\{A, C\}$  are admissible but all other subsets of  $\{A, B, C\}$  are not admissible.

#### *Argument extensions*

In terms of the notions of acceptability and admissibility several notions of ‘argument extensions’ can be defined, which are what we above called ‘status assignments’. The following notion of a stable extension is equivalent to Definition 26 above.

DEFINITION 38. A conflict-free set  $S$  is a *stable extension* iff every argument that is not in  $S$ , is defeated by some argument in  $S$ .

In Example 2,  $TT$  has only one stable extension, viz.  $\{A, C\}$ . Consider next an AF called  $ND$  (the Nixon Diamond), corresponding to Example 3, with  $Args = \{A, B\}$ , and  $defeat = \{(A, B), (B, A)\}$ .  $ND$  has two stable extensions,  $\{A\}$  and  $\{B\}$ .

Since a stable extension is conflict-free, it reflects in some sense a coherent point of view. It is also a maximal point of view, in the sense that every possible argument is either accepted or rejected. In fact, stable semantics is the most ‘aggressive’ type of semantics, since a stable extension defeats every argument not belonging to it, whether or not that argument is hostile to the extension. This feature is the reason why not all AF’s have stable extensions, as Example 30 has shown.

To give such examples also a multiple-assignment semantics, Dung defines the notion of a preferred extension.

DEFINITION 39. A conflict-free set is a *preferred extension* iff it is a maximal (with respect to set inclusion) admissible set.

Let us go back to Definition 26 of a status assignment and define a *partial status assignment* in the same way as a status assignment, but without the condition that it assigns a status to all arguments. Then it is easy to verify that preferred extensions correspond to maximal partial status assignments.

Dung shows that every AF has a preferred extension. Moreover, he shows that stable extensions are preferred extensions, so in the Nixon Diamond and the Tweety Triangle the two semantics coincide. However, not all preferred extensions are stable: in Example 30 the empty set is a (unique) preferred extension, which is not stable. Preferred semantics leaves all arguments in an odd defeat cycle out of the extension, so none of them is defeated by an argument in the extension.

Preferred and stable semantics are an instance of the multiple-status-assignments approach of Section 4.2: in cases of an irresolvable conflict as

in the Nixon diamond, two incompatible extensions are obtained. Dung also explores the unique-status-assignment approach, with his notion of a *grounded* extension, already presented above as Definition 7. To build a bridge between the various semantics, Dung also defines ‘complete semantics’.

DEFINITION 40. An admissible set of arguments is a *complete extension* iff each argument that is acceptable with respect to  $S$  belongs to  $S$ .

This definition implies that a set of arguments is a complete extension iff it is a fixed point of the operator  $F$  defined in Definition 7. According to Dung, a complete extension captures the beliefs of a rational person who believes everything s/he can defend.

#### *Self-defeating arguments*

How do Dung’s various semantics deal with self-defeating arguments? It turns out that all semantics have some problems. For stable semantics they are the most serious, since an AF with a self-defeating argument might have no stable extensions. For preferred semantics this problem does not arise, since preferred extensions are guaranteed to exist. However, this semantics still has a problem, since self-defeating arguments can prevent other arguments from being justified. This can be illustrated with Example 15 (an AF with two arguments  $A$  and  $B$  such that  $A$  defeats  $A$  and  $A$  defeats  $B$ ). The set  $\{B\}$  is not admissible, so the only preferred extension is the empty set. Yet intuitively it seems that instead  $\{B\}$  should be the only preferred extension, since  $B$ ’s only defeater is self-defeating. It is easy to see that the same holds for complete semantics. In Section 4.1 we already saw that this example causes the same problems for grounded semantics, but that for finitary AF’s Pollock [1987] provides a solution. Both Dung [1995] and Bondarenko *et al.* [1997] recognise the problem of self-defeating arguments, and suggest that solutions in the context of logic programming of Kakas *et al.* [1994] could be generalised to deal with it. Dung also acknowledges Pollock’s [1995] approach, to be discussed in Subsection 5.2.

#### *Formal results*

Both Dung [1995] and Bondarenko *et al.* [1997] establish a number of results on the existence of extensions and the relation between the various semantics. We now summarise some of them.

1. Every stable extension is preferred, but not vice versa.
2. Every preferred extension is a complete extension, but not vice versa.
3. The grounded extension is the least (with respect to set inclusion) complete extension.

4. The grounded extension is contained in the intersection of all preferred extensions (Example 24 is a counterexample against ‘equal to’.)
5. If an AF contains no infinite chains  $A_1, \dots, A_n, \dots$  such that each  $A_{i+1}$  defeats  $A_i$  then AF has exactly one complete extension, which is grounded, preferred and stable. (Note that the even loop of Example 3 and the odd loop of Example 30 form such an infinite chain.)
6. Every AF has at least one preferred extension.
7. Every AF has exactly one grounded extension.

Finally, Dung [1995] and Bondarenko *et al.* [1997] identify several conditions under which preferred and stable semantics coincide.

*Assumption-based formulation of the framework*

As mentioned above, Bondarenko *et al.* [1997] have developed a different version of the BDKT approach. This version is less abstract than the one of Dung [1995], in that it embodies a particular view on the structure of arguments. Arguments are seen as sets of assumptions that can be added to a theory in order to (monotonically) derive conclusions that cannot be derived from the theory alone. Accordingly, Bondarenko *et al.* [1997] define a more concrete version of Dung’s [1995] argumentation frameworks as follows:

DEFINITION 41. Let  $\mathcal{L}$  be a formal language and  $\vdash$  a monotonic logic defined over  $\mathcal{L}$ . An *assumption-based framework* with respect to  $(\mathcal{L}, \vdash)$  is a tuple  $\langle T, Ab, \bar{\ } \rangle$  where

- $T, Ab \subseteq \mathcal{L}$
- $\bar{\ }$  is a mapping from  $Ab$  into  $\mathcal{L}$ , where  $\bar{\alpha}$  denotes the *contrary* of  $\alpha$ .

The notion of *defeat* is now defined for sets of assumptions (below we leave the assumption-based framework implicit).

DEFINITION 42. A set of assumptions  $A$  *defeats* an assumption  $\alpha$  iff  $T \cup A \vdash \bar{\alpha}$ ; and  $A$  *defeats* a set of assumptions  $\Delta$  iff  $A$  *defeats* some assumption  $\alpha \in \Delta$ .

The notions of argument extensions are then defined in terms of sets of assumptions. For instance,

DEFINITION 43. A set of assumptions  $\Delta$  is *stable* iff

- $\Delta$  is closed, i.e.,  $\Delta = \{\alpha \in Ab \mid T \cup \Delta \vdash \alpha\}$
- $\Delta$  does not defeat itself

- $\Delta$  defeats each assumption  $\alpha \notin \Delta$

A *stable extension* is a set  $Th(T \cup \Delta)$  for some stable set  $\Delta$  of assumptions.

As remarked above, Bondarenko *et al.*'s [1997] main aim is to reformulate existing nonmonotonic logics in their general framework. Accordingly, what an assumption is, and what its contrary is, is determined by the choice of nonmonotonic logic to be reformulated. For instance, in applications of preferential entailment where abnormality predicates  $\text{ab}_i$  are to be minimised (see Section 2.1), the assumptions will include expressions of the form  $\neg \text{ab}_i(c)$ , where  $\overline{\neg \text{ab}_i(c)} = \text{ab}_i(c)$ . And in default logic (see also Section 2.1), an assumption is of the form  $M\varphi$  for any 'middle part'  $\varphi$  of a default, where  $\overline{M\varphi} = \neg\varphi$ ; moreover, all defaults  $\varphi:\psi/\chi$  are added to the rules defining  $\vdash$  as monotonic inference rules  $\varphi, M\psi/\chi$ .

### Procedure

The developers of the BDKT approach have also studied procedural forms for the various semantics. Dung *et al.* [1996; 1997] propose two abstract proof procedures for computing admissibility (Definition 37), where the second proof procedure is a computationally more efficient refinement of the first. Both procedures are based upon a proof procedure originally intended for computing stable semantics in logic programming. And they are both formulated as logic programs that are derived from a formal specification. The derivation guarantees the correctness of the proof procedures. Further, Dung *et al.* [1997] show that both proof procedures are complete. Here, the first procedure is discussed.

It is defined in the form of a meta-level logic program, of which the top-level clause defines admissibility. This concept is captured in a predicate *adm*:

$$(1) \quad \text{adm}(\Delta_0, \Delta) \longleftrightarrow [\Delta_0 \subseteq \Delta \text{ and } \Delta \text{ is admissible}]$$

$\Delta$  and  $\Delta_0$  are sets of assumptions, where ' $\Delta$  is admissible' is a low-level concept that is defined with the help of auxiliary clauses. In this manner, (1) provides a specification for the proof procedure. Similarly, a top-level predicate *defends* is defined

$$\text{defends}(D, \Delta) \longleftrightarrow [D \text{ defeats } \Delta' - \Delta, \text{ for every } \Delta' \text{ that defeats } \Delta]$$

The proof procedure that Dung *et al.* propose can be understood in procedural terms as repeatedly adding defences to the initially given set of assumptions  $\Delta_0$  until no further defences need to be added. More precisely,

given a current set of assumptions  $\Delta$ , initialised as  $\Delta_0$ , the proof procedure repeatedly

1. finds a set of assumptions  $D$  such that  $defends(D, \Delta)$ ;
2. replaces  $\Delta$  by  $\Delta \cup D$

until  $D = \Delta$ , in which case it returns  $\Delta$ .

Step (1) is non-deterministic, since there might be more than one set of assumptions  $D$  defending the current  $\Delta$ . The proof procedure potentially needs to explore a search tree of alternatives to find a branch which terminates with a self-defending set. The logic-programming formulation of the proof procedure is:

$$\begin{aligned} adm(\Delta, \Delta) &\leftarrow defends(\Delta, \Delta) \\ adm(\Delta, \Delta') &\leftarrow defends(D, \Delta), adm(\Delta \cup D, \Delta') \end{aligned}$$

The procedural characterisation of the proof procedure is obtained by applying SLD resolution to the above clauses with a left-to-right selection rule, with an initial query of the form  $\leftarrow adm(\Delta_0, \Delta)$  with  $\Delta_0$  as input and  $\Delta$  as output.

The procedure is proved correct with respect to the admissibility semantics, but it is shown to be incorrect for stable semantics in general. According to Dung *et al.*, this is due to the above-mentioned ‘epistemic aggressiveness’ of stable semantics, viz. the fact that a stable extension defeats every argument not belonging to it. Dung *et al.* remark that, besides being counterintuitive, this property is also computationally very expensive, because it necessitates a search through the entire space of arguments to determine, for every argument, whether or not it is defeated. Subsequent evaluation by Dung *et al.* of the proof procedure has suggested that it is the semantics, rather than the proof procedure, which was at fault, and that preferred semantics provides an improvement. This insight is also formulated by Dung [1995].

Finally, it should be noted that recently, Kakas & Toni [1999] have developed proof procedures in dialectical style (see Section 6 below) for the various semantics of Bondarenko *et al.* [1997] and for Kakas *et al.* [1994]’s acceptability semantics.

### *Evaluation*

As remarked above, the abstract BDKT approach was a major innovation in the study of defeasible argumentation, in that it provided an elegant general framework for investigating the various argumentation systems. Moreover, the framework also applies to other nonmonotonic logics, since Dung and Bondarenko *et al.* extensively show how many of these logics can be translated into argumentation systems. Thus it becomes very easy to formulate alternative semantics for nonmonotonic logics. For instance, default logic, which was shown by Dung [1995] to have a stable semantics, can very easily

be given an alternative semantics in which extensions are guaranteed to exist, like preferred or grounded semantics. Moreover, the proof theories that have been or will be developed for the various argument-based semantics immediately apply to the systems that are an instance of these semantics. Because of these features, the BDKT framework is also very useful as guidance in the development of new systems, as, for instance, Prakken & Sartor have used it in developing the system of Subsection 5.7 below.

On the other hand, the level of abstractness of the BDKT approach (especially in Dung's version) also leaves much to the developers of particular systems. In particular, they have to define the internal structure of an argument, the ways in which arguments can conflict, and the origin of the defeat relation. Moreover, it seems that at some points the BDKT approach needs to be refined or extended. We already mentioned the treatment of self-defeating arguments, and Prakken & Sartor [1997b] have extended the BDKT framework to let it cope with reasoning about priorities (see Subsection 5.7 below).

## 5.2 Pollock

John Pollock was one of the initiators of the argument-based approach to the formalisation of defeasible reasoning. Originally he developed his theory as a contribution to philosophy, in particular epistemology. Later he turned to artificial intelligence, developing a computer program called OSCAR, which implements his theory. Since the program falls outside the scope of this handbook, we shall only discuss the logical aspects of Pollock's system; for the architecture of the computer program the reader is referred to e.g. Pollock [1995]. The latter also discusses other topics, such as practical reasoning, planning and reasoning about action.

### *Reasons, arguments, conflict and defeat*

In Pollock's system, the underlying logical language is standard first-order logic, but the notion of an argument has some nonstandard features. What still conforms to accounts of deductive logic is that arguments are sequences of propositions linked by inference rules (or better, by instantiated inference schemes). However, Pollock's formalism begins to deviate when we look at the kinds of inference schemes that can be used to build arguments. Let us first concentrate on linear arguments; these are formed by combining so-called *reasons*. Technically, reasons connect a set of propositions with a proposition. Reasons come in two kinds, conclusive and *prima facie* reasons.

*Conclusive reasons* still adhere to the common standard, since they are reasons that logically entail their conclusions. In other words, a conclusive reason is any valid first-order inference scheme (which means that Pollock's system includes first-order logic). Thus, examples of conclusive reasons are

$\{p, q\}$  is a conclusive reason for  $p \wedge q$   
 $\{\forall xPx\}$  is a conclusive reason for  $Pa$

*Prima facie* reasons, by contrast have no counterpart in deductive logic; they only create a presumption in favour of their conclusion, which can be defeated by other reasons, depending on the strengths of the conflicting reasons. Based on his work in epistemology, Pollock distinguishes several kinds of *prima facie* reasons: for instance, principles of perception, such as<sup>12</sup>

$[x \text{ appears to me as } Y]$  is a *prima facie* reason for believing  $[x \text{ is } Y]$ .

(For the objectification-operator  $[\cdot]$  see page 231 and page 265.)

Another source of *prima facie* reasons is the statistical syllogism, which says that:

If  $(r > 0.5)$  then  $[x \text{ is an } F \text{ and } \text{prob}(G/F) = r]$  is a *prima facie* reason of strength  $r$  for believing  $[x \text{ is a } G]$ .

Here  $\text{prob}(G/F)$  stands for the conditional probability of  $G$  given  $F$ .

*Prima facie* reasons can also be based on principles of induction, for example,

$[X \text{ is a set of } m \text{ } F\text{'s and } n \text{ members of } X \text{ have the property } G \text{ (} n/m > 0.5)]$  is a *prima facie* reason of strength  $n/m$  for believing  $[ \text{all } F\text{'s have the property } G]$ .

Actually, Pollock adds to these definitions the condition that  $F$  is *projectible* with respect to  $G$ . This condition, introduced by Goodman, 1954, is meant to prevent certain ‘unfounded’ probabilistic or inductive inferences. For instance, the first observed person from Lanikai, who is a genius, does not permit the prediction that the next observed Lanikaian will be a genius. That is, the predicate ‘intelligence’ is not projectible with respect to ‘birthplace’. Projectibility is of major concern in probabilistic reasoning.

To give a simple example of a linear argument, assume the following set of ‘input’ facts  $\text{INPUT} = \{A(a), \text{prob}(B/A) = 0.8, \text{prob}(C/B) = 0.7\}$ . The following argument uses reasons based on the statistical syllogism, and the first of the above-displayed conclusive reasons.

---

<sup>12</sup>When a reason for a proposition is a singleton set, we drop the brackets.



- |   |   |
|---|---|
| 1. $\langle A(a), \infty \rangle$                                 | $(A(a) \text{ is in INPUT})$                                  |
| 2. $\langle [\text{prob}(B/A) = 0.8], \infty \rangle$             | $([\text{prob}(B/A) = 0.8] \text{ is in INPUT})$              |
| 3. $\langle A(a) \wedge [\text{prob}(B/A) = 0.8], \infty \rangle$ | (1,2 and $\{p, q\}$ is a conclusive reason for $p \wedge q$ ) |
| 4. $\langle B(a), 0.8 \rangle$                                    | (3 and the statistical syllogism)                             |
| 5. $\langle [\text{prob}(C/B) = 0.7], \infty \rangle$             | $(\lceil \text{prob}(C/B) = 0.7 \rceil \text{ is in INPUT})$  |
| 6. $\langle B(a) \wedge [\text{prob}(C/B) = 0.7], 0.8 \rangle$    | (4,5 and $\{p, q\}$ is a conclusive reason for $p \wedge q$ ) |
| 7. $\langle C(a), 0.7 \rangle$                                    | (6 and the statistical syllogism)                             |

So each line of a linear argument is a pair, consisting of a proposition and a numerical value that indicates the strength, or degree of justification of the proposition. The strength  $\infty$  at lines 1,2 and 5 indicates that the conclusions of these lines are put forward as absolute facts, originating from the epistemic base 'INPUT'. At line 4, the *weakest link* principle is applied, with the result that the strength of the argument line is the minimum of the strength of the reason for  $B(a)$  (0.8) and the argument line 3 from which  $C(a)$  is derived with this reason ( $\infty$ ). At lines 6 and 7 the weakest link principle is applied again.

Besides linear arguments, Pollock also studies *suppositional* arguments. In suppositional reasoning, we 'suppose' something that we have not inferred from the input, draw conclusions from the supposition, and then 'discharge' the supposition to obtain a related conclusion that no longer depends on the supposition. In Pollock's system, suppositional arguments can be constructed with inference rules familiar from natural deduction. Accordingly, the propositions in an argument have sets of propositions attached to them, which are the *suppositions* under which the proposition can be derived from earlier elements in the sequence.

The following definition (based on [Pollock, 1995]) summarises this informal account of argument formation.

**DEFINITION 44.** In OSCAR, an *argument* based on INPUT is a finite sequence  $\sigma_1, \dots, \sigma_n$ , where each  $\sigma_i$  is a line of argument. A *line of argument*  $\sigma_i$  is a triple  $\langle X_i, p_i, \nu_i \rangle$ , where  $X_i$ , a set of propositions, is the set of *suppositions* at line  $i$ ,  $p_i$  is a proposition, and  $\nu_i$  is the *degree of justification* of  $\sigma$  at line  $i$ . A line of argument is obtained from earlier lines of argument according to one of the following rules of argument formation.

**Input.** If  $p$  is in INPUT and  $\sigma$  is an argument, then for any  $X$  it holds that  $\sigma, \langle X, p, \infty \rangle$  is an argument.

**Reason.** If  $\sigma$  is an argument,  $\langle X_1, p_1, \eta_1 \rangle, \dots, \langle X_n, p_n, \eta_n \rangle$  are members of  $\sigma$ , and  $\{p_1, \dots, p_n\}$  is a reason of strength  $\nu$  for  $q$ , and for each  $i$ ,  $X_i \subset X$ , then  $\sigma, \langle X, q, \min\{\eta_1, \dots, \eta_n, \nu\} \rangle$  is an argument.

**Supposition.** If  $\sigma$  is an argument,  $X$  a set of propositions and  $p \in X$ , then  $\sigma, \langle X, p, \infty \rangle$  is also an argument.

**Conditionalisation.** If  $\sigma$  is an argument and some line of  $\sigma$  is  $\langle X \cup \{p\}, q, \nu \rangle$ , then  $\sigma, \langle X, (p \supset q), \nu \rangle$  is also an argument.

**Dilemma.** If  $\sigma$  is an argument and some line of  $\sigma$  is  $\langle X, p \vee q, \nu \rangle$ , and some line of  $\sigma$  is  $\langle X \cup \{p\}, r, \mu \rangle$ , and some line of  $\sigma$  is  $\langle X \cup \{q\}, r, \xi \rangle$ , then  $\sigma, \langle X, r, \min\{\nu, \mu, \xi\} \rangle$  is also an argument.

Pollock [1995] notes that other inference rules could be added as well.

It is the use of *prima facie* reasons that makes arguments defeasible, since these reasons can be defeated by other reasons. This can take place in two ways: by *rebutting* defeaters, which are at least as strong reasons with the opposite conclusion, and by *undercutting* defeaters, which are at least as strong reasons of which the conclusion denies the connection that the undercut reason states between its premises and its conclusion. A typical example of rebutting defeat is when an argument using the reason ‘Birds fly’ is defeated by an argument using the reason ‘Penguins don’t fly’. Pollock’s favourite example of an undercutting defeater is when an object looks red because it is illuminated by a red light: knowing this undercuts the reason for believing that this object is red, but it does not give a reason for believing that the object is not red.

Before we can explain how Pollock formally defines the relation of defeat among arguments, some extra notation must be introduced. In the definition of defeat among arguments, Pollock uses a, what may be called, *objectification* operator,  $[\cdot]$ . (This operator was also used in Fig. 4 on page 230 and in the *prima facie* reasons on page 263.) With this operator, expressions in the meta-language are transformed into expressions in the object language. For example, the meta-level rule

$\{p, q\}$  is a conclusive reason for  $p$

may be transformed into the object-level expression

$[\{p, q\}]$  is a conclusive reason for  $p$ .

If the object language is rich enough, then the latter expression is present in the object language, in the form  $(p \wedge q) \supset p$ . Evidently, a large fraction of the meta-expressions cannot be conveyed to the object language, because the object language lacks sufficient expressibility. This is the case, for example, if corresponding connectives are missing in the object language.

Pollock formally defines the relation of defeat among arguments as follows.

**Defeat among arguments.** An argument  $\sigma$  defeats another argument  $\eta$  if and only if:

1.  $\eta$ ’s last line is  $\langle X, q, \alpha \rangle$  and is obtained by the argument formation rule *Reason* from some earlier lines  $\langle X_1, p_1, \alpha_1 \rangle, \dots, \langle X_n, p_n, \alpha_n \rangle$  where  $\{p_1, \dots, p_n\}$  is a *prima facie* reason for  $q$ ; and

2.  $\sigma$ 's last line is  $\langle Y, r, \beta \rangle$  where  $Y \subseteq X$  and either:

- (a)  $r$  is  $\neg q$  and  $\beta \geq \alpha$ ; or
- (b)  $r$  is  $\neg[\{p_1, \dots, p_n\} \gg q]$  and  $\beta \geq \alpha$ .

(1) determines the weak spot of  $\eta$ , while (2) determines whether that weak spot is (2a) a conclusion (in this case  $q$ ), or (2b) a reason (in this case  $\{p_1, \dots, p_n\} \gg q$ ). For Pollock, (2a) is a case of *rebutting defeat*, and 2b is a case of *undercutting defeat*: if  $\sigma$  undercuts the last reason of  $\eta$ , it blocks the derivation of  $q$ , without supporting  $\neg q$  as alternative conclusion. The formula  $[\{p_1, \dots, p_n\} \gg q]$  stands for the translation of ‘ $\{p_1, \dots, p_n\}$  is a *prima facie* reason for  $q$ ’ into the object language.

Pollock leaves the notion of conflicting arguments implicit in this definition of defeat. Note also that a defeater of an argument always defeats the last step of an argument; Pollock treats ‘subargument defeat’ by a recursive definition of a justified argument, i.e., in the manner explained above in Section 4.1.

### *Suppositional reasoning*

As noted above, the argument formation rules *supposition*, *conditionalisation* and *dilemma* can be used to form suppositional arguments. OSCAR is one of the very few nonmonotonic logics that allow for suppositional reasoning. Pollock finds it necessary to introduce suppositional reasoning because, in his opinion, this type of reasoning is ubiquitous not only in deductive, but also in defeasible reasoning. Pollock mentions, among other things, the reasoning form ‘reasoning by cases’, which is notoriously hard for many non-monotonic logics. An example is ‘presumably, birds fly, presumably, bats fly, Tweety is a bird or a bat, so, presumably, Tweety flies’. In Pollock’s system, this argument can be formalised as follows.

EXAMPLE 45. Consider the following reasons.

- (1)  $\text{Bird}(x)$  is a *prima facie* reason of strength  $\nu$  for  $\text{Flies}(x)$
- (2)  $\text{Bat}(x)$  is a *prima facie* reason of strength  $\mu$  for  $\text{Flies}(x)$

And consider  $\text{INPUT} = \{\text{Bird}(t) \vee \text{Bat}(t)\}$ . The conclusion  $\text{Flies}(t)$  can be defeasibly derived as follows.

- 1.  $\langle \emptyset, \text{Bird}(t) \vee \text{Bat}(t), \infty \rangle$  ( $\text{Bird}(t) \vee \text{Bat}(t)$  is in INPUT)
- 2.  $\langle \{\text{Bird}(t)\}, \text{Bird}(t), \infty \rangle$  (Supposition)
- 3.  $\langle \{\text{Bird}(t)\}, \text{Flies}(t), \nu \rangle$  (2 and *prima facie* reason (1))
- 4.  $\langle \{\text{Bat}(t)\}, \text{Bat}(t), \infty \rangle$  (Supposition)
- 5.  $\langle \{\text{Bat}(t)\}, \text{Flies}(t), \mu \rangle$  (4 and *prima facie* reason (2))
- 6.  $\langle \emptyset, \text{Flies}(t), \min\{\nu, \mu\} \rangle$  (3,5 and Dilemma)

At line 1, the proposition  $\text{Bird}(t) \vee \text{Bat}(t)$  is put forward as an absolute fact. At line (2), the proposition  $\text{Bird}(t)$  is temporarily supposed to be true. From this assumption, at the following line the conclusion  $\text{Flies}(t)$  is defeasibly derived with the first *prima facie* reason. Line (4) is an alternative continuation of line 1. At line (4),  $\text{Bat}(t)$  is supposed to be true, and at line (5) it is used to again defeasibly derive  $\text{Flies}(t)$ , this time from the second *prima facie* reason. Finally, at line (6) the Dilemma rule is applied to (3) and (5), discharging the assumptions in the alternative suppositional arguments, and concluding to  $\text{Flies}(t)$  under no assumption.

According to Pollock, another virtue of his system is that it validates the defeasible derivation of a material implication from a *prima facie* reason. Consider again the ‘birds fly’ reason (1), and assume that INPUT is empty.

1.  $\langle \{\text{Bird}(t)\}, \text{Bird}(t), \infty \rangle$  (Supposition)
2.  $\langle \{\text{Bird}(t)\}, \text{Flies}(t), \nu \rangle$  (1 and *prima facie* reason (1))
3.  $\langle \emptyset, \text{Bird}(t) \supset \text{Flies}(t), \nu \rangle$  (2 and Conditionalisation)

Pollock regards the validity of these inferences as desirable. On the other hand, Vreeswijk has argued that suppositional defeasible reasoning, in the way Pollock proposes it, sometimes enables incorrect inferences. Vreeswijk’s argument is based on the idea that the strength of a conclusion obtained by means of conditionalisation is incomparable to the reason strength of the implication occurring in that conclusion. For a discussion of this problem the reader is further referred to Vreeswijk [1993a, pp. 184–7].

Having seen how Pollock defines the notions of arguments, conflicting arguments, and defeat among arguments, we now turn to what was the main topic of Section 4 and the main concern of Dung [1995], defining the status of arguments.

### *The status of arguments*

Over the years, Pollock has more than once changed his definition of the status of arguments. One change is that while earlier versions (e.g. Pollock, 1987) dealt with (successful) attack on a subargument in an implicit way via the definition of defeat, the latest version makes this part of the status definition, by explicitly requiring that all subarguments of an ‘undefeated’ argument are also undefeated (cf. Section 4.1). Another change is in the form of the status definition. Earlier Pollock took the unique-status-assignment approach, in particular, the fixed-point variant of Definition 16 which, as shown by Dung [1995], (almost) corresponds to the grounded semantics of Definition 7. However, his most recent work is in terms of multiple status assignments, and very similar to the preferred semantics of Definition 39. Pollock’s thus combines the recursive style of Definition 17 with the multiple-status-assignments approach. We present the most recent definition, of [Pollock, 1995]. To maintain uniformity in our terminology, we

state it in terms of arguments instead of, as Pollock, in terms of an ‘inference graph’. To maintain the link with inference graphs, we make the definition relative to a *closed* set of arguments, i.e., a set of arguments containing all subarguments of all its elements.

Since we deviate from Pollock’s inference graphs, we must be careful in defining the notion of subarguments. Sometimes a later line of an argument depends on only some of its earlier lines. For instance, in Example 45 line (5) only depends on (4). In fact, the entire argument (1-6) has three independent, or parallel subarguments, viz. a linear subargument (1), and two suppositional subarguments (2,3) and (4,5). Pollock’s inference graphs nicely capture such dependencies, since their nodes are argument lines and their links are inferences. However, with our sequential format of an argument this is different, for which reason we cannot define a subargument as being any subsequence of an argument. Instead, they are only those subsequences of  $A$  that can be transformed into an inference tree.

**DEFINITION 46** (subarguments). An argument  $A$  is a *subargument* of an argument  $B$  iff  $A$  is a subsequence of  $B$  and there exists a tree  $T$  of argument lines such that

1.  $T$  contains all and only lines from  $A$ ; and
2.  $T$ ’s root is  $A$ ’s last element; and
3.  $l$  is a child of  $l'$  iff  $l$  was inferred from a set of lines one of which was  $l'$ .

A *proper* subargument of  $A$  is any subargument of  $A$  unequal to  $A$ .

Now we can give Pollock’s [1995] definition of a status assignment.

**DEFINITION 47.** An assignment of ‘defeated’ and ‘undefeated’ to a closed set  $S$  of arguments is a *partial defeat status assignment* iff it satisfies the following conditions.

1. All arguments in  $S$  with only lines obtained by the *input* argument formation rule are assigned ‘undefeated’;
2.  $A \in S$  is assigned ‘undefeated’ iff:
  - (a) All proper sub-arguments of  $A$  are assigned ‘undefeated’; and
  - (b) All arguments in  $S$  defeating  $A$  are assigned ‘defeated’.
3.  $A \in S$  is assigned ‘defeated’ iff:
  - (a) One of  $A$ ’s proper sub-arguments is assigned ‘defeated’;
  - or
  - (b)  $A$  is defeated by an argument in  $S$  that is assigned ‘undefeated’.

A *defeat status assignment* is a maximal (with respect to set inclusion) partial defeat status assignment.

Observe that the conditions (2a) and (3a) on the sub-arguments of  $A$  make the weakest link principle hold by definition.

The similarity of defeat status assignments to Dung's preferred extensions of Definition 39 shows itself as follows: the conditions (2b) and (3b) on the defeaters of  $A$  are the analogues of Dung's notion of acceptability, which make a defeat status assignment an admissible set; then the fact that a defeat status assignment is a maximal partial assignment induces the similarity with preferred extensions.

It is easy to verify that when two arguments defeat each other (Example 3), an input has more than one status assignment. Since Pollock wants to define a sceptical consequence notion, he therefore has to consider the intersection of all assignments. Pollock does so in a variant of Definitions 27 and 28.

**DEFINITION 48.** (The status of arguments.) Let  $S$  be a closed set of arguments based on INPUT. Then, relative to  $S$ , an argument is *undefeated* iff every status assignment to  $S$  assigns 'undefeated' to it; it is *defeated outright* iff no status assignment to  $S$  assigns 'undefeated' to it; otherwise it is *provisionally defeated*.

In our terms, 'undefeated' is 'justified', 'defeated outright' is 'overruled', and 'provisionally defeated' is 'defensible'.

#### *Direct vs. indirect reinstatement*

It is now the time to come back to the discussion in Section 4.1 on reinstatement. Example 19 showed that there is reason to invalidate the direct version of this principle, viz. when the conflicts are about the same issue. We remarked that the explicitly recursive Definition 17 of justified arguments indeed invalidates direct reinstatement while preserving its indirect version. However, we also promised to explain that both versions of reinstatement can be retained if Example 19 is represented in a particular way. In fact, Pollock (personal communication) would represent the example as follows:

- (1) Being a bird is a prima facie reason for being able to fly
- (2a) Being a penguin is an undercutting reason for (1)
- (2b) Being a penguin is a defeasible reason for not being able to fly
- (3) Being a genetically altered penguin is an undercutting reason for (2b)
- (4) Tweety is a genetically altered penguin

It is easy to verify that Definitions 47 and 48, which validate both direct and indirect of reinstatement, yield the intuitive outcome, viz. that it is neither justified that Tweety can fly, nor that it cannot fly. A similar representation

is possible in systems that allow for abnormality or exception clauses, e.g. in [Geffner & Pearl, 1992; Bondarenko *et al.*, 1997; Prakken & Sartor, 1997b].

### *Self-defeating arguments*

Pollock has paid much attention to the problem of self-defeating arguments. In Pollock's system, an argument defeats itself iff one of its lines defeats another of its lines. Above in Section 4.1 we already discussed Pollock's treatment of self-defeating arguments within the unique-status-assignment approach. However, he later came to regard this treatment as incorrect, and he now thinks that it can only be solved in the multiple-assignment approach (personal communication).

Let us now see how Pollock's Definitions 47 and 48 deal with the problem. Two cases must be distinguished. Consider first two defeasible arguments  $A$  and  $B$  rebutting each other. Then  $A$  and  $B$  are 'parallel' subarguments of a deductive argument  $A+B$  for any proposition. Then (if no other arguments interfere with  $A$  or  $B$ ) there are two status assignments, one in which  $A$  is assigned 'undefeated' and  $B$  assigned 'defeated', and one the other way around. Now  $A+B$  is in both of these assignments assigned 'defeated', since in both assignments one of its proper subarguments is assigned 'defeated'. Thus the self-defeating argument  $A+B$  turns out to be defeated outright, which seems intuitively plausible.

A different case is the following, with the following reasons

- (1)  $p$  is a *prima facie* reason of strength 0.8 for  $q$
- (2)  $q$  is a *prima facie* reason of strength 0.8 for  $r$
- (3)  $r$  is a conclusive reason for  $[\neg(p \gg q)]$

and with INPUT =  $\{p\}$ . The following (linear) argument can be constructed.

- |    |                                      |  |
|----|--------------------------------------|--|
| 1. | $\langle p, \infty \rangle$          | $(p \text{ is in INPUT})$                                  |
| 2. | $\langle q, 0.8 \rangle$             | $(1 \text{ and } \textit{prima facie} \text{ reason (1)})$ |
| 3. | $\langle r, 0.8 \rangle$             | $(2 \text{ and } \textit{prima facie} \text{ reason (2)})$ |
| 4. | $\langle \neg(p \gg q), 0.8 \rangle$ | $(3 \text{ and conclusive reason (3)})$                    |

Let us call this argument  $A$ , with proper subarguments  $A_1, A_2, A_3$  and  $A_4$ , respectively. Observe first that, according to Pollock's definition of self-defeat,  $A_4$  is self-defeating. Further, according to Pollock's earlier approach with Definition 16,  $A_4$  is, as being self-defeating, overruled, or 'defeated', while  $A_1, A_2$  and  $A_3$  are justified, or 'undefeated'. Pollock now regards this outcome as incorrect: since  $A_4$  is a deductive consequence of  $A_3$ ,  $A_3$  should also be 'defeated'.

This result is obtained with Definitions 47 and 48. Firstly,  $A_1$  is clearly undefeated. Consider next  $A_2$ . This argument is undercut by  $A_4$ , so if  $A_4$  is assigned 'undefeated', then  $A_2$  must be assigned 'defeated'. But then  $A_4$

must also be assigned ‘defeated’, since one of its proper subarguments is assigned ‘defeated’. Contradiction. If, on the other hand,  $A_4$  is assigned ‘defeated’, then  $A_2$  and so  $A_3$  must be assigned ‘undefeated’. But then  $A_4$  must be assigned ‘undefeated’. Contradiction. In conclusion, no partial status assignment will assign a status to  $A_4$  and, consequently, no status assignment will assign a status to  $A_2$  or  $A_3$  either. And since this implies that no status assignment assigns the status ‘undefeated’ to any of these arguments, they are by Definition 48 all defeated outright.

Two remarks about this outcome can be made. Firstly, it might be doubted whether  $A_2$  should indeed be defeated outright, i.e., overruled. It is not self-defeating, its only defeater is self-defeating, and this defeater is not a deductive consequence of  $A_2$ ’s conclusion. Other systems, e.g. those of Vreeswijk (Section 5.5) and Prakken & Sartor (Section 5.7), regard  $A_2$  as justified. In these systems Pollock’s intuition about  $A_3$  is formalised by regarding  $A_3$  as self-defeating because its conclusion deductively, not just defeasibly, implies a conclusion incompatible with itself. This makes it possible to regard  $A_3$  as overruled but  $A_2$  as justified.

Furthermore, even if Pollock’s outcome is accepted, the situation is not quite the same as with the previous example. Consider another defeasible argument  $B$  which rebuts and is rebutted by  $A_3$ . Then no assignment assigns a status to  $B$  either, for which reason  $B$  is also defeated outright. Yet this shows that the ‘defeated outright’ status of  $A_2$  is not the same as the ‘defeated outright’ status of an argument that has an undefeated defeater: apparently,  $A_2$  is still capable of preventing other arguments from being undefeated. In fact, the same holds for arguments involved in an odd defeat cycle (as in Example 30).

In conclusion, Pollock’s definitions leave room for a fourth status of arguments, which might be called ‘seemingly defeated’. This status holds for arguments that according to Definition 48 are defeated outright but still have the power to prevent other arguments from being ultimately defeated. The four statuses can be partially ordered as follows: ‘undefeated’ is better than ‘provisionally defeated’ and than ‘seemingly defeated’, which both in turn are better than ‘defeated outright’. This observation applies not only to Pollock’s definition, but to all approaches based on partial status assignments, like Dung [1995] preferred semantics.

However, this is not yet all: even if the notion of seeming defeat is made explicit, there still is an issue concerning floating arguments (cf. Example 24). To see this, consider the following extension of Example 30 (formulated in terms of [Dung, 1995]).

**EXAMPLE 49.** Let  $A, B$  and  $C$  be three arguments, represented in a triangle, such that  $A$  defeats  $C$ ,  $B$  defeats  $A$ , and  $C$  defeats  $B$ . Furthermore, let  $D$  and  $E$  be arguments such that all of  $A, B$  and  $C$  defeat  $D$ , and  $D$  defeats  $E$ .



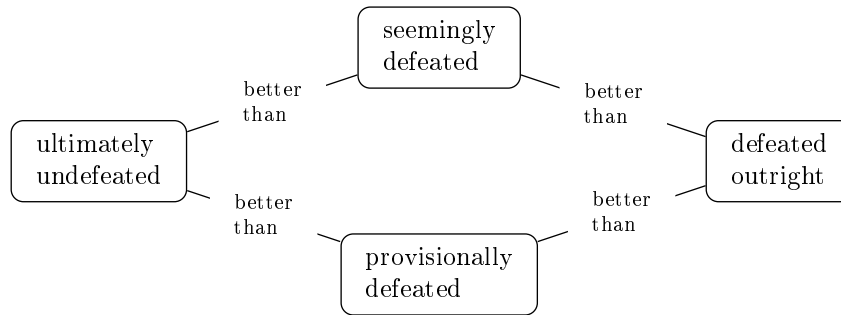
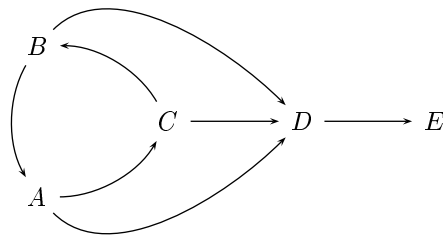


Figure 8. Partial ordering of defeat statuses.



The difference between Example 24 and this example is that the even defeat loop between two arguments is replaced by an odd defeat loop between three arguments. One view on the new example is that this difference is inessential and that, for the same reasons as why in Example 24 the argument  $D$  is justified, here the argument  $E$  is ultimately undefeated: although  $E$  is strictly defeated by  $D$ , it is reinstated by all of  $A$ ,  $B$  and  $C$ , since all these arguments strictly defeat  $D$ . On this account Definitions 47 and 48 are flawed since they render all five arguments defeated outright (and in our terms seemingly defeated). However, an alternative view is that odd defeat loops are of an essentially different kind than even defeat loops, so that our analysis of Example 24 does not apply here and that the outcome in Pollock's system reflects a flaw in the available input information rather than in the system.

#### *Ideal and resource-bounded reasoning*

We shall now see that Definition 48 is not yet all that Pollock has to say on the status of arguments. In the previous section we saw that the BDKT approach leaves the origin of the set of 'input' arguments unspecified. At

this point Pollock develops some interesting ideas. At first sight it might be thought that the set  $S$  of the just-given definitions is just the set of all arguments that can be constructed with the argument formation rules of Definition 44. However, this is only one of the possibilities that Pollock considers, in which Definition 48 captures so-called *ideal warrant*.

DEFINITION 50. (Ideal warrant.) Let  $S$  be the set of all arguments based on INPUT. Then an argument  $A$  is *ideally warranted* relative to INPUT iff  $A$  is undefeated relative to  $S$ .

Pollock wants to respect that in actual reasoning the construction of arguments takes time, and that reasoners have no infinite amount of time available. Therefore, he also considers two other definitions, both of which have a computational flavour. To capture an actual reasoning process, Pollock makes them relative to a sequence  $\mathcal{S}$  of closed finite sets  $S_0 \subseteq \dots \subseteq S_i \dots$  of arguments. Let us call this an *argumentation sequence*. Such a sequence contains all arguments constructed by a reasoner, in the order in which they are produced. It (and any of its elements) is based on INPUT if all its arguments are based on INPUT.<sup>13</sup>

Now the first ‘computational’ status definition determines what a reasoner must believe at any given time.

DEFINITION 51. (Justification.) Let  $\mathcal{S}$  be an argumentation sequence based on INPUT, and  $S_i$  an element of  $\mathcal{S}$ . Then an argument  $A$  is *justified* relative to INPUT at stage  $i$  iff  $A$  is undefeated relative to  $S_i$ .

In this definition the set  $S_i$  contains just those arguments that have actually been constructed by a reasoner. Thus this definition captures the *current* status of a belief; it may be that further reasoning (without adding new premises) changes the status of a conclusion.

This cannot happen for the other ‘computational’ consequence notion defined by Pollock, called *warrant*. Intuitively, an argument  $A$  is warranted iff eventually in an argumentation sequence a stage is reached where  $A$  remains justified at every subsequent stage. To define this, the notion of a ‘maximal’ argumentation sequence is needed, i.e., a sequence that cannot be extended. Thus it contains all arguments that a reasoner with unlimited resources would construct (in a particular order).

DEFINITION 52. (Warrant.) Let  $\mathcal{S}$  be a maximal argumentation sequence  $S_0 \subseteq \dots \subseteq S_i \dots$  based on INPUT. Then an argument  $A$  is *warranted* (relative to INPUT) iff there is an  $i$  such that for all  $j > i$ ,  $A$  is undefeated relative to  $S_j$ .

The difference between warrant and ideal warrant is subtle: it has to do with the fact that, while in determining warrant every set  $S_j \supseteq S_i$  that

---

<sup>13</sup>Note that we again translate Pollock’s inference graphs into (structured) sets of arguments.

is considered is *finite*, in determining ideal warrant the set of all possible arguments has to be considered, and this set can be infinite.

EXAMPLE 53. (Warrant does not entail ideal warrant.) Suppose  $A_1, A_2, A_3, \dots$  are arguments such that every  $A_i$  is defeated by its successor  $A_{i+1}$ . Further, suppose that the arguments are produced in the order  $A_2, A_1, A_4, A_3, A_6, A_5, A_8, \dots$ . Then

Stage	Produced	Justified
1	$A_2$	$A_2$
2	$A_2, A_1$	$A_2$
3	$A_2, A_1, A_4$	$A_2, A_4$
4	$A_2, A_1, A_4, A_3$	$A_2, A_4$
5	$A_2, A_1, A_4, A_3, A_6$	$A_2, A_4, A_6$
6	$A_2, A_1, A_4, A_3, A_6, A_5$	$A_2, A_4, A_6$
7	$A_2, A_1, A_4, A_3, A_6, A_5, A_8$	$A_2, A_4, A_6, A_8$
$\vdots$	$\vdots$	$\vdots$

From stage 1,  $A_2$  is justified and stays justified. Thus,  $A_2$  is warranted. At the same time, however,  $A_2$  is not ideally warranted, because there exist two status assignments for all  $A_i$ 's. One assignment in which all and only all odd arguments are 'in', and one assignment in which all and only all odd arguments are 'out'. Hence, according to ideal warrant, every argument is only provisionally defeated. In particular,  $A_2$  is provisionally defeated. A remarkable aspect of this example is that, eventually, every argument will be produced, but without reaching the right result for  $A_2$ .

EXAMPLE 54. (Ideal warrant does not imply warrant.) Suppose that  $A, B_1, B_2, B_3, \dots$  and  $C_1, C_2, C_3, \dots$  are arguments such that  $A$  is defeated by every  $B_i$ , and every  $B_i$  is defeated by  $C_i$ . Further, suppose that the arguments are produced in the order  $A, B_1, C_1, B_2, C_2, B_3, C_3, \dots$ . Then

Stage	Produced	Justified
1	$A$	$A$
2	$A, B_1$	$B_1$
3	$A, B_1, C_1$	$A, C_1$
4	$A, B_1, C_1, B_2$	$C_1, B_2$
5	$A, B_1, C_1, B_2, C_2$	$A, C_1, C_2$
6	$A, B_1, C_1, B_2, C_2, B_3$	$C_1, C_2, B_3$
$\vdots$	$\vdots$	$\vdots$

Thus, in this sequence,  $A$  is provisionally defeated. However, according to the definition of ideal warrant, every  $B_i$  is defeated by  $C_i$ , so that  $A$  remains undefeated.

Although the notion of warrant is computationally inspired, as Pollock observes there is no automated procedure that can determine of any war-

ranted argument that it is warranted: even if in fact a warranted argument stays undefeated after some finite number  $n$  of computations, a reasoner can in state  $n$  not *know* whether it has reached a point where the argument stays undefeated, or whether further computation will change its status.

*Pollock's reasoning architecture*

We now discuss Pollock's reasoning architecture for computing the ideally warranted propositions, i.e. the propositions that are the conclusion of an ideally warranted argument. (According to Pollock, ideal warrant is what every reasoner should ultimately strive for.) In deductive logic such an architecture would be called a 'proof theory', but Pollock rejects this term. The reason is that one condition normally required of proof theories, viz. that the set of theorems is recursively enumerable, cannot in general be satisfied for a defeasible reasoner. Pollock assumes that a reasoner reasons by constantly updating its beliefs, where an update is an elementary transition from one set of propositions to the next set of propositions. According to this view, a reasoner would be adequate if the resulting sequence is a recursively enumerable approximation of ideal warrant. However, this is impossible. Ideal warrant contains all theorems of predicate logic, and it is known that all theorems of predicate logic form a set that is not recursive. And since in defeasible reasoning some conclusions depend on the failure to derive other conclusions, the set of defeasible conclusions is not recursively enumerable. Therefore, Pollock suggests an alternative criterion of adequacy. A reasoner is called *defeasibly adequate* if the resulting sequence is a defeasibly enumerable approximation of ideal warrant.

DEFINITION 55. A set  $A$  is *defeasibly enumerable* if there is a sequence of sets  $\{A_i\}_{1 \leq i}$  such that for all  $x$

1. If  $x \in A$ , then there is an  $N$  such that  $x \in A_i$  for all  $i > N$ .
2. If  $x \notin A$ , then there is an  $M$  such that  $x \notin A_i$  for all  $i > M$ .

If  $A$  is recursively enumerable, then a reasoner who updates his beliefs in Pollock's way can approach  $A$  'from below': the reasoner can construct sets that are all supersets of the preceding set and subsets of  $A$ . However, when  $A$  is only defeasibly enumerable, a reasoner can only approach  $A$  from below and above simultaneously, in the sense that the sets  $A_i$  the reasoner constructs may contain elements not contained in  $A$ . Every such element must eventually be taken out of the  $A_i$ 's, but there need not be any point at which they have *all* been removed.

To ensure defeasible adequacy, Pollock introduces the following three operations:

1. The reasoner must adopt beliefs in response to constructing arguments, provided no counterarguments have already been adopted for

any step in the argument. If a defeasible inference occurs, a check must be made whether a counterargument for it has not already been adopted as a belief.

2. The reasoner must keep track of the bases upon which its beliefs are held. When a new belief is adopted that is a defeater for a previous inference step, then the reasoner must retract that inference step and all beliefs inferred from it.
3. The reasoner must keep track of defeated inferences, and when a defeater is itself retracted (2), this should reinstate the defeated inference.

To achieve the functions just described, Pollock introduces a so-called *flag-based* reasoner. A flag-based reasoner consists of an inference engine that produces all arguments eventually, and a component computing the defeat status of arguments.

```

LOOP   BEGIN
        make-an-inference
        recompute-defeat-statuses
      END

```

The procedure `recompute-defeat-statuses` determines which arguments are defeated outright, undefeated and provisionally defeated at each iteration of the loop. That is, at each iteration it determines justification.

Pollock then identifies certain conditions under which a flag-based reasoner is defeasibly adequate. For these conditions, the reader is referred to [Pollock, 1995, ch. 4].

### *Evaluation*

Pollock's theory of defeasible reasoning is based on more than thirty years of research in logic and epistemology. This large time span perhaps explains the richness of his theory. It includes both linear and suppositional arguments, and deductive as well as non-deductive (mainly statistical and inductive) arguments, with a corresponding distinction between two types of conflicts between arguments. Pollock's definition of the status of arguments takes the multiple-status-assignments approach, being related to Dung's preferred semantics. This semantics can deal with certain types of floating statuses and conclusions, but we have seen that certain other types are still ignored. In fact, this seems one of the main unsolved problems in argument-based semantics. An interesting aspect of Pollock's work is his study of the resource-bounded nature of practical reasoning, with the idea of partial computation embodied in the notions of warrant and especially

justification. And for artificial intelligence it is interesting that Pollock has implemented his system as a computer program.

Since Pollock focuses on epistemological issues, his system is not immediately applicable to some specific features of practical (including legal) reasoning. For instance, the use of probabilistic notions seems to make it difficult to give an account of reasoning with and about priority relations between arguments (see below in Subsection 5.7). Moreover, it would be interesting to know what Pollock would regard as suitable reasons for normative reasoning. It would also be interesting to study how, for instance, analogical and abductive arguments can be analysed in Pollock's system as giving rise to *prima facie* reasons.

### 5.3 Inheritance systems

A forerunner of argumentation systems is work on so-called inheritance systems, especially of Horty *et al.*, e.g. [1990], which we shall briefly discuss. Inheritance systems determine whether an object of a certain kind has a certain property. Their language is very restricted. The network is a directed graph. Its initial nodes represent individuals and its other nodes stand for classes of individuals. There are two kinds of links,  $\rightarrow$  and  $\nrightarrow$ , depending on whether something does or does not belong to a certain class. Links from an individual to a class express class membership, and links between two classes express class inclusion.

A path through the graph is an *inheritance path* iff its only negative link is the last one. Thus the following are examples of inheritance paths.

$P_1$ : Tweety  $\rightarrow$  Penguin  $\rightarrow$  Bird  $\rightarrow$  Canfly  
 $P_2$ : Tweety  $\rightarrow$  Penguin  $\nrightarrow$  Canfly

Another basic notion is that of an *assertion*, which is of the form  $x \rightarrow y$  or  $x \nrightarrow y$ , where  $y$  is a class. Such an assertion is *enabled* by an inheritance path if the path starts with  $x$  and ends with the same link to  $y$  as the assertion. Above, an assertion enabled by  $P_1$  is Tweety  $\rightarrow$  Canfly, and an assertion enabled by  $P_2$  is Tweety  $\nrightarrow$  Canfly.

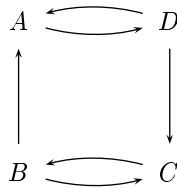
As the example shows, two paths can be conflicting. They are compared on specificity, which is read off from the syntactic structure of the net, resulting in relations of *neutralisation* and *preemption* between paths. The assignment of a status to a path (whether it is *permitted*) is similar to the recursive variant of the unique-status-assignment approach of Definition 17. This means that the system has problems with *Zombie paths* and *floating conclusions* (as observed by Makinson & Schlechta [1991]).

Although Horty *et al.* present their system as a special-purpose formalism, it clearly has all the elements of an argumentation system. An inheritance path corresponds to an argument, and an assertion enabled by a path to a conclusion of an argument. Their notion of conflicting paths

corresponds to rebutting attack. Furthermore, neutralisation and preemption correspond to defeat, while a permitted path is the same as a justified argument.

Because of the restricted language and the rather complex definition of when an inheritance path is permitted, we shall not present the full system. However, Horty *et al.* should be credited for anticipating many distinctions and discussions in the field of defeasible argumentation. In particular, their work is a rich source of benchmark examples. We shall discuss one of them.

EXAMPLE 56. Consider four arguments  $A, B, C$  and  $D$  such that  $B$  strictly defeats  $A$ ,  $D$  strictly defeats  $C$ ,  $A$  and  $D$  defeat each other and  $B$  and  $C$  defeat each other.



Here is a natural-language version (due to Horty, personal communication), in which the defeat relations are based on specificity considerations.

$A$  = Larry is rich because he is a public defender, public defenders are lawyers, and lawyers are rich;

$B$  = Larry is not rich because he is a public defender, and public defenders are not rich;

$C$  = Larry is rich because he lives in Brentwood, and people who live in Brentwood are rich;

$D$  = Larry is not rich because he rents in Brentwood, and people who rent in Brentwood are not rich.

If we apply the various semantics of the BDKT approach to this example, we see that since no argument is undefeated, none of them is in the grounded extension. Moreover, there are preferred extensions in which Larry is rich, and preferred extensions in which Larry is not rich. Yet it might be argued that since both arguments that Larry is rich are strictly defeated by an argument that Larry is not rich, the sceptical conclusion should be that Larry is not rich. This is the outcome obtained by Horty *et al.* [1990]. We note that if this example is represented in the way Pollock proposes for Example 19 (see page 269 above), this outcome can also be obtained in the BDKT approach.

### 5.4 *Lin and Shoham*

Before the BDKT approach, an earlier attempt to provide a unifying framework for nonmonotonic logics was made by Lin & Shoham [1989]. They show how any logic, whether monotonic or not, can be reformulated as a system for constructing arguments. However, in contrast with the other theories in this section, they are not concerned with comparing incompatible arguments, and so their framework cannot be used as a theory of defeat among arguments.

The basic elements of Lin & Shoham's abstract framework are an unspecified logical language, only assumed to contain a negation symbol, and an also unspecified set of inference rules defined over the assumed language. Arguments can be constructed by chaining inference rules into trees.

Inference rules are either monotonic or nonmonotonic. For instance,

$$\begin{aligned} & \text{Penguin}(a) \rightarrow \text{Bird}(a) \\ & \text{Penguin}(a), \neg \text{ab}(\text{penguin}(a)) \rightarrow \neg \text{Fly}(a) \end{aligned}$$

are monotonic rules, and

$$\begin{aligned} & \text{True} \Rightarrow \neg \text{ab}(\text{penguin}(a)) \\ & \text{True} \Rightarrow \neg \text{ab}(\text{bird}(a)) \end{aligned}$$

are nonmonotonic rules. Note that these inference rules are, as in default logic, domain specific. In fact, Lin & Shoham do not distinguish between general and domain-dependent inference rules, as is shown by their reconstruction of default logic, to be discussed below.

Although the lack of a notion of defeat is a severe limitation, in capturing nonmonotonic consequence Lin & Shoham introduce a notion which for defeasible argumentation is very relevant viz. that of an *argument structure*.

**DEFINITION 57.** (argument structures) A set  $T$  of arguments is an *argument structure* if  $T$  satisfies the following conditions:

1. The set of 'base facts' (which roughly are the premises) is in  $T$ ;
2. Of every argument in  $T$  all its subarguments are in  $T$ ;
3. The set of conclusions of arguments in  $T$  is deductively closed and consistent.

Note that the notion of a 'closed' set of arguments that we used above in Pollock's Definition 47 satisfies the first two but not the third of these conditions. Note also that, although argument structures are closed under monotonic rules, they are not closed under defeasible rules.

Lin & Shoham then reformulate existing nonmonotonic logics in terms of monotonic and nonmonotonic inference rules, and show how the alternative



sets of conclusions of these logics can be captured in terms of argument structures with certain completeness properties. Bondarenko *et al.* [1997] remark that structures with these properties are very similar to their stable extensions.

The claim that existing nonmonotonic logics can be captured by an argument system is an important one, and Lin & Shoham were among the first to make it. The remainder of this section is therefore devoted to showing with an example how Lin & Shoham accomplish this, viz. for default logic [Reiter, 1980].

In default logic (see also Subsection 2.1), a *default theory* is a pair  $\Delta = (W, D)$ , where  $W$  is a set of first-order formulas, and  $D$  a set of defaults. Each default is of the form  $A : B_1, \dots, B_n / C$ , where  $A$ ,  $B_i$  and  $C$  are first-order formulas. Informally, a default reads as ‘If  $A$  is known, and  $B_1, \dots, B_n$  are consistent with what is known, then  $C$  may be inferred’. An *extension* of a default theory is any set of formulas  $E$  satisfying the following conditions.  $E = \bigcup_{i=0}^{\infty} E_i$ , where

$$\begin{aligned} E_0 &= W, \\ E_{i+1} &= Th(E_i) \cup \{C \mid A : B_1, \dots, B_n / C \in D \\ &\quad \text{where } A \in E_i \text{ and } \neg B_1, \dots, \neg B_n \notin E\} \end{aligned}$$

We now discuss the correspondence between default logic and argument systems by providing a global outline of the translation and proof. Lin & Shoham perform the translation as follows. Let  $\Delta = (W, D)$  be a closed default theory. Define  $R(\Delta)$  to be the set of the following rules:

1. True is a base fact.
2. If  $A \in W$ , then  $A$  is a base fact of  $R(\Delta)$ .
3. If  $A_1, \dots, A_n$ , and  $B$  are first-order sentences and  $B$  is a consequence of  $A_1, \dots, A_n$  in first-order logic, then  $A_1, \dots, A_n \rightarrow B$  is a monotonic rule.
4. If  $A$  is a first-order sentence, then  $\neg A \rightarrow \text{ab}(A)$  is a monotonic rule.
5. If  $A : B_1, \dots, B_n / C$  is a default in  $D$ , then

$$A, \neg \text{ab}(B_1), \dots, \neg \text{ab}(B_n) \rightarrow C$$

is a monotonic rule.

6. If  $B$  is a first-order sentence, then  $\text{True} \Rightarrow \neg \text{ab}(B)$  is a nonmonotonic rule.

Lin & Shoham proceed by introducing the concept of DL-complete argument structures.

DEFINITION 58. An argument structure  $T$  of  $R(\Delta)$  is said to be *DL-complete* if for any first-order sentence  $A$ , either  $\text{ab}(A)$  or  $\neg\text{ab}(A)$  is in  $\text{Wff}(T)$ .

Thus, a DL-complete argument structure is explicit about the abnormality of every first-order sentence. For DL-complete argument structures, the following lemma is established.

LEMMA 59. *If  $T$  is a DL-complete argument structure of  $R(\Delta)$ , then for any first-order sentence  $A$ ,  $\text{ab}(A) \in \text{Wff}(T)$  iff  $\neg A \in \text{Wff}(T)$ .*

On the basis of this result, Lin & Shoham are able to establish the following correspondence between default logic and argument systems.

THEOREM 60. *Let  $E$  be a consistent set of first-order sentences.  $E$  is an extension of  $\Delta$  iff there is a DL-complete argument structure  $T$  of  $R(\Delta)$  such that  $E$  is the restriction of  $\text{Wff}(T)$  to the set of first-order sentences.*

This theorem is proven by constructing extensions for given argument structures and *vice versa*. If  $E$  is an extension of  $\Delta$ , Lin & Shoham define  $T$  as the set of arguments with all nodes in  $E'$ , where

$$E' = E \cup \{\text{ab}(B) \mid \neg B \in E\} \cup \{\neg\text{ab}(B) \mid \neg B \notin E\}$$

and prove that  $\text{Wff}(T) = E'$ . Conversely, for a DL-complete argument structure  $T$  of  $R(\Delta)$ , Lin & Shoham prove that the first-order restriction  $E$  of  $\text{Wff}(T)$  is a default extension of  $\Delta$ . This is proven by induction on the definition of an extension.

Two features in the translation are worth noticing. First, default logic makes a distinction between meta-logic default rules and first-order logic, while argument systems do not. Second, the notion of groundedness of default extensions corresponds to that of an argument in argument systems, and the notion of fixed points in default logic corresponds to that of DL-completeness of argument structures.

Lin & Shoham further show that, for normal default theories, the translation can be performed without second-order predicates, such as  $\text{ab}$ . This result however falls beyond the scope of this chapter.

### 5.5 Vreeswijk's Abstract Argumentation Systems

Like the BDKT approach and Lin & Shoham [1989], Vreeswijk [1993a; 1997] also aims to provide an abstract framework for defeasible argumentation. His framework builds on the one of Lin & Shoham, but contains the main elements that are missing in their system, namely, notions of conflict and defeat between arguments. As Lin & Shoham, Vreeswijk also assumes an unspecified logical language  $\mathcal{L}$ , only assumed to contain the symbol  $\perp$ , denoting 'falsum' or 'contradiction,' and an unspecified set of monotonic and

nonmonotonic inference rules (which Vreeswijk calls ‘strict’ and ‘defeasible’). This also makes his system an abstract framework rather than a particular system. A point in which Vreeswijk’s work differs from Lin & Shoham is that Vreeswijk’s inference rules are not domain specific but general logical principles.

DEFINITION 61. (Rule of inference.) Let  $\mathcal{L}$  be a language.

1. A *strict rule of inference* is a formula of the form  $\phi_1, \dots, \phi_n \rightarrow \phi$  where  $\phi_1, \dots, \phi_n$  is a finite, possibly empty, sequence in  $\mathcal{L}$  and  $\phi$  is a member of  $\mathcal{L}$ .
2. A *defeasible rule of inference* is a formula of the form  $\phi_1, \dots, \phi_n \Rightarrow \phi$  where  $\phi_1, \dots, \phi_n$  is a finite, possibly empty, sequence in  $\mathcal{L}$  and  $\phi$  is a member of  $\mathcal{L}$ .

A *rule of inference* is a strict or a defeasible rule of inference.

Another aspect taken from Lin & Shoham is that in Vreeswijk’s framework, arguments can also be formed by chaining inference rules into trees.

DEFINITION 62. (Argument.) Let  $R$  be a set of rules. An argument has *premises*, a *conclusion*, *sentences* (or propositions), *assumptions*, *subarguments*, *top arguments*, a *length*, and a *size*. These are abbreviated by corresponding prefixes. An *argument*  $\sigma$  is

1. A member of  $\mathcal{L}$ ; in that case,

$$\begin{aligned} \text{prem}(\sigma) &= \{\sigma\}, \text{conc}(\sigma) = \sigma, \text{sent}(\sigma) = \{\sigma\}, \text{asm}(\sigma) = \emptyset, \\ \text{sub}(\sigma) &= \{\sigma\}, \text{top}(\sigma) = \{\sigma\}, \text{length}(\sigma) = 1, \text{and } \text{size}(\sigma) = \\ &1; \end{aligned}$$

or

2. A formula of the form  $\sigma_1, \dots, \sigma_n \rightarrow \phi$  where  $\sigma_1, \dots, \sigma_n$  is a finite, possibly empty, sequence of arguments, such that  $\text{conc}(\sigma_1) = \phi_1, \dots, \text{conc}(\sigma_n) = \phi_n$  for some rule  $\phi_1, \dots, \phi_n \rightarrow \phi$  in  $R$ , and  $\phi \notin \text{sent}(\sigma_1) \cup \dots \cup \text{sent}(\sigma_n)$ —in that case,

$$\begin{aligned} \text{prem}(\sigma) &= \text{prem}(\sigma_1) \cup \dots \cup \text{prem}(\sigma_n), \\ \text{conc}(\sigma) &= \phi, \\ \text{sent}(\sigma) &= \text{sent}(\sigma_1) \cup \dots \cup \text{sent}(\sigma_n) \cup \{\phi\}, \\ \text{asm}(\sigma) &= \text{asm}(\sigma_1) \cup \dots \cup \text{asm}(\sigma_n), \\ \text{sub}(\sigma) &= \text{sub}(\sigma_1) \cup \dots \cup \text{sub}(\sigma_n) \cup \{\sigma\}, \\ \text{top}(\sigma) &= \{\tau_1, \dots, \tau_n \rightarrow \phi \mid \tau_1 \in \text{top}(\sigma_1), \dots, \tau_n \in \text{top}(\sigma_n)\} \cup \\ &\{\phi\}, \\ \text{length}(\sigma) &= \max\{\text{length}(\sigma_1), \dots, \text{length}(\sigma_n)\} + 1, \text{and} \\ \text{size}(\sigma) &= \text{size}(\sigma_1) + \dots + \text{size}(\sigma_n) + 1; \end{aligned}$$

or

3. A formula of the form  $\sigma_1, \dots, \sigma_n \Rightarrow \phi$  where  $\sigma_1, \dots, \sigma_n$  is a finite, possibly empty, sequence of arguments, such that  $\text{conc}(\sigma_1) = \phi_1, \dots, \text{conc}(\sigma_n) = \phi_n$  for some rule  $\phi_1, \dots, \phi_n \Rightarrow \phi$  in  $R$ , and  $\phi \notin \text{sent}(\sigma_1) \cup \dots \cup \text{sent}(\sigma_n)$ ; for assumptions we have

$$\text{asm}(\sigma) = \text{asm}(\sigma_1) \cup \dots \cup \text{asm}(\sigma_n) \cup \{\phi\};$$

premises, conclusions, and other attributes are defined as in (2).

Arguments of type (1) are *atomic* arguments; arguments of type (2) and (3) are *composite* arguments. Thus, atomic arguments are language elements. An argument  $\sigma$  is said to be *in contradiction* if  $\text{conc}(\sigma) = \perp$ . An argument is *defeasible* if it contains at least one defeasible rule of inference; else it is *strict*.

Unlike Lin & Shoham, Vreeswijk assumes an ordering on arguments, indicating their difference in strength (on which more below).

As for conflicts between arguments, a difference from all other systems of this section (except [Verheij, 1996]; see below in subsection 5.10) is that a counterargument is in fact a *set* of arguments: Vreeswijk defines a set  $\Sigma$  of arguments *incompatible* with an argument  $\tau$  iff the conclusions of  $\Sigma \cup \{\tau\}$  give rise to a strict argument for  $\perp$ . Sets of arguments are needed because the language in Vreeswijk's framework is unspecified and therefore lacks the expressive power to 'recognise' inconsistency. The consequence of this lack of expressiveness is that a set of arguments  $\sigma_1, \dots, \sigma_n$  that is incompatible with  $\tau$ , cannot be joined to one argument  $\sigma$  that contradicts, or is inconsistent, with  $\tau$ . Therefore, it is necessary to take sets of arguments into account.

Vreeswijk has no explicit notion of undercutting attacks; he claims that this notion is implicitly captured by his notion of incompatibility, viz. as arguments for the denial of a defeasible conditional used by another argument. This requires some extra assumptions on the language of an abstract argumentation system, viz. that it is closed under negation ( $\neg$ ), conjunction ( $\wedge$ ), material implication ( $\supset$ ), and defeasible implication ( $\triangleright$ ). For the latter connective Vreeswijk defines the following defeasible inference rule.

$$\varphi, \varphi \triangleright \psi \Rightarrow \psi$$

With these extra language elements, it is possible to express rules of inference (which are meta-linguistic notions) in the object language. Meta-level rules using  $\rightarrow$  (strict rule of inference) and  $\Rightarrow$  (defeasible rule of inference) are then represented by corresponding object language implication symbols  $\supset$  and  $\triangleright$ . Under this condition, Vreeswijk claims to be able to define rebutting and undercutting attackers in a formal fashion. For example, let  $\sigma$  and  $\tau$  be arguments in Vreeswijk's system with conclusions  $\varphi$  and  $\psi$ , respectively. Let  $\varphi_1, \dots, \varphi_n \Rightarrow \varphi$  be the top rule of  $\sigma$ .

**Rebutting attack.** If  $\psi = \neg\varphi$ , then Vreeswijk calls  $\tau$  a *rebutting* attacker of  $\sigma$ . Thus, the conclusion of a rebutting attacker contradicts the conclusion of the argument it attacks.

**Undercutting attack.** If  $\psi = \neg(\varphi_1 \wedge \dots \wedge \varphi_n \supset \varphi)$ , i.e. if  $\psi$  is the negation of the last rule of  $\sigma$  stated in the object language, then  $\tau$  is said to be an *undercutting* attacker of  $\sigma$ . Thus, the conclusion of an undercutting attacker contradicts the last inference of the argument it attacks.

Vreeswijk's notion of defeat rests on two basic concepts, viz. the above-defined notion of incompatibility and the notion of undermining. An argument is said to *undermine* a set of arguments, if it dominates at least one element of that set. Formally, a set of arguments  $\Sigma$  is *undermined* by an argument  $\tau$  if  $\sigma < \tau$  for some  $\sigma \in \Sigma$ . If a set of arguments is undermined by another argument, it cannot uphold or maintain all of its members in case of a conflict.

Vreeswijk then defines the notion of a defeater as follows:

DEFINITION 63. (Defeater.) Let  $P$  be a base set, and let  $\sigma$  be an argument. A set of arguments  $\Sigma$  is a *defeater* of  $\sigma$  if it is incompatible with  $\sigma$  and not undermined by it; in this case  $\sigma$  is said to be *defeated* by  $\Sigma$ , and  $\Sigma$  *defeats*  $\sigma$ .  $\Sigma$  is a *minimal defeater* of  $\sigma$  if all its proper subsets do not defeat  $\sigma$ .

As for the assessment of arguments, Vreeswijk's declarative definition, (which he says is about "warrant") is similar to Pollock's definition of a defeat status assignment: both definitions have an explicit recursive structure and both lead to multiple status assignments in case of irresolvable conflicts. However, Vreeswijk's status assignments cannot be partial, for which reason Vreeswijk's definition is closer to stable semantics than to preferred semantics.

DEFINITION 64. (Defeasible entailment.) Let  $P$  be a base set. A relation  $\sim$  between  $P$  and arguments based on  $P$  is a *defeasible entailment relation* if, for every argument  $\sigma$  based on  $P$ , we have  $P \sim \sigma$  ( $\sigma$  is in force on the basis of  $P$ ) if and only if

1. The set  $P$  contains  $\sigma$ ; or
2. For some arguments  $\sigma_1, \dots, \sigma_n$  we have  $P \sim \sigma_1, \dots, \sigma_n$  and  $\sigma_1, \dots, \sigma_n \rightarrow \sigma$ ; or
3. For some arguments  $\sigma_1, \dots, \sigma_n$  we have  $P \sim \sigma_1, \dots, \sigma_n$  and  $\sigma_1, \dots, \sigma_n \Rightarrow \sigma$  and every set of arguments  $\Sigma$  such that  $P \sim \Sigma$  does not defeat  $\sigma$ .

In the Nixon Diamond of Example 3 this results in 'the Quaker argument is in force iff the Republican argument is not in force'. To deal with such

circularities Vreeswijk defines for every  $\sim$  satisfying the above definition an extension

$$(1) \quad \Sigma = \{\phi \mid P \sim \phi\}$$

On the basis of Definition 64 it can be proven that (1) is stable, i.e., it can be proven that  $\phi \notin \Sigma$  iff  $\Sigma'$  defeats  $\phi$  for some  $\Sigma' \subseteq \Sigma$ . With equally strong conflicting arguments, as in the Nixon Diamond, this results in multiple stable extensions (cf. Definition 38).

Just as in Dung's stable semantics, in Vreeswijk's system examples with odd defeat loops might have no extensions. However, an exception holds for the special case of self-defeating arguments, since Definition 63 implies that every argument of which the conclusion strictly implies  $\perp$  is defeated by the empty set.

### *Argumentation sequences*

Vreeswijk extensively studies various other characterisations of defeasible argumentation. Among other things, he develops the notion of an 'argumentation sequence'. An argumentation sequence can be regarded as a sequence

$$\Sigma_1 \longrightarrow \Sigma_2 \longrightarrow \dots \longrightarrow \Sigma_n \longrightarrow \dots$$

of Lin & Shoham's [1989] argument structures, but without the condition that these structures are closed under deduction. Each following structure is constructed by applying an inference rule to the arguments in the preceding structure. An important addition to Lin & Shoham's notion is that a newly constructed argument is only appended to the sequence if it survives all counterattacks from the argument structure developed thus far. Thus the notion of an argumentation sequence embodies, like Pollock's notion of 'justification', the idea of partial computation, i.e., of assessing arguments relative to the inferences made so far. Vreeswijk's argumentation sequences also resemble BDKT's procedure for computing admissible semantics. The difference is that BDKT adopt arguments that are defended (admissible semantics), while Vreeswijk argumentation sequences adopt arguments that are not defeated (stable semantics).

Vreeswijk also develops a procedural version of his framework in dialectical style. It will be discussed below in Section 6.

### *Plausible reasoning*

Vreeswijk further discusses a distinction between two kinds of nonmonotonic reasoning, 'defeasible' and 'plausible' reasoning. According to him, the above definition of defeasible entailment captures defeasible reasoning, which is unsound (i.e., defeasible) reasoning from firm premises, like in 'typically birds fly, Tweety is a bird, so presumably Tweety flies'. Plausible

reasoning, by contrast, is sound (i.e., deductive) reasoning from uncertain premises, as in ‘all birds fly (we think), Tweety is a bird, so Tweety flies (we think)’ [Rescher, 1976]. The difference is that in the first case a default proposition is accepted categorically, while in the second case a categorical proposition is accepted by default. In fact, Vreeswijk would regard reasoning with ordered premises, as studied in many nonmonotonic logics, not as defeasible but as plausible reasoning.

One element of this distinction is that for defeasible reasoning the ordering on arguments is not part of the input theory, reflecting priority relations between, or degrees of belief in premises, but a general ordering of *types* of arguments, such as ‘deductive arguments prevail over inductive arguments’ and ‘statistical inductive arguments prevail over generic inductive arguments’. Accordingly, Vreeswijk assumes that the ordering on arguments is the same for all sets of premises (although relative to a set of inference rules). Vreeswijk formalises plausible reasoning independent of defeasible reasoning, with the possibility to define input orderings on the premises, and he then combines the two formal treatments. To our knowledge, Vreeswijk’s framework is unique in treating these two types of reasoning in one formalism as distinct forms of reasoning; usually the two forms are regarded as alternative ways to look at the same kind of reasoning.

Evaluating Vreeswijk’s framework, we can say that it has little attention for the details of comparing arguments and that, as Pollock but in contrast to BDKT, it formalises only one type of defeasible consequence, but that it is philosophically well-motivated, and quite detailed with respect to the structure of arguments and the process of argumentation.

### 5.6 Simari & Loui

Simari & Loui [1992] present a declarative system for defeasible argumentation that combines ideas of Pollock [1987] on the interaction of arguments with ideas of Poole [1985] on specificity and ideas of Loui [1987] on defaults as twoplace meta-linguistic rules. Simari & Loui divide the premises into sets of contingent first-order formulas  $\mathcal{K}_C$ , and necessary first-order formulas  $\mathcal{K}_N$ , and one-directional default rules  $\Delta$ , e.g.

$$\begin{aligned}\mathcal{K}_C &= \{P(a)\} \\ \mathcal{K}_N &= \{\forall x.P(x) \supset B(x)\} \\ \Delta &= \{B(x) \succ F(x), P(x) \succ \neg F(x)\}.\end{aligned}$$

Note that Simari & Loui’s default rules are not threeplace as Reiter’s defaults, but twoplace. The set of *grounded instances* of  $\Delta$ , i.e., of defeasible rules without variables, is denoted by  $\Delta^\downarrow$ . The notion of argument that Simari & Loui maintain is somewhat uncommon:

DEFINITION 65. (Arguments.) Given a context  $\mathcal{K} = \mathcal{K}_N \cup \mathcal{K}_C$  and a set  $\Delta$  of defeasible rules we say that a subset  $T$  of  $\Delta^\downarrow$  is an *argument* for  $h \in \text{Sent}_C(\mathcal{L})$  in the context  $\mathcal{K}$ , denoted by  $\langle T, h \rangle_{\mathcal{K}}$  if and only if

1.  $\mathcal{K} \cup T \vdash h$
2.  $\mathcal{K} \cup T \not\vdash \perp$
3.  $\nexists T' \subset T : \mathcal{K} \cup T' \vdash h$

An argument  $\langle T, h_1 \rangle_{\mathcal{K}}$  is a subargument of an argument  $\langle S, h_2 \rangle_{\mathcal{K}}$  iff  $T \subseteq S$ .

That  $\mathcal{K} \cup T \vdash h$  means that  $h$  is derivable from  $\mathcal{K} \cup T$  with first-order inferences applied to first-order formulas and modus ponens applied to defaults. Thus, an argument  $T$  is a set of grounded instances of defeasible rules containing sufficient rules to infer  $h$  (1), containing no rules irrelevant for inferring  $h$  (3), and not making it possible to infer  $\perp$  (2). This notion of argument is somewhat uncommon because it does not refer to a tree or chain of inference rules. Instead, Definition 65 merely demands that an argument is a unordered collection of rules that together imply a certain conclusion.

Simari & Loui define conflict between arguments as follows. An argument  $\langle T, h_1 \rangle_{\mathcal{K}}$  *counterargues* an argument  $\langle S, h_2 \rangle_{\mathcal{K}}$  iff the latter has a subargument  $\langle S', h \rangle_{\mathcal{K}}$  such that  $\langle T, h_1 \rangle_{\mathcal{K}}$  *disagrees* with  $\langle S', h \rangle_{\mathcal{K}}$ , i.e.,  $\mathcal{K} \cup \{h_1, h\} \vdash \perp$ .

Arguments are compared with Poole's [1985] definition of specificity: an argument  $A$  *defeats* an argument  $B$  iff  $A$  disagrees with a subargument  $B^-$  of  $B$  and  $A$  is more specific than  $B^-$ . Note that this allows for subargument defeat: this is necessary since Simari & Loui's definition of the status of arguments is not explicitly recursive. In fact, they use Pollock's theory of level- $n$  arguments. Since they exclude self-defeating arguments by definition, they can use the version of Definition 11.

An important component of Simari & Loui's system is the  $\Sigma^k$ -operator. Of all the conclusions that can be argued, the  $\Sigma^k$ -operator returns the conclusions that are supported by level- $k$  arguments. Simari & Loui prove that arguments for which  $\Sigma^k = \Sigma^{k+1}$ , are justified. The main theorem of the paper states that the set of justified conclusions is uniquely determined, and that a repeated application of the  $\Sigma$ -operator will bring us to that set.

A strong point of Simari & Loui's approach is that it combines the ideas of specificity (Poole) and level- $n$  arguments (Pollock) into one system. Another strong point of the paper is that it presents a convenient calculus of arguments, that possesses elegant mathematical properties. Finally, Simari & Loui sketch an interesting architecture for implementation, which has a dialectical form (see below, Section 6 and, for a full description, [Simari *et al.*, 1994; Garcia *et al.*, 1998]). However, the system also has some limitations. Most of them are addressed by Prakken & Sartor [1996; 1997b], to be discussed next.



### 5.7 Prakken & Sartor

Inspired by legal reasoning, Prakken & Sartor [1996; 1997b] have developed an argumentation system that combines the language (but not the rest) of default logic with the grounded semantics of the BDKT approach.<sup>14</sup> Actually, Prakken & Sartor originally used the language of extended logic programming, but Prakken [1997] generalised the system to default logic's language. Below we present the latter version. The main contributions to defeasible argumentation are a study of the relation between rebutting and assumption attack, and a formalisation of argumentation about the criteria for defeat. The use of default logic's language and grounded semantics make Prakken & Sartor's system rather similar to Simari & Loui's. However, as just noted, they extend and revise it in a number of respects, to be indicated in more detail below.

As for the logical language, the premises are divided into factual knowledge  $\mathcal{F}$ , a set of first-order formulas subdivided into the necessary facts  $\mathcal{F}_n$  and the contingent facts  $\mathcal{F}_c$ , and defeasible knowledge  $\Delta$ , consisting of Reiter-defaults. The set  $\mathcal{F}$  is assumed consistent. Prakken & Sartor write defaults as follows.

$$d: \varphi_1 \wedge \dots \wedge \varphi_j \wedge \sim \varphi_k \wedge \dots \wedge \sim \varphi_n \Rightarrow \psi$$

where  $d$ , a term, is the informal name of the default, and each  $\varphi_i$  and  $\psi$  is a first-order formula. The part  $\sim \varphi_k \wedge \dots \wedge \sim \varphi_n$  corresponds to the middle part of a Reiter-default. The symbol  $\sim$  can be informally read as 'not provable that'. For each  $\sim \varphi_i$  in a default,  $\neg \varphi_i$  is called an *assumption* of the default. The language is defined such that defaults cannot be nested, nor combined with other formulas.

Arguments are, as in [Simari & Loui, 1992], chains of defaults 'glued' together by first-order reasoning. More precisely, consider the set  $\mathcal{R}$  consisting of all valid first-order inference rules plus the following rule of *defeasible modus ponens (DMP)*:

$$\frac{d: \quad \varphi_0 \wedge \dots \wedge \varphi_j \wedge \sim \varphi_k \wedge \dots \wedge \sim \varphi_m \Rightarrow \varphi_n, \quad \varphi_0 \wedge \dots \wedge \varphi_j}{\varphi_n}$$

where all  $\varphi_i$  are first-order formulas. Note that *DMP* ignores a default's assumptions; the idea is that such an assumption is untenable, this will be reflected by a successful attack on the argument using the default.

An argument is defined as follows.

**DEFINITION 66.** (Arguments.) Let  $\Gamma$  be any default theory ( $\mathcal{F}_c \cup \mathcal{F}_n \cup \Delta$ ). An *argument based on*  $\Gamma$  is a sequence of distinct first-order formulas and/or ground instances of defaults  $[\varphi_1, \dots, \varphi_n]$  such that for all  $\varphi_i$ :

<sup>14</sup>A forerunner of this system was presented in [Prakken, 1993].

- $\varphi_i \in \Gamma$ ; or
- There exists an inference rule  $\psi_1, \dots, \psi_m / \varphi_i$  in  $\mathcal{R}$  such that  $\psi_1, \dots, \psi_m \in \{\varphi_1, \dots, \varphi_{i-1}\}$

For any argument  $A$

- $\varphi \in A$  is a *conclusion* of  $A$  iff  $\varphi$  is a first-order formula;
- $\varphi \in A$  is an *assumption* of  $A$  iff  $\varphi$  is an assumption of a default in  $A$ ;
- $A$  is *strict* iff  $A$  does not contain any default;  $A$  is *defeasible* otherwise.

The set of conclusions of an argument  $A$  is denoted by  $CONC(A)$  and the set of its assumptions by  $ASS(A)$ .

Note that unlike in Simari & Loui, arguments are not assumed consistent. Here is an example of an argument:

$$[a, r_1: a \wedge \sim \neg b \Rightarrow c, c, a \wedge c, r_2: a \wedge c \Rightarrow d, d, d \vee e]$$

$CONC(A) = \{a, c, a \wedge c, d, d \vee e\}$  and  $ASS(A) = \{b\}$ .

The presence of assumptions in a rule gives rise to two kinds of conflicts between arguments, conclusion-to-conclusion attack and conclusion-to-assumption attack.

**DEFINITION 67.** (Attack.) Let  $A$  and  $B$  be two arguments.  $A$  *attacks*  $B$  iff

1.  $CONC(A) \cup CONC(B) \cup \mathcal{F}_n \vdash \perp$ ; or
2.  $CONC(A) \cup \mathcal{F}_n \vdash \neg\varphi$  for any  $\varphi \in ASS(B)$ .

Prakken & Sartor's notion of defeat among arguments is built up from two other notions, 'rebutting' and 'undercutting' an argument. An argument  $A$  *rebutts* an argument  $B$  iff  $A$  conclusion-to-conclusion attacks  $B$  and either  $A$  is strict and  $B$  is defeasible, or  $A$ 's default rules involved in the conflict have no lower priority than  $B$ 's defaults involved in the conflict. Identifying the involved defaults and applying the priorities to them requires some subtleties for which the reader is referred to Prakken & Sartor [1996; 1997b] and Prakken [1997]. The source of the priorities will be discussed below.

An argument  $A$  *undercuts* an argument  $B$  precisely in case of the second kind of conflict (attack on an assumption). Note that it is not necessary that the default(s) responsible for the attack on the assumption has/have no lower priority than the default containing the assumption. Note also that Prakken & Sartor's undercutters capture a different situation than Pollock's: their undercutters attack an explicit non-provability assumption of another argument (in Section 3 called 'assumption attack'), while Pollock's undercutters deny the relation between premises and conclusion in a non-deductive argument.

Prakken & Sartor's notion of defeat also differs from that of Pollock [1995]. An inessential difference is that their notion allows for 'subargument defeat';

this is necessary since their definition of the status of arguments is not explicitly recursive (cf. Subsection 4.1). More importantly, Prakken & Sartor regard undercutting defeat as prior to rebutting defeat.

DEFINITION 68. (Defeat.) An argument  $A$  *defeats* an argument  $B$  iff  $A = \square$  and  $B$  attacks itself, or else if

- $A$  undercuts  $B$ ; or
- $A$  rebuts  $B$  and  $B$  does not undercut  $A$ .

As mentioned above in Subsection 4.1, the empty argument serves to adequately deal with self-defeating arguments. By definition the empty argument is not defeated by any other argument.

The rationale for the precedence of undercutters over rebutters is explained by the following example.

EXAMPLE 69. Consider

$$\begin{aligned} r_1 : & \sim \neg \textit{Brutus is innocent} \Rightarrow \textit{Brutus is innocent} \\ r_2 : & \varphi \Rightarrow \neg \textit{Brutus is innocent} \end{aligned}$$

Assume that for some reason  $r_2$  has no priority over  $r_1$  and consider the arguments  $[r_1]$  and  $[\dots, r_2]$ .<sup>15</sup> Then, although  $[r_1]$  rebuts  $[\dots, r_2]$ ,  $[r_1]$  does not defeat  $[\dots, r_2]$ , since  $[\dots, r_2]$  undercuts  $[r_1]$ . So  $[\dots, r_2]$  strictly defeats  $[r_1]$ .

Why should this be so? According to Prakken & Sartor, the crux is to regard the assumption of a rule as one of its conditions (albeit of a special kind) for application. Then the only way to accept both rules is to believe that Brutus is not innocent: in that case the condition of  $r_1$  is not satisfied. By contrast, if it is believed that Brutus is innocent, then  $r_2$  has to be rejected, in the sense that its conditions are believed but its consequent is not ('believing an assumption' here means not believing its negation). Note that this line of reasoning does not naturally apply to undercutters Pollock-style, which might explain why in Pollock's [1995] rebutting and undercutting defeaters stand on equal footing.

Finally, we come to Prakken & Sartor's definition of the status of arguments. As remarked above, they use the grounded semantics of Definition 7. However, they change it in one important respect. This has to do with the origin of the default priorities with which conflicting arguments are compared.

In artificial intelligence research the question where these priorities can be found is usually not treated as a matter of common-sense reasoning. Either a fixed ordering is simply assumed, or use is made of a specificity ordering, read off from the syntax or semantics of an input theory. However, Prakken

<sup>15</sup>We abbreviate arguments by omitting their conclusions and only giving the names of their defaults. Furthermore, we leave implicit that  $r_2$ 's antecedent  $\varphi$  is derived by a subargument of possibly several steps.

& Sartor want to capture that in many domains of common-sense reasoning, like the law or bureaucracies, priority issues are part of the domain theory. This even holds for specificity; although checking which argument is more specific may be a logical matter, deciding to prefer the most specific argument is an extra-logical decision. Besides varying from domain to domain, the priority sources can also be incomplete or inconsistent, in the same way as ‘ordinary’ domain information can be. In other words, reasoning about priorities is defeasible reasoning. (This is why our example of the introduction contains a priority argument, viz.  $A$ ’s use of (9) and (10).) For these reasons, Prakken & Sartor want that the status of arguments does not only *depend* on the priorities, but also *determines* the priorities. Accordingly, priority conclusions can be defeasibly derived within their system in the same way as conclusions like ‘Tweety flies’.<sup>16</sup>

To formalise this, Prakken & Sartor need a few technicalities. First the first-order part of the language is extended with a special twoplace predicate  $\prec$ . That  $x \prec y$  means that  $y$  has priority over  $x$ . The variables  $x$  and  $y$  can be instantiated with default names. This new predicate symbol should denote a strict partial order on the set of defaults that is assumed by the metatheory of the system. For this reason, the set  $\mathcal{F}_n$  must contain the axioms of a strict partial order:

$$\begin{aligned} \textit{transitivity:} \quad & \forall x, y, z. x \prec y \wedge y \prec z \supset x \prec z \\ \textit{asymmetry:} \quad & \forall x, y. x \prec y \supset \neg y \prec x \end{aligned}$$

For simplicity, some restrictions on the syntactic form of priority expressions are assumed.  $\mathcal{F}_c$  may not contain any priority expressions, while in the defaults priority expressions may only occur in the consequent, and only in the form of conjunctions of literals (a literal is an atomic formula or a negated atomic formula). This excludes, for instance, disjunctive priority expressions.

Next, the rebut and defeat relations must be made relative to an ordering relation that might vary during the reasoning process.

DEFINITION 70. For any set  $S$  of arguments

- $\prec_S = \{r \prec r' \mid r \prec r' \text{ is a conclusion of some } A \in S\}$
- $A$  (strictly)  $S$ -defeats  $B$  iff, assuming the ordering  $\prec_S$  on  $\Delta$ ,  $A$  (strictly) defeats  $B$ .

The idea is that when it must be determined whether an argument is acceptable with respect to a set  $S$  of arguments, the relevant defeat relations are verified relative to the priority conclusions drawn by the arguments in  $S$ .

---

<sup>16</sup>For some non-argument-based nonmonotonic logics that deal with this phenomenon, see Grosz [1993], Brewka [1994a; 1996], Prakken [1995] and Hage [1997]; see also Gordon’s [1995] use of [Geffner & Pearl, 1992].

DEFINITION 71. An argument  $A$  is *acceptable* with respect to a set  $S$  of arguments iff all arguments  $S$ -defeating  $A$  are strictly  $S$ -defeated by some argument in  $S$ .

Note that this definition also replaces the second occurrence of defeat in Definition 6 with strict defeat. This is because otherwise it cannot be proven that no two justified arguments are in conflict with each other.

Prakken & Sartor then apply the construction of Proposition 9 with Definition 71. They prove that the resulting set of justified arguments is unique and conflict-free and that, when  $S$  is this set, the ordering  $<_S$  is a strict partial order. They also prove that if an argument is justified, all its sub-arguments are justified.

We illustrate the system with the following example.

EXAMPLE 72. Consider an input theory with empty  $\mathcal{F}_c$ ,  $\mathcal{F}_n$  containing the above axioms for  $\prec$ , and  $\Delta$  containing the following defaults.

$$\begin{array}{ll} r_0: \Rightarrow a & r_4: \Rightarrow r_0 \prec r_3 \\ r_1: a \Rightarrow b & r_5: \Rightarrow r_3 \prec r_0 \\ r_2: \sim b \Rightarrow c & r_6: \Rightarrow r_5 \prec r_4 \\ r_3: \Rightarrow \neg a & \end{array}$$

The set of justified arguments is constructed as follows (for simplicity we ignore combinations of the listed arguments).

$$\begin{array}{ll} F^0 = \emptyset & <_0 = \emptyset \\ F^1 = \{\emptyset, [r_6]\} & <_1 = \{r_5 \prec r_4\} \\ F^2 = F^1 \cup \{[r_4]\} & <_2 = \{r_5 \prec r_4, r_0 \prec r_3\} \\ F^3 = F^2 \cup \{[r_3]\} & <_3 = <_2 \\ F^4 = F^3 \cup \{[r_2]\} & <_4 = <_3 \\ F^5 = F^4 & <_5 = <_4 \end{array}$$

Kowalski & Toni [1996] propose an alternative formalisation of reasoning about priorities, which does not require a change of the logic. They show how within the BDKT approach priority statements can be encoded with assumptions. This method requires that the notion of conflicting rules is expressed in the logical language of the system. Similar methods in non-argument-based approaches have been proposed by Gordon [1995] and Hage [1997].

### *Procedural form*

Like several other systems, Prakken & Sartor define a procedural version of their system in dialectical form. Compared to the other systems, its main feature is that it also covers debates about priorities. It will be discussed in some detail in Section 6.

*Comparison with Simari & Loui [1992]*

As remarked above, Prakken & Sartor's system is (in the version of [Prakken, 1997]) similar to Simari & Loui's. They both use the language of default logic, and their notions of an argument are quite similar: in particular, both systems use a modus ponens rule for defaults. Finally, both systems use grounded semantics and both have a procedural version in dialectical form. However, we have also seen that Prakken & Sartor extend Simari & Loui's system in a number of respects: their defaults are not twoplace but threeplace, which makes it possible to distinguish rebutting from assumption attack; they allow for comparing arguments on any ground, and they allow for debates on these grounds.

### 5.8 Nute's Defeasible Logic

A development closely related to defeasible argumentation is so-called 'defeasible logic', initiated by Donald Nute, e.g. [1994].<sup>17</sup> In both fields the notion of defeat is central. However, while in defeasible argumentation defeat is among arguments, in defeasible logic it happens between rules. Nevertheless, the approaches are sufficiently similar to warrant a discussion of defeasible logic in this chapter.

In several publications Nute has developed a family of such logics. For explanatory purposes we discuss the simplest version, described in [Nute & Erk, 1995]. In a way this is unfair, since this version has a problem that is absent in the other versions. However, it is instructive to see what the problem is, and we shall indicate how Nute deals with it in his other work.

Nute's systems are based on the idea that defaults are not propositions but inference licenses. Thus Nute's defeasible rules are, like Reiter's defaults, one-directional. However, unlike Reiter's defaults they are twoplace; assumption attacks are dealt with by an explicit category of defeater rules, which are comparable to Pollock's undercutting defeaters, although in Nute's case they are, like his defeasible rules, not intended to express general principles of inference but, as in default logic, domain specific generalisations.

As for the underlying logical language, since Nute's aim is to develop a logic that is efficiently implementable, he keeps the language as simple as possible. It has three categories of one-direction rules, viz. *strict* rules  $A \rightarrow p$ , *defeasible rules*  $A \Rightarrow p$  and *defeaters*  $A \rightsquigarrow p$ . In all three cases  $p$  is a strong literal, i.e., an atomic proposition or a classically negated atomic proposition, and  $A$  is a finite set of strong literals. Defeaters must be read as 'if  $A$  then it might be that  $p$ '. Defeaters cannot be used to derive formulas; they can only be used to block an application of a rule  $B \Rightarrow \neg p$ . An example

---

<sup>17</sup>In fact, Nute [1994] also counts systems for defeasible argumentation as defeasible logics.

is ‘Genetically altered penguins might fly’, which undercuts ‘Penguins don’t fly’. Thus Nute has, like Pollock, both rebutting and undercutting conflicts between arguments.

Arguments can be formed by chaining rules into trees, and conflicting arguments are compared with the help of an ordering on the rules. Actually, Nute does not work with an explicit notion of argument; instead he incorporates it in two notions of derivability, strict ( $\vdash$ ) and defeasible ( $\vdash\sim$ ) derivability, to be explained below. To capture non-derivability, Nute does not use the familiar notions  $\nmid$  (meaning ‘not  $\vdash$ ’) and  $\nmid\sim$  (meaning ‘not  $\vdash\sim$ ’). Instead, his aim of designing a tractable system leads him to define two notions of *demonstrable* non-derivability  $\dashv$  and  $\dashv\sim$ , which require that a proof of a formula fails after finitely many steps.

As just stated, Nute’s assessment of arguments is implicit in his definitions of derivability. Nute has two core definitions, depending on when the last rule of the tree is strict or defeasible. (He has similar rules for  $\dashv$  and  $\dashv\sim$ .) The first definition detaches consequences of strict rules.

**DEFINITION 73.** (Strict derivability.)  $T \vdash p$  if

1.  $p \in T$ , or
2. There is a  $A \rightarrow p \in T$  such that for every  $a \in A$ ,  $T \vdash a$ .

The second definition detaches consequences of defeasible rules, taking into account all nonmonotonic proofs that derive the contrary:

**DEFINITION 74.** (Defeasible derivability.)  $T \vdash\sim p$  if there is a rule  $A \Rightarrow p \in T$  such that

1.  $T \dashv \neg p$ , and
2. for each  $a \in A$ ,  $T \vdash\sim a$ , and
3. for each  $B \rightarrow \neg p \in T$  there is  $b \in B$  such that  $T \dashv\sim b$ , and
4. for each  $C \Rightarrow \neg p \in T$  or  $C \rightsquigarrow \neg p \in T$ , either
  - (a) there is a  $c \in C$  such that  $T \dashv\sim c$  or
  - (b)  $A \Rightarrow p$  has higher priority than  $C \rightarrow \neg p$  (or than  $C \rightsquigarrow \neg p$ ).

Condition (1) says that the opposite of  $p$  must demonstrably be not strictly derivable. This gives strict arguments priority over defeasible arguments. For the rest, this definition has the recursive structure discussed above in Section 4.1. There must be a defeasible rule for  $p$  which, firstly, ‘fires’, i.e., of which all antecedents are themselves defeasibly derivable (condition 2) and which, secondly, is of higher priority than any conflicting rule which also fires: for any rule which is not lower, at least one antecedent must

be demonstrably non-derivable (conditions 3–4). As a special case, condition (3) implicitly gives priority to strict rules over defeasible rules; for the rest these priorities must be defined by the user (condition 4), although Nute pays much attention to the specificity criterion. Note that like Pollock [1995], Nute applies priorities to decide whether undercutting attack succeeds.

A literal can also be derived defeasibly from a strict rule, namely, when one of its antecedents is itself derived defeasibly. When there is a strict rule for  $p$ , the definition of defeasible derivability is simpler: since strict rules have priority over the other two categories, condition 4 can be dropped. In consequence, defeasible derivability from a strict rule can only be blocked by derivability from a conflicting strict rule.

Since Definitions 73 and 74 have the recursive structure of Definition 17, they share with this definition the problem that multiple assignments are not always avoided. Consider the following variant of Example 23.

EXAMPLE 75. Assume we have the following rules

1.  $\Rightarrow p$
2.  $p \Rightarrow q$
3.  $\Rightarrow \neg q$
4.  $\neg q \Rightarrow \neg p$

Three status assignments satisfy the above definitions.

- Status assignment 1:*  $T \mid\sim p, T \mid\sim q, T \not\sim \neg p, T \not\sim \neg q$ ;  
*Status assignment 2:*  $T \mid\sim \neg p, T \mid\sim \neg q, T \not\sim p, T \not\sim q$   
*Status assignment 3:*  $T \not\sim p, T \not\sim q, T \not\sim \neg p, T \not\sim \neg q$

Only the third assignment is intended by Nute. In his other work, e.g. [Nute, 1994], he reformulates Definitions 73 and 74, and also the rules for  $\not\sim$ , as conditions on finite proof trees for a formula. This solves the problem, since for the unintended status assignments no proof trees can be constructed. The crux is that  $\not\sim$  must also be established by constructing a finite proof tree (being a finite proof that a formula cannot be derived). And in the above example this is impossible.

Another problem inherited from Definition 17 is that Nute's system cannot capture floating conclusions (cf. Example 24). This is since an inference of  $p$  can only be blocked by a rule for  $\neg p$  if all antecedents for that rule are derivable. Since Nute has no third category 'defensible' in between '(demonstrably) derivable' and '(demonstrably) not derivable', two rules that are in an irresolvable conflict do not give rise to conclusions and thus cannot block other inferences.

Finally, Nute's system behaves in a somewhat peculiar way when a conflict involves strict rules, as in the following example:



1.  $x$  has children  $\Rightarrow x$  is married
2.  $x$  lives alone  $\Rightarrow x$  is a bachelor
3.  $x$  is married  $\rightarrow \neg x$  is a bachelor

In Nute's system only rules with directly contradicting heads are compared, and since strict rules prevail over defeasible rules, the outcome is that  $x$  is a bachelor, even if the first defeasible rule has priority over the second. This seems counterintuitive. In [Simari & Loui, 1992] and [Prakken & Sartor, 1997b] this problem does not occur, since there (3) is in the necessary facts  $\mathcal{F}_n$ , which count in testing whether conclusions contradict each other, for which reason the conflict is recognised as being between (1) and (2). It should be noted that in his most recent work Nute deals with this problem [Nute, 1997].

### *Evaluation*

Evaluating Nute's defeasible logic, we see that it is an instance of the recursive-definition variant of the multiple-status-assignments approach, without an intermediate notion of defensible arguments. Consequently, his system has some problems with zombie arguments and floating conclusions. On the positive side, Nute's system gives intuitive results for a large class of benchmark examples and is, due to its simple language and its transparent definitions, very suitable for implementation.

As for the relation with defeasible argumentation, although Nute never introduced 'argument' as a concept in his defeasible logics, his theory can easily be recast in terms of arguments. One way to do this is to chain Nute's rules into trees (analogously to Lin & Shoham or Vreeswijk) and call them arguments (these trees must not be confused with the above-mentioned proof trees, which are proofs that a formula is defeasibly derivable). With this definition of arguments, Definition 74 can be stated alternatively in the way Vreeswijk defines defeat among arguments. A first conclusion that may be drawn from such a translation is that Nute's logic for defeasible reasoning is closely related to other approaches discussed here. This close relation justifies the discussion of defeasible logic in this chapter. For work on formalising this relation see [Governatori & Maher, 2000]. Another conclusion is that arguments in Nute's logic defeat each other on the basis of information in top-rules only. This is due to the fact that a strong literal in Nute's system is defeasibly derivable only if the antecedent of the last rule applied is defeasibly derivable. This is in contrast with Vreeswijk's theory, in which arguments are compared and defeated in their entirety.

### 5.9 Defeasible argumentation in reasoning about events (Konolige, 1988)

Konolige's [1988] system ARGH (Argumentation with Hypotheses) was presented as a solution to the Yale Shooting Problem (YSP) [Hanks & McDermott, 1987]. Although the resulting formalism is still rather rudimentary, Konolige's discussion anticipates many issues and distinctions of later work, so that ARGH can be regarded as one of the forerunners of the field of defeasible argumentation.

The YSP concerns reasoning about events. The main problem to be dealt with is that sometimes the tendency of facts to 'persist' over time conflicts with the change of these facts by certain events. Konolige uses argumentation to allow various types of arguments based on considerations of persistence or change, and to adjudicate between conflicting arguments by means of principles of defeat. One such principle says that arguments based on change caused by events defeat arguments based on persistence.

The logical language of ARGH resembles McCarthy's [1969] situation calculus, where properties are attached to situations and events bring us in new situations, with new properties. This language is used for giving *world descriptions*. An example of a world-description is

$$W = \left\{ \begin{array}{lll} p, q, \neg r, s \mid s_0, & s_0 \rightarrow_{\alpha} s_1, & p, \neg q \mid s_1 \\ \text{The propositions} & \text{At situation} & \text{At situation} \\ p, q, \neg r \text{ and } s \text{ hold} & s_0, \text{ action } \alpha & s_1, \text{ the propo-} \\ \text{at situation } s_0. & \text{brings us to} & \text{situation } p \text{ still} \\ & \text{situation } s_1. & \text{holds, but } q \\ & & \text{does not.} \end{array} \right\}$$

This scheme forms a single world description, consisting of three statements. The second statement is an *event description*, connecting the two *situation descriptions* that are stated on the first and the third line. Thus, typically, the letters  $s_0, s_1, \dots$  denote *situations*, the letters  $p, q, r, \dots$  denote propositions or *properties* that hold at situations, the letters  $\alpha, \beta, \dots$  denote actions or *events*. In ARGH, a world description can be partial: in the example above, neither  $\neg r, s$ , nor their negations are specified at  $s_1$ .

The purpose of argumentation in ARGH is to fill in partially described worlds as much as possible, by drawing conclusions regarding missing propositional values. Konolige considers three elementary types of inference rules for constructing arguments (which because of their generality are comparable to Pollock's notion of defeasible reasons).

	<i>Notation</i>	<i>Meaning</i>
Forward persis- tence:	$p   s_i \rightarrow_{\text{persist}} p   s_i + 1$	If $p$ holds at $s_i$ , then it is likely that $p$ holds at the next situation $s_{i+1}$ .
Backward per- sistence:	$p   s_i + 1 \rightarrow_{\text{persist}} p   s_i$	If $p$ holds at $s_{i+1}$ , then it is likely that $p$ is inherited from the previous situation $s_i$ .
A $p$ -establishing action:	$ s_i \rightarrow_{\alpha} p   s_i + 1$	Doing $\alpha$ in $s_i$ results in $p$ at $s_i + 1$ , defeasibly.

Labels such as ‘persist’,  $\alpha$  and  $\beta$ , are not typed, that is, do not belong to a certain class of actions or propositions.

The above notation is used as a basis for constructing compound arguments and for performing defeasible reasoning. For example, the world

$W =$	$p   s_1$	Proposition $p$ holds at $s_1$
	$s_1 \rightarrow_{\text{wait}} s_2 \rightarrow_{\beta} s_3$	At $s_1$ , waiting brings us in $s_2$ ; then, performing $\beta$ in $s_2$ , brings us in $s_3$ .

enables a number of arguments such as

	<i>Argument</i>	<i>For</i>
$A$	$p   s_1 \rightarrow_{\text{persist}} p   s_2$	$p   s_2$
$B$	$p   s_2 \rightarrow_{\text{persist}} p   s_3$	$p   s_3$
$A; B$	$A$ followed by $B$	$p   s_3$
$B'$	$p   s_2 \rightarrow_{\beta} \neg p   s_3$	$\neg p   s_3$
$A; B'$	$A$ followed by $B'$	$\neg p   s_3$
$C$	$\neg p   s_3 \rightarrow_{\text{persist}} \neg p   s_2$	$\neg p   s_2$
$A; B'; C$	$A, B'$ followed by $C$	$\neg p   s_2$

$A; B$  and  $A; B'$  are conflicting arguments. An argument for  $\neg p | s_2$  is  $A; B'$ , followed by  $C$  (backward persistence). In this way, the arguments  $A$  and  $A; B'; C$  compete for  $p$ .

To adjudicate among competing arguments, Konolige formulates a number of *rules of defeat*, such as the rule that event arguments have priority over persistence arguments. However, he also observes that this priority rule is defeasible, by giving an example in which backwards persistence is stronger than that change-by-event. In fact, one of Konolige’s main observations is that any general, domain-independent priority principle will be very weak, and that information from the semantics of the domain will be the most important way of deciding among competing arguments. Such semantic information could, for instance, express the strength of the tendency of certain facts to persist over time. For example, the fact that a house will remain at its place is more likely to persist over time than the fact that a car will remain at its place. Thus Konolige anticipates later research on reasoning with and about domain specific priorities (see above, Section 5.7).

### *Evaluation*

Evaluating Konolige's formalism, we can say that it is tailored to one particular problem, viz. reasoning about a changing world. However, for defeasible argumentation the main value of Konolige's system is not this application but the fact that it was one of the earliest argument-based accounts of defeasible reasoning, anticipating many of the issues arising in later work.

### *5.10 A brief overview of other work*

We end this section with a brief overview of other work on logics for defeasible argumentation.

#### *Loui [1987]*

One of the initiators of the field of defeasible argumentation was Loui [1987]. On the basis of the same language as later used in [Simari & Loui, 1992], Loui defines arguments as graphs in which the links are formed by first-order inferences or default applications. Since defaults are twoplace, Loui only has rebutting attack. In particular, an argument  $A$  is a counterargument of an argument  $B$  if the root of  $A$  is inconsistent with some node in  $B$ . Loui orders conflicting arguments in terms of four syntactic specificity criteria, and then defines an argument  $A$  to be justified iff it is undefeated (with respect to its top node) and all its counterarguments (i.e., all argument attacking another node of  $A$ ) are defeated by a counterargument.

As this brief description shows, Loui's [1987] system already has all the elements of an argumentation system. The ideas of this paper have been very influential, but the formalism has some technical flaws, for which reason it has not survived. His paper with Simari was Loui's own attempt to overcome the flaws. Loui's most recent work (e.g. [Loui & Norman, 1995; Loui, 1998]) addresses the procedural aspects of argumentation.

#### *Connection with truth-maintenance systems*

Systems for defeasible argumentation are related to so-called truth-maintenance systems (TMSs). A TMS is a bookkeeping system for a reasoning system, in which logical dependencies among propositional beliefs, or assertions, are represented and maintained to preserve consistency of the reasoning system. There exists several TMSs, such as Justification-Based [Doyle, 1979], Assumption-Based, [De Kleer, 1986] and Logic-Based TMSs. Basically, in all TMSs all assertions are connected via a network of dependencies and all TMSs do some form of dependency-directed backtracking. In Justification-Based TMSs, for example,

- The structure of the assertions themselves is left unspecified. Each supported belief (assertion) has a so-called *justification*.

- Each justification has two parts:
  1. An IN-List which supports beliefs held.
  2. An OUT-List which supports beliefs not held.
- An assertion is connected to its justification by an arrow; one assertion can feed another justification thus creating the network.
- Assertions may be labelled with a belief status.
- An assertion is valid if every assertion in its IN-List is believed and none in its OUT-List are believed.
- An assertion is *non-monotonic* if the OUT-List is not empty or if any assertion in the IN-List is *non-monotonic*.

Thus, the concepts and ideas are similar in spirit to those underlying argumentation systems. For instance, the issues of multiple and nonexisting status assignments have been studied in the literature on Justification-Based TMSs as the issues of multiple and nonexisting labellings of a dependency network. Since [Doyle, 1979], a variety of TMSs have been developed as a means of implementing nonmonotonic reasoning. The relation between TMSs and nonmonotonic reasoning is further discussed in [Martins & Reinfrank, 1991]. Baker & Ginsberg [1989] establish a connection with argument and debate.

*Krause et al. [1995]*

Recently, the system of Krause *et al.* [1995], further explored by Elvang-Gøransson & Hunter [1995], has attracted some attention in the multi-agent community, as a component of models of negotiation; cf. [Parsons *et al.*, 1998]. In this system, arguments are essentially a (Premises, Conclusion) pair, where the conclusion follows from the set Premises according to a system of intuitionistic logic. The conclusion of an argument can, as in Pollock's system, have a degree of belief, which allows arguments to be ordered using numerical (e.g. probabilistic) information. The only type of conflict is conclusion-to-conclusion attack. However, Krause *et al.* distinguish two subtypes, "rebutting" and "undercutting" conflict, with a deviating use of the term 'undercutter': in their terms,  $A$  undercuts  $B$  iff  $A$  rebuts (i.e., conclusion-to-conclusion-attacks) a subargument of  $B$ . (An argument  $(S, \varphi)$  is a subargument of an argument  $(T, \psi)$  iff  $S \subseteq T$ .)

The main feature that sets this system apart from other systems, is the definition of the status of arguments. Given rebutting and undercutting relations between arguments, arguments are divided into the following categories (relative to a certain input theory  $\Delta$ ).

DEFINITION 76. (Argument classes.)

- A1 is the class of all arguments that can be made from  $\Delta$ .
- A2 is the class of all consistent arguments that can be made from  $\Delta$ .
- A3 is the class of all consistent arguments from  $\Delta$  without rebutting arguments.
- A4 is the class of all consistent arguments from  $\Delta$  without undercutting arguments.
- A5 is the class of all arguments with empty set of Premises.

Observe that  $A5 \subseteq A4 \subseteq A3 \subseteq A2 \subseteq A1$ . (Note that rebutting an argument implies undercutting it.) Accordingly, arguments in smaller classes are regarded as better than arguments in larger classes. Krause *et al.* also consider a refinement of this ordering in terms of the degrees of belief of arguments.

In our opinion, a drawback of this definition is that it does not capture reinstatement.

#### *Argument-based proof theories for preferential entailment*

Two argumentation-theoretic proof theories have been proposed for a preferred-model semantics. As explained in Section 2, in preferential entailment defaults are represented as first-order material implications with special ‘normality conditions’, as in

- (1)  $\forall x. \text{Bird}(x) \wedge \neg \text{ab}_1(x) \supset \text{Canfly}(x)$
- (2)  $\forall x. \text{Penguin}(x) \wedge \neg \text{ab}_2(x) \supset \neg \text{Canfly}(x)$

First-order theories containing such defaults are then semantically interpreted by only looking at those models where the extension of the  $\text{ab}_i$  predicates are minimal (with respect to set inclusion), which captures the assumption that the world is as normal as possible.

The proof-theoretic idea is that arguments are (in their simplest form) a set of normality statements that can be added to a certain theory to derive certain conclusions. (This is essentially a special case of Bondarenko *et al.*'s [1997] assumption-based definition of an argument.) For instance, suppose that the defaults (1) and (2) are part of a first-order theory

$$T = \{1, 2\} \cup \{\text{Penguin}(\text{Tweety}), \forall x. \text{Penguin}(x) \supset \text{Bird}(x)\}$$

Then  $A = \{\neg \text{ab}_1(\text{Tweety})\}$  is an argument for the conclusion  $\text{Canfly}(\text{Tweety})$ , since  $T \cup A \vdash \text{Canfly}(\text{Tweety})$ , and  $B = \{\neg \text{ab}_2(\text{Tweety})\}$  is an argument for  $\neg \text{Canfly}(\text{Tweety})$ , since  $T \cup B \vdash \neg \text{Canfly}(\text{Tweety})$ . In order to capture floating conclusions (cf. Example 25), the general form of an argument is

not that of a set but of a collection of sets of normality assumptions (an alternative form is that of a disjunction of conjunctions of such assumptions). Conflicting arguments can be compared in terms of an ordering of the normality assumptions.

*Baker & Ginsberg [1989]*

Baker & Ginsberg [1989] have applied this idea to the semantics of so-called prioritised circumscription. In their proof theory, an argument  $A$  *rebuts* another argument  $B$  if  $A$  and  $B$  have contradictory conclusions, and if  $A$ 's least default is not inferior to  $B$ 's least default, while  $A$  *refutes*  $B$  if in addition its least default has priority over the least default of  $B$ . A defeasible proof then has a dialectical form, which form will be discussed in detail in Section 6. Baker & Ginsberg prove that this proof theory is sound and complete with respect to the model theory of prioritised circumscription.

*Geffner & Pearl [1992]*

Geffner & Pearl [1992] have proposed similar ideas, in a proof theory for their “conditional entailment” (see also [Geffner, 1991], for an application to logic programming’s negation as failure). When representing default rules, a minor difference with Baker & Ginsberg is that they use positive ‘applicability’ atoms  $\delta_i$  instead of negated abnormality atoms. In their preferred model semantics they then prefer those models which make as few applicability atoms false as possible. In ordering applicability atoms, Geffner and Pearl define a class of “admissible orderings” which, if respected by the preference relation on models, reflects the notion of specificity. Although this notion is the only source of priorities that Geffner & Pearl consider, their formalism seems not to exclude orderings on the  $\delta_i$ 's based on other standards.

Geffner & Pearl’s proof theory is sound and complete with respect to conditional entailment. They also define an architecture for (incompletely) implementing the proof theory as a computer program, which has the dialectical flavour that will be the topic of Section 6. Bondarenko *et al.* [1997] conjecture that it computes the grounded semantics of Definition 7.

*Evaluation*

The idea of providing a model-theoretic foundation for defeasible argumentation is interesting, but as we remarked at the end of Section 3, a critical test for such approaches is whether the resulting criteria for model preference are sufficiently natural. For certain restricted applications this test might succeed, but it remains to be seen to what extent this approach can be generalised; for instance, to argumentation systems that allow for inductive, analogical or abductive arguments.

Verheij [1996]

Verheij combines ideas of Lin & Shoham and Vreeswijk on the structure of arguments with Pollock's partial status assignments into a formalism called CumulA. This system has three distinctive features. The first is a new type of argument called 'coordinated argument', which combines two arguments for the same conclusion. For instance, from the arguments '*The sun is shining. So, it is a beautiful day*' and '*The sky is blue. So, it is a beautiful day*' it is possible to construct a new argument '*The sun is shining; the sky is blue. So, it is a beautiful day*'. Verheij stresses that this is not the same as an ordinary argument with two premises: the semicolon expresses that each premise on its own also supports the conclusion. With coordinated arguments Verheij wants to capture the 'accrual of arguments', i.e., the phenomenon that a combination of arguments that are individually defeated by another argument, possibly defeats that argument.

EXAMPLE 77. (Accrual of arguments.) Consider the arguments

- A: Peter robbed a person, therefore Peter should be punished.
- B: Peter injured a person, therefore Peter should be punished.
- C: Peter is a minor offender and should therefore not be punished.
- D: Peter robbed a person. He injured that person too. Therefore, Peter should be punished.

According to Verheij, it is conceivable that the coordination *D* of *A* and *B* prevails over *C*, even if *C* would prevail over *A* and *B* when these are considered individually. Accordingly, Verheij allows that a status assignment makes a coordinated argument 'in' even when any of its components would be 'out' when present without the others. On the other hand, any status assignment should make a coordinated argument 'in' if already one of its components is 'in'.

A second feature of CumulA is that it generalises other argumentation systems by making defeat a relation between *sets* of arguments. According to Verheij this enables a more natural formalisation of certain types of defeat. Verheij also argues that several types of defeat, such as Pollock's undercutters, cannot be defined in terms of inconsistency between conclusions of arguments. For this reason, in CumulA the relation of 'defeat' is, as in [Dung, 1995], a primitive notion and can be further defined in various ways, which may but need not be triggered by inconsistency of conclusions. Verheij claims that his treatment of defeaters is able to capture a wide range of types of defeat proposed in the literature.

A final feature of CumulA is its further development of Lin & Shoham's and Vreeswijk's notions of argument structures and sequences. In particular,



Cumula models the replacement of a premise with an argument that has this premise as conclusion. Such a move is very common in actual debates but has not yet received much attention in the field of defeasible argumentation (but see Loui, 1998). Verheij also develops an elegant notation that shows how the status of arguments can change when more arguments are taken into account (Figure 9).

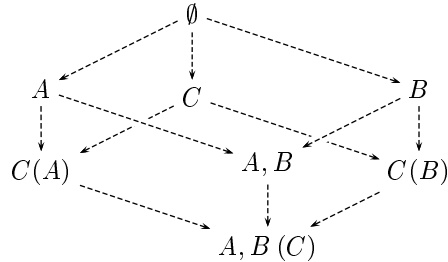


Figure 9. Stages of argumentation when  $C$  defeats  $A$ ,  $C$  defeats  $B$ , but  $\{A, B\}$  defeats  $C$ . Each node represents a partial defeat status assignment (cf. Definition 47), and reflects a ‘stage’ in the argumentation process. Arguments between parentheses have the status ‘defeated’, the other arguments have the status ‘undefeated’.

#### Other work

Finally, we mention other relevant work on logics for defeasible argumentation.

Marek *et al.* [1990; 1992] aim to capture the main existing nonmonotonic logics in a general framework of so-called ‘nonmonotonic rule systems’. The basic notion is not that of an argument but that of a (one-direction) rule. They define a notion of extensions of a given rule system as a set of formulas that has certain closure and completeness properties with regard to rule application. Bondarenko *et al.* [1997] prove that these extensions correspond to stable semantics. Marek *et al.*’s ideas bear some resemblance to Lin & Shoham’s system. Both systems aim to be a general framework for capturing nonmonotonic logics, both work with one-direction rules, and Marek *et al.*’s notion of extensions is related to Lin & Shoham’s notion of a complete argument structure. Finally, neither have a mechanism of defeat among arguments (or proofs).

Benferhat *et al.* [1993] study argumentative reasoning with inconsistent databases. An argument for a formula is a consistent subset of a database (which is a set of logical formulas) that classically entails the formula. Conflicts between arguments are resolved with an ordering on the elements of

the database. The approach and its relation with inconsistency handling approaches (cf. Section 2.1) and other argumentation systems is further investigated by Benferhat *et al.* [1995], Cayrol [1995] and Amgoud & Cayrol [1997].

The BDKT framework has triggered further work in the area of logic programming. For instance, Dung [1993] has applied his own framework to the semantics of extended logic programming. Thielscher [1996] has defined a semantics and proof theory for drawing sceptical conclusions from multiple status assignments, based on an adapted version of Dung's [1995] framework. And Jakobovits [Jakobovits & Vermeir, 1999; Jakobovits, 2000] has generalised Dung's version of the BDKT framework by defining several weak notions of argument extensions, and examining the relation with the various BDKT semantics.

Finally, Starmans [1996] carries the ideas of defeasible argumentation to a multi-agent environment, where more than two parties participate in a dispute. Part of this endeavour is to show that  $n$ -party disputes, where  $n \geq 3$ , involve a richer arsenal of speech acts (question, demand for clarification, refusal of adduced evidence) and other types of attack than just rebutting or undercutting counterarguments (such as such as just refusing to accept a certain claim). As debate proceeds, on the basis of the individual theories a so-called *aggregated* theory is formed, which contains the claims that are supported collectively by the group of disputants. This process can be constrained by so-called *principles of preservation*. Starmans discusses a number of such principles analogous to choice principles in the theory of social choice.

## 6 DIALECTICAL FORMS OF ARGUMENTATION SYSTEMS

So far mainly semantical aspects have been discussed, where the main focus was on properties of *sets* of arguments. In this section we shall go deeper into proof-theoretical, or procedural aspects of argumentation, where the chief concern is to establish the status of *individual* arguments. Several argumentation systems have been formulated in dialectical style [Baker & Ginsberg, 1989; Simari & Loui, 1992; Vreeswijk, 1993b; Simari *et al.*, 1994; Dung, 1994; Brewka, 1994b; Prakken & Sartor, 1996; Loui, 1998; Garcia *et al.*, 1998; Prakken, 1999; Kakas & Toni, 1999; Jakobovits, 2000]. (It should be noted that Loui [1998] does not regard the dialectical style merely as a reformulation of declarative nonmonotonic logics, but as a formalism in its own right, capturing the “essentially constructive” nature of defeasible reasoning, which, Loui argues, cannot be captured by declarative formalisms.)

The common idea can be explained in terms of a dialogue game between two players, a proponent and an opponent of an argument. A dialogue is an alternating series of moves by the two players. The proponent starts with an

argument to be tested, and each following move consists of an argument that attacks the last move of the other party with a certain minimum force. The initial argument provably has a certain status if the proponent has a winning strategy, i.e., if he can make the opponent run out of moves whatever moves the opponent makes. The exact rules of the game depend on the semantics it is meant to capture. A natural idea here is that of dialectical asymmetry. For instance, if the game reflects sceptical reasoning, i.e., if it is meant to test whether an argument is justified, the proponent's arguments can be required to be strictly defeating while the opponent's moves may be just defeating. If, on the other hand, the game reflects credulous reasoning, these rules can be reversed (as suggested by Prakken [1999]).

Let us introduce the concept of dispute more formally by making use of an adapted version of what is called a 'dialogue' in [Prakken & Sartor, 1996] and 'argument game' in [Loui, 1998]. It is meant to capture sceptical reasoning.

DEFINITION 78. (Disputes.) A *dispute* on an argument  $A$  is a non-empty sequence of arguments  $move_i = (Player_i, A_i)$  ( $i > 0$ ) with  $A_1 = A$ , in which one player, denoted by **PRO**, uses odd-numbered moves to try to establish  $A$  and another player, denoted by **CON**, uses even-numbered moves to try to prevent  $Player_1$ 's success.

1.  $Player_i = \mathbf{PRO}$  iff  $i$  is odd; and  $Player_i = \mathbf{CON}$  iff  $i$  is even;
2. If  $Player_i = Player_j = \mathbf{PRO}$  and  $i \neq j$ , then  $A_i \neq A_j$ ;
3. If  $Player_i = \mathbf{PRO}$  ( $i > 1$ ), then  $A_i$  strictly defeats  $A_{i-1}$ ;
4. If  $Player_i = \mathbf{CON}$ , then  $A_i$  defeats  $A_{i-1}$ .

The first condition stipulates that **PRO** begins and then the players take turns, while the second condition prevents the proponent from repeating its attacks. The remaining two conditions form the heart of the definition: they state the burdens of proof for **PRO** and **CON**. Thus, **PRO** is required to establish  $A$  while **CON** need only provide nuisance defeaters.

The various authors format their disputes in different ways. Vreeswijk [1993b; 1995] uses a format that displays the depth of the proof tree and is able to represent exhaustive disputes. (See below.) Here we have instead used a simplified version of the format used by [Dung, 1994; Prakken & Sartor, 1997b; Loui, 1998]. This format is simple and compact, but does not represent the depth of the proof tree.

EXAMPLE 79. Let  $A, B, C$  and  $D$  be arguments such that  $B$  and  $D$  defeat  $A$ , and  $C$  defeats  $B$ . Then a dispute on  $A$  may run as follows:

**PRO**:  $A$ , **CON**:  $B$ , **PRO**:  $C$

In this dispute **PRO** advances  $A$  as an argument supporting the main thesis. (Arguments are conceived as primitive concepts here, so that the main thesis is left unspecified.) Both  $B$  and  $D$  defeat  $A$ , which means that **CON** has two choices in response to  $A$ . **CON** chooses to respond with  $B$  in the second move. Then  $C$  is the only argument defeating  $B$ , so that **PRO** has no choice than to respond with  $C$  in the third move. There are no arguments against  $C$ , so that **CON** cannot move and loses the dispute. As a result,  $A$  and  $C$  are established, and  $B$  is overruled by  $C$ .

A dispute in which **CON** follows an optimal strategy is

**PRO:**  $A$ , **CON:**  $D$

So in this game, under these rules, there is no winning strategy for player 1, **PRO**. The only reason why **PRO** wins the first dispute is that **CON** chooses the wrong argument, viz.  $B$ , in response to  $A$ . In fact, **CON** is in the position to win every game, provided it chooses the right moves. In other words, **CON** possesses a winning strategy.

The concept of dispute presently discussed can be characterised as a so-called *argument game*. An argument game is a ‘one-dimensional’ dispute in which each player may respond only once to each argument advanced by the opponent, and if that argument turns out to be ineffective, that player may not try a second reply to the same argument. Thus, no backtracking is allowed. This fact makes argument games into what is officially known as *two-player zero-sum games*, including the concepts that come with it, the most important of which is strategy.

#### *Exhaustive dispute*

The opposite of an argument game is a so-called *exhaustive dispute*. An exhaustive dispute is a dialogue in which each player is allowed to try out every possible rebuttal in reply to the arguments of its opponent. If a player discovers that it has put forward the wrong argument, it can recover from its mistake by trying another argument, provided there are such alternatives.

In displaying exhaustive disputes, we follow the format of Vreeswijk [1993b; 1995], in which the depth of the proof tree is represented by vertical bars in the left column:

- |    |  |                           |
|----|--|---------------------------|
| 1. | <b>PRO</b> : argument 1                  | [justification]           |
| 2. | <b>CON</b> : reply                       | [justification for reply] |
| 3. | <b>PRO</b> : reply to reply              | ...                       |
| 4. | <b>CON</b> : 2nd reply to argument 1     | ...                       |
| 5. | <b>PRO</b> : reply to 2nd reply          | ...                       |
| 6. | <b>CON</b> : reply to reply to 2nd reply | ...                       |
| ⋮  | ⋮  | ⋮                         |

With the arguments presented in Example 79, **CON** has two strategies: one employing *B* and one employing *D*; let us refer to these strategies as *strategy B* and *strategy D*, respectively. As remarked above, when the players are engaged in an argument game, **CON** must choose between strategy *B* and strategy *D*. What **CON** cannot do is deploying *B* and *D* one after the other. In an exhaustive dispute, on the other hand, **CON** has the opportunity to try both strategies in succession:

1. | **PRO** : *A* [A]
2. || **CON** : *B* [*B* defeats *A*]
3. ||| **PRO** : *C* [*C* defeats *B*]
4. ||| **CON** : *D* [*D* defeats *A*]

At line 1, **PRO** advances *A* as an argument supporting the main thesis. (The main thesis is left unspecified here.) Both *B* and *D* defeat *A*, so that **CON** has two choices in response to *A*. **CON** chooses to respond with *B* at line 2. *C* is the only argument defeating *B*, so that **PRO** responds with *C* at line 3. There are no counterarguments to *C*, so that **CON** backtracks and searches new counterarguments to *A*. **CON** finds *D* as a new counterargument to *A*. At line 4, **CON** advances *D* in reply to *A*. There are no arguments against *D*, so that **PRO** cannot move and loses the dispute. As a result, we know that *A* cannot be established as justified.

Had **CON** responded with *D* instead of *B* at line 2, then the dispute would be settled within 2 moves:

1. | **PRO** : *A* [A]
2. || **CON** : *D* [*D* defeats *C*]

The choice and order of moves is determined by the players.

In the above definition as well as in most approaches a move consists of a complete argument. This means that the search for an individual argument is conducted in a ‘monological’ fashion, determined by the nature of the underlying logic; only the process of considering counterarguments is modelled dialectically. A notable exception is [Loui, 1998], in which arguments are constructed piecewise (beginning with the top-rule) and dialogue moves consist of

- attacking the conclusion of an unfinished argument,
- challenging an unfinished argument, or
- extending an unfinished argument in a top-down fashion on request of the opponent.

Another feature of Loui’s protocol is that, to reflect the idea of resource-bounded reasoning, every move consumes resources except requests to the opponent to extend unfinished arguments.

*Completeness results*

An important objective in the dialectic approach is a correspondence between the various argument-based semantics and the different forms of dispute.

Dung [1994] establishes a correspondence between the semantics defined in Definition 7 (grounded semantics) and his notion of argument game. Dung's game is similar to the one of Definition 78, but it is different in two respects: it does not have the nonrepetition rule (2), and it allows that **PRO**'s moves are, like **CON**'s moves, just defeating. On the other hand, Prakken & Sartor [1997b] show that Dung's result also holds for Definition 78. Thus they give a justification to the nonrepetition rule and the dialectical asymmetry, in the sense that these features make debate more efficient while preserving semantical soundness of the game. Intuitively, this is since the only effect of these features is the termination of dialogues that could otherwise go on forever: thus they do not deny **PRO** any chance of winning the debate.

As for some details, Dung's idea is to establish a mapping for which

- arguments in the set  $F^{2i}$  map to arguments for which **PRO** has a winning strategy that results in an argument game of at most  $2i$  moves
- arguments *not* in  $F^{2i+1}$  map to arguments for which **CON** has a winning strategy that results in an argument game of at most  $2i + 1$  moves

Another completeness result is established by Vreeswijk [1995], between a particular form of exhaustive dispute and a variant of his argumentation system with grounded instead of stable semantics (in the 'levelled' form of Definition 11). Furthermore, Kakas & Toni [1999] define dialectical versions of most of the assumption-based semantics proposed by Bondarenko *et al.* [1997], while Jakobovits [2000] does the same for several of her generalisations of the BDKT semantics. Finally, Vreeswijk & Prakken [2000] define a dialectical version of preferred semantics, both for sceptical and for credulous reasoning.

*Disputes with defeasible priorities*

Prakken & Sartor [1997b] extend their dialectical proof theory (see Definition 78) to the case with defeasible priorities. The main problem is on the basis of which priorities the defeating force of the moves should be determined. In fact, a few very simple conditions suffice. **CON** may completely ignore priorities: it suffices that its moves  $\emptyset$ -defeat **PRO**'s previous move. And for **PRO** only those priorities count that are stated by **PRO**'s move itself, i.e., moving with an argument  $A$  is allowed for **PRO** if  $A$  strictly  $A$ -defeats **CON**'s previous move; in addition, **PRO** has a new move available,

viz. moving a priority argument  $A$  such that **CON**'s last move does not  $A$ -defeat **PRO**'s previous move.

This results in the following change of conditions (3) and (4) of Definition 78.

- (3) If  $Player_i = \mathbf{PRO}$  ( $i > 1$ ), then
- $Arg_i$  strictly  $Arg_i$ -defeats  $Arg_{i-1}$ ; or
  - $Arg_{i-1}$  does not  $Arg_i$ -defeat  $A_{i-2}$ .
- (4) If  $Player_i = \mathbf{CON}$  then  $Arg_i \emptyset$ -defeats  $Arg_{i-1}$ .

Prakken & Sartor [1997b] show that their correctness and completeness results also hold for this definition (although in this case dialectical asymmetry is necessary). The main feature of their system that ensures this is the following property of the defeat relation: if  $A$   $S$ -defeats  $B$  and  $S' \subseteq S$ , then  $A$   $S'$ -defeats  $B$ .

Consider by way of illustration the dialectical version of Example 72.

$$\begin{array}{ll} \mathbf{PRO}_1 : [r_2 : \sim b \Rightarrow c] & \mathbf{CON}_1 : [r_0 : \Rightarrow a, r_1 : a \Rightarrow b] \\ \mathbf{PRO}_2 : [r_3 : \Rightarrow \neg a, r_4 : \Rightarrow r_0 \prec r_3] & \mathbf{CON}_2 : [r_5 : \Rightarrow r_3 \prec r_0] \\ \mathbf{PRO}_3 : [r_6 : \Rightarrow r_5 \prec r_4] & \end{array}$$

Here, **PRO**<sub>2</sub> uses the first available type of move, while **PRO**<sub>3</sub> uses the second type.

## 7 FINAL REMARKS

As we remarked in the introduction, the field of defeasible argumentation is still young, with a proliferation of systems and disagreement on many issues. Nevertheless, we have also observed many similarities and connections between the various systems, and we have seen that a formal meta-theory is emerging. In particular the BDKT approach has shown that a unifying account is possible; not only has it shown that many differences between argument-based systems are variations on just a few basic themes, but also has it shown how many nonmonotonic logics can be reformulated in argument-based terms. And Pollock's work on partial computation and adequacy criteria for defeasible reasoners paves the way for more meta-theoretical research. This also holds for the work of Lin & Shoham, Vreeswijk and Verheij on argumentation sequences, and for the just-discussed work on argument games and disputes.

In addition, several differences between the various systems appear to be mainly a matter of design, i.e., the systems are, to a large extent, translatable into each other. This holds, for instance, for the conceptions of arguments as sets (Simari & Loui), sequences (Prakken & Sartor) or trees (Lin & Shoham, Nute, Vreeswijk), and for the implicit (BDKT, Simari &

Loui, Prakken & Sartor), or explicit (Pollock, Nute, Vreeswijk) stepwise assessment of arguments. Moreover, other differences result from different levels of abstraction, notably with respect to the underlying logical language, the structure of arguments and the grounds for defeat. And some systems extend other systems: for example, Vreeswijk extends Lin & Shoham by adding the possibility to compare conflicting arguments, and Prakken & Sartor extend Simari & Loui with priorities from any source and with assumption attack, and they extend both Simari & Loui and Dung [1995] with reasoning about priorities. Finally, the declarative form of some systems and the procedural form of other systems are two sides of the same coin, as are the semantics and proof theory of standard logic.

The main substantial differences between the systems are probably the various notions of defeasible consequence described in Section 4, often reflecting a clash of intuitions in particular examples. Although the debate on the best definitions will probably continue for some time, in our opinion the BDKT approach has nevertheless shown that to a certain degree a unifying account is possible here also. Moreover, as already explained at the end of Section 4, some of the different consequence notions are not mutually exclusive but can be used in parallel, as capturing different senses in which belief in a proposition can be supported by a body of information. And each of these notions may be useful in a different context or for different purposes. Of course, in some cases this is otherwise. For instance, we would regard a definition as flawed if it does not capture indirect reinstatement (cf. p. 245). However, in general the existence of different definitions is not a problem for, but a feature of the field of defeasible argumentation. An important consequence of this is that the choice between the notions might depend on pragmatic considerations, as is, for instance, the case in legal procedure for the standards of proof. For example, the distinction in Anglo-Saxon jurisdictions between ‘beyond reasonable doubt’ in criminal cases and ‘on the balance of probabilities’ in civil cases is of a pragmatic nature; there are no intrinsic reasons to prefer one standard over the other as being ‘the’ standard of rational belief.

Another important difference is that while some systems formalise ‘logically ideal’ reasoners, other systems embody the idea of partial computation, i.e., of evaluating arguments not with respect to all possible arguments but only with respect to the arguments that have actually been constructed by the reasoner (Pollock, Loui, Vreeswijk, Verheij). However, here, too, we can say that these notions are not rivals, but capture different senses of support for beliefs, perhaps useful in different contexts.

We end with listing some of the main open problems in defeasible argumentation.

- Some examples do not receive a fully adequate treatment in any of the semantics that we have discussed. This holds, for instance, for



the ‘seemingly defeated’ arguments discussed in Section 5.2, and for Horty’s example discussed in Section 5.3. And perhaps other ‘critical’ examples can be discovered.

- Verheij’s work raises the question whether the conflict types that have been discussed in this chapter are all types of conflict that can exist between arguments.
- Another question raised by Verheij is what the best treatment is of accrual of arguments.
- Our informal remarks on the relation between the various systems should, where possible, be turned into a formal meta-theory of defeasible argumentation, making use of the work that has already been done.
- The procedural form of defeasible argumentation must be further developed; most current systems only have a semantic form.
- The notion of partial computation should be further studied. This notion is not only relevant for artificial intelligence but also for philosophy. The essence of defeasible reasoning is that it is reasoning under less than perfect conditions, where it is difficult or even impossible to obtain complete and reliable information. Since these conditions are very common in daily life, the correctness conditions for reasoning in such circumstances should be of interest to any logician who wants to study the formal structure of ordinary reasoning.
- Finally, it would be interesting to connect argumentation systems with research in so-called ‘formal dialectics’, which studies formal systems of procedural rules for dialogues; see e.g. Hamblin [1971], MacKenzie [1979] and Walton & Krabbe [1995]. Both fields would be enriched by such a connection. The argument games discussed in Section 6 are, unlike those of formal dialectics, not rules for real discussions between persons, but just serve as a proof theory for a (nonmonotonic) logic, i.e. they determine the (defeasible) consequences of a given set of premises. The ‘players’ of these argument games are not real actors but stand for the alternate search for arguments and counterarguments that is required by the proof theory. An embedding of argumentation systems in formal dialectics would yield an account of how their input theories are constructed dynamically during disputes between real discussants, instead of given in advance and fixed. On the other hand, argumentation systems could also enrich formal dialectics, which lacks notions of counterargument and defeat; its underlying logic is still deductive and its main dialectical speech act is

asking for premises that support a certain claim; ‘real’ counterarguments are impossible. Defeasible argumentation can provide formal dialectics with stronger dialectical features.

Some work of this nature has already been done, much of it in the area of artificial intelligence and law [Loui, 1998; Hage *et al.*, 1994; Gordon, 1995; Loui & Norman, 1995; Starmans, 1996; Prakken & Sartor, 1998; Lodder, 1999; Vreeswijk, 1999; Prakken, 2000]. Such work could provide a key in meeting Toulmin’s [1958] challenge to logicians to study how the properties of disputational procedures influence the validity of arguments. Perhaps in 1958 Toulmin’s challenge seemed odd, but 40 years of work in logic, philosophy, artificial intelligence and argumentation theory have brought an answer within reach.

### ACKNOWLEDGEMENTS

We thank all those with whom over the years we have had fruitful discussions on the topic of defeasible argumentation. Useful comments on an earlier version of this chapter were given by Jaap Hage, Simon Parsons, John Pollock and Bart Verheij.

Henry Prakken and Gerard Vreeswijk  
*Department of Information and Computing Sciences, Utrecht University,  
 The Netherlands.*

### BIBLIOGRAPHY

- [Amgoud & Cayrol, 1997] L. Amgoud & C. Cayrol, Integrating preference orderings into argument-based reasoning. *Proceedings of the International Conference on Qualitative and Quantitative Practical Reasoning (ECSQARU-FAPR’97)*. Lecture Notes in Artificial Intelligence 1244, 159–170. Berlin: Springer Verlag, 1997.
- [Asher & Morreau, 1990] N. Asher & M. Morreau, Commonsense entailment: a modal theory of nonmonotonic reasoning. *Proceedings of the Second European Workshop on Logics in Artificial Intelligence (JELIA’90)*. Lecture notes in Artificial Intelligence 478, 1–30. Berlin: Springer Verlag, 1990.
- [Baker & Ginsberg, 1989] A.B. Baker & M.L. Ginsberg, A theorem prover for prioritized circumscription. *Proceedings of the 11th International Joint Conference on Artificial Intelligence*, 463–467, 1989.
- [Benferhat *et al.*, 1993] S. Benferhat, D. Dubois & H. Prade, Argumentative inference in uncertain and inconsistent knowledge bases. *Proceedings of the 9th International Conference on Uncertainty in Artificial Intelligence*, 411–419. San Mateo, CA: Morgan Kaufman Publishers Inc, 1993.
- [Benferhat *et al.*, 1995] S. Benferhat, D. Dubois & H. Prade, How to infer from inconsistent beliefs without revising? *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1449–1455, 1995.
- [Bondarenko *et al.*, 1997] A. Bondarenko, P.M. Dung, R.A. Kowalski & F. Toni, An abstract argumentation-theoretic approach to default reasoning. *Artificial Intelligence* 93:63–101, 1997.

- [Brewka, 1989] G. Brewka, Preferred subtheories: an extended logical framework for default reasoning. *Proceedings of the Eleventh International Joint Conference on Artificial Intelligence*, 1043–1048, 1989.
- [Brewka, 1991] G. Brewka, *Nonmonotonic Reasoning: Logical Foundations of Commonsense*. Cambridge: Cambridge University Press, 1991.
- [Brewka, 1994a] G. Brewka, Reasoning about priorities in default logic. *Proceedings of the Twelfth National Conference on Artificial Intelligence*, 940–945, 1994.
- [Brewka, 1994b] G. Brewka, A logical reconstruction of Rescher's theory of formal disputation based on default logic. *Proceedings of the 11th European Conference on Artificial Intelligence*, 366–370, 1994.
- [Brewka, 1996] G. Brewka, Well-founded semantics for extended logic programs with dynamic preferences. *Journal of Artificial Intelligence Research* 4:19–30, 1996.
- [Cayrol, 1995] C. Cayrol, On the relation between argumentation and non-monotonic coherence-based entailment. *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1443–1448, 1995.
- [Chesñevar *et al.*, 1999] C.I. Chesñevar, A.G. Maguitman & R.P. Loui, Logical models of argument. *Submitted*.
- [Clark, 1990] P. Clark, Representing knowledge as arguments: Applying expert system technology to judgemental problem-solving. In *Research and Development in Expert Systems VII*, eds. T. R. Addis and R. M. Muir, 147–159. Cambridge University Press, 1990.
- [Das *et al.*, 1996] S. Das, J. Fox, & P. Krause, A unified framework for hypothetical and practical reasoning (1): theoretical foundations. *Proceedings of the International Conference on Formal and Applied Practical Reasoning (FAPR'96)*. Lecture Notes in Artificial Intelligence 1085, 58–72. Berlin: Springer Verlag, 1996.
- [De Kleer, 1986] J. De Kleer, An assumption-based TMS. *Artificial Intelligence* 28:127–162, 1986.
- [Delgrande, 1988] J. Delgrande, An approach to default reasoning based on a first-order conditional logic: revised report. *Artificial Intelligence* 36:63–90, 1988.
- [Doyle, 1979] J. Doyle, Truth Maintenance Systems. *Artificial Intelligence* 12:231–272, 1979.
- [Dung, 1993] P.M. Dung, An argumentation semantics for logic programming with explicit negation. *Proceedings of the Tenth Logic Programming Conference*, 616–630. Cambridge, MA: MIT Press, 1993.
- [Dung, 1994] P.M. Dung, Logic programming as dialogue games. Report Division of Computer Science, Asian Institute of Technology, Bangkok, 1994.
- [Dung, 1995] P.M. Dung, On the acceptability of arguments and its fundamental role in nonmonotonic reasoning, logic programming, and  $n$ -person games. *Artificial Intelligence* 77:321–357, 1995.
- [Dung *et al.*, 1996] P.M. Dung, R.A. Kowalski & F. Toni, Synthesis of proof procedures for default reasoning. *Proceedings International Workshop on Logic Program Synthesis and Transformation (LOPSTR'96)*, ed. J. Gallagher. Lecture Notes in Computer Science 1207, 313–324. Berlin: Springer Verlag, 1996.
- [Dung *et al.*, 1997] P.M. Dung, R.A. Kowalski & F. Toni, Argumentation-theoretic proof procedures for default reasoning. Report Department of Computing, Imperial College London, 1997.
- [Elvang-Gøransson & Hunter, 1995] M. Elvang-Gøransson & A. Hunter, Argumentative logics: reasoning with classically inconsistent information. *Data & Knowledge Engineering* 16:125–145, 1995.
- [Freeman & Farley, 1996] K. Freeman & A.M. Farley, A model of argumentation and its application to legal reasoning. *Artificial Intelligence and Law* 4:163–197, 1996. Reprinted in [Prakken & Sartor, 1997a].
- [Gabbay, 1985] D.M. Gabbay, Theoretical Foundations for Non-monotonic Reasoning in Expert Systems, in: *Logics and Models of Concurrent Systems*, ed. K.R. Apt, 439–457. Berlin, Springer-Verlag, 1985.
- [Gabbay *et al.*, 1994] D.M. Gabbay, C.J. Hogger & J.A. Robinson, *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3, Nonmonotonic Reasoning and Uncertain Reasoning*. Oxford: Oxford University Press, 1994.

- [Garcia *et al.*, 1998] A.J. Garcia, G.R. Simari & C.I. Chesñevar, An argumentative framework for reasoning with inconsistent and incomplete information. *Proceedings of the ECAI'98 Workshop on Practical Reasoning and Rationality*, Brighton, UK, 1998.
- [Geffner, 1991] H. Geffner, Beyond negation as failure. *Proceedings of the Third International Conference on Knowledge Representation and Reasoning*, 218–229. San Mateo, CA: Morgan Kaufmann Publishers Inc., 1991.
- [Geffner & Pearl, 1992] H. Geffner & J. Pearl, Conditional entailment: bridging two approaches to default reasoning. *Artificial Intelligence* 53:209–244, 1992.
- [Goodman, 1954] N. Goodman, *Fact, Fiction, and Forecast*. Cambridge, MA: Harvard University Press, 1954.
- [Gordon, 1995] T.F. Gordon, *The Pleadings Game. An Artificial Intelligence Model of Procedural Justice*. Dordrecht etc.: Kluwer Academic Publishers, 1995.
- [Gordon & Karacapilidis, 1997] T.F. Gordon & N. Karacapilidis, The Zeno argumentation framework. In *Proceedings of the Sixth International Conference on Artificial Intelligence and Law*, 10–18. New York: ACM Press, 1997.
- [Governatori & Maher, 2000] G. Governatori & M. Maher, An argumentation-theoretic characterization of defeasible logic. *Proceedings of the 14th European Conference on Artificial Intelligence*, 469–473, 2000.
- [Grosz, 1993] B.N. Grosz, Prioritizing multiple, contradictory sources in common-sense learning by being told; or, advice-taker meets bureaucracy. *Proceedings Common Sense '93: The Second Symposium on Logical formalisations of Common-Sense Reasoning*, Austin, Texas, 1993.
- [Hage, 1997] J.C. Hage, *Reasoning With Rules. An Essay on Legal Reasoning and Its Underlying Logic*. Dordrecht etc.: Kluwer Law and Philosophy Library, 1997.
- [Hage *et al.*, 1994] J.C. Hage, R. Leenes & A.R. Lodder, Hard cases: a procedural approach. *Artificial Intelligence and Law* 2:113–166, 1994.
- [Hamblin, 1971] C.L. Hamblin, Mathematical models of dialogue. *Theoria* 37:130–155, 1971.
- [Hanks & McDermott, 1987] S. Hanks & D. McDermott, Nonmonotonic Logic and Temporal Projection. *Artificial Intelligence* 33:379–412, 1987.
- [Hart, 1949] H.L.A. Hart, The ascription of responsibility and rights. *Proceedings of the Aristotelean Society*, n.s. 49 (1948–9), 171–194. Reprinted in *Logic and Language. First Series*, ed. A.G.N. Flew, 145–166. Oxford: Basil Blackwell, 1951.
- [Horty *et al.*, 1990] J.F. Horty, R.H. Thomasson & D.S. Touretzky, A skeptical theory of inheritance in nonmonotonic semantic networks. *Artificial Intelligence* 42:311–348, 1990.
- [Hunter, 1993] A. Hunter, Using priorities in non-monotonic proof theory. Report Department of Computing, Imperial College London, 1993.
- [Jakobovits, 2000] H. Jakobovits, *On the Theory of Argumentation Frameworks*. Doctoral dissertation, Department of Computer Science, Free University Brussels, 2000.
- [Jakobovits & Vermeir, 1999] H. Jakobovits, & D. Vermeir, Robust Semantics for Argumentation Frameworks. *Journal of Logic and Computation* 9:215–262, 1999.
- [Kakas *et al.*, 1994] A.C. Kakas, P. Mancarella & P.M. Dung, The acceptability semantics for logic programs. *Proceedings of the Eleventh International Conference on Logic Programming*, 509–514. Cambridge, MA: MIT Press, 1994.
- [Kakas & Toni, 1999] A.C. Kakas & F. Toni, Computing argumentation in logic programming. *Journal of Logic and Computation* 9:515–562, 1999.
- [Konolige, 1988] K. Konolige, Defeasible argumentation in reasoning about events. In *Methodologies for Intelligent Systems*, eds. Z.W. Ras and L. Saitta, 380–390. Amsterdam: Elsevier, 1988.
- [Kowalski & Toni, 1996] R.A. Kowalski & F. Toni, Abstract argumentation. *Artificial Intelligence and Law* 4:275–296, 1996. Reprinted in [Prakken & Sartor, 1997a].
- [Kraus *et al.*, 1990] S. Kraus, D. Lehmann & M. Magidor, Nonmonotonic reasoning, preferential models, and cumulative logics. *Artificial Intelligence* 44:167–207, 1990.
- [Krause *et al.*, 1995] P. Krause, S.J. Ambler, M. Elvang-Gøransson & J. Fox, A logic of argumentation for uncertain reasoning. *Computational Intelligence* 11:113–131, 1995.

- [Lewis, 1973] D.K. Lewis, *Counterfactuals*. Cambridge, MA: Harvard University Press, 1973.
- [Lin & Shoham, 1989] F. Lin & Y. Shoham, Argument systems. A uniform basis for nonmonotonic reasoning. *Proceedings of the First International Conference on Principles of Knowledge Representation and Reasoning*, 245–255. San Mateo, CA: Morgan Kaufmann Publishers Inc, 1989.
- [Lodder, 1999] A.R. Lodder, *DiaLaw. On Legal Justification and Dialog Games*. To appear in Kluwer's Law and Philosophy Library, 1999.
- [Loui, 1987] R.P. Loui, Defeat among arguments: a system of defeasible inference. *Computational Intelligence* 2:100–106, 1987.
- [Loui, 1995] R.P. Loui, Hart's critics on defeasible concepts and ascriptivism. *Proceedings of the Fifth International Conference on Artificial Intelligence and Law*, 21–30. New York: ACM Press, 1995.
- [Loui, 1998] R.P. Loui, Process and policy: resource-bounded non-demonstrative reasoning. *Computational Intelligence* 14:1–38, 1998.
- [Loui et al., 1993] R.P. Loui, J. Norman, J. Olson & A. Merrill, A design for reasoning with policies, precedents, and rationales. *Proceedings of the Fourth International Conference on Artificial Intelligence and Law*, 202–211. New York: ACM Press, 1993.
- [Loui & Norman, 1995] R.P. Loui & J. Norman, Rationales and argument moves. *Artificial Intelligence and Law* 3:159–189, 1995.
- [MacKenzie, 1979] J.D. MacKenzie, Question-begging in non-cumulative systems. *Journal of Philosophical Logic* 8:117–133, 1979.
- [Makinson, 1989] D. Makinson, General Theory of Cumulative Inference, *Proceedings of the 2nd Workshop on Nonmonotonic Reasoning*, eds. M. Reinfrank et al., Lecture Notes in Artificial Intelligence 346, 1–18. Berlin: Springer Verlag, 1989.
- [Makinson & Schlechta, 1991] D. Makinson & K. Schlechta, Floating conclusions and zombie paths: two deep difficulties in the 'directly sceptical' approach to inheritance nets. *Artificial Intelligence* 48:199–209, 1991.
- [Marek et al., 1990] W. Marek, A. Nerode & J. Remmel, A theory of non-monotonic rule systems I. *Annals of Mathematics and Artificial Intelligence* 1:241–273, 1990.
- [Marek et al., 1992] W. Marek, A. Nerode & J. Remmel, A theory of non-monotonic rule systems II. *Annals of Mathematics and Artificial Intelligence* 5:229–263, 1992.
- [Martins & Reinfrank, 1991] J.P. Martins & M. Reinfrank (eds.), *Truth Maintenance Systems*. Springer Lecture Notes in Artificial Intelligence, 515, Berlin: Springer Verlag, 1991.
- [McCarthy et al., 1969] J. McCarthy & P.J. Hayes, Some Philosophical Problems from the Standpoint of Artificial Intelligence, *Machine Intelligence* 4, eds. B. Meltzer et al., 463–502. Edinburgh University Press, 1969.
- [Nute, 1994] D.N. Nute, Defeasible logic. In *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 3, Nonmonotonic Reasoning and Uncertain Reasoning*, eds. D.M. Gabbay, C.J. Hogger & J.A. Robinson, 355–395. Oxford: Oxford University Press, 1994.
- [Nute, 1997] D.N. Nute, Apparent obligation. In *Defeasible Deontic Logic*, ed. D.N. Nute, 287–315. Dordrecht etc.: Kluwer Synthese Library, 1997.
- [Nute & Erk, 1995] D.N. Nute & K. Erk, Defeasible logic. Report AI Center, University of Georgia, Athens, GA, 1995.
- [Parsons et al., 1998] S. Parsons, C. Sierra & N.R. Jennings, Agents that reason and negotiate by arguing. *Journal of Logic and Computation* 8:261–292, 1998.
- [Pollock, 1970] J.L. Pollock, The Structure of Epistemic Justification. *American Philosophical Quarterly*, monograph series, vol. 4, 62–78, 1970.
- [Pollock, 1974] J.L. Pollock, *Knowledge and Justification*. Princeton: Princeton University Press, 1974.
- [Pollock, 1987] J.L. Pollock, Defeasible reasoning. *Cognitive Science* 11:481–518, 1987.
- [Pollock, 1991] J.L. Pollock, A theory of defeasible reasoning. *International Journal of Intelligent Systems* 6:33–54, 1991.
- [Pollock, 1992] J.L. Pollock, How to reason defeasibly. *Artificial Intelligence* 57:1–42, 1992.

- [Pollock, 1995] J.L. Pollock, *Cognitive Carpentry. A Blueprint for How to Build a Person*. Cambridge, MA: MIT Press, 1995.
- [Poole, 1985] D.L. Poole, On the comparison of theories: Preferring the most specific explanation. *Proceedings of the Ninth International Joint Conference on Artificial Intelligence*, 144–147, 1985.
- [Poole, 1988] D.L. Poole, A logical framework for default reasoning. *Artificial Intelligence* 36:27–47, 1988.
- [Prakken, 1993] H. Prakken, An argumentation framework in default logic. *Annals of Mathematics and Artificial Intelligence* 9:91–132, 1993.
- [Prakken, 1995] H. Prakken, A semantic view on reasoning about priorities (extended abstract). *Proceedings of the Second Dutch/German Workshop on Nonmonotonic Reasoning*, Utrecht, 152–159, 1995.
- [Prakken, 1997] H. Prakken, *Logical Tools for Modelling Legal Argument. A Study of Defeasible Reasoning in Law*. Dordrecht etc.: Kluwer Law and Philosophy Library, 1997.
- [Prakken, 1999] H. Prakken, Dialectical proof theory for defeasible argumentation with defeasible priorities (preliminary report). *Proceedings of the 4th ModelAge Workshop 'Formal Models of Agents'*, Springer Lecture Notes in Artificial Intelligence 1760, 202–215. Berlin: Springer Verlag, 1999.
- [Prakken, 2000] H. Prakken, On dialogue systems with speech acts, arguments, and counterarguments. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'2000)*, Springer Lecture Notes in AI 1919, 224–238. Berlin: Springer Verlag, 2000.
- [Prakken & Sartor, 1996] H. Prakken & G. Sartor, A dialectical model of assessing conflicting arguments in legal reasoning. *Artificial Intelligence and Law* 4:331–368, 1996. Reprinted in [Prakken & Sartor, 1997a].
- [Prakken & Sartor, 1997a] H. Prakken & G. Sartor, (eds.) 1997a. *Logical Models of Legal Argument*. Dordrecht etc.: Kluwer Academic Publishers, 1997. (reprint of *Artificial Intelligence and Law* 4, 1996).
- [Prakken & Sartor, 1997b] H. Prakken & G. Sartor, Argument-based extended logic programming with defeasible priorities. *Journal of Applied Non-classical Logics* 7:25–75, 1997.
- [Prakken & Sartor, 1998] H. Prakken & G. Sartor, Modelling reasoning with precedents in a formal dialogue game. *Artificial Intelligence and Law* 6:231–287, 1998.
- [Reiter, 1980] R. Reiter, A logic for default reasoning. *Artificial Intelligence* 13:81–132, 1980.
- [Rescher, 1976] N. Rescher, *Plausible Reasoning*. Assen: Van Gorcum, 1976.
- [Rescher, 1977] N. Rescher, *Dialectics: a Controversy-oriented Approach to the Theory of Knowledge*. Albany, N.Y.: State University of New York Press, 1977.
- [Ross, 1930] W.D. Ross, *The Right and the Good*. Oxford: Oxford University Press, 1930.
- [Sartor, 1994] G. Sartor, A formal model of legal argumentation. *Ratio Juris* 7:212–226, 1994.
- [Shoham, 1988] Y. Shoham, *Reasoning about Change. Time and Causation from the Standpoint of Artificial Intelligence*. Cambridge, MA: MIT Press, 1988.
- [Simari et al., 1994] G.R. Simari, C.I. Chesñevar & A.J. Garcia, The role of dialectics in defeasible argumentation. *Proceedings of the XIV International Conference of the Chilean Computer Science Society*, Concepción, Chile, 1994.
- [Simari & Loui, 1992] G.R. Simari & R.P. Loui, A mathematical treatment of defeasible argumentation and its implementation. *Artificial Intelligence* 53:125–157, 1992.
- [Starmans, 1996] R.J.C.M. Starmans, *Logic, Argument, and Commonsense*. Doctoral Dissertation, Tilburg University, 1996.
- [Thielscher, 1996] M. Thielscher, A nonmonotonic disputation-based semantics and proof procedure for logic programs. *Proceedings of the Joint International Conference and Symposium on Logic Programming*, 483–497. Cambridge, MA: MIT Press, 1996.
- [Toulmin, 1958] S.E. Toulmin, *The Uses of Argument*. Cambridge: Cambridge University Press, 1958.

- [Verheij, 1996] B. Verheij, *Rules, Reasons, Arguments. Formal Studies of Argumentation and Defeat*. Doctoral Dissertation, University of Maastricht, 1996.
- [Vreeswijk, 1989] G.A.W. Vreeswijk, The Feasibility of Defeat in Defeasible Reasoning, *Proceedings of the Second International Conference on Knowledge Representation and Reasoning*, 526–534. San Mateo, CA: Morgan Kaufmann Publishers Inc., 1991. Also published in *Diamonds and Defaults*, Studies in Language, Logic, and Information, Vol. 1, 359–380. Dordrecht: Kluwer, 1993.
- [Vreeswijk, 1993a] G.A.W. Vreeswijk, *Studies in Defeasible Argumentation*. Doctoral dissertation, Department of Computer Science, Free University Amsterdam, 1993.
- [Vreeswijk, 1993b] G.A.W. Vreeswijk, Defeasible dialectics: a controversy-oriented approach towards defeasible argumentation. *Journal of Logic and Computation* 3:317–334, 1993.
- [Vreeswijk, 1995] G.A.W. Vreeswijk, The computational value of debate in defeasible reasoning. *Argumentation* 9:305–342, 1995.
- [Vreeswijk, 1997] G.A.W. Vreeswijk, Abstract argumentation systems. *Artificial Intelligence* 90:225–279, 1997.
- [Vreeswijk, 1999] G.A.W. Vreeswijk, Representation of formal dispute with a standing order. To appear in *Artificial Intelligence and Law*, 1999.
- [Vreeswijk & Prakken, 2000] G.A.W. Vreeswijk & H. Prakken, Credulous and sceptical argument games for preferred semantics. In *Proceedings of the 7th European Workshop on Logic for Artificial Intelligence (JELIA'2000)*, Springer Lecture Notes in AI 1919, 239–253. Berlin: Springer Verlag, 2000.
- [Walton & Krabbe, 1995] D.N. Walton & E.C.W. Krabbe, *Commitment in Dialogue. Basic Concepts of Interpersonal Reasoning*. Albany, NY: State University of New York Press, 1995.

SVEN OVE HANSSON

## PREFERENCE LOGIC

### 1 INTRODUCTION

The study of general principles for preferences can, if we so wish, be traced back to Book III of Aristotle's *Topics*. Since the early twentieth century several philosophers have approached the subject of preferences with logical tools, but it is probably fair to say that the first complete systems of preference logic were those proposed by Sören Halldén in 1957 and Georg Henrik von Wright in 1963. [Rescher, 1968, pp. 287–288; Halldén, 1967; von Wright, 1963]. The subject also has important roots in utility theory and in the theory of games and decisions.

Preferences and their logical properties have a central role in rational choice theory, a subject that in its turn permeates modern economics, as well as other branches of formalized social science. Some of the most important recent developments in moral philosophy make essential use of preference logic [Fehige and Wessels, 1998]. At the same time, preference logic has turned out to be an indispensable tool in studies of belief revision and non-monotonic logic [Rott, 1999]. Preference logic has become so integrated into both philosophy and social science that we run the risk of taking it for granted and not noticing its influence.

This chapter is devoted to the philosophical foundations, rather than the applications, of preference logic. The emphasis is on fundamental results and their interpretation. Section 2 treats the basic case in which the objects of preferences form a set of mutually exclusive alternatives. In Section 3, such preferences are related to choice functions. In Section 4, the requirement of mutual exclusivity is relaxed. In Section 5, preferences are related to monadic concepts such as 'best', 'good', and 'ought'.

### 2 PREFERENCES OVER INCOMPATIBLE ALTERNATIVES

In most applications of preference logic, the objects that preferences refer to are assumed to be mutually exclusive. This assumption will also be made in the present section.

#### *2.1 Preference, indifference, and other value concepts*

From a logical point of view, the major value concepts of ordinary language can be divided into two major categories. The *monadic* (classificatory) value concepts, such as 'good', 'very bad', and 'worst' report how we evaluate a



single referent. The *dyadic* (comparative) value concepts, such as ‘better’, ‘worse’, and ‘equal in value to’, indicate a relation between two referents. In less colloquial contexts we can also find three-termed value predicates, such as ‘if  $x$ , then  $y$  is better than  $z$ ’ (conditional preferences) and even four-termed ones, such as ‘ $x$  is preferred to  $y$  more than  $z$  is preferred to  $w$ ’ [Packard, 1987]. This chapter is primarily devoted to the dyadic value concepts.

There are two fundamental comparative value concepts, namely ‘better’ (strict preference) and ‘equal in value to’ (indifference) [Halldén, 1957, p. 10]. The relations of preference and indifference between alternatives are usually denoted by the symbols  $>$  and  $\equiv$  or by the symbols  $P$  and  $I$ . Here, the former notation will be used.

There is a long-standing philosophical tradition to take  $A > B$  to represent ‘ $B$  is worse than  $A$ ’ as well as ‘ $A$  is better than  $B$ ’. [Brogan, 1919, p. 97]. This is not in exact accordance with ordinary English. We tend to use ‘better’ when focusing on the goodness of the higher-ranked of the two alternatives, and ‘worse’ when emphasizing the badness of the lower-ranked one [Halldén, p. 13; von Wright, 1963, p. 10; Chisholm and Sosa, 1966, p. 244]. However, the distinction between betterness and converse worseness can only be made at the price of a much more complex formal structure. The distinction does not seem to have enough philosophical significance to be worth this complexity, at least not in a general-purpose treatment of the subject.

When describing the preferences of others, we tend to use the word ‘preferred’. The word ‘better’ is used when we express our own preferences and also when we refer to purportedly impersonal evaluations. Although these are important distinctions, not very much has been made of them in preference logic. ‘Logic of preference’ and ‘logic of betterness’ are in practice taken as synonyms.

The preferences studied in preference logic are the preferences of rational individuals. Since none of us is fully rational, this means that we are dealing with an idealization. If a proposed principle for preference logic does not correspond to how we actually think and behave, the reason may be either that the principle is wrong or that we are not fully rational when our behaviour runs into conflicts with it.

The objects of preference are represented by the relata of the preference relation. ( $A$  and  $B$  in  $A > B$ .) In order to make the formal structure determinate enough, every preference relation is assumed to range over a specified set of relata. As already indicated, in this section, the relata are assumed to be mutually exclusive, i.e. none of them is compatible with, or included in, any of the others. No further assumptions are made about their internal structure. They may be physical objects, types or properties of such objects, states of affairs, possible worlds—just about anything.

Preferences over a set of mutually exclusive relata will be referred to as *exclusionary* preferences.

The following four properties of the two exclusionary comparative relations will be taken to be part of the meaning of the concepts of (strict) preference and of indifference:

- (1) If  $A$  is better than  $B$ , then  $B$  is not better than  $A$ .
- (2) If  $A$  is equal in value to  $B$ , then is  $B$  equal in value to  $A$ .
- (3)  $A$  is equal in value to  $A$ .
- (4) If  $A$  is better than  $B$ , then  $A$  is not equal in value to  $B$ .

It follows from (1) that preference is irreflexive, i.e. that  $A$  is not better than  $A$ . The following is a restatement of the four properties in formal language.

DEFINITION 1. A (*triplex*) *comparison structure* is a triple  $\langle \mathcal{A}, >, \equiv \rangle$ , in which  $\mathcal{A}$  is a set of alternatives, and  $>$  and  $\equiv$  are relations in  $\mathcal{A}$  such that for all  $A, B \in \mathcal{A}$ :

- (1)  $A > B \rightarrow \neg(B > A)$  (*asymmetry of preference*)
- (2)  $A \equiv B \rightarrow B \equiv A$  (*symmetry of indifference*)
- (3)  $A \equiv A$  (*reflexivity of indifference*)
- (4)  $A > B \rightarrow \neg(A \equiv B)$  (*incompatibility of preference and indifference*)

Furthermore:

$$A \geq B \leftrightarrow (A > B) \vee (A \equiv B) \text{ (weak preference)}$$

The intended reading of  $\geq$  is 'at least as good as' (or more precisely: 'better than or equal in value to'). As an alternative to  $\geq$ , it can also be denoted ' $R$ '. Weak preference can replace (strict) preference and indifference as primitive relations in comparison structures:

OBSERVATION 2. Let  $\langle \mathcal{A}, >, \equiv \rangle$  be a triplex comparison structure, and let  $\geq$  be the union of  $>$  and  $\equiv$ . Then:

- (1)  $A > B \leftrightarrow (A \geq B) \ \& \ \neg(B \geq A)$
- (2)  $A \equiv B \leftrightarrow (A \geq B) \ \& \ (B \geq A)$

**Proof.**

*Part 1: Left-to-right:* From  $A > B$  it follows by the definition of  $\geq$  that  $A \geq B$ . Furthermore, it follows from the asymmetry of preference that  $\neg(B > A)$  and from the incompatibility of preference and indifference that  $\neg(A \equiv B)$ , i.e., by the symmetry of indifference,  $\neg(B \equiv A)$ . Thus  $\neg((B > A) \vee (B \equiv A))$ , i.e., by the definition of  $\geq$ ,  $\neg(B \geq A)$ . *Right-to-left:* It follows from  $A \geq B$ , according to the definition of  $\geq$ , that either  $A > B$  or  $A \equiv B$ . By the same definition, it follows from  $\neg(B \geq A)$  that  $\neg(B \equiv A)$ . By the symmetry of indifference,  $\neg(A \equiv B)$ , so that  $A > B$  may be concluded.

*Part 2: Left-to-right:* It follows from  $A \equiv B$ , by the definition of  $\geq$ , that  $A \geq B$ . By the symmetry of indifference,  $A \equiv B$  yields  $B \equiv A$  so that, by the definition of  $\geq$ ,  $B \geq A$ . *Right-to-left:* It follows from the definition of  $\geq$  and  $(A \geq B) \& (B \geq A)$  that  $((A > B) \vee (A \equiv B)) \& ((B > A) \vee (B \equiv A))$ . By the symmetry of indifference,  $((A > B) \vee (A \equiv B)) \& ((B > A) \vee (A \equiv B))$ . By the asymmetry of preference,  $A > B$  is incompatible with  $B > A$ . We may conclude that  $A \equiv B$ . ■

The choice of primitives (either  $\geq$  or both  $>$  and  $\equiv$ ) is a fairly inconsequential choice between formal simplicity ( $\geq$ ) and conceptual clarity ( $>$  and  $\equiv$ ). (Cf. [Burros, 1976].) The following is an alternative to Definition 1.

**DEFINITION 3.** A (*duplex*) *comparison structure* is a pair  $\langle A, \geq \rangle$ , in which  $A$  is a set of alternatives and  $\geq$  a reflexive relation on  $A$ . The derived relations  $>$  and  $\equiv$  are defined as follows:

$$\begin{aligned} A > B & \text{ if and only if } A \geq B \text{ and } \neg(B \geq A) \\ A \equiv B & \text{ if and only if } A \geq B \text{ and } B \geq A \end{aligned}$$

It will be seen that the defined relation  $\geq$  of Definition 1 is reflexive and that the defined relations  $>$  and  $\equiv$  of Definition 3 satisfy conditions (1)–(4) of Definition 7. It follows that the two definitions are interchangeable. Given our definitions, the four conditions of Definition 1 are in combination equivalent to the reflexivity of weak preference.

The relations  $>$  and  $\equiv$  that are defined from  $\geq$  in the manner of Definition 3 are called the *strict part*, respectively the *symmetric part*, of  $\geq$ .

**NOTATIONAL CONVENTIONS N1:**

- (1) Chains of relations can be contracted. Hence,  $A \geq B \geq C$  abbreviates  $(A \geq B) \& (B \geq C)$ , and  $A > B > C \equiv D$  abbreviates  $(A > B) \& (B > C) \& (C \equiv D)$ .
- (2)  $>^*$  stands for  $>$  repeated any finite non-zero number of times (and similarly for the other relations). Thus  $A >^* C$  denotes that either  $A > C$  or there are  $B_1, \dots, B_n$  such that  $(A > B_1) \& (B_1 > B_2) \& \dots \& (B_{n-1} > B_n) \& (B_n > C)$ .

## 2.2 Completeness

In most applications of preference logic, it is taken for granted that the following property, called *completeness* or *connectedness*, should be satisfied:

$$(A \geq B) \vee (B \geq A), \text{ or equivalently:} \\ (A > B) \vee (A \equiv B) \vee (B > A)$$

As we will see later on, the assumption of completeness is often extremely helpful in terms of simplifying the formal structure. In terms of interpretation, however, it is much more problematic. In many everyday cases, we do not have, and do not need, complete preferences. In the choice between three brands of canned soup,  $A$ ,  $B$ , and  $C$ , I clearly prefer  $A$  to both  $B$  and  $C$ . As long as  $A$  is available I do not need to make up my mind whether I prefer  $B$  to  $C$ , prefer  $C$  to  $B$  or consider them to be of equal value. Similarly, a voter in a multi-party or multi-candidate election can do without ranking the parties or candidates that she does not vote for.

From the viewpoint of interpretation, we can distinguish between three major types of preference incompleteness. First, incompleteness may be *uniquely resolvable*, i.e. resolvable in exactly one way. The most natural reason for this to be the case is that incompleteness is due to lack of knowledge or reflection. Behind what we perceive as an incomplete preference relation there may be a complete preference relation that we can arrive at through observation, logical inference, or some other means of discovery.

Secondly, incompleteness may be *multiply resolvable*, i.e. possible to resolve in several different ways. In this case it is genuinely undetermined what will be the outcome of extending the relation to cover the previously uncovered cases.

Thirdly, incompleteness may be *irresolvable*. The most natural reason for this is that the alternatives differ in terms of advantages or disadvantages that we are unable to put on the same footing. I may be unable to say which I prefer—the death of two specified acquaintances or the death of a specified friend [Hansson, 1998a]. I may be unable to say which I prefer—the destruction of the pyramids in Giza or the extinction of the giant panda. I may also be unable in many cases to compare monetary costs to environmental damage.

It is established terminology to call two alternatives ‘incomparable’ whenever the preference relation is incomplete with respect to them. The term ‘incommensurable’ ‘can be reserved for cases when the incompleteness is irresolvable.

## 2.3 Transitivity and acyclicity

By far the most discussed logical property of preferences is the following:

$$A \geq B \geq C \rightarrow A \geq C \text{ (transitivity of weak preference)}$$

The corresponding properties of the other two relations are defined analogously:

$$\begin{aligned} A \equiv B \equiv C &\rightarrow A \equiv C \text{ (transitivity of indifference)} \\ A > B > C &\rightarrow A > C \text{ (transitivity of strict preference)} \end{aligned}$$

A weak preference relation  $\geq$  is called *quasi-transitive* if its strict part  $>$  is transitive.

'Mixed' transitivity properties can be also defined. The most important of these are:

$$\begin{aligned} A \equiv B > C &\rightarrow A > C \text{ (IP-transitivity)} \\ A > B \equiv C &\rightarrow A > C \text{ (PI-transitivity)} \end{aligned}$$

The relation  $\geq$  is *acyclic* if its strict part  $>$  satisfies the following property:

$$\text{There is no series } A_1, \dots, A_n \text{ of alternatives such that } A_1 > \dots > A_n > A_1.$$

These properties are logically related as follows:

**OBSERVATION 4.** Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\geq$  satisfies transitivity (of weak preference). Then it also satisfies:

1. Transitivity of indifference.
2. Transitivity of strict preference.
3. IP-transitivity.
4. PI-transitivity.

**Proof.**

*Part 1:* Let  $A \equiv B$  and  $B \equiv C$ . Then  $A \geq B$  and  $B \geq C$ , and  $\geq$ -transitivity yields  $A \geq C$ . Similarly,  $C \geq B$  and  $B \geq A$ , so that  $C \geq A$ . Hence  $A \equiv C$ .

*Parts 2 and 3:* Let  $A \geq B$  and  $B > C$ . Then  $A \geq B$  and  $B \geq C$ , and  $\geq$ -transitivity yields  $A \geq C$ . Suppose that  $A > C$  is not the case. It then follows from  $A \geq C$  that  $A \equiv C$ , hence  $C \geq A$ . From this and  $A \geq B$  we obtain that  $C \geq B$ , contrary to  $B > C$ . It follows from this contradiction that  $A > C$ .

*Part 4:* Let  $A > B$  and  $B \equiv C$ . Then  $A \geq B$  and  $B \geq C$ , and  $\geq$ -transitivity yields  $A \geq C$ . Suppose that  $A > C$  is not the case. It then follows from  $A \geq C$  that  $A \equiv C$ , hence  $C \geq A$ . From this and  $B \geq C$  we obtain that  $B \geq A$ , contrary to  $A > B$ . It follows from this contradiction that  $A > C$ . ■

**OBSERVATION 5.** Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\geq$  satisfies quasi-transitivity (transitivity of  $>$ ). Then it also satisfies acyclicity.

**Proof.** Let  $>$  be transitive and suppose that  $A_1 > \dots A_n > A_1$ . It follows by repeated use of  $>$ -transitivity that  $A_1 > A_1$ , contrary to the asymmetry of  $>$ . ■

## 2.4 Anti-cyclic properties

Acyclicity implies each member of the following series of properties that are specified with respect to the number of elements of the prohibited cycles:

- 1-acyclicity prohibits  $A_1 > A_1$
- 2-acyclicity prohibits  $A_1 > A_2 > A_1$
- 3-acyclicity prohibits  $A_1 > A_2 > A_3 > A_1$
- etc.

1-acyclicity is better known as irreflexivity and 2-acyclicity as asymmetry. Furthermore, just like the mixed transitivity properties referred to in the previous subsection, we can introduce mixed variants of acyclicity. The term *anti-cyclic properties* can be used for this more general category of properties. The definition is as follows:

**DEFINITION 6** (Hansson, 1993b). Let  $\pi_1, \dots, \pi_n$  be binary relations with the common domain  $\mathcal{A}$ . Then  $\pi_1, \dots, \pi_n$ -anticyclicity, denoted  $[\pi_1 \dots \pi_n]$ , is the property that there is no series  $A_1, \dots, A_n$  of elements of  $\mathcal{A}$  such that  $A_1 \pi_1 \dots A_n \pi_n A_1$ .

Hence, irreflexivity of a relation  $\pi$  can be written  $[\pi]$ , and asymmetry can be written  $[\pi\pi]$ . The following notation is convenient:

NOTATIONAL CONVENTIONS N2:

- (1)  $\Pi$  and  $\Psi$  denote series of relation symbols.
- (2)  $\pi^n$ , with  $\pi$  a relation symbol and  $n \geq 1$ , denotes the repetition of  $\pi$   $n$  times. Similarly,  $\Pi^n$  denotes the repetition of  $\Pi$   $n$  times.
- (3) In the notation of anti-cyclic properties,  $\geq$  is replaced by  $R$ ,  $>$  by  $P$ , and  $\equiv$  by  $I$ .

Hence,  $[P^3]$  denotes 3-acyclicity, and  $[P^2 I^2]$  denotes that there are no  $A_1, A_2, A_3, A_4$ , such that  $A_1 > A_2 > A_3 \equiv A_4 \equiv A_1$ .

Anticyclic properties are useful in preference logic. A major reason for their usefulness is that if the weak preference relation is complete, then the common transitivity properties are all equivalent to an anti-cyclic property. To begin with, consider transitivity of weak preference. When  $\geq$  is complete, then  $A \geq C$  is equivalent with  $\neg(C > A)$ , and we have the following equivalences:

For all  $A, B, C$ :  $(A \geq B) \& (B \geq C) \rightarrow A \geq C$   
 iff: For all  $A, B, C$ :  $(A \geq B) \& (B \geq C) \rightarrow \neg(C > A)$   
 iff: For all  $A, B, C$ :  $\neg((A \geq B) \& (B \geq C) \& (C > A))$   
 iff:  $[RRP]$

Hence, transitivity of a complete relation is equivalent to the anticyclic property  $[RRP]$ . The following more general translation rules can be used to replace transitivity-related properties of a complete preference relation by equivalent anticyclic properties.

OBSERVATION 7 (Hansson, 1993b). Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\geq$  satisfies completeness. Then:

- (T1)  $A \Pi C \rightarrow A \geq C$  is equivalent to  $[\Pi P]$   
 (T2)  $A \Pi C \rightarrow A > C$  is equivalent to  $[\Pi R]$   
 (T3)  $A \equiv^n C \rightarrow A \equiv C$  is equivalent to  $[I^n P]$   
 (T4)  $A \Pi C \rightarrow (A > B) \vee (B > C)$  is equivalent to  $[RR\Pi]$   
 (T5)  $(A \Pi B) \& (C \Psi D) \rightarrow (A > D) \vee (C > B)$  is equivalent to  $[R\Pi R\Psi]$

**Proof.** *T1* follows in the same way as the translation of  $\geq$ -transitivity that was given in the text. So does *T2*; note that  $A > C$  is equivalent with  $\neg(C \geq A)$ .

For *T3*, we have the following series of equivalent statements:

For all  $A_1, \dots, A_n$ :  $A_1 \equiv A_2 \equiv \dots \equiv A_n \rightarrow A_1 \equiv A_n$ .  
 For all  $A_1, \dots, A_n$ :  $A_1 \equiv A_2 \equiv \dots \equiv A_n \rightarrow \neg(A_1 > A_n) \& \neg(A_n > A_1)$ .  
 For all  $A_1, \dots, A_n$ :  $\neg(A_1 \equiv A_2 \equiv \dots \equiv A_n \& (A_1 > A_n))$  and  
 $\neg(A_1 \equiv A_2 \equiv \dots \equiv A_n \& (A_n > A_1))$ .  
 For all  $A_1, \dots, A_n$ :  $\neg(A_n \equiv A_{n-1} \equiv \dots \equiv A_1 > A_n)$  and  
 $\neg(A_1 \equiv A_2 \equiv \dots \equiv A_n > A_1)$   
 $[I^n P]$

For *T4*:

For all  $A, B, C$ :  $A \Pi C \rightarrow (A > B) \vee (B > C)$   
 For all  $A, B, C$ :  $A \Pi C \rightarrow \neg((B \geq A) \& (C \geq B))$   
 For all  $A, B, C$ :  $\neg((A \Pi C) \& (B \geq A) \& (C \geq B))$   
 For all  $A, B, C$ :  $\neg(C \geq B \geq A \Pi C)$   
 $[RR\Pi]$

For *T5*:

For all  $A, B, C, D$ :  $(A \Pi B) \& (C \Psi D) \rightarrow (A > D) \vee (C > B)$   
 For all  $A, B, C, D$ :  $(A \Pi B) \& (C \Psi D) \rightarrow \neg((D \geq A) \& (B \geq C))$   
 For all  $A, B, C, D$ :  $\neg((A \Pi B) \& (C \Psi D) \& (D \geq A) \& (B \geq C))$   
 For all  $A, B, C, D$ :  $\neg((D \geq A) \& (A \Pi B) \& (B \geq C) \& (C \Psi D))$   
 $[R\Pi R\Psi]$  ■

One important instance of T4 refers to the following property:

$$A > C \rightarrow (A > B) \vee (B > C) \text{ (virtual connectivity)}$$

It follows directly from T4 that virtual connectivity is equivalent to  $[RRP]$ , or in other words to transitivity of  $\geq$ . Another important instance of T4 is the translation of the following property:

$$A > B > C \rightarrow (A > D) \vee (D > C) \text{ (semi-transitivity)}$$

to  $[RRPP]$ . The following property:

$$(A > B) \& (C > D) \rightarrow (A > D) \vee (C > B) \text{ (interval order property)}$$

can be translated to  $[RPRP]$ , using T5. In summary, some of the major transitivity-related properties can be translated as follows:

$$[RRP] \quad A \geq B \geq C \rightarrow A \geq C \text{ (transitivity of weak preference)}$$

$$[IIP] \quad A \equiv B \equiv C \rightarrow A \equiv C \text{ (transitivity of indifference)}$$

$$[PPR] \quad A > B > C \rightarrow A > C \text{ (transitivity of strict preference)}$$

$$[IPR] \quad A \equiv B > C \rightarrow A > C \text{ (IP-transitivity)}$$

$$[PIR] \quad A > B \equiv C \rightarrow A > C \text{ (PI-transitivity)}$$

$$[RRPP] \quad A > B > C \rightarrow (A > D) \vee (D > C) \text{ (semi-transitivity)}$$

$$[RPRP] \quad (A > B) \& (C > D) \rightarrow (A > D) \vee (C > B) \text{ (interval order property)}$$

$$[P^n] \quad \text{n-acyclicity}$$

$$[P^*] \quad \text{acyclicity}$$

The major reason for undertaking these translations is that a series of simple derivation rules are available for proving the logical interrelations of anticyclic properties.

OBSERVATION 8 (Hansson, 1993b). Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\geq$  satisfies completeness. Then the following derivation rules hold for anticyclic properties of  $\geq$ .

$$(DR1) \quad [\Pi\Psi] \text{ iff } [\Psi\Pi].$$

$$(DR2) \quad [\Pi R] \text{ iff } [\Pi P] \& [\Pi I].$$

$$(DR3) \quad \text{If } [\Pi^n], \text{ then } [\Pi]. \text{ (} n \geq 1 \text{)}$$

$$(DR4) \quad \text{If } [\Pi I], \text{ then } [\Pi].$$



(DR5) If  $[\Pi R]$  &  $[\Psi P]$ , then  $[\Pi\Psi]$ .

**Proof.** The validity of DR1, DR2, and DR3 is obvious. For DR4, note that  $A\Pi A \equiv A$  follows from  $A\Pi A$ .

For DR5, suppose that  $[\Pi R]$  holds. We need to show that if  $[\Pi\Psi]$  is violated, then so is  $[\Psi P]$ . Suppose that  $(A\Pi B)$  &  $(B\Psi A)$ . From  $A\Pi B$  it follows by  $[\Pi R]$  that  $B \geq A$  does not hold, thus  $A > B$ . We therefore have  $(B\Psi A)$  &  $(A > B)$ , violating  $[\Psi P]$ . ■

Derivation rules DR1–DR5 have turned out to be sufficient to prove the major connections between the common transitivity-related properties. However, it remains an open issue how to construct a complete set of rules, i.e. a set of rules that is sufficient to prove all valid logical connections between anticyclic properties involving a reflexive relation  $\geq$  and its strict and symmetric parts.

The proofs of the standard logical connections between the transitivity-related properties of complete preference relations are quite simple:

OBSERVATION 9 (Sen 1969). Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\geq$  satisfies completeness, and let  $>$  and  $\equiv$  be the strict and symmetric parts of  $\geq$ . Then:

- (1) IP-transitivity and PI-transitivity are equivalent.
- (2) IP-transitivity implies  $\equiv$ -transitivity.
- (3)  $>$ -transitivity and  $\equiv$ -transitivity together imply PI-transitivity.
- (4)  $>$ -transitivity and PI-transitivity together imply  $\geq$ -transitivity.

**Proof.**

*Part 1:* We can use DR1 and DR2 to show that  $[PIR]$  iff  $[PII]$  &  $[PIP]$ , iff  $[IPI]$  &  $[IPP]$ , iff  $[IPR]$ .

*Part 2:* From  $[IPR]$  we obtain  $[IPI]$  by DR2 and  $[IIP]$  by DR1.

*Part 3:* Let  $[PPR]$  and  $[IIP]$ . It follows from  $[PPR]$  by DR2 that  $[PPI]$ , hence by DR1  $[PIP]$ . Applying DR1 to  $[IIP]$  we obtain  $[PII]$ , and applying DR2 to  $[PIP]$  and  $[PII]$  we obtain  $[PIR]$ .

*Part 4:* Let  $[PPR]$  and  $[PIR]$ . Applying DR1 to both of them we obtain  $[RPP]$  and  $[RPI]$ . From this we obtain  $[RPR]$  by DR2 and  $[RRP]$  by DR1. ■

The following results refer to longer cycles:

OBSERVATION 10 (Hansson, 1993b). Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\geq$  satisfies completeness. Then:

- (1) If  $[\Pi PR]$  then  $[\Pi^n PR]$  for all  $n \geq 1$ .

- (2) If  $[\Pi RP]$  then  $[\Pi^n RP]$  for all  $n \geq 1$ .
- (3) If  $[\Pi PR]$  and  $[\Psi R]$ , then  $[\Pi^n \Psi R]$  for all  $n \geq 1$ .
- (4) If  $[\Pi RP]$  and  $[\Psi P]$ , then  $[\Pi^n \Psi P]$  for all  $n \geq 1$ .
- (5) For all  $n \geq 2$ :  $[R^n P]$  iff  $[RRP]$ .
- (6) If  $\Pi$  contains at least one instance of  $P$ , then: If  $[RRP]$  then  $[\Pi]$
- (7) If  $[PPR]$  then  $[P^n R]$  for all  $n \geq 2$ .
- (8) If  $[I^n P]$  with  $n \geq 2$ , then  $[IIP]$ .
- (9) If  $[IPP]$  and  $[IIP]$ , then  $[I^n PP]$  and  $[I^{n+1} P]$  for all  $n \geq 1$ .
- (10) If  $[PPP]$  and  $[IPP]$ , then  $[IP^n]$  and  $[P^{n+1}]$  for all  $n \geq 2$ .
- (11) If  $[RPRP]$ , then  $[(RP)^n]$  for all  $n \geq 1$ .
- (12) If  $[RRPP]$ , then  $[R^m P^n]$  for all  $m, n$  such that  $m \leq n$  and  $n \geq 2$ .
- (13) If  $[RPRP]$  and  $[RRPP]$ , then  $[R^k P^l R^m P^n]$  for all  $k, l, m, n \geq 1$  such that  $k + m \leq l + n$ .

**Proof.**

*Part 1:* The proof is by induction. Let  $[\Pi PR]$  and  $[\Pi^k PR]$ . We are going to show that  $[\Pi^{k+1} PR]$ . Through DR1,  $[\Pi PR]$  yields  $[R\Pi P]$ . We can then apply DR5 to  $[\Pi^k PR]$  and  $[R\Pi P]$  and obtain  $[\Pi^k PR\Pi]$ , hence through DR1  $[\Pi^{k+1} PR]$ .

*Part 2.* This proof is similar to the previous one. Let  $[\Pi RP]$  and  $[\Pi^k RP]$ . We use DR1 to obtain  $[P\Pi R]$ , DR5 to obtain  $[P\Pi\Pi^k R]$ , and finally DR1 to obtain  $[\Pi^{k+1} RP]$ .

*Part 3:* Let  $[\Pi PR]$  and  $[\Psi R]$ . DR1 yields  $[R\Pi P]$ , and then DR5 can be used to obtain  $[\Psi R\Pi]$  and DR1 to obtain  $[\Pi\Psi R]$ . For induction, let  $[\Pi^k \Psi R]$ . We can apply DR5 to  $[\Pi^k \Psi R]$  and  $[R\Pi P]$  to obtain  $[\Pi^k \Psi R\Pi]$ , and then use DR1 to obtain  $[\Pi^{k+1} \Psi R]$ .

*Part 4:* Let  $[\Pi RP]$  and  $[\Psi P]$ . We can use DR1 to obtain  $[P\Pi R]$ , DR5 to obtain  $[P\Pi\Psi]$  and then DR1 to obtain  $[\Pi\Psi P]$ . For induction, let  $[\Pi^k \Psi P]$ . We can apply DR5 to  $[P\Pi R]$  and  $[\Pi^k \Psi P]$  to obtain  $[P\Pi\Pi^k \Psi]$  and then DR1 to obtain  $[\Pi^{k+1} \Psi P]$ .

*Part 5:* For one direction, let  $[R^n P]$  with  $n \geq 2$ . Then DR1 yields  $[R^{n-1} PR]$ , and DR2 and DR4 yield  $[R^{n-1} P]$ . By repetition,  $[R^2 P]$  will be obtained. For the other direction, let  $[R^2 P]$  and  $[R^k P]$  with  $k \geq 2$ . DR1 yields  $[RPR]$ , and DR5 can be applied to  $[RPR]$  and  $[R^k P]$  to obtain  $[RPR^k]$ . We can use DR1 to obtain  $[R^{k+1} P]$ .

*Part 6:* We are first going to show that if  $[RRP]$  then  $[\Pi RP]$  where  $\Pi$  is a possibly empty sequence. To see that this holds if  $\Pi$  is empty, use DR1 to obtain  $[RPR]$  and then DR2 and DR4 to obtain  $[RP]$ . Next suppose that  $[\Psi RP]$  holds for all sequences with  $n$  elements, and let  $\Psi'$  be a sequence with  $n+1$  elements. Then  $\Psi'$  has one of the three forms  $\Psi R$ ,  $\Psi I$ , and  $\Psi P$ , where  $\Psi$  has  $n$  elements.

We can apply DR1 to  $[\Psi RP]$  and obtain  $[P\Psi R]$ . Applying DR5 to this and  $[RRP]$  we obtain  $[P\Psi RR]$ . From DR1 follows  $[RP\Psi R]$  and then from DR2  $[RP\Psi I]$  and  $[RP\Psi P]$ . DR1 yields  $[\Psi RRP]$ ,  $[\Psi IRP]$ , and  $[\Psi PRP]$ , hence  $[\Psi' RP]$  in all three cases. Hence, if  $[RRP]$  then  $[\Pi RP]$ .

It follows by DR1 from  $[\Pi RP]$  that  $[P\Pi R]$ . DR2 and DR4 can be used to obtain  $[P\Pi]$ . Since every sequence that contains at least one instance of  $P$  is equivalent in the sense of DR1 to a sequence of the form  $[P\Pi]$ , this concludes the proof.

*Part 7:* The proof will be by induction. Let  $[PPR]$  and  $[P^k R]$  with  $k \geq 2$ . We can use DR1 to obtain  $[PRP]$  and then DR5 to obtain  $[P^k PR]$ , i.e.  $[P^{k+1} R]$ .

*Part 8:* Let  $[I^n P]$ . We can use DR1 to obtain  $[IPI^{n-2}]$  and then DR4  $n-2$  times to obtain  $[IIP]$ .

*Part 9:* Let  $[IPP]$  and  $[IIP]$ . We are first going to show by induction that  $[I^n PP]$ . Let  $[I^n PP]$  with  $n \geq 1$ . We can use DR1 to obtain  $[PIP]$  and  $[PII]$  and then DR2 to obtain  $[PIR]$ . From this and  $[I^n PP]$  we obtain  $[PII^n P]$  with DR5 and  $[I^{n+1} PP]$  with DR1.

Next, let  $[I^{n+1} P]$  with  $n \geq 1$ . We can use DR5 and combine this with  $[PIR]$  to obtain  $[PII^{n+1}]$ , and DR1 yields  $[I^{n+2} P]$ .

*Part 10:* Let  $[PPP]$  and  $[IPP]$ . We are first going to show by induction that  $[IP^n]$  for all  $n \geq 2$ . Let  $[IP^n]$  and  $n \geq 2$ . DR1 yields  $[PPI]$  and DR2  $[PPR]$ , that can be combined with  $[IP^n]$  to obtain, through DR5,  $[PPIP^{n-1}]$ , and then through DR1,  $[IP^{n+1}]$ .

Next, we are going to show that  $[P^{n+1}]$  for all  $n \geq 2$ . Let  $[P^{n+1}]$  and  $n \geq 2$ . Equivalently,  $[P^n P]$ . Since we also have  $[PPR]$ , DR5 yields  $[PPP^n]$ , or equivalently  $[P^{n+2}]$ .

*Part 11:* The proof proceeds by induction. It follows from DR3 that  $[RP]$ . Let  $[(RP)^n]$  with  $n \geq 2$ . Equivalently,  $[(RP)^{n-1} RP]$ . From  $[RPRP]$  it follows, via DR1, that  $[PRPR]$ . We can use DR5 and combine this with  $[(RP)^{n-1} RP]$  to obtain  $[PRP(RP)^{n-1} R]$ , and DR1 yields  $[RPRP(RP)^{n-1}]$  or equivalently  $[(RP)^{n+1}]$ .

*Part 12:* Let  $[RRPP]$ . We are first going to show by induction that  $[R^n P^n]$  for all  $n \geq 2$ . Let  $[R^n P^n]$  and  $n \geq 2$ . From  $[RRPP]$  follows by DR1 that  $[PPRR]$ . DR5 yields  $[PPRR^n P^{n-1}]$ , and then DR1 yields  $[R^{n+1} P^{n+1}]$ .

Next, we are going to show that if  $[R^n P^n]$  and  $m \leq n$ , then  $[R^m P^n]$ . Let  $[R^n P^n]$ . It follows from DR1 that  $[P^n R^n]$ , from repeated use of DR2 and DR4 that  $[P^n R^m]$  and then from DR1 that  $[R^m P^n]$ .

*Part 13:* Let  $[RPRP]$  and  $[RRPP]$ . We are first going to show that  $[R^k P^l R^m P^n]$  for all  $k, l, m, n \geq 1$  such that  $k + m = l + n$ . The proof will be by induction. We have  $[RPRP]$ , and for the induction step we need to show that if  $[R^k P^l R^m P^n]$ , then (A)  $[R^{k+1} P^l R^m P^{n+1}]$  and (B)  $[R^k P^l R^{m+1} P^{n+1}]$ . (There are two additional cases, but they can be excluded due to DR1.) For (A), use DR1 to obtain  $[PPRR]$ . We can use DR5 to combine  $[PPRR]$  and  $[R^k P^l R^m P^n]$ , and obtain  $[PPRR^k P^l R^m P^{n-1}]$ . DR1 yields  $[R^{k+1} P^l R^m P^{n+1}]$ . For (B), apply DR1 to  $[R^k P^l R^m P^n]$  to obtain  $[P^n R^k P^l R^m]$ . Then use DR5 to combine this with  $[RRPP]$  and obtain  $[P^n R^k P^l R^{m-1} RRP]$ . DR1 yields  $[R^k P^l R^{m+1} P^{n+1}]$ .

To complete the proof it is sufficient to show that if  $[R^k P^l R^m P^n]$ , then  $[R^{k-1} P^l R^m P^n]$  and  $[R^k P^l R^{m-1} P^n]$ . Due to DR1 it is sufficient to prove one of these. Let  $[R^k P^l R^m P^n]$ . DR1 yields  $[P^l R^m P^n R^k]$ , DR2 and DR4 yield  $[P^l R^m P^n R^{k-1}]$  and then DR1 yields  $[R^{k-1} P^l R^m P^n]$ . ■

## 2.5 Preference cycles exemplified

Part 6 of Observation 10 is particularly interesting since it shows that  $[RRP]$ , transitivity of weak preference, implies all anticyclic properties that can hold if  $\geq$  is reflexive and has a non-empty domain. (Let  $\pi_1, \dots, \pi_n$  be a series, each element of which is either  $\geq$  or  $\equiv$ . Then  $A\pi_1 A \dots A\pi_n A\pi_1$  holds for any  $A \in \mathcal{A}$  due to the reflexivity of  $\geq$  and consequently of  $\equiv$ . It follows that  $[\pi_1 \dots \pi_n]$  does not hold.)

Should transitivity of weak preference hold, or at least some of the weaker transitivity-related properties mentioned above? This is probably the most debated issue in preference logic. Since completeness has mostly been assumed to hold, this controversy can also be expressed in terms of anticyclic properties: What types of cycles are acceptable? As we have just seen, the controversial cycles are those that contain at least one instance of  $P$ . Therefore, this is more precisely a matter of which if any  $P$ -containing cycles should be allowed. All  $P$ -containing cycles contradict transitivity of weak preference. In addition, IIP-cycles contradict transitivity of indifference, PPR-cycles contradict transitivity of strict preference, etc. In what follows, preferences with a  $P$ -containing cycle will be called *cyclic preferences*.

Quite a few examples of preference cycles have been proposed in the literature for various philosophical purposes. Most of these examples belong to the following four categories:  $I^n P$ -cycles with  $n \geq 2$ , IPIP-cycles, IPP-cycles, and  $P^n$ -cycles with  $n \geq 3$ .

$I^n P$ -cycles, with  $n \geq 2$ , are often used as arguments against transitivity. The classic construction employs a series of objects that are so arranged that we cannot distinguish between two adjacent members of the series, whereas we can distinguish between members at greater distance [Armstrong, 1939; Armstrong, 1948; Luce, 1954]. Let us think of 1000 cups of coffee, numbered  $C_0, C_1, C_2, \dots$  up to  $C_{999}$ . Cup  $C_0$  contains no sugar, cup  $C_1$  one grain of

sugar, cup  $C_2$  two grains etc. Since I cannot taste the difference between  $C_{999}$  and  $C_{998}$ , they are equally good (or rather equally terrible) in my taste,  $C_{999} \equiv C_{998}$ . For the same reason, we have  $C_{998} \equiv C_{997}$ , etc. all the way up to  $C_1 \equiv C_0$ , but clearly  $C_0 > C_{999}$ , so that we have an  $I^{999}P$ -cycle.

With suitably adjusted thresholds of discrimination, it is also possible to construct shorter cycles of the same kind, including IIP-cycles. Michael Dummett [1984, p. 34] proposed that a subject may be incapable of distinguishing between wine  $A$  and wine  $B$  or between wine  $B$  and wine  $C$ , but able to distinguish between  $A$  and  $C$ , and likes  $A$  better. A somewhat different construction of IIP-cycles has been ascribed to W. Armstrong [Lehrer and Wagner, 1985]: A boy may be indifferent between receiving a bicycle or a pony, and also between receiving a bicycle with a bell and a pony, whereas he prefers receiving a bicycle with a bell to receiving just a bicycle. The reason is of course that the bell is too small an advantage to be significant in the uncertain choice between a bicycle and a horse. A similar example was proposed by Frank Restle [1961, pp. 62–63]: replace the pony by a trip to Florida, the bicycle by a trip to California and the bell by a very small amount of money.

An IPIP-cycle was constructed by Sven Danielsson [1998] through the addition of one more alternative to Restle's example:

- $X$  A trip to California plus an apple
- $Y$  A trip to California
- $Z$  A trip to Florida
- $U$  A trip to Florida plus an apple

We can then have  $Z \equiv X > Y \equiv U > Z$ , i.e. an IPIP-cycle.

Next, let us turn to IPP-cycles. A simple way to construct an IPP-cycle is to combine two IIP-cycles in different dimensions. This was done by Ng [1977], whose (unusually undramatic example) refers to three samples of paper,  $x$ ,  $y$ , and  $z$ . It can be observed that  $x$  is thicker than  $y$ , but no difference in thickness can be observed between  $x$  and  $z$  (which is intermediate in thickness) or between  $z$  and  $y$ . Similarly,  $y$  is perceptibly whiter than  $z$ , but there is no noticeable difference in whiteness between  $y$  and  $x$  or between  $x$  and  $z$ . Assuming that I prefer thick white paper, 'I prefer  $x$  to  $y$  as I can observe that  $x$  is thicker than  $y$  but cannot observe any difference in whiteness. Similarly, I prefer  $y$  to  $z$ . But I am indifferent to the choice between  $x$  and  $z$ ' [Ng, 1977, p. 52].

Our next category is  $P^n$ -cycles with  $n > 3$ . To construct them, we can make use of an  $IP^n$ -cycle with  $n > 2$ . The construction method is exemplified by the 'lawn-crossing example' that has been much discussed in the literature on utilitarianism [Harrison, 1953, p. 107; Österberg, 1989]. Let  $c_0, \dots, c_{1000}$  denote the number of times that you cross a particular lawn. A single crossing makes no (perceptible) difference in the condition of the lawn, but it results in a perceptible time gain. Therefore,  $c_{1000} > c_{999} >$

...  $> c_2 > c_1 > c_0$ . However, a large number of crossings will cause a complete damage of the lawn that is not outweighed by the total time gain. Therefore,  $c_0 > c_{1000}$ , and we have a  $P^{1001}$ -cycle (based on the  $I^{1000}P$ -cycle for the condition of the lawn).

A famous example by Warren S. Quinn [Quinn 1990] has the same structure. He assumed that a medical device has been implanted in the body of a person (the self-torturer). The device has 1001 settings, from 0 (off) to 1000. Each increase leads to a negligible increase in pain. Each week, the self-torturer 'has only two options—to stay put or to advance the dial one setting. But he may advance only one step each week, and he may never retreat. *At each advance he gets \$ 10,000.*' In this way he may 'eventually reach settings that will be so painful that he would then gladly relinquish his fortune and return to 0' [Quinn, 1990, p. 79].

Our final category of cycles is PPP-cycles. They differ from IPP-cycles in being direct arguments against acyclicity. One way to construct a PPP-cycle is to combine three IIP-cycles in the same way that two such cycles were used to obtain an IPP-cycle. This was done by George Schumm [1987], who invites us to consider a Mr. Smith who chooses between three boxes of Christmas tree ornaments. Each box contains one red, one blue, and one green ball. The balls of box 1 are denoted  $R_1$ ,  $B_1$ , and  $G_1$ , etc. in the obvious way. 'Suppose that any difference in color between  $R_1$  and  $R_3$  falls below Smith's threshold of discrimination, as does that between  $R_3$  and  $R_2$ . But he can see a difference between  $R_1$  and  $R_2$ , and he prefers the former. Likewise, suppose that while Smith sees no difference between  $B_3$  and  $B_2$ , or between  $B_2$  and  $B_1$ , he prefers the hue of  $B_3$  to that of  $B_1$ . Finally, although being unable to detect any difference between  $G_2$  and  $G_1$ , or between  $G_1$  and  $G_3$ , he prefers  $G_2$  to  $G_3$ ... Smith prefers Box 1 to Box 2 since, to his eye, they contain equally attractive blue balls and green balls, while Box 1 contains the prettier red ball. Analogously, he prefers Box 2 to Box 3, and Box 3 to Box 1.'

Schumm argued that 'given any proposed counterexample to the transitivity of indifference... one can always construct, on the foregoing model, an equally compelling counterexample to the transitivity of strict preference. Thus, those who would have us shun the transitivity of indifference should have the courage of their convictions to forsake both principles.'

R. G. Hughes [1980] constructed a PPP-cycle in a quite different but not less plausible way. In his example, a voter assesses three political candidates  $A$ ,  $B$ , and  $C$ , as follows: In terms of political views,  $A$  is better than the other two, and  $B$  is better than  $C$ . In terms of honesty,  $C$  is better than the other two, and  $B$  is better than  $A$ . A difference in corruptibility is important to this voter only when it exceeds a critical level, but when it does so, this issue becomes more important than all other considerations. The difference between  $A$  and  $B$  and that between  $B$  and  $C$  are below this critical level, but that between  $A$  and  $C$  is above it. (The voter therefore

acts as if she were indifferent between  $A$  and  $B$ , and also between  $B$  and  $C$ , but prefers  $C$  to  $A$ , in terms of honesty.) Thus, the voter, both aspects considered, prefers  $A$  to  $B$ ,  $B$  to  $C$ , and  $C$  to  $A$ . (An example with the same structure, that Hughes does not seem to have been aware of, can be found in [Tversky, 1969, p. 321].)

## 2.6 *Why cycles are problematic*

At least some of the examples cited in the foregoing subsection can be used to show that actual human beings may have cyclic preferences. It does not necessarily follow, however, that the same applies to the idealized *rational* agents of preference logic. Perhaps such patterns are due to irrationality or to factors, such as lack of knowledge or discrimination, that prevent us from being rational. There is a strong tradition, not least in economic applications, to regard full  $\geq$ -transitivity as a necessary prerequisite of rationality.

Some authors have argued for transitivity through direct appeal to intuition. According to Savage, whenever I find a PPP-cycle among my own preferences, 'I feel uncomfortable in much the same way that I would do when it is brought to my attention that some of my beliefs are logically contradictory. Whenever I examine such a triple of preferences on my own part, I find that it is not at all difficult to reverse one of them. In fact, I find on contemplating the three alleged preferences side by side that at least one of them is not a preference at all, at any rate not any more' [Savage, 1954, p. 21]. There is also some empirical evidence that when people are faced with their own intransitivities, they tend to modify their preferences to make them transitive [Tversky, 1969].

Two other, somewhat more substantial types of argument have been put forward in favour of transitivity: The money-pump argument and the choice-guidance argument.

The money-pump argument originates with F. P. Ramsey [1931, p. 182]. Ramsey pointed out that if a subject's relation of preference violates transitivity, then '[h]e could have a book made against him by a cunning better and would then stand to lose in any event'. The non-probabilistic version of this argument, the 'money-pump', runs as follows:

'Suppose an individual prefers  $y$  to  $x$ ,  $z$  to  $y$ , and  $x$  to  $z$ . It is reasonable to assume that he is willing to pay a sum of money to replace  $x$  by  $y$ . Similarly, he should be willing to pay some amount of money to replace  $y$  by  $z$  and still a third amount to replace  $z$  by  $x$ . Thus, he ends up with the alternative he started with but with less money.'

[Tversky, 1969, p. 45]

In order to see more in detail how the argument works, consider the following example [Hansson, 1993a]. A certain stamp-collector has cyclic preferences with respect to three stamps, denoted  $a$ ,  $b$ , and  $c$ . She prefers  $a$  to  $b$ ,  $b$  to  $c$ ,

and  $c$  to  $a$ . Following Ramsey, we may assume that there is an amount of money, say 10 cents, that she is prepared to pay for exchanging  $b$  for  $a$ ,  $c$  for  $b$ , or  $a$  for  $c$ . She comes into a stamp shop with stamp  $a$ . The stamp-dealer offers her to trade in  $a$  for  $c$ , if she pays 10 cents. She accepts the deal.

For a precise notation, let  $\langle x, v \rangle$  denote that the collector owns stamp  $x$  and has paid  $v$  cents to the dealer. She has now moved from the state  $\langle a, 0 \rangle$  to the state  $\langle c, 10 \rangle$ .

Next, the stamp-dealer takes out stamp  $b$  from a drawer, and offers her to swap  $c$  for  $b$ , against another payment of 10 cents. She accepts, thus moving from the state  $\langle c, 10 \rangle$  to  $\langle b, 20 \rangle$ .

When she is just on her way out of the shop, the dealer calls her back, and advises her that it only costs 10 cents to change back to  $a$ , the very stamp that she had in her pocket when she entered the shop. Since she prefers it to  $b$ , she pulls out a third dime, thus moving from  $\langle b, 20 \rangle$  to  $\langle a, 30 \rangle$ . Since her original state was  $\langle a, 0 \rangle$ , this does not seem to be much of an achievement.

To summarize the argument, the following sequence of preferences caused the trouble:

$$\begin{aligned} \langle c, 10 \rangle &> \langle a, 0 \rangle \\ \langle b, 20 \rangle &> \langle c, 10 \rangle \\ \langle a, 30 \rangle &> \langle b, 20 \rangle \end{aligned}$$

The trouble does not end here. Presumably, the sequence continues:

$$\begin{aligned} \langle c, 40 \rangle &> \langle a, 30 \rangle \\ \langle b, 50 \rangle &> \langle c, 40 \rangle \\ \langle a, 60 \rangle &> \langle b, 50 \rangle \end{aligned}$$

...

If the poor customer stays long enough in the stamp shop, she will be bereft of all her money, to no avail.

The money-pump argument relies on the following two assumptions: (1) The primary alternatives (the stamps) can be combined with some other commodity (money) to form composite alternatives. (2) For every preferred change of primary alternatives, there is some non-zero loss of the auxiliary commodity (money) that is worth that change. The money-pump can be used to extract money from a subject with cyclic preferences only if these two conditions are satisfied.

The money-pump presented above requires a  $P^n$ -cycle. There is also another type of money-pump that can operate on any type of cyclic preferences. Let's go back to the stamp shop.

A new customer enters the shop. She is indifferent between stamps  $a$  and  $b$ , and also between stamps  $b$  and  $c$ , but prefers  $c$  to  $a$ . Contrary to the first customer, she only has an IIP-cycle (intransitive indifference). Strangely enough, just like the first customer she enters the shop carrying stamp  $a$ .



Can the stamp-dealer extract money from this customer as well? It turns out that he can, but he must apply a modified strategy. The first move is identical. He offers her to exchange stamp  $a$  for stamp  $c$  against a modest fee of 10 cents (or whatever sum is small enough to make her accept the deal). In this way, he makes her move from  $\langle a, 0 \rangle$  to  $\langle c, 10 \rangle$ . Next, he offers to pay her 1 cent if she is willing to take stamp  $b$  instead of stamp  $c$ . Since, presumably, the customer is absolutely indifferent between  $b$  and  $c$ , she is—or so we may expect—willing to accept this bid, thus moving from  $\langle c, 10 \rangle$  to  $\langle b, 9 \rangle$ . After that he offers her another cent for changing to  $a$ , thus bringing her to  $\langle a, 8 \rangle$ . Just like the previous customer, she has given away money to no avail. The vicious sequence of preferences was:

$$\begin{aligned}\langle c, 10 \rangle &> \langle a, 0 \rangle \\ \langle b, 9 \rangle &> \langle c, 10 \rangle \\ \langle a, 8 \rangle &> \langle b, 9 \rangle.\end{aligned}$$

Presumably, the sequence continues:

$$\begin{aligned}\langle c, 18 \rangle &> \langle a, 8 \rangle \\ \langle b, 17 \rangle &> \langle c, 18 \rangle \\ \langle a, 16 \rangle &> \langle b, 17 \rangle \\ \dots\end{aligned}$$

In this way, the second customer, just like the first, will be ruined unless her dealings with the cunning stamp-dealer are interrupted. In order for this type of money-pump to operate we only need an R\*P-cycle. Therefore, this combination can be used as a fully general argument against all types of cyclic preferences.

But how convincing are the money-pumps? It should be noted that they rely on a particular way to combine preferences in two dimensions. A critic can argue that the construction of preferences for the combined alternative set  $(\{\langle a, 0 \rangle, \langle c, 10 \rangle, \langle b, 20 \rangle, \langle a, 30 \rangle, \langle c, 40 \rangle, \langle b, 50 \rangle \dots\})$  in our first example) out of preferences over the primary alternative set  $(\{a, b, c\})$  should not be performed in the straightforward simple way that was indicated in the examples. When forming their preferences over the new alternative sets created by the cunning dealer, the collectors must consider the totality of the situation, and therefore—according to the critic—they must construct these preferences in a way that avoids the absurd result. Most of us would prefer  $\langle b, 20 \rangle$  to  $\langle a, 30 \rangle$  in the first example, even if we would have preferred  $\langle a, 10 \rangle$  to  $\langle b, 0 \rangle$ . Arguably, the example only works if the agent (the stamp-collector) can be brought to make each decision in isolation, without taking into account the total situation. What the example shows, it can be argued, is only that a rational subject's preference-guided behaviour can be 'manipulated' by persons, institutions or impersonal conditions that control her agenda (decision horizon). This may be seen as an extension of the well-known result

from social decision theory that '[a] clever agenda setter, with knowledge of all voters' preferences could design an agenda to reach virtually any point in the alternative space' [McKelvey, 1979, p. 1087]; cf. [McKelvey, 1976; McKelvey and Wendell, 1976; Plott, 1967]. Even if the logical structure of our preferences is rational, agenda-setting mechanisms may very well drive us to irrational behaviour.

The second major argument in favour of transitivity is the choice-guidance argument. It is based on the assumption that the logical properties of preferences should be compatible with their use as guides to choice or action. It is easy to find examples of how cycles make preferences unsuitable as guides for choices. Our first stamp-collector, who prefers stamp  $a$  to stamp  $b$ , stamp  $b$  to stamp  $c$ , and stamp  $c$  to stamp  $a$ , cannot use these preferences as a guide to choose one of these stamps.

However, the choice-guidance argument cannot be used directly against all forms of cyclic preferences. Suppose that there is also a fourth stamp  $d$ , that she prefers to all the other three. Then her preference relation can be used without problem to guide a choice among the set  $\{a, b, c, d\}$ , in spite of the cycle. Cycles among defeated elements do not prevent rational choice.

For a preference relation to be choice-guiding, it must supply at least one alternative that is eligible, i.e. can reasonably be chosen. The minimal formal criterion for eligibility is that the chosen alternative is no worse than any other alternative:

*Weak eligibility*

There is at least one alternative  $A$  such that for all  $B$ ,  $\neg(B > A)$ .

Let  $A$  be a weakly eligible alternative, and let  $B$  be an alternative that is *not* (weakly) eligible. Furthermore, suppose that  $A$  and  $B$  are comparable, i.e. that either  $A > B$ ,  $A \equiv B$ , or  $B > A$ . Then, by the definition of weak eligibility,  $B > A$  does not hold. It would also be very strange for  $A$  and  $B$  to be equal in value, i.e., for  $A \equiv B$  to hold. If preferences are choice-guiding, then two alternatives should not be considered to be of equal value if one of them is eligible and the other is not. We may therefore conclude, as a consequence of the principle of choice-guidance, that if  $A$  but not  $B$  is weakly eligible, then  $A$  and  $B$  are not equal in value. In an equivalent formulation:

*Top-transitivity of weak eligibility*

If  $A \equiv B$ , and  $\neg(C > A)$  for all  $C$ , then  $\neg(C > B)$  for all  $C$ .

If the preference relation  $\geq$  is complete, then weak eligibility is equivalent with the following condition:

*Strong eligibility*

There is at least one alternative  $A$  such that for all  $B$ ,  $A \geq B$ .

Top-transitivity can be rewritten as follows:

*Top-transitivity of strong eligibility:*

If  $A \equiv B$ , and  $A \geq C$  for all  $C$ , then  $B \geq C$  for all  $C$ .

The sets of alternatives that our preferences refer to are not immutable. To the contrary, new alternatives can become available, and old ones can be lost. If no alternative is considered to be exempt from possibly being lost in the future, then it may be a cost-minimizing strategy to pursue one's deliberations until (weak or strong) eligibility holds for all non-empty subsets of the original alternative set. A rationality criterion will be said to hold *restrictably* for a set of alternatives if and only if it holds for all its non-empty subsets. It must be emphasized that restrictability does not always hold for rational preferences. The preference relation best suited for guiding choices among a certain set of alternatives need not be a suitable guide for choosing among a particular subset of that set. (For a counterexample, see Subsection 3.1.)

As will be seen from the following theorem, if the eligibility properties are required to hold restrictably, then we obtain rationality criteria of the more well-known types, such as completeness, acyclicity, and various types of transitivity.

**THEOREM 11** (Hansson, 1997a). *Let  $\geq$  be a relation over some finite set  $A$  with at least two elements.*

1. *It satisfies restrictable weak eligibility if and only if it satisfies  $[P^*]$  (acyclicity).*
2. *It satisfies restrictable strong eligibility if and only if it satisfies completeness and  $[P^*]$  (acyclicity).*
3. *It satisfies restrictable top-transitive weak eligibility if and only if it satisfies  $[P^*]$  (acyclicity) and  $[PIR]$  (PI-transitivity).*
4. *It satisfies restrictable top-transitive strong eligibility if and only if it satisfies completeness and  $[RRP]$  (transitivity).*

**Proof.**

*Part 1:* For one direction, suppose that acyclicity does not hold. Then there are  $A_1, \dots, A_n \in \mathcal{A}$  such that  $A_1 > A_2 > \dots > A_{n-1} > A_n$  and  $A_n > A_1$ . Weak eligibility is not satisfied for the subset  $\{A_1, \dots, A_n\}$  of  $\mathcal{A}$ .

For the other direction, suppose for *reductio* that acyclicity but not restrictable weak eligibility is satisfied. We are going to show that  $\mathcal{A}$  is infinite, contrary to the assumptions. Since restrictable weak eligibility is violated, there is some subset  $\mathcal{B}$  of  $\mathcal{A}$  for which weak eligibility does not hold. Let  $A_1 \in \mathcal{B}$ . Since weak eligibility is not satisfied, there is some  $A_2 \in \mathcal{B}$  such

that  $A_2 > A_1$ . Similarly, there is some  $A_3$  such that  $A_3 > A_2$ , etc... If any two elements on the list  $A_1, A_2, A_3, \dots$  are identical, then acyclicity is violated. Thus,  $\mathcal{B}$  is infinite, and consequently so is  $\mathcal{A}$ , contrary to the conditions.

*Part 2:* For one direction, suppose that  $\geq$  satisfies restrictable strong eligibility. To see that it satisfies completeness let  $A, B \in \mathcal{A}$ . Since strong eligibility holds restrictably for  $\mathcal{A}$ , strong eligibility holds for the subset  $\{A, B\}$  of  $\mathcal{A}$ , so that either  $A \geq B$  or  $B \geq A$ . Since restrictable strong eligibility implies restrictable weak eligibility, acyclicity follows from part 1.

For the other direction, suppose for *reductio* that  $\geq$  is complete and acyclic but violates restrictable strong eligibility. There must be some subset  $\mathcal{B}$  of  $\mathcal{A}$  for which strong eligibility does not hold. Let  $A_1 \in \mathcal{B}$ . There is then some  $A_2 \in \mathcal{B}$  such that  $\neg(A_1 \geq A_2)$ . By completeness,  $A_2 > A_1$ . Similarly, there is some  $A_3 \in \mathcal{B}$  such that  $\neg(A_2 \geq A_3)$  and consequently  $A_3 > A_2$ , etc. Suppose that any two elements of the list  $A_1, A_2, A_3, \dots$  are identical. Then acyclicity is violated. Thus  $\mathcal{B}$  is infinite, and so is  $\mathcal{A}$ , contrary to the conditions.

*Part 3:* First suppose that  $\geq$  satisfies restrictable top-transitive weak eligibility. Acyclicity follows from part 1. For PI-transitivity, let  $A, B$ , and  $C$  be three elements of  $\mathcal{A}$  such that  $A > B$  and  $B \equiv C$ . Suppose that  $\neg(A > C)$ . Then  $\neg(X > C)$  for all  $X \in \{A, B, C\}$ , and by top-transitive weak eligibility for that set  $B \equiv C$  yields  $\neg(X > B)$  for all  $X \in \{A, B, C\}$ , contrary to  $A > B$ . We may conclude that  $A > C$ .

For the other direction, suppose that acyclicity and PI-transitivity are satisfied. It follows from part 1 that  $\geq$  satisfies restrictable weak eligibility. For top-transitivity, let  $\mathcal{B}$  be a subset of  $\mathcal{A}$  and  $A$  and  $B$  two elements of  $\mathcal{B}$  such that  $A \equiv B$  and that for all  $X \in \mathcal{B}$ ,  $\neg(X > B)$ . For *reductio*, suppose that for some  $C \in \mathcal{B}$ ,  $C > A$ . Then it follows from  $A \equiv B$  and PI-transitivity that  $C > B$ , contrary to the conditions. We may conclude that  $\neg(X > A)$  holds for all  $X \in \mathcal{B}$ , so that top-transitivity of weak eligibility holds for  $\geq$  in  $\mathcal{B}$ . Since this applies to all subsets  $\mathcal{B}$  of  $\mathcal{A}$ , top-transitivity of weak eligibility holds restrictably in  $\mathcal{A}$ .

*Part 4:* First suppose that restrictable top-transitive strong eligibility is satisfied. Completeness follows from part 2. For transitivity, let  $A \geq B$  and  $B \geq C$ . Since top-transitivity of strong eligibility holds restrictably for  $\mathcal{A}$ , top-transitive strong eligibility holds for  $\{A, B, C\}$ . There are three cases:

*Case i,  $A \equiv B$ :* By completeness,  $B \geq B$ , so that  $B \geq X$  for all  $X \in \{A, B, C\}$ . It follows from top-transitivity of strong eligibility, as applied to  $\{A, B, C\}$ , that  $A \geq X$  for all  $X \in \{A, B, C\}$ , so that  $A \geq C$ .

*Case ii,  $A > B > C$ :* Suppose that  $C > A$ . Then strong eligibility does not hold for  $\{A, B, C\}$ , contrary to the conditions. It follows that  $\neg(C > A)$  and by completeness that  $A \geq C$ .

*Case iii,  $A > B \equiv C$ :* Suppose that  $C \geq A$ . By completeness  $C \geq C$ , so that  $C \geq X$  for all  $X \in \{A, B, C\}$ . By top-transitivity and  $B \equiv C$ ,

$B \geq X$  for all  $X \in \{A, B, C\}$ , so that  $B \geq A$ , contrary to  $A > B$ . By this contradiction,  $\neg(C \geq A)$ . By completeness,  $A \geq C$ .

For the other direction, suppose that completeness and transitivity are satisfied. Transitivity implies acyclicity, so that restrictable strong eligibility follows from part 2. For top-transitivity, let  $\mathcal{B}$  be a subset of  $\mathcal{A}$  with  $A, B \in \mathcal{B}$  and such that  $A \equiv B$  and that  $A \geq C$  for all  $C \in \mathcal{B}$ . Then for all  $C, B \geq A$  and  $A \geq C$  yield  $B \geq C$ , so that top-transitivity of strong eligibility holds in  $\mathcal{B}$ . Since this applies to all subsets  $\mathcal{B}$  of  $\mathcal{A}$ , top-transitivity of strong eligibility holds restrictably in  $\mathcal{A}$ . ■

In summary, the two major anticyclic (and protransitive) arguments, money-pumps and choice-guidance, both depend on manipulations of the alternative set. Money-pumps require the construction of composite alternative sets and the choice-guidance argument depends on the restriction of alternative sets. Since neither of these manipulations is uncontroversial, we do not have an uncontroversial argument in favour of preference transitivity.

## 2.7 Numerical representation

Preferences can be interpreted as expressions of value.  $A > B$  then means that more value is assigned to  $A$  than to  $B$ , and  $A \equiv B$  that the same value is assigned to the two. Values, we may assume, can be adequately expressed in numerical terms. Let  $u$  (as in *utility*) be a value function, that assigns a real number to each element of the alternative set. We can then construct a model of preference logic in the following way: ( $\mathfrak{R}$  is the set of real numbers.)

### *Exact value representation*

$A > B$  iff  $u(A) > u(B)$ , where  $u$  is a function from  $\mathcal{A}$  to  $\mathfrak{R}$ .

Since completeness is assumed to hold,  $A \geq B$  is defined to hold if and only if  $\neg(B > A)$ . This construction has been characterized in terms of postulates as follows:

**THEOREM 12** (Roberts, 1979). *Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\mathcal{A}$  is countable. Then the following two conditions are equivalent:*

1.  $\geq$  is satisfies completeness and transitivity ( $[RRP]$ ).
2. There is a function  $u$  from  $\mathcal{A}$  to  $\mathfrak{R}$  such that  $A > B$  iff  $u(A) > u(B)$ .

**Proof.** See [Roberts, 1979, pp. 109–110]. ■

As can be seen from  $I^nP$ -cycles such as the ‘cups of coffee’ example mentioned in Subsection 2.5, the exact value representation of preferences is for some purposes too demanding. If  $u(A) > u(B)$ , but  $u(A) - u(B)$  is so small

that it cannot be discerned, then we should not expect  $A > B$  to hold. One interesting way to represent this feature is to introduce a fixed limit of indiscernibility, such that  $A > B$  holds if and only if  $u(A) - u(B)$  is larger than that limit. Such a limit is commonly called a *just noticeable difference* (JND).

*JND representation*

$A > B$  iff  $u(A) - u(B) > \delta$ , where  $\delta$  is a positive real number.

**THEOREM 13** (Scott and Suppes, 1958). *Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\mathcal{A}$  is finite. Then the following two conditions are equivalent:*

1.  $\geq$  satisfies completeness and the two properties

$A > B > C \rightarrow (A > D) \vee (D > C)$  ([RRPP], semi-transitivity)

$(A > B) \ \& \ (C > D) \rightarrow (A > D) \vee (C > B)$  ([RPRP], interval order property)

2. There is a function  $u$  from  $\mathcal{A}$  to  $\Re$  and a positive real number  $\delta$  such that  $A > B$  iff  $u(A) - u(B) > \delta$ .

**Proof.** See [Scott and Suppes, 1958; Suppes and Zinnes, 1963] or [Roberts, 1979, p. 260–264]. ■

A relation  $\geq$  that satisfies condition 1 of this theorem is called a *semiorder*. Semiorders were introduced in [Luce, 1954]. The present axioms and the above representation theorem were given in [Scott and Suppes, 1958]. The theorem cannot in general be extended to infinite alternative sets; for the infinite case see [Manders, 1981].

Semiorders can be generalized by relaxing the condition that the threshold of discrimination be the same for all comparisons of alternatives:

*Variable threshold representation*

$A > B$  iff  $u(A) - u(B) > \sigma(A)$ , where  $\sigma(A) > 0$  for all  $A$ .

Another interesting construction is to assign to each alternative an interval instead of a single number. We then need two functions from  $\mathcal{A}$  to  $\Re$ ,  $u_{\max}$  and  $u_{\min}$ , such that for all  $A \in \mathcal{A}$ ,  $u_{\max}(A) \geq u_{\min}(A)$ . Here,  $u_{\max}$  represents, of course, the upper limit of the interval assigned to  $A$ , and  $u_{\min}$  its lower limit.  $A > B$  holds if and only if all elements of the interval assigned to  $A$  have higher value than all elements of the  $B$  interval:

*Interval representation:*

$A > B$  iff  $u_{\min}(A) > u_{\max}(B)$

It is easy to see that the variable threshold representation and the interval representation are equivalent. Just let:

$$u(A) = u_{\max}(A) \text{ and } \sigma(A) = u_{\max}(A) - u_{\min}(A).$$

The following representation theorem has been obtained for these constructions:

**THEOREM 14** (Fishburn, 1970a). *Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure. Then the following two conditions are equivalent:*

1.  $\geq$  satisfies completeness and:  
 $(A > B) \ \& \ (C > D) \rightarrow (A > D) \vee (C > B)$  ( $\lceil RPRP \rceil$ , interval order property)
2. There is a function  $u$  from  $\mathcal{A}$  to  $\mathfrak{R}$  and a function  $\sigma$  from  $\mathcal{A}$  to the positive part of  $\mathfrak{R}$  such that for all  $A$  and  $B$  in  $\mathcal{A}$ ,  $A > B$  iff  $u(A) - u(B) > \sigma(A)$ .
3. There are two functions  $u_{\max}$  and  $u_{\min}$  from  $\mathcal{A}$  to  $\mathfrak{R}$  such that for all  $A$  and  $B$  in  $\mathcal{A}$ ,  $u_{\max}(A) \geq u_{\min}(A)$  and  $A > B$  iff  $u_{\min}(A) > u_{\max}(B)$ .

**Proof.** See [Fishburn, 1970a]. ■

A relation  $\geq$  is called an *interval order* if it satisfies the conditions of Theorem 14. Interval orders were introduced by Fishburn as a generalization of semiorders [Fishburn, 1970a]. One further step of generalization can be taken: We can let the threshold of discrimination depend on both relata.

*Doubly variable threshold representation*

$A > B$  iff  $u(A) - u(B) > \sigma(A, B)$ , with  $\sigma(A, B) > 0$  for all  $A$  and  $B$ .

**THEOREM 15** (Abbas, 1995). *Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure such that  $\mathcal{A}$  is finite. Then the following two conditions are equivalent:*

1.  $\geq$  satisfies acyclicity ( $\lceil P^* \rceil$ )
2. There is a function  $u$  from  $\mathcal{A}$  to  $\mathfrak{R}$  and a function  $\sigma$  from  $\mathcal{A} \times \mathcal{A}$  to the positive part of  $\mathfrak{R}$  such that  $A > B$  iff  $u(A) - u(B) > \sigma(A, B)$ .

**Proof.** See [Abbas, 1995]. ■

Relations satisfying acyclicity, the condition referred to in Theorem 15, are often called *suborders* [Fishburn, 1970b].

### 3 PREFERENCE AND CHOICE

There is a strong tradition, particularly in economics, to equate preference with choice. Preference is considered to be hypothetical choice, and choice to be revealed preference. Given an alternative set  $\mathcal{A}$ , we can represent (hypothetical) choice as a function  $C$  that, for any given subset  $\mathcal{B}$  of  $\mathcal{A}$ , turns out the chosen ('preferred') elements of  $\mathcal{B}$ .

Subsection 3.1 introduces some rationality criteria for choice functions. In Subsection 3.2 these are related to properties of the preference relation.

#### 3.1 Postulates for choice functions

The formal definition of choice functions is as follows:

DEFINITION 16.  $C$  is a choice function for  $\mathcal{A}$  if and only if it is a function from and to  $\wp(\mathcal{A})$ , such that for all  $\mathcal{B} \in \wp(\mathcal{A})$ :

- (1)  $C(\mathcal{B}) \subseteq \mathcal{B}$ , and
- (2) if  $\mathcal{B} \neq \emptyset$ , then  $C(\mathcal{B}) \neq \emptyset$ .

A large number of rationality properties for preferences have been proposed. Only three of the most important of these will be discussed here.

*Chernoff* (property  $\alpha$ ) [Chernoff 1954]  
If  $\mathcal{B}_1 \subseteq \mathcal{B}_2$  then  $\mathcal{B}_1 \cap C(\mathcal{B}_2) \subseteq C(\mathcal{B}_1)$ .

Amartya Sen has called Chernoff 'a very basic requirement of rational choice' [Sen, 1969, p. 384]. It 'states that if the world champion in some game is Pakistani, then he must also be the champion in Pakistan'. However, it is far from self-evident that this property should hold on all occasions. Two types of examples showing this are well-known from the literature. First, the alternative set may carry information, as in Amartya Sen's example: '[G]iven the choice between having tea at a distant acquaintance's home ( $x$ ), and not going there ( $y$ ), a person who chooses to have tea ( $x$ ) may nevertheless choose to go away ( $y$ ), if offered—by that acquaintance—a choice over having tea ( $x$ ), going away ( $y$ ), and having some cocaine ( $z$ )' [Sen, 1993, p. 502]. See also [Kirchsteiger and Puppe, 1996]. Secondly, choice may be positional. In a choice between a big apple, a small apple, and an orange, you may choose the big apple, but in a choice between only the two apples you may nevertheless opt for the smaller one [Anand, 1993, p. 344], cf. [Gärdenfors, 1973].

Property  $\beta$   
If  $\mathcal{B}_1 \subseteq \mathcal{B}_2$  and  $X, Y \in C(\mathcal{B}_1)$ , then  $X \in C(\mathcal{B}_2)$  iff  $Y \in C(\mathcal{B}_2)$



According to Sen, property  $\beta$  is ‘also appealing, though... perhaps somewhat less intuitive than Property  $\alpha$ ’. It ‘states that if some Pakistani is a world champion, then *all* champions of Pakistan must be champions of the world’ [Sen, 1969, p. 384]. Property  $\beta$  is not either unproblematic, as can be seen from a modification of Sen’s cocaine example. I may be indifferent between staying for tea and going away ( $C(\{x, y\}) = \{x, y\}$ ), but prefer to leave if cocaine is offered ( $C(\{x, y, z\}) = \{y\}$ ).

$$\begin{aligned} & \text{Expansion (property } \gamma) \\ & C(\mathcal{B}_1) \cap \dots \cap C(\mathcal{B}_n) \subseteq C(\mathcal{B}_1 \cup \dots \cup \mathcal{B}_n) \end{aligned}$$

To see that expansion does not always hold, let  $\mathcal{B}_1 = \{\text{small apple, big apple}\}$  and  $\mathcal{B}_2 = \{\text{small apple, orange}\}$ . It may very well be that  $C(\mathcal{B}_1) = C(\mathcal{B}_2) = \{\text{small apple}\}$  whereas  $C(\mathcal{B}_1 \cup \mathcal{B}_2) = \{\text{big apple}\}$ .

### 3.2 Connecting choice and preference

The most obvious way to construct a choice function out of a preference relation  $\geq$  is to have the function always choose the elements that are best according to  $\geq$ :

$$\begin{aligned} & \text{The best choice connection} \\ & C(\mathcal{B}) = \{X \in \mathcal{B} \mid (\forall Y \in \mathcal{B})(X \geq Y)\} \end{aligned}$$

A choice function is *relational* if it is based on some preference relation  $\geq$  in this way. It can be seen from the definition that  $\geq$  must then be complete. (If it is incomplete, then there are elements  $X$  and  $Y$  such that neither  $X \geq Y$  nor  $Y \geq X$ . It follows that  $C(\{X, Y\}) = \emptyset$ , contrary to Definition 16.) It can also be seen that the connection does not work if  $\geq$  violates acyclicity ( $[P^*]$ ). Let  $X_1 > X_2 > \dots > X_n > X_1$ . Then it holds for each  $X_k$  that there is some  $X_m$  such that  $X_m > X_k$ , so that  $X_k \notin C(\{X_1, X_2, \dots, X_n\})$ . Hence  $C(\{X_1, X_2, \dots, X_n\}) = \emptyset$ , again contrary to Definition 16. Indeed, these two conditions can also be shown to be sufficient for the workability of the best choice connection. The following theorems show how various properties of choice functions correspond to properties of underlying preference relations.

**THEOREM 17.** *Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure, and let  $C$  be the function constructed from  $\geq$  according to the best choice connection. Then:*

1.  *$C$  is a choice function if and only if  $\geq$  satisfies completeness and acyclicity.*

*Furthermore, if  $C$  is a relational choice function, then:*

2.  *$C$  satisfies Chernoff [Sen 1969 p. 384]*

3.  $C$  satisfies property  $\beta$  if and only if  $\geq$  satisfies  $PI$ -transitivity ( $[PIR]$ ).  
[Sen 1969 p. 384]
4.  $C$  satisfies property  $\beta$  if and only if  $\geq$  satisfies  $\geq$ -transitivity ( $[RRP]$ ).  
[Sen 1969 p. 385]

**Proof.**

*Part 1:* This is essentially a restatement of Theorem 11, part 2.

*Part 2:* Let  $\mathcal{B}_1 \subseteq \mathcal{B}_2$  and  $X \in \mathcal{B}_1 \cap C(\mathcal{B}_2)$ . Then it holds for all  $Y \in \mathcal{B}_2$  that  $X \geq Y$ , hence this holds for all  $Y$  in  $\mathcal{B}_1$ , hence  $X \in C(\mathcal{B}_1)$ .

*Part 3:* For one direction, let  $[PIR]$  be violated. Then there are  $X, Y$ , and  $Z$  such that  $X > Y \equiv Z \geq X$ . It follows that  $C(\{Y, Z\}) = \{Y, Z\}$ ,  $Z \in C(\{X, Y, Z\})$  and  $Y \notin C(\{X, Y, Z\})$ , contrary to property  $\beta$ .

For the other direction, let property  $\beta$  be violated. Then there are alternatives  $X$  and  $Y$  and sets  $\mathcal{B}_1$  and  $\mathcal{B}_2$  such that  $\mathcal{B}_1 \subseteq \mathcal{B}_2$ ,  $X, Y \in C(\mathcal{B}_1)$ ,  $X \in C(\mathcal{B}_2)$  and  $Y \notin C(\mathcal{B}_2)$ . It follows from  $X, Y \in C(\mathcal{B}_1)$  that  $Y \equiv X$ , from  $Y \notin C(\mathcal{B}_2)$  that there is some  $Z \in \mathcal{B}_2$  such that  $Z > Y$ , and from  $X \in C(\mathcal{B}_2)$  that  $X \geq Z$ . Hence,  $Z > Y \equiv X \geq Z$ , contrary to  $[PIR]$ .

*Part 4:* Due to Part 1 of the present theorem, it is sufficient to show that if  $R$  is complete and satisfies  $[P^*]$ , then  $[PIR]$  and  $[RRP]$  are equivalent. It follows from part 6 of Observation 10 that  $[RRP]$  implies  $[PIR]$ . For the other direction, apply DR2 to  $[PIR]$  to obtain  $[PIP]$ , and then DR1 to obtain  $[PPI]$ .  $[P^*]$  yields  $[PPP]$  that we can combine with  $[PPI]$ , using DR2, to obtain  $[PPR]$ . Applying DR1 to  $[PPR]$  and  $[PIR]$  we obtain  $[RPP]$  and  $[RPI]$ , and DR2 yields  $[RPR]$ . Finally, through DR1 we obtain  $[RRP]$ . ■

**THEOREM 18** (Sen, 1971). *Let  $C$  be a choice function for  $\mathcal{A}$ . Then the following two conditions are equivalent:*

- (1)  $C$  satisfies Chernoff and Expansion.
- (2) There is a relation  $\geq$  on  $\mathcal{A}$  such that  $C$  coincides with the function constructed from  $\geq$  via the best choice connection.

**Proof.**

*1-to-2:* Let  $\geq$  be the relation such that  $X \geq Y$  iff  $X \in C(\{X, Y\})$ . For one direction, let  $X \in C(\mathcal{B})$  and  $\mathcal{B} \subseteq \mathcal{A}$ . We need to show that  $X \geq Y$  for all  $Y \in \mathcal{B}$ . Suppose to the contrary that for some  $Y$ ,  $\neg(X \geq Y)$ . Then  $X \notin C(\{X, Y\})$ , and according to Chernoff,  $C(\mathcal{B}) \cap \{X, Y\} \subseteq C(\{X, Y\}) = \{Y\}$ . Contradiction.

For the other direction, let  $X \in C(\{X, Y\})$  for all  $Y \in \mathcal{B}$ . It follows from Expansion that  $X \in C(\mathcal{B})$ .

*2-to-1.* To prove that Chernoff holds, let  $\mathcal{B}_1 \subseteq \mathcal{B}_2$  and  $X \in \mathcal{B}_1 \cap C(\mathcal{B}_2)$ . Then it holds for all  $Y \in \mathcal{B}_2$  that  $X \geq Y$ , hence this holds for all  $Y \in \mathcal{B}_1$ ,

hence  $X \in C(\mathcal{B}_1)$ . To prove that Expansion holds, let  $X \in C(\mathcal{B}_1) \cap \dots \cap C(\mathcal{B}_n)$ . Then it holds for all  $Y \in \mathcal{B}_1 \cup \dots \cup \mathcal{B}_n$  that  $X \geq Y$ . Hence,  $X \in C(\mathcal{B}_1 \cup \dots \cup \mathcal{B}_n)$ . ■

For a more extensive review of connections between choice and preference, the reader is referred to [Moulin, 1985]. For applications to epistemic choice and preference, see [Rott, 1993; Rott, 1999].

The results that connect choice functions with preference relations are so elegant that it may be somewhat unwelcome to question their meaningfulness. Nevertheless, the very idea of regarding choice as based on preference is quite problematic. This can be seen in two ways. First, as we saw above, the Chernoff property holds for all relational choice functions, but it nevertheless has counterintuitive consequences in realistic cases. Secondly, on a more basic, conceptual level, choices and preferences are entities of quite different categories. Preferences are *states of mind*. That I prefer  $x$  to  $y$  means that I consider  $x$  to be better than  $y$ . Choices are *actions*. That I have chosen  $x$  means that I have actually selected  $x$  (irrespective of whether I consider myself or anyone else to be better off through this choice).

Obviously, examples can easily be found in which choice and preference coincide, but there are also situations in which they clearly do *not* coincide. (Cf. [Sen, 1973].) We can, for instance, make choices that are not guided by preferences. A person may be indifferent between two alternatives, but still have to choose between them. This is exemplified by my recent choice between a match with a red head and one with a black head. Although I actually chose the red one, this does not mean that I prefer it, other than in some technical sense of ‘prefer’ that has been constructed to conciliate it with choice. A similar situation obtains when we have to choose between incommensurable alternatives. As was noted by Sen, it is particularly odd in this latter case to claim that choice reveals preference [Sen, 1973].

Clearly, preference can be defined (technically) as binary choice, but then difficulties arise in ‘interpreting preference thus defined as preference in the usual sense with the property that if a person prefers  $x$  to  $y$  then he must regard himself to be better off with  $x$  than with  $y$ ’ [Sen, 1973, p. 15].

#### 4 PREFERENCES WITH COMPATIBLE RELATA

In the previous sections, we have studied preferences that refer to a set of mutually exclusive alternatives that are taken as primitive units (exclusionary preferences). In actual discourse on preferences, we often make statements that transgress these limitations. In a discussion on musical pieces, someone may express preferences for orchestral music over chamber music, and also for Baroque over Romantic music. We may then ask her how she rates Baroque chamber music versus orchestral music from the Romantic

period. Assuming that these comparisons are all covered by one and the same preference relation, some of the relata of this preference relation are not mutually exclusive. Preferences with compatible relata may be called *combinative preferences*.

In Subsection 4.1, the use of sentential representation for the relata of combinative preferences is introduced, and in Subsection 4.2 some postulates for this type of preferences are discussed. Subsections 4.3–4.7 are devoted to the stepwise construction of a model in which preferences with compatible relata are based on exclusionary preferences. Some logical properties emerging from this construction are discussed in Subsection 4.8. An alternative construction of combinative preferences is discussed in Subsection 4.9.

#### 4.1 Sentential representation

In non-regimented language, all sorts of abstract and concrete entities can serve as the relata of preference relations. Thus, one may prefer butter to margarine, democracy to tyranny, or Bartok's fourth string quartet to his third. In spite of this, logical analyses of combinative preferences have been almost exclusively concerned with relata that represent states of affairs.

This practice is based on the assumption that combinative preferences over other types of entities can be adequately expressed as preferences over states of affairs. R. Lee went as far as to saying that 'all preferences can be understood in terms of preference among states of affairs or possible circumstances. A preference for bourbon, for example, may be a general preference that one drink bourbon instead of drinking scotch' [von Wright, 1963, p. 12; von Wright, 1972, pp. 143–144; Trapp, 1985, p. 303]. It is probably not quite as simple as that, but no other general-purpose representation of combinative preferences seems to be available. Therefore, combinative preferences will be taken to have states of affairs as relata. States of affairs, in their turn, will be represented in the usual way by sentences in sentential logic. The logical relationships among these sentences are assumed to include classical sentential logic. These choices are in line with tradition in philosophical logic.

Furthermore, it will be assumed that logically equivalent expressions can be substituted for each other. This assumption makes way for certain counter-intuitive inferences. Let  $p$  denote that you receive \$100 tomorrow,  $q$  that you receive \$50 tomorrow, and  $r$  that you are robbed of all the money that you own the day after tomorrow. Presumably, you prefer  $p$  to  $q$ . By intersubstitutivity, you then also prefer  $(p \& r) \vee (p \& \neg r)$  to  $q$ . However, the direct translation of  $(p \& r) \vee (p \& \neg r)$  into natural language does not seem to be preferable to the direct translation of  $q$  into natural language. The reason for this is that the disjunctive formulation of the comparison gives the impression that each of the disjuncts is preferred to  $q$  [Hansson, 1998b].

This and other counter-intuitive inferences can only be avoided by giving up intersubstitutivity, thereby losing much of the simplicity and logical strength of the formal structure. It is, on balance, better for most purposes to endure the somewhat strange consequences of intersubstitutivity than to pay the high price for getting rid of them.

Sentences will be denoted by lower-case letters  $p, q, \dots$ . The relations of weak preference, strict preference, and indifference will be denoted  $\geq, >, \equiv$  as before, with indices added to distinguish between different relations whenever needed.

#### 4.2 *Preference postulates for sentences*

The postulates for exclusionary preferences discussed in Subsections 2.2–2.4 can also be applied to combinative preferences. We therefore have the following properties:

$$\begin{aligned} p &\geq p \text{ (reflexivity)} \\ (p &\geq q) \vee (q \geq p) \text{ (completeness)} \\ p &\geq q \geq r \rightarrow p \geq r \text{ (transitivity)} \end{aligned}$$

and the various anticyclic properties discussed in Subsection 2.4. Reflexivity is clearly a desirable property. Everything that we can compare—not only complete alternatives—should be equal in value to itself. Completeness, on the other hand, is even more problematic for combinative than for exclusionary preferences. Even if your preferences are sufficiently developed to cover all possible comparisons between complete alternatives, they do not in general also cover all other possible comparisons. To see this, consider the four meals that can be composed out of the two dishes and the two drinks served at a small market stand. Suppose that you like each of the meals on the following list better than all those below it:

hamburger and beer  
sandwich and coffee  
sandwich and beer  
hamburger and coffee

It does not follow that you, in this context, prefer a meal with coffee to a meal with beer, or a meal with beer to a meal with coffee, or that you are indifferent between these two (incomplete) alternatives. You may very well lack a determinate preference between the two.

Some of the logical issues that arise in connection with combinative preferences do not arise for exclusionary preferences, for the simple reason that they are not defined for the latter. In particular, this applies to logical principles that refer to negated or disjunctive states of affairs.

Sören Halldén introduced the postulates  $(p \equiv q) \rightarrow (\neg q \equiv \neg p)$  and  $(p > q) \rightarrow (\neg q > \neg p)$  [Halldén, 1957, pp. 27–29 and 36]. von Wright used

the phrase ‘the principle of contraposition’ for the latter of the two principles [von Wright, 1972, pp. 147–149]. A similar postulate,  $(p \geq q) \rightarrow (\neg q \geq \neg p)$ , can be formed for weak preference. The term ‘contraposition’ can be used as a common term for all postulates of this general form. Thus,  $(p \geq q) \rightarrow (\neg q \geq \neg p)$  is contraposition of weak preference,  $(p \equiv q) \rightarrow (\neg q \equiv \neg p)$  is contraposition of indifference, etc.

The principles of contraposition have a clear intuitive appeal. If you prefer playing the piano to playing football, then not playing the piano should be worse for you than not playing football. But convincing counterexamples are not either difficult to find. Bengt Hansson provided an example against contraposition of strict preference [Hansson, 1968, pp. 428–429]. Let  $p$  denote that you win the first prize and  $q$  that you win some prize. Then  $p > q$  may reasonably hold, but it does not hold that  $\neg q > \neg p$ . To the contrary,  $\neg p$  is preferable to  $\neg q$ , since it leaves open the possibility of winning some other prize than the first prize. The same example can also be used against contraposition of weak preference. ( $p \geq q$  holds, but not  $\neg q \geq \neg p$ .)

The following example can be used against contraposition of indifference [Hansson, 1996a]. Let  $p$  denote that I have at least two copies of Rousseau’s *Du contrat social* on my bookshelf and  $q$  that I have at least one copy of it. Since I need the book, but cannot use more than one copy,  $p$  and  $q$  are of equal value, i.e.  $p \equiv q$ . However, it does not hold that  $\neg q \equiv \neg p$ . To the contrary,  $\neg q$  is worse than  $\neg p$ , since it means that I am in the precarious situation of not having access to *Du contrat social*.

The most widely quoted argument against contraposition was provided by Chisholm and Sosa. They claimed that ‘although that state of affairs consisting of there being happy egrets ( $p$ ) is better than that one that consists of there being stones ( $q$ ), that state of affairs that consists of there being no stones ( $\neg q$ ) is no better, nor worse, than that state of affairs consisting of there being no happy egrets ( $\neg p$ )’ [Chisholm and Sosa, 1966, p. 245]. More will be said about this example in Subsection 4.5.

Halldén also introduced the two principles  $(p > q) \leftrightarrow ((p \& \neg q) > (q \& \neg p))$  and  $(p \equiv q) \leftrightarrow ((p \& \neg q) \equiv (q \& \neg p))$  [Halldén, 1957, p. 28]. They have been accepted by von Wright. [von Wright 1963, pp. 24–25, 40, and 60.] The postulate  $(p > q) \leftrightarrow ((p \& \neg q) > (q \& \neg p))$  has been called conjunctive expansion’ [Jennings, 1967]. This term can be used for all relationships of the same form. (Thus,  $(p \equiv q) \leftrightarrow ((p \& \neg q) \equiv (q \& \neg p))$  is conjunctive expansion of indifference, etc.)

Conjunctive expansion is based on the reasonable assumption that ‘when one is to decide between two situations  $p$  and  $q$ , one does not actually compare these alternatives, but the situation that  $p$  is true while  $q$  is not on one hand and that  $q$  is true while  $p$  is not on the other’ [Hansson, 1968, p. 428]. However, as has been pointed out by several authors, conjunctive expansion cannot hold unrestrictedly since it would involve preferences with contradictory relata [Castañeda, 1958; Chisholm and Sosa, 1966; Quinn,

1974]. For concreteness, let  $p$  denote that a certain person is blind in her left eye and  $q$  that she is blind in two eyes. It is clearly worse to be blind in two eyes ( $q$ ) than to be blind in the left eye ( $p$ ). However, it does not follow that being blind only in the left eye ( $p \& \neg q$ ) is better than contradiction ( $q \& \neg p$ ).

Chisholm and Sosa chose to reject conjunctive expansion altogether, and so did Quinn [Chisholm and Sosa, 1966, p. 245; Quinn, 1974, p. 125]. It should be noted, however, that the arguments that led up to this conclusion refer to examples in which one of the relata implies the other. This was pointed out by Saito, who therefore claimed that conjunctive expansion of indifference and strict preference hold ‘only when both  $p \& \neg q$  and  $\neg p \& q$  are logically possible, i.e.,  $p$  and  $q$  do not imply each other’ [Saito, 1973, p. 388]. Cf. [Trapp, 1985, p. 318].

Intuitively, we would expect  $p \vee q$  to be intermediate in value between  $p$  and  $q$ . Suppose that I prefer that the painter paints the house white rather than that she paints it yellow. Then the information that she painted it either white or yellow should be at most as welcome as the information that she painted it white, and at least as welcome as the information that she painted it yellow. More generally speaking, the following should hold:

$$(p \geq q) \rightarrow (p \geq (p \vee q) \geq q) \text{ (disjunctive interpolation)}$$

von Wright argued that ‘[d]isjunctive preferences are conjunctively distributive’ in the sense that preferring  $p \vee q$  to  $r$  is essentially the same as preferring  $p$  to  $r$  and also  $q$  to  $r$  [von Wright, 1963, p. 26]. See also [Hansson, 1968, pp. 433–439]. This standpoint is expressed in the following two distributive axioms:

$$\begin{aligned} ((p \vee q) \geq r) &\leftrightarrow ((p \geq r) \& (q \geq r)) \text{ (left disjunctive distribution of } \geq) \\ (p \geq (q \vee r)) &\leftrightarrow ((p \geq q) \& (p \geq r)) \text{ (right disjunctive distribution of } \geq) \end{aligned}$$

Close connections hold between disjunctive distribution principles for  $\geq$  and  $>$ :

**OBSERVATION 19** (Bengt Hansson, 1968). Let  $\geq$  be a relation over a set  $\mathcal{L}$  of sentences that is closed under truth-functional operations. Let  $>$  be the strict part of  $\geq$ . Furthermore, let  $\geq$  be complete. Then

- (1)  $(p \vee q) \geq r \rightarrow (p \geq r) \& (q \geq r)$  is valid iff  $(p > q) \vee (p > r) \rightarrow p > (q \vee r)$  is valid.
- (2)  $(p \geq r) \& (q \geq r) \rightarrow (p \vee q) \geq r$  is valid iff  $p > (q \vee r) \rightarrow (p > q) \vee (p > r)$  is valid.
- (3)  $p \geq (q \vee r) \rightarrow (p \geq q) \& (p \geq r)$  is valid iff  $(p > r) \vee (q > r) \rightarrow (p \vee q) > r$  is valid.

(4)  $(p \geq q) \& (p \geq r) \rightarrow p \geq (q \vee r)$  is valid iff  $(p \vee q) > r \rightarrow (p > r) \vee (q > r)$  is valid.

If  $\geq$  is both complete and transitive, then:

(5) If  $(p \geq r) \& (q \geq r) \rightarrow (p \vee q) \geq r$  is valid, then so is  $p \geq (q \vee r) \rightarrow (p \geq q) \vee (p \geq r)$

(6) If  $(p \geq q) \& (p \geq r) \rightarrow p \geq (q \vee r)$  is valid, then so is  $(p \vee q) \geq r \rightarrow (p \geq r) \vee (q \geq r)$

**Proof.**

*Part 1:*

$(p \vee q) \geq r \rightarrow (p \geq r) \& (q \geq r)$

iff  $\neg((p \geq r) \& (q \geq r)) \rightarrow \neg((p \vee q) \geq r)$

iff  $\neg(p \geq r) \vee \neg(q \geq r) \rightarrow \neg((p \vee q) \geq r)$

iff  $(r > p) \vee (r > q) \rightarrow r > (p \vee q)$

Substitution yields  $(p > q) \vee (p > r) \rightarrow p > (q \vee r)$ .

*Parts 2-4* are proved in the same way as part 1.

*Part 5:* Suppose to the contrary that  $p \geq (q \vee r) \rightarrow (p \geq q) \vee (p \geq r)$  does not hold. Then we have  $p \geq (q \vee r)$  and, due to completeness,  $q > p$  and  $r > p$ . Transitivity yields  $q > (q \vee r)$  and  $r > (q \vee r)$ .

Due to part (2), since  $(p \geq r) \& (q \geq r) \rightarrow (p \vee q) \geq r$  is valid, so is  $p > (q \vee r) \rightarrow (p > q) \vee (p > r)$ . Applying the appropriate substitution instance to  $q > (q \vee r)$  we obtain  $(q > q) \vee (q > r)$ , and since  $>$  is irreflexive it follows that  $q > r$ . In the same way,  $r > (q \vee r)$  yields  $r > q$ . Since  $>$  is asymmetric, this is impossible, and we can conclude from the contradiction that  $p \geq (q \vee r) \rightarrow (p \geq q) \vee (p \geq r)$ .

*Part 6:* Suppose to the contrary that  $(p \vee q) \geq r \rightarrow (p \geq r) \vee (q \geq r)$  does not hold. Then we have  $(p \vee q) \geq r$  and, due to completeness,  $r > p$  and  $r > q$ . Transitivity yields  $(p \vee q) > p$  and  $(p \vee q) > q$ .

Due to part (4), since  $(p \geq q) \& (p \geq r) \rightarrow p \geq (q \vee r)$  is valid, so is  $(p \vee q) > r \rightarrow (p > r) \vee (q > r)$ . Applying the appropriate substitution instance to  $(p \vee q) > p$  we obtain  $(p > p) \vee (q > p)$ , hence  $q > p$ . In the same way,  $(p \vee q) > q$  yields  $p > q$ . This contradiction concludes the proof. ■

The following argument against  $(p \geq q) \& (p \geq r) \rightarrow p \geq (q \vee r)$  was proposed by Sven Danielsson and reported by Bengt Hansson [1968, p. 439]: A person who is away from home receives a letter. The following are statements about the letter:

- $p$  the letter says that the family's dog is feeling well
- $q$  the letter says that his son is feeling well
- $r$  the letter says that his daughter is feeling well
- $s$  the letter says that his wife has been killed in an accident



We can then reasonably expect to have  $p \geq (q \vee (r \& s))$  and  $p \geq ((q \& s) \vee r)$ . It follows from the postulate under discussion that  $p \geq (q \vee (r \& s) \vee (q \& s) \vee r)$ , or equivalently  $p \geq (q \vee r)$ , which is much less plausible. It should be observed, though, that this argument depends on the substitution of  $q \vee r$  for the logically equivalent sentence  $q \vee (r \& s) \vee (q \& s) \vee r$ . This substitution, rather than the application of disjunctive distribution, is the problematic step.

A close connection has been shown to hold between disjunctive interpolation and one direction of the disjunctive distribution principles:

**OBSERVATION 20** (Bengt Hansson, 1968). Let  $\geq$  be a relation over a set  $\mathcal{L}$  of sentences that is closed under truth-functional operations. Consider the following postulates:

- (i)  $(p \geq r) \& (q \geq r) \rightarrow ((p \vee q) \geq r)$
- (ii)  $(p \geq q) \& (p \geq r) \rightarrow (p \geq (q \vee r))$
- (iii) If  $p \geq q$  then  $p \geq (p \vee q) \geq q$  (*disjunctive interpolation*)

(1) If  $\geq$  is complete, and (i) and (ii) both hold, then so does (iii).  
If  $\geq$  is complete and transitive, then (iii) holds if and only if both (i) and (ii) hold.

**Proof.**

*Part 1:* Suppose to the contrary that (iii) does not hold. Then  $p \geq q$ , and it follows from completeness that either  $q > (p \vee q)$  or  $(p \vee q) > p$ . In the former case, it follows from (i) and part (2) of Observation 19 that  $(q > p) \vee (q > q)$ , and by the irreflexivity of  $>$  that  $q > p$ , contrary to  $p \geq q$ . In the latter case, it follows from (ii) and part (4) of Observation 19 that  $(p > p) \vee (q > p)$ , which is contradictory in the same way.

*Part 2:* Due to part 1, only one direction of the equivalence remains to prove.

In order to prove (i), let  $(p \geq r) \& (q \geq r)$ . Due to completeness, either  $p \geq q$  or  $q \geq p$ . In the former case, (iii) yields  $(p \vee q) \geq q$ , and with  $q \geq r$  and transitivity we obtain  $(p \vee q) \geq r$ . In the latter case, (iii) yields  $(p \vee q) \geq p$ , and with  $p \geq r$  and transitivity we again obtain  $(p \vee q) \geq r$ .

In order to prove (ii), let  $(p \geq q) \& (p \geq r)$ . Due to completeness, either  $q \geq r$  or  $r \geq q$ . In the former case, (iii) yields  $q \geq (q \vee r)$ , and with  $p \geq q$  transitivity yields  $p \geq (q \vee r)$ . In the latter case, (iii) yields  $r \geq (q \vee r)$ , and with  $p \geq r$  transitivity again yields  $p \geq (q \vee r)$ . ■

### 4.3 Connecting the two levels

We can expect strong connections to hold between the preferences that refer to a set of (mutually exclusive) alternatives and the preferences that

refer to incomplete relata that are associated with those same alternatives. In the formal representation, there are two major ways to construct these connections.

One of these is the *holistic* approach, that takes preferences over wholes for basic and uses them to derive combinative preferences. The other may be called the *aggregative* approach. It takes smaller units (expressible as incomplete relata) to be the fundamental bearers of value, and the values of complete alternatives are obtained by aggregating these units. A precise aggregative model was developed by Warren Quinn, on the basis of a proposal by Gilbert Harman. In Quinn's model, (intrinsic) values are assigned to certain basic propositions, which come in groups of mutually exclusive propositions. A conjunction of basic propositions is assigned the sum of the intrinsic values of its conjuncts. Various proposals have been made for the calculation of other truth-functional combinations of basic propositions [Harman, 1967; Quinn, 1974; Oldfield, 1977; Carlson, 1997; Danielsson, 1997].

The aggregative approach requires that there be isolable units of value and that these can be aggregated in some exact way, such as arithmetic addition. These conditions are satisfied in some utilitarian theories of moral betterness. This was indeed what Quinn had in mind; he considered it 'natural to suppose that the most evaluatively prior of all states of affairs are those which locate a specific sentient individual at a specific point along an evaluatively relevant dimension such as happiness, virtue, wisdom, etc. Thus for each pair consisting of an individual and a dimension there will be a distinct basic proposition for each point on that dimension which that individual may occupy' [Quinn, 1974, p. 131]. Cf. [Harman, 1967, p. 799].

The forms of utilitarianism that lend themselves to this mathematization are not the only reasonable theories of moral value. Furthermore, there are non-moral preference relations for which the aggregative approach does not seem at all suitable. Although many different factors may influence our judgment of the overall aesthetic value of a theatre performance, we cannot expect its overall value to be derivable in a mechanical way (such as addition) from these factors. The aesthetic value of the whole cannot be reduced in a summative way into isolable constituents. An analogous argument can be made against applying the aggregative approach to moral value according to intuitionist moral theories. 'The value of a whole must not be assumed to be the same as the sum of the values of its parts' [Moore, 1903, p. 28].

The holistic approach avoids these difficulties. Furthermore, it allows us to make use of the results already obtained for exclusionary preferences. An underlying exclusionary preference relation for (complete) alternatives can be used to derive preferences over the incomplete relata associated with these alternatives. Due to this, and to the implausibility in many cases of the decomposition required in aggregative models, the holistic approach

will be followed here. In other words, exclusionary preferences over a set of (mutually exclusive) alternatives are taken to be basic, and from them preferences over other relata can be derived.

This is not an unusual choice; the holistic approach has been chosen by most philosophical logicians dealing with combinative preferences. It must be borne in mind that it is a logical reconstruction rather than a faithful representation of actual deliberative or evaluative processes. In everyday life, combinative preferences do not seem to need the support of underlying exclusionary preferences. I prefer chess to boxing *simpliciter*. Only as a result of philosophical reflection do I prefer certain alternatives in which I watch or take part in chess to certain other such alternatives in which I watch or take part in pugilism. [Pollock, 1983, esp pp. 413–414; Beck, 1941, esp. p. 12]. This assumption, and the additional assumption that preferences over combinative relata can be reconstructed from the exclusionary preference relation (although, of course, they did not originate that way) have been made since they provide us with the basis for a series of fruitful formal explications of preference.

#### 4.4 *Constructing the alternatives*

What is the nature of the underlying alternatives that are used as a basis for modelling combinative preferences? Clearly, to each such alternative should be assigned a set of sentences, namely those sentences that hold in that alternative. This can be achieved through the introduction of a function that assigns a set of sentences to each alternative. However, an even simpler construction is possible. We may assume that if two alternatives support the same sentences, then they are treated in the same way by the preference relation. Under this assumption, we can dispense with the function that was just mentioned, and simplify the notation by identifying alternatives with their supported sets of sentences.

We will therefore assume that there is a non-empty language  $\mathcal{L}$  that is closed under the truth-functional operations  $\neg$  (negation),  $\vee$  (disjunction),  $\&$  (conjunction),  $\rightarrow$  (implication), and  $\leftrightarrow$  (equivalence). In order to express the logical relations between sentences in the formal language, an operator of logical consequence (Cn) will be used, such that for any set  $X$  of sentences,  $\text{Cn}(X)$  is the set of logical consequences of  $X$ . Cn includes classical sentential logic. (On consequence operators, see [Hansson, 1999a].)

Logically equivalent sets represent the same states of affairs, i.e., if  $\text{Cn}(S) = \text{Cn}(S')$  for some  $S, S' \subseteq \mathcal{L}$ , then  $S$  and  $S'$  represent the same state of affairs. Therefore, nothing is lost by requiring that all alternatives be logically closed, i.e. that if  $A \in \mathcal{A}$ , then  $A = \text{Cn}(A)$ . Clearly, the set of alternatives should be non-empty (and arguably, it should have at least two elements). This gives rise to the following definition:

DEFINITION 21. A subset  $\mathcal{A}$  of  $\wp(\mathcal{L})$  is a *sentential alternative set* (a set of *sentential alternatives*) if and only if:

- (1)  $\mathcal{A} \neq \emptyset$ , and
- (2) If  $A \in \mathcal{A}$ , then  $A$  is consistent and logically closed ( $A = \text{Cn}(A)$ ).

A comparison structure  $\langle \mathcal{A}, \geq \rangle$  is a *sentential comparison structure* if and only if  $\mathcal{A}$  is a sentential alternative set.

This definition allows for alternative sets such as  $\{\text{Cn}(\{p\}), \text{Cn}(\{p, q\})\}$  in which one alternative is a proper subset of another. Such sets should be excluded, and we also have reasons to exclude alternative sets such as  $\{\text{Cn}(\{p\}), \text{Cn}(\{q\})\}$  in which two alternatives are logically compatible. Mutual exclusivity is a characteristic feature of complete alternatives that distinguishes them from *relata* in general. These requirements can be summarized as follows:

DEFINITION 22. A subset  $\mathcal{A}$  of  $\wp(\mathcal{L})$  is a *set of mutually exclusive alternatives* if and only if:

- (1)  $\mathcal{A} \neq \emptyset$ ,
- (2) If  $A \in \mathcal{A}$ , then  $A$  is consistent and logically closed ( $A = \text{Cn}(A)$ ), and
- (3) If  $A, A' \in \mathcal{A}$  and  $A \neq A'$ , then  $A \cup A'$  is inconsistent. (*mutual exclusivity*)

This definition still allows for an alternative set such as the following:

$$\{\text{Cn}(\{p, q\}), \text{Cn}(\{p, \neg q\}), \text{Cn}(\{\neg p\})\}$$

For concreteness, consider the alternative set containing the following three alternatives, referring to possible ways of spending an evening:

- (1) Eating out ( $p$ ) and going to the theatre ( $q$ ).
- (2) Eating out ( $p$ ) and not going to the theatre ( $\neg q$ ).
- (3) Not eating out ( $\neg p$ ).

This is a somewhat strange set of alternatives, since the third alternative is less specified than the other two. If neither  $\text{Cn}(\{\neg p, q\})$  nor  $\text{Cn}(\{\neg p, \neg q\})$  has to be excluded from consideration, then the two of them should replace  $\text{Cn}(\{\neg p\})$ . If only one of them is available, then that one alone should replace  $\text{Cn}(\{\neg p\})$ . The outcome of amending the set of alternatives in either of these ways is a new alternative set in which all alternatives have been specified in the same respects. This makes it possible to compare them in a more uniform way. In the above case, such uniformity seems to be a

prerequisite for exhaustiveness in deliberation. On the other hand, there are also cases in which such exhaustiveness is not needed. This can be seen from an alternative interpretation of the above example that was proposed by Wlodek Rabinowicz. Let  $p$  denote that I go out and  $q$  that I wear a tie. Then  $\{\text{Cn}(\{p, q\}), \text{Cn}(\{p, \neg q\}), \text{Cn}(\{\neg p\})\}$  is an adequate alternative set, provided that  $q$  is value-relevant in the presence of  $p$  but not of  $\neg p$ . When exhaustiveness of deliberation is required, then the alternative set should satisfy the following condition:

**DEFINITION 23.** A subset  $\mathcal{A}$  of  $\wp(\mathcal{L})$  is a *set of contextually complete alternatives* if and only if:

- (1)  $\mathcal{A} \neq \emptyset$ ,
- (2) If  $A \in \mathcal{A}$ , then  $A$  is consistent and logically closed ( $A = \text{Cn}(A)$ ), and
- (3) If  $p \in A \in \mathcal{A}$  and  $A' \in \mathcal{A}$ , then either  $p \in A'$  or  $\neg p \in A'$ . (*relative negation-completeness* [Hansson, 1992])

**OBSERVATION 24.** Any set of contextually complete alternatives is also a set of mutually exclusive alternatives.

**Proof.** Conditions (1) and (2) of Definition 23 coincide with the equally numbered conditions of Definition 22. To see that condition (3) of Definition 22 is satisfied, let  $A, A' \in \mathcal{A}$  and  $A \neq A'$ . Without loss of generality, we can assume that there is some  $p \in A \setminus A'$ . It follows from condition (3) of Definition 23 that  $\neg p \in A'$ . Hence,  $\{p, \neg p\} \subseteq A \cup A'$ , so that condition (3) of Definition 22 is satisfied. ■

In most applications of the holistic approach to combinative preferences, the underlying alternatives have been possible worlds, represented by maximal consistent subsets of the language [Rescher, 1967; Åqvist, 1968; Cresswell, 1971; von Wright, 1972; van Dalen, 1974; von Kutschera, 1975; Trapp, 1985; Hansson, 1989; Hansson, 1996a].

**DEFINITION 25.** A subset  $\mathcal{A}$  of  $\wp(\mathcal{L})$  is a *set of possible worlds* if and only if:

- (1)  $\mathcal{A} \neq \emptyset$ ,
- (2) If  $A \in \mathcal{A}$ , then  $A$  is a maximal consistent subset of  $\mathcal{L}$ .

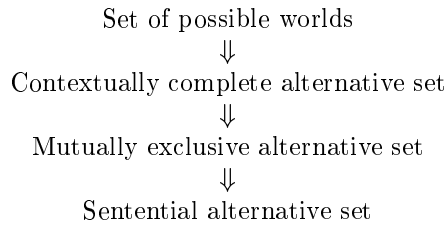
**OBSERVATION 26.** Any set of possible worlds is a set of contextually complete alternatives.

**Proof.** We need to show that if  $A$  is a maximal consistent subset of  $\mathcal{L}$ , then it is logically closed and satisfies relative negation-completeness. Both are standard results. For logical closure, suppose to the contrary that  $p \in \text{Cn}(A)$

and  $p \notin A$ . Then  $A \cup \{p\}$  is a superset of  $A$  and a consistent subset of  $\mathcal{L}$ , contrary to the assumption that  $A$  is a maximal consistent subset of  $\mathcal{L}$ . Relative negation-completeness, follows from the well-known fact that if  $p \in \mathcal{L}$  and  $A' \in \mathcal{A}$ , then either  $p \in A'$  or  $\neg p \in A'$ . (Suppose not. Then due to the logical closure of  $A$ ,  $p \notin \text{Cn}(A')$ , hence  $A' \cup \{\neg p\}$  is a superset of  $A'$  and a consistent subset of  $\mathcal{L}$ .) ■

Possible world modelling has the advantages of generality and logical beauty, but it also has the disadvantage of cognitive unrealism. In practice, we are not capable of deliberating on anything approaching the size of completely determinate possible worlds. Instead, we restrict our deliberations to objects of manageable size. It can therefore be argued that a more realistic holism should be based on smaller wholes, namely alternatives that cover all the aspects under consideration—but not all the aspects that might have been considered. This approach may be seen as an application of Simon’s ‘bounded rationality view’. Alternatives smaller than possible worlds are referred to in decision theory as ‘small worlds’ [Savage, 1954; Simon, 1957; Toda and Shuford, 1965; Toda, 1976; Schoemaker, 1982; Humphreys, 1983; Mendola, 1987; Hansson, 1993c; Hansson, 1996b].

In summary, we have the following series of increasingly general representations of (holistic) alternatives:



The following notation will turn out to be useful:

**DEFINITION 27.** Let  $\mathcal{A}$  be a set of sentential alternatives in  $\mathcal{L}$ . The subset  $\mathcal{L}_{\mathcal{A}}$  of  $\mathcal{L}$  is the set consisting exactly of (1)  $\cup\mathcal{A}$ , and (2) the truth-functional combinations of elements of  $\cup\mathcal{A}$ .

$\mathcal{L}_{\mathcal{A}}$  is called the  $\mathcal{A}$ -language. Its elements are the  $\mathcal{A}$ -sentences.

**DEFINITION 28.** Let  $\mathcal{A}$  be a set of sentential alternatives, and let  $p$  and  $q$  be elements of  $\cup\mathcal{A}$ . Then:

- $\models_{\mathcal{A}} q$  denotes that  $q \in A$  for all  $A \in \mathcal{A}$ .
- $p \models_{\mathcal{A}} q$  denotes that  $q \in A$  for all  $A \in \mathcal{A}$  such that  $p \in A$ .
- $p$  and  $q$  are  $\mathcal{A}$ -incompatible if and only if  $\models_{\mathcal{A}} \neg(p \& q)$

#### 4.5 Comparing compatible alternatives

We have now constructed the holistic preference structure. Before connecting it with combinative preferences, we need to have a closer look at the

characteristic feature of combinative preferences—namely that they allow for comparisons of compatible relata.

There is nothing strange or unusual with an utterance such as ‘It is better to have a cat than to have a dog’—although it is possible to have both a cat and a dog. We need to make the conventions explicit that guide our understanding of such utterances. A child may very well protest against the quoted sentence, saying: ‘No, it is better to have a dog, if you have a cat too.’ This we perceive as a sign that the child has misunderstood what it means to make this comparison. But why is it so, and what is a correct analysis?

There are at least two plausible answers to this question. According to one approach, that we may call the *adjustment account*, having both a cat and a dog is not under consideration. The sentence expresses a comparison between cat-and-no-dog and dog-and-no-cat. As proposed by Castañeda, ‘[w]hen St. Paul said “better to marry than to burn” he meant “it is better to marry and not to burn than not to marry and to burn” ’ [Castañeda, 1958, Cf. 1 Cor 7:9].

According to the other approach, that we may call the *totality account*, the comparison is between all-ways-to-have-a-dog and all-ways-to-have-a-cat. Since the alternatives in which one has both a dog and a cat are elements of both these sets of alternatives, their influence is cancelled out.

At first view, the difference between the adjustment and the totality account may seem rather inconsequential. In the first approach, the cat-and-dog cases are excluded for both relata, and in the second approach they are included in both relata but their effects are cancelled out. The difference will be more clearly seen when a third option is included in the comparison, such as ‘having a canary’. First consider the totality account. When we compare having a cat to having dog, the former alternative is represented by all-ways-to-have-a-cat. Similarly, when we compare having a cat to having canary, it is represented by all-ways-to-have-a-cat. The representation of having a cat is unaffected by what we compare it to. Next, consider the adjustment account. When we compare having a cat to having dog, the former alternative is represented by cat-and-no-dog alternatives. Similarly, when we compare having a cat to having canary, the former alternative is represented by cat-and-no-canary alternatives. Hence, the representation of cat-owning alternatives is constant according to the totality view, but according to the adjustment account it differs depending on what comparison is being made.

It has been argued that the adjustment account is better suited to express *ceteris paribus* preferences, whereas the totality view is better suited for decision-theoretical applications. The two approaches require different constructions and give rise to different logical properties. The adjustment approach will be developed in the rest of this subsection and in Subsections 4.6–4.8. We will return to the totality approach in Subsection 4.9.

Probably the first statement of the adjustment approach was given by

Halldén in his pioneering work on preference logic. He observed: ‘If we say that it would be better if  $p$  than if  $q$ , then we mean that it would be better if  $p \& \neg q$  than if  $q \& \neg p$ ’ [Halldén, 1957, p. 28]. (Cf. Subsection 4.2.) The same standpoint was taken by von Wright in his analysis of *ceteris paribus* preference [von Wright, 1963, pp. 24–25; von Wright, 1972, pp. 146–147]. The following has become a standard procedure in preference logic:

TRANSLATION PROCEDURE 1 (Halldén): The informal statement ‘ $p$  is better than  $q$ ’ is translated into  $(p \& \neg q) > (q \& \neg p)$ , and ‘ $p$  is equal in value to  $q$ ’ is translated into  $(p \& \neg q) \equiv (q \& \neg p)$ .

This is by no means bad as a first approximation. It works in cases such as the one just cited, when the alternatives are compatible and neither of them logically implies the other. It also works when the alternatives are logically incompatible. (Then  $p \& \neg q$  is equivalent to  $p$  and  $q \& \neg p$  to  $q$ .)

Halldén’s translation procedure runs into serious trouble when at least one of  $p$  and  $q$  logically implies the other. Then it forces us to compare a state of affairs to a contradictory state of affairs. This problem was observed by Kron and Milovanovic, who decided to accept the translation procedure but left as an open question ‘what it could mean to prefer a contradiction to something else or to prefer a state of affairs to a contradiction’ [Kron and Milovanovic, 1975, p. 187]. Cf. [Trapp, 1985, pp. 314–318]. The translation procedure breaks down completely when a sentence  $p$  is compared to itself; this comparison will be reduced to comparing logical contradiction to itself. Arguably, logical contradiction is equal in value to itself, but this does not seem to be the right reason why a non-contradictory statement  $p$  should be equal in value to itself. The right reason must be concerned with comparing  $p$  to itself, not contradiction to itself.

A remedy for this breakdown can be found simply by observing how the problematic cases are treated in informal discourse. Let  $p$  denote ‘I work hard and earn a lot of money’ and  $q$  ‘I work hard’. A case can be made for the viewpoint that  $p$  and  $q$  are incomparable. However, it should be clear that *if the comparison can be made in a meaningful way*, then it does not invoke the contradictory state of affairs  $p \& \neg q$ . Rather, the actual comparison takes place between  $p$  and  $q \& \neg p$ . It would seem correct to say that since  $p \& \neg q$  is contradictory, it is not used to replace  $p$ .

Similarly, a comparison between  $p$  and itself does not involve a comparison between  $p \& \neg p$  and itself. Since  $p \& \neg p$  is contradictory, it is not used to replace  $p$ . We are thus led to the following definition and translation procedure:

DEFINITION 29 (Hansson, 1989).  $p/q$  (‘ $p$  and if possible not  $q$ ’) is equal to  $p$  if  $p \& \neg q$  is logically contradictory, and otherwise it is equal to  $p \& \neg q$ .



TRANSLATION PROCEDURE 2: [Hansson 1989] The informal statement ‘ $p$  is better than  $q$ ’ is translated into  $(p/q) > (q/p)$ , and ‘ $p$  is equal in value to  $q$ ’ is translated into  $(p/q) \equiv (q/p)$ .

This procedure yields the same result as Halldén’s in the two cases when the latter turns out to be satisfactory, namely when  $p$  and  $q$  are incompatible and when they are compatible and neither of them implies the other. In the remaining cases, namely when one or both of  $p$  and  $q$  implies the other, the second procedure yields an intuitively more reasonable result than Halldén’s procedure.

However, we are not yet finished. The use of logical contradiction in the definition of  $/$  leads to undesired results. Let  $p$  denote ‘I go to the moon’ and  $q$  ‘I travel by spaceship’. A comparison between  $p$  and  $q$  will, according to translation procedure 2, be conceived as a comparison between  $p \& \neg q$  and  $q \& \neg p$ . However,  $p \& \neg q$  is not a serious possibility, although it is clearly *logically* possible. The only reasonable way to perform this comparison (outside of certain science fiction contexts) is to compare  $p$  to  $q \& \neg p$ . More generally,  $p/q$  should be defined as  $p$  not only when  $p \& \neg q$  is logically impossible but also when it is for other reasons not to be counted as possible, or more precisely: not included in any element of the alternative set.

DEFINITION 30.  $p/Aq$  (‘ $p$  and if  $\mathcal{A}$ -possible not  $q$ ’) is equal to  $p \& \neg q$  if  $p \not\vdash_{\mathcal{A}} q$ . If  $p \vdash_{\mathcal{A}} q$ , then  $p/Aq$  is equal to  $p$ .

TRANSLATION PROCEDURE 3: The informal statement ‘ $p$  is better than  $q$ ’ is translated into  $(p/Aq) > (q/AP)$ , and ‘ $p$  is equal in value to  $q$ ’ is translated into  $(p/Aq) \equiv (q/AP)$ .

This is the translation procedure that will be used in what follows.

We can now return to Chisholm’s and Sosa’s argument against contraposition, that was referred to in Subsection 4.2. They argued that ‘although that state of affairs consisting of there being happy egrets ( $p$ ) is better than that one that consists of there being stones ( $q$ ), that state of affairs that consists of there being no stones ( $\neg q$ ) is no better, nor worse, than that state of affairs consisting of there being no happy egrets ( $\neg p$ )’ [Chisholm and Sosa, 1966, p. 245].

Since stones and happy egrets can coexist, this is a comparison between compatible alternatives. Therefore, we can apply translation procedure 3. In other words, when comparing the existence of happy egrets with that of stones, we should compare alternatives in which there are happy egrets but no stones to alternatives in which there are stones but no happy egrets, i.e.,  $p \& \neg q$  to  $q \& \neg p$ . Next, let us compare  $\neg q$  to  $\neg p$ . By the same argument, this should be a comparison between, on the one hand, there being no stones and not being no happy egrets and, on the other hand, there being no happy egrets and not being no stones. This is, hidden behind double negations, the same comparison between  $p \& \neg q$  and  $q \& \neg p$  that we have just

made. Thus, from a logical point of view, it is unavoidable—once we have accepted translation procedure 3—that  $p > q$  holds if and only if  $\neg q > \neg p$ . What makes the example seem strange is that although we apply translation procedure 3 spontaneously to  $p$  and  $q$ , unaided intuition halts before the negated statements and does not perform the same operation.

#### 4.6 Representation functions

As we have just seen, an informal comparison between the relata  $p$  and  $q$  should be translated into a formal comparison between the relata  $p/_{\mathcal{A}}q$  and  $q/_{\mathcal{A}}p$ . Therefore, it should be derivable from a comparison between alternatives in which  $p/_{\mathcal{A}}q$  is true and alternatives in which  $q/_{\mathcal{A}}p$  is true. A pair  $\langle A_1, A_2 \rangle$  of alternatives, such that  $p/_{\mathcal{A}}q$  is true in  $A_1$  and  $q/_{\mathcal{A}}p$  is true in  $A_2$  will be called a *representation* of the pair  $\langle p/_{\mathcal{A}}q, q/_{\mathcal{A}}p \rangle$ .

DEFINITION 31. Let  $\mathcal{A}$  be a set of sentential alternatives. An element  $A$  of  $\mathcal{A}$  is a *representation* in  $\mathcal{A}$  of a sentence  $x$  if and only if  $x \in A$ .

An element  $\langle A, B \rangle$  of  $\mathcal{A} \times \mathcal{A}$  is a *representation* in  $\mathcal{A}$  of the pair  $\langle x, y \rangle$  of sentences if and only if  $x \in A$  and  $y \in B$ .

A sentence  $x$  or a pair  $\langle x, y \rangle$  of sentences is *representable* in  $\mathcal{A}$  if and only if it has a representation in  $\mathcal{A}$ .

More concisely,  $x$  is representable in  $\mathcal{A}$  if and only if  $x \in \cup \mathcal{A}$ , and  $\langle x, y \rangle$  if and only if  $x, y \in \cup \mathcal{A}$ .

Not all representations of  $\langle p/_{\mathcal{A}}q, q/_{\mathcal{A}}p \rangle$  need to be relevant to the comparison between  $p$  and  $q$ . Those that are relevant will be picked out by a *representation function*.

DEFINITION 32 (Hansson, 1989). A *representation function* for a set  $\mathcal{A}$  of sentential alternatives is a function  $f$  such that:

- (1) If  $\langle x, y \rangle$  is representable in  $\mathcal{A}$ , then  $f(\langle x, y \rangle)$  is a non-empty set of representations of  $\langle x, y \rangle$  in  $\mathcal{A}$ .
- (2) Otherwise,  $f(\langle x, y \rangle) = \emptyset$ .

Representation functions provide a general format for deriving combinative preference relations from exclusionary preference relations:

DEFINITION 33 (Hansson, 1989). Let  $\geq$  be a relation on the set  $\mathcal{A}$  of sentential alternatives, and  $f$  a representation function for  $\mathcal{A}$ . The weak preference relation  $\geq_f$ , the *f-extension* of  $\geq$ , is defined as follows:

$$p \geq_f q \text{ if and only if } A \geq B \text{ for all } \langle A, B \rangle \in f(\langle p/_{\mathcal{A}}q, q/_{\mathcal{A}}p \rangle).$$

$>_f$  is the strict part of  $\geq_f$ , and  $\equiv_f$  its symmetric part.

For most purposes it can be assumed that a comparison between  $p$  and  $q$  and one between  $q$  and  $p$  are based on comparisons between the same

pairs of complete alternatives. This assumption corresponds to the following symmetry property of representation functions:

**DEFINITION 34.** A representation function  $f$  for a set  $\mathcal{A}$  of sentential alternatives is *symmetric* if and only if for all sentences  $x, y \in \cup\mathcal{A}$  and all elements  $A$  and  $B$  of  $\mathcal{A}$ :

$$\langle A, B \rangle \in f(\langle x, y \rangle) \text{ if and only if } \langle B, A \rangle \in f(\langle y, x \rangle)$$

Another plausible property of a representation function is that reflexive comparisons of states of affairs (comparisons of a state of affairs to itself) should only be represented by reflexive comparisons of complete alternatives (comparisons of such an alternative to itself). This can also be required for comparisons between states of affairs that are coextensive, i.e. hold in exactly the same alternatives:

**DEFINITION 35.** A representation function  $f$  for a set  $\mathcal{A}$  of sentential alternatives satisfies *weak centering* if and only if for all sentences  $x \in \cup\mathcal{A}$  and all elements  $A_1$  and  $A_2$  of  $\mathcal{A}$ :

$$\text{If } \langle A_1, A_2 \rangle \in f(\langle x, x \rangle), \text{ then } A_1 = A_2.$$

Furthermore, it satisfies *centering* if and only if for all sentences  $x, y \in \cup\mathcal{A}$ :

$$\text{If } \models_{\mathcal{A}} x \leftrightarrow y, \text{ and } \langle A_1, A_2 \rangle \in f(\langle x, y \rangle), \text{ then } A_1 = A_2.$$

We should expect a derived combinative preference relation to say about the complete alternatives exactly what the underlying exclusionary preference relation says about them. If there is a sentence  $a$  that has  $A$  as its only representation, and a sentence  $b$  that has  $B$  as its only representation, then  $a \geq_f b$  should hold if and only if  $A \geq B$  holds. Indeed, this condition holds for all representation functions.

**OBSERVATION 36** (Hansson, 1989). Let  $\geq$  be a relation on the set  $\mathcal{A}$  of sentential alternatives and  $f$  a representation function for  $\mathcal{A}$ . Furthermore, let  $A$  and  $B$  be elements of  $\mathcal{A}$ , and  $a$  and  $b$  sentences such that  $A$  is the only representation of  $a$  in  $\mathcal{A}$ , and  $B$  the only representation of  $b$  in  $\mathcal{A}$ . Then:

$$a \geq_f b \text{ if and only if } A \geq B.$$

**COROLLARY:** If  $\mathcal{A}$  is a mutually exclusive alternative set, and  $A = \text{Cn}(\{a\})$  and  $B = \text{Cn}(\{b\})$ , then  $a \geq_f b$  if and only if  $A \geq B$ .

**Proof.** Since  $A$  is the only representation of  $a$  in  $\mathcal{A}$ , it is also the only representation of  $a/\mathcal{A}b$  in  $\mathcal{A}$ . Similarly,  $B$  is the only representation of  $b/\mathcal{A}a$  in  $\mathcal{A}$ . Definition 32 yields  $f(\langle a/\mathcal{A}b, b/\mathcal{A}a \rangle) = \{\langle A, B \rangle\}$ . According to Definition 33,  $a \geq_f b$  iff  $A \geq B$ . ■

#### 4.7 *Ceteribus paribus preferences*

The more precise construction of a representation function will have to depend on the type of preferences that we aim at representing. This subsection is devoted to the construction of representation functions for *ceteris paribus* preferences.

A recipe for this construction can be extracted from von Wright's early work. He defined *ceteris paribus* preferences as follows:

'[A]ny given total state of the world, which contains  $p$  but not  $q$ , is preferred to a total state of the world, which differs from the first in that it contains  $q$  but not  $p$ , but otherwise is identical with it.' [von Wright, 1963, p. 31]. Cf. [Quinn, 1974, p. 124; von Wright, 1972, pp 140 and 147].

This recipe needs some modifications before it can be put to use. Where von Wright refers to ' $p$  but not  $q$ ', i.e. to  $p \& \neg q$ , we should instead refer to  $p/\Delta q$ , as explained in Subsection 4.5. Furthermore, von Wright's concept of 'identity' is problematic. It is more reasonable to require that the alternatives are, given the differences required for them to represent the respective sentences, as similar as possible in all other respects.

With these modifications, the quoted passage can be rephrased as follows:

Any given alternative which contains  $p/\Delta q$  is preferred to an alternative which differs from the first in that it contains  $q/\Delta p$ , but is otherwise as similar as possible it.

Before this recipe can be formalized, we need to operationalize 'as similar as possible'. In a follow-up article, von Wright attempted to solve this problem (under another description) by means of an arithmetical count of differences in terms of logically independent atomic states of the world [von Wright, 1972, pp. 146–147]. He assumed that there are  $n$  logically independent states of affairs  $p_1, \dots, p_n$ , and  $2^n$  possible states of the world  $w_1, \dots, w_{2^n}$  that can be compared in terms of the  $n$  atomic states. If two states of affairs  $q$  and  $r$  are molecular combinations of in all  $m$  out of the  $n$  atomic states, then a *ceteris paribus* comparison of  $q$  and  $r$  keeps the other  $n - m$  states constant.

Unfortunately, this simple construction is not as promising as it might seem at first sight. Its major weakness is that the choice of atomic states can be made in different ways that give rise to different relations of similarity. For an example of this, consider the following four sentential alternatives:

- (1a)  $\text{Cn}(\{p, q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}\})$
- (1b)  $\text{Cn}(\{\neg p, q_1, q_2, q_3, q_4, q_5, q_6, q_7, q_8, q_9, q_{10}\})$
- (2a)  $\text{Cn}(\{p, r_1, r_2, r_3, r_4, r_5, r_6, r_7, r_8, r_9, r_{10}\})$

$$(2b) \text{ Cn}(\{\neg p, \neg r_1, \neg r_2, \neg r_3, \neg r_4, \neg r_5, \neg r_6, \neg r_7, \neg r_8, \neg r_9, \neg r_{10}\})$$

Intuitively, (1a) and (1b) seem to represent a *ceteris paribus* comparison between  $p$  and  $\neg p$ , whereas (2a) and (2b) do not. But suppose that  $r_1, \dots, r_{10}$  are definable in terms of  $p, q_1, \dots, q_{10}$  as follows:

$$\begin{aligned} r_1 &\leftrightarrow (p \leftrightarrow q_1) \\ \dots & \\ r_{10} &\leftrightarrow (p \leftrightarrow q_{10}) \end{aligned}$$

Then, in going from (1a) and (1b) to (2a) and (2b), we shift to another, expressively equivalent set of atomic sentences. Since there are no objectively given logical atoms, there is in general ample scope for choosing among sets of atomic sentences that are equivalent in terms of what can be expressed in the language, but not in terms of von Wright's similarity measure.

It seems inescapable that a non-trivial explication of similarity will have to make use of more information than what is inherent in the logic. Probably the most transparent way to represent similarity is by means of a similarity relation, as follows:

**DEFINITION 37** (Williamson, 1988). For any set  $\Psi$ , the four-place relation  $T$  is a similarity relation over  $\Psi$  if and only if, for all  $U, V, W, X, Y, Z \in \Psi$ :

$$(T1) \quad T(W, X, Y, Z) \vee T(Y, Z, W, X) \quad (\text{completeness})$$

$$(T2) \quad T(U, V, W, X) \ \& \ T(W, X, Y, Z) \rightarrow T(U, V, Y, Z) \quad (\text{transitivity})$$

$$(T3) \quad T(X, X, Y, Z)$$

$$(T4) \quad T(X, Y, Y, Y) \rightarrow X = Y$$

$$(T5) \quad T(X, Y, Y, X) \quad (\text{symmetry})$$

The strict part of  $T$  is defined as follows:

$$\hat{T}(W, X, Y, Z) \leftrightarrow T(W, X, Y, Z) \ \& \ \neg T(Y, Z, W, X)$$

$T(W, X, Y, Z)$  should be read 'W is at least as similar to X as is Y to Z', and  $\hat{T}(W, X, Y, Z)$  'W is more similar to X than is Y to Z'. This axiomatization of the four-termed similarity relation was proposed by T. Williamson [1988], see also [Hansson, 1992]. It is a generalization of a three-termed similarity relation that was introduced earlier by David Lewis [Lewis, 1973a, pp. 48ff; 1973b, p. 560; 1981]. Lewis's relation  $S(X, Y, Z)$  should be read 'X is more similar to Y than is Z'. It can be defined from the four-termed relation through the relationship  $S(X, Y, Z) \leftrightarrow T(X, Y, Z, Y)$ .

(T1) and (T2) combine to say that similarity is a weak ordering (complete and transitive). (T3) and (T4) combine to say that maximal similarity obtains between two arguments if and only if they are identical, and (T5)

states that the degree of similarity between two arguments does not depend on the order in which they are taken.

How can similarity be used to explicate *ceteris paribus* preferences? Two proposals are available in the literature. One of these is based on the intuition that when comparing  $p$  and  $q$  we should look for pairs of alternatives  $\langle A_1, A_2 \rangle$  that satisfy the following two conditions:

(1) *The representation condition*

$A_1$  is a representation of  $p$ , and  $A_2$  a representation of  $q$ .

(2) *The unfocused similarity condition*

$A_1$  and  $A_2$  are maximally similar to each other, as compared to other pairs of alternatives that satisfy the representation condition.

These assumptions give rise to the following definition of *ceteris paribus* preferences. It can be seen as a formalized version of the basic ideas behind von Wright's explication of *ceteris paribus* preferences, as quoted above.

DEFINITION 38. Let  $\mathcal{A}$  be a set of sentential alternatives and  $T$  a similarity relation over  $\mathcal{A}$ . Then  $f$  is the *unfocused similarity-maximizing* representation function that is based on  $T$ , if and only if it is a representation function and, for all  $x, y \in \cup \mathcal{A}$  and  $A, B \in \mathcal{A}$ :

$\langle A, B \rangle \in f(\langle x, y \rangle)$  if and only if  $x \in A, y \in B$ , and  $T(A, B, A', B')$  holds for all  $A', B' \in \mathcal{A}$  such that  $x \in A'$  and  $y \in B'$ .

Furthermore, if  $\geq$  is a reflexive relation on  $\mathcal{A}$ , then  $\geq_f$  is an *unfocused similarity-maximizing* preference relation if and only if it is based on an unfocused similarity-maximizing representation function.

The adequacy criteria introduced in Subsection 4.6 are satisfied by unfocused similarity-maximizing representation functions.

OBSERVATION 39. Let  $\mathcal{A}$  be a set of sentential alternatives and  $f$  an unfocused similarity-maximizing representation function over  $\mathcal{A}$ . Then  $f$  satisfies centring and symmetry.

**Proof.**

*Centring:* Suppose to the contrary that centring does not hold for  $f$ . Then, according to Definition 35, there are  $A_1, A_2 \in \mathcal{A}$  and  $x, y \in \cup \mathcal{A}$  such that  $\models_{\mathcal{A}} x \leftrightarrow y$ ,  $\langle A_1, A_2 \rangle \in f(\langle x, y \rangle)$ , and  $A_1 \neq A_2$ . It follows from  $A_1 \neq A_2$ , using (T4), that  $\neg T(A_1, A_2, A_2, A_2)$ . On the other hand, according to Definition 38, it follows from  $x, y \in A_2$  that  $T(A_1, A_2, A_2, A_2)$ . Contradiction.

*Symmetry:* Suppose to the contrary that symmetry is not satisfied. Then there are  $x, y \in \cup \mathcal{A}$  and  $A, B \in \mathcal{A}$  such that  $\langle A, B \rangle \in f(\langle x, y \rangle)$  and  $\langle B, A \rangle \notin$

$f(\langle y, x \rangle)$ . It follows by Definition 38 from  $\langle B, A \rangle \notin f(\langle y, x \rangle)$  that there are  $A', B' \in \mathcal{A}$  such that  $x \in A', y \in B'$ , and  $\neg T(B, A, B', A')$ .

On the other hand, it follows according to Definition 38 from  $\langle A, B \rangle \in f(\langle x, y \rangle)$ ,  $x \in A'$ , and  $y \in B'$ , that  $T(A, B, A', B')$ . We can use (T5) to obtain  $T(B, A, A, B)$  and  $T(A', B', B', A')$ . Two applications of (T2) to  $T(B, A, A, B)$ ,  $T(A, B, A', B')$ , and  $T(A', B', B', A')$  provide us with  $T(B, A, B', A')$ , contrary to what was just shown. This contradiction concludes the proof. ■

The other similarity-based approach to *ceteris paribus* preferences is based on the assumption that there is a privileged alternative  $A_0$  that can serve as a reference point. If the alternative set consists of possible worlds, then the actual world can be used as such a reference point. This amounts to the following alternative to (2):

(2') *The focused similarity condition*

$A_1$  is maximally similar to  $A_0$ , as compared to other alternatives that satisfy the representation condition with respect to  $p$ . In the same way,  $A_2$  is maximally similar to  $A_0$ , as compared to other alternatives that satisfy the representation condition with respect to  $q$ .

This is an approach with some tradition in the literature on preference logic [von Kutschera, 1975; Trapp, 1985; Hansson, 1989]. In the present formal framework it can be expressed as follows:

DEFINITION 40. Let  $\mathcal{A}$  be a set of sentential alternatives,  $A_0$  an element of  $\mathcal{A}$ , and  $T$  a similarity relation over  $\mathcal{A}$ . Then  $f$  is the  $A_0$ -focused similarity-maximizing representation function that is based on  $T$ , if and only if it is a representation function  $f$  such that, for all  $x, y \in \cup \mathcal{A}$  and  $A, B \in \mathcal{A}$ :

$$\begin{aligned} \langle A, B \rangle \in f(\langle x, y \rangle) & \text{ if and only if } x \in A, y \in B, \\ T(A, A_0, A', A_0) & \text{ holds for all } A' \text{ such that } x \in A' \in \mathcal{A}, \text{ and} \\ T(B, A_0, B', A_0) & \text{ holds for all } B' \text{ such that } y \in B' \in \mathcal{A}. \end{aligned}$$

Furthermore, if  $\geq$  is a reflexive relation on  $\mathcal{A}$ , then  $\geq_f$  is an  $A_0$ -focused similarity-maximizing preference relation if and only if it is based on an  $A_0$ -focused similarity-maximizing representation function.

Perhaps surprisingly, from a formal point of view the focused approach can be subsumed under the unfocused approach.

OBSERVATION 41 (Hansson, 1998b). Let  $\langle \mathcal{A}, \geq \rangle$  be a sentential comparison structure such that  $\mathcal{A}$  is finite and that  $\geq$  is complete and transitive. Then:

1. If  $\geq_f$  is a focused similarity-maximizing preference relation, based on  $\langle \mathcal{A}, \geq \rangle$  and a similarity relation  $T$ , then it is also an unfocused

similarity-maximizing preference relation, based on  $\langle \mathcal{A}, \geq \rangle$  and another similarity relation  $T'$ .

2. The converse relationship does not hold in general.

**Proof.**

*Part 1:* Let  $\geq_f$  be focused on  $A_0$ . For each  $X \in \mathcal{A}$ , let  $\delta(X)$  be the number of elements of  $Y$  in  $\mathcal{A}$  such that  $\hat{T}(Y, A_0, X, A_0)$ . Let  $T'(X, Y, Z, W)$  hold if and only if *either*  $X = Y$  *or*  $\delta(X) + \delta(Y) \leq \delta(Z) + \delta(W)$  and  $Z \neq W$ . Then  $T'$  satisfies conditions (T1)-(T5) of Definition ???. Furthermore, if  $x$  and  $y$  are  $\mathcal{A}$ -incompatible, then:

$T'(A, B, A', B')$  whenever  $x \in A' \in \mathcal{A}$  and  $y \in B' \in \mathcal{A}$ ,  
 iff  $\delta(A) + \delta(B) \leq \delta(A') + \delta(B')$  whenever  $x \in A' \in \mathcal{A}$  and  $y \in B' \in \mathcal{A}$ ,  
 iff  $\delta(A) \leq \delta(A')$  whenever  $x \in A' \in \mathcal{A}$  and  $\delta(B) \leq \delta(B')$   
 whenever  $y \in B' \in \mathcal{A}$ ,  
 iff  $T(A, A_0, A', A_0)$  whenever  $x \in A' \in \mathcal{A}$  and  $T(B, A_0, B', A_0)$   
 whenever  $y \in B' \in \mathcal{A}$ .

It follows that  $T'$  gives rise to the same preference relation via Definition 38 as does  $T$  via Definition 40.

*Part 2:* We are going to exhibit an unfocused similarity-maximizing preference relation that cannot be reconstructed as a focused similarity-maximizing preference relation. For that purpose, let  $\geq$  be transitive and complete, and let  $p, q,$  and  $r$  be mutually exclusive relata. Let  $\mathcal{A} = \{A, B, C, D\}$  be contextually complete, with  $r \in A, p \in B, q \in C,$  and  $r \in D$ . Furthermore, let  $\geq$  be a weak ordering (complete and transitive) over  $\mathcal{A}$ , such that  $A > B > C > D$ . Let  $T$  be a similarity relation over  $\mathcal{A}$  such that similarity coincides with closeness in the following diagram:



(The distances  $A - B, B - C,$  and  $C - D$  are the same.) Let  $f$  be the unfocused representation function based on  $T$  in the manner of Definition 38. Then  $p \geq_f q$  and  $q \geq_f r$  but not  $p \geq_f r$ . It is easy to show that a focused similarity-maximizing preference relation always satisfies transitivity for mutually exclusive relata, if the underlying exclusionary preference relation is transitive. Therefore,  $\geq_f$  cannot be reconstructed as focused. ■

Due to its greater generality, the unfocused approach will be used in what follows. To simplify the terminology, it will be called ‘similarity-maximizing’ rather than ‘unfocused similarity-maximizing’.



#### 4.8 Logical properties of combinative preferences

It is natural to ask to what extent various logical properties of the underlying exclusionary preference relation  $\geq$  are reflected in the logic of the derived preference relation  $\geq_f$ . More precisely, a logical property is *transmitted* by  $f$  if and only if: If  $\geq$  has this property, then so does  $\geq_f$  [Hansson, 1996a].

Reflexivity is not transmitted by all representation functions, but it is transmitted by a wide range of representation functions, including those that are similarity-maximizing.

**OBSERVATION 42** (Hansson, 1998b). Let  $\geq$  be a reflexive relation on the sentential alternative set  $\mathcal{A}$ , and let  $f$  be a representation function for  $\mathcal{A}$ . Then  $\geq_f$  is reflexive if and only if for all sentences  $x$  and all elements  $A_1$  and  $A_2$  of  $\mathcal{A}$ : If  $\langle A_1, A_2 \rangle \in f(\langle x, x \rangle)$ , then  $A_1 \geq A_2$ .

**COROLLARY.** If  $f$  satisfies weak centring, then  $\geq_f$  is reflexive.

**Proof.** Immediate from Definitions 33 and 35. ■

Completeness of the exclusionary preference relation  $((A \geq B) \vee (B \geq A))$  is not transmitted to similarity-maximizing preference relations. Indeed, a fairly strong negative result can be obtained that holds for all types of representation functions.

**OBSERVATION 43** (Hansson, 1998b). Let  $f$  be a representation function for the sentential alternative set  $\mathcal{A}$ , such that there are two elements  $p$  and  $q$  of  $\cup\mathcal{A}$  and four pairwise distinct elements  $A_1, A_2, B_1,$  and  $B_2$  of  $\mathcal{A}$  such that  $\langle A_1, B_1 \rangle \in f(\langle p/\mathcal{A}q, q/\mathcal{A}p \rangle)$  and  $\langle B_2, A_2 \rangle \in f(\langle q/\mathcal{A}p, p/\mathcal{A}q \rangle)$ . Then there is a complete relation  $\geq$  over  $\mathcal{A}$  such that  $(p \geq_f q) \vee (q \geq_f p)$  does not hold.

**Proof.** Let  $\geq$  be complete and such that  $A_2 > B_2$  and  $B_1 > A_1$ . Then it follows from  $\langle B_2, A_2 \rangle \in f(\langle q/\mathcal{A}p, p/\mathcal{A}q \rangle)$  and  $\neg(B_2 \geq A_2)$  that  $\neg(q \geq_f p)$ . Similarly, it follows from  $\langle A_1, B_1 \rangle \in f(\langle p/\mathcal{A}q, q/\mathcal{A}p \rangle)$  and  $\neg(A_1 \geq B_1)$  that  $\neg(p \geq_f q)$ . ■

Transitivity is not in general transmitted by similarity-maximizing representation functions, not even for pairwise incompatible relations.

**OBSERVATION 44** (Hansson, 1998b). Let  $\geq$  be a transitive and complete relation on the contextually complete alternative set  $\mathcal{A}$ , and let  $f$  be a representation function on  $\mathcal{A}$ . Then  $p \geq_f q \geq_f r \rightarrow p \geq_f r$  does not hold in general if  $\geq_f$  is similarity-maximizing, not even if  $p, q,$  and  $r$  are pairwise incompatible.

**Proof.** See part 2 of the proof of Observation 41. ■

We can now turn to such logical properties of combinative preferences that cannot be transmitted since they are not defined for exclusionary preferences. The principles of contraposition and conjunctive expansion hold for

$\geq_f$ ,  $\equiv_f$ , and  $>_f$  in the principal case when neither of the relata contextually implies the other. These results apply to all preference relations that are based on a representation function in the manner of Definition 33.

**OBSERVATION 45** (Hansson 1998b). Let  $\geq$  be a reflexive relation on the sentential alternative set  $\mathcal{A}$  and  $f$  a representation function for  $\mathcal{A}$ . Furthermore, let  $p$  and  $q$  be elements of  $\cup\mathcal{A}$  such that  $p \not\leq_{\mathcal{A}} q$  and  $q \not\leq_{\mathcal{A}} p$ . Then:

- (1)  $p \geq_f q \rightarrow \neg q \geq_f \neg p$ ,
- (2)  $p \equiv_f q \rightarrow \neg q \equiv_f \neg p$ , and
- (3)  $p >_f q \rightarrow \neg q >_f \neg p$ .

**Proof.**

*Part 1:* Let  $p \geq_f q$ , and let  $\langle A, B \rangle \in f(\langle \neg q /_{\mathcal{A}} \neg p, \neg p /_{\mathcal{A}} \neg q \rangle)$ . It follows from  $p \not\leq_{\mathcal{A}} q$  that  $\neg q /_{\mathcal{A}} \neg p$  is equivalent to  $p /_{\mathcal{A}} q$ , and from  $q \not\leq_{\mathcal{A}} p$  that  $\neg p /_{\mathcal{A}} \neg q$  is equivalent to  $q /_{\mathcal{A}} p$ . Thus,  $\langle A, B \rangle \in f(\langle p /_{\mathcal{A}} q, q /_{\mathcal{A}} p \rangle)$ . It follows from  $p \geq_f q$  that  $A \geq B$ . Since this holds for all  $\langle A, B \rangle \in f(\langle \neg q /_{\mathcal{A}} \neg p, \neg p /_{\mathcal{A}} \neg q \rangle)$ , we may conclude that  $\neg q \geq_f \neg p$ .

*Part 2:* From part 1.

*Part 3:* Suppose that  $p >_f q$ , i.e.,  $p \geq_f q$  and  $\neg(q \geq_f p)$ . It follows from  $p \geq_f q$ , in the same way as in part 1, that  $\neg q \geq_f \neg p$ .

It follows from  $\neg(q \geq_f p)$  that there is some  $A$  and some  $B$  such that  $\langle B, A \rangle \in f(\langle q /_{\mathcal{A}} p, p /_{\mathcal{A}} q \rangle)$  and  $\neg(B \geq A)$ . Then,  $\langle B, A \rangle \in f(\langle \neg p /_{\mathcal{A}} \neg q, \neg q /_{\mathcal{A}} \neg p \rangle)$ . From this and  $\neg(B \geq A)$  follows  $\neg(\neg p \geq_f \neg q)$ .

From  $\neg q \geq_f \neg p$  and  $\neg(\neg p \geq_f \neg q)$  it follows that  $\neg q >_f \neg p$ . ■

**OBSERVATION 46** (Hansson, 1998b). Let  $\geq$  be a reflexive relation on the sentential alternative set  $\mathcal{A}$  and  $f$  a representation function for  $\mathcal{A}$ . Furthermore, let  $p$  and  $q$  be elements of  $\cup\mathcal{A}$  such that  $p \not\leq_{\mathcal{A}} q$  and  $q \not\leq_{\mathcal{A}} p$ . Then:

- (1)  $p \geq_f q \leftrightarrow (p \& \neg q) \geq_f (q \& \neg p)$ ,
- (2)  $p \equiv_f q \leftrightarrow (p \& \neg q) \equiv_f (q \& \neg p)$ , and
- (3)  $p >_f q \leftrightarrow (p \& \neg q) >_f (q \& \neg p)$ .

**Proof.** For all  $\langle A, B \rangle \in \mathcal{A} \times \mathcal{A}$ ,  $\langle A, B \rangle \in f(p /_{\mathcal{A}} q, q /_{\mathcal{A}} p)$  iff  $\langle A, B \rangle \in f(p \& \neg q /_{\mathcal{A}} q \& \neg p, q \& \neg p /_{\mathcal{A}} p \& \neg q)$ . The proof proceeds as that of Observation 45. ■

Disjunctive interpolation does not hold in general for similarity-maximizing preference relations, but if  $p$  and  $q$  are  $\mathcal{A}$ -incompatible then it holds for all preference relations that are based on representation functions.

OBSERVATION 47 (Hansson, 1998b). Let  $\mathcal{A}$  be a sentential alternative set,  $\geq$  a reflexive relation on  $\mathcal{A}$  and  $f$  a representation function on  $\mathcal{A}$ . Let  $p$  and  $q$  be  $\mathcal{A}$ -incompatible elements of  $\cup\mathcal{A}$ . Then:

- (1)  $(p \geq_f (p \vee q)) \leftrightarrow (p \geq_f q)$
- (2)  $((p \vee q) \geq_f p) \leftrightarrow (q \geq_f p)$
- (3)  $(p \geq_f q) \rightarrow (p \geq_f (p \vee q) \geq_f q)$

**Proof.** For *part 1*, we have:  $f(\langle p/\mathcal{A}(p \vee q), (p \vee q)/\mathcal{A}p \rangle) = f(\langle p, q \rangle) = f(\langle p/\mathcal{A}q, q/\mathcal{A}p \rangle)$ . *Part 2* is proved in the same way, and *part 3* follows from parts 1 and 2. ■

OBSERVATION 48 (Hansson, 1998b). Let  $\geq$  be a transitive and complete relation on the contextually complete alternative set  $\mathcal{A}$ . Let  $\geq_f$  be a similarity-maximizing extension of  $\geq$ . Then:

- (1)  $(p \geq_f q) \rightarrow (p \geq_f (p \vee q))$  does not hold in general.
- (2)  $(p \geq_f q) \rightarrow ((p \vee q) \geq_f q)$  does not hold in general.

**Proof.**

*Part 1:* Let  $\mathcal{A} = \{A, B, C\}$ , with  $p, \neg q \in A$ ,  $\neg p, q \in B$ , and  $p, q \in C$ . Let  $A > B > C$ . Let  $f$  be based on a similarity relation  $T$  such that for all  $X, Y, Z$ , and  $W$ , if  $X \neq Y$  and  $Z \neq W$  then  $T(X, Y, Z, W)$ . Then  $f(\langle p/\mathcal{A}q, q/\mathcal{A}p \rangle) = \{\langle A, B \rangle\}$  and  $A \geq B$ , so that  $p \geq_f q$ . However, it follows from  $\langle C, B \rangle \in f(\langle p/\mathcal{A}p \vee q, p \vee q/\mathcal{A}p \rangle)$  and  $B > C$  that  $\neg(p \geq_f (p \vee q))$ .

*Part 2:* Let  $\mathcal{A} = \{A, B, C\}$ , with  $p, q \in A$ ,  $p, \neg q \in B$ , and  $\neg p, q \in C$ . Let  $A > B > C$ . Let  $f$  be based on a similarity relation  $T$  such that for all  $X, Y, Z$ , and  $W$ , if  $X \neq Y$  and  $Z \neq W$  then  $T(X, Y, Z, W)$ . Then  $f(\langle p/\mathcal{A}q, q/\mathcal{A}p \rangle) = \{\langle B, C \rangle\}$  and  $B \geq C$ , so that  $p \geq_f q$ . However, it follows from  $\langle B, A \rangle \in f(\langle p \vee q/\mathcal{A}q, q/\mathcal{A}p \vee q \rangle)$  and  $A > B$  that  $\neg((p \vee q) \geq_f q)$ . ■

The properties of disjunctive distribution referred to in Subsection 4.2 do not hold in general for similarity-maximizing preference relations [Hansson, 1998b]. However, the following much weaker properties for pairwise incompatible relata can be shown to hold:

OBSERVATION 49. Let  $\geq$  be a transitive and complete relation on the contextually complete alternative set  $\mathcal{A}$ . Let  $p, q, r \in \cup\mathcal{A}$ , and let  $\geq_f$  be a similarity-maximizing extension of  $\geq$ . Then:

- (1)  $((p \vee q) \geq_f r) \rightarrow (p \geq_f r) \vee (q \geq_f r)$  holds if  $p, q$ , and  $r$  are pairwise  $\mathcal{A}$ -incompatible elements of  $\mathcal{A}$ .

(2)  $(p \geq_f (q \vee r)) \rightarrow (p \geq_f q) \vee (p \geq_f r)$  holds if  $p, q,$  and  $r$  are pairwise  $\mathcal{A}$ -incompatible elements of  $\mathcal{A}$ .

**Proof.**

*Part 1:* Let  $(p \vee q) \geq_f r$ . Then there is at least one pair  $\langle X, Y \rangle$  of elements of  $\mathcal{A}$  such that  $p \vee q \in X, r \in Y,$  and  $T(X, Y, X', Y')$  for all  $X', Y' \in \mathcal{A}$  such that  $p \vee q \in X'$  and  $r \in Y'$ . Clearly, either  $p \in X$  or  $q \in X$ .

If  $p \in X,$  let  $\langle X'', Y'' \rangle$  be any pair such that  $p \in X''$  and  $r \in Y''$ . Then  $p \vee q \in X'',$  and it follows that  $T(X, Y, X'', Y'')$ . Since this holds for all pairs  $\langle X'', Y'' \rangle$  with  $p \in X''$  and  $r \in Y'', p \geq_f r$ .

If  $q \in X,$  then  $q \geq_f r$  follows in the same way.

*Part 2:* Let  $p \geq_f (q \vee r)$ . Then there is at least one pair  $\langle X, Y \rangle$  of elements of  $\mathcal{A}$  such that  $p \in X, q \vee r \in Y,$  and  $T(X, Y, X', Y')$  for all  $X', Y' \in \mathcal{A}$  such that  $p \in X'$  and  $q \vee r \in Y'$ . Clearly, either  $q \in Y$  or  $r \in Y$ .

If  $q \in Y,$  let  $\langle X'', Y'' \rangle$  be a pair such that  $p \in X''$  and  $q \in Y''$ . Then  $q \vee r \in Y'',$  and it follows that  $T(X, Y, X'', Y'')$ . Since this holds for all pairs  $\langle X'', Y'' \rangle$  with  $p \in X''$  and  $q \in Y'', p \geq_f q$ .

If  $r \in Y,$  then  $p \geq_f r$  follows in the same way. ■

In summary, similarity-maximizing preference relations have very weak, perhaps disappointingly weak, logical properties. However, it does not follow that they are inadequate to represent *ceteris paribus* preferences. Counterexamples to several of the rejected principles were given in Subsection 4.2, and it can reasonably be argued that an adequate logic for *ceteris paribus* preferences should be quite weak.

#### 4.9 The totality approach

In this subsection, we are going to return to the alternative approach to combinative preferences that was mentioned in Subsection 4.5, namely the totality approach. It is based on the representation of (single) sentences rather than of pairs of sentences. Each sentence is represented by the set of alternatives to which it is applied.

DEFINITION 50. Let  $p \in \mathcal{L}_{\mathcal{A}}$ . Then:

$$repr_{\mathcal{A}}(p) = \{X \in \mathcal{A} \mid p \in X\}$$

The index of  $repr_{\mathcal{A}}$  is deleted whenever convenient.

Preferences over sentences can be derived from preferences over sets of alternatives, according to the simple principle that  $p \geq' q$  holds if and only if  $repr(p) \geq' repr(q)$  holds. More precisely:

DEFINITION 51. Let  $\langle \mathcal{A}, \geq \rangle$  be a comparison structure. Then a relation  $\geq'$  over  $\wp(\mathcal{A}) \setminus \{\emptyset\}$  is a *subset-extension* of  $\geq$  if and only if it holds for all  $A, B \in \mathcal{A}$  that  $\{A\} \geq' \{B\}$  iff  $A \geq B$ .

If  $\langle \mathcal{A}, \geq \rangle$  is a sentential comparison structure, then  $p \geq' q$  is an abbreviated notation for  $\text{repr}_{\mathcal{A}}(p) \geq' \text{repr}_{\mathcal{A}}(q)$ .

Several types of subset-extensions have been investigated. Among the simplest are those that are based on the decision-theoretical principles of maximin and maximax. To express them we need some additional notation and terminology.

DEFINITION 52. Let  $\emptyset \neq \mathcal{B} \subseteq \mathcal{A}$ , let  $p \in \cup \mathcal{A}$ , and let  $\geq$  be a relation on  $\mathcal{A}$ . Then:

$$\begin{aligned} \max(\mathcal{B}) &= \{X \in \mathcal{B} \mid (\forall Y \in \mathcal{B})(X \geq Y)\} \\ \min(\mathcal{B}) &= \{X \in \mathcal{B} \mid (\forall Y \in \mathcal{B})(Y \geq X)\} \end{aligned}$$

The elements of  $\max(\mathcal{B})$  are the ( $\geq$ -)maximal elements of  $\mathcal{B}$ , and those of  $\min(\mathcal{B})$  are its ( $\geq$ -)minimal elements.

$\max(p)$  is an abbreviation of  $\max(\text{repr}(p))$ , and  $\min(p)$  an abbreviation of  $\min(\text{repr}(p))$ . Furthermore:

$$\begin{aligned} \max(\mathcal{B}) \geq \max(\mathcal{D}) &\text{ holds if and only if } X \geq Y \\ &\text{ for all } X \in \max(\mathcal{B}) \text{ and } Y \in \max(\mathcal{D}). \\ \min(\mathcal{B}) \geq \min(\mathcal{D}) &\text{ holds if and only if } X \geq Y \\ &\text{ for all } X \in \min(\mathcal{B}) \text{ and } Y \in \min(\mathcal{D}). \end{aligned}$$

DEFINITION 53. Let  $\langle \mathcal{A}, \geq \rangle$  be a sentential comparison structure. The *maximin preference relation* that is based on  $\geq$  is the relation  $\geq_{\mathbf{i}}$  on  $\cup \mathcal{A}$  such that:

$$\mathcal{B} \geq_{\mathbf{i}} \mathcal{D} \text{ if and only if } \min(\mathcal{B}) \geq \min(\mathcal{D}).$$

Furthermore, the *maximax preference relation* based on  $\geq$  is the relation  $\geq_{\mathbf{x}}$  on  $\cup \mathcal{A}$  such that:

$$\mathcal{B} \geq_{\mathbf{x}} \mathcal{D} \text{ if and only if } \max(\mathcal{B}) \geq \max(\mathcal{D}).$$

$\mathcal{B} >_{\mathbf{i}} \mathcal{D}$  is an abbreviation of  $(\mathcal{B} \geq_{\mathbf{i}} \mathcal{D}) \ \& \ \neg(\mathcal{D} \geq_{\mathbf{i}} \mathcal{B})$ , and  $\mathcal{B} \equiv_{\mathbf{i}} \mathcal{D}$  of  $(\mathcal{B} \geq_{\mathbf{i}} \mathcal{D}) \ \& \ (\mathcal{D} \geq_{\mathbf{i}} \mathcal{B})$ .  $\mathcal{B} >_{\mathbf{x}} \mathcal{D}$  and  $\mathcal{B} \equiv_{\mathbf{x}} \mathcal{D}$  are defined analogously.

In the indices,  $\mathbf{x}$  refers to maximization of the maximum and  $\mathbf{i}$  to maximization of the minimum.

Neither completeness nor transitivity is transmitted from an exclusionary preference relation  $\geq$  to  $\geq_{\mathbf{i}}$  and  $\geq_{\mathbf{x}}$ . However, the combined property of being both complete and transitive is transmitted.

OBSERVATION 54 (Hansson, 1998b). Let  $\mathcal{A}$  be a finite and sentential alternative set.

- (1) Let  $\geq$  be a complete relation on  $\mathcal{A}$ . It does not follow in general that  $\geq_{\mathbf{i}}$  and  $\geq_{\mathbf{x}}$  are complete.

- (2) Let  $\geq$  be a reflexive and transitive relation on  $\mathcal{A}$ . It does not follow in general that  $\geq_i$  and  $\geq_x$  are transitive.
- (3) Let  $\geq$  be a complete and transitive relation on  $\mathcal{A}$ . Then  $\geq_i$  and  $\geq_x$  are complete and transitive.

**Proof.**

*Part 1:* Let  $\mathcal{A} = \{X, Y, Z\}$  and let  $repr(p) = \{X, Y\}$ ,  $repr(q) = \{Z\}$ , and  $Z > X \equiv Y > Z$ . Then  $min(p) = \{X, Y\}$  and  $min(q) = \{Z\}$ . It follows from  $Z > X$  that  $p \geq_i q$  does not hold and from  $Y > Z$  that  $q \geq_i p$  does not hold. The same example can be used to prove the incompleteness of  $\geq_x$ .

*Part 2:* Let  $\mathcal{A} = \{X, Y_1, Y_2, Z\}$ ,  $repr(p) = \{X\}$ ,  $repr(q) = \{Y_1, Y_2\}$ ,  $repr(r) = \{Z\}$ , and  $\geq = \{\langle X, X \rangle, \langle Y_1, Y_1 \rangle, \langle Y_2, Y_2 \rangle, \langle Z, Z \rangle, \langle Z, X \rangle\}$ . Then  $min(p) = \{X\}$ ,  $min(q) = \emptyset$ , and  $min(r) = \{Z\}$ . Since  $min(q) = \emptyset$ ,  $p \geq_i q$  and  $q \geq_i r$  hold vacuously, whereas  $p \geq_i r$  does *not* hold. The same example can be used to show that  $\geq_x$  is not transitive.

*Part 3:* For completeness of  $\geq_i$ , it is sufficient to note that due to the completeness and transitivity of  $\geq$ , either  $min(p) \geq min(q)$  or  $min(q) \geq min(p)$ . For the transitivity of  $\geq_i$ , let  $p \geq_i q \geq_i r$ . Let  $X \in min(p)$  and  $Z \in min(r)$ . Since  $q \in \cup \mathcal{A}$ ,  $repr(q)$  is non-empty. Since  $\mathcal{A}$  is finite and  $\geq$  is complete and transitive, so is  $min(q)$ . Let  $Y \in min(q)$ . Then  $X \geq Y$  follows from  $p \geq_i q$  and  $Y \geq Z$  from  $q \geq_i r$ . Due to the transitivity of  $\geq$ ,  $X \geq Z$ . Since this holds for all elements  $X$  of  $min(p)$  and  $Z$  of  $min(r)$ , we may conclude that  $p \geq_i r$ .

The completeness and transitivity of  $\geq_x$  follows in the same way. ■

Contraposition does not hold for either maximin nor maximax preferences, but conjunctive expansion of strict preference holds in both cases.

OBSERVATION 55. Let  $\geq$  be a transitive and complete relation on the contextually complete alternative set  $\mathcal{A}$ , and let  $p, \neg p, q, \neg q \in \cup \mathcal{A}$ . Then:

- (1a)  $p \geq_i q \rightarrow \neg q \geq_i \neg p$  does not hold in general.
- (1b)  $p \equiv_i q \rightarrow \neg q \equiv_i \neg p$  does not hold in general.
- (1c)  $p >_i q \rightarrow \neg q >_i \neg p$  does not hold in general.
- (2a)  $p \geq_x q \rightarrow \neg q \geq_x \neg p$  does not hold in general.
- (2b)  $p \equiv_x q \rightarrow \neg q \equiv_x \neg p$  does not hold in general.
- (2c)  $p >_x q \rightarrow \neg q >_x \neg p$  does not hold in general.

**Proof.**

*Parts 1a, 1b, 2a, and 2b:* Let  $\mathcal{A} = \{A, B, C, D\}$ ,  $p, q \in A$ ,  $\neg p, q \in B$ ,  $p, \neg q \in C$ , and  $p, q \in D$ . Let  $A \equiv B > C \equiv D$ .

*Part 1c and 2c:* Let  $\mathcal{A} = \{A, B, C, D\}$ ,  $\neg p, \neg q \in A$ ,  $p, \neg q \in B$ ,  $\neg p, q \in C$ , and  $\neg p, \neg q \in D$ . Let  $A \equiv B > C \equiv D$ . ■

**OBSERVATION 56.** Let  $\geq$  be a transitive and complete relation on the contextually complete alternative set  $\mathcal{A}$ , and let  $p \& \neg q, q \& \neg p \in \cup \mathcal{A}$ . Then:

(1a)  $(p \geq_i q) \rightarrow ((p \& \neg q) \geq_i (q \& \neg p))$  does not hold in general.

(1b)  $(p \equiv_i q) \rightarrow ((p \& \neg q) \equiv_i (q \& \neg p))$  does not hold in general.

(1c)  $(p >_i q) \rightarrow ((p \& \neg q) >_i (q \& \neg p))$

(2a)  $(p \geq_x q) \rightarrow ((p \& \neg q) \geq_x (q \& \neg p))$  does not hold in general.

(2b)  $(p \equiv_x q) \rightarrow ((p \& \neg q) \equiv_x (q \& \neg p))$  does not hold in general.

(2c)  $(p >_x q) \rightarrow ((p \& \neg q) >_x (q \& \neg p))$

**Proof.**

*Parts 1a, 1b, 2a, and 2b:* Let  $\mathcal{A} = \{A, B, C, D\}$ ,  $p, q \in A$ ,  $\neg p, q \in B$ ,  $p, \neg q \in C$ , and  $p, q \in D$ . Let  $A \equiv B > C \equiv D$ .

*Part 1c:* Let  $p >_i q$ . Then  $\min(p) > \min(q)$ . Clearly  $\min(p \& \neg q) \geq \min(p)$ . Furthermore, it follows from  $\min(p) > \min(q)$  that  $\neg p \in \min(q)$ , hence  $q \& \neg p \in \min(q)$ , hence  $\min(q) \equiv \min(q \& \neg p)$ . We can apply transitivity to  $\min(p \& \neg q) \geq \min(p)$ ,  $\min(p) > \min(q)$ , and  $\min(q) \equiv \min(q \& \neg p)$ , and obtain  $\min(p \& \neg q) > \min(q \& \neg p)$ , so that  $(p \& \neg q) >_i (q \& \neg p)$ .

*Part 2c:* Let  $p >_x q$ . Then  $\max(p) > \max(q)$ . Clearly  $\max(q) \geq \max(q \& \neg p)$ . Furthermore, it follows from  $\max(p) > \max(q)$  that  $\neg q \in \max(p)$ , hence  $p \& \neg q \in \max(p)$ , hence  $\max(p \& \neg q) \equiv \max(p)$ . We can apply transitivity to  $\max(p \& \neg q) \equiv \max(p)$ ,  $\max(p) > \max(q)$ , and  $\max(q) \geq \max(q \& \neg p)$ , and obtain  $\max(p \& \neg q) > \max(q \& \neg p)$ , so that  $(p \& \neg q) >_x (q \& \neg p)$ . ■

Disjunctive interpolation holds for both  $\geq_i$  and  $\geq_x$ .

**OBSERVATION 57.** Let  $\geq$  be a transitive and complete relation on the contextually complete alternative set  $\mathcal{A}$ , and let  $p, q \in \cup \mathcal{A}$ . Then:

(1)  $(p \geq_i q) \rightarrow (p \geq_i (p \vee q) \geq_i q)$

(2)  $(p \geq_x q) \rightarrow (p \geq_x (p \vee q) \geq_x q)$

**Proof.**

*Part 1:* Let  $p \geq_i q$ . Then  $\min(p) \geq \min(p \vee q)$  and  $\min(p \vee q) \equiv \min(q)$ .

*Part 2:* Let  $p \geq_x q$ . Then  $\max(p) \equiv \max(p \vee q)$  and  $\max(p \vee q) \geq \max(q)$ . ■

Fairly strong principles of disjunctive distribution can be obtained for  $\geq_i$  and  $\geq_x$ :

OBSERVATION 58. Let  $\geq$  be a transitive and complete relation over the contextually complete alternative set  $\mathcal{A}$ . Then:

$$(1a) ((p \vee q) \geq_i r) \leftrightarrow (p \geq_i r) \& (q \geq_i r)$$

$$(1b) (p \geq_i (q \vee r)) \leftrightarrow (p \geq_i q) \vee (p \geq_i r)$$

$$(2a) ((p \vee q) \geq_x r) \leftrightarrow (p \geq_x r) \vee (q \geq_x r)$$

$$(2b) (p \geq_x (q \vee r)) \leftrightarrow (p \geq_x q) \& (p \geq_x r)$$

**Proof.**

*Part 1a:* Left to right: Let  $(p \vee q) \geq_i r$ . Then  $\min(p \vee q) \geq \min(r)$ . Since  $\min(p) \geq \min(p \vee q)$ , transitivity yields  $\min(p) \geq \min(r)$ , hence  $p \geq_i r$ . We can prove  $q \geq_i r$  in the same way.

Right to left: Let  $p \geq_i r$  and  $q \geq_i r$ . Then  $\min(p) \geq \min(r)$  and  $\min(q) \geq \min(r)$ . Since either  $\min(p \vee q) \equiv \min(p)$  or  $\min(p \vee q) \equiv \min(q)$ , we can use transitivity to obtain  $\min(p \vee q) \geq \min(r)$ , hence  $(p \vee q) \geq_i r$ .

*Part 1b:* Left to right: Let  $p \geq_i (q \vee r)$ . Then  $\min(p) \geq \min(q \vee r)$ . Since either  $\min(q \vee r) \equiv \min(q)$  or  $\min(q \vee r) \equiv \min(r)$ , transitivity yields either  $\min(p) \geq \min(q)$  or  $\min(p) \geq \min(r)$ , hence either  $p \geq_i q$  or  $p \geq_i r$ .

Right to left: For symmetry reasons, we may assume that  $p \geq_i q$ . Then  $\min(p) \geq \min(q)$ . Since  $\min(q) \geq \min(q \vee r)$ , transitivity yields  $\min(p) \geq \min(q \vee r)$ , hence  $p \geq_i (q \vee r)$ .

*Part 2a:* Left to right: Let  $(p \vee q) \geq_x r$ , i.e.  $\max(p \vee q) \geq \max(r)$ . Since either  $\max(p) \equiv \max(p \vee q)$  or  $\max(q) \equiv \max(p \vee q)$ , we can use transitivity to obtain either  $\max(p) \geq \max(r)$  or  $\max(q) \geq \max(r)$ , hence either  $p \geq_x r$  or  $q \geq_x r$ .

Right to left: Let  $p \geq_x r$ , i.e.  $\max(p) \geq \max(r)$ . We have  $\max(p \vee q) \geq \max(p)$ , and transitivity yields  $\max(p \vee q) \geq \max(r)$ , i.e.  $(p \vee q) \geq_x r$ . If  $q \geq_x r$ , then  $(p \vee q) \geq_x r$  follows in the same way.

*Part 2b:* Left to right: Let  $p \geq_x (q \vee r)$ . Then  $\max(p) \geq \max(q \vee r)$ . Since  $\max(q \vee r) \geq \max(q)$ , transitivity yields  $\max(p) \geq \max(q)$ , so that  $p \geq_x q$ . We can obtain  $p \geq_x r$  in the same way.

Right to left: Let  $p \geq_x q$  and  $p \geq_x r$ . Then  $\max(p) \geq \max(q)$  and  $\max(p) \geq \max(r)$ . Furthermore, either  $\max(q) \equiv \max(q \vee r)$  or  $\max(r) \equiv \max(q \vee r)$ . In either case it follows by transitivity that  $\max(p) \geq \max(q \vee r)$ , hence  $p \geq_x (q \vee r)$ . ■



The following observation introduces a couple of fairly problematic properties for maximin and maximax preferences.

**OBSERVATION 59.** Let  $\geq$  be a relation on the sentential alternative set  $\mathcal{A}$ , and let  $p, q \in \cup \mathcal{A}$ . Then:

- (1) If  $\models_{\mathcal{A}} p \rightarrow q$ , then  $p \geq_i q$ .
- (2) If  $\models_{\mathcal{A}} p \rightarrow q$ , then  $q \geq_x p$ .

**COROLLARY:** Let  $\geq$  be a relation on the sentential alternative set  $\mathcal{A}$ , and let  $p, q \in \cup \mathcal{A}$ . Then:

- (1)  $p \geq_i (p \vee q)$
- (2)  $(p \vee q) \geq_x p$

**Proof.**

*Part 1:* It follows from  $\models_{\mathcal{A}} p \rightarrow q$  that  $\min(p) \geq \min(q)$ , hence  $p \geq_i q$ .

*Part 2:* It follows from  $\models_{\mathcal{A}} p \rightarrow q$  that  $\max(q) \geq \max(p)$ , hence  $q \geq_x p$ . ■

Part (1) of this observation has been called the ‘Nobel peace prize postulate’. [Hansson 1998b] Let  $q$  denote that a certain statesman stops a war, and  $p$  that he first starts a war and then stops it. Let  $\mathcal{A}$  be an alternative set that contains representations of  $p$  and  $q$ . Then  $\models_{\mathcal{A}} p \rightarrow q$  is satisfied, and we can conclude that  $p \geq_i q$ , i.e.  $p$  is (in the maximin sense) at least as good a behaviour as  $q$ . It is not difficult, either, to find examples that bring out the strangeness of part (2). We may, for instance, let  $q$  denote some violent action and  $p$  the same action, performed in self-defence.

The properties listed in the Corollary of Observation 59 were used by Packard in axiomatic characterizations. Maximin preference is characterized by transitivity, completeness,  $p \geq_i (p \vee q)$ , and  $(p \geq_i r) \& (q \geq_i r) \rightarrow ((p \vee q) \geq_i r)$ . Maximax preference is characterized by transitivity, completeness,  $(p \vee q) \geq_x p$ , and  $(p \geq_x q) \& (p \geq_x r) \rightarrow (p \geq_x (q \vee r))$ . [Packard 1979]

Maximin and maximax preferences are not they only subset-extended preference relations of interest. To begin with, they are not the only such relations that are determined exclusively by the best and worst elements of a set. (On such relations, see [Barbera *et al.*, 1984].) Two other such preference relations are the interval maximin and interval maximax relations:

**DEFINITION 60** (Hansson, 1998b). Let  $\geq$  be a relation on the sentential alternative set  $\mathcal{A}$ . The *interval maximin preference relation*  $\geq_{ix}$  based on  $\geq$  is the relation on  $\wp(\mathbf{A}) \setminus \emptyset$  such that for all  $\mathcal{B}, \mathcal{D} \in \wp(\mathbf{A}) \setminus \emptyset$ :

- (1) If  $\min(\mathcal{B}) > \min(\mathcal{D})$ , then  $\mathcal{B} >_{ix} \mathcal{D}$ .
- (2) If  $\min(\mathcal{B}) \equiv \min(\mathcal{D})$ , then  $\mathcal{B} \geq_{ix} \mathcal{D}$  if and only if  $\max(\mathcal{B}) \geq \max(\mathcal{D})$ .

$\mathcal{B} >_{ix} \mathcal{D}$  is an abbreviation of  $(\mathcal{B} \geq_{ix} \mathcal{D}) \ \& \ \neg(\mathcal{D} \geq_{ix} \mathcal{B})$ , and  $\mathcal{B} \equiv_{ix} \mathcal{D}$  of  $(\mathcal{B} \geq_{ix} \mathcal{D}) \ \& \ (\mathcal{D} \geq_{ix} \mathcal{B})$ .

DEFINITION 61 (Hansson, 1998b). Let  $\geq$  be a relation on the sentential alternative set  $\mathcal{A}$ . The *interval maximax preference relation*  $\geq_{xi}$  based on  $\geq$  is the relation on  $\wp(\mathbf{A}) \setminus \emptyset$  such that for all  $\mathcal{B}, \mathcal{D} \in \wp(\mathbf{A}) \setminus \emptyset$ :

- (1) If  $\max(\mathcal{B}) > \max(\mathcal{D})$ , then  $\mathcal{B} >_{xi} \mathcal{D}$ .
- (2) If  $\max(\mathcal{B}) \equiv \max(\mathcal{D})$ , then  $\mathcal{B} \geq_{xi} \mathcal{D}$  if and only if  $\min(\mathcal{B}) \geq \min(\mathcal{D})$ .

$\mathcal{B} >_{xi} \mathcal{D}$  is an abbreviation of  $(\mathcal{B} \geq_{xi} \mathcal{D}) \ \& \ \neg(\mathcal{D} \geq_{xi} \mathcal{B})$ , and  $\mathcal{B} \equiv_{xi} \mathcal{D}$  of  $(\mathcal{B} \geq_{xi} \mathcal{D}) \ \& \ (\mathcal{D} \geq_{xi} \mathcal{B})$ .

$\geq_{ix}$  maximizes first the minimum and after that the maximum, whereas  $\geq_{xi}$  does this in the reverse order.

Interval maximin preference is a modification of the maximin preference relation. The latter not only gives precedence to the avoidance of bad worst outcomes (which is an expression of cautiousness), but also refrains from making any difference between two relata that both satisfy this criterion. In contrast, the interval maximin preference relation maximizes both worst and best alternatives, but gives maximization of the former absolute priority over maximization of the latter. Similarly, the interval maximax preference relation maximizes both worst and best alternatives, but gives maximization of the latter absolute priority over maximization of the former.

Another interesting group of subset-extensions are those that rank sets of alternatives according to their medians. If a set has an odd number of elements, then the set consisting of the element in the middle according to the  $\geq$ -ranking is the median according to  $\geq$ . If there is an even number of elements, then the two elements closest to the middle form the median [Nitzan and Prasanta, 1984]. Hence, in this case as well, one or two elements determine the value of the whole set.

A good case can be made that all elements of a set of alternatives should have an influence on the value of the set as a whole. This can easily be achieved if a numerical value (utility) is assigned to each element of  $\mathcal{A}$ . Fishburn has provided an axiomatic characterization of preferences over  $\wp(\mathcal{A}) \setminus \emptyset$  that are derived from utility assignments to  $\mathcal{A}$  by means of even-chance lotteries [Fishburn, 1972].

## 5 PREFERENCES AND MONADIC CONCEPTS

In addition to the comparative notions, ‘better’ and ‘of equal value’, informal discourse on values contains monadic (one-place) value predicates, such as ‘good’, ‘best’, ‘very bad’, ‘fairly good’, etc. It also contains monadic normative concepts such as ‘ought’, ‘may’, ‘forbidden’, etc. This section

is devoted to the connection between preference relations and some major types of monadic predicates. Throughout this section,  $\geq'$  denotes a (weak) combinative preference relation that operates on the union  $\cup\mathcal{A}$  of some contextually complete alternative set  $\mathcal{A}$ .  $>'$  and  $\equiv'$  are its strict and symmetric parts, respectively. The construction of  $\geq'$  will be left open, but the constructions discussed in Section 4 are obvious candidates.

Subsection 5.1 introduces two general categories of monadic predicates. Subsections 5.2–5.3 are devoted to ‘good’ and ‘bad’, Subsection 5.4 to some other monadic value predicates, and Subsection 5.5 to normative predicates.

### 5.1 Positive and negative predicates

What is better than something good is itself good. Many other value predicates—such as ‘best’, ‘not worst’, ‘very good’, ‘excellent’, ‘not very bad’, ‘acceptable’, etc.—have the same property. If one of these predicates holds for  $p$ , then it also holds for everything that is better than  $p$  or equal in value to  $p$ . This property will be called ‘ $\geq'$ -positivity’, or (when there is no risk of confusion), simply ‘positivity’.

DEFINITION 62 (Hansson, 1990). A monadic predicate  $H$  is  $\geq'$ -positive if and only if for all  $p$  and  $q$ :

$$Hp \ \& \ (q \geq' p) \rightarrow Hq.$$

Similarly, ‘bad’ has the converse property that if  $p$  is bad, then whatever is worse than or equal in value to  $p$ , is also bad. Other predicates that share this property are ‘very bad’, ‘worst’, and ‘not best’. This property will be called ‘( $\geq'$ -)negativity’.

DEFINITION 63 (Hansson, 1990). A monadic predicate  $H$  is  $\geq'$ -negative if and only if for all  $p$  and  $q$ :

$$Hp \ \& \ (p \geq' q) \rightarrow Hq.$$

Intuitively, we expect the negation ‘not good’ of the positive predicate ‘good’ to be negative. Indeed, this can easily be shown to be a general pattern that holds for all positive and negative predicates.

OBSERVATION 64 (Hansson, 1990). A monadic predicate  $H$  satisfies  $\geq'$ -positivity if and only if its negation  $\neg H$  satisfies  $\geq'$ -negativity.

#### Proof.

*Left-to-right:* Let  $H$  be a  $\geq'$ -positive predicate. Suppose that  $\neg H$  does not satisfy  $\geq'$ -negativity. Then there are relata  $p$  and  $q$  such that  $\neg Hp$ ,  $p \geq' q$ , and  $\neg(\neg Hq)$ . Hence,  $Hq$ ,  $p \geq' q$ , and  $\neg Hp$ , contrary to the positivity of  $H$ .

*Right-to-left:* Let  $\neg H$  be a  $\geq'$ -negative predicate. Suppose that  $H$  does not satisfy  $\geq'$ -positivity. Then there are relata  $p$  and  $q$  such that  $Hp, q \geq' p$ , and  $\neg(Hq)$ . Hence,  $\neg Hq, q \geq' p$ , and  $\neg(\neg Hp)$ , contrary to the negativity of  $\neg H$ . ■

An important class of positive predicates are those that represent ‘best’. They are mirrored at the other end of the value-scale by negative predicates that represent ‘worst’:

DEFINITION 65. Let  $\geq'$  be a combinative preference relation. The following are monadic predicates defined from  $\geq'$ :

$$\begin{aligned} Hp &\leftrightarrow (\forall q)(p \geq' q) \text{ (strongly best)} \\ Hp &\leftrightarrow \neg(\exists q)(q >' p) \text{ (weakly best)} \\ Hp &\leftrightarrow (\forall q)(q \geq' p) \text{ (strongly worst)} \\ Hp &\leftrightarrow \neg(\exists q)(p >' q) \text{ (weakly worst)} \end{aligned}$$

The first two of these definitions correspond to the notions of strong and weak eligibility, that were introduced in Subsection 2.6. The first of them also corresponds to the best choice connection discussed in Subsection 3.2.

## 5.2 Good and bad: definitions

Definitions of ‘good’ and ‘bad’ in terms of a preference relation are a fairly common theme in the value-logical literature. There are two major traditions. One of these may be called *indifference-related* since it bases the definitions of ‘good’ and ‘bad’ on a set of indifferent or neutral propositions. Goodness is predicated of everything that is better than something neutral, and badness of everything that is worse than something neutral.

This construction requires a sentence that represents neutral value. Such a sentence can of course be introduced as a primitive notion, but it would be more interesting to identify it among the sentences already available.

Some authors have made use of tautologies or contradictions as neutral propositions. Tautologies have been used for this purpose by Danielsson [1968, p. 37] and contradictions by von Wright [1972, p. 164]. However, it is far from clear how something contingent can be compared in terms of value to a tautology or a contradiction. It would be more intuitively appealing to have neutral sentences that represent contingent states of affairs. Such an approach was proposed by Chisholm and Sosa. According to these authors, a state of affairs is indifferent if and only if it is neither better nor worse than its negation. Then ‘a state of affairs is *good* provided it is better than some state of affairs that is indifferent, and... a state of affairs is *bad* provided some state of affairs that is indifferent is better than it’ [Chisholm and Sosa, 1966, p. 246]. (These authors distinguish between indifference and neutrality. To be neutral means, in their terminology, to be equal in

value to something that is indifferent.) The definitions of ‘good’ and ‘bad’ proposed by Chisholm and Sosa can be introduced into the present formal framework as follows:

DEFINITION 66.

$$\begin{aligned} G_I p &\leftrightarrow (\exists q)(p >' q \equiv' \neg q) \text{ (indifference-related good)} \\ B_I p &\leftrightarrow (\exists q)(\neg q \equiv' q >' p) \text{ (indifference-related bad)} \end{aligned}$$

For the definitions of  $G_I$  and  $B_I$  to be at all useful, there should be at least one indifferent element, i.e., at least one  $q$  such that  $q \equiv' \neg q$ . Furthermore, it can be required that all indifferent elements should be interchangeable in comparisons. This amounts to the following requirement on the preference relation:

DEFINITION 67 (Hansson, 1990).  $\geq'$  satisfies *calibration* if and only if:

- (1) There is some  $q$  such that  $q \equiv' \neg q$ , and
- (2) If  $q \equiv' \neg q$  and  $s \equiv' \neg s$ , then for all  $p$ :  $p \geq' q \leftrightarrow p \geq' s$  and  $q \geq' p \leftrightarrow s \geq' p$ .

The other major approach to defining ‘good’ and ‘bad’ has no need for neutral propositions. According to this definition, ‘good’ means ‘better than its negation’ and ‘bad’ means ‘worse than its negation’. The first clear statement of this idea seems to be due to Brogan [1919]. It has been accepted by many other authors [Mitchell, 1950, pp. 103–105; Halldén, 1957, p. 109; von Wright, 1963, p. 34; von Wright, 1972, p. 162; Åqvist, 1968]. We can express it in the present framework as follows:

DEFINITION 68.

$$\begin{aligned} G_N p &\leftrightarrow p >' \neg p \text{ (negation-related good)} \\ B_N p &\leftrightarrow \neg p >' p \text{ (negation-related bad)} \end{aligned}$$

This definition has a strong intuitive appeal, but unfortunately  $G_N$  and  $B_N$  do not always satisfy positivity, respectively negativity. For an example, let  $q \equiv' \neg q \geq' p >' \neg p$ . Then  $G_N p$ ,  $q \geq' p$  and  $\neg G_N q$ , contrary to positivity. In order to avoid this deficiency, a modified version of the negation-related definition has been proposed.

DEFINITION 69 (Hansson, 1990).

$$\begin{aligned} G_C p &\leftrightarrow (\forall q)(q \geq'^* p \rightarrow q >' \neg q) \text{ (canonical good)} \\ B_C p &\leftrightarrow (\forall q)(p \geq'^* q \rightarrow \neg q >' q) \text{ (canonical bad)} \end{aligned}$$

It is easy to show that  $G_C$  satisfies  $\geq'$ -positivity and  $B_C$   $\geq'$ -negativity. Since the positivity of ‘good’ and the negativity of ‘bad’ are indispensable properties of these predicates,  $G_N$  and  $B_N$  can be plausible formalizations

of ‘good’ and ‘bad’ only if  $\geq'$  is such they satisfy positivity, respectively negativity. It turns out that this is so exactly when  $G_N$  coincides with  $G_C$  and  $B_N$  with  $B_C$ .

**OBSERVATION 70** (Hansson, 1990). Let  $\geq'$  satisfy ancestral reflexivity ( $p \geq'^* p$ ). Then  $G_N$  coincides with  $G_C$  and  $B_N$  with  $B_C$  if and only if  $G_N$  satisfies positivity and  $B_N$  satisfies negativity.

**Proof.** For one direction, note that if  $G_N$  and  $B_N$  do not satisfy positivity respectively negativity, then they cannot be identical with  $G_C$  and  $B_C$  that satisfy these conditions.

For the other direction, let  $G_N$  and  $B_N$  satisfy positivity and negativity. It follows from Part 2 of Theorem 73 (to be proved in the next subsection) that  $G_N p \rightarrow G_C p$  and  $B_N p \rightarrow B_C p$ . It follows directly from Definitions 68 and 69 that  $G_C p \rightarrow G_N p$  and  $B_C p \rightarrow B_N p$ . ■

Hence,  $G_C$  and  $B_C$  may be seen as extentions of  $G_N$  and  $B_N$  that coincide with the latter in all cases when the latter provide a reasonable account of ‘good’ and ‘bad’.

### 5.3 Good and bad: The axiomatic approach

Another approach to defining ‘good’ and ‘bad’ is to identify a set of reasonable axioms that a pair of predicates representing these notions should satisfy. The following are such axioms:

**DEFINITION 71** (Hansson, 1990). Let  $\langle G, B \rangle$  be a pair of monadic predicates.

- (1) It satisfies *positivity – negativity* (PN) with respect to  $\geq'$  if and only if  $G$  satisfies  $\geq'$ -positivity and  $B$  satisfies  $\geq'$ -negativity.
- (2) It satisfies *negation-comparability* (NC) with respect to  $\geq'$  if and only if, for all  $p$ :

$$Gp \rightarrow (p \geq' \neg p) \vee (\neg p \geq' p)$$

$$Bp \rightarrow (p \geq' \neg p) \vee (\neg p \geq' p)$$

- (3) It satisfies *mutual exclusiveness* (ME) if and only if for all  $p$ :  $\neg(Gp \ \& \ Bp)$ .
- (4) It satisfies *non-duplication* (ND) if and only if for all  $p$ :  $\neg(Gp \ \& \ G\neg p)$  and  $\neg(Bp \ \& \ B\neg p)$ .
- (5) It satisfies *closeness* if and only if for all  $p$  and  $q$ ,  $p >' q \rightarrow Gp \vee Bq$ .

These postulates are fairly self-explanatory. Perhaps it should be mentioned that NC can be seen as a (much) weakened form of completeness. In favour

of this postulate it can be argued that a sentence that is not comparable to its negation is deficient in determinate value information. Therefore, predicates such as ‘good’ and ‘bad’ are not applicable to such states of affairs.

According to closeness, ‘good’ and ‘bad’ come so close to each other that they only have ‘neutral’ values between them. One way to express this is that ‘if two things are of unequal value, then at least one of them must be good or at least one of them bad’ [von Wright, 1972, p. 161].

As the following observation shows, ME is redundant in the presence of three of the other postulates:

**OBSERVATION 72.** If  $\langle G, B \rangle$  satisfies PN, ND, and NC, then it also satisfies ME.

**Proof.** Suppose to the contrary that  $\langle G, B \rangle$  satisfies PN, ND, and NC, but not ME. Then, since ME does not hold, there is some  $p$  such that  $Gp \& Bp$ . It follows from NC that  $(p \geq' \neg p) \vee (\neg p \geq' p)$ . From PN follows  $(\neg p \geq' p) \& Gp \rightarrow G\neg p$  and  $(p \geq' \neg p) \& Bp \rightarrow B\neg p$ . By sentential logic,  $(Gp \& G\neg p) \vee (Bp \& B\neg p)$ , contrary to ND. This contradiction completes the proof. ■

We have already seen that  $\langle G_N, B_N \rangle$  does not always satisfy PN. It is easy to check that it satisfies ND, NC, and ME.  $\langle G_C, B_C \rangle$  satisfies all these postulates, and it can also be shown to be maximal among the predicate pairs that satisfy them.

**THEOREM 73** (Hansson, 1990). *Let  $\geq'$  be a relation that satisfies ancestral reflexivity ( $p \geq'^* p$ ). Let  $\langle G_C, B_C \rangle$  be as in Definition 69. Then:*

- (1)  $\langle G_C, B_C \rangle$  satisfies PN, ND and NC.
- (2) Let  $\langle G, B \rangle$  be a pair of monadic predicates that satisfies PN, ND and NC. Then for all  $p$ :

$$Gp \rightarrow G_C p \text{ and } Bp \rightarrow B_C p.$$

- (3) If there is a pair  $\langle G, B \rangle$  of predicates that satisfies PN, ND, NC, and closeness, then  $\langle G_C, B_C \rangle$  satisfies (PN, ND, NC, and) closeness.

**Proof.**

*Part 1:* That PN holds follows directly from Definition 69. To see that ND is satisfied, suppose to the contrary that  $G_C p$  and  $G_C \neg p$ . Due to ancestral reflexivity,  $p \geq'^* p$ , and since  $G_C p$ , Definition 69 yields  $p >' \neg p$ . In the same way it follows from  $G_C \neg p$  that  $\neg p >' p$ . This contradiction is sufficient to ensure that  $\neg(G_C p \& G_C \neg p)$ . The proof that  $\neg(B_C p \& B_C \neg p)$  is similar.

To see that NC is satisfied, note that due to ancestral reflexivity,  $G_C p$  implies  $p >' \neg p$  and  $B_C p$  implies  $\neg p >' p$ .

*Part 2:* Let  $Gp$  and  $q \geq'^* p$ . Then there is a series of sentences  $s_0, \dots, s_n$ , such that  $s_0 \leftrightarrow p$ ,  $s_n \leftrightarrow q$  and for all integers  $k$ , if  $0 \leq k < n$ , then  $s_{k+1} \geq' s_k$ . Clearly,  $Gs_0$ . From  $Gs_k$  and  $s_{k+1} \geq' s_k$  it follows by PN that  $Gs_{k+1}$ . Thus, by induction,  $Gs_n$ , i.e.  $Gq$ .

From  $Gq$  it follows by NC that  $(q >' \neg q) \vee (\neg q \geq' q)$ . Suppose that  $\neg q \geq' q$ . Then by PN follows  $G\neg q$ , so that  $Gq \ \& \ G\neg q$ , contrary to ND. It follows that  $q >' \neg q$ .

Thus, if  $Gp$ , then for all  $q$ , if  $q \geq'^* p$ , then  $q >' \neg q$ . The corresponding property for  $Bp$  can be proved in the same way.

*Part 3:* Let  $\langle G, B \rangle$  satisfy PN, ND, NC, and closeness. Due to part 1, it remains to show that  $\langle G_C, B_C \rangle$  satisfies closeness. Let  $p >' q$ . Since  $\langle G, B \rangle$  satisfies closeness, we have  $Gp \vee Bp$ , and by part (2) of the present theorem we have  $G_C p \vee B_C p$ . ■

The indifference-related approach fares worse with respect to the postulates.

**OBSERVATION 74.** Let  $\geq'$  be a relation that satisfies ancestral reflexivity ( $p \geq'^* p$ ). Let  $\langle G_I, B_I \rangle$  be as in Definition 66. Then:

- (1) If  $\geq'$  satisfies calibration, then  $\langle G_I, B_I \rangle$  satisfies ME.
- (2) If  $\geq'$  satisfies transitivity, then  $\langle G_I, B_I \rangle$  satisfies PN.
- (3) If  $\geq'$  satisfies completeness, then  $\langle G_I, B_I \rangle$  satisfies NC.
- (4) ND does not follow even if calibration, transitivity, and completeness are all satisfied.

**Proof.**

*Part 1:* Let ME be violated. Then there is some  $p$  such that  $G_I p$  and  $B_I p$ , i.e. there are  $q$  and  $r$  such that  $p >' q \equiv' \neg q$  and  $\neg s \equiv' s >' p$ . It follows from  $p >' q \equiv' \neg q$  and  $s \equiv' \neg s$ , due to calibration, that  $p >' s$ . Contradiction.

*Part 2:* For the positivity of  $G_I$ , let  $G_I p$  and  $q \geq' p$ . Then there is some  $s$  such that  $p >' s \equiv' \neg s$ . Transitivity yields  $q >' s$ , hence  $G_I q$ . The negativity of  $B_I$  is proved in the same way.

*Part 3:* Directly from the definition of NC.

*Part 4:* Let  $\geq'$  be transitive and complete, and such that  $p >' \neg p >' s \equiv' \neg s$  and that calibration is satisfied. Then  $Gp \ \& \ G\neg p$ . ■

ND is an essential property of 'good' and 'bad', and  $\langle G_I, B_I \rangle$  can hardly be a satisfactory account of these concepts unless this property holds. Fairly strong additional conditions are needed to ensure that it holds [Hansson 1990]. In the light of this, the axiomatic analysis is much more favourable to  $\langle G_N, B_N \rangle$  and, in particular to its generalization  $\langle G_C, B_C \rangle$ , than to  $\langle G_I, B_I \rangle$ .



#### 5.4 *Some other value predicates*

Common language contains many value predicates in addition to ‘best’, ‘worst’, ‘good’, and ‘bad’, as defined above. For a couple of these, precise definitions have been proposed:

very good = good among those that are good [Wheeler, 1972]  
 very bad = bad among those that are bad  
 fairly good = good but not very good [Wheeler, 1972]  
 fairly good = good among those that are not very good [Klein, 1980, pp. 24–25]  
 almost worst = very bad but not worst [Hansson, 1998b]

The last three of these are neither positive nor negative predicates, but belong to a third category of predicates, namely those that are, intuitively speaking, bounded both upwards and downwards. From a formal point of view, they can be defined as the meets of one positive and one negative predicate. Thus, as indicated above, ‘ $p$  is almost worst’ may be defined as ‘ $p$  is very bad and  $p$  is not worst’, employing the negative predicate ‘very bad’ and the positive predicate ‘not worst’.

DEFINITION 75 (Hansson, 1998b). A monadic predicate  $H$  is  $\geq'$ -circumscriptive if and only if there is a  $\geq'$ -positive predicate  $H^+$  and a  $\geq'$ -negative predicate  $H^-$  such that for all  $p$ :

$$Hp \leftrightarrow H^+p \ \& \ H^-p.$$

A  $\geq'$ -circumscriptive predicate is *properly*  $\geq'$ -circumscriptive if and only if it is neither  $\geq'$ -positive nor  $\geq'$ -negative.

#### 5.5 *Deontic concepts*

It is generally recognized that there are three major groups of normative expressions in ordinary language, namely prescriptive, prohibitive, and permissive expressions. In the formal language, they are represented by the corresponding three types of predicates. Here, prescriptive predicates will be denoted by ‘ $O$ ’, permissive predicates by ‘ $P$ ’, and prohibitive predicates by ‘ $W$ ’. (These are abbreviations of ‘ought’, ‘permitted’, and ‘wrong’.) The arguments of these predicates are in general taken to be sentences that represent states of affairs or actions. The three categories of predicates are also generally taken to be interdefinable:  $Oq$  holds if and only if  $W\neg q$ , and it also holds if and only if  $\neg P\neg q$ .

Modern deontic logic began with a seminal paper by Georg Henrik von Wright in 1951 [von Wright, 1951]. (On the origins of deontic logic, see also [Føllesdal and Hilpinen, 1970; von Wright, 1998].) The literature in this area is at least as extensive as that on preference logic. The purpose of this subsection is not to give an overview of this vast subject, but only

to point out two alternative ways in which deontic logic can be connected with preference logic. (For an overview of deontic logic, see [Åqvist, 1987].)

The first of these is the standard semantical construction that dominates the subject [Føllesdal and Hilpinen, 1970]. It is assumed that there is a subset of the set of possible worlds (the ‘ideal worlds’) such that for any sentence  $p$ ,  $Op$  holds if and only if  $p$  holds in all these worlds. Although there is some leeway in the meaning of the term *standard deontic logic* (SDL), the following definition seems to capture the gist of the matter:

**DEFINITION 76.** A model  $\langle \mathcal{A}, \mathcal{I} \rangle$  for *non-iterative standard deontic logic* (non-iterative SDL) consists of a set  $\mathcal{A}$  of possible worlds and a non-empty subset  $\mathcal{I}$  of  $\mathcal{A}$ .

A *non-iterative deontic sentence* in  $\langle \mathcal{A}, \mathcal{I} \rangle$  is a truth-functional combination of sentences of the form  $O\alpha$ , with  $\alpha \in \mathcal{L}_{\mathcal{A}}$ . Such a sentence is *true* in  $\langle \mathcal{A}, \mathcal{I} \rangle$  if and only if it follows by classical truth-functional logic from the set  $\{O\alpha \mid \alpha \in \cap \mathcal{I}\} \cup \{\neg O\alpha \mid \alpha \notin \cap \mathcal{I}\}$ . It is *valid* if and only if it is true in all models.

No explicit preference relation is involved here, but  $\mathcal{I}$  can be interpreted as consisting of the best alternatives according to some preference relation. An explicit preference relation is used in corresponding accounts of conditional obligation. A sentence such as ‘If you borrow his lawn-mower then you ought to return it’ is held to be true if and only if you return the lawn-mower in all those worlds that are best among the worlds in which you borrow the lawn-mower in question.

The valid sentences of non-iterative SDL coincide with the theorems that are derivable from the following three axioms [Føllesdal and Hilpinen, 1970]:

$$\begin{aligned} Op &\rightarrow \neg O\neg p, \\ Op \ \& \ Oq &\leftrightarrow O(p\&q), \text{ and} \\ O(p \vee \neg p) &. \end{aligned}$$

The term ‘non-iterative’ in Definition 76 refers to the fact that sentences containing iterations of the deontic predicate (such as  $OOp$  and  $\neg O(Op \vee Oq)$ ) have been excluded. To cover them, modal semantics (with an accessibility relation) can be used [Føllesdal and Hilpinen, 1970, pp. 15–19].

Unfortunately, it is an immediate consequence of the basic semantic idea of SDL—that of identifying obligatory status with presence in all elements of a certain subset of the alternative set—that the following property will hold:

$$\text{If } \vdash p \rightarrow q, \text{ then } \vdash Op \rightarrow Oq$$

This property may be called *necessitation* since it says that whatever is necessitated by a moral requirement is itself a moral requirement. (It has also been called ‘the inheritance principle’ [Vermazen, 1977, p. 14], ‘Becker’s law’

[McArthur, 1981, p. 149], ‘transmission’ [Routley and Plumwood, 1984], ‘the consequence principle’ [Hilpinen, 1985, p. 191], and ‘entailment’ [Jackson, 1985, p. 178].) As an example, suppose that I am morally required to take a boat without the consent of its owner and use it to rescue a drowning person. Let  $p$  denote this composite action that I am required to perform, and let  $q$  denote the part of it that consists in taking the boat without leave. Since  $q$  follows logically from  $p$ , I am logically necessitated to perform  $q$  in order to perform  $p$ . According to the postulate of necessitation, I then also have an obligation to  $q$ . This is contestable, since it can be argued that I have no obligation to  $q$  in isolation.

Necessitation is the source of all the major deontic paradoxes. We may call them the *necessitation paradoxes*. Four of the most prominent are Ross’s paradox, the paradox of commitment, the Good Samaritan, and the Knower. Ross’s paradox is based on the instance  $Op \rightarrow O(p \vee q)$  of necessitation. (‘If you ought to mail the letter, then you ought to either mail or burn it.’) [Ross, 1941, p. 62] The paradox of commitment is based on the instance  $O\neg p \rightarrow O(p \rightarrow q)$ , which is interpreted as saying that if you do what is forbidden, then you are required to do anything whatsoever. (‘If it is forbidden for you to steal this car, then if you steal it you ought to run over a pedestrian’) [Prior, 1954]. The Good Samaritan operates on two sentences  $p$  and  $q$ , such that  $q$  denotes some atrocity and  $p$  some good act that can only take place if  $q$  has taken place. We then have  $\vdash p \rightarrow q$ , and it follows by necessitation that if  $Op$  then  $Oq$ . (‘You ought to help the assaulted person. Therefore, there ought to be an assaulted person’) [Prior, 1958, p. 144]. Åqvist’s Knower paradox makes use of the epistemic principle that only that which is true can be known. Here,  $q$  denotes some wrongful action, and  $p$  denotes that  $q$  is known by someone who is required to know it. Again, we have  $\vdash p \rightarrow q$  and  $Op$ , and it follows by necessitation that  $Oq$ . (‘If the police officer ought to know that Smith robbed Jones, then Smith ought to rob Jones’) [Åqvist, 1967].

A quite different approach, introduced in [Hansson, 1993c] and further developed in [Hansson, 1997b; Hansson, 1998b; Hansson, 1999b] is based on the assumption that prescriptive predicates (ought-predicates) should satisfy the following property:

**DEFINITION 77.** A (monadic) predicate  $H$  is *contranegative* with respect to a given relation  $\geq'$  if and only if the following holds for all  $p$  and  $q$ :

$$Hp \ \& \ (\neg p \geq' \neg q) \rightarrow Hq.$$

**OBSERVATION 78.** Let  $O$ ,  $P$ , and  $W$  be predicates with a common domain that is closed under negation, and such that for all  $p$ ,  $Op$  if and only if  $\neg P\neg p$ , and  $Op$  if and only if  $W\neg p$ . Let  $\geq'$  be a relation over this domain. Then the following three conditions are equivalent:

- (1)  $O$  satisfies  $\geq'$ -contranegativity,

- (2)  $P$  satisfies  $\geq'$ -positivity, and
- (3)  $W$  satisfies  $\geq'$ -negativity.

**Proof.** Left to the reader. ■

Since both  $\geq'$ -positivity of  $P$  and  $\geq'$ -negativity of  $W$  are reasonable properties, we have good reasons to accept the equivalent requirement that  $\geq'$  be contranegative. At first sight, one might also wish to require that  $O$  be  $\geq'$ -positive, but it is easy to show with examples that this is not a plausible property. For instance, let  $q$  denote that you give your hungry visitor something to eat and  $p$  that you serve her a gourmet meal. It is quite plausible to claim both that  $p$  is better than  $q$  and that  $q$  is morally required whereas  $p$  is not.

In a deontic logic based on contranegativity of  $O$ , the logical properties of  $O$  will depend on those of the underlying preference relation. The more implausible properties of SDL turn out to correspond to rather implausible properties of the preference relation. In particular, this applies to necessitation.

**OBSERVATION 79.** Let  $\mathcal{A}$  be a set of contextually complete alternatives. The following are two conditions on a relation  $\geq'$  in  $\mathcal{L}_{\mathcal{A}}$ :

- (1) If  $\models_{\mathcal{A}} q \rightarrow p$ , then  $p \geq'^* q$ .
- (2) Every  $\geq'$ -contranegative predicate  $O$  on  $\mathcal{L}_{\mathcal{A}}$  satisfies necessitation (If  $\models_{\mathcal{A}} p \rightarrow q$ , then  $Op \rightarrow Oq$ .)

If (1) holds, then so does (2). If  $\geq'$  satisfies ancestral reflexivity ( $p \geq'^* p$ ) then (1) and (2) are equivalent.

**Proof.**

*From (1) to (2):* Let (1) hold. Let  $O$  be a predicate that is contranegative with respect to  $\geq'$ , and such that  $\models_{\mathcal{A}} p \rightarrow q$  and  $Op$ . Then, equivalently:  $\models_{\mathcal{A}} \neg q \rightarrow \neg p$  and  $Op$ . It follows from (1) that  $\neg p \geq'^* \neg q$  and from the contranegativity of  $O$  that  $Oq$ .

*From ancestral reflexivity and (2) to (1):* We are going to assume that ancestral reflexivity holds, but (1) does not hold, and prove that then (2) is violated. Since (1) is not satisfied there are  $p$  and  $q$  such that  $\models_{\mathcal{A}} q \rightarrow p$  and  $\neg(p \geq'^* q)$ .

Let  $W$  be the predicate such that for all  $r \in \mathcal{L}_{\mathcal{A}}$ ,  $Wr$  holds if and only if  $p \geq'^* r$ . Then  $W$  is  $\geq'$ -negative. Since  $\geq'$  satisfies ancestral reflexivity, we have  $p \geq'^* p$  and thus  $Wp$ . It follows from  $\neg(p \geq'^* q)$  that  $\neg Wq$ . We therefore have  $\models_{\mathcal{A}} q \rightarrow p$ ,  $Wp$ , and  $\neg Wq$ , or equivalently for the corresponding  $\geq'$ -contranegative predicate  $O$ :  $\models_{\mathcal{A}} \neg p \rightarrow \neg q$ ,  $O(\neg p)$ , and  $\neg O(\neg q)$ . This is sufficient to show that (2) is violated. ■

On the other hand, some of the more plausible properties of deontic predicates turn out to correspond to more plausible preference postulates [Hansson, 1997b; Hansson, 1998b]. The following postulate was proposed by von Wright [1972, p. 44].

$$P(p \& q) \ \& \ P(p \& \neg q) \ \rightarrow \ Pp$$

It has been called *permissive cancellation* since it allows for the cancellation of  $q$  and  $\neg q$  from the two permissions [Hansson, 1998b]. As the following observation shows, permissive cancellation holds for a wide range of contranegative predicates. (Note that a permissive predicate  $P$  is  $\geq'$ -positive if and only if the corresponding prescriptive predicate  $O$  is  $\geq'$ -contranegative.)

**OBSERVATION 80.** Let  $\mathcal{A}$  be a set of contextually complete alternatives. The following are two conditions on a relation  $\geq'$  in  $\mathcal{L}_{\mathcal{A}}$ :

- (1)  $(p \geq'^* (p \& q)) \vee (p \geq'^* (p \& \neg q))$
- (2) Every  $\geq'$ -positive predicate  $P$  on  $\mathcal{L}_{\mathcal{A}}$  satisfies permissive cancellation  $(P(p \& q) \ \& \ P(p \& \neg q) \ \rightarrow \ Pp)$ .

If (1) holds, then so does (2). Furthermore, if  $\geq'$  satisfies completeness, then (1) and (2) are equivalent.

**Proof.**

*From (1) to (2):* If  $p \geq'^* (p \& q)$ , then we can use  $P(p \& q)$  and the positivity of  $P$  to obtain  $Pp$ . If  $p \geq'^* (p \& \neg q)$ , then we can use  $P(p \& \neg q)$  and the positivity of  $P$  to obtain  $Pp$ .

*From (2) and completeness to (1):* Let  $\geq'$  satisfy completeness. We are going to assume that (1) does not hold, and prove that then neither does (2). Since (1) does not hold, there are  $p$  and  $q$  such that  $\neg(p \geq'^* (p \& q))$  and  $\neg(p \geq'^* (p \& \neg q))$ . Due to completeness, there are two cases.

Case *i*,  $(p \& q) \geq' (p \& \neg q)$ : Let  $P$  be the predicate such that for all  $r$ ,  $Pr$  iff  $r \geq' (p \& \neg q)$ . Then  $P$  is  $\geq'$ -positive, and it follows directly that  $P(p \& q)$  and  $P(p \& \neg q)$ . It follows from  $\neg(p \geq'^* (p \& \neg q))$  that  $\neg Pp$ .

Case *ii*,  $(p \& \neg q) \geq' (p \& q)$ : The proof proceeds in the same way as in case *i*. ■

The condition  $(p \geq'^* (p \& q)) \vee (p \geq'^* (p \& \neg q))$  used in the observation follows from completeness and disjunctive interpolation:

$$\begin{aligned} & ((p \& \neg q) \geq' (p \& q)) \vee ((p \& q) \geq' (p \& \neg q)) && \text{(completeness)} \\ & (((p \& q) \vee (p \& \neg q)) \geq' (p \& q)) \vee (((p \& q) \vee (p \& \neg q)) \geq' (p \& \neg q)) \\ & && \text{(disjunctive interpolation)} \\ & (p \geq' (p \& q)) \vee (p \geq' (p \& \neg q)) && \text{(intersubstitutivity)} \\ & (p \geq'^* (p \& q)) \vee (p \geq'^* (p \& \neg q)) && \text{(definition of ancestral)} \end{aligned}$$

More details on contranegative logic can be found in [Hansson, 1998b; Hansson, 1999b]. This is probably only one of many examples of how new applications of preference logic can lead to new insights in other branches of philosophical logic.

*Philosophy Unit, Royal Institute of Technology, Stockholm, Sweden.*

## BIBLIOGRAPHY

- [Abbas, 1995] M. Abbas. Any complete preference structure without circuit admits an interval representation, *Theory and Decision*, **39**, 115–126, 1995.
- [Anand, 1993] P. Anand. The philosophy of intransitive preference, *Economic Journal*, **103**, 337–346, 1993.
- [Åqvist, 1967] L. Åqvist. Good samaritans, contrary-to-duty imperatives, and epistemic obligations, *Noûs*, **1**, 361–379, 1967.
- [Åqvist, 1968] L. Åqvist. Chisholm–Sosa logics of intrinsic betterness and value, *Noûs*, **2**, 253–270, 1968.
- [Åqvist, 1987] L. Åqvist. *Introduction to Deontic Logic and the Theory of Normative Systems*, Bibliopolis, Napoli, 1987.
- [Armstrong, 1939] W. E. Armstrong. The determinateness of the utility function, *Economic Journal*, **49**, 453–467, 1939.
- [Armstrong, 1948] W. E. Armstrong. Uncertainty and the utility function, *Economic Journal*, **58**, 1–10, 1948.
- [Barbera *et al.*, 1984] S. Barbera, C. R. Barrett and P. K. Pattanaik. On some axioms for ranking sets of alternatives, *Journal of Economic Theory*, **33**, 301–308, 1984.
- [Beck, 1941] L. W. Beck. The formal properties of ethical wholes, *Journal of Philosophy*, **38**, 5–14, 1941.
- [Brogan, 1919] A. P. Brogan. The fundamental value universal, *Journal of Philosophy, Psychology, and Scientific Methods*, **16**, 96–104, 1919.
- [Burros, 1976] R. H. Burros. Complementary relations in the theory of preference, *Theory and Decision*, **7**, 181–190, 1976.
- [Carlson, 1997] E. Carlson. The intrinsic value of non-basic states of affairs, *Philosophical Studies*, **85**, 95–107, 1997.
- [Castañeda, 1958] H. N. Castañeda. Review of Halldén, ‘On the Logic of “Better”’, *Philosophy and Phenomenological Research*, **19**, 266, 1958.
- [Chernoff, 1954] H. Chernoff. Rational selection of decision functions, *Econometrica*, **22**, 422–443, 1954.
- [Chisholm and Sosa, 1966] R. M. Chisholm and E. Sosa. On the logic of ‘intrinsically better’, *American Philosophical Quarterly*, **3**, 244–249, 1966.
- [Cresswell, 1971] M. J. Cresswell. A semantics for a logic of ‘better’, *Logique et Analyse*, **14**, 775–782, 1971.
- [Danielsson, 1968] S. Danielsson. *Preference and Obligation*, Filosofiska Föreningen, Uppsala, Sweden, 1968.
- [Danielsson, 1997] S. Danielsson. Harman’s equation and the additivity of intrinsic value. In *For Good Measure. Philosophical Essays Dedicated to Jan Odelstad on the Occasion of His Fiftieth Birthday*, L. Lindahl, P. Needham and R. Sliwinski, eds. pp. 23–24. Uppsala Philosophical Studies 46, Uppsala, 1997.
- [Danielsson, 1998] S. Danielsson. Numerical representations of value-orderings: some basic problems. In *Preferences*, C. Fehige and U. Wessels, eds. pp. 114–122. Walter de Gruyter, Berlin, 1998.
- [Dummett, 1984] M. Dummett. *Voting Procedures*, Clarendon Press, Oxford, 1984.
- [Fehige and Wessels, 1998] C. Fehige and U. Wessels. *Preferences*, Walter de Gruyter, Berlin, 1998.
- [Fishburn, 1970a] P. C. Fishburn. Intransitive indifference with unequal indifference intervals, *Journal of Mathematical Psychology*, **7**, 144–149, 1970.

- [Fishburn, 1970b] P. C. Fishburn. Intransitive indifference in preference theory: a survey, *Operations Research*, **8**, 207–228, 1970.
- [Fishburn, 1972] P. C. Fishburn. Even-chance lotteries in social choice theory. *Theory and Decision*, **3**, 18–40, 1972.
- [Føllesdal and Hilpinen, 1970] D. Føllesdal and R. Hilpinen. (1970) ” Deontic logic: an introduction. In *Deontic Logic: Introductory and Systematic Readings*, R. Hilpinen, ed. pp. 1–35. Reidel, Dordrecht, 1970.
- [Gärdenfors, 1973] P. Gärdenfors. Positionalist voting functions, *Theory and Decision*, **4**, 1–24, 1973.
- [Halldén, 1957] S. Halldén. *On the Logic of 'Better'*, Lund, 1957.
- [Hansson, 1968] B. Hansson. Fundamental axioms for preference relations, *Synthese*, **18**, 423–442, 1968.
- [Hansson, 1989] S. O. Hansson. A new semantical approach to the logic of preferences, *Erkenntnis*, **31**, 42, 1989.
- [Hansson, 1990] S. O. Hansson. Defining ‘good’ and ‘bad’ in terms of ‘better’, *Notre Dame Journal of Formal Logic*, **31**, 136–149, 1990.
- [Hansson, 1992] S. O. Hansson. Similarity semantics and minimal changes of belief, *Erkenntnis*, **37**, 401–429, 1992.
- [Hansson, 1993a] S. O. Hansson. Money-pumps, self-torturers and the demons of real life, *Australasian Journal of Philosophy*, **71**, 476–485, 1993.
- [Hansson, 1993b] S. O. Hansson. A note on anti-cyclic properties of complete binary relations, *Reports on Mathematical Logic*, **27**, 41–44, 1993.
- [Hansson, 1993c] S. O. Hansson. The false promises of risk analysis, *Ratio*, **6**, 16–26, 1993.
- [Hansson, 1996a] S. O. Hansson. What is ceteris paribus preference?, *Journal of Philosophical Logic*, **25**, 307–332, 1996.
- [Hansson, 1996b] S. O. Hansson. Decision-making under great uncertainty, *Philosophy of the Social Sciences*, **26**, 369–386, 1996.
- [Hansson, 1997a] S. O. Hansson. Decision-theoretic foundations for axioms of rational preference, *Synthese*, **109**, 401–412, 1997.
- [Hansson, 1997b] S. O. Hansson. Situationist deontic logic, *Journal of Philosophical Logic*, **26**, 423–448, 1997.
- [Hansson, 1998a] S. O. Hansson. Should we avoid moral dilemmas?, *Journal of Value Inquiry*, **32**, 407–416, 1998.
- [Hansson, 1998b] S. O. Hansson. Structures of value, Lund Philosophy Reports 1998:1, Lund University.
- [Hansson, 1999a] S. O. Hansson. *A Textbook of Belief Dynamics*, Kluwer, 1999.
- [Hansson, 1999b] S. O. Hansson. Representation theorems for contranegative deontic logic, manuscript, 1999.
- [Harman, 1967] G. Harman. Toward a theory of intrinsic value, *Journal of Philosophy*, **64**, 792–804, 1967.
- [Harrison, 1952] J. Harrison. Utilitarianism, universalisation, and our duty to be just, *Proceedings of the Aristotelian Society*, **53**, 105–134, 1952.
- [Hilpinen, 1985] R. Hilpinen. Normative conflicts and legal reasoning In *Man, Law and Modern Forms of Life*, E. Bulygin *et al.*, eds. pp. 191–208. Reidel, Dordrecht, 1985.
- [Hughes, 1980] R. G. Hughes. Rationality and intransitive preferences, *Analysis*, **40**, 132–134, 1980.
- [Humphreys, 1983] P. Humphreys. Decision aids: aiding decisions. In *Human Decision Making*, L. Sjöberg, T. Tyszka and J. A. Wise, eds. pp. 14–44. Doxa, Bodafors, Sweden, 1983.
- [Jackson, 1985] F. Jackson. On the semantics and logic of obligation, *Mind*, **94**, 177–195, 1985.
- [Jennings, 1967] R. E. Jennings. Preference and choice as logical correlates, *Mind*, **76**, 556–567, 1967.
- [Kirchsteiger and Puppe, 1996] G. Kirchsteiger and C. Puppe. Intransitive choices based on transitive preferences: the case of menu-dependent information, *Theory and Decision*, **41**, 37–58, 1996.

- [Klein, 1980] E. Klein. A semantics for positive and comparative adjectives, *Linguistics and Philosophy*, **4**, 1–45, 1980.
- [Kron and Milovanovic, 1975] A. Kron and V. Milovanovic. Preference and choice, *Theory and Decision*, **6**, 185–196, 1975.
- [Lee, 1984] R. Lee. Preference and transitivity, *Analysis*, **44**, 129–134, 1984.
- [Lehrer and Wagner, 1985] K. Lehrer and C. Wagner. Intransitive indifference: the semi-order problem, *Synthese*, **65**, 249–256, 1985.
- [Lewis, 1973a] D. Lewis. *Counterfactuals*, Harvard University Press, 1973.
- [Lewis, 1973b] D. Lewis. Causation, *Journal of Philosophy*, **70**, 556–567, 1973.
- [Lewis, 1981] D. Lewis. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophy*, **10**, 217–234, 1981.
- [Luce, 1954] R. D. Luce. Semiorders and a theory of utility discrimination, *Econometrica*, **24**, 178–191, 1954.
- [Manders, 1981] K. L. Manders. On JND representation of semiorders, *Journal of Mathematical Psychology*, **24**, 224–248, 1981.
- [McArthur, 1981] R. P. McArthur. Anderson's deontic logic and relevant implication, *Notre Dame Journal of Formal Logic*, **22**, 145–154, 1981.
- [McKelvey, 1976] R. D. McKelvey. Intransitivities in multidimensional voting models and some implications for agenda control, *Journal of Economic Theory*, **12**, 472–482, 1976.
- [McKelvey, 1979] R. D. McKelvey. General conditions for global intransitivities in formal voting models, *Econometrica*, **47**, 1085–1112, 1979.
- [McKelvey and Wendell, 1976] R. D. McKelvey and R. E. Wendell. Voting equilibria in multidimensional choice spaces, *Mathematics of Operations Research*, **1**, 144–158, 1976.
- [Mendola, 1987] J. Mendola. The indeterminacy of options, *American Philosophical Quarterly*, **24**, 125–136, 1987.
- [Mitchell, 1950] E. T. Mitchell. *A System of Ethics*, Charles Scribner's Sons, New York, 1950.
- [Moore, 1903] G. E. Moore. *Principia Ethica*. Cambridge University Press, 1903; reprinted 1951.
- [Moulin, 1985] H. Moulin. Choice functions over a finite set: a summary, *Social Choice and Welfare*, **2**, 147–160, 1985.
- [Ng, 1977] Y. Ng. Sub-semiorder: a model of multidimensional choice with preference intransitivity, *Journal of Mathematical Psychology*, **16**, 51–59, 1977.
- [Nitzan and Prasanta, 1984] S. I. Nitzan and P. Pattanaik. Median-based extensions of an ordering over a set to the power set: an axiomatic characterization, *Journal of Economic Theory*, **34**, 252–261, 1984.
- [Oldfield, 1977] E. Oldfield. An approach to a theory of intrinsic value, *Philosophical Studies*, **32**, 233–249, 1977.
- [Österberg, 1989] J. Österberg. One more turn on the lawn. In *In So Many Words. Philosophical Essays dedicated to Sven Danielsson on the Occasion of his Fiftieth Birthday*, S. Lindström and W. Rabinowicz, eds. pp. 125–133. Uppsala University, Sweden, 1989.
- [Packard, 1979] D. J. Packard. Preference relations, *Journal of Mathematical Psychology*, **19**, 295–306, 1979.
- [Packard, 1987] D. J. Packard. Difference logic for preference, *Theory and Decision*, **22**, 71–76, 1987.
- [Plott, 1967] C. R. Plott. A notion of equilibrium and its possibility under majority rule, *American Economic Review*, **57**, 787–806, 1967.
- [Pollock, 1983] J. L. Pollock. How do you maximize expectation value?, *Noûs*, **17**, 409–421, 1983.
- [Prior, 1954] A. N. Prior. The paradoxes of derived obligation, *Mind*, **63**, 64–65, 1954.
- [Prior, 1958] A. N. Prior. Escapism, In *Essays in Moral Philosophy*, A. I. Melden, ed., pp. 135–146. University of Washington Press, Seattle, 1958.
- [Quinn, 1974] W. S. Quinn. Theories of intrinsic value, *American Philosophical Quarterly*, **11**, 123–132, 1974.
- [Quinn, 1990] W. S. Quinn. The puzzle of the self-torturer, *Philosophical Studies*, **59**, 79–90, 1990.



- [Ramsey, 1931] F. P. Ramsey. *The Foundations of Mathematics and other Logical Essays*, Kegan Paul, Trench, Trubner & Co, London, 1931; reprinted 1950.
- [Rescher, 1967] N. Rescher. Semantic Foundations for the Logic of Preference, pp. 37–62 in Nicholas Rescher (ed.) *The Logic of Decision and Action*. University of Pittsburgh Press, Pittsburgh.
- [Rescher, 1968] N. Rescher. *Topics in Philosophical Logic*. Reidel, Dordrecht, 1968.
- [Restle, 1961] F. Restle. *Psychology of Judgment and Choice*, NY, 1961.
- [Roberts, 1979] F. S. Roberts. Measurement theory, *Encyclopedia of Mathematics and its Applications*, Vol. 7, G.-C. Rota, ed. Addison-Wesley, Reading, MA, 1979.
- [Ross, 1941] A. Ross. Imperatives and logic, *Theoria*, **7**, 53–71, 1941.
- [Rott, 1993] H. Rott. Belief contraction in the context of the general theory of rational choice, *Journal of Symbolic Logic*, **58**, 1426–1450, 1993.
- [Rott, 1999] H. Rott. *Change, Choice and Inference*, Oxford University Press, in press.
- [Routley and Plumwood, 1984] R. Routley and V. Plumwood. Moral dilemmas and the logic of deontic notions, *Discussion Papers in Environmental Philosophy*, number 6, Philosophy Department, Australian National University, 1984.
- [Saito, 1973] S. Saito. Modality and preference relation, *Notre Dame Journal of Formal Logic*, **14**, 387–391, 1973.
- [Savage, 1954] L. J. Savage. *The Foundations of Statistics*. Wiley, New York, 1954.
- [Schoemaker, 1982] P. J. H. Schoemaker. The expected utility model: its variants, purposes, evidence and limitations, *Journal of Economic Literature*, **20**, 529–563, 1982.
- [Schumm, 1987] G. F. Schumm. Transitivity, preference and indifference. *Philosophical Studies*, **52**, 435–437, 1987.
- [Scott and Suppes, 1958] D. Scott and P. Suppes. Foundational aspects of theories of measurement, *Journal of Symbolic Logic*, **23**, 113–128, 1958.
- [Sen, 1969] A. Sen. Quasi-transitivity, rational choice and collective decisions, *Review of Economic Studies*, **35**, 381–393, 1969.
- [Sen, 1971] A. Sen. Choice functions and revealed preference, *Review of Economic Studies*, **38**, 307–317, 1971.
- [Sen, 1973] A. Sen. Behaviour and the concept of preference, London School of Economics, London, 1973.
- [Sen, 1993] A. Sen. Internal consistency of choice, *Econometrica*, **61**, 495–521, 1993.
- [Simon, 1957] H. A. Simon. *Models of Man*, John Wiley & Sons, New York, 1957.
- [Suppes and Zinnes, 1963] P. Suppes and J. L. Zinnes. Basic measurement theory. In *Handbook of Mathematical Psychology*, vol I, R. D. Luce, R. R. Bush and E. Galanter, eds. pp. 1–76. John Wiley and Sons, New York, 1963.
- [Toda, 1976] M. Toda. The decision process: a perspective, *International Journal of General Systems*, **3**, 79–88, 1976.
- [Toda and Shuford, 1965] M. Toda and E. H. Shuford. Utility, induced utilities and small worlds, *Behavioral Science*, **10**, 238–254, 1965.
- [Trapp, 1985] R. W. Trapp. Utility theory and preference logic, *Erkenntnis*, **22**, 301–339, 1985.
- [Tversky, 1969] A. Tversky. Intransitivity of preferences, *Psychological Review*, **76**, 31–48, 1969.
- [Vermazen, 1977] B. Vermazen. The logic of practical ‘ought’-sentences, *Philosophical Studies*, **32**, 1–71, 1977.
- [Wheeler, 1972] S. C. Wheeler. Attributives and their modifiers, *Noûs*, **6**, 310–334, 1972.
- [Williamson, 1988] T. Williamson. First-order logics for comparative similarity, *Notre Dame Journal of Formal Logic*, **29**, 457–481, 1988.
- [van Dalen, 1974] D. van Dalen. Variants of Rescher’s semantics for preference logic and some completeness theorems. *Studia Logica*, **33**, 163–181, 1974.
- [von Kutschera, 1975] F. von Kutschera. Semantic analyses of normative concepts, *Erkenntnis*, **9**, 195–218, 1975.
- [von Wright, 1951] G. H. von Wright. Deontic logic, *Mind*, **60**, 1–15, 1951.
- [von Wright, 1963] G. H. von Wright. *The Logic of Preference*, Edinburgh University Press, Edinburgh, 1963.
- [von Wright, 1972] G. H. von Wright. The logic of preference reconsidered, *Theory and Decision*, **3**, 140–169, 1972.

[von Wright, 1998] G. H. von Wright. Deontic logic—as I see it. Paper presented at the *Fourth International Workshop on Deontic Logic in Computer Science (DEON'98)*, Bologna, 1988.



## DIAGRAMMATIC LOGIC

The many diagrammatic systems in use include Euler circles, Venn diagrams, state diagrams, control-flow diagrams, line graphs, circuit diagrams, category-theory diagrams, Hasse diagrams, and geometry diagrams. A *diagrammatic logic* seeks to describe the syntax, semantics, proof theory, etc., of some such diagrammatic system.

The diagrams of a diagrammatic system have a (typically two-dimensional) syntactic structure that can be described using concepts such as labeling, connectedness, inclusion, direction, etc. They also have a meaning that can be described using techniques from model theory or algebra. Thus, a diagrammatic logic differs from an ordinary logic only in the type of well-formed representations it describes (though these may well have properties not common to more familiar logics).

Diagrams can have unusual properties that distinguish them from expressions of many languages, properties that might motivate the formulation and analysis of a diagrammatic logic. The structure of a diagram might have a close correspondence with what they represent. Its meaning might be invariant under certain topological transformations. It might be unusually easy to understand. A diagrammatic logic *need* illuminate none of these matters (though some of them may be connected to the system's logical properties and hence addressed by the logic). In particular, philosophical and psychological questions about the nature of the diagrammatic system that is the target of a logic could be left to philosophy and psychology.

To reveal the typical characteristics of diagrammatic logics more directly, several examples will be presented. These include Venn diagrams, a variation due to Peirce that will be called *Peirce-Venn diagrams*, and a historically important system developed by Peirce called *existential graphs*. Other diagrammatic logics that have been developed include logics of state transition diagrams,<sup>1</sup> blocks world diagrams,<sup>2</sup> circuit diagrams,<sup>3</sup> conceptual graphs,<sup>4</sup> and geometry diagrams.<sup>5</sup> Relevant collections include Allwein and Barwise [1996] and Glasgow, Narayanan, and Chandrasekaran [1995].

### 1 FOUNDATIONS

Venn diagrams and Peirce-Venn diagrams (covered in the next two sections) are constructed from circles or, more generally, closed curves, that overlap in

---

<sup>1</sup>Harel [1988].

<sup>2</sup>Barwise and Etchemendy [1995].

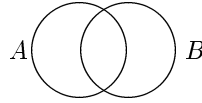
<sup>3</sup>Johnson, Barwise and Allwein [1996].

<sup>4</sup>Sowa [1984].

<sup>5</sup>Luengo [1995].

all combinations. Some simple syntactic and semantic concepts are common to both of these systems and so are handled jointly in this section.

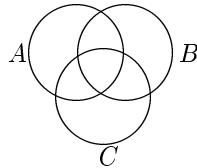
The circles of Venn diagrams represent sets, and the overlapping combinations of the circles represent combinations of the sets. For example, in the case of two circles the four combinations of circles represent the intersection, the two differences, and the complement of the union.



In particular, this diagram consists of four *minimal regions*<sup>6</sup> which can be described by four corresponding combinations of the two labels:

Term	Corresponds to minimal region
$AB$	within both
$A\bar{B}$	within A, not B
$\bar{A}B$	within B, not A
$\bar{A}\bar{B}$	within neither

A term such as  $A\bar{B}$  is said to *correspond* to the minimal region of the diagram within left one circle but outside of the right circle.<sup>7</sup> Likewise,  $\bar{A}\bar{B}$  corresponds to the minimal region outside of both circles,  $AB$  corresponds to the minimal region within both circles, and  $B\bar{A}$  corresponds to the minimal region within the right but not the left circle. A three-circle diagram such as



has eight corresponding terms:

$$ABC \quad \bar{A}BC \quad AB\bar{C} \quad \bar{A}B\bar{C} \quad A\bar{B}C \quad \bar{A}\bar{B}C \quad A\bar{B}\bar{C} \quad \bar{A}\bar{B}\bar{C}$$

The term  $A\bar{B}C$  corresponds to the minimal region within both A and C but outside of B, etc. More generally, with an  $n$ -circle diagram labeled by  $n$

<sup>6</sup>Minimal regions are described in Shin [1994], p. 51.

<sup>7</sup>Correspondence is described in Hammer [1994], pp. 77–78.

letters, there should be a minimal region and a corresponding term for each of the  $2^n$  combinations of circles. One way to think of this is that there should be a term for each row of an  $n$ -variable truth table, the variables of which are the letters labeling the circles, with truth indicating that the region falls within the circle and falsity indicating that it falls outside of the circle.

For the purposes of logic, minimal regions are entirely described by which of the circles they fall within (and hence also which they fall outside of). So any subset of the  $n$  circles should describe a minimal region: that minimal region falling within all the circles in the subset and outside of the rest of the circles of the diagram.

Given  $n$  circles, the following are the conditions desired for a Venn-type diagram:

1. For each of the  $2^n$  terms, there is a minimal region corresponding to it.
2. There is no more than one region corresponding to any term.

The first condition ensures that every Boolean combination of the  $n$  sets is represented in the diagram. The second prevents any redundancy by ensuring that each combination is represented only once.

For logical purposes, these two conditions are really the only desiderata of a (formal or informal) syntax of the circles of a system of Venn-type diagrams. All that is relevant is that there is exactly one minimal region for each term representing each combination of circles.<sup>8</sup>

A *region* of a diagram consists of one or more minimal regions. Hence, a region can be entirely represented as a set of one or more of the terms corresponding to the minimal regions of a diagram.<sup>9</sup> In the case of a two-circle diagram with labels  $A$  and  $B$ , the set  $\{A\bar{B}, \bar{A}B\}$  represents the region outside of the circle labeled by  $B$ .

Since a region consists of any one or more minimal regions, there are as many regions as there are sets of minimal region, minus the empty set. So there are  $2^{(2^n)} - 1$  regions.

If two regions of two diagrams are represented by the same set of terms, they are said to be *counterparts*.<sup>10</sup> Because regions that are counterparts have to be assigned the same set by any model, for convenience below they are sometimes spoken of as though they were the same region. This makes some discussions and proofs easier to read.

---

<sup>8</sup>Formal models of the syntax of overlapping circles have been provided for which these two conditions are satisfied for any finite number of circles, though the concept of circle must be extended to include non-convex closed curves. An example of such a model is presented in More [1959].

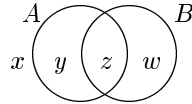
<sup>9</sup>See Shin [1994], p. 51.

<sup>10</sup>The counterpart relation is defined in Shin [1994], pp. 53–57.

A *model* has a domain of discourse which can be an arbitrary set, and assigns subsets of the domain to the circles of the diagrams in question, assigning the same subset to circles labeled by the same letter. For example, a model might assign  $\{x, y\}$  to the domain, assign  $\{x\}$  to one circle of a diagram and  $\{x, y\}$  to the other circle.

A model can also be understood as assigning subsets of the domain to minimal regions. A minimal region such as  $\overline{A}BC\overline{D}E$  would be assigned  $\overline{A} \cap B \cap C \cap \overline{D} \cap E$  (where  $\overline{A}$  is the domain minus the set assigned to the circle labeled by  $A$ ,  $B$  is the set assigned to the circle labeled  $B$ , etc.).<sup>11</sup> Likewise, a region can be understood as being assigned the union of the sets assigned to the minimal regions composing it.

Just as a model determines the sets assigned to minimal regions, conversely, an assignment to minimal regions can be used to specify a model. For example, suppose the four minimal regions of the following diagram are assigned sets  $x$ ,  $y$ ,  $z$ , and  $w$ , as shown:



This specifies the model:

$$\left\{ \begin{array}{l} A = y \cup z \\ B = z \cup w \\ \text{domain} = x \cup y \cup z \cup w \end{array} \right.$$

The two systems, Venn diagrams and Peirce-Venn diagrams, discussed in the next two sections build on the basic diagrams described here by adding additional syntactic devices that can be used to mark various regions and thereby make assertions about the sets they represent.

## 2 VENN DIAGRAMS

This section presents the logical theory of Venn diagrams. Venn diagrams were introduced by John Venn in 1880 for the purpose of clearly representing categorical sentences and syllogistic reasoning.<sup>12</sup> Venn's system is a modification of a previous, incompleting system of Leonhard Euler's developed in 1761.<sup>13</sup>

<sup>11</sup>This definition of model is given in Hammer and Danner [1996]. A similar concept is defined in Shin [1994], pp. 64–68.

<sup>12</sup>See Venn [1880] and Venn [1894].

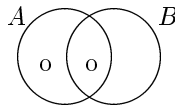
<sup>13</sup>Euler [1846]. For an analysis of Euler's system see Hammer and Shin [1996].

The particular version of Venn diagrams presented here is based on modifications made by Peirce in 1903<sup>14</sup> and Shin in 1994.<sup>15</sup> Peirce provided syntactic rules of inference for manipulating his variation on Venn diagrams while Shin formulated a coherent fragment of Peirce's system and reconstructed and analyzed it in modern form.

Venn diagrams are based on the syntax and semantics developed in the previous section. In addition, the system allows any region of a diagram to be marked as either representing an empty set or a non-empty set (more briefly: to be marked as empty or non-empty).

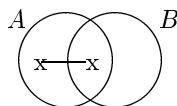
To assert that a region (rather, the set it represents) is empty is simply to assert that each of the minimal regions that make it up is empty. A minimal region is marked as empty by adding the symbol 'o' to it. This is Peirce's notation replacing Venn's shading of the minimal region.

For example, the following diagram asserts that  $A$  is empty (that both  $\overline{AB}$  and  $AB$  are empty):



It is redundant to mark a minimal region with more than one 'o'. If the region is empty it's empty. Therefore well-formed diagrams will be required to have at most one 'o' in each minimal region.

To assert that a region is non-empty (rather, the set it represents) is not the same as asserting that each of the minimal regions composing it is empty. Rather, it is to assert that at least one of them is non-empty. With Venn diagrams, this is done by adding a chain of 'x's connected by lines to the region, with one 'x' falling in each of its minimal regions. For example, the following diagram asserts that  $A$  is non-empty (that either  $\overline{AB}$  or  $AB$  is non-empty):



The region consisting of all the minimal regions with 'x's of the chain is said to *have* the chain. In particular, larger regions will not be said to have a chain falling in some proper subregion of it. For example in the

---

<sup>14</sup>Peirce [1958], pp. 294-319.

<sup>15</sup>Shin [1994].



above diagram the region  $\{\overline{AB}, AB\}$  has the 'x'-chain but the larger region  $\{\overline{AB}, AB, \overline{AB}\}$  does not.

Because it is redundant for any one chain to have more than one 'x' in a minimal region, all chains of a well-formed diagram are required to have no more than one 'x' in each minimal region.

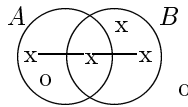
Likewise, because it is redundant to have two chains in the same region, a well-formed diagram is allowed to have no more than one 'x'-chain in each region.

Thus, the well-formed Venn diagrams can be summarized by the following four constructions:

1. Any  $n$  circles drawn to overlap in all combinations as described in the previous section and labeled by  $n$  names is a well-formed Venn diagrams.
2. Given any Venn diagram, the result of adding an 'o' to any minimal region not already containing an 'o' results in a well-formed Venn diagram.
3. Given any Venn diagram, the result of adding an 'x'-chain to any region not already having an 'x'-chain results in a well-formed Venn diagram.
4. Nothing else is a well-formed Venn diagram.

A Venn diagram is *consistent* just in case no minimal region has both an 'o' and an unconnected 'x' in it.

For logical purposes, the syntactic granularity that is relevant for defining diagrams is at the level of which regions have x-chains and which minimal regions have an 'o'. Thus, a diagram is entirely determined by (i) the set of letters labeling the circles, (ii) the minimal regions with an 'o', and (iii) the sets of minimal regions constituting a region with an 'x'-chain. For example, the following diagram is specified by (i) the set  $\{A, B\}$  of letters labeling the circles, (ii) the minimal regions with an 'o'  $\overline{AB}$  and  $\overline{AB}$ , and (iii) the regions with 'x'-chains  $\{\overline{AB}, AB, \overline{AB}\}$  and  $\{\overline{AB}\}$ .<sup>16</sup>



<sup>16</sup>This type of approach to the identity conditions between diagram is presented in Hammer and Danner [1996].

Any other diagram having the same such descriptions is just another instance of the same diagram.

Consider the number of distinct syntactically correct diagrams there are having  $n$  curves and some fixed set of  $n$  predicates. A diagram can have an ‘o’ in any number of its minimal regions, so there are  $2^n$  possibilities for adding o’s to each diagram. A diagram can have a chain of ‘x’s in any number of its regions, so there are  $2^{(2^n)} - 1$  possible chains to consider. This leaves a choice of  $2^n + 2^{(2^n)} - 1$  ‘o’s and chains of ‘x’s to choose from for each diagram. Since a diagram can include any combination of these, there are  $2^{(2^n + 2^{(2^n)} - 1)}$  distinct diagrams possible. In the case of  $n = 1$  there are 16 distinct diagrams possible, with  $n = 2$  there are 524,288 diagrams possible, and with  $n = 3$  and up the number is huge but finite.<sup>17</sup> A more difficult task is that of specifying the precise number of logically distinct diagrams that can be constructed from  $n$  curves and some fixed set of  $n$  labels, that is, the number of equivalence classes (the relation being logical equivalence) of diagrams constructible from the  $n$  curves and labels.

The definition of the conditions under which a model satisfies a Venn diagram are as was intuitively described:

DEFINITION 1 (Satisfies).

1. A model *satisfies* an ‘x’ occurring in some minimal region just in case the set assigned to that minimal region is non-empty.
2. A model *satisfies* an ‘o’ occurring in some minimal region just in case the set assigned to that minimal region is empty.
3. A model *satisfies* an ‘x’-chain occurring in some region just in case the set satisfies at least one ‘x’ in the chain.
4. A model *satisfies* a Venn diagram just in case it satisfies each ‘x’-chain and each ‘o’ in the diagram.<sup>18</sup>

A diagram is a *logical consequence* of a set of diagrams just in case the diagram is satisfied by every model satisfying each diagram in the set. A diagram is *logically equivalent* to another diagram just in case the two are satisfied by the same models.

The following rules of inference govern the manipulation of ‘x’-chains and ‘o’s.

RULE 2 (Addition). An ‘x’-chain can be extended with an additional ‘x’ in a new minimal region.<sup>19</sup>

---

<sup>17</sup>Various calculations of this sort are given in Peirce [1960], pp. 306–307 and analyzed in Hammer [1995b], pp. 811–813.

<sup>18</sup>This definition is given in Hammer [1995b], pp. 817–818.

<sup>19</sup>Peirce [1958], p. 310.

The validity of Addition can be seen from the fact that if a region is assigned a non-empty set any region containing it will be assigned a superset, and hence will be non-empty.

RULE 3 (Contraction). If an 'x'-chain has an 'x' in a region also having an 'o', that 'x' can be erased. If the 'x' does not occur on an end, the two halves of the chain must be reconnected.<sup>20</sup>

The validity of Contraction can be seen from the fact that if a minimal region is empty and some region containing it is non-empty, then some other minimal region of the larger region must be non-empty.

RULE 4 (Simplification). Any 'o' can be erased. Any entire 'x'-chain can be erased.<sup>21</sup>

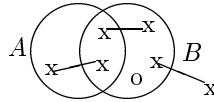
The validity of Simplification can be seen from the fact that the various 'o's and 'x'-chains of a diagram must all be satisfied for a diagram to be satisfied.

RULE 5 (Contradiction). Any diagram can be inferred from a diagram having a minimal region with both an 'o' and an unconnected 'x'.<sup>22</sup>

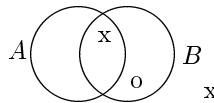
The validity of Contradiction can be seen from the fact that no diagram of this type can be satisfied.

PROPOSITION 6. Addition, Simplification, and Contradiction result in diagrams that are logical consequences of the diagrams they are applied to. Contraction results in a logically equivalent diagram.

For two diagrams having the same labels, logical equivalence can be characterized in terms of the two rules of Contraction and Addition. Define the *summary* of a diagram to be the result of applying Contraction as many times as possible, then erasing any chain that could be obtained by Addition. For example, the summary of the diagram



is the diagram:



<sup>20</sup>Peirce [1958], pp. 310–311.

<sup>21</sup>Peirce [1958], p. 310.

<sup>22</sup>Shin [1994], pp. 87–88.

The two chains are shortened, and then the remaining 2-link chain is erased because it could be obtained from the unconnected 'x' by Addition. A summary is said to be *inconsistent* if the result is an inconsistent diagram.

**THEOREM 7.** Two Venn diagrams having the same labels (and number of circles) are logically equivalent if and only if they either have the same summary or else both have inconsistent summaries.

**Proof.** The right-to-left direction of the theorem follows from the fact that Contraction results in a logically equivalent diagram and Addition is a valid rule of inference. For the contrapositive of the other direction, suppose that two diagrams have distinct (consistent) summaries  $d$  and  $e$ . Then some minimal region has an 'o' in one but not the other summary, or else some region has an 'x'-chain in one but not the other. The 'x'-chain case is handled.

Case 1: A region of  $d$  has an 'x'-chain but no subregion of  $e$  has one. Assign the empty set to each minimal region of the chain. Because  $e$  does not have an 'x' in any of those regions, it can still be satisfied by extending this model. The result satisfies  $e$  but not  $d$ . Likewise with  $d$  and  $e$  reversed.

Case 2: A region of  $d$  has an 'x'-chain and some proper subregion of  $e$  has an 'x'-chain. Let  $r$  be a minimal region with an 'x' of the chain in  $d$  but not  $e$ . Assign to  $r$  a non-empty set, but to all other regions of the 'x'-chain the empty set. Extend this model to satisfy  $d$ . The model does not satisfy the 'x'-chain of  $e$  in the subregion and so does not satisfy  $e$ . Likewise with  $d$  and  $e$  reversed. ■

The following completeness result for Venn diagrams shows that if a diagram  $e$  is a consequence of a diagram  $d$  with the same labels, then  $e$  can be obtained from  $d$  by applying Contraction a number of times followed by either one application of Contradiction or else a number of applications of Addition and Simplification.

**THEOREM 8 (Completeness).** If diagram  $e$  is a logical consequence of  $d$  and both have the same labels (and number of circles), then  $e$  is provable from  $d$ .<sup>23</sup>

**Proof.** Assume  $e$  is a logical consequence of  $d$ . Apply Contraction to  $d$  as many times as possible. It can be assumed that at no time during this process is an inconsistent diagram obtained. For if one were,  $e$  would be immediately obtainable by Contradiction, establishing  $e$ 's provability. By applying Contraction a number of times, it can be assumed without loss of generality that no minimal region of  $d$  has both an 'x' and an 'o'.

---

<sup>23</sup>This result is essentially a special case of the completeness result proved in Shin [1994], pp. 98–110.

First note that every minimal region with an ‘o’ in  $e$  has an ‘o’ in  $d$ . Suppose otherwise for some minimal region  $r$ . Construct a model which assigns

$$\begin{cases} \text{the empty set} & \text{to any minimal region of } d \text{ with an ‘o’} \\ \text{a non-empty set} & \text{to all other minimal regions of } d \end{cases}$$

Such a model satisfies  $d$ . Because  $r$  is assigned the empty set, the model does not satisfy  $e$ , a contradiction.

Next note that for every region with an ‘x’-chain in  $e$ , some subregion has an ‘x’-chain in  $d$ . Suppose otherwise for some region  $r$ . Construct a model which assigns

$$\begin{cases} \text{the empty set} & \text{to region } r \\ \text{the empty set} & \text{to any minimal region of } d \text{ with an ‘o’} \\ \text{a non-empty set} & \text{to all other minimal regions of } d \end{cases}$$

Because no subregion of  $r$  has an ‘x’-chain, this model satisfies  $d$ . However, because  $r$  is assigned the empty set, the model does not satisfy  $e$ , a contradiction.

These two observations imply that  $e$  can now be obtained from  $d$  by several applications of Addition and Simplification. ■

More general completeness results extending Theorem 8 can be proved by formulating rules of Merge, Add Circle, and Remove Circle. This section concludes with a formulation of these three rules of inference.

**RULE 9 (Merge).** Two diagrams having the same labels may be combined into a single diagram as follows:

1. A new diagram is drawn with circles labeled by each of the letters occurring in the two premises.
2. For each minimal region of either premise with an ‘o’, add an ‘o’ to each of its counterparts in the conclusion to which an ‘o’ has not already been added.<sup>24</sup>
3. For each region of either premise with an ‘x’-chain, add an ‘x’-chain to its counterpart in the conclusion if one has not already been added to that region.<sup>25</sup>

---

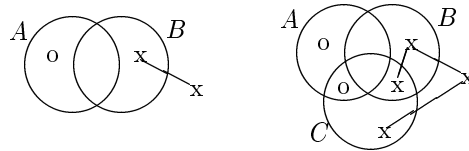
<sup>24</sup>The formulation of this rule uses the more general concept of any two regions being counterparts. This is defined as follows: (i) If two circles are labeled by the same letter, the two regions within the circles are counterparts. (ii) If two regions are counterparts then their two ‘complements’ are counterparts, where the ‘complement’ of a region is the combination of minimal region of the diagram that are not a part of the region. (iii) If two pairs of regions are counterparts, then the two ‘unions’ of the two pairs are counterparts, where the ‘union’ of a pair of regions is the combination of all minimal regions that are a part of either region. (iv) No other two regions are counterparts.

<sup>25</sup>Examples of this rule are given in Peirce [1958], e.g., p. 312, however the rule is not stated explicitly. It is stated in essentially this form in Shin [1994], pp. 88–92.

Next is addition of new circles to Venn diagrams. First, the new circle must be drawn so that the well-formedness of the overlapping circles is preserved, that is, so that all Boolean combinations of the circles are represented. In doing this, any minimal region of the original diagram is broken into two parts, one within the new circle and the other outside of the new circle. Hence, any 'o' occurring in a minimal region needs to be replaced by two connected 'o's, one in each of the two new subregions. Similarly, any 'x' occurring in a minimal region needs to be split into two parts, one within the new circle and the other outside of the new circle, with the two being connected by a line.

RULE 10 (Add Circle). A new circle may be added to a Venn diagram in such a way that well-formedness is preserved, provided all 'x's and 'o's are split as described.<sup>26</sup>

The following is an example of an application of Add Circle:

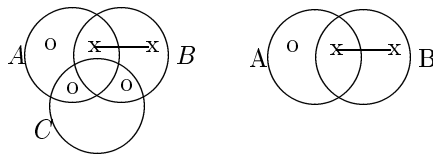


The 'o' is replaced by two 'o's, and the two 'x's are each replaced by two 'x's that are connected by lines.

The final rule is that allowing the removal of a circle. The removal of a circle from a diagram throws pairs of *adjoining* minimal regions together: one within the circle and one outside of the new circle.

RULE 11 (Remove Circle). A circle may be erased provided any two 'o's in adjoining minimal regions are replaced by a single 'o' when the two regions are thrown together, and any 'o's without an 'o' in the adjoining region are erased.<sup>27</sup>

The following is an example of Remove Circle:



<sup>26</sup>Peirce [1958], p. 311 and Shin [1994], pp. 86–87.

<sup>27</sup>Peirce [1958], p. 311 and Shin [1994], pp. 82–85.

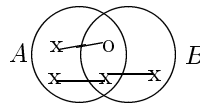
The two ‘o’s in the two regions that are thrown together when the circle is erased are replaced by a single ‘o’ while the ‘o’ without an adjoining ‘o’ is erased. The ‘x’-chain is left as is.

The repeated application of Merge allows any finite set of diagrams to be combined into a single, logically equivalent ‘conjunction’. Add Circle allows new circles to be added to any diagram, the result being logically equivalent. Remove Circle has the property that if  $e$  is a logical consequence of  $d$  but a circle in  $e$  is labeled by a letter not occurring in  $d$ , then the result of removing that circle using Remove Circle is a diagram that still implies  $e$ . The earliest general completeness result using these additional rules is due to Shin.<sup>28</sup> Another is in Hammer and Danner [1996].

### 3 PEIRCE–VENN DIAGRAMS

This section presents the logic of *Peirce–Venn diagrams*, Peirce’s variation and extension of Venn diagrams developed in 1903.<sup>29</sup> Peirce’s system is equivalent to the monadic fragment of first-order logic in expressive power. It also is based on what amounts to a conjunctive normal form. In fact, the key rules of inference formulated by Peirce are practically identical to the resolution proof procedure for propositional logic.

All Venn diagrams are also Peirce–Venn diagrams. However, Peirce–Venn diagrams allow any combination of ‘x’s and ‘o’s to be connected by lines to form a disjunctive chain. For example, the following is a Peirce–Venn diagram with two chains:



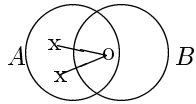
It asserts that either some  $A$  are not  $B$  or no  $A$  is  $B$  (by means of the upper chain) and something is either  $A$  or  $B$  (by means of the lower chain).

Because it is redundant to have a single chain with more than one ‘x’ in one minimal region or more than one ‘o’ in one minimal region, well-formed Peirce–Venn diagrams are required to have at most one ‘x’ and at most one ‘o’ in each minimal region. This rules out the following diagram as not well-formed:

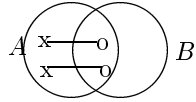
---

<sup>28</sup> Shin[1994], pp. 98–110.

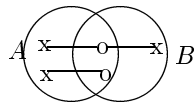
<sup>29</sup>Peirce [1958], pp. 294–319.



Likewise, well-formed Peirce–Venn diagrams may not have two chains in the same region that have ‘x’s and ‘o’s in the same minimal regions. This rules out the following diagram:



However, the following diagram is not ruled out:



The syntax of well-formed Peirce–Venn diagrams can be summarized by the following conditions:

1. Any  $n$  circles drawn to overlap in all combinations as described above and labeled by  $n$  names is a well-formed Peirce–Venn diagrams.
2. Given any Venn diagram, the result of adding a chain of ‘x’s and ‘o’s to any region not already having such a chain results in a well-formed Peirce–Venn diagram.
3. Nothing else is a well-formed Peirce–Venn diagram.

A Peirce–Venn diagram is *consistent* just in case no minimal region has both an unconnected ‘o’ and an unconnected ‘x’ in it.

The semantics for Peirce–Venn diagrams are given by the following conditions:<sup>30</sup>

DEFINITION 12 (Satisfies).

1. A model *satisfies* an ‘x’ occurring in some minimal region just in case the set assigned to that minimal region is non-empty.

---

<sup>30</sup>Hammer [1995b], pp. 817–818.



2. A model *satisfies* an ‘o’ occurring in some minimal region just in case the set assigned to that minimal region is empty.
3. A model *satisfies* a chain occurring in some region just in case the set satisfies at least one of the ‘x’s or ‘o’s in the chain.
4. A model *satisfies* a Peirce–Venn diagram just in case it satisfies each chain in the diagram.

There is a very close similarity between Peirce–Venn diagram and propositional sentences in conjunctive normal form. A Peirce–Venn diagram is interpreted as a conjunction of disjunctions, each of the distinct chains being a ‘conjunct’ and each link of such a chain being a ‘disjunct’.

For some purposes, it is convenient to represent Peirce–Venn diagrams in tabular form. Lower-case letters are used to represent the minimal regions of a diagram having either an ‘x’ or an ‘o’. One row of a table represents one chain of the diagram. The left side of a row consists the minimal regions that have an ‘x’ from the chain while the right side consists of the minimal regions that have an ‘o’ from the chain.

As an example, the following table could be used to represent a Peirce–Venn diagram with two chains, the various links of which fall in five different minimal regions (temporarily referred to as  $a$ ,  $b$ ,  $c$ ,  $d$ , and  $e$ ):

‘x’	‘o’
$a, b$	$c, d$
$c, e$	$b$

The first chain has two ‘x’s in regions  $a$  and  $b$  and two ‘o’s in regions  $c$  and  $d$ . The second chain has ‘x’s in  $c$  and  $e$  and an ‘o’ in  $b$ .

Notice that the conditions on well-formed diagram prevent such a table from having two duplicate rows. They also prevent a table from having any row where the same letter appears twice on the left or twice on the right.

The two rows of the above table can also be expressed as the two propositional sentences

$$a \vee b \vee \neg c \vee \neg d$$

and

$$c \vee e \vee \neg b$$

where  $\neg b$  represents an ‘o’ in minimal region  $b$  and  $b$  represents an ‘x’ in region  $b$ , and a disjunction of such literals represents a chain of such ‘x’s and ‘o’s in those minimal regions.

The first task is to show that Peirce–Venn diagrams are equivalent in expressive power to the monadic fragment of first-order logic.<sup>31</sup> The sentences

<sup>31</sup>An extension of Venn diagrams is formulated in Shin [1994], pp. 111–152, with the same expressive power. That system allows any finite disjunction of Venn diagrams to qualify as a well-formed diagram.

of *monadic logic* are those sentences of first-order logic without identity constructible from languages involving only one-place predicate symbols.

**THEOREM 13.** Peirce–Venn diagrams are equivalent to monadic logic.

**Proof.** It is clear how to derive an equivalent monadic sentence from a given Peirce–Venn diagram, so only the converse will be shown. Let  $\phi$  be a monadic sentence and let  $\mathcal{L}(\phi)$  be the set of all predicates occurring in  $\phi$ . First, the quantifiers of  $\phi$  are driven inwards so that the scope of each quantifier is a truth function of atomic formulas each involving the quantified variable.<sup>32</sup> It can be assumed that only existential quantifiers occur. Consider such a subformula  $\exists x\psi$ . The truth-function  $\psi$  can be put into disjunctive normal form, resulting in

$$\exists x(\phi_1 \vee \dots \vee \phi_n)$$

which is equivalent to

$$\exists x\phi_1 \vee \dots \vee \exists x\phi_n$$

Thus the scope of each quantifier is a conjunction of atomic formulas and negated atomic formulas. Notice that each such conjunction, say  $A(x) \wedge \neg B(x) \wedge C(x)$ , can be written as a term such as  $A\overline{B}C$ , the notation used above. Now expand each such existentially quantified term into a disjunction of existentially quantified terms each of which involves every predicate in  $\mathcal{L}(\phi)$ . For example, if the subformula is  $\exists xA\overline{B}C$  and  $D$  is the only other predicate in  $\mathcal{L}(\phi)$ , the result would be

$$\exists xA\overline{B}CD \vee \exists xA\overline{B}C\overline{D}$$

Call each such disjunct a *complete atom*. Thus, complete atoms are existentially quantified conjunction of atomic formulas and negated atomic formulas in which each predicates in  $\mathcal{L}(\phi)$  occurs once. Now put the entire sentence into conjunctive normal form using the complete atoms as atomic elements. Each conjunct of the resulting sentence corresponds to one chain, and each disjunct of a conjunct corresponds to one ‘x’ or ‘o’ of the chain, depending on whether the disjunct is negated or not. The minimal region the ‘x’ or ‘o’ should be drawn in depends on which region corresponds to the quantified term. Thus, a Peirce–Venn diagram can be drawn for  $\phi$  by drawing one circle for each predicate in  $\mathcal{L}(\phi)$  and adding one chain for each conjunct of the derived sentence. ■

The rules of inference for Peirce–Venn diagrams include parallels of the rules of Addition, Simplification, Contraction, and Contradiction from Venn diagrams. The primary rule, Peirce’s Rule, is new.

---

<sup>32</sup>A procedure for this deriving from Behmann [1922] is described in Quine [1982].

RULE 14 (Addition). A chain can be extended with an additional ‘x’ or ‘o’ in a new minimal region.

RULE 15 (Simplification). Any entire chain can be erased.

RULE 16 (Contradiction). Any diagram can be inferred from a diagram having a minimal region with both an unconnected ‘o’ and an unconnected ‘x’.

RULE 17 (Contraction).

Case 1: If a chain has an ‘x’ in a minimal region having an unconnected ‘o’, the ‘x’ and the two halves reconnected.

Case 2: If a chain has an ‘o’ in a minimal region having an unconnected ‘x’, the ‘o’ can be erased and the two halves reconnected.

RULE 18 (Peirce’s Rule). If a chain has an ‘x’ in a minimal region and another chain has an ‘o’ in that same region, the ‘x’ and the ‘o’ can be that ‘x’ can be erased provided the four halves of the remaining chains are all connected to each other.<sup>33</sup>

Expressed in terms of tables, Peirce’s Rule states that two rows that share a letter that is on the right in one and on the left in the other can be combined into a larger row, with the two letters in common erased (unless they were the only letter on that side of the row). As an example, a Peirce–Venn diagram represented as

$$\begin{array}{c|c} \text{‘x’} & \text{‘o’} \\ \hline a, b & c \\ d & b, e \end{array}$$

implies

$$\begin{array}{c|c} \text{‘x’} & \text{‘o’} \\ \hline a, d & c, e \end{array}$$

In minimal region  $B$ , one chain has an ‘x’ and the other has an ‘o’. The ‘x’ and ‘o’ are erased, and the resulting pieces from the two chains are connected together.

Peirce’s Rule is essentially identical to the Resolution Rule of the Resolution proof procedure for propositional logic, which operates on propositional sentences in conjunctive normal form.

The following lemma shows that a trivial test determines whether or not a Peirce–Venn diagram is satisfiable.

DEFINITION 19 (Peirce Closure). The *Peirce closure* of a diagram is the result of applying Peirce’s Rule and Contraction to it as many times as possible.

---

<sup>33</sup> Peirce [1958], pp. 310–311.

LEMMA 20. A Peirce–Venn diagram is satisfiable if and only if its Peirce closure is consistent.

**Proof.** One direction follows from the validity of Peirce’s Rule. For the other direction, note first that a consistent Peirce closure is satisfiable because no minimal region has both an ‘x’ and an ‘o’. Next note that an application of Peirce’s Rule on two chains that conflict in some minimal region results in two subchains of the original chains. Therefore any model satisfying the new chains must also satisfy the two original chains. Hence, by induction, any model satisfying the Peirce closure also satisfies the original diagram. ■

Lemma 20 provides a simple decision procedure for propositional logic. Given a propositional sentence  $\phi$ , construct the conjunctive normal form of  $\neg\phi$ . Draw a Peirce–Venn diagram with at least as many minimal regions as propositional variables in  $\phi$ . Assign each variable  $P$  to a fixed minimal region and let  $P$  translate to an ‘x’ in that region,  $\neg P$  translate to an ‘o’ in that region, and each conjunct of the CNF translate to a connected chain of these ‘x’s and ‘o’s. Let the assignment of a non-empty set to a minimal region translates to the assignment of truth to the variable corresponding to it, and the empty set to false. The resulting Peirce–Venn diagram is unsatisfiable if and only if  $\phi$  is valid. Hence  $\phi$  is valid if and only if the Peirce closure of the diagram is inconsistent.

Next a completeness result for Peirce’s diagrammatic logic is proved. A somewhat different completeness result for a natural deduction formulation of Peirce’s system is given in Hammer [1995b].<sup>34</sup>

THEOREM 21 (Completeness). If diagram  $e$  is a logical consequence of  $d$  and both have the same labels (and number of circles), then  $e$  is provable from  $d$ .

**Proof.** Assume that  $e$  is not provable from  $d$ . Take the Peirce closure of  $d$ , which we can assume is consistent. Some chain

$$p \vee q \vee \neg r \vee \neg s$$

occurs in  $e$  but no subchain of it occurs in the Peirce closure. (The same argument will work for other types of chains). We construct a model satisfying  $d$  but not  $e$ . Add the two unconnected ‘o’s  $\neg p, \neg q$  and the two unconnected ‘x’s  $r, s$  to the Peirce closure, obtaining  $d'$ . We construct a model of  $d'$  using Lemma 20. Suppose the Peirce closure of  $d'$  were inconsistent, say resulting in an ‘x’ and ‘o’  $z$  and  $\neg z$ . The presence of  $\neg p, \neg q, r, s$  allow chains having any of  $p, q, \neg r$ , or  $\neg s$  as links to be shortened by Contraction. Consider now the same proof with all uses of the added ‘x’s and ‘o’s  $\neg p, \neg q, r, s$  removed (this is a proof from the Peirce closure of  $d$ .) The result is two chains

---

<sup>34</sup>pp. 821–825.

$p \vee q \vee \neg r \vee \neg s \vee z$  and  $p \vee q \vee \neg r \vee \neg s \vee \neg z$ , including at least the last link but possibly not all of the other links, depending on which of  $\neg p, \neg q, r, s$  were used in the original proof. One application of Peirce's Rule to these two chains results in either  $p \vee q \vee \neg r \vee \neg s$  or a subchain of  $p \vee q \vee \neg r \vee \neg s$  (from which  $p \vee q \vee \neg r \vee \neg s$  is obtainable by Addition). This contradicts that  $p \vee q \vee \neg r \vee \neg s$  is not provable from  $d$ . Hence the Peirce closure of  $d'$  is consistent, and so by Lemma 20  $e$  is not a logical consequence of  $d$ . ■

As with the Venn system, rules of Merge, Add Circle, and Remove Circle can be formulated that allow more general completeness results to be proved for Peirce–Venn diagrams.

#### 4 EXISTENTIAL GRAPHS

This section describes the logic of *existential graphs* developed by Peirce.<sup>35</sup> Existential graphs, a system arising from Peirce's work on the calculus of relations and predicate logic, is a graphical system for representing logical sentences and inferences.

Peirce wavered somewhat on the purpose of existential graphs. In 1911 he describes the system as a 'system of logical symbols' whose 'purpose and end is simply and solely the investigation of the theory of logic, and not at all the construction of a calculus to aid the drawing of inferences'.<sup>36</sup> Likewise, in 1903 he writes of the system that 'the whole effort has been to dissect the operations of inference into as many distinct steps as possible'.<sup>37</sup> The system is presented in these statements as an analytical device rather than a practical tool. On the other hand, in 1906 Peirce describes the system as a practical reasoning tool: 'The system of Existential Graphs which I have now sufficiently described - or, at any rate, have described as well as I know how, leaving the further perfection of it to others - greatly facilitates the solution of problems of Logic...'.<sup>38</sup> This statement describes the system as a practical tool designed to assist in logical reasoning.

The system of existential graphs was divided by Peirce into several natural fragments. The *alpha* fragment is equivalent to propositional logic, and forms a very elegant and workable substitute. The *beta* fragment is equivalent to first-order logic with identity. Its rules are much more complex than those of the alpha fragment, and is a system that is not readily analyzable. These are the two most polished fragments of the system of existential graphs, and are the two examined here. The *gamma* fragment allows expressions of modality, abstraction, higher-order quantification, and state-

---

<sup>35</sup> Peirce [1958].

<sup>36</sup> Peirce [1958], p. 320.

<sup>37</sup> Peirce [1958], p. 343.

<sup>38</sup> Peirce [1958], pp. 458–459.

ments about existential graphs themselves.<sup>39</sup> A good description of Peirce's entire system is Roberts [1973].

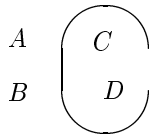
Graphs are drawn on the *sheet of assertion*: a blank, empty area of the page on which is drawn all that is asserted. The blank sheet of assertion is logically true, since nothing is, in that case, being asserted.

Several graphs drawn on the sheet of assertion are interpreted conjunctively. Thus,



is equivalent to  $A \wedge B \wedge C$ .

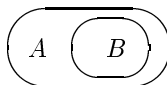
A closed curve (as with the circles of Venn diagrams) called a *cut* negates the subgraph that it encloses. Thus,



is equivalent to  $A \wedge B \wedge \neg(C \wedge D)$ .

A cut that encloses no subgraph other than a part of the sheet of assertion is logically false because it denies the empty subgraph consisting of no assertion.

A common idiom is used frequently by Peirce to graph implications. To graph an implication, first two concentrically nested cuts are drawn. Then the antecedent is drawn in the area within the outer cut but outside of the inner cut, and the consequent drawn within the inner cut. Thus, the following graph is equivalent to 'if  $A$  then  $B$ '.



The first rule of inference allows a double negation to be added or removed from any subgraph.

**RULE 22 (Double Cut).** Two concentrically nested cuts may be erased or added around any subgraph.

<sup>39</sup>See Peirce [1958], pp. 401–410.

RULE 23 (Insertion in Odd). Any graph may be drawn on an area of the sheet of assertion that is enclosed by an odd number of cuts.

Insertion in Odd can be thought of as allowing additional assumptions to be added to subproofs.<sup>40</sup>

RULE 24 (Erasure in Even). Any subgraph drawn on an area of the sheet of assertion that is enclosed by an even number of cuts may be erased.

Erasure in Even can be thought of as a generalized version of simplification, the rule allowing any conjunct to be eliminated from a conjunction.

RULE 25 (Iteration). A subgraph may be copied to any other area on the sheet of assertion that falls within all of the cuts enclosing the original subgraph.

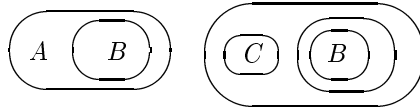
Iteration can be understood as allowing one to reiterate or use assumptions or facts in subproofs within their scope.

RULE 26 (Deiteration). Any subgraph that could have been drawn as a result of the rule of Iteration may be erased.

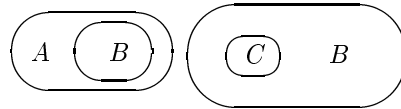
Conversely, Deiteration encodes the principle that if a previously established fact occurs in a subproof, there is no need to reestablish it in the subproof and so any such redundant occurrence can be eliminated.

The *alpha* fragment of Peirce's system is equivalent to propositional logic. Completeness results for various formulations of the system have been provided by Zeman [1964], Roberts [1964], Roberts [1973], White [1984], and Hammer [1995a].

The following is an example of a proof that uses all five inference rules. The conclusion is a graph of 'if  $A$  and  $D$ , then  $C$ '. The premises are graphs of 'if  $A$  then  $B$ ' and 'if not- $C$ , then not- $B$ ':

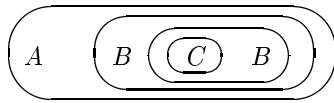


By Double Cut:

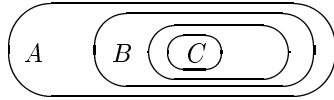


<sup>40</sup>The connection between the rules of existential graphs and natural deduction proofs is made in Roberts [1964].

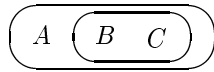
By Iteration of the right graph into the left graph and then Erasure of the right graph:



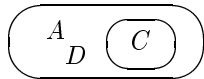
By Deiteration:



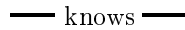
By Double Cut:



Finally, by Erasure in Even (of  $B$ ) and then Insertion in Odd (of  $D$ ):



Peirce's next fragment of existential graphs, the *beta* system, is much more complicated. Peirce uses what he calls *lines of identity* instead of variables. The formula  $x$  knows  $y$  would be approximated by the graph:



Actually, lines of identity also have quantificational import, so the graph is really the equivalent of  $\exists x \exists y (x \text{ knows } y)$ .

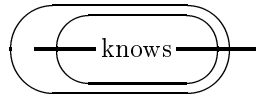
The next graph is equivalent to  $\exists x \neg \exists y (x \text{ knows } y)$ , or equivalently, 'Someone knows nobody.'





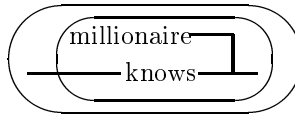
As this example shows, the scope of quantification associated with a line of identity is determined by the portion of the line that is the least deeply enclosed within cuts. In particular, the order in which the elements of a graph are interpreted is (i) lines of identity on the sheet of assertion (i.e., with parts enclosed by no cuts), (ii) cuts on the sheet of assertion, (iii) lines of identity on the sheet of assertion (i.e., with parts enclosed by no cuts), (iv) cuts on the sheet of assertion, etc.

The next graph is equivalent to ‘someone knows everyone’.



Notice that lines of identity enclosed by an odd number of cuts are naturally interpreted as universally quantified.

Cross-reference (indicated in first-order logic by the same variable occurring more than once) is accomplished in the system of existential graphs by allowing lines of identity to branch. For example, the following graph is equivalent to ‘everyone knows a millionaire’.



The following is a selection of most of the rules of inference for the beta fragment.<sup>41</sup> For a more complete list see Peirce [1958], Zeman [1964], Roberts [1973] and Roberts [1992]. Most of the rules are generalizations of the alpha rules, now taking into account lines of identity.

**RULE 27 (Double Cut).** Concentrically nested cuts may be added or removed around any subgraph as long as no graphs occur in the area within the outer cut but outside of the inner cut except possibly lines of identity that pass directly from within the inner cut to outside of the other cut.

This is the same rule as in the alpha system, with the only exception being that lines of identity are allowed to pass directly through the two cuts.

**RULE 28 (Erasure).** Any subgraph occurring within an even number of cuts may be erased, including an evenly enclosed portion of a line of identity.

<sup>41</sup>See Peirce [1958], pp. 395–396.

RULE 29 (Iteration). A subgraph of a graph can be copied to any other part of the graph which falls within the same or additional cuts.<sup>42</sup>

This rule is not stated in its entirety, which allows lines of identity to be connected to their iterated counterparts.

RULE 30 (Deiteration). A subgraph that could be the result of an application of Iteration can be erased.

RULE 31 (Connect in Odd). Two loose ends of lines of identity that occur in the same, oddly enclosed area can be connected.

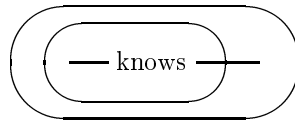
Graphs in oddly enclosed areas behave like assumptions. Connecting two loose ends in such an area has something of the effect of making a stronger assumption, namely that the two objects are identical.

RULE 32 (Retraction Outwards). A loose end of a line of identity can be retracted as long as the only cuts it is retracted across are in the direction of within the cut to outside of the cut.

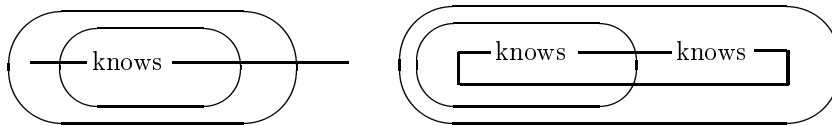
RULE 33 (Extension Inwards). A loose end can be extended inwards through zero or more additional cuts.

RULE 34 (Branch). A branch can be added to any portion of a line of identity.

Here is an example of a non-trivial proof in the beta system. The conclusion is a graph of ‘everyone is known by someone’:

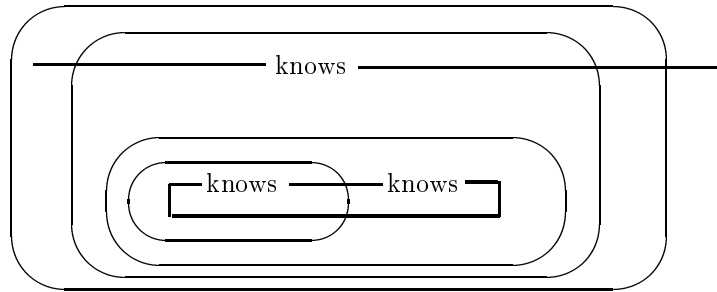


The two premises are graphs of ‘someone is known by everyone’ and ‘if someone knows another, that person also knows the first’ (or ‘knows is symmetric’):



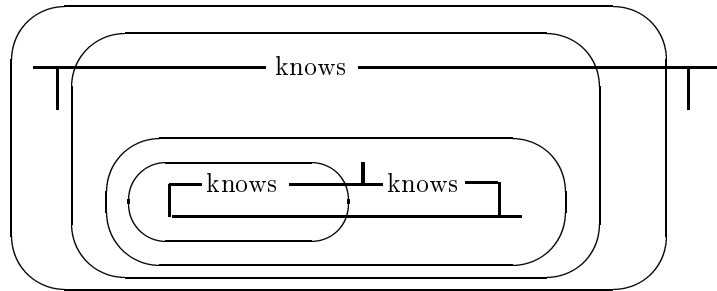
To begin the proof, an application of Iteration to the graph of ‘knows is symmetric’ (and then an application of Erasure to the original) gives:

<sup>42</sup>Peirce [1958], p. 396.

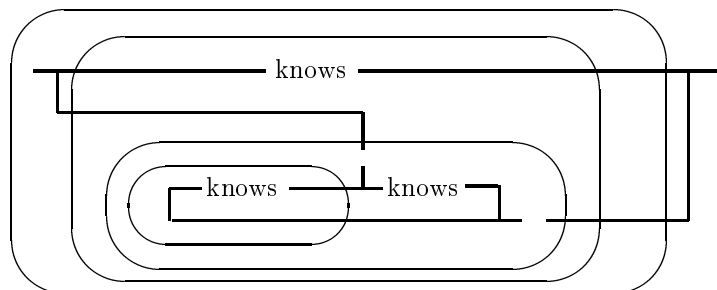


The effect of this is to bring the premise 'knows is symmetric' within the scope of the other premise so that the two can be combined.

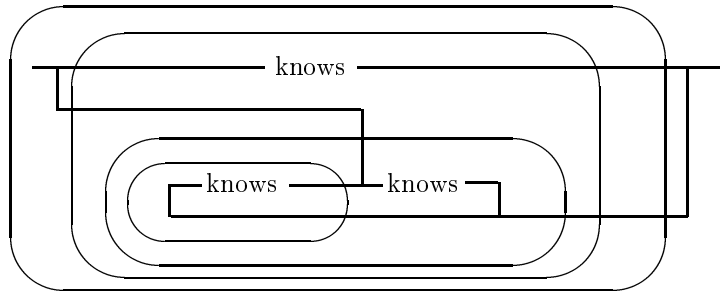
The next step is to connect the two lines of identity of each premise to a line of the other premise, thereby identifying the variables of the two premises. This is done using Branch, Extension Inwards, and Connect in Odd. First, four applications of Branch gives:



Second, two applications of Extension Inwards to the two new outer branches gives:

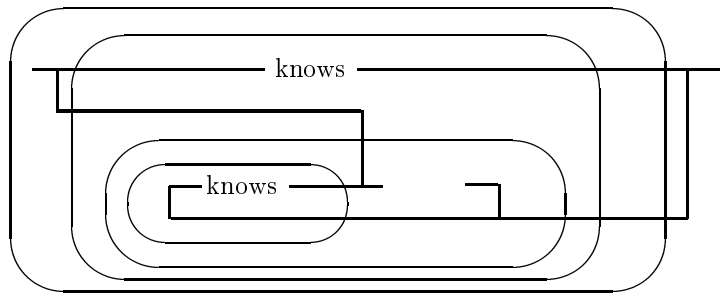


Finally, two applications of Connect in Odd gives:

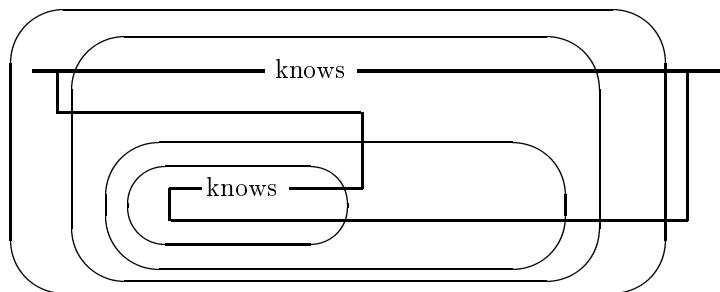


The effect of these operations is that the two pairs of 'variables' of the two premises have been identified, allowing the lines of identity of the two premises to interact.

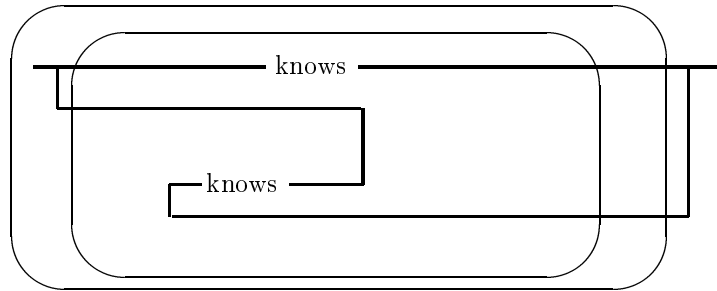
Eventually, the innermost occurrence of 'knows' will be the predicate of the conclusion, the other two being eliminated once they have been used. An application of Deiteration (to the subgraph 'knows') gives:



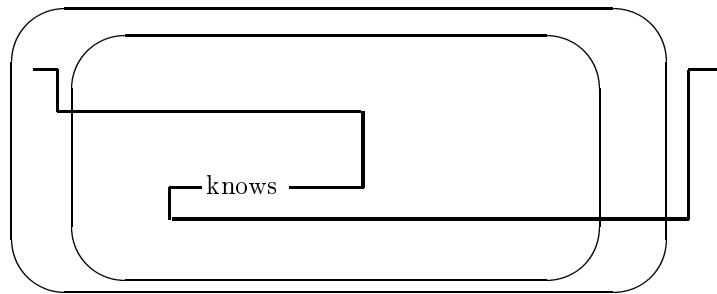
Two applications of Retraction to the loose ends results in:



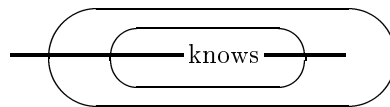
Then, by Double Cut:



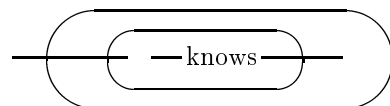
Because the outermost 'knows' occurs within an even number of cuts it can be eliminated. Thus, an application of Erasure and then two applications of Retraction on the loose ends yields:



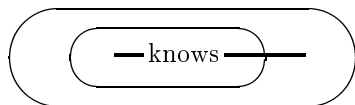
Restructuring this graph for readability gives:



This graph says that the person who was known by everyone knows everybody. To get the conclusion, an application of Erasure gives:



Finally, an application of Retraction Outwards to the unconnected line of identity and then Erasure yields the conclusion:



Analysis of the beta system of existential graphs remains uncompleted at this time partly because the system's unusual topological syntax resists many standard techniques. However, completeness results have been reported in Zeman [1964] and Roberts [1973], and consistency results have been reported in Zeman [1964] and Roberts [1973, 1992].

## 5 CONCLUSION

A diagrammatic logic is simply a logic whose target objects are diagrams rather than sentences. Other than this, diagrammatic logics and logics involving expressions of some language are not different in kind. In either case, the logic should provide an adequate description of the class of representations being studied, their meaning, and the principles behind their use and purpose within the system of which they are a part. The reasons for formulating and analyzing a diagrammatic logic are also the same as for a language-based logic. If for any reason the grammatical, semantical, or inferential properties of a diagrammatic system need to be determined precisely, say for computational or psychological purposes, a diagrammatic logic will do exactly that.

*Peoplesoft Inc., USA.*

## BIBLIOGRAPHY

- [Allwein and Barwise, 1996] G. Allwein and J. Barwise. *Logical Reasoning with Diagrams*, Oxford University Press, 1996.
- [Barwise and Etchemendy, 1991] J. Barwise and J. Etchemendy. Visual Information and Valid Reasoning. In *Visualization in Teaching and Learning Mathematics*. Mathematical Association of America, 1991.
- [Barwise and Etchemendy, 1995] J. Barwise and J. Etchemendy. Heterogeneous Logic. In Glasgow, Narayanan, and Chandrasekaran [1995].
- [Behmann, 1922] H. Behmann. Beiträge zur Algebra der Logik, insbesondere zum Entscheidungsproblem. *Mathematische Annalen* 86. pp. 163–229, 1922.
- [Euler, 1846] L. Euler. *Lettres à une Princesse d'Allemagne sur Divers Sujets de Physique et de Philosophie*. New York: Harper and Bros, 1846.
- [Gardner, 1982] M. Gardner. Logic Diagrams. *Logic Machines and Diagrams*, Chapter 2. University of Chicago Press, 1982.

- [Glasgow *et al.*, 1995] J. Glasgow, N. Narayanan, and B. Chandrasekaran. *Diagrammatic Reasoning*. Menlo Park, Cambridge and London: AAAI Press/The MIT Press, 1995.
- [Hammer, 1994] E. Hammer. Reasoning with Sentences and Diagrams, *The Notre Dame Journal of Formal Logic*, 35, 1994.
- [Hammer, 1995a] E. Hammer. *Logic and Visual Information*. Stanford: CSLI Publications and the European Association for Logic, Language and Information, 1995.
- [Hammer, 1995b] E. Hammer. Peirce on Logical Diagrams. *Transactions of the Charles S. Peirce Society*, 31(3): 807–828, 1995.
- [Hammer and Danner, 1996] E. Hammer and N. Danner. Towards a Model Theory of Diagrams. *Journal of Philosophical Logic*, 1996.
- [Hammer, 1997] E. Hammer. Semantics for Existential Graphs. *Journal of Philosophical Logic*, 1997.
- [Hammer and Shin, 1996] E. Hammer and S. Shin. Euler and the Role of Visualization in Logic. In J. Seligman and D. Westerstahl (Eds.), *Language, Logic and Computation: The 1994 Moraga Proceedings*. Stanford: CSLI Publications, 1996.
- [Harel, 1988] D. Harel. On Visual Formalisms. *Communications of the ACM*, 31(5): 514–530. Reprinted in: *Visual Programming Environments, Vol. I: Paradigms and Systems*, E.P. Glinert (ed.), IEEE Computer Society Press, Washington DC, 1990, pp. 171–187, 1988.
- [Johnson *et al.*, 1996] S. Johnson, J. Barwise, and G. Allwein. Toward the Rigorous Use of Diagrams in Reasoning about Hardware. In Allwein and Barwise [1996].
- [Luengo, 1995] I. Luengo. *Diagrams in Geometry*. PhD Thesis, Indiana University, 1995.
- [More, 1959] T. More. On the Construction of Venn Diagrams. *Journal of Symbolic Logic*, 24, 1959.
- [Peirce *et al.*, 1958] C. Peirce, C. Hartshorne, P. Weiss and A. Burks (eds.) *The Collected Papers of C. S. Peirce*, Volume 4, Book 2. Cambridge: Harvard University Press, 1958.
- [Quine, 1982] W. V. O. Quine. *Methods of Logic*, 4th ed. Cambridge: Harvard University Press, 1982.
- [Roberts, 1973] D. Roberts. *The Existential Graphs of Charles S. Peirce*. The Hague: Mouton and Co, 1973.
- [Roberts, 1964] D. Roberts. The Existential Graphs and Natural Deduction, *Studies in the Philosophy of Charles Sanders Peirce*, Edward Moore and Richard Robin (eds.), University of Massachusetts Press, 1964.
- [Roberts, 1992] D. Roberts. The Existential Graphs. *Computers in Mathematics with Applications*, 23(6-9): 639–663, 1992.
- [Shin, 1991] S. Shin. A Situation-Theoretic Account of Valid Reasoning with Venn Diagrams. In J. Barwise *et al.*, editors, *Situation Theory and Its Applications, Vol. 2*. Stanford: CSLI Publications, 1991.
- [Shin, 1994] S. Shin. *The Logical Status of Diagrams*. Cambridge University Press, 1994.
- [Shin, 1994b] S. Shin. Peirce and the Logical Status of Diagrams. *History and Philosophy of Logic*, 15, 1994.
- [Sowa, 1984] J. Sowa. *Conceptual Structures: Information Processing in Mind and Machine*. Addison–Wesley, 1984.
- [Venn, 1880] J. Venn. On the Diagrammatic and Mechanical Representation of Propositions and Reasonings. *Philosophical Reasonings*, 1880.
- [Venn, 1894] J. Venn. *Symbolic Logic*. Revised 2nd edition, 1894.
- [White, 1984] R. White. Peirce’s Alpha Graphs: The Completeness of Propositional Logic and the Fast Simplification of Truth Functions. *Transactions of the Charles S. Peirce Society*, 20, 1984.
- [Zeman, 1964] J. Zeman. *The Graphical Logic of C. S. Peirce*. PhD thesis, University of Chicago, 1964.

# INDEX

- \*-continuity, 193
- \*-continuous dynamic algebra, *see*  
dynamic algebra
- $\mathfrak{A}$ -validity, 161
- abstraction, 412
- acceptable structure, 202
- accessibility relation, 385
- acyclicity, 323, 327
- Adian structure, 203
- adjustment account, 358
- aggregative approach, 353
- AL, *see* Algorithmic Logic
- algebra
  - dynamic, *see* dynamic algebra
  - Kleene, *see* Kleene algebra
- algebraic stack, 154
- Algorithmic Logic, 180–181, 204
- Allwein, G., 395
- alternative set, 340
- ancestral reflexivity, 381
- anti-cyclic properties, 325, 348
- Aristotle, 319
- arithmetical
  - completeness, 174, 175, 202
  - structure, 165, 174, 202
- arity, 148, 152
- array, 152
  - assignment, 153
  - variable, 152
    - nullary, 156
- as expressive as, 134, 162
- assignment
  - array, 153
  - nondeterministic, *see* wildcard
  - random, 155
  - simple, 103, 147, 149
  - wildcard, 99, 152, 155, 179, 203
- associativity, 114
- asymmetry, 321
- atomic
  - formula, 149
  - program, 103, 147, 149
  - symbol, 113
  - test, 103
- automata PDL, 141
- automaton
  - finite, 141
  - $\omega$ -, 141
  - pushdown, 132
- axiomatization
  - DL, 170–176
  - equational theory of regular sets, 193
  - $\mu$ -calculus, 191
  - PDL, 128–129
  - PL, 188
- bad, 379
- Barwise, J., 395
- basic propositions, 353
- Behmann, H., 409
- belief change, 47
- belief revision, 319
- belief states, 48
- best, 379
- best choice connection, 344, 345
- better value concept: strict preference, 320
- binary
  - nondeterminism, 179
  - relation, 193
- blocks world diagrams, 395



- Boolean satisfiability, *see* satisfiability, propositional
- bounded memory, 152, 178
- bounded rationality, 357
- box
  - operator, 113, 185
- branching-time TL, 184
- calibration, 380
- canonical good, 380
- capture
  - of a complexity class by a spectrum, 169
- carrier, 155
- category-theory diagrams, 395
- cautiousness, 377
- centring, 362
- ceteris paribus, 358, 363, 371
- Chandrasekaran, B., 395
- Chernoff, 343–345
- choice, 343
  - operator, 113
- choice function, 343
- choice-guidance, 337
- circuit diagrams, 395
- circumscriptive, 384
- closed curves, 395
- closeness, 381
- closure
  - Fischer–Ladner, 125–126
- CLPDL, 145
- combination preferences, 347
- compactness, 121, 131, 162
- comparative schematology, 203
- comparison structure, 321
- compatible relata, 346
- complete atom, 409
- completeness, 162, 323, 348, 403, 404, 411
  - DL, 202
  - for termination assertions, 202
  - LED, 184
  - $\mu$ -calculus, 191
  - PDL, 128–129, 199
  - relative, 202
  - TL, 187
- complexity
  - of DL, 167
  - of DL, 167–170
  - of PDL, 130–132
  - of spectra, 169
- composition, 103
  - operator, 113
  - relational, 116
  - rule, 124
- compositionality, 111
- computation
  - sequence, 105, 118
- conceptual graphs, 395
- concurrent PDL, 146
- concurrent systems, 186
- conditional, 103, 115
  - rule, 124
- conditional logic that approximates AGM belief revision, 82
- conditional preference, 320
- conditionals, 5
- conjunctive expansion, 349, 373
- conjunctive normal form, 410
- connectedness, 323
- consequence operators, 354
- consistency, 129
- consistent, 400, 407
- constant, 148
  - test, 103
- constructive  $L_{\omega_1\omega}$ , 164
- context-free
  - DL, 201
  - language
    - simple-minded, 137
  - PDL, 134
  - program, 132, 134
  - set of seqs, 106
- continuity, 190
- contraction and revision, 51
- contradiction, 359, 379
- contranegativity, 386
- contraposition, 360, 373

- control-flow diagrams, 395
- converse, 121, 128, 142
- correctness
  - partial, *see* partial correctness
  - specification, 107
  - total, *see* total correctness
- correspondence between term and
  - minimal region, 396
- cotenability, 5
- counterparts of regions, 397
- CRPDL, 145
- cut, 413
- cyclic preferences, 331
  
- Danner, N., 398, 400, 406
- decidability
  - of PDL, 125, 127
- deduction theorem, 129
- $\Delta$ PDL, 144
- deontic concepts, 384
- deontic logic, 384
- deontic paradoxes, 386
- deterministic
  - Kripke frame, 139
  - semantically, 139
  - while** program, 103
- diagrammatic logic, 395
- diamond
  - operator, 115, 185
- direction of time (and conditionals), 33
- disjunctive interpolation, 350, 352, 369
- disjunctive normal form, 409
- domain
  - of computation, 101, 147, 155
- DPDL, 139
- duality, 115, 119
- DWP, 139
- dyadic value concepts: comparative, 320
- dynamic
  - formula, 181
  - term, 181
- dynamic algebra, 196–198
  - \*-continuous, 197
  - separable, 197
- Dynamic Logic
  - axiomatization, 171
  - basic, 149
  - context-free, 201
  - poor test, 103, 149
  - probabilistic, 183
  - rich test, 103, 149, 151
    - of r.e. programs, 162, 163
- effective definitional scheme, 183
- eligibility, 337
- endogenous, 111, 184
- epistemic choice, 346
- epistemic conditionals, 46
- equal expressive power, 163
- equality symbol, 148
- equivalence
  - of logics, 163, 204
- Etchemendy, J., 395
- Euler circles, 395
- Euler, L., 398
- exclusionary preferences, 321
- existential graphs, 395, 412
  - alpha fragment, 412, 414
  - beta fragment, 412
  - gamma fragment, 412
- exogenous, 111
- expansion, 344, 345
- expressive structure, 174
- expressiveness
  - relative, 176, 203
  - over  $\mathbb{N}$ , 166
- fairness, 155
- filtration, 125–128
  - for nonstandard models, 127–129
- finite
  - automaton, 141
  - model property, *see* small model property

- model theorem, *see* small model theorem
  - theorem
  - variant, 157
- first-order
  - spectrum, 202
  - test, 204
  - vocabulary, 148
- first-order logic, 412
- Fischer–Ladner closure, 125–126
- formula, 148
  - atomic, 149
  - DL, 150, 160
  - dynamic, 181
  - positive in a variable, 189
- free
  - occurrence of a variable
    - in DL, 171
- function
  - patching, 156
  - symbol, 148
- generalization rule, 129
- geometry diagrams, 395
- Glasgow, J., 395
- good, 379
- Good Samaritan, 386
- guarded command, 115, 199
- Halldén, S., 319
- halt**, 144, 181
- halting problem
  - over finite interpretations, 169
- Hammer, E., 396, 398, 400, 401, 406, 411, 414
- Harel, D., 395
- Hasse diagrams, 395
- Herbrand-like state, 168
- higher-order quantification, 412
- Hoare Logic, 124
- holistic approach, 353
- ideal worlds, 385
- idealization, 320
- incommensurable, 323
- incompatibility, 321
- incompleteness, 323
- indifference, 320, 379
- indifference-related good, 380
- individual
  - variable, 148, 151
- induction
  - axiom
    - PDL, 123, 128, 129
    - Peano arithmetic, 124
  - principle
    - for temporal logic, 185
  - structural, 111
- infinitary completeness
  - for DL, 202
- initial state, 157, 163, 168
- input variable, 102
- input/output
  - pair, 116, 155
  - relation, 102, 117, 151, 157
  - specification, 108, 110
- intermittent assertions method, 185
- interpreted reasoning, 164, 173
- intersubstitutivity, 347
- interval maximax, 376
- interval maximin, 376
- interval order, 342
- interval order property, 327, 341, 342
- invariant assertions method, 111
- IP-transitivity, 324, 328
- iteration operator, 113, 121, 197
- Johnson, S., 395
- just noticeable difference, 341
- KA, *see* Kleene algebra
- KAT, *see* Kleene algebra with tests
- Kleene algebra, 191–196
  - \*-continuous, 193
  - typed, 196
  - with tests, 194
- Knaster–Tarski theorem, 189
- knower, 386

- Kripke frame, 116, 155
  - nonstandard, 128, 129
- language
  - first-order DL, 148
- LDL, 181
- least fixpoint, 189
- LED, 183
- line graphs, 395
- linear
  - recurrence, 138
- linear-time TL, 184
- lines of identity, 415
- Logic of Effective Definitions, 183
- logical consequence, 119, 401
  - in PDL, 129–132
- logical equivalence, 112
- logically equivalent, 401
- logics for Ramsey test conditionals, 77
- $L_{\omega_1\omega}$ , 164
  - constructive, 164
- $L_{\omega_1^{\text{ck}}\omega}$ , 162, 164
- loop
  - invariance rule, 123
- loop**, 144, 181
- Löwenheim–Skolem theorem, 162
- lower bound
  - for PDL, 130
- LPDL, 144
- Luengo, I., 395
- $m$ -state, 168
- maximal change theories, 25
- maximax, 372
- maximin, 372
- meaning
  - function, 116, 155
- medians, 377
- min,+ algebra, 193
- minimal change theories, 9
- minimal regions, 396, 397
- modal logic, 100, 113, 115, 125
- modal  $\mu$ -calculus, *see*  $\mu$ -calculus
- modality, 412
- model, 398
- model checking, 127, 130
  - for the  $\mu$ -calculus, 191
- modus ponens, 129
- monadic concepts, 377
- monadic logic, 409
- monadic value concepts: classificatory, 319
- money-pump, 334
- mono-unary vocabulary, 148
- monotonicity, 189
- moral philosophy, 319
- More, T., 397
- $m^{\text{th}}$  spectrum, 169
- $\mu$  operator, 189
- $\mu$ -calculus, 143, 189, 190
- multiprocessor systems, 186
- mutual exclusiveness, 381
- Narayanan, N., 395
- natural chain, 202
- necessitation, 385
- negation-comparability, 381
- negation-related good, 380
- negativity, 378
- neutrality, 379
- nexttime operator, 185
- non-duplicity, 381
- non-monotonic logic, 319
- nondeterminism, 106, 198
  - binary, 179
  - unbounded, 179
- nondeterministic
  - assignment, *see* wildcard assignment
  - while** program, 150
- nonstandard
  - Kripke frame, 128, 129
- Nonstandard DL, 204
- NP
  - completeness, 131
- nullary
  - array variable, 156

- numerical representation, 340
- $\omega$ -automaton, 141
- output variable, 102
- parameterless recursion, 132, 154
- parentheses, 114
- partial
  - correctness, 110
  - assertion, 115, 167, 168, 199
- PDL, 111–147
  - automata, 141
  - concurrent, 146
  - regular, 113
  - rich test, 114
- Peano arithmetic
  - induction axiom of, 124
- Peirce closure, 410
- Peirce's Rule, 409
- Peirce, C., 395
- Peirce, C. S., 401, 402, 404, 406, 410–412, 416
- Peirce, C. S.<sub>i</sub>, 399
- Peirce–Venn diagrams, 395, 398, 406–409
  - tabular form, 408
- permissive cancellation, 388
- PI-transitivity, 324, 328, 345
- $\Pi_1^1$ 
  - completeness, 132
- PL, *see* Process Logic
- polyadic, 148
- poor
  - test, 103, 149
  - vocabulary, 148
- positional choice, 343
- positivity, 378
- possible worlds, 356, 385
- postcondition, 110
- precedence, 114
- precondition, 110
- predicate symbol, 148
- premissive predicates, 384
- prescriptive predicates, 384
- probabilistic program, 182–183
- Process Logic, 188
- program, 101, 148, 151
  - atomic, 103, 147, 149
  - DL, 149
  - operator, 103
  - probabilistic, 182–183
  - r.e., 152, 159
  - regular, 104, 117, 149
  - schematology, 201
  - variable, 151
  - while**, 105
  - with Boolean arrays, 203
- prohibitive predicates, 384
- propositional
  - satisfiability, *see* satisfiability
- Propositional Dynamic Logic, *see* PDL
- propositional logic, 410, 412
- pushdown
  - automaton, 132
  - store, *see* stack
- quasi-transitive, 324
- Quine, W. V. O., 409
- Ramsey test, 56
  - negative, 65
  - triviality results, 58
- random assignment, *see* assignment
- rational choice, 319
- RDL, 181
- r.e.
  - program, 152, 159
- reasoning
  - interpreted, 164, 173
  - uninterpreted, 161, 170
- recursion, 105, 153
  - parameterless, 132, 154
- recursive
  - call, 105
- reflexive transitive closure, 123
- reflexivity, 321, 348, 368
- region, 397

- regular
  - expression, 113, 117
  - program, 104, 117, 149
    - with arrays, 153
    - with Boolean stack, 177
    - with stack, 154
  - set, 117, 193
- relation
  - symbol, 148
- relational choice function, 344
- relative
  - completeness, 202
  - expressiveness, 176, 203
    - over  $\mathbb{N}$ , 166
- repeat**, 144, 181
- representation, 361
- representation function, 361
- requirement of temporal priority, 36
- resolution, 410
- rich
  - test, 103, 114, 149, 151
  - vocabulary, 148
- Roberts, D., 413, 414, 416, 421
- Ross's paradox, 386
- RPDL, 144
- satisfaction
  - PDL, 116
- satisfiability
  - algorithm for PDL, 125
  - Boolean, *see* satisfiability, propositional
  - DL, 161
  - PDL, 119, 125
  - propositional, 131
- satisfies, 401, 407
- scalar multiplication, 197
- schematology, 203
- SDPDL, 139
- Segerberg axioms, *see* axiomatization, PDL
- semantic determinacy, 139
- semantics
  - DL, 155–161
  - PDL, 115–117
- semi-transitivity, 327, 341
- semiorder, 341
- Sen, A., 343
- sentential representation, 347
- separable dynamic algebra, *see* dynamic algebra
- seq, 105, 118, 151
- Shin, S., 396–399, 402–404, 406, 408
- similarity, 364
- simple assignment, *see* assignment
- simple-minded
  - context-free language, 137
  - pushdown automaton, 137
- small change theories, 20
- small model
  - property, 125, 127, 132
  - theorem, 127, 130
- soundness
  - PDL, 119, 129
- Sowa, J., 395
- SPDL, 139
- specification
  - correctness, 107
  - input/output, 108, 110
- spectral
  - complexity, 168, 169
- spectral theorem, 203
- spectrum
  - first-order, 202
  - $m^{\text{th}}$ , 169
  - of a formula, 169
  - second-order, 202
- stack, 105, 152, 153
  - algebraic, 154
  - automaton, 137
  - Boolean, 154
  - higher-order, 202
  - operation, 153
- standard deontic logic, 385
- standard Kripke frame, 197
- \*-continuity, 193

- \*-continuous dynamic algebra, *see*
  - dynamic algebra
- state, 101, 116, 155, 157
  - Herbrand-like, 168
  - initial, 157, 163, 168
  - $m$ -, 168
- state diagrams, 395
- state transition diagrams, 395
- states of affairs, 347
- static logic, 162
- strict conditionals, 6
- strict part, 322
- strictly more expressive than, 134, 163
- structure
  - acceptable, 202
  - arithmetical, 165
  - expressive, 174
- subexpression relation, 125
- suborders, 342
- substitution
  - in DL formulas, 171
  - instance, 132
- summary, 402
- symbol
  - atomic, 113
  - constant, 148
  - equality, 148
  - function, 148
  - predicate, 148
  - relation, 148
- symmetric, 362
- symmetric part, 322
- symmetry, 321
- syntactic
  - continuity, 190
  - monotonicity, 189
- tail recursion, 105
- tautologies, 379
- Temporal Logic, 111, 184
- tense and conditional, 38
- term
  - dynamic, 181
- termination, 115
  - assertion, 170
- test, 103
  - atomic, 103
  - first-order, 204
  - operator, 113, 149
  - poor, *see* poor test
  - rich, *see* rich test
- top-transitivity, 337
- total
  - correctness, 110, 143, 170
  - assertion, 167
- totality account, 358
- totality approach, 371
- trace, 102
  - quantifier, 186
- transitivity, 323, 348, 368
- transmitted, 368, 372
- Turing machine
  - alternating, 130
- typed Kleene algebra, *see* Kleene algebra
- unbounded nondeterminism, 179
- uninterpreted reasoning, 161, 170
- until operator, 185
- utilitarianism, 353
- utility, 340
- validity
  - $\mathfrak{A}$ -, 161
  - DL, 161, 167
  - PDL, 119
- valuation, 102, 147, 155
- value function, 340
- variable, 101
  - array, 152
  - individual, 148, 151
  - program, 151
  - work, 102
- Venn diagrams, 395, 398
- Venn, J., 398
- Venn-type diagram, 397
- verification conditions, 173

virtual connectivity, 327  
vocabulary  
  first-order, 148  
  monadic, 148  
  mono-unary, 148  
  polyadic, 148  
  poor, 148  
  rich, 148  
  
weak centering, 362  
weak centring, 368  
weak preference, 321  
weakening rule, 124  
well-founded, 143  
well-foundedness, 181–182  
**wf**, 144, 181  
**while**  
  loop, 115, 139  
  operator, 103  
  program, 105, 139  
    deterministic, 150  
    nondeterministic, 150  
    with arrays, 153  
    with stack, 154  
  rule, 124  
White, R., 414  
wildcard assignment, *see* assignment  
work variable, 102  
worse, 320  
worst, 379  
WP, 139  
Wright, G. H. von, 319, 384  
  
Zeman, J., 414, 416, 421



# Handbook of Philosophical Logic

2nd Edition

Volume 5

edited by Dov M. Gabbay and F. Guentner



## CONTENTS

Editorial Preface	vii
<b>Dov M. Gabbay</b>	
Intuitionistic Logic	1
<b>Dirk van Dalen</b>	
Dialogues as a Foundation for Intuitionistic Logic	115
<b>Walter Felscher</b>	
Free Logics	147
<b>Ermanno Bencivenga</b>	
Advanced Free Logic	197
<b>Scott Lehmann</b>	
Partial Logic	261
<b>Stephen Blamey</b>	
Index	354



## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good.!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London

Logic	IT			
	Natural language processing	Program control specification, verification, concurrency	Artificial intelligence	Logic programming
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs



<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical frag- ments</b>	Basic back- ground lan- guage	Program syn- thesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely use- ful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calcu- lus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syn- tax	Modules. Combining languages	Logics of space and time	Combining fea- tures
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game seman- tics gaining ground		
<b>Object level/ metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action- revision mod- els</b>			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model



DIRK VAN DALEN

## INTUITIONISTIC LOGIC

### INTRODUCTION

Among these logics that deal with the familiar connectives and quantifiers two stand out as having a solid philosophical–mathematical justification. On the one hand there is a classical logic with its ontological basis and on the other hand intuitionistic logic with its epistemic motivation. The case for other logics is considerably weaker; although one may consider intermediate logics with more or less plausible principles from certain viewpoints none of them is accompanied by a comparably compelling philosophy. For this reason we have mostly paid attention to pure intuitionistic theories.

Since Brouwer, and later Heyting, considered intuitionistic reasoning, intuitionistic logic has grown into a discipline with a considerable scope. The subject has connections with almost all foundational disciplines, and it has rapidly expanded.

The present survey is just a modest cross-section of the presently available material. We have concentrated on a more or less semantic approach at the cost of the proof theoretic features. Although the proof theoretical tradition may be closer to the spirit of intuitionism (with its stress on *proofs*), even a modest treatment of the proof theory of intuitionistic logic would be beyond the scope of this chapter. The reader will find ample information on this particular subject in the papers of, e.g. Prawitz and Troelstra.

For the same reason we have refrained from going into the connection between recursion theory and intuitionistic logic. Section 8 provides a brief introduction to realizability.

Intuitionistic logic is, technically speaking, just a subsystem of classical logic; the matter changes, however, in higher-order logic and in mathematical theories. In those cases specific intuitionistic principles come into play, e.g. in the theory of choice sequences the meaning of the prefix  $\forall\xi\exists x$  derives from the nature of the mathematical objects concerned. Topics of the above kind are dealt with in Section 9.

The last sections touch on the recent developments in the area of categorical logic. We do not mention categories but consider a very special case. There has been an enormous proliferation in the semantics of intuitionistic second-order and higher-order theories. The philosophical relevance is quite often absent so that we have not paid attention to the extensive literature on independence results. For the same reason we have not incorporated the intuitionistic ZF-like systems.

Intuitionistic logic can be arrived at in many ways—e.g. physicalistic or materialistic—we have chosen to stick to the intuitionistic tradition in considering mathematics and logic as based on human mental activities. Not surprisingly, intuitionistic logic plays a role in constructive theories that do not share the basic principles of intuitionism, e.g. Bishop’s constructive mathematics. There was no room to go into the foundations of these alternatives to intuitionism. In particular we had to leave out Feferman’s powerful and elegant formalisations of operations and classes. The reader is referred to Beeson [1985] and Troelstra and van Dalen [1988] for this and related topics.

We are indebted for discussions and comments to C.P.J. Koymans, A.S. Troelstra and A. Visser.

## 1 A SHORT HISTORY

Intuitionism was conceived by Brouwer in the early part of the twentieth century when logic was still in its infancy. Hence we must view Brouwer’s attitude towards logic in the light of a rather crude form of theoretical logic. It is probably a sound conjecture that he never read Frege’s fundamental expositions and that he even avoided Whitehead and Russell’s *Principia Mathematica*. Frege was at the time mainly known in mathematical circles for his polemics with Hilbert and others, and one could do without the *Principia Mathematica* by reading the fundamental papers in the journals. Taking into account the limited amount of specialised knowledge Brouwer had of logic, one might well be surprised to find an astute appraisal of the role of logic in Brouwer’s Dissertation [Brouwer, 1907]. Contrary to most traditional views, Brouwer claims that logic does not precede mathematics, but, conversely, that logic depends on mathematics. The apparent contradiction with the existing practice of establishing strings of ‘logical’ steps in mathematical reasoning, is explained by pointing out that each of these steps represents a sequence of mathematical constructions. The logic, so to speak, is what remains if one takes away the specific mathematical constructions that lead from one stage of insight to the next.

Here it is essential to make a short excursion into the mathematical and scientific views that Brouwer held and that are peculiar to intuitionism. Mathematics, according to Brouwer, is a mental activity, sometimes described by him as the exact part of human thought. In particular, mathematical objects are mental constructions, and properties of these objects are established by, again, mental constructions. Hence, in this view, something holds for a person if he has a construction (or proof) that establishes it. Language does not play a role in this process but may be (and in practice: is) introduced for reasons of communication. ‘People try by means of sounds and symbols to originate in other copies of mathematical constructions and

reasonings which they have made themselves; by the same means they try to aid their own memory. In this way *mathematical language* comes into being, and as its special case *the language of logical reasoning*. The next step taken by man is to consider the language of logical reasoning mathematically, i.e. to study its mathematical properties. This is the birth of *theoretical logic*.

Brouwer's criticism of logic is two-fold. In the first place, logicians are blamed for giving logic precedence over mathematics, and in the second place, logic is said to be unreliable (Brouwer [1907; 1908]). In particular, Brouwer singled out the *principle of the excluded third* as incorrect and unjustified. The criticism of this principle is coupled to the criticism of Hilbert's famous dictum that 'each particular mathematical problem can be solved in the sense that the question under consideration can either be affirmed, or refuted' [Brouwer, 1975, pp. 101 and 109].

Let us, by way of example, consider Goldbach's Conjecture,  $G$ , which states that each even number is the sum of two odd primes. A quick check tells us that for small numbers the conjecture is borne out:  $12 = 5 + 7$ ,  $26 = 13 + 13$ ,  $62 = 3 + 59$ ,  $300 = 149 + 151$ . Since we cannot perform an infinite search, this simple method of checking can at best provide, with luck, a counter example, but not a proof of the conjecture. At the present stage of mathematical knowledge no proof of Goldbach's conjecture, or of its negation, has been provided. So can we affirm  $G \vee \neg G$ ? If so, we should have a construction that would decide which of the two alternatives holds and provide a proof for it. Clearly we are in no position to exhibit such a construction, hence we have no grounds for accepting  $G \vee \neg G$  as correct.

The undue attention paid to the principle of the excluded third, had the unfortunate historical consequence that the issues of the foundational dispute between the Formalists and the Intuitionists were obscured. An outsider might easily think that the matter was a dispute of two schools—one with, and one without, the *principle of the excluded third* (or *middle*), PEM for short. Brouwer himself was in no small degree the originator of the misunderstanding by choosing the far too modest and misleading title of 'Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten' for his first fundamental paper on intuitionistic mathematics. For the philosophical-mystical background of Brouwer's views, see [van Dalen, 1999a]; a foundational exposition can be found in [van Dalen, 2000].

The logic of intuitionism was not elaborated by Brouwer, although he proved its first theorem:  $\neg\varphi \leftrightarrow \neg\neg\neg\varphi$ .

The first mathematicians to consider the logic of intuitionism in a more formal way were Glivenko and Kolmogorov.

The first presented a fragment of propositional logic and the second a fragment of predicate logic. In 1928 Heyting independently formalised intuitionist predicate logic and the fundamental theories of arithmetic and 'set

theory' [Heyting, 1930]. For historical details, cf. Troelstra [1978; 1981]. Heyting's formalization opened up a new field to adventurous logicians, but it did not provide a 'standard' or 'intended' interpretation, thus lacking the inner coherence of a conceptual explanation. In a couple of papers (cf. [Heyting, 1934]), Heyting presented from 1931 on the interpretation that we have come to call the *proof-interpretation* (cf. [Heyting, 1956, Chapter VII]). The underlying idea traces back to Brouwer: the truth of a mathematical statement is established by a proof, hence the meaning of the logical connective has to be explained in terms of proofs and constructions (recall that a proof is a kind of construction). Let us consider one connective, by way of example: A proof of  $\varphi \rightarrow \psi$  is a construction which converts any proof of  $\varphi$  into a proof of  $\psi$ .

Note that this definition is in accord with the conception of mathematics (and hence logic) as a mental constructive activity. Moreover it does not require statements to be bivalent, i.e. to be either true or false. For example,  $\varphi \rightarrow \varphi$  is true independent of our knowledge of the truth of  $\varphi$ . The proof-interpretation provided at least an informal insight into the mysteries of intuitionistic truth, but it lacked the formal clarity of the notion of truth in classical logic with its completeness property.

An analogue of the classical notion of truth value was discovered by Tarski, Stone and others who had observed the similarities between intuitionistic logic and the closure operation of topology (cf. [Rasiowa and Sikorski, 1963]). This so-called *topological interpretation* of intuitionistic logic also covers a number of interpretations that at first sight might seem to be totally devoid of topological features. Among these are the lattice (like) interpretations of Jaskowski, Rieger and others, but also the more recent interpretations of Beth and Kripke. All these interpretations are grouped together as semantical interpretations, in contrast to interpretations that are based on algorithms, one way or another.

A breakthrough in intuitionistic logic was accomplished by Gentzen in 1934 in his system of *Natural Deduction* (and also his *calculus of sequents*), which embodied the meaning of the intuitionistic connectives far more accurately than the existing Hilbert-type formalizations. The eventual recognition of Gentzen's insights is to a large extent due to the efforts of Prawitz who reintroduced Natural Deduction, and considerably extended Gentzen's work [1965; 1971].

In the beginning of the thirties the first meta-logical results about intuitionistic logic and its relation to existing logics appeared. Gödel, and independently Gentzen, formulated a translation of classical predicate logic into a fragment of intuitionistic predicate logic, thus extending early work of Glivenko [Glivenko, 1929; Gentzen, 1933; Gödel, 1932].

Gödel also established the connection between the modal logic **S4** and intuitionistic logic [Gödel, 1932].



The period after the Second World War brought new researchers to intuitionistic logic and mathematics. In particular Kleene, who based an ‘effective’ interpretation of intuitionistic arithmetic on the notion of recursive function. His interpretation is known as *realizability* (Kleene [1952; 1973]). In 1956 Beth introduced a new semantic interpretation with a better foundational motivation than the earlier topological interpretations, and Kripke presented a similar, but more convenient interpretation in 1963 [Kripke, 1965]. These new semantics showed more flexibility than the earlier interpretations and lent themselves better to the model theory of concrete theories. General model theory in the lattice and topological tradition had already been undertaken by the Polish school (cf. [Rasiowa and Sikorski, 1963]).

In the meantime Gödel had presented his Dialectica Interpretation [1958], which like Kleene’s realizability, belongs to the algorithmic type of interpretations. Both the realizability and the Dialectica Interpretation have shown to be extremely fruitful for the purpose of Proof Theory.

Another branch at the tree of semantic interpretations appeared fairly recently, when it was discovered that sheaves and topoi present a generalisation of the topological interpretations [Goldblatt, 1979; Troelstra and van Dalen, 1988].

The role of a formal semantics will be expounded in Section 3. Its most obvious and immediate use is the establishing of underivability results in a logical calculus. However, even before a satisfactory semantics was discovered, intuitionists used to show that certain classical theorems were not valid by straightforward intuitive methods. We will illustrate the naive approach for two reasons. In the first place it is direct and the first thing one would think of, in the second place it has its counterparts in formal semantics and can be useful as a heuristics.

The traditional counterexamples are usually formulated in terms of a particular unsolved problem. The problem in the following example goes back to Brouwer. Consider the decimal expansion of  $\pi : 3,14\dots$ , hardly anything is known about regularities in this expansion, e.g. it is not known if it contains a sequence of 9 nines. Let  $A(n)$  be the statement ‘the  $n$ th decimal of  $\pi$  is a nine and it is preceded by 8 nines’.

1. The principle of the excluded third is not valid.  
Suppose  $\exists xA(x) \vee \neg\exists xA(x)$ , then we would have a proof that either provides us with a natural number  $n$  such that  $A(n)$ , or that shows us that no such  $n$  exists. Since there is no such evidence available we cannot accept the principle of the excluded third.
2. The double negation principle is not valid. Observe that  $\neg\neg(\exists xA(x) \vee \neg\exists xA(x))$  holds. In general the double negation of the principle of the excluded third holds, since  $\neg\neg(\varphi \vee \neg\varphi)$  is equivalent to  $\neg(\neg\varphi \wedge \neg\neg\varphi)$  and the latter is correct on the intuitive interpretations.

Since  $\exists xA(x) \vee \neg\exists xA(x)$  does not hold, we see that  $\neg\neg\varphi \rightarrow \varphi$  is not valid.

### 3. One version of De Morgan's Law fails.

The suspect case is  $\neg(\varphi \wedge \psi) \rightarrow \neg\varphi \vee \neg\psi$ , since its conclusion is strong and its premise is weak. Consider  $\neg(\neg\exists xA(x) \wedge \exists xA(x)) \rightarrow \neg\neg\exists xA(x) \vee \neg\exists xA(x)$ . The premise is true, but the conclusion cannot be asserted, since we do not know if it is impossible that there is no sequence of 9 nines or it is impossible that there is such a sequence.

Counterexamples of the above kind show that our present state of knowledge does not permit us to affirm certain logical statements that are classically true. They represent evidence of implausibility, all the same it is not the strongest possible result. Of course we cannot expect to establish the negation of the principle of the excluded third because that is a downright contradiction. By means of certain strong intuitionistic, or alternatively algorithmic, principles one can establish a strongly non-classical theorem like  $\neg\forall x(\varphi(x) \vee \neg\varphi(x))$  for a suitable  $\varphi(x)$ .

We will now present an informal version of the proof interpretation. For convenience we will suppose that the variables of our language range over natural numbers. This is not strictly necessary, but it suffices to illustrate the working of the interpretation. Recall that we understand the primitive notion 'a is a proof of  $\varphi$ ', where a proof is a particular kind of (mental) construction. We will now proceed to explain what it means to have a proof of a non-atomic formula  $\varphi$  in terms of proofs of its components.

- (i)  $a$  is a proof of  $\varphi \wedge \psi$  iff  $a$  is a pair  $(a_1, a_2)$  such that  $a_1$  is a proof of  $\varphi$  and  $a_2$  is a proof of  $\psi$ .
- (ii)  $a$  is a proof of  $\varphi \vee \psi$  iff  $a$  is a pair  $(a_1, a_2)$  such that  $a_1 = 0$  and  $a_2$  is a proof of  $\varphi$  or  $a_1 = 1$  and  $a_2$  is a proof of  $\psi$ .
- (iii)  $a$  is a proof of  $\varphi \rightarrow \psi$  iff  $a$  is a construction that converts each proof  $b$  of  $\varphi$  into a proof  $a(b)$  of  $\psi$ .
- (iv) nothing is a proof of  $\perp$  (falsity).
- (v)  $a$  is a proof of  $\exists x\varphi(x)$  iff  $a$  is a pair  $(a_1, a_2)$  such that  $a_1$  is a proof of  $\varphi(a_2)$ .
- (vi)  $a$  is a proof of  $\forall x\varphi(x)$  iff  $a$  is a construction such that for each natural number  $n$ ,  $a(n)$  is a proof of  $\varphi(\bar{n})$ .

Note that intuitionists consider  $\neg\varphi$  as an abbreviation for  $\varphi \rightarrow \perp$ . The clause that a trained logician will immediately look for is the one dealing with the atomic case. We cannot provide a definition for that case since it

must essentially depend on the specific theory under consideration. In the case of ordinary arithmetic the matter is not terribly important as the closed atoms are decidable statements of the form  $5 = 7 + 6$ ,  $23.16 = 5(3 + 2.8)$ , etc. We can ‘start’ the definition in a suitable fashion.

REMARK. If one wishes to preserve the feature that from a proof one can read off the result, then some extra care has to be taken, e.g. according to clause (iii)  $(0, p)$  proves  $\varphi \vee \psi$  for all possible  $\psi$ , where  $\varphi$  is a proof of  $\varphi$ . One may beef up the ‘proof’ by adding the disjunction to it: replace  $(0, p)$  by  $(0, p, \varphi \vee \psi)$ , etc.

The above version is due to Heyting (cf. [Heyting, 1956; Troelstra, 1981]). Refinements have been added by Kreisel for the clauses involving the implication and universal quantification [Kreisel, 1965]. His argument being: the definition contains a part that is not immediately seen to be of the ultimate simple and lucid form we wish it to be. In particular one could ask oneself ‘does this alleged construction do what it purports to do?’ For this reason Kreisel modified clause (iii) as follows:  $a$  is a proof of  $\varphi \rightarrow \psi$  iff  $a$  is a pair  $(a_1, a_2)$  such that  $a_1$  is a construction that converts any proof  $b$  of  $\varphi$  into a proof  $a_1(b)$  of  $\psi$ , and  $a_2$  is a proof of the latter fact. A similar modification is provided for (vi). The situation is akin to that of the correctness of computer programs. In particular we use Kreisel’s clause if we want the relation ‘ $a$  is a proof of  $\varphi$ ’ to be decidable. Clauses (iii) and (vi) clearly do not preserve decidability, moreover they do not yield ‘logic free’ conditions.

It must be pointed out however that the decidability of the proof-relations has been criticised and that the ‘extra clauses’ are not universally accepted.

Sundholm [1983] contains a critical analysis of the various presentations of the ‘proof interpretation’. In summing up the views of Brouwer, Heyting and Kreisel, he notes a certain confusion in terminology. In particular he points out that constructions (in particular proofs) can be viewed as *processes* and differ from the resulting construction-object. The latter is a mathematical object, and can be operated upon, not so the former. The judgements at the right-hand side, explaining the meaning of the logical constants, are taken by Kreisel to be mathematical objects, a procedure that is objected to by Sundholm. indeed, on viewing the judgement ‘ $a$  converts each proof of  $\varphi$  into a proof of  $\psi$ ’ as extra-mathematical, the need for a second clause disappears.

In Beeson [1979] a theory of constructions and proofs is presented violating the decidability of the proof relation. Troelstra and Diller [1982] study the relation between the proof interpretation and Martin-Löf’s type theory.

The proofs inductively defined above are called *canonical* by Martin-Löf, Prawitz and others. Of course there are also non-canonical proofs, and some of them are preferable to canonical ones. Consider, e.g.  $10^{11} + 11^{10} = 11^{10} + 10^{11}$  in arithmetic. One knows how to get a canonical proof: by simply carrying out the addition according to the basic rules  $(x + 0 = x$

and  $x + Sy = S(x + y)$ , where  $S$  is the successor function). An obvious non-canonical (and shorter) proof would be: first show  $\forall xy(x + y = y + x)$  by mathematical induction and then specialise.

We will now proceed to illustrate the rules in use.

$$(1) \quad (\varphi \wedge \psi \rightarrow \sigma) \rightarrow (\varphi \rightarrow (\psi \rightarrow \sigma)).$$

Let  $a$  be a proof of  $\varphi \wedge \psi \rightarrow \sigma$ , i.e.  $a$  is a construction that converts any proof  $(b, c)$  of  $\varphi \wedge \psi$  into a proof  $a((b, c))$  of  $\sigma$ . We want a proof of  $\varphi \rightarrow (\psi \rightarrow \sigma)$ . So let  $p$  be a proof of  $\varphi$  and  $q$  a proof of  $\psi$ .

Define a construction  $k$  such that  $k(p)$  is a proof of  $\psi \rightarrow \sigma$ , i.e.  $(k(p))(q)$  is a proof of  $\sigma$ . Evidently we should put  $(k(p))(q) = a((p, q))$ ; so, using the functional abstraction operator,  $k(p) = \lambda q.a((p, q))$  and  $k = \lambda p.\lambda q.a((p, q))$ . The required proof is a construction that carries  $a$  into  $k$ , i.e.  $\lambda apq.a((p, q))$ .

$$(2) \quad \neg(\varphi \vee \psi) \rightarrow (\neg\varphi \wedge \neg\psi).$$

Let  $a$  be a proof of  $\neg(\varphi \vee \psi)$ , a construction that carries a proof of  $\varphi \vee \psi$  into a proof of  $\perp$ . Suppose now that  $p$  is a proof of  $\varphi$ , then  $(0, p)$  is a proof of  $\varphi \vee \psi$ , and hence  $a((0, p))$  is a proof of  $\perp$ . So  $\lambda p.a((0, p))$  is a proof of  $\neg\varphi$ . Likewise  $\lambda q.a((1, q))$  is a proof of  $\neg\psi$ . By definition  $(\lambda p.a((0, p)), \lambda q.a((1, q)))$  is a proof of  $\neg\varphi \wedge \neg\psi$ . So the construction that carries  $a$  into  $(\lambda p.a((0, p)), \lambda q.a((1, q)))$ , i.e.  $\lambda a.(\lambda p.a((0, p)), \lambda q.a((1, q)))$ , is the required proof.

$$(3) \quad \exists x\neg\varphi(x) \rightarrow \neg\forall x\varphi(x).$$

Let  $(a_1, a_2)$  be a proof of  $\exists x\neg\varphi(x)$ , i.e.  $a_1$  is a proof of  $\varphi(\bar{a}_2) \rightarrow \perp$ . Suppose  $p$  is a proof of  $\forall x\varphi(x)$ , then in particular  $p(a_2)$  is a proof of  $\varphi(\bar{a}_2)$ , and hence  $a_1(p(a_2))$  is a proof of  $\perp$ . So  $\lambda p.a_1(p(a_2))$  is a proof of  $\neg\forall x\varphi(x)$ . Therefore  $\lambda(a_1, a_2)\lambda p.a_1(p(a_2))$  is the required proof.

The history of intuitionistic logic is not as stirring as the history of intuitionism itself. The logic itself was not controversial, Heyting's formalization showed it to be a subsystem of classical logic. Moreover, it convinced logicians that there was a coherent notion of 'constructive reasoning'. In the following sections we will show some of the rich structure of this logic. One problem in intuitionistic logical theories is how to codify and exploit typically intuitionistic principles. These are to be found in particular in the second-order theories where the concepts of set (species) and function play a role.

Despite Brouwer's scorn for logic, some of the finer distinctions that are common today were introduced by him. In his thesis we can already find the fully understood notions of language, logic, metalanguage, metalogic, etc. (cf. Brouwer [1907; 1975]).

The Brouwer–Hilbert controversy seems from our present viewpoint to be one of those deplorable misunderstandings. Hilbert wanted to justify by

metamathematical means the mathematics of infinity with all its idealizations. He considered mathematics as based on the bedrock of its finitistic part, which is just a very concrete part of intuitionistic mathematics. The latter transcends finitism by its introduction of abstract notions, such as *set* and *sequence*.

## 2 PROPOSITIONAL AND PREDICATE LOGIC

The syntax of intuitionistic logic is identical to that of classical logic (cf. Wilfrid Hodges' chapter in Volume 1 of this *Handbook*). As in classical logic, we have the choice between a formalisation in a Hilbert-type system or in a Gentzen-type system. Heyting's original formalisation used the first kind. We will exhibit a Hilbert-type system first.

### 2.1 An Axiom System for Intuitionistic Logic

#### *Axioms*

1.  $\varphi \rightarrow (\psi \rightarrow \varphi)$
2.  $(\varphi \rightarrow \psi) \rightarrow ((\varphi \rightarrow (\psi \rightarrow \sigma)) \rightarrow (\varphi \rightarrow \sigma))$
3.  $\varphi \rightarrow (\psi \rightarrow \varphi \wedge \psi)$
4.  $\varphi \wedge \psi \rightarrow \varphi \quad \varphi \wedge \psi \rightarrow \psi$
5.  $\varphi \rightarrow \varphi \vee \psi \quad \psi \rightarrow \varphi \vee \psi$
6.  $(\varphi \rightarrow \sigma) \rightarrow ((\psi \rightarrow \sigma) \rightarrow (\varphi \vee \psi \rightarrow \sigma))$
7.  $(\varphi \rightarrow \psi) \rightarrow ((\varphi \rightarrow \neg\psi) \rightarrow \neg\varphi)$
8.  $\varphi(t) \rightarrow \exists x\varphi(x)$
9.  $\forall x\varphi(x) \rightarrow \varphi(t)$
10.  $\varphi \rightarrow (\neg\varphi \rightarrow \psi)$

#### *Rules*

Modus Ponens

$$\frac{\varphi \quad \varphi \rightarrow \psi}{\psi}$$

## Quantifier rules

$$\frac{\varphi \rightarrow \psi(x)}{\varphi \rightarrow \forall x\psi(x)}$$

$$\frac{\varphi(x) \rightarrow \psi}{\exists x\varphi(x) \rightarrow \psi}$$

The quantifier axioms and rules are subject to the usual variable conditions:  $t$  is free for  $x$  and  $x$  does not occur free in  $\psi$ .

The deducibility relation,  $\vdash$ , is defined as in Hodges' chapter (Vol. 1) of the *Handbook*. As in classical logic, we have the *Deduction theorem*:

$$\psi_1, \dots, \psi_n \vdash \varphi \Leftrightarrow \psi_1, \dots, \psi_{n-1} \vdash \psi_n \rightarrow \varphi.$$

If we add to the axioms the principle of the excluded third,  $\varphi \vee \neg\varphi$ , or the double negation principle,  $\neg\neg\varphi \rightarrow \varphi$ , we obtain the familiar classical logic.

We should note that the axioms contain all connectives, and not, as in classical logic, just  $\vee$ ,  $\neg$  and  $\exists$  (or whatever your favourite choice may be). The reason is that the definability of the connectives in terms of some of them (Hodges Chapter in Volume 1 of this *Handbook*) fails, as we will see later.

Since intuitionistic logic is more of an epistemic than of an ontological nature, we will study it mainly by means of Gentzen's Natural Deduction, as this latter system reflects the specific constructive reasoning of the intuitionist best.

This particular system has only rules and no axioms. The simplest rules have the form  $\frac{\dots}{\varphi}$ , and are to be read as  $\varphi$  follows (immediately) from the premises above the line. Some of the rules, however, involve manipulations with the so-called *assumptions*. The prime example is the rule that corresponds to the deduction theorem in Hilbert-type systems. Suppose we can derive  $\psi$  by means of a derivation  $\mathcal{D}$  from a number of assumptions among which is a formula  $\varphi$ , then we can derive  $\varphi \rightarrow \psi$  from the mentioned assumptions *without*  $\varphi$ . We denote this by

$$\frac{\begin{array}{c} [\varphi] \\ \mathcal{D} \\ \psi \end{array}}{\varphi \rightarrow \psi}$$

we say that the assumption  $\varphi$  is *cancelled*, this is indicated by the use of square brackets.

It appears to be convenient to employ a choice of connectives that includes  $\perp$  and excludes  $\neg$ . Of course  $\neg\varphi$  can be introduced as an abbreviation for  $\varphi \rightarrow \perp$ . We will also use the traditional abbreviation  $\varphi \leftrightarrow \psi$ .

The rules come in two kinds, *Introduction rules* and *Elimination rules*.

<i>Introduction rules</i>	<i>Elimination rules</i>
$\wedge I$ $\frac{\varphi \quad \psi}{\varphi \wedge \psi}$	$\wedge E$ $\frac{\varphi \wedge \psi}{\varphi} \quad \frac{\varphi \wedge \psi}{\psi}$
$\vee I$ $\frac{\varphi}{\varphi \vee \psi} \quad \frac{\psi}{\varphi \vee \psi}$	$\vee E$ $\frac{\begin{array}{c} [\varphi] \\ \mathcal{D}_1 \end{array} \quad \begin{array}{c} [\psi] \\ \mathcal{D}_2 \end{array}}{\varphi \vee \psi} \quad \frac{\sigma \quad \sigma}{\sigma}$
$\rightarrow I$ $\frac{\begin{array}{c} [\varphi] \\ \mathcal{D} \\ \psi \end{array}}{\varphi \rightarrow \psi}$	$\rightarrow E$ $\frac{\varphi \quad \varphi \rightarrow \psi}{\psi}$
$\forall I$ $\frac{\varphi(x)}{\forall x\varphi(x)}$	$\forall E$ $\frac{\forall x\varphi(x)}{\varphi(t)}$
$\exists I$ $\frac{\varphi(t)}{\exists x\varphi(x)}$	$\exists E$ $\frac{\begin{array}{c} [\varphi(y)] \\ \mathcal{D} \end{array} \quad \sigma}{\exists x \quad \sigma}$
	$\perp$ $\frac{\perp}{\varphi}$

For the quantifier rules we have to add a few conditions: in the rules  $\exists I$  and  $\forall E$ ,  $t$  has to be ‘free for  $x$ ’. An application of  $\forall I$  is allowed only if the variable  $x$  does not occur in any of the assumptions in the derivation of  $\varphi(x)$ . Similarly the free variable  $y$  in the cancelled formula  $\varphi(y)$  may not occur free in  $\sigma$  or any of the assumptions in the right-hand derivation of  $\sigma$  (in  $\exists E$ ).

The rules of Gentzen’s system of Natural Deduction are intended to represent the meaning of connectives as faithfully as possible (cf. [Gentzen, 1935] or [Szabo, 1969, p. 74]). Gentzen’s goals have recently been made more precise in [Dummett, 1973] and [Prawitz, 1977]. We will set ourselves a specific goal by showing that the natural deduction rules are in accordance with the meaning of the logical connectives as put forward in Heyting’s proof interpretation.

We will consider a few representative cases.

$$\wedge I : \frac{\varphi_1 \quad \varphi_2}{\varphi_1 \wedge \varphi_2}.$$

Let proofs  $p_i$  of  $\varphi_i$  be given. Then we can form the ordered pair  $(p_1, p_2)$  which is a proof of  $\varphi_1 \wedge \varphi_2$ . This is the step that, given canonical proofs of the conjuncts, provides the canonical proof of the conjunction.

$$\wedge E : \frac{\varphi_1 \wedge \varphi_2}{\varphi_i}.$$

Given a canonical proof  $p$  of  $\varphi_1 \wedge \varphi_2$ , we know that it must be an ordered pair  $(p_1, p_2)$ . The projection  $\pi_i$  yields the required canonical proof of  $\varphi_i$ .

$$\rightarrow I : \frac{\begin{array}{c} [\varphi] \\ \mathcal{D} \\ \psi \end{array}}{\varphi \rightarrow \psi}.$$

Suppose that we have a proof of  $\psi$  under a number of assumptions, including  $\varphi$ . Then this proof, when supplemented by a proof of  $\varphi$  yields a proof of  $\psi$ , i.e. we have a construction that transforms any proof of  $\varphi$  into a proof of  $\psi$ , but that means that we have a proof of  $\varphi \rightarrow \psi$ .

$$\forall I : \frac{\varphi(x)}{\forall x \varphi(x)}.$$

Suppose that we have a proof of  $\varphi(x)$ , i.e. a proof schema, that for each instance  $\varphi(n)$  of  $\varphi(x)$  yields a proof of it. Since  $x$  does not occur in the assumptions, the proof is uniform in  $x$ , i.e. it is a method for converting  $n$  into a proof of  $\varphi(n)$ . Again we have found a proof of  $\forall x \varphi(x)$ , along the lines of Heyting's interpretation.

The reader will now be able to continue this line of argument. We will only dwell for a moment on the *ex falso* rule.

$$\perp : \frac{\perp}{\varphi}.$$

The justification in terms of constructions is not universally accepted, e.g. [Johansson, 1936] rejected the rule and formulated his so-called minimal logic, which has the same rules as intuitionistic logic with deletion of the *ex falso* rule.

Now,  $\perp$  has, in the intuitionistic conception, no proof. What we have to provide is a construction that automatically yields for every proof of  $\perp$  a



proof of  $\varphi$ . Nothing is simpler; take for example the identity construction  $i : p \mapsto p$ ,  $i$  promises to give a proof of  $\varphi$  as output as soon as it gets a proof of  $\perp$  as input. Obviously,  $i$  keeps its promise because it is never asked to fulfill it.

Note that there is an alternative way of looking at the Natural Deduction system, we could consider it as a concrete illustration of Heyting's proof interpretation. For instance, the actual formal derivations *are* the proofs and/or constructions. In that sense they realized Heyting's clauses.

Let us, by way of illustration, make a few derivations.

1.  $(\varphi \rightarrow \psi) \rightarrow ((\psi \rightarrow \sigma) \rightarrow \varphi \rightarrow \sigma)$

$$\begin{array}{c}
 \frac{{}^1[\varphi] \quad [\varphi \rightarrow \psi]^3}{\psi \quad [\psi \rightarrow \sigma]^2} \rightarrow E \\
 \frac{\sigma}{\varphi \rightarrow \sigma} \rightarrow I \\
 (1) \quad \frac{\sigma}{\varphi \rightarrow \sigma} \rightarrow I \\
 (2) \quad \frac{\psi \quad [\psi \rightarrow \sigma]^2}{\psi \rightarrow \sigma} \rightarrow E \\
 (3) \quad \frac{(\psi \rightarrow \sigma) \rightarrow (\varphi \rightarrow \sigma)}{(\varphi \rightarrow \psi) \rightarrow ((\psi \rightarrow \sigma) \rightarrow (\varphi \rightarrow \sigma))} \rightarrow I
 \end{array}$$

2. By substitution of  $\perp$  for  $\sigma$  we obtain the law of contraposition  $(\varphi \rightarrow \psi) \rightarrow (\neg\psi \rightarrow \neg\varphi)$ .

3.  $\varphi \rightarrow \neg\neg\varphi$

$$\begin{array}{c}
 \frac{{}^1[\neg\varphi] \quad [\varphi]^2}{\perp} \rightarrow E \quad (\text{recall that } \neg\varphi \text{ stands for } \varphi \rightarrow \perp) \\
 (1) \quad \frac{\perp}{\neg\neg\varphi} \\
 (2) \quad \frac{\perp}{\varphi \rightarrow \neg\neg\varphi} \rightarrow I
 \end{array}$$

4.  $\neg\neg\neg\varphi \rightarrow \neg\varphi$

$$\begin{array}{c}
 \frac{{}^1[\neg\varphi] \quad [\varphi]^2}{\perp} \\
 (1) \quad \frac{\perp}{\neg\neg\neg\varphi} \quad [\neg\neg\neg\varphi]^3 \\
 (2) \quad \frac{\perp}{\neg\varphi} \\
 (3) \quad \frac{\perp}{\neg\neg\neg\varphi \rightarrow \neg\varphi}
 \end{array}$$

5. From 3. we get  $\neg\varphi \rightarrow \neg\neg\neg\varphi$ , combining this with 4. we have  $\neg\varphi \leftrightarrow \neg\neg\neg\varphi$ .
6.  $\neg\neg\forall x\varphi(x) \rightarrow \forall x\neg\neg\varphi(x)$

$$\frac{\frac{^1[\forall x\varphi(x)]}{\varphi(x)} \quad [\neg\varphi(x)]^2}{\perp}$$

$$(1) \frac{\perp}{\neg\neg\forall x\varphi(x) \quad [\neg\neg\forall x\varphi(x)]^3}$$

$$(2) \frac{\perp}{\neg\neg\varphi(x)}$$

$$(3) \frac{\forall x\neg\neg\varphi(x)}{\neg\neg\forall x\varphi(x) \rightarrow \forall x\neg\neg\varphi(x)}$$

7.  $\neg(\varphi \vee \psi) \leftrightarrow (\neg\varphi \wedge \neg\psi)$

$$\frac{\frac{[\varphi]^1}{\varphi \vee \psi} \quad [\neg(\varphi \vee \psi)]^3}{\neg\varphi} \quad \frac{\frac{[\psi]^2}{\varphi \vee \psi} \quad [\neg(\varphi \vee \psi)]^3}{\neg\psi}$$

$$(1) \quad (2)$$

$$(3) \frac{\neg\varphi \wedge \neg\psi}{\neg(\varphi \vee \psi) \rightarrow \neg\varphi \wedge \neg\psi}$$

The arrow from right to left is trivial.

8.  $\varphi \vee \neg\varphi$  and  $\neg\neg\varphi \rightarrow \varphi$  are equivalent as schema's, i.e. all instances of PEM follow from all instances of the double negation principle and vice versa. We will consider one direction. the proof requires a number of derivations, each of which is simple.

(a)  $\vdash \neg\neg(\varphi \vee \neg\varphi)$  (use (7))

$$(b) \frac{\neg\neg(\varphi \vee \neg\varphi) \rightarrow (\varphi \vee \neg\varphi) \quad \mathcal{D}}{\varphi \vee \neg\varphi} \quad \frac{\mathcal{D}}{\neg\neg(\varphi \vee \neg\varphi)}$$

where  $\mathcal{D}$  is a derivation obtained in (a).

The other direction is left to the reader.

The following list of provable statements will come in handy (relevant variables are shown)

1.  $\varphi \rightarrow \neg\neg\varphi$
2.  $\neg\varphi \leftrightarrow \neg\neg\neg\varphi$
3.  $\neg(\varphi \wedge \neg\varphi)$

4.  $\neg\neg(\varphi \vee \neg\varphi)$
5.  $\neg(\varphi \vee \psi) \leftrightarrow \neg\varphi \wedge \neg\psi$
6.  $(\varphi \vee \neg\varphi) \rightarrow (\neg\neg\varphi \rightarrow \varphi)$
7.  $(\varphi \rightarrow \psi) \rightarrow \neg(\varphi \wedge \neg\psi)$
8.  $(\varphi \rightarrow \neg\psi) \leftrightarrow \neg(\varphi \wedge \psi)$
9.  $(\neg\neg\varphi \wedge \neg\neg\psi) \leftrightarrow \neg\neg(\varphi \wedge \psi)$
10.  $(\neg\neg\varphi \rightarrow \neg\neg\psi) \leftrightarrow \neg\neg(\varphi \rightarrow \psi)$
11.  $(\neg\neg\varphi \rightarrow \psi) \rightarrow (\neg\psi \rightarrow \neg\varphi)$
12.  $\exists x\neg\varphi(x) \rightarrow \neg\forall x\varphi(x)$
13.  $\neg\exists x\varphi(x) \leftrightarrow \forall x\neg\varphi(x)$
14.  $\varphi \vee \forall x\psi(x) \rightarrow \forall x(\varphi \vee \psi(x))$
15.  $\forall x(\varphi \rightarrow \psi(x)) \leftrightarrow (\varphi \rightarrow \forall x\psi(x))$
16.  $\forall x(\varphi(x) \rightarrow \psi) \leftrightarrow (\exists\varphi(x) \rightarrow \psi)$
17.  $\exists x(\varphi \rightarrow \psi(x)) \rightarrow (\varphi \rightarrow \exists x\psi(x))$
18.  $\neg\neg\forall x\varphi(x) \rightarrow \forall x\neg\neg\varphi(x)$ .

Furthermore, conjunction and disjunction have the familiar associative, commutative and distributive properties.

For counterexamples to invalid propositions and sentences see Section 3.11.

The systems of intuitionistic *propositional* and *predicate* (or *quantificational*) logic are, without consideration of their formalisations, denoted by **IPC** and **IQC**.

Derivability will pedantically be denoted by  $\Gamma \vdash_{\text{IPC}} \varphi$  (resp.  $\Gamma \vdash_{\text{IQC}} \varphi$ ), or **IPC**  $\vdash \varphi$  (resp. **IQC**  $\vdash \varphi$ ), for empty  $\Gamma$ . When no confusion arises, we will however delete the subscripts. The derivations are in tree form, but one can easily represent them in linear form (cf. [Prawitz, 1965, p. 89 ff]). The present form, however, is more suggestive and since there is nothing sacrosanct about linearity we will stick to Gentzen's notation.

There is, nonetheless, a good reason for a more complete notation that makes the cancellation of assumptions explicit.

As usual, we write  $\Gamma \vdash \varphi$  for 'there is a derivation of  $\varphi$  from uncanceled assumptions that belong to the set  $\Gamma$ '. The rules of natural deduction can be formulated in terms of  $\vdash$ . For convenience we write  $\Gamma, \varphi_1, \dots, \varphi_n$  for  $\Gamma \cup \{\varphi_1, \dots, \varphi_n\}$  and  $\Gamma, \Delta$  for  $\Gamma \cup \Delta$ .

The following facts follow immediately from our rules:

1.  $\Gamma \vdash \varphi$  if  $\varphi \in \Gamma$
2.  $\Gamma \vdash \varphi$  and  $\Delta \vdash \psi \Rightarrow \Gamma, \Delta \vdash \varphi \wedge \psi$
3.  $\Gamma \vdash \varphi \wedge \psi \Rightarrow \Gamma \vdash \varphi$   
 $\Gamma \vdash \varphi \wedge \psi \Rightarrow \Gamma \vdash \psi$
4.  $\Gamma \vdash \varphi \Rightarrow \Gamma \vdash \varphi \vee \psi$   
 $\Gamma \vdash \psi \Rightarrow \Gamma \vdash \varphi \vee \psi$
5.  $\Gamma \vdash \varphi \vee \psi$  and  $\Delta, \varphi \vdash \sigma$  and  $\Delta', \psi \vdash \sigma \Rightarrow \Gamma, \Delta, \Delta' \vdash \sigma$
6.  $\Gamma, \varphi \vdash \psi \Rightarrow \Gamma \vdash \varphi \rightarrow \psi$
7.  $\Gamma \vdash \varphi$  and  $\Delta \vdash \varphi \rightarrow \psi \Rightarrow \Gamma, \Delta \vdash \psi$
8.  $\Gamma \vdash \perp \Rightarrow \Gamma \vdash \varphi$
9.  $\Gamma \vdash \varphi(x) \Rightarrow \Gamma \vdash \forall x\varphi(x)$ , where  $x$  is not free in  $\Gamma$
10.  $\Gamma \vdash \forall x\varphi(x) \Rightarrow \Gamma \vdash \varphi(t)$
11.  $\Gamma \vdash \varphi(t) \Rightarrow \Gamma \vdash \exists x\varphi(x)$
12.  $\Gamma \vdash \exists x\varphi(x)$  and  $\Delta, \varphi(y) \vdash \sigma \Rightarrow \Gamma, \Delta \vdash \sigma$ , where  $y$  is not free in  $\Delta$  and  $\sigma$ .

The above presentation of natural deduction can be viewed as a kind of sequent calculus, cf. [Troelstra and Schwichtenberg, 1996, §2.1.4]

We can now turn the tables and define  $\Gamma \vdash \varphi$  inductively by the preceding clauses.  $D$  is the least class of pairs  $(\Gamma, \varphi)$  (denoted by  $\Gamma \vdash \varphi$ ) such that

$$\begin{aligned} &\Gamma \vdash \varphi \in D \text{ if } \varphi \in \Gamma \\ &\Gamma \vdash \varphi \in D; \Delta \vdash \psi \in D \Rightarrow \Gamma, \Delta \vdash \varphi \wedge \psi \in D \\ &\vdots \\ &\Gamma \vdash \exists x\varphi(x) \in D; \Delta, \varphi(y) \vdash \sigma \in D \Rightarrow \Gamma, \Delta \vdash \sigma \in D, \\ &\text{where } y \text{ is not free in } \Delta \text{ and } \sigma. \end{aligned}$$

Observe that a derivation in  $D$  corresponds to a derivation in tree form, as presented before. The linearisation of natural deduction derivations that some authors have practised obscures the perspicuity of the derivations and we will stick to the tree form (remember what Frege said about ‘the convenience of the printer’).

EXAMPLE 1. Take the string

$$\begin{aligned} \varphi, \psi, \sigma \vdash \varphi & \quad (\text{by 1}) \\ \psi, \sigma \vdash \varphi \rightarrow \varphi & \quad (\text{by 6}) \\ \psi \vdash \sigma \rightarrow (\varphi \rightarrow \varphi) & \quad (\text{by 6}) \end{aligned}$$

It shows that  $\psi \vdash \sigma \rightarrow (\varphi \rightarrow \varphi)$  and we can recover the derivation in tree form from it:

$$\begin{array}{lcl}
 \text{first derivation} & \varphi & \text{second derivation} \quad \frac{[\varphi]}{\varphi \rightarrow \varphi} \\
 \\
 \text{third derivation} & \frac{[\varphi]}{\varphi \rightarrow \varphi} & \\
 & \frac{\quad}{\sigma \rightarrow (\varphi \rightarrow \varphi)} & 
 \end{array}$$

All this calls for some clarification.

1. The matter of cancellation is somewhat delicate, you don't have to cancel *all* occurrences of the relevant formula, not even *any* occurrence. This is made explicit in, e.g. rule 6,  $\Gamma, \varphi \vdash \psi \Rightarrow \Gamma \vdash \varphi \rightarrow \psi$ .  $\Gamma$  may still contain  $\varphi$ .
2. The tree derivation shows only the assumptions that actually play a role, but in  $\Gamma \vdash \varphi$  there may be lots of superfluous assumptions (infinitely many if you wish!).

It is for example quite simple to show, on the basis of the rules 1–12  $\Gamma \vdash \varphi \Rightarrow \Gamma, \Delta \vdash \varphi$ .

Natural Deduction, or for that matter its sister system of the *Sequent Calculus*, lends itself well to study derivations for their own sake. This particular branch of logic has in the case of Natural Deduction been rigorously practised and promoted by Dag Prawitz, who established the main facts of the system and who demonstrated its flexibility and usefulness (cf. Prawitz [1965; 1971]).

The fundamental theorem in the subject is concerned with derivations without superfluous parts. The following is evidently awkward.

$$\frac{\frac{\frac{\sigma \quad \sigma \rightarrow \varphi}{\varphi \quad [\psi]} \rightarrow E}{\varphi \wedge \psi} \wedge I}{\varphi} \wedge E}{\psi \rightarrow \varphi} \rightarrow I$$

We have introduced the superfluous conjunction  $\varphi \wedge \psi$  only in order to eliminate it again. A more efficient proof is

$$\frac{\frac{\sigma \quad \sigma \rightarrow \varphi}{\varphi} \rightarrow E}{\psi \rightarrow \varphi} \rightarrow I$$

We have eliminated the introduction followed by an elimination, thus simplifying the derivation. A derivation in which an introduction is never followed by an elimination is called *normal*. Here it has to be explained what ‘follow’ means. For this purpose a special partial ordering is introduced; e.g. in  $\rightarrow I$   $\varphi \rightarrow \psi$  follows after  $\psi$ , in  $\forall E$   $\varphi$  follows after  $\varphi \forall \psi$ , etc. See [van Dalen, 1997, p. 199,203].

Prawitz proved the

**THEOREM 2** (Normal Form Theorem). *If  $\Gamma \vdash \varphi$ , then there is a normal derivation of  $\varphi$  from  $\Gamma$  (cf. [Prawitz, 1965]).*

There is a better result called the

**THEOREM 3** (Normalisation Theorem). *Any derivation reduces to a normal derivation.*

Here a reduction step consists in the removal of a superfluous introduction followed by an elimination (cf. [Prawitz, 1971]).

There is even a still stronger form, the

**THEOREM 4** (Strong Normalisation Theorem). *Every sequence of reduction steps terminates in a normal form.*

The whole tradition of normalisation and reduction is traditionally a part of combinatory logic and  $\lambda$ -calculus, a systematic account is given in [Klop, 1980] and [Barendregt, 1984].

There is an interesting interplay between natural deduction derivations and  $\lambda$ -terms, and hence between normalisation in natural deduction and in  $\lambda$ -calculus (cf. [Gallier, 1995; Howard, 1980; Pottinger, 1976; Troelstra and van Dalen, 1988]).

One of the pleasant corollaries of the normal form of a derivation is the

**PROPERTY 5** (Subformula property). *In a normal derivation of  $\Gamma \vdash \varphi$  only subformulas of  $\Gamma$  and  $\varphi$  occur.*

In particular only connectives from  $\Gamma$  and  $\varphi$  can occur. As a consequence we have

**THEOREM 6.** *Intuitionistic predicate logic is conservative over intuitionistic propositional logic.*

**Proof.** Let  $\vdash \varphi$  where  $\varphi$  is a proposition. Consider a normal derivation  $\pi$  of  $\varphi$ . By the subformula property only propositional connectives can occur, hence we have a derivation using only propositional rules. ■

Natural deduction was given an interesting extension by Schroeder-Heister, [1984]; an exposition and applications can be found in [Negri and von Plato, 2001].

## 3 PROOF TERMS AND THE CURRY–HOWARD ISOMORPHISM

Since natural deduction is so close in nature to the proof interpretation, it is perhaps not surprising that a formal correspondence between a term calculus and natural deduction can be established.

We will first demonstrate this for a small fragment, containing only the connective ‘ $\rightarrow$ ’. Consider an  $\rightarrow$  introduction:

$$\frac{\frac{[\varphi]}{\mathcal{D}} \quad \psi}{\varphi \rightarrow \psi} \quad \frac{\frac{[x : \varphi]}{\mathcal{D}} \quad t : \psi}{\lambda x \cdot t : \varphi \rightarrow \psi}$$

We assign in a systematic way proof-terms to formulas in the derivation. Since  $\varphi$  is an assumption, it has a hypothetical proof term, say  $x$ . On cancelling the hypotheses; we introduce a  $\lambda x$  in front of the (given) term  $t$  for  $\psi$ . By binding  $x$ , the proof term for  $\varphi \rightarrow \psi$  no longer depends on the hypothetical proof  $x$  of  $\varphi$ . Note that this corresponds exactly to our intuitive proof interpretation.

The elimination runs as follows:

$$\frac{\varphi \rightarrow \psi \quad \varphi}{\psi} \quad \frac{t : \varphi \rightarrow \psi \quad s : \varphi}{t(s) : \psi}$$

Observe the analogy to the proof interpretation. Let us consider a particular derivation.

$$\frac{\frac{[\varphi]}{\psi \rightarrow \varphi}}{\varphi \rightarrow (\psi \rightarrow \varphi)} \quad \frac{\frac{[x : \varphi]}{\lambda y \cdot x : \psi \rightarrow \varphi}}{\lambda x \cdot \lambda y \cdot x : \varphi \rightarrow (\psi \rightarrow \varphi)}$$

Thus the proof term of  $\varphi \rightarrow (\psi \rightarrow \varphi)$  is  $\lambda xy.x$ , this is Curry combinator  $K$ .

A cut elimination conversion now should give us information about the conversion of the proof term.

$$\frac{\frac{x : \psi}{\mathcal{D}} \quad t : \varphi}{\lambda x \cdot t : \psi \rightarrow \varphi} \quad \frac{\mathcal{D}' \quad s : \psi}{s : \psi} \quad \text{reduces to} \quad \frac{\mathcal{D}' \quad s : \psi}{\mathcal{D}[s/x]} \quad t[s/x] : \varphi$$

$$\frac{\lambda x \cdot t : \psi \rightarrow \varphi \quad s : \psi}{(\lambda x \cdot t)(s) : \varphi}$$

The proof theoretic conversion corresponds to the  $\beta$ -reduction of the  $\lambda$ -calculus.

In order to deal with full predicate logic we have to introduce specific operations in order to render the meaning of the connectives and their derivation rules:

$$\begin{cases} p & \text{--- pairing} \\ p_0, p_1 & \text{--- projections} \end{cases}$$

$$\begin{cases} D & \text{--- discriminator ("case dependency")} \\ k & \text{--- case obliteration} \end{cases}$$

E – witness extractor

$\perp$  – ex falso operator

$$\begin{array}{ll} \wedge I & \frac{t_0 : \varphi_0 \quad t_1 : \varphi_1}{p(t_0, t_1) : \varphi_0 \wedge \varphi_1} & \wedge E & \frac{t : \varphi_0 \wedge \varphi_1}{p_i(t) : \varphi_i} \quad (i = 0, 1) \\ \vee I & \frac{t : \varphi_i}{k_i(t) : \varphi_0 \vee \varphi_1} \quad (i = 0, 1) & \vee E & \frac{t : \varphi \vee \psi \quad t_0[x^\varphi] : \sigma \quad t_1[x^\psi] : \sigma}{D_{u,v}(t, t_0[u], t_1[v]) : \sigma} \\ \rightarrow I & \frac{t[x^\varphi] : \psi}{\lambda y^\varphi \cdot t[y^\varphi] : \varphi \rightarrow \psi} & \rightarrow E & \frac{t : \varphi \rightarrow \psi \quad t' : \varphi}{t(t') : \psi} \\ \forall I & \frac{t[x] : \varphi(x)}{\lambda y \cdot t[y] : \forall y \varphi(y)} & \forall E & \frac{t : \forall x \varphi(x)}{t(t') : \varphi(t')} \\ \exists I & \frac{t_1 : \varphi(t_0)}{p(t_0, t_1) : \exists x \varphi(x)} & \exists E & \frac{t : \exists x \varphi(x) \quad t_1[y, z^{\varphi(y)}] : \sigma}{E_{u,v}(t, t_1[u, v]) : \sigma} \end{array}$$

There are a number of details that we have to mention.

- (i) In  $\rightarrow I$  the dependency on the hypothesis has to be made explicit in the term. We do this by assigning to each hypothesis its own variable. E.g.  $x^\varphi : \varphi$ .
- (ii) In  $\forall E$  (and similarly  $\exists E$ ) the dependency on the particular (auxilliary) hypotheses  $\varphi$  and  $\psi$  disappears. This is done by a variable binding technique. In  $D_{u,v}$  the variables  $u$  and  $v$  are bound.
- (iii) In the falsum rule the result, of course, depends on the conclusion  $\varphi$ . So  $\varphi$  has its own ex falso operator  $\perp_\varphi$ .



Now the conversion rules for the derivation automatically suggest the conversion for the term.

We have seen that the term calculus corresponds with the natural deduction system. This suggests a correspondence between proofs and propositions on the one hand and elements (given by the terms) and types (the spaces where these terms are to be found). This correspondence was first observed for a simple case (the implication fragment) by Haskell Curry, [Curry and Feys, 1958], ch. 9, § E, and extended to full intuitionistic logic by W. Howard, [Howard, 1980]. Let us first look at a simple case, the one considered by Curry.

Since the meaning of proposition is expressed in terms of possible proofs — we know the meaning of  $\varphi$  if we know what things qualify as proofs — one may take an abstract view and consider a proposition as its collection of proofs. From this viewpoint there is a striking analogy between propositions and sets. A set has elements, and a proposition has proofs. As we have seen, proofs are actually a special kind of constructions, and they operate on each other. E.g. if we have a proof  $p : \varphi \rightarrow \psi$  and a proof  $q : \varphi$  then  $p(q) : \psi$ . So proofs are naturally typed objects.

Similarly one may consider sets as being typed in a specific way. If  $\varphi$  and  $\psi$  are typed sets then the set of all mappings from  $\varphi$  to  $\psi$  is of a higher type, denoted by  $\varphi \rightarrow \psi$  or  $\psi^\varphi$ . Starting from certain basic sets with types, one can construct higher types by iterating this ‘function space’-operation. Let us denote ‘ $a$  is in type  $\varphi$ ’ by  $a \in \varphi$ .

Now there is this striking parallel.

Propositions	Types
$a : \varphi$	$a \in \varphi$
$p : \varphi \rightarrow \psi, q : \varphi$ $\Rightarrow p(q) : \psi$	$p \in \varphi \rightarrow \psi, q \in \varphi$ $\Rightarrow p(q) \in \psi$
$x : \varphi \Rightarrow t(x) : \psi$ then $\lambda x \cdot t : \varphi \rightarrow \psi$	$x : \varphi \Rightarrow t(x) \in \psi$ then $\lambda x \cdot t \in \varphi \rightarrow \psi$

It now is a matter of finding the right types corresponding to the remaining connectives. For  $\wedge$  and  $\vee$  we introduce a product type and a disjoint sum type. For the quantifiers generalizations are available. The reader is referred to the literature, cf. [Howard, 1980], [Gallier, 1995].

The main aspect of the Curry-Howard isomorphism, (also known as “proofs as types”), is the faithful correspondence:

$$\frac{\text{proofs}}{\text{propositions}} = \frac{\text{elements}}{\text{types}}$$

with their conversion and normalization properties.

The importance of the connection between intuitionistic logic and type theory was fully grasped and exploited by Per Martin-Löf. Indeed, in his approach the two are actually merge into one master system. His type systems are no mere technical innovations, but they intend to capture the foundational meaning of intuitionistic logic and the corresponding mathematical universe. Expositions of ‘proofs as types’ and the Martin-Löf type theories can be found in e.g. [Gallier, 1995], [Girard *et al.*, 1989], [Martin-Löf, 1977], [Martin-Löf, 1984], [Troelstra and van Dalen, 1988], [Sommaruga, 2000].

#### 4 SEMANTICS

The intended interpretation of intuitionistic logic as presented by Heyting, Kreisel and others so far has proved to be rather elusive, in as much that the completeness properties that are on every logicians shopping list, have not (yet) been established. Even in the case of the interpretation of arithmetic the results are far from final. The Curry–Howard isomorphism, also known by the name ‘formulas as types’, in a sense fulfills the promise of the proof interpretation for intuitionistic logic, in the sense that there is a precise correspondence between natural deductions and proof terms, [Troelstra and van Dalen, 1988, p. 556].

However, ever since Heyting’s formalisation, various, more or less artificial, semantics have been proposed. In the thirties the topological interpretation was introduced by Tarski, and in the fifties and sixties Beth and Kripke formulated two closely related semantics.

We will first consider the topological interpretation.

DEFINITION 7. A *topological space* is a pair  $\langle X, \mathcal{O} \rangle$  where  $\mathcal{O} \subseteq \mathcal{P}(X)$  such that

1.  $\emptyset, X \in \mathcal{O}$
2.  $U, V \in \mathcal{O} \rightarrow U \cap V \in \mathcal{O}$
3.  $U_i \in \mathcal{O} (i \in I) \rightarrow \cup \{U_i \mid i \in I\} \in \mathcal{O}$ .

In plain words, a topological space is a set that comes with a family  $\mathcal{O}$  of open subsets that is closed under arbitrary unions and finite intersections and that contains  $\emptyset$  and  $X$ . A familiar example is the Euclidean plane, where  $\mathcal{O}$  consists of unions of open discs.

In general we can define a topological space when a *basis* is given, i.e. a collection  $\mathcal{B}$  of subsets such that

1.  $A_i \in \mathcal{B}, p \in A_i (i = 1, 2) \Rightarrow \exists A \in \mathcal{B} (p \in A \subseteq A_1 \cap A_2)$
2.  $\forall p \in X \exists A \in \mathcal{B} (p \in A)$ .

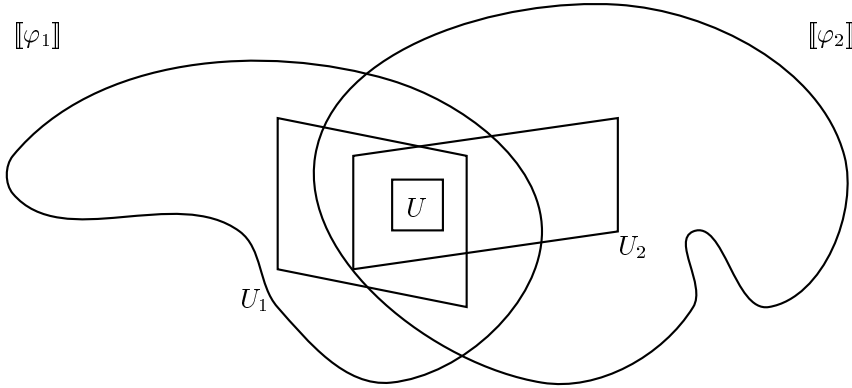
We now define open sets as arbitrary unions of basis-elements. It is a simple exercise to show that the open sets, thus introduced, indeed satisfy the condition of Definition 7. The open discs of the Euclidean plane evidently form a basis for the natural topology.

We call  $U$  a *neighbourhood* of a point  $p$  if  $U$  is open and  $p \in U$ , and if, for a given basis  $\mathcal{B}$ ,  $U \in \mathcal{B}$ , we say that  $U$  is a basic neighbourhood of  $p$ .

Now we will interpret sentences as open subsets (*opens*, for short) of a topological space. In order to motivate the interpretation we recall that, when a fixed basis  $\mathcal{B}$  is given, the evidence for  $p \in U$  is a basic neighbourhood  $A$  of  $p$  such that  $A \subseteq U$ .

Let us now assign to each statement  $\varphi$  an open subset  $\llbracket \varphi \rrbracket$  of  $X$ . We will try to motivate the topological operations that accompany the connectives.

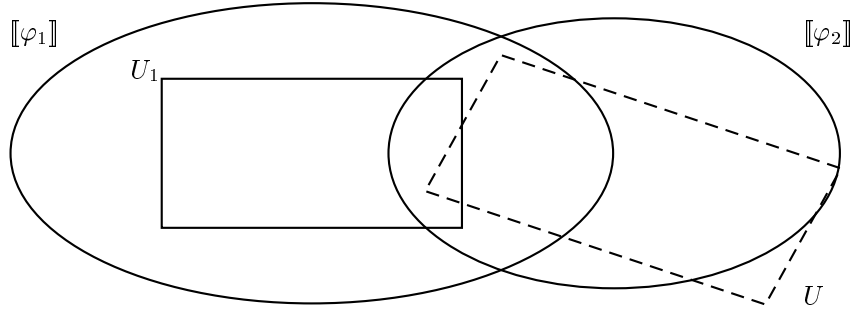
Let us say that a basic neighbourhood  $U$  *proves*  $\varphi$  if  $U \subseteq \llbracket \varphi \rrbracket$ . Suppose that  $U_i$  proves  $\varphi_i$  then by the definition of basis we can find  $U \in \mathcal{B}$  such that  $U \subseteq U_1 \cap U_2$ ,  $U$  proves both  $\varphi_1$  and  $\varphi_2$ . The union of all those  $U$ 's that prove both  $\varphi_1$  and  $\varphi_2$  is  $\llbracket \varphi_1 \rrbracket \cap \llbracket \varphi_2 \rrbracket$ , so let us put  $\llbracket \varphi_1 \wedge \varphi_2 \rrbracket := \llbracket \varphi_1 \rrbracket \cap \llbracket \varphi_2 \rrbracket$ . Similarly we put  $\llbracket \varphi_1 \vee \varphi_2 \rrbracket := \llbracket \varphi_1 \rrbracket \cup \llbracket \varphi_2 \rrbracket$ . Since  $\perp$  should not have a proof, we put  $\llbracket \perp \rrbracket := \emptyset$ . Note that this leaves  $\emptyset$  as a proof of  $\perp$ , therefore we consider  $\emptyset$  as the empty proof (or a kind of degenerate proof that carries no evidence). The interesting case is the implication.



A proof of  $\varphi_1 \rightarrow \varphi_2$  should give us a method to convert a proof of  $\varphi_1$  into a proof of  $\varphi_2$ . Therefore we take a basic neighbourhood  $U$  in  $\llbracket \varphi_1 \rrbracket^c \cup \llbracket \varphi_2 \rrbracket$ , now for any proof  $U_1$  that intersects  $U$  we can find a proof  $U_2$  of  $\varphi_2$  in  $U \cap U_1$ : So  $U$  indeed provides the required method.

The  $U$ 's with that property make up the largest open subset of  $\llbracket \varphi_1 \rrbracket^c \cup \llbracket \varphi_2 \rrbracket$ , which we call the interior of that set. So let us put

$$\llbracket \varphi_1 \rightarrow \varphi_2 \rrbracket := \text{Int} (\llbracket \varphi_1 \rrbracket^c \cup \llbracket \varphi_2 \rrbracket) \quad (= \text{Int}\{x \mid x \in \llbracket \varphi_1 \rrbracket \Rightarrow x \in \llbracket \varphi_2 \rrbracket\}).$$



In order to interpret quantified statements we assume that a domain  $A$  of individuals is given. Then we put

$$\begin{aligned} \llbracket \exists x\varphi(x) \rrbracket &:= \cup \{ \llbracket \varphi(a) \rrbracket \mid a \in A \} \\ \llbracket \forall x\varphi(x) \rrbracket &:= \text{Int} \cap \{ \llbracket \varphi(a) \rrbracket \mid a \in A \}.^1 \end{aligned}$$

Let us now accept the above as an inductive definition of the *value*  $\llbracket \varphi \rrbracket_X$  of  $\varphi$  in  $X$  under a given assignment of open sets to atomic sentences. When no confusion arises we will delete the index  $X$ . The notation suppresses  $\mathcal{O}$ , a better notation would be  $\llbracket \varphi \rrbracket_{\mathcal{O}}$ , but the reader will have no difficulty finding the correct meaning. A formula  $\varphi$  is said to be *true in the topological space*  $X$ , notation  $\vDash_X \varphi$ , if for all valuations  $\llbracket \text{cl}(\varphi) \rrbracket = X$ , where  $\text{cl}(\varphi)$  is the universal closure of  $\varphi$ .  $\varphi$  is true,  $\vDash \varphi$ , if  $\varphi$  is true in all topological spaces.

For the consequence relation,  $\vDash$ , we define  $\Gamma \vDash_X \varphi := \text{Int} \cap \{ \llbracket \psi \rrbracket_X \mid \psi \in \Gamma \} \subseteq \llbracket \varphi \rrbracket_X$  and  $\Gamma \vDash \varphi$  iff  $\Gamma \vDash_X \varphi$  for all  $X$ . Observe that for finite  $\Gamma (= \{ \psi_1, \dots, \psi_n \})$ ,  $\Gamma \vDash \varphi \Leftrightarrow \vDash \psi_1 \wedge \dots \wedge \psi_n \rightarrow \varphi$ . Observe that nothing has been said about the topological space  $X$ , in particular  $X$  could be the one-point space with a resulting two-valued, classical logic! This shows that the above motivation has not enough special assumptions on ‘constructions’, or ‘evidence’ to lead to a specifically intuitionistic logic. The explanation is too liberal.

The topological interpretation is complete in the following sense:

**THEOREM 8.**  $\Gamma \vdash \varphi \Leftrightarrow \Gamma \vDash \varphi$ .

The implication from left to right (the soundness with respect to the topological interpretation) is easily verified by the reader. Just check all the axioms of the Hilbert-type system and show that the derivation rules preserve truth, or do the latter for the rules of natural deduction.

<sup>1</sup>For convenience we will abuse notation and use the same symbol for the individual and its name.

We will treat the  $\rightarrow I$  rule.

Let us abbreviate  $\llbracket \Gamma \rrbracket_X$ ,  $\llbracket \varphi \rrbracket_X$  and  $\llbracket \psi \rrbracket_X$  as  $U, V, W$  (where  $\llbracket \Gamma \rrbracket_X = \bigcup \{ \llbracket \sigma \rrbracket_X \mid \sigma \in \Gamma \}$ ). Then the induction hypothesis is  $U \cap V \subseteq W$  (note that we use the formulation of p. 20). Since  $U$  is open,  $U \subseteq \text{Int}(V^c \cup W) \Leftrightarrow U \subseteq V^c \cup W$ . Now it is a matter of elementary set theory to show  $U \cap V \subseteq W \Leftrightarrow U \subseteq V^c \cup W$ .

The implication from right to left will follow from a later result.

EXAMPLE 9.  $\llbracket \neg \varphi \rrbracket = \llbracket \varphi \rightarrow \perp \rrbracket = \text{Int} \llbracket \varphi \rrbracket^c$ . Let  $\varphi$  be an atom and assign to it the complement of a point  $p$  (in the plane), then  $\llbracket \neg \varphi \rrbracket = \emptyset$  and  $\llbracket \varphi \vee \neg \varphi \rrbracket = X - \{p\} \neq X$ . By the soundness of the logic we have  $\not\vdash \varphi \vee \neg \varphi$ .

The topological interpretation is extensively studied in [Rasiowa and Sikorski, 1963] (cf. also [Schütte, 1968; Dummett, 1977]).

We will move on to a semantics that belongs to the same family as the topological interpretation but that has certain advantages. Beth and Kripke have each introduced a semantics for intuitionistic logic and shown its completeness. The semantics that we present here is a common generalisation introduced for metamathematical purposes in [van Dalen, 1984].

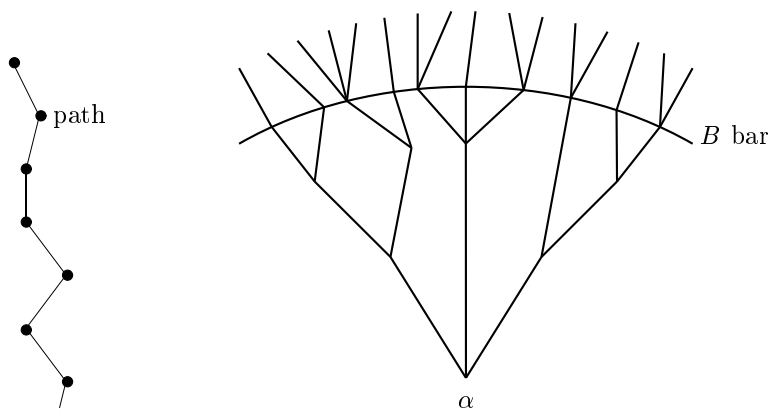
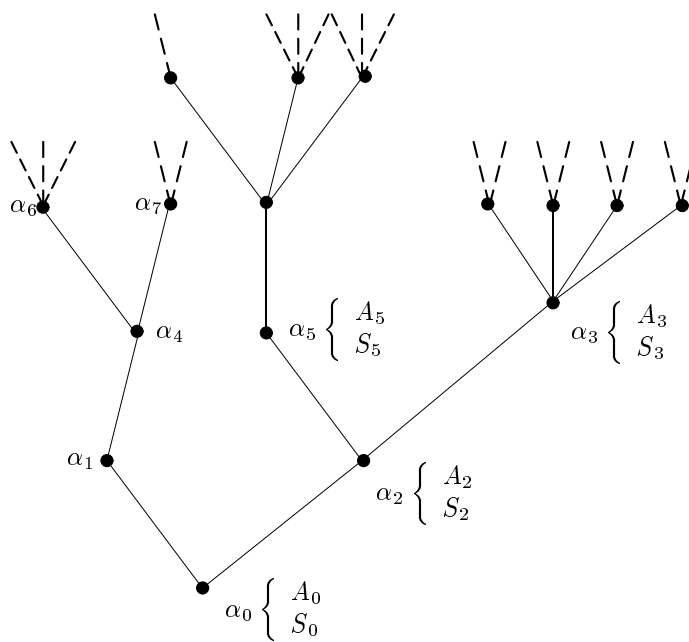
The underlying heuristics is based on the conception of mathematics (and hence logic) as a mental activity of an (idealised) mathematician (or logician if you like). Consider the mental activity of this person,  $S$ , as structured in linear time of type  $\omega$ , i.e. time  $t$  runs through  $0, 1, 2, 3, \dots$ . At each time  $t$   $S$  has acquired a certain body of facts, knowledge. It seems reasonable to assume that  $S$  has perfect memory, so that the body of facts increases monotone in time. Furthermore  $S$  has at each time  $t$ , in general, a number of possibilities to increase his knowledge in the transition to time  $t+1$ . So if we present 'life' graphically for  $S$ , it turns out to fork. However,  $S$  not only collects, experiences or establishes truths, but he also constructs objects, the elements of his universe. Here also is considerable freedom of choice for  $S$ , going from time  $t$  to  $t+1$  he may decide to construct the next prime, or to construct  $\sqrt{2}$ . This yields a treelike picture of  $S$ 's possible histories.

Each node of the tree represents a stage of knowledge of  $S$  and a stage in his construction of his universe. So to each node  $\alpha_i$  we have assigned a set of sentences  $S_i$  and a set of objects  $A_i$ , subject to the condition that  $S_i$  and  $A_i$  increase, i.e.

$$\alpha_i \leq \alpha_j \Rightarrow S_i \subseteq S_j \text{ and } A_i \subseteq A_j.$$

Given this picture of  $S$ 's activity, let us find out how he interprets the logical constants. First, two auxiliary notions: a *path* through  $\alpha$  is a maximal linearly ordered subset, a *bar* for  $\alpha$  is a subset  $B$  such that each path through  $\alpha$  intersects  $B$ .

It is suggestive to picture bars above  $\alpha_i$ , i.e. to situate them in the future. It is no restriction to restrict ourselves to this kind of bars we will see. Now let  $\varphi$  be an atomic sentence. How can  $S$  know  $\varphi$  at state  $\alpha$ ? He could



require that  $\varphi$  were then and there given to him. That however seems a bit restrictive. He might know how to establish  $\varphi$ , but need more time to do so. In that case we say that  $S$  knows  $\varphi$  at stage  $\alpha$  if for each path through  $\alpha$  (so to speak each ‘research’) there is a stage  $\beta$  such that at  $\beta$   $\varphi$  is actually established (or, maybe, experienced). In other words, if there is a bar  $B$  for  $\alpha$  such that at each  $\beta \in B$   $\varphi$  is given. The following clauses fix the knowledge of  $S$  concerning composite statements.

**CONJUNCTION.**  $S$  knows  $\varphi \wedge \psi$  at stage  $\alpha$  if he knows both  $\varphi$  and  $\psi$  at stage  $\alpha$ .

**DISJUNCTION.** For  $S$  to know that  $\varphi \vee \psi$  holds at stage  $\alpha$  he need not know right away which one holds, he may again need a bit more time. All he needs to know is that eventually  $\varphi$  or  $\psi$  will hold. To be precise, that there is a bar  $B$  for  $\alpha$  such that for each  $\beta \in B$   $S$  knows  $\varphi$  at stage  $\beta$  or he knows  $\psi$  at stage  $\beta$ .

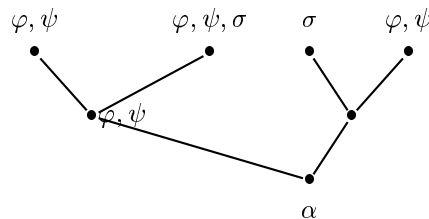
**IMPLICATION.** For  $S$  to know  $\varphi \rightarrow \psi$  at stage  $\alpha$ , he need not know anything about  $\varphi$  or  $\psi$  at stage  $\alpha$ , all he must be certain of is that if he comes to know  $\varphi$  in any later stage  $\beta$ , he must also know  $\psi$  at that stage.

**FALSITY.**  $S$ , being an idealised person, never establishes a falsity.

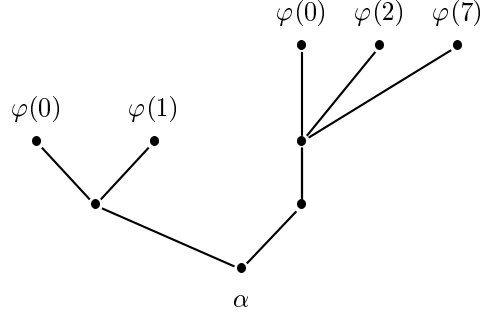
**UNIVERSAL QUANTIFICATION.** For  $S$  to know  $\forall x\varphi(x)$  at stage  $\alpha$  it does not suffice to know  $\varphi(a)$  for al objects  $a$  that exist at stage  $\alpha$ , but also for all objects that will be constructed in the future.

**EXISTENTIAL QUANTIFICATION.**  $S$  knows  $\exists x\varphi(x)$  at stage  $\alpha$  if eventually he will construct an element  $a$  such that he knows  $\varphi(a)$ . To be precise, if there is a bar  $B$  for  $\alpha$  such that for each  $\beta \in B$  there exists an element  $a$  at stage  $\beta$  such that  $S$  knows  $\varphi(a)$  at that stage.

Examples.



\* knows  $\varphi \rightarrow \psi$  at  $\alpha$



\* knows  $\exists x\varphi(x)$  at  $\alpha$

We will now give a formal definition of a model for a given similarity type (without functions).

**DEFINITION 10.**

1. A model is a quadruple  $\mathcal{M} = \langle M, \leq, D, \Vdash \rangle$  where  $M$  is partially ordered by  $\leq$ , and  $D$  is a function that assigns to each element of  $M$  a structure of the given type, such that for  $\alpha, \beta \in M, \alpha \leq \beta \Rightarrow D(\alpha) \subseteq D(\beta)$ . *Warning:* we mean literally ‘subset’, not ‘substructure’.  $D(\alpha) \subseteq D(\beta)$  is used as a shorthand for ‘the universe of  $D(\alpha)$  is a subset of that of  $D(\beta)$ , and the relations of  $D(\alpha)$  are subsets of the corresponding relations of  $D(\beta)$ ’. We write  $a \in D(\alpha)$  for ‘ $a$  is in the universe of  $D(\alpha)$ ’.
2. The relation  $\Vdash$  between elements of  $M$  and sentences, called the *forcing relation* is inductively defined by
  - (a)  $\alpha \Vdash \varphi$ , for  $\varphi$  atomic, if there is a bar  $B$  for  $\alpha$  such that  $\forall \beta \in B, D(\beta) \models \varphi$
  - (b)  $\alpha \Vdash \varphi \wedge \psi$  if  $\alpha \Vdash \varphi$  and  $\alpha \Vdash \psi$
  - (c)  $\alpha \Vdash \varphi \vee \psi$  if there is a bar  $B$  for  $\alpha$  such that  $\forall \beta \in B, \beta \Vdash \varphi$  or  $\beta \Vdash \psi$
  - (d)  $\alpha \Vdash \varphi \rightarrow \psi$  if  $\forall \beta \geq \alpha, \beta \Vdash \varphi \Rightarrow \beta \Vdash \psi$
  - (e)  $\alpha \Vdash \forall x\varphi(x)$  if  $\forall \beta \geq \alpha \forall b \in D(\beta), \beta \Vdash \varphi(b)$
  - (f)  $\alpha \Vdash \exists x\varphi(x)$  if there is a bar  $B$  for  $\alpha$  such that  $\forall \beta \in B, \exists b \in D(\beta), \beta \Vdash \varphi(b)$ .

Observe that for no  $\alpha, \alpha \Vdash \perp$ , so by defining  $\neg\varphi := \varphi \rightarrow \perp$  we get

3.  $\alpha \Vdash \neg\varphi$  if  $\forall \beta \geq \alpha, \beta \not\Vdash \varphi$  (where  $\beta \not\Vdash \varphi$  stands for  $\beta \Vdash \varphi$ ).



Our definition used the approach with auxiliary names for elements of the structures  $D(\alpha)$ . The alternative approach with assignments works just as well.

We say that a formula  $\varphi$  holds (is true) in a model  $\mathcal{M}$  if  $\alpha \Vdash \text{cl}(\varphi)$  for all  $\alpha \in M$ . If we also allow for the language to contain proposition letters, then the interpretation of propositional logic is contained as a special case.

The following lemma is rather convenient for practical purposes

LEMMA 11.

1.  $\alpha \leq \beta, \alpha \Vdash \varphi \Rightarrow \beta \Vdash \varphi$
2.  $\alpha \not\Vdash \varphi \Leftrightarrow$  *there is a path  $P$  through  $\alpha$  such  $\forall \beta \in P(\beta \not\Vdash \varphi)$*
3.  $\alpha \Vdash \varphi \Leftrightarrow$  *there is a bar  $B$  for  $\alpha$  such that  $\forall \beta \in B(\beta \Vdash \varphi)$ .*

**Proof.** Induction on  $\varphi$ . Note that (2) is obtained from (3) by negating both sides. ■

For sentences we have

LEMMA 12 (Soundness).  $\Gamma \vdash \varphi \Rightarrow \Gamma \Vdash \varphi$ .

**Proof.**  $\Gamma \Vdash \varphi$  stands for ‘for each  $\mathcal{M}$  and each  $\alpha \in M, \alpha \Vdash \psi$  for all  $\psi \in \Gamma \Rightarrow \alpha \Vdash \varphi$ ’. The proof proceeds by induction on the derivation of  $\Gamma \vdash \varphi$ . We consider one case:  $\Gamma, \varphi \vdash \psi \rightarrow \Gamma \vdash \varphi \rightarrow \psi$ . Let, in a model  $\mathcal{M}, \alpha \Vdash \sigma$  for all  $\sigma \in \Gamma$ . Suppose that  $\alpha \not\Vdash \varphi \rightarrow \psi$ , then there is a  $\beta \geq \alpha$  such that  $\beta \Vdash \varphi$  but  $\beta \not\Vdash \psi$ . This conflicts with the induction hypothesis  $\Gamma, \varphi \Vdash \psi$ . Hence  $\Gamma \Vdash \varphi \rightarrow \psi$ . ■

We obtain the Beth models and Kripke models by specialisation:

DEFINITION 13.

1.  $\mathcal{M}$  is a *Beth model* if  $|D(\alpha)|$  is a fixed set  $D$  for all  $\alpha$ .
2.  $\mathcal{M}$  is a *Kripke model* if in (a), (c) and (f)  $B = \{\alpha\}$ . To spell it out:

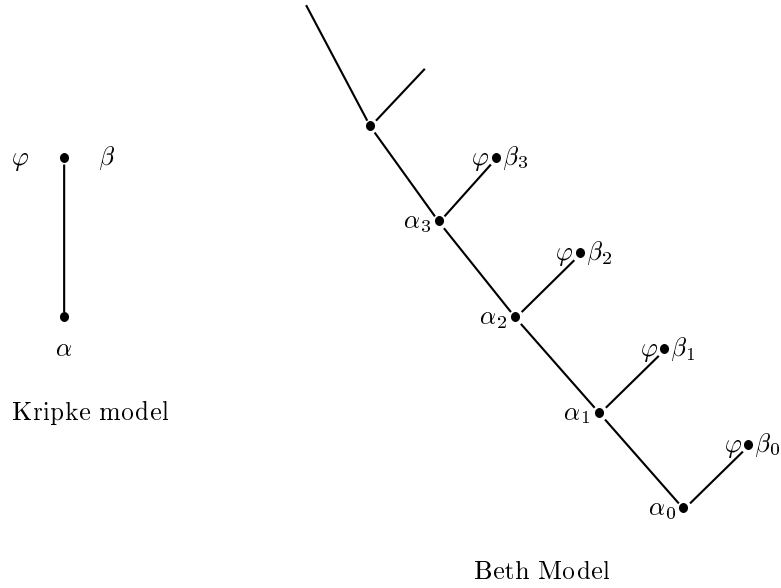
- (a')  $\alpha \Vdash \varphi$  if  $D(\alpha) \vDash \varphi$
- (c')  $\alpha \Vdash \varphi \vee \psi$  if  $\alpha \Vdash \varphi$  or  $\alpha \Vdash \psi$
- (f')  $\alpha \Vdash \exists x\varphi(x)$  if  $\exists a \in D(\alpha), \alpha \Vdash \varphi(a)$ .

For a Beth model we can simplify clause 5:

- (a')  $\forall x\varphi(x) \Leftrightarrow \forall a \in D, \alpha \Vdash \varphi(a)$  (repeat the proof of Lemma 11(a)).

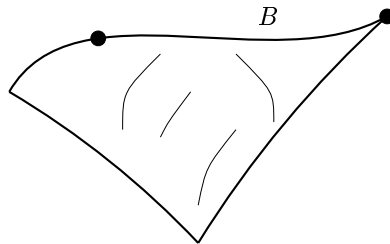
Generally speaking, Kripke models are somewhat superior to Beth models. A small example may serve to illustrate this.

We will summarily present models by a simple diagram. For each node we list the propositions that are forced by it.



The Kripke model is a counter-example to  $\varphi \vee \neg\varphi$ , and so is the Beth model. Note that the Beth model has to be infinite in order to refute a classical tautology, since in a well-founded model all classical tautologies are true. One sees this by observing that in a well-founded model (i.e. there are no infinite *ascending* sequences; if we had turned the model upside down, we would have had the proper well-foundedness) there is a bar of maximal nodes.

Now consider a maximal node  $\alpha$ , if  $\alpha \not\Vdash \varphi$ , then  $\alpha \Vdash \neg\varphi$ . So  $\alpha \Vdash \varphi \vee \neg\varphi$ . So  $\varphi \vee \neg\varphi$  is forced on the bar  $B$  and hence in each node of the model.



So, as a rule, we have simpler Kripke models for our logical purposes than Beth models.

A Beth model is a special case of our model, so we automatically have soundness for Beth models. For Kripke models, however, we have to show soundness separately.

Each class of models is complete for intuitionistic logic. This can be shown as follows, first show the Model Existence Lemma for Kripke semantics, then modify a Kripke model into a model and finally a model into a Beth model.

**LEMMA 14** (Model Existence Lemma for Kripke Semantics). *If  $\Gamma \not\vdash \varphi$  then there is a Kripke model  $\mathcal{K}$  with a bottom node  $\alpha_0$  such that  $\alpha_0 \Vdash \psi$  for all  $\psi \in \Gamma$  and  $\alpha_0 \not\vdash \varphi$ .*

**Proof.** We'll use a Henkin-style proof after Aczel, Fitting and Thomason. For simplicity's sake we'll treat the case of a denumerable language, i.e. we have denumerably many individual variables and individual constants. A set  $\Gamma$  of sentences is called a *prime* (also, saturated) *theory* if

1. it is closed under derivability
2.  $\varphi \vee \psi \in \Gamma \Rightarrow \varphi \in \Gamma$  or  $\psi \in \Gamma$
3.  $\exists x\varphi(x) \in \Gamma \Rightarrow \varphi(c) \in \Gamma$  for some constant  $c$ . ■

The fundamental fact about prime theories is the following:

**LEMMA 15.** *If  $\Gamma \not\vdash \varphi$  then there is a prime theory  $\Gamma^p \supseteq \Gamma$  such that  $\varphi \notin \Gamma^p$ .*

**Proof.** We have to make a harmless little assumption, namely that there are enumerably many constants  $c_i$ , not in  $\Gamma$ . We approximate the  $\Gamma^p$ , as in the case of the Hintikka sets. To start, we add enumerably many new constants to the language of  $\Gamma, \varphi$ . Since we have a countable language, we may assume that the sentences are given in some fixed enumeration. We will treat these sentences one by one. This 'treatment' consists of adding witnesses (as in the case of the Hintikka set) and deciding disjunctions. We, so to speak, approximate the required  $\Gamma^p$ .

*step 0*      $\Gamma_0 = \Gamma$

*step  $k + 1$*   *$k$  is even.* Let  $\exists x\psi(x)$  be the first existential sentence such that  $\Gamma_k \vdash \exists x\psi(x)$ , that has not been treated, and let  $c$  be the first fresh constant not in  $\Gamma_k$ . then put  $\Gamma_{k+1} = \Gamma_k, \psi(c)$ .

*$k$  is odd.* Let  $\psi_1 \vee \psi_2$  be the first disjunction that has not been treated, such that  $\Gamma_k \vdash \psi_1 \vee \psi_2$ . Pick an  $i$  such that  $\Gamma_k, \psi_i \not\vdash \varphi$ , then put  $\Gamma_{k+1} = \Gamma_k, \psi_i$ . By 2. below, at least one of  $\psi_1, \psi_2$  will do.

The prime theory we are looking for is

$$\Gamma^p = \bigcup_{k \geq 0} \Gamma_k.$$

We will check the properties.

1.  $\Gamma \subseteq \Gamma^p$ , trivially.

2.  $\Gamma^p \not\vdash \varphi$ . This amounts to  $\Gamma_k \not\vdash \varphi$  for all  $k$ . We use induction on  $k$ .

*Case 1.*  $\Gamma_{2k+1} = \Gamma_{2k}, \psi(c)$ . Assume  $\Gamma_{2k} \not\vdash \varphi$ . If  $\Gamma_{2k+1} \vdash \varphi$  then by  $\exists E, \Gamma_{2k} \vdash \varphi$ . Contradiction.

*Case 2.* We have to show that  $\Gamma_{2k+1}, \psi_1 \not\vdash \varphi$  or  $\Gamma_{2k+1}, \psi_2 \not\vdash \varphi$ . Suppose both are false, then by  $(\forall E)$   $\Gamma_{2k+1} \vdash \varphi$ . Contradiction.

So, we proved  $\Gamma_k \not\vdash \varphi$  for all  $k$ .

3.  $\Gamma^p$  is a prime theory.

(a) Let  $\psi_1 \vee \psi_2 \in \Gamma^p$ , then  $\psi_1 \vee \psi_2 \in \Gamma_k$  for some  $k$ , and hence  $\Gamma_h \vdash \psi_1 \vee \psi_2$  for all  $h \geq k$ . Now look for the first  $h$  such that  $\psi_1 \vee \psi_2$  is treated at step  $h$ ; then by definition  $\psi_1 \in \Gamma_{h+1}$  or  $\psi_2 \in \Gamma_{h+1}$ . And so at least one of the  $\psi_i$ 's is in  $\Gamma^p$ .

(b)  $\exists x\psi(x) \in \Gamma^p$  implies by a similar argument that  $\psi(c) \in \Gamma^p$  for some  $c$ .

(c) If  $\Gamma^p \vdash \psi$ , then  $\Gamma^p \vdash \psi \vee \psi$  and, as in (1),  $\psi \in \Gamma^p$ . ■

We now can construct the required Kripke model. In order to obtain elements for the various domains we consider denumerably many disjoint sets  $V_i$  of denumerably many constants  $\{c_m^i \mid m \geq 0\}$ . By joining these  $V_i$ 's we get a denumerable family of languages  $L_i$  partially ordered by inclusion. The nodes of our Kripke model are prime theories  $\Gamma \supseteq \Gamma_0$ , which are prime with respect to some  $L_i$ , and the partial ordering is the inclusion relation. The domain of such a  $\Gamma$  is the set of constants of its language  $L_i$ . The forcing relation is defined by

$$\Gamma \Vdash \psi \Leftrightarrow \psi \in \Gamma \text{ for atomic } \psi.$$

*Claim:*  $\Gamma \Vdash \psi \Leftrightarrow \psi \in \Gamma$  holds for all sentences  $\psi$ .

We use induction on  $\psi$ . For  $\psi_1 \vee \psi_2, \exists x\psi(x)$  we apply the prime property of  $\Gamma$ . Consider  $\Gamma \Vdash \psi_1 \rightarrow \psi_2$ , if  $\psi_1 \rightarrow \psi_2 \notin \Gamma$ , then  $\Gamma, \psi_1 \not\vdash \psi_2$ , so we can find  $\Delta \supseteq \Gamma, \psi_1$  such that  $\Delta \not\vdash \psi_2$  and  $\Delta$  is prime with respect to  $L_{i+1}$  (where  $L_i$  belongs to  $\Gamma$ ). So, by induction hypothesis,  $\Delta \Vdash \psi_1$  and  $\Delta \not\vdash \psi_2$ . Contradiction. Hence  $\psi_1 \rightarrow \psi_2 \in \Gamma$ . The converse is simple.

A similar argument is used for  $\forall x\psi(x)$ . Let  $\Gamma \Vdash \forall x\psi(x)$ , i.e.  $\forall \Delta \supseteq \Gamma, \forall c \in D(\Delta), \Delta \Vdash \psi(c)$ , and by induction hypothesis  $\psi(c) \in \Delta$ . Now if

$\forall x\psi(x) \notin \Gamma$ , then  $\Gamma \not\vdash \forall x\psi(x)$  and hence  $\Gamma \not\vdash \psi(c)$  for a fresh constant  $c$  of the next language  $L_i$ . But then we can find a prime theory  $\Delta$  with respect to  $L_i$  that contains  $\Gamma$  and  $\Delta \not\vdash \psi(c)$ , so  $\psi(c) \notin \Delta$ . Contradiction. So  $\forall x\psi(x) \in \Gamma$ . Again the converse is simple.

We now finally can finish our proof: the model that we have constructed satisfies the requirements. To be precise, we first extend  $\Gamma$  to a prime theory  $\Gamma_0$  and then construct the model with  $\Gamma_0$  as bottom node.

As a corollary we have the

**THEOREM 16** (Strong Completeness Theorem for Kripke Semantics).

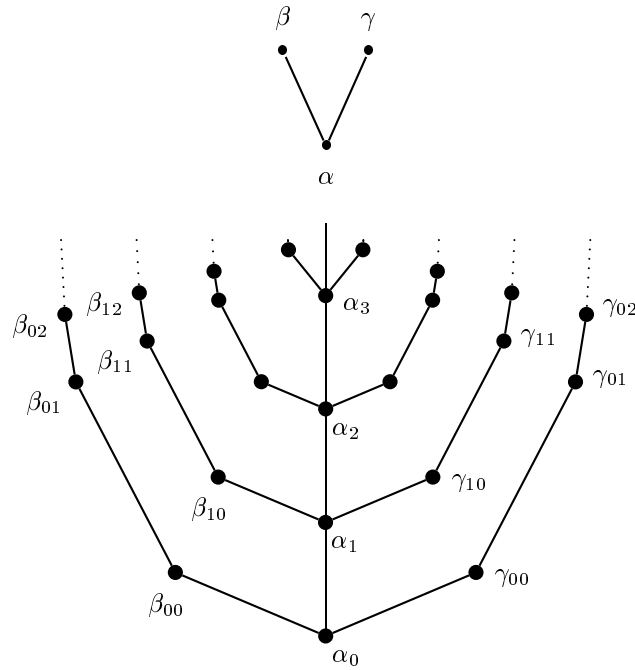
$$\Gamma \vdash \varphi \Leftrightarrow \Gamma \Vdash \varphi.$$

**Proof.**  $\Rightarrow$  is the soundness property.

$\Leftarrow$  If  $\Gamma \not\vdash \varphi$ , then we have a Kripke model such that its bottom node  $\alpha_0 \not\vdash \varphi$  and  $\alpha_0 \Vdash \psi$  for all  $\psi \in \Gamma$ . Hence  $\Gamma \not\vdash \varphi$ . ■

In order to carry the result over to the other two semantics it suffices to modify a Kripke model so that we obtain a (Beth) model that does the trick of Lemma 14.

Kripke has indicated how to do this. In one step we obtain a general model, and in one more step a Beth model. We will indicate only the first modification



We basically repeat each node infinitely often, complete with its domain. If we look at the Kripke model and its modification below, then we see that

each  $\alpha_i$  forces the same atoms as  $\alpha$  in the Kripke model, since any bar intersects the path  $\alpha_0\alpha_1\alpha_2\dots$ .

An inductive argument shows that

$$\delta \Vdash_K \varphi \Leftrightarrow \delta' \Vdash \varphi,$$

where  $\delta = \alpha, \beta, \gamma$  and  $\delta'$  is one of the indexed  $\delta$ 's below and where  $\Vdash_K$  stands for Kripke-forcing, and  $\Vdash$  for general forcing. In order to make the procedure general, we introduce finite sequences  $\langle \alpha_1, \alpha_2, \dots, \alpha_n \rangle$  of nodes of the Kripke model, with  $\alpha_i \leq \alpha_{i+1}$ , as nodes of the new model. Put  $D(\langle \alpha_1, \dots, \alpha_n \rangle) = D(\alpha_n)$ . It is a simple exercise to show that the new model serves to establish Lemma 14. This suffices to show the completeness of our semantics. In order to obtain a Beth model we have to collect everything into one domain. This is worked out in [Kripke, 1965, p. 112 ff] or [Schütte, 1968].

As a result we have

$$\Gamma \vdash \varphi \Leftrightarrow \Gamma \Vdash_K \varphi \Leftrightarrow \Gamma \Vdash \varphi \Leftrightarrow \Gamma \Vdash_B \varphi,$$

where  $\Vdash_K$  and  $\Vdash_B$  stand for Kripke and Beth forcing.

Let us finally return to the topological interpretation. We will show that each Beth model can be viewed as a topological model. Consider a Beth model  $\langle B, \leq, D, \Vdash \rangle$ , the poset  $B$  gives rise to a topological space as follows: the points of  $T_B$  are paths in  $B$ . We define a topology by indicating the basic open sets  $U_\alpha$ , where  $U_\alpha = \{P \mid \exists \beta \geq \alpha, P \text{ passes through } \beta\}$ . The opens (short for ‘open sets’) of  $T_B$  are unions of  $U_\alpha$ 's. In the terminology of topology:  $\{U_\alpha \mid \alpha \in B\}$  is a basis for the topology on  $T_B$ . We check the properties of a basis 4:

1.  $P \in U_\alpha \cap U_\beta$ , then there are  $\gamma \geq \alpha$  and  $\delta \geq \beta$  such that  $\gamma, \delta \in P$ . Let  $\delta \geq \gamma$  then  $P \in U_\delta$  and  $U_\delta \subseteq U_\alpha \cap U_\beta$ .
2. For any path  $P$  and any  $\alpha \in P$  we have  $P \in U_\alpha$ .

We next turn to the definition of the truth values. Put  $\llbracket \varphi \rrbracket = \cup\{U_\alpha \mid \alpha \Vdash \varphi\}$  for atomic  $\varphi$ . We thus obtain a canonical topological model  $T_B$ .

**THEOREM 17.** *For the topological model  $T_B$  the identity  $\llbracket \varphi \rrbracket = \cup\{U_\alpha \mid \alpha \Vdash \varphi\}$  holds for all sentences  $\varphi$ .*

**Proof.** Induction on  $\varphi$ .

For atoms the identity holds by definition.  $\vee$  and  $\wedge$  are simple.

Consider  $\rightarrow$ . We must show  $U_\alpha \subseteq \llbracket \varphi \rightarrow \psi \rrbracket \Leftrightarrow \alpha \Vdash \varphi \rightarrow \psi$ .

We use a small topological lemma:  $U \subseteq \text{Int}(V^c \cup W) \Leftrightarrow U \cap V \subseteq W$ , cf. the proof of Theorem 8.

$$\text{So, } U_\alpha \subseteq \llbracket \varphi \rightarrow \psi \rrbracket \Leftrightarrow U_\alpha \subseteq \text{Int}(\llbracket \varphi \rrbracket^c \cup \llbracket \psi \rrbracket) \Leftrightarrow U_\alpha \cap \llbracket \varphi \rrbracket \subseteq \llbracket \psi \rrbracket.$$

We want to show  $\beta \Vdash \varphi \Rightarrow \beta \Vdash \psi$  for all  $\beta \geq \alpha$ . So let  $\beta \Vdash \varphi$ . then by induction hypothesis,  $U_\beta \subseteq \llbracket \varphi \rrbracket$ , and by  $\beta \geq \alpha$   $U_\beta \subseteq U_\alpha$ . Therefore  $U_\beta \cap \llbracket \varphi \rrbracket = U_\beta \subseteq \llbracket \psi \rrbracket$ , i.e.  $\beta \Vdash \psi$ . Conversely we have to show  $U_\alpha \cap \llbracket \varphi \rrbracket \subseteq \llbracket \psi \rrbracket$ . since the  $U_\alpha$ 's form a basis it suffices to show  $U_\beta \subseteq U_\alpha \cap \llbracket \varphi \rrbracket \rightarrow U_\beta \subseteq \llbracket \psi \rrbracket$ , but  $U_\beta \subseteq \llbracket \varphi \rrbracket$  implies  $\beta \Vdash \varphi$ , and hence  $\beta \Vdash \psi$ , which in turn implies,  $U_\beta \subseteq \llbracket \psi \rrbracket$ .

The quantifier cases are simple, we leave them to the reader. ■

**COROLLARY 18.** *For the topological interpretation the completeness theorem holds, i.e.  $\Gamma \vdash \varphi \Leftrightarrow \Gamma \Vdash \varphi$ .*

**Proof.** Soundness is shown by a routine induction. Completeness follows from the completeness of the Beth semantics and Theorem 17. ■

We have introduced a number of semantics each of which has certain drawbacks. For designing counterexamples and straightforward theoretical applications the Kripke semantics is the most convenient one. We will demonstrate this below in a few examples.

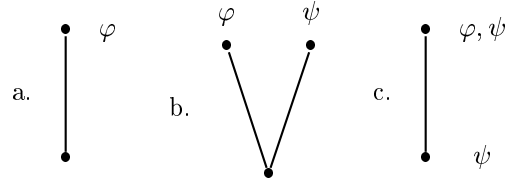
**EXAMPLE 19.** The following, classically valid, sentences are not derivable.

1.  $\varphi \vee \neg\varphi$  (principle of the excluded middle, PEM)
2.  $\neg\neg\varphi \rightarrow \varphi$  (double negation principle)
3.  $\neg(\varphi \wedge \psi) \rightarrow \neg\varphi \vee \neg\psi$  (De Morgan's Law)
4.  $\neg\varphi \vee \neg\neg\varphi$
5.  $(\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$  (Dummett's axiom)
6.  $(\neg\neg\varphi \rightarrow \varphi) \rightarrow \varphi \vee \neg\varphi$
7.  $(\neg\varphi \rightarrow \neg\psi) \rightarrow (\psi \rightarrow \varphi)$
8.  $(\varphi \rightarrow \psi) \rightarrow \neg\varphi \vee \psi$
9.  $\neg\forall x\varphi(x) \rightarrow \exists x\neg\varphi(x)$
10.  $\forall x\neg\neg\varphi(x) \rightarrow \neg\neg\forall x\varphi(x)$  (double negation shift, DNS)
11.  $\forall x(\varphi \vee \psi(x)) \rightarrow \varphi \vee \forall x\psi(x)$  (constant domain axiom)
12.  $(\varphi \rightarrow \exists x\psi(x)) \rightarrow \exists x(\varphi \rightarrow \psi(x))$  (independence of premiss principle, IP)
13.  $(\forall x\varphi(x) \rightarrow \psi) \rightarrow \exists x(\varphi(x) \rightarrow \psi)$
14.  $\forall x(\varphi(x) \vee \neg\varphi(x)) \wedge \neg\neg\exists x\varphi(x) \rightarrow \exists x\varphi(x)$

15.  $\neg\neg\forall xy(x = y \vee x \neq y)$

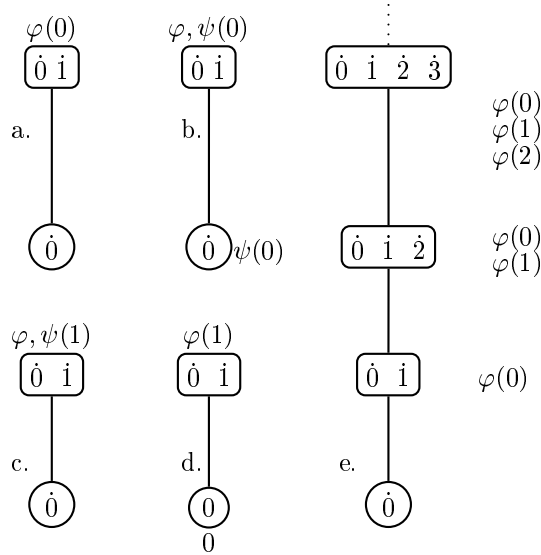
16.  $\neg\neg\forall xy(\neg x \neq y \rightarrow x \neq y)$

**Proof.** Consider the following Kripke models (where the nodes are labelled with the forced atoms and forced formulas).



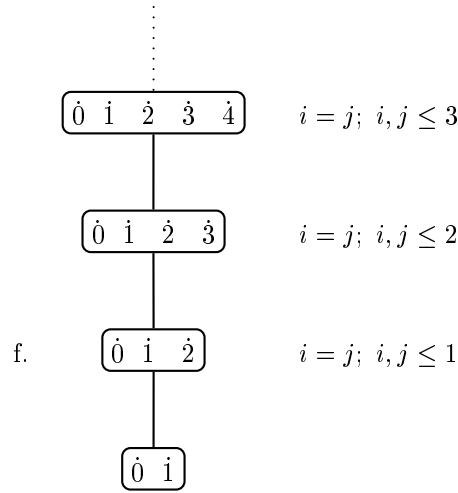
1 and 2 are refuted by model a.  
 4 and 6 are refuted by model b. (forget about the  $\psi$ ).  
 3 and 5 are refuted by model b.  
 7 is refuted by model c.  
 8 is refuted by model a (take  $\psi := \varphi$ ).

For the quantified sentences we need to indicate universes.



9 and 13 are refuted in model a.  
 10 is refuted in model e.  
 11 is refuted in model b.





- 12 is refuted in model c.
- 14 is refuted in model d.
- 15 and 16 are refuted in model f.

The identity relation satisfies the obvious axioms of reflexivity, symmetry, transitivity and compatibility with basic relations. Model f clearly satisfies these axioms.

Observe that we could have refuted 9, 11, 12, 13, 14 by the familiar reduction of a quantified statement to a proposition mimicking a finite domain. Sentence 10 is of a different ilk, we can even show that 10 is true in all finite Kripke models (i.e. with a finite tree).

In a finite tree each node is dominated by an end (or top) node. Suppose that  $\forall x \neg \neg \varphi(x)$  holds, then in an end node  $\alpha$  we have  $\alpha \Vdash \forall x \neg \neg \varphi(x)$ , i.e.  $\forall a \in D\alpha, \alpha \Vdash \neg \neg \varphi(a)$ . But, since  $\alpha$  is an end node, this implies  $\alpha \Vdash \varphi(a)$  hence  $\alpha \Vdash \forall x \varphi(x)$ . As a result we get  $\alpha_0 \Vdash \neg \neg \forall x \varphi(x)$  for the bottom node.

As we will show in the next section, **IPC** is complete for finite Kripke models, so **IQC** essentially needs a wider class of partially ordered sets for its Kripke semantics. ■

*Heyting algebras, the common generalization of the preceding semantics.*

Boole's discovery of the algebraic nature of the logical laws and operations was repeated for the case of intuitionistic logic by McKinsey, Stone, Tarski and others. The resulting algebra has been called *closure algebra*, *Brouwerian algebra*, *pseudo-Boolean algebra*, but nowadays the term *Heyting algebra* is generally accepted.

There are various axiomatisations for the theory of Heyting algebras (cf. [Rasiowa and Sikorski, 1963; Johnstone, 1982]), we will use one that stays very close to the axioms of **IPC**.

For the formulation it is convenient to use the notion of lattice.

DEFINITION 20.  $\langle A, \leq \rangle$  is a lattice if it is a poset in which each pair of elements has a sup and an inf.

We denote the sup and inf of  $x$  and  $y$  by  $x \sqcup y$  and  $x \sqcap y$ . by definition  $\sqcap$  and  $\sqcup$  satisfy

$$\begin{aligned} x \sqcap y &\leq x, \quad y \leq x \sqcup y \\ x, y &\leq z \rightarrow x \sqcup y \leq z \\ z &\leq x, y \rightarrow z \leq x \sqcap y. \end{aligned}$$

We can alternatively obtain a lattice from a structure  $\langle A, \sqcup, \sqcap \rangle$  satisfying

$$\begin{aligned} x \sqcup y &= y \sqcup x & x \sqcap y &= y \sqcap x \\ x \sqcup (y \sqcup z) &= (x \sqcup y) \sqcup z & x \sqcap (y \sqcap z) &= (x \sqcap y) \sqcap z \\ x \sqcap (x \sqcup y) &= x & x \sqcup (x \sqcap y) &= x. \end{aligned}$$

We define the relation ‘ $\leq$ ’ by  $x \leq y := x \sqcap y = x$ . It is a simple exercise to show that  $\leq$  defines a lattice (cf. [Rasiowa and Sikorski, 1963, pp. 35,36]). A lattice with top  $\top$  and bottom  $\perp$  is a lattice with two elements  $\top$  and  $\perp$  satisfying  $\perp \leq x \leq \top$  for all  $x$ .

Note that we can show  $x \sqcap y = x \Leftrightarrow x \leq y \Leftrightarrow x \sqcup y = y$ , so the ordering can also be expressed by  $\sqcup$ .

DEFINITION 21. A Heyting-algebra is a structure  $\langle A, \sqcap, \sqcup, \Rightarrow, \top, \perp \rangle$  such that

1. it is a distributive lattice with respect to  $\sqcap, \sqcup$  and with top and bottom.
2.  $x \sqcap (x \Rightarrow y) = x \sqcap y$
3.  $(x \Rightarrow y) \sqcap y = y$
4.  $(x \Rightarrow y) \sqcap (x \Rightarrow z) \Rightarrow (x \Rightarrow (y \sqcap z))$
5.  $\perp \sqcap x = \perp$
6.  $\perp \Rightarrow \perp = \top$ .

Any Boolean algebra obviously is a Heyting algebra. The paradigm of a Heyting algebra is  $\mathcal{O}(X)$ , the set of opens of a topological space  $X$ , where  $U \Rightarrow V$  is defined as in Section 3:  $\text{Int}(U^c \cup V)$ .

We have the following key properties

LEMMA 22.

1.  $x \Rightarrow x = \top$
2.  $x \sqcap y \leq z \leftrightarrow x \leq y \Rightarrow z$ .

We define the complement by  $-x := x \Rightarrow \perp$ .

The obvious connection with logic is via the *Lindenbaum algebra* of a theory. Consider some theory  $T$  in **IPC**, then

$$\varphi \sim \psi := T \vdash_{\mathbf{IPC}} \varphi \leftrightarrow \psi$$

is a congruence relation, as one easily shows.

On the equivalence classes we define a Heyting algebra, by putting

$$\begin{aligned} \varphi / \sim \sqcap \psi / \sim &:= (\varphi \wedge \psi) / \sim \\ \varphi / \sim \sqcup \psi / \sim &:= (\varphi \vee \psi) / \sim \\ \varphi / \sim \Rightarrow \psi / \sim &:= (\varphi \rightarrow \psi) / \sim \\ \perp &:= \perp / \sim \\ \top &:= (\perp \rightarrow \perp) / \sim. \end{aligned}$$

It is a routine matter to show that one thus obtains a Heyting algebra, the so-called Lindenbaum algebra of  $T$ .

*Examples of Heyting algebras*

1.

$$\begin{array}{ccccccc} 0 & 1 & 2 & 3 & \dots & \dots & \omega \\ \bullet & \bullet & \bullet & \bullet & \dots & \dots & \bullet \end{array}$$

Consider the set of natural numbers with a sup  $\omega$  (i.e. the ordinal  $\omega + 1$ ) and define  $n \sqcap m := \min(n, m)$ ,  $n \sqcup m := \max(n, m)$ ,

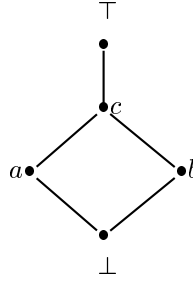
$$\begin{aligned} n \Rightarrow m &:= \begin{cases} m & \text{if } n > m \\ \omega & \text{if } n \leq m, \text{ for } n, m \leq \omega \end{cases} \\ \perp &:= 0, \top := \omega. \end{aligned}$$

The ordering is the natural one. In this Heyting algebra the excluded third fails:

$$-n = n \Rightarrow \perp = \begin{cases} \perp & \text{if } n \neq 0 \\ \top & \text{else.} \end{cases}$$

For  $n \neq \perp, \top$  we get  $n \sqcup -n = n \sqcup \perp = n \neq \top$ .

2. From the diagram below we can read off the operations. The non-trivial one is the ‘implication’ (relative complement).



The relation  $x \leq y \Rightarrow z \Leftrightarrow x \sqcap y \leq z$  tells us that  $y \Rightarrow z$  is the greatest element  $x$  such that  $x \sqcap y \leq z$ , so we can write down the table for  $\Rightarrow$ .

$\Rightarrow$	$\perp$	$a$	$b$	$c$	$\top$
$\perp$	$\top$	$\top$	$\top$	$\top$	$\top$
$a$	$b$	$\top$	$b$	$\top$	$\top$
$b$	$a$	$a$	$\top$	$\top$	$\top$
$c$	$\perp$	$a$	$b$	$\top$	$\top$
$\top$	$\perp$	$a$	$b$	$c$	$\top$

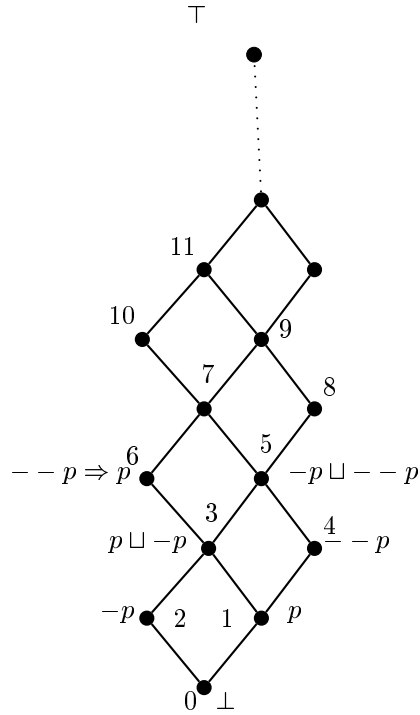
The first column yields the negation. One can view the Heyting algebras as a suitable generalisation of the classical truth table. In this form Heyting algebras occur already in Heyting’s paper of 1930. Truth tables also occur in [Jaskowski, 1936].

3. *The Rieger–Nishimura lattice* [Nishimura, 1966]

In the diagram below one of the two points immediately above the bottom, is the complement of the other. If we call the right hand one  $p$ , we can compute the remaining elements. We enumerate the points as indicated. We put

$$\begin{aligned}
 \varphi_0 &:= \perp \\
 \varphi_1 &:= p \\
 \varphi_2 &:= \neg p \\
 \varphi_{2n+3} &:= \varphi_{2n+1} \sqcup \varphi_{2n+2} \\
 \varphi_{2n+4} &:= \varphi_{2n+2} \Rightarrow \varphi_{2n+1}.
 \end{aligned}$$

The operations on the lattice follow from its order. The Rieger–Nishimura lattice is the free Heyting algebra with one generator, i.e. in logical terms it is the Lindenbaum algebra of **IPC** with just one atom.

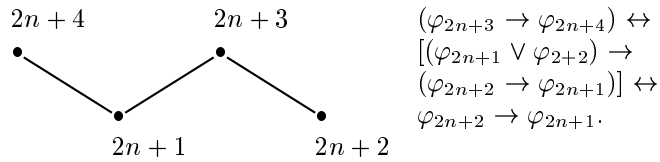


There are two things to be shown:

- (a) each proposition in  $p$  is one of the  $\varphi_i$ 's;
- (b) the dependencies between the  $\varphi_i$ 's are as shown in the diagram.

(a) is shown by induction on  $\varphi$ . We'll do one case. Let  $\varphi = \psi \wedge \sigma$ . By induction hypothesis  $\vdash \psi \leftrightarrow \varphi_i, \vdash \sigma \leftrightarrow \varphi_j$  for some  $i, j$ . If the elements  $i, j$  are comparable, then we immediately see that  $\varphi$  is a  $\varphi_k$ . So the interesting cases are  $i = 2n + 1, j = 2n + 2$  and  $i = 2n + 3, j = 2n + 4$ . In the first case  $\vdash \varphi \leftrightarrow \varphi_{2n-1}$ , in the second case  $\vdash \varphi \leftrightarrow \varphi_{2n+1}$ .

The proof of (b) is a matter of tedious bookkeeping. Given the dependencies between  $\varphi_0, \varphi_1, \varphi_2$ , one checks the dependencies for higher  $\varphi_n$ 's inductively. Consider for example  $\varphi_{2n+3}$  and  $\varphi_{2n+4}$ .



$$\begin{aligned}
 & (\varphi_{2n+3} \rightarrow \varphi_{2n+4}) \leftrightarrow \\
 & [(\varphi_{2n+1} \vee \varphi_{2n+2}) \rightarrow \\
 & (\varphi_{2n+2} \rightarrow \varphi_{2n+1})] \leftrightarrow \\
 & \varphi_{2n+2} \rightarrow \varphi_{2n+1}.
 \end{aligned}$$

So, from the induction hypothesis  $\not\vdash \varphi_{2n+2} \rightarrow \varphi_{2n+1}$ , we obtain  $\not\vdash \varphi_{2n+3} \rightarrow \varphi_{2n+4}$ , i.e.  $\varphi_{2n+3}/\sim \Rightarrow \varphi_{2n+4}/\sim \neq \top$ , i.e.  $\varphi_{2n+3}/\sim \not\leq \varphi_{2n+4}/\sim$ .

An interpretation of **IPC** in a Heyting algebra  $A$  is given by a mapping  $h$  from the atoms into  $A$ .  $h$  is then extended to all propositions in the canonical way i.e.  $h(\varphi \wedge \psi) = h(\varphi) \sqcap h(\psi)$ ,  $h(\varphi \vee \psi) = h(\varphi) \sqcup h(\psi)$ ,  $h(\varphi \rightarrow \psi) = h(\varphi) \Rightarrow h(\psi)$ ;  $\varphi$  is true in  $A$  if for all interpretations  $h$ ,  $h(\varphi) = \top$ . A simple inductive proof shows the

**SOUNDNESS THEOREM IPC**  $\vdash \varphi \Rightarrow \varphi$  is true in all Heyting algebras.

The converse also holds, for consider the Lindenbaum algebra of **IPC** and interpret each proposition canonically:  $h(\sigma) = \sigma/\sim$ , then **IPC**  $\vdash \sigma \Leftrightarrow h(\sigma) = \top$ . So  $\varphi$  is true in the Lindenbaum algebra.

Hence we have the

**COMPLETENESS THEOREM FOR HEYTING ALGEBRAS. IPC**  $\vdash \varphi \Leftrightarrow \varphi$  is true in all Heyting algebras.

There is a simple connection between Kripke models and Heyting algebras. We can associate to a Kripke model a topological space as follows. The points of the space are the nodes of the poset; the opens are the sets  $U$  with the property  $\alpha \in U \wedge \beta \geq \alpha \Rightarrow \beta \in U$ . As in the case of the topological model associated to a Beth model over a tree, the sets  $U_\alpha = \{\beta \mid \beta \geq \alpha\}$  form a basis for this topology.

For atoms we define  $\llbracket \varphi \rrbracket = \{\alpha \mid \alpha \Vdash \varphi\} (*)$ .

One shows by induction on  $\varphi$  that  $(*)$  holds for all propositions (cf. also [Fitting, 1969, p.23]). Thus we have associated to each Kripke model an interpretation in the Heyting algebra of the opens of the associated topological space.

Instead of considering Kripke or Beth models with a prescribed interpretation (forcing) of the atoms, we can also consider the underlying poset only. We then speak of a *Kripke (Beth) frame*. A frame is thus turned into a model by assigning structures to the nodes.

There is an alternative formulation of Kripke (Beth, etc.) models, that sticks closer to the language. Instead of assigning classical structures to nodes, one can just as well assign sets of atoms to nodes, e.g. think of  $V(\alpha)$  as the set of atomic sentences that are true in  $D(\alpha)$ . So  $V$  is a function from  $M$  to the power set of the set of closed atoms, subject to the condition that  $\alpha \leq \beta \Rightarrow V(\alpha) \subseteq V(\beta)$ .

Alternatively one can define a binary interpretation function  $i : At \times M \rightarrow \{0, 1\}$  (where  $At$  is the set of closed atoms), such that  $\alpha \leq \beta$  and  $i(\varphi, \alpha) = 1 \Rightarrow i(\varphi, \beta) = 1$  (think of  $i(\varphi, \alpha) = 1$  as  $D(\alpha) \models \varphi$ ).

### 4.1 An External View of Kripke Models

If one looks at a Kripke model from the outside, then it appears as a complicated concoction of classical structures, and hence as a classical structure itself. Such a structure has its own language and we can handle it by ordinary, classical, model-theoretical means.

What is involved in this ‘master structure’ of  $\mathcal{K}$ ? (i) the partially ordered set of nodes, (ii) the relations between these structures. We can simply describe this master structure  $\mathcal{K}^*$  by a language, containing two sorts of individuals (or alternatively one sort, but two predicates  $N(x)$  and  $E(x)$ , for ‘ $x$  is a node’ and ‘ $x$  is an element’). Let us use  $\alpha, \beta, \gamma, \dots$  for the ‘node-sort’ and  $x, y, z, \dots$  for the ‘element sort’. Then we add  $\leq$  to the original language, and replace each predicate symbol  $P$  by  $P^*$  with one more argument than  $P$  and add a domain predicate symbol. The structure  $\mathcal{K}^*$  validates the following laws (referred to by  $\Sigma$ ):

$$\begin{aligned} \alpha \leq \beta \wedge \beta \leq \gamma &\rightarrow \alpha \leq \gamma \\ \alpha \leq \beta \wedge \beta \leq \alpha &\rightarrow \alpha = \beta \\ \forall \alpha \beta \vec{x} (\alpha \leq \beta \wedge P^*(\alpha, \vec{x}) &\rightarrow P^*(\beta, \vec{x})) \\ \forall \alpha \beta x (D(\alpha, x) &\rightarrow D(\beta, x)) \end{aligned}$$

Now we can mimic the forcing clauses in the extended language. Consider the translation of  $\alpha \Vdash \varphi$  given by the inductive definition:

1.  $(\alpha \Vdash P(\vec{t}))^* := P^*(\alpha, \vec{t})$  and  $(\alpha \Vdash \perp)^* := \perp$ .
2.  $(\alpha \Vdash \varphi \wedge \psi)^* := (\alpha \Vdash \varphi)^* \wedge (\alpha \Vdash \psi)^*$ .
3.  $(\alpha \Vdash \varphi \vee \psi)^* := (\alpha \Vdash \varphi)^* \vee (\alpha \Vdash \psi)^*$ .
4.  $(\alpha \Vdash \varphi \rightarrow \psi)^* := \forall \beta \geq \alpha ((\beta \Vdash \varphi)^* \rightarrow (\beta \Vdash \psi)^*)$ .
5.  $(\alpha \Vdash \exists x \varphi(x))^* := \exists x (D(\alpha, x) \wedge (\alpha \Vdash \varphi(x))^*)$ .
6.  $(\alpha \Vdash \forall x \varphi(x))^* := \forall \beta \geq \alpha \forall x (D(\beta, x) \rightarrow (\beta \Vdash \varphi(x))^*)$ .

It is obvious that:

1.  $\alpha \Vdash \varphi \Leftrightarrow \mathcal{K}^* \models (\alpha \Vdash \varphi)^*$
2. each model of  $\Sigma$  corresponds uniquely to a Kripke model.

Now we can apply the full force of classical model theory to the models of  $\Sigma$  in order to obtain results about Kripke models. For example, one gets for free the ultraproduct theorem and the Hilbert–Bernays completeness theorem (consistent RE theories have  $\Delta_2^0$  models, cf. [Kleene, 1952, Ch XIV]).

Similar ‘translations’ can be applied to Beth semantics or the general semantics (cf. [van Dalen, 1978] for an application to lawless sequences).

#### 4.2 Model theory of intuitionistic logic in an intuitionistic setting

If one is willing to give up the strong results of all the artificial semantics (completeness, Skolem–Löwenheim, etc.), there is no reason why one should not practise model theory of intuitionistic theories as an ordinary part of intuitionistic mathematics. That is to say, to adopt an intuitionistic variant of the Tarskian semantics. A number of interesting results have been obtained for specific theories and structures, e.g., the continuum and the irrationals are elementarily equivalent for the theories of equality, apartness and linear order. Note that even a seemingly trivial theory, such as that of equality, turns out to be highly complicated—in contrast to the classical case. Also, strong classical theorems cannot always be upheld in an intuitionistic setting. e.g. the existence of winning strategies for Ehrenfeucht–Fraïssé games implies elementary equivalence (cf. [van Dalen, 1993]), but the converse fails (cf. [Veldman and Waaldijk, 1996]). The last mentioned paper contains a wealth of interesting methods and results, it is recommended for getting acquainted with the field.

### 5 SOME METALOGICAL PROPERTIES OF IPC AND IQC

Intuitionistic logic is in a sense richer in metalogical properties than classical logic. There are common properties, such as completeness, compactness and deduction theorem, but soon the logics start to diverge. Classical logic has phenomena such as prenex normal forms, Skolem form, and Herbrand’s theorem which are absent in intuitionistic logic. Intuitionistic logic on the other hand is more blessed with derived rules.

The first example is the

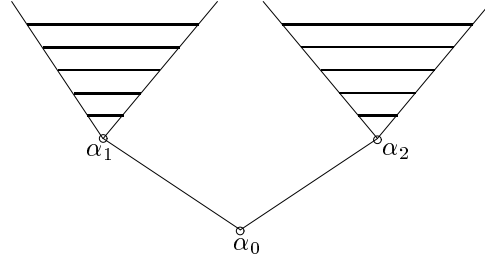
DISJUNCTION PROPERTY, DP.  $\Gamma \vdash \varphi \vee \psi \Rightarrow \Gamma \vdash \varphi$  or  $\Gamma \vdash \psi$ .

Clearly, the nature of  $\Gamma$  is relevant, for if  $\Gamma$  contains all instances of PEM, then DP is false, since in **CPC**  $\varphi \vee \neg\varphi$  is a tautology, but neither  $\varphi$ , nor  $\neg\varphi$  needs to be a tautology.

A sufficient condition on  $\Gamma$  is that it exists of *Harrop formulas*, i.e. formulas without dangerous occurrences of  $\vee$  or  $\exists$ . To be precise, the class of Harrop formulas is inductively defined by

1.  $\varphi \in H$  for atomic  $\varphi$
2.  $\varphi, \psi \in H \rightarrow \varphi \wedge \psi \in H$
3.  $\varphi \in H \Rightarrow \forall x\varphi \in H$
4.  $\psi \in H \Rightarrow \varphi \rightarrow \psi \in H$ .





**THEOREM 23.** *The disjunction property holds for sets  $\Gamma$  of Harrop formulas.*

For a proof using natural deduction, see [Prawitz, 1965, p. 55], [van Dalen, 1997, p. 209]. In Aczel [1968] a proof is given using a metamathematical device ‘Aczel’s slash’. See also [Gabbay, 1981, Ch. 2, Section 3].

The intuitionistic reading of the disjunction property is: given a proof of  $\varphi \vee \psi$  we can effectively find a proof of  $\varphi$  or a proof of  $\psi$ . The proof theoretical demonstrations of DP have this intuitionistic character, not however the model-theoretic proof below. The proof uses classical meta-theory, to be specific, it uses *reductio ad absurdum*.

To demonstrate the use of Kripke models, we give the proof for a simple case,  $\Gamma = \emptyset$ .

Let  $\vdash \varphi \vee \psi$  and suppose  $\not\vdash \varphi$  and  $\not\vdash \psi$ . Then there are Kripke models  $\mathcal{K}_1$  and  $\mathcal{K}_2$  with bottom node  $\alpha_1$  and  $\alpha_2$  such that  $\alpha_1 \not\vdash \varphi$  and  $\alpha_2 \not\vdash \psi$ . We construct a new Kripke model  $\mathcal{K}$  by taking the disjoint union of  $\mathcal{K}_1$  and  $\mathcal{K}_2$  and placing an extra node  $\alpha_0$  at the bottom, see figure above. We stipulate that nothing is forced at  $\alpha_0$ . Clearly, the result is a Kripke model.

$\alpha_0 \Vdash \varphi \vee \psi$ , so  $\alpha_0 \Vdash \varphi$  or  $\alpha_0 \Vdash \psi$ . If  $\alpha_0 \Vdash \varphi$ , then  $\alpha_1 \Vdash \varphi$ . Contradiction. And if  $\alpha_0 \Vdash \psi$ , then  $\alpha_2 \Vdash \psi$ . Contradiction. Hence we have  $\vdash \varphi$  or  $\vdash \psi$ .

For predicate logic we can also establish the *Existence Property*:  $\Gamma \vdash \exists x\varphi(x) \Rightarrow \Gamma \vdash \varphi(t)$  for a chosen term  $t$ , where  $\Gamma$  consists of Harrop formulas ( $\exists x\varphi(x)$  is closed). See [Prawitz, 1965; Aczel, 1968; Gabbay, 1981; van Dalen, 1997].

Since the only closed terms in our present approach are constants, we can replace the conclusion of EP by ‘ $\Gamma \vdash \varphi(c)$  for a constant  $c$ ’.

In the case that there are no constants at all the conclusion is rather surprising:  $\Gamma \vdash \forall x\varphi(x)$ .

Like its classical counterpart **IPC** is decidable; there are various proofs for this fact. In [Kleene, 1952, Section 80], [Troelstra and van Dalen, 1988, p. 541] and [Szabo, 1969, p. 103], a sequent calculus is used. The use of normal derivations in natural deduction likewise yields a decision procedure. In [Rasiowa, 1974, p. 266] decidability is derived from the completeness of **IPC** for finite Heyting algebras. We will use a similar argument based on Kripke models.

Our first step is to reduce Kripke models for **IPC** to finite models, following [Smoryński, 1973].

We consider a Kripke model  $\mathcal{K}$  with a tree as its underlying poset such that  $K \not\vdash \varphi$ ; a suitable refining will yield a ‘submodel’  $\mathcal{K}^*$ , such that

1.  $\mathcal{K}^*$  is finite
2.  $\alpha \Vdash^* \psi \Leftrightarrow \alpha \Vdash \psi$ , for all subformulas of  $\varphi$ .

Let  $S$  be the set of subformulas of  $\varphi$ , and put  $S_\beta = \{\psi \in S \mid \beta \Vdash \psi\}$ . We define a sequence of sets  $K_n : K_0 = \{\alpha_0\}$  ( $\alpha_0$  is the bottom node of  $\mathcal{K}$ ).

Let  $K_n$  be defined, and  $\beta \in K_n$ . We consider sets  $\{\delta_1, \dots, \delta_k\} \subseteq K$  such that

1.  $\beta \leq \delta_i$
2.  $S_\beta \neq S_{\delta_i}$
3. the  $S_\delta$  jumps only once between  $\beta$  and  $\delta_i$ , i.e.  $S_\delta = S_\beta$  or  $S_\delta = S_{\delta_i}$  for  $\beta \leq \delta \leq \delta_i$
4.  $S_{\delta_i} \neq S_{\delta_j}$  for  $i \neq j$ .

Since there are only finitely many  $S_\delta$ ’s we can find a maximal such set say  $\{\beta'_1, \dots, \beta'_k\}$ , if there are such  $\delta$ ’s at all.

Define

$$\begin{aligned} K_1 &= \{\alpha'_{0,1}, \dots, \alpha'_{0,k}\} \cup \{\alpha_0\} \\ K_{n+1} &= K_n \cup \bigcup \{\{\beta'_2, \dots, \beta'_k\} \mid \beta \in K_n - K_{n-1}\}, n \geq 1. \end{aligned}$$

As the  $S_\beta$ ’s increase, and there are only finitely many subformulas, the sequence  $K_n$  stops eventually. Clearly each  $K_n$  is finite, hence  $K^* = \bigcup K_n$  is finite.

*Claim:*  $K^*$  with its inherited  $\Vdash^*$  is the required finite submodel. Property (2) is shown by induction on  $\psi$ . For atomic  $\psi$  (2) holds by definition. For  $\vee$  and  $\wedge$  the result follows immediately. Let us consider  $\psi_1 \rightarrow \psi_2$ . Suppose that for  $\beta \in K^*$ ,  $\beta \not\vdash \psi_1 \rightarrow \psi_2$ , then there is a  $\gamma \geq \beta$  in  $K$  such that  $\gamma \Vdash \psi_1$  and  $\gamma \not\vdash \psi_2$ . If  $\psi_1 \in S_\beta$  we are done. Else we find by our construction a  $\delta \in K^*$  with  $\beta < \delta \leq \gamma$  such that  $\psi_1 \in S_\delta$  and  $\psi_2 \notin S_\delta$ , hence  $\beta \not\vdash^* \psi_1 \rightarrow \psi_2$ . The converse is simple.

We now may conclude.

**THEOREM 24.** *IPC is complete for finite Kripke models over trees.*

**Proof.** By the above and Lemma 14. ■

As a consequence we get

**COROLLARY 25.** *IPC is decidable.*

**Proof.** We can effectively enumerate all finite Kripke models over trees, and hence effectively enumerate all refutable propositions. By enumerating all proofs in **IPC** we also obtain an effective enumeration of all provable propositions. By performing these enumerations simultaneously we obtain an effective test for provability in **IPC**. ■

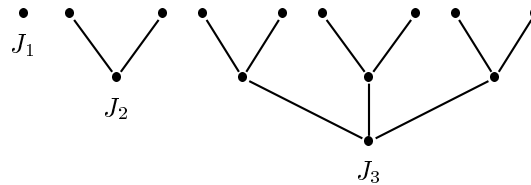
Theorem 24 is also paraphrased as ‘**IPC** has the Finite Model Property (FMP)’, i.e. **IPC**  $\not\vdash \varphi \Rightarrow \varphi$  is false in a finite model. The FMP is the key concept in our decidability proof. Note that the decision procedure of Corollary 25 is horribly inefficient. The procedures based on sequent calculus or natural deduction are much more practical.

Corollary 25 can be considerably improved, in the sense that narrower classes of Kripke models can be indicated for which **IPC** is complete.

EXAMPLES.

1. **IPC** is complete for the Jaskowski sequence  $J_n$ . The sequence  $J_n$  is defined inductively.  $J_1$  is the one point tree.

$J_{n+1}$  is obtained from  $J_n$  by taking  $n+1$  disjoint copies  $J_n$  and adding an extra bottom node.



Cf. [Gabbay, 1981, p. 70 ff.]. The Jaskowski sequence is the Kripke model version of Jaskowski’s original sequence of truth tables, [Jaskowski, 1936].

2. **IPC** is complete for the full binary tree (cf. [Gabbay, 1981, p. 72]).

Strictly speaking we have given classes of Kripke frames, where completeness with respect to a class **K** of frames means ‘completeness with respect to all Kripke models over frames from **K**’.

During the early childhood of intuitionism and its logic it was put forward by some mathematicians that intuitionistic logic actually is a three-valued logic with values true, false, undecided. This proposal is wrong on two counts, it is philosophically wrong and by a result of Gödel no finite truth table completely characterizes intuitionistic logic (see Section 5).

Our comments on the failure of the *double negation shift*, DNS, (Section 3.11-10) have already made it clear that **IQC** is not complete for finite

Kripke frames. The usual refinement of the completeness proof tells us that (for a countable language) **IQC** is complete for countable Kripke models over trees.

Intuitionistic predicate calculus differs in a number of ways from its classical counterpart. Although both **IQC** and **CQC** are undecidable, monadic **IQC** is undecidable (Kripke) (cf. [Gabbay, 1981, p. 234]), whereas the monadic fragment of **CQC** is decidable (Behmann). Another remarkable result is the decidability of the prenex fragment of **IQC**, which implies that not every formula has a prenex normal form to which it is equivalent in **IQC**.

We will consider the class of prenex formulas below.

**LEMMA 26.**  $\mathbf{IQC} \vdash \exists y\varphi(x_1, \dots, x_n, y) \Rightarrow \mathbf{IQC} \vdash \forall x_1, \dots, x_n\varphi(x_1, \dots, x_n, t)$ , where all variables in  $\varphi$  are shown, and where  $t$  is either a constant or one of the variables  $x_1, \dots, x_n$ .

**Proof.** Add new constants  $a_1, \dots, a_n$ , then  $\mathbf{IQC} \vdash \exists y\varphi(a_1, \dots, a_n, y)$  and apply EP. ■

We now get the following intuitionistic version of the Herbrand Theorem.

**THEOREM 27.** Let  $Q_1x_1, \dots, Q_nx_n\varphi$  be a prenex sentence, then  $\mathbf{IQC} \vdash Q_1x_1, \dots, Q_nx_n\varphi$  iff  $\mathbf{IPC} \vdash \varphi'$ , where  $\varphi'$  is obtained from  $\varphi$  by replacing the universally quantified variables by distinct new constants, and the existentially quantified variables by suitable old or new constants.

**Proof.** Induction on  $n$ . Use EP and Lemma 26. ■

As a corollary of Theorem 27 and Corollary 25 we get

**THEOREM 28.** *The prenex fragment of **IQC** is decidable.*

and

**COROLLARY 29.** *There is not for every  $\varphi$  a prenex  $\psi$  such that  $\mathbf{IQC} \vdash \varphi \leftrightarrow \psi$ .*

Among the properties that classical and intuitionistic logic share is the so-called

**THEOREM 30 (Interpolation Theorem).** *If  $\mathbf{IQC} \vdash \varphi \rightarrow \psi$ , then there exists a  $\sigma$ , called an interpolant of  $\varphi \rightarrow \psi$ , such that*

1.  $\mathbf{IQC} \vdash \varphi \rightarrow \sigma$  and  $\mathbf{IQC} \vdash \sigma \rightarrow \psi$
2. all non-logical symbols in  $\sigma$  occur in  $\varphi$  and in  $\psi$ .

The interpolation theorem was established by proof theoretical means by [Schütte, 1962] and [Prawitz, 1965]. Gabbay [1971] proved the theorem by

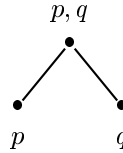
model theory, he also established a suitable form of Robinson’s consistency theorem.

For proofs and refinements the reader is referred to [Gabbay, 1981, Chapter 8], and [Troelstra and Schwichtenberg, 1996, §4.3], whereas in **CPC** the interpolation theorem holds in all fragments. Zucker has shown this not to be the case for **IPC** (cf. [Renardel de Lavalette, 1981]).

### 5.1 Independence of the Propositional Connectives

Whereas in classical logic the propositional connectives are interdefinable, this is not the case in **IPC**, a fact already known to McKinsey [1939]. There are a number of ways to show the independence of the intuitionistic connectives. A proof theoretical argument, based on the normal form theorem, is given by [Prawitz, 1965, p. 59 ff]. We will use some *ad hoc* considerations.

1. The independence of  $\vee$  from  $\rightarrow, \wedge, \neg, \perp$  is clear, since  $\rightarrow, \wedge, \neg, \perp$  are preserved under the double negation translation (up to provable equivalence), but  $\vee$  is not.
2.  $\neg$  is independent from  $\vee, \rightarrow, \wedge$  already in **CPC**, so let alone in **IPC**.
3.  $\rightarrow$  is independent from  $\wedge, \vee, \neg$ . We use the simple fact that for  $\rightarrow$ -free  $\varphi, \vdash (p \rightarrow q) \rightarrow \varphi \Rightarrow \vdash (p \rightarrow \neg\neg q) \rightarrow \varphi$ . Definability of  $\rightarrow$  would yield  $\vdash (p \rightarrow \neg\neg q) \rightarrow (p \rightarrow q)$ .
4.  $\wedge$  is independent of  $\vee, \rightarrow, \neg, \perp$ . Consider the Kripke model



A simple inductive argument shows that the  $\wedge$ -free formulas are either equivalent to  $\perp$  or are forced in at least one of the lower nodes.

Although even the traditional definability result fail in intuitionistic logic, there is a completeness of the sets  $\{\rightarrow, \wedge, \vee, \perp\}$  for **IPC** or  $\{\rightarrow, \wedge, \vee, \perp, =, \exists, \forall\}$  for **IQC** under special assumptions. Zucker and Tragesser [1978] showed that logical constants, given by Natural Deduction rules are definable in the above sets. A similar result is to be found in [Prawitz, 1979].

In view of the incompleteness of the intuitionistic connectives there have been a number of definitions of new connectives, e.g. by model theoretic means (cf. Gabbay [1977; 1981, p. 130 ff], Goad [1978] and de Jongh [1980]). Kreisel introduced the connective  $*$  by a second-order propositional condition:  $*(\varphi := \exists\psi(\varphi \leftrightarrow \neg\psi \vee \neg\neg\psi))$ . Matters of definability, etc. of  $*$  have been extensively investigated in [Troelstra, 1980].

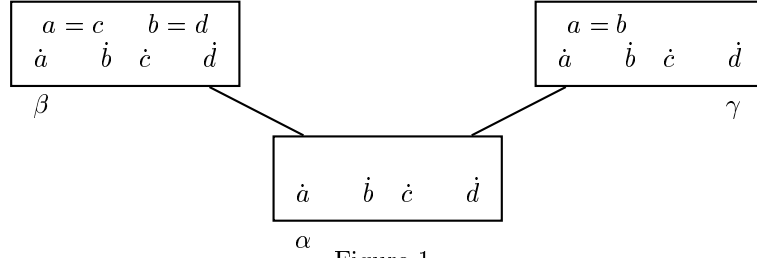


Figure 1.

### 5.2 The Addition of Skolem Functions is not Conservative

It is a fact of classical logic that the extension of a theory by Skolem functions does not essentially strengthen  $T$  (Vol 1, p. 89), i.e. (a simple case) if  $T \vdash \forall x \exists y \varphi(x, y)$  then we may form  $T^S$  by adding a function symbol  $f$  and the axiom  $\forall x \varphi(x, f(x))$  and  $T^S$  is conservative over  $T$ : if  $T^S \vdash \sigma$  where  $\sigma$  does not contain  $f$ , then  $T \vdash \sigma$ . In general this is not true in intuitionistic logic [Minc, 1966]. We will show this by means of a simple counter example of Smoryński [1978].

Consider the theory  $T$  of equality **EQ** plus the extra axiom  $\forall x \exists y (x \neq y)$ , and its Skolem extension  $T^S = \mathbf{EQ} + \forall x (x \neq f(x)) \wedge \forall xy (x = y \rightarrow f(x) = f(y))$ , then  $T^S$  is not conservative over  $T$ .

It suffices to find a statement  $\sigma$  in the language of **EQ** such that  $T^S \vdash \sigma$  and  $T \not\vdash \sigma$ . We take  $\sigma := \forall x_1 \exists y_1 \forall x_2 \exists y_2 [x_1 \neq y_1 \wedge x_2 \neq y_2 \wedge (x_1 = x_2 \rightarrow y_1 = y_2)]$ .

Clearly  $T^S \vdash \sigma$ . The Kripke model of figure 1 establishes  $T \not\vdash \sigma$ .

Clearly  $\alpha \Vdash \forall x \exists y (x \neq y)$ .

Now suppose  $\alpha \Vdash \sigma$ . Take  $a, b$  for  $x_1, x_2$  then we must take  $d, c$  for  $y_1, y_2$  (in that order). However  $\alpha \not\vdash a = b \rightarrow c = d$ .

The equality fragment of  $T^S$  is axiomatised in [Smoryński, 1978].

### 5.3 Fragments of IPC

The situation in intuitionistic logic radically changes if one leaves out some connectives. We mention the following result: (Diego, McKay) there are only finitely many non-equivalent propositions built from finitely many atoms in the  $\vee$ -free fragment (cf. [Gabbay, 1981, p. 80]).

### 5.4 Some Remarks on Completeness and Intuitionistically Acceptable Semantics

This section uses notions of later sections, in particular Section 9. the reader is suggested to consult those sections.

As we have argued in Section 1, an interpretation of the logical constants based on intuitionistic principles must somehow exploit the notion of construction. This has been proposed by Heyting, and extended by Kreisel. It has not (so far), however, led to a flexible semantics that provided logic with completeness. The more successful semantics have provided completeness theorems, but at the price of importing classical metamathematics. This is a matter of considerable philosophical interest. As Intuitionism is a legitimate, well-motivated philosophy, it should at least have a semantics for its logic that stands up to the criteria of the underlying philosophy; unless one adopts Brouwer's radical view that 'mathematics is an essentially languageless activity'. The traditional semantics lend themselves perfectly well to an intuitionistic formulation. One has to select among the various classically equivalent formulations the intuitionistically correct one (e.g. in the topological interpretation  $\llbracket \varphi \rightarrow \psi \rrbracket = \text{Int}\{x \mid x \in \llbracket \varphi \rrbracket \rightarrow x \in \llbracket \psi \rrbracket\}$  and *not*  $\text{Int}(\llbracket \varphi \rrbracket^c \cup \llbracket \psi \rrbracket)$ ). Soundness does not present problems, so independence results can usually be obtained by Intuitionistic means. For the more sophisticated applications of semantics one usually needs completeness, and the original completeness proofs relied heavily on classical logic. For propositional logic the problem is relatively simple.

The first positive result was provided by Kreisel, who in [Kreisel, 1958] interpreted **IPC** by means of lawless sequences, and showed by intuitionistic means **IPC** to be complete for this particular interpretation.

The basic idea is to relate Beth models (which are special cases of topological models) to lawless sequences, considered as paths through the underlying trees; one assigns sets of lawless sequences to propositions,  $\varphi \mapsto \llbracket \varphi \rrbracket$ , cf. Theorem 17, such that the logical operations correspond to the Heyting algebra operations. Since one can restrict oneself to finitely branching trees in this context, one can show completeness for the topological space of lawless sequences using only the simple properties of lawless sequences (including the fan theorem). Kripke [1965] indicates a similar procedure on the basis of Kripke models.

A more serious matter is the completeness of predicate calculus. The plausible approach, i.e. to interpret 'validity' as 'validity in structure *à la* Tarski', called *internal validity* by Dummett [1977, p. 215], led to an unexpected obstacle. Kreisel [1962], following Gödel, established the following result: if **IQC** is complete for internal validity, then  $\forall \xi \neg \neg \exists x \varphi(\xi, x) \rightarrow \forall \xi \exists x \varphi(\xi, x)$  holds for all primitive recursive predicates  $\varphi$ .

So validity of the above kind would give us Markov's 'Principle' (cf. Section 6.5.3), a patently non-intuitionistic principle. It does not do any good to consider Beth semantics, for one can obtain the same fact for validity in all Beth models [Dyson and Kreisel, 1961]. Even worse, under the assumption of Church's Thesis (i.e. all functions from  $\mathbb{N} \rightarrow \mathbb{N}$  are recursive, cf. Chapter 4 of Vol. 1 of this *Handbook*) **IQC** is *incomplete* in the sense that the set

of valid formulae is not recursively enumerable, as established by [Kreisel, 1970] (cf. [van Dalen, 1973; Leivant, 1976]).

The strongest result so far is McCarty's theorem; constructive validity is nonarithmetic, [McCarty, 1988]. This bleak situation in semantics for **IQC** changed when Veldman in 1974 introduced a technical device that allowed for a modified Kripke (and similarly, Beth) semantics for which the completeness of **IQC** can be established in an intuitionistically acceptable manner. Although Veldman's proposal can be implemented in more than one way, its main feature is relaxation of the forcing conditions for atoms:  $\alpha \Vdash \perp$  is in general allowed. For these more general models intuitionistic completeness proofs have been given for the Kripke version by [Veldman, 1976], and for the Beth version by [Swart, 1976].

Extensive discussions of the aspects of intuitionistic completeness of **IQC** are to be found in [Dummett, 1977] and [Troelstra, 1977]. H. Friedman [1977; 1977a] has sketched intuitionistically correct completeness proofs for **MQC** and the  $\perp$  (and  $\neg$ )-free part of **IQC**. The details of a slightly upgraded version can be found in [Troelstra and van Dalen, 1988, §13.2], there the result is cast in the form of a universal Beth model:

1. There is a Beth model  $\mathcal{M}$  such that  $\mathcal{M} \Vdash \varphi \Leftrightarrow \mathbf{IQC} \vdash \varphi$  for all  $\perp$ -free formulas  $\varphi$ .
2. There is a Beth model  $\mathcal{M}$  for minimal logic such that  $\mathcal{M} \Vdash \varphi \Leftrightarrow \mathbf{MQC} \vdash \varphi$  for all  $\varphi$ .
3. there is a *modified* Beth model  $\mathcal{M}^*$  such that  $\mathcal{M}^* \Vdash \varphi \Leftrightarrow \mathbf{IQC} \vdash \varphi$  for all  $\varphi$ .

### 5.5 *The Intuitionistic View of Non-intuitionistic Model Theoretic Methods*

It should not come as a surprise that for intuitionists such semantical proofs as employed, e.g. in the case of DP (cf. Theorem 23) do not carry much weight. After all, one wants to extract a proof of either  $\varphi$  or  $\psi$  from a proof of  $\varphi \vee \psi$ , and the gluing proof does not provide means for doing so. There is however a roundabout way of having one's cake and eating it. For example, in the case of the proof of DP one shows classically that ' $\varphi$  has no proof in **IQC**' ' $\psi$  has no proof in **IQC**' then ' $\varphi \vee \psi$  has no proof in **IQC**', and hence (classically)  $\mathbf{IQC} \vdash \varphi \vee \psi \Rightarrow \mathbf{IQC} \vdash \varphi \vee \mathbf{IQC} \vdash \psi$ .

One formalizes this statement in Peano's Arithmetic, so

$$\mathbf{PA} \vdash \exists x \text{Pr}_{\mathbf{IQC}}(x, \ulcorner \varphi \vee \psi \urcorner) \rightarrow \exists y \text{Pr}_{\mathbf{IQC}}(y, \ulcorner \varphi \urcorner) \vee \exists z \text{Pr}_{\mathbf{IQC}}(z, \ulcorner \psi \urcorner)$$

or

$$\mathbf{PA} \vdash \forall x \exists yz (\text{Pr}_{\mathbf{IQC}}(x, \ulcorner \varphi \vee \psi \urcorner) \rightarrow \text{Pr}_{\mathbf{IQC}}(y, \ulcorner \varphi \urcorner) \vee \text{Pr}_{\mathbf{IQC}}(z, \ulcorner \psi \urcorner)).$$



Now one uses the fact that **PA** is conservative over **HA** for  $\Pi_2^0$  statements, so that  $\mathbf{HA} \vdash \forall x \exists y z (\text{_____})$ .

This shows that DP is intuitionistically correct. In [Smoryński, 1982] problems of this kind is considered in a more general setting. Of course, one might wonder why go through all this rigmarole when direct proofs (e.g. via natural deduction, or slash operations) are available. A matter of taste maybe.

## 6 INTERMEDIATE LOGICS

By adding the principle of the excluded middle to **IPC** we obtain full classical propositional logic. It is a natural question what logics one gets by adding other principles. We will consider extensions of **IPC** by schemas, e.g.  $\mathbf{IPC} + (\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$ . First we remark that all such extensions are subsystems of **CPC**, for let  $T$  be such an extension and suppose that  $T \not\subseteq \mathbf{CPC}$ , then there is a  $\varphi$  such that  $T \vdash \varphi$  (and hence all substitution instances) and  $\varphi$  is not a tautology. but then we find by substituting, say  $p_0 \wedge \neg p_0$  and  $p_0 \rightarrow p_0$  for suitable atoms of  $\varphi$  an instance  $\varphi'$  which is false. therefore  $\mathbf{CPC} \vdash \neg \varphi'$  and, by Glivenko's theorem (Corollary 51)  $\mathbf{IPC} \vdash \neg \varphi'$ . This contradicts  $T \vdash \varphi'$ .

So there are only logics between **IPC** and **CPC** to consider.

The study of intermediate logics is mainly a matter for pure technical logic, dealing with completeness, finite model property, etc. There are however certain intermediate logics that occur more or less naturally in real life (e.g. in the context of Gödel's Dialectica interpretation, or of realizability), so that their study is not merely *l'art pour l'art*. One such instance is Dummett's logic **LC**, which turns up in the provability logic of Heyting's arithmetic (cf. [Visser, 1982]).

One of the most popular topics in intermediate logic was the investigation of classes of semantics for which various logics are complete. Furthermore there is the problem to determine the structure of the family of all intermediate logics under inclusion.

The field has extensively been studied and an even moderately complete treatment is outside the scope of this chapter. the reader is referred to [Rautenberg, 1979] and [Gabbay, 1981].

### 6.1 Dummett's Logic **LC**

DEFINITION.  $\mathbf{LC} = \mathbf{IPC} + (\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$ .

Theorem. **LC** is complete for linearly ordered Kripke models.

One direction is simple, one just checks that  $(\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$  holds in all linearly ordered Kripke models. For the converse, consider the model,

obtained in the Model Existence Lemma 14, consisting of prime theories, ordered by inclusion. The bottom node  $\Gamma_0$  forces all instances of the schema  $(\varphi \rightarrow \psi) \vee (\psi \rightarrow \varphi)$ .

Consider  $\Gamma_1, \Gamma_2$  with  $\varphi \in \Gamma_1 - \Gamma_2$  for some  $\varphi$ . We will show that  $\Gamma_2 \subseteq \Gamma_1$ . Let  $\psi \in \Gamma_2$ . Since  $\Gamma_0 \Vdash \varphi \rightarrow \psi$  or  $\Gamma_0 \Vdash \psi \rightarrow \varphi$  and  $\Gamma_0 \subseteq \Gamma_i (i = 1, 2)$  we have  $\psi \in \Gamma_1$  or  $\varphi \in \Gamma_2$ . As the latter is ruled out we find  $\psi \in \Gamma_1$ . Hence for any two  $\Gamma_1, \Gamma_2$ , we have  $\Gamma_1 \subseteq \Gamma_2$  or  $\Gamma_2 \subseteq \Gamma_1$ .

This establishes the semantic characterisation of **LC**.

## 6.2 Filtration and Minimalisation

Some models are needlessly complicated because some of their nodes are in a sense redundant. A simple case is a model with two nodes  $\alpha < \beta$ , which force exactly the same formulas. The idea to collapse nodes that force the same formulas presents itself naturally. Scott and Lemmon introduced such a procedure in modal logic under the name of filtration [Lemmon and Scott, 1966], and Smoryński did something similar in intuitionistic logic under the name of minimalisation [Smoryński, 1973; Segerberg, 1968]. Let a Kripke model  $\mathcal{K} = \langle K, \leq, \Vdash \rangle$  be given. We consider forcing on  $\mathcal{K}$  for a class of formulas  $\Gamma$  closed under subformulas. For  $\alpha \in K$  define  $[\alpha]_\Gamma := \{\varphi \in \Gamma \mid \alpha \Vdash \varphi\}$ . Put  $K_\Gamma = \{[\alpha]_\Gamma \mid \alpha \in K\}$ ,  $[\alpha]_\Gamma \leq [\beta]_\Gamma$  if  $[\alpha]_\Gamma \subseteq [\beta]_\Gamma$  and  $[\alpha]_\Gamma \Vdash \varphi$  if  $\varphi \in [\alpha]_\Gamma$  for atomic  $\varphi$ . Observe that the mapping  $\alpha \rightarrow [\alpha]_\Gamma$  is a homomorphism of posets.

Obviously  $\mathcal{K}_\Gamma = \langle K_\Gamma, \leq, \Vdash_\Gamma \rangle$  is a Kripke model.

**THEOREM 31.**  $[\alpha]_\Gamma \Vdash_\Gamma \varphi \Leftrightarrow \alpha \Vdash \varphi$  for  $\varphi \in \Gamma$ .

**Proof.** Induction on  $\varphi$ . The only non-trivial case is the implication.

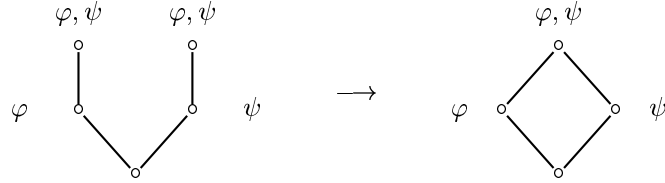
(i)  $\alpha \not\Vdash \varphi \rightarrow \psi \Leftrightarrow \exists \beta \geq \alpha \beta \Vdash \varphi$  and  $\beta \not\Vdash \psi \Leftrightarrow$  (induction hypothesis)  $\Leftrightarrow \exists \beta \geq \alpha ([\beta]_\Gamma \Vdash_\Gamma \varphi$  and  $[\beta]_\Gamma \not\Vdash_\Gamma \psi)$ .

Since  $\beta \geq \alpha$  implies  $[\beta]_\Gamma \supseteq [\alpha]_\Gamma$ , we have  $\alpha \not\Vdash \varphi \rightarrow \psi$ .

(ii)  $\alpha \Vdash \varphi \rightarrow \psi$ . Let  $[\alpha]_\Gamma \subseteq [\beta]_\Gamma$  and  $[\beta]_\Gamma \Vdash_\Gamma \varphi$ . By induction hypothesis  $\beta \Vdash \varphi$  and hence  $\varphi \in [\beta]_\Gamma$ . But  $\varphi \rightarrow \psi \in [\alpha]_\Gamma \subseteq [\beta]_\Gamma$ , so  $\beta \Vdash \psi$  and again by induction hypothesis  $[\beta]_\Gamma \Vdash_\Gamma \psi$ . This shows  $[\alpha]_\Gamma \Vdash_\Gamma \varphi \rightarrow \psi$ . ■

Observe that this procedure does not preserve all desirable properties, e.g. being a tree.

**EXAMPLE.**



Gabbay has refined the notion of filtration in order to obtain models with special properties. For this *selective filtration* cf. [Gabbay, 1981, p. 87 ff.].

### 6.3 The Finite Model Property, FMP

An intermediate logic is said to have the Finite Model Property if it is complete for a class of finite models. We have already seen the importance of the FMP for logic: if  $T$  is effectively axiomatised (RE will do) and has the FMP, then  $T$  is decidable [Harrop, 1958].

The following facts may be helpful in establishing the FMP in some cases.

**THEOREM 32** (Smoryński [1973]).

1. Let  $T$  be complete for a class of Kripke models with posets characterised by positive sentences in a language extended by individual constants, then  $T$  has the FMP.
2. as (1) but with universal sentences and finitely many constants.

**Proof.**

1. Let  $\alpha_0 \not\models \varphi$  for  $\alpha_0$  bottom node of  $\mathcal{K}$ . Apply filtration to  $\mathcal{K}$  and call the result  $\mathcal{K}'$ .  $\mathcal{K}'$  is a homomorphic image of  $\mathcal{K}$  and since positive sentences are preserved under homomorphic images (a simple fact of model theory),  $\mathcal{K}'$  belongs to the given class. Since we only have to consider subformulas of  $\varphi$ ,  $\mathcal{K}'$  evidently is finite.
2. Use the fact that universal sentences are preserved under substructures (cf. [van Dalen, 1997, p. 141, ex. 3]) and apply the construction given in the proof of Theorem 24. ■

**COROLLARY 33.** **LC** has the FMP and is decidable.

### 6.4 The ‘Bounded Height’ models

A Kripke frame (model) is said to have height  $n$  if the maximum length of its chains is  $n$ . If the length of the chains is unbounded, we say that the height is  $\omega$ . Can we find an intermediate logic such that it is complete for all frames of height at most  $n$ ?

We define a sequence of propositions  $\varphi_i$  by:

$$\begin{aligned} \varphi_1 &:= p_1 \vee \neg p_1 \\ \varphi_{n+1} &:= p_{n+1} \vee (p_{n+1} \rightarrow \varphi_n), \end{aligned}$$

where  $p_n$  is the  $n$ th atom.

Let  $\mathbf{BH}_n = \mathbf{IPC} + \varphi_n$ , where we take  $\varphi_n$  to be a *schema* (i.e. we add all substitution instances of  $\varphi_n$  to  $\mathbf{IPC}$ ).

**THEOREM 34.**  $\mathbf{BH}_n$  is complete for all Kripke frames of height  $\leq n$ .

**Proof.** Suppose that  $\mathcal{K}$  has height  $\leq n$  and for some  $\alpha_0 \in K, \alpha_0 \not\Vdash \varphi_n$ . So  $\alpha_0 \not\Vdash \tau_n \rightarrow \varphi_{n-1}, \alpha_0 \not\Vdash \tau_n$  for some  $\tau_n$ . by definition of forcing we find an  $\alpha_1 > \alpha_0$  such that  $\alpha_1 \Vdash \tau_n$  and  $\alpha_1 \not\Vdash \varphi_{n-1}$ . By iterating this step we find an increasing sequence  $\alpha_0 < \alpha_1 < \dots < \alpha_n$ . This contradicts the condition on the heights, so  $\mathcal{K} \Vdash \mathbf{BH}_n$ .

Conversely, we have to show that if  $\mathbf{BH}_n \not\Vdash \sigma$  then there exists a model of height  $\leq n$  which falsifies  $\sigma$ . So let  $\mathcal{K}$  be a Kripke model of  $\mathbf{BH}_n$  and not of  $\sigma$ . We obtain  $\mathcal{K}'$  from it by filtration. It remains to show that  $\mathcal{K}'$  has height  $\leq n$ . Suppose  $\mathcal{K}'$  has a chain  $\alpha_0 < \alpha_1 < \dots < \alpha_n$ . Since  $\mathcal{K}'$  is filtrated we can find atoms  $p_i (i = 1, \dots, n)$ , such that  $\alpha_{n-i+1} \Vdash p_i$  and  $\alpha_{n-i} \not\Vdash p_i$ . So  $\alpha_j \Vdash p_i$  if and only if  $j > n - i$ .

*Claim:*  $\alpha_{n-i} \not\Vdash \varphi_i$ . We show this by induction on  $i$ . By definition  $\alpha_{n-1} \not\Vdash \varphi_1$ .  $\alpha_{n-i-1} \Vdash \varphi_{i+1} \Leftrightarrow \alpha_{n-i-1} \Vdash p_{i+1} \vee (p_{i+1} \rightarrow \varphi_i)$ . Now  $\alpha_{n-i-1} \not\Vdash p_{i+1}$ , and  $\alpha_{n-1} \Vdash p_{i+1}$  but  $\alpha_{n-1} \not\Vdash \varphi_i$ , by induction hypothesis; so  $\alpha_{n-i-1} \not\Vdash \varphi_{i+1}$ . We now may apply the induction principle:  $\alpha_0 \not\Vdash \varphi_n$ . Contradiction.

So  $\mathcal{K}'$  has height  $\leq n$ . ■

**COROLLARY 35.**  $\mathbf{BH}_n$  has the FMP and is decidable.

**Proof.** The posets of height  $\leq n$  are axiomatised by

$$\forall x_0 \dots x_n \left( \bigwedge_{i=0}^{n-1} x_i \leq x_{i+1} \rightarrow \bigvee_{i=0}^{n-1} x_i = x_{i+1} \right).$$

Apply Theorem 32. ■

It is obvious that  $\mathbf{BH}_\omega$  coincides with  $\mathbf{IPC}$ , so only the finite  $\mathbf{BH}_n$ 's are relevant here for us.

Another approach to the bounded height logics is via a sequence of generalisations of *Peirce's law*:

$$\begin{aligned} \pi_1 &= ((p_1 \rightarrow p_0) \rightarrow p_1) \rightarrow p_1 \\ \pi_{n+1} &= ((p_{n+1} \rightarrow \pi_n) \rightarrow p_{n+1}) \rightarrow p_{n+1}. \end{aligned}$$

Put  $\mathbf{LP}_n = \mathbf{IPC} + \pi_n, \mathbf{LP}_\omega = \mathbf{IPC}$ .

Ono [1972] and Smoryński [1973] showed that  $\mathbf{BH}_n = \mathbf{LP}_n$ . The notion of *nth slice* was introduced by [Hosoi, 1967] to capture logic of exact height  $n$ :  $\mathfrak{S}_n$  is the class of logics that are complete for models of height  $n$ , but not for models of smaller height.

## 6.5 Cardinality Conditions

Consider the statement

$$C_n := \bigvee_{0 \leq i < j \leq n} p_i \leftrightarrow p_j;$$

if  $C_n$  holds in a model with bottom node then two atoms must be forced on exactly the same set of nodes. So if we have  $n$  nodes, then there are  $2^n$  subsets and hence  $C_{2^n}$  holds in all models of  $\leq n$  elements. This bound is in general too crude. Let us therefore specialise the class of models to linearly ordered frames.

Put  $\mathbf{S}_n = \mathbf{LC} + C_{n+1}$  (as a schema).

**THEOREM 36.**  *$\mathbf{S}_n$  is complete for all linear models with  $n$  nodes.*

**Proof.** If the model has  $n$  nodes then by a simple inspection one sees that  $C_{n+1}$  holds. Conversely, if a model of  $\mathcal{K}$  of  $C_{n+1}$ , obtained by filtration, has at least  $k + 1$  nodes  $\alpha_0 < \alpha_1 < \dots < \alpha_n$ , then by filtration we can find  $\varphi_1, \dots, \varphi_n$  such that  $\alpha_i \not\Vdash \varphi_{i+1}$  and  $\alpha_{i+1} \Vdash \varphi_{i+1}$ . Putting  $\varphi_0 := p \rightarrow p$  and  $\varphi_{n+1} := p \wedge \neg p$ , we obtain an instance of  $C_{n+1}$  that is not forced by  $\alpha_0$ . This shows that  $\mathbf{S}_n$  is complete for models with linear poset of length  $\leq n$ . But, since we can always add some nodes for free, it is also complete for all linear models with exactly  $n$  nodes.

Converting the linear, finite Kripke frames into truth tables one obtains Gödel's many-valued logic, used to establish the fact that  $\mathbf{IPC}$  is not a (finite) many-valued logic [1932]. ■

### 6.6 Some More Intermediate Logics

A number of intermediate logics have found their way into the literature. We will mention some of them, with their main properties.

$\mathbf{KC}$  is axiomatised by  $\neg\varphi \vee \neg\neg\varphi$ .

1.  $\mathbf{KC}$  is strongly complete for the class of directed Kripke frames. (A poset is direct if it satisfies  $\forall\alpha\beta\exists\gamma(\alpha \leq \gamma \wedge \beta \leq \gamma)$ .)
2.  $\mathbf{KC}$  is strongly complete for the class of Kripke frames with a maximum.

For proofs see [Gabbay, 1981, p. 66 ff.], [Smoryński, 1973].  $\mathbf{KC}$  can alternatively be axiomatised by  $(\neg\varphi \rightarrow \neg\psi) \vee (\neg\psi \rightarrow \neg\varphi)$ . This shows that  $\mathbf{LC}$  is an extension of  $\mathbf{KC}$ .

The Kreisel–Putnam system  $\mathbf{KP}$  is axiomatised by  $(\neg\sigma \rightarrow \varphi \vee \psi) \rightarrow [(\neg\sigma \rightarrow \varphi) \vee (\neg\sigma \rightarrow \psi)]$  [Kreisel and Putnam, 1957].

$\mathbf{KP}$  has the DP and FMP, and Gabbay has shown it to be complete for the class of Kripke models satisfying the condition # below [Gabbay, 1970].

For a poset  $\langle K, \leq \rangle$  with bottom element we define for a subset  $E$  of  $K$ :

$$E^+ = \{p \mid \exists q \in E (q \leq p)\}; E^- = \{p \mid \exists q \in E (p \leq q)\}.$$

# For every  $E \subseteq P$  the set  $P - (E^+)^-$  is either empty or has a first element.

For a proof see [Gabbay, 1981, p. 96 ff.].

### *Notions of width*

The width of a frame can be conceived in various ways. One can look for the length of maximal anti-chains, or the maximal number of successors of individual nodes (say in finite frames). We will consider some notions below.

(a) *Anti-chain width.* A Kripke frame  $\mathcal{K}$  has *a.c. width*  $n$  if it has an anti-chain length  $n$ , but no anti-chain of length  $n + 1$ . Define

$$\varphi_n = \bigvee_{i=0}^n \left( p_i \rightarrow \bigvee_{j \neq i} p_j \right) \text{ and } \mathbf{BA}_n = \mathbf{IPC} + \varphi_n \text{ (as a schema).}$$

**THEOREM.** [Smoryński, 1973]  $\mathbf{BA}_n$  is strongly complete for the class of frames of a.c. width at most  $n$ .

The proof is routine, use filtration.

**COROLLARY.**  $\mathbf{BA}_n$  has the FMP and is decidable.

(b) *Top-width.* In finite frames one can just count the number of top nodes, this gives a maximum width for trees, but not for posets in general.

Define the *top-width* of a frame as the number of maximal nodes and put

$$\begin{aligned} \delta_n &:= \bigvee_{i=0}^n \left( \neg p_i \rightarrow \bigvee_{j \neq i} \neg p_j \right), \\ \psi_n &:= \left( \bigwedge_{0 \leq i < j \leq n} \neg(\neg p_i \wedge \neg p_j) \right) \rightarrow \delta_n \end{aligned}$$

and  $\mathbf{BTW}_n := \mathbf{IPC} + \psi_n$  (as a schema).

**THEOREM.** [Smoryński, 1973]  $\mathbf{BTW}_n$  is complete for the class of all frames of top-width at most  $n$ .

**COROLLARY.**  $\mathbf{BTW}_n$  has the FMP and is decidable.

(c) *Local width* [Gabbay and de Jongh, 1974]. We consider finite trees and define the *local width* of a tree frame as the maximum number of successors of its nodes. Define

$$\alpha_n := \left[ \bigwedge_{i=0}^{n+1} \left( \left( p_i \rightarrow \bigvee_{i \neq j} p_j \right) \rightarrow \bigvee_{i \neq j} p_j \right) \right] \rightarrow \bigvee_{i=0}^{n+1} p_i$$

and  $\mathbf{LW}_n = \mathbf{IPC} + \alpha_n$  (as a schema).

**THEOREM.**  $\mathbf{LW}_n$  is complete for tree frames of local width at most  $n$ .

**COROLLARY.**  $\mathbf{LW}_n$  has the FMP and is decidable.

Furthermore one can show that

1.  $\mathbf{LW}_n$  has the DP (use the gluing trick)
2.  $\bigcap \mathbf{LW}_n = \mathbf{IPC}$
3.  $\mathbf{LW}_n \neq \mathbf{LW}_{n+1}$ .

For proofs see [Gabbay, 1981, p. 83 ff.].

### 6.7 The Lattice of Intermediate Logics

Intermediate logics constitute a poset under the natural order of inclusion. Let us agree to consider intermediate logics as being given by axiom schemata. Observe that any such consistent extension of  $\mathbf{IPC}$  is a subsystem of  $\mathbf{CPC}$ . Hence we can safely form the meet and join of intermediate logics as follows: let  $T_i$  be axiomatised by the schemas  $\varphi_j^i (j \geq 0)$ , then  $T_1 \sqcap T_2 (T_1 \sqcup T_2)$  are axiomatised by  $\varphi_1^j \vee \varphi_2^k (\varphi_1^j \wedge \varphi_2^k)$ . It immediately follows that the intermediate logics constitute a distributive lattice.

This lattice has extensively been investigated. We have already met some properties: e.g. there is a descending sequence of logics with intersection  $\mathbf{IPC}$  (Section 5.6.3 (c)).

Further properties are:

There are  $2^{\aleph_0}$  many intermediate logics [Jankov, 1968].

There are intermediate logics, that are not finitely axiomatisable.

There exists a sequence of formulas  $\varphi_i$  such that the logics  $T_A$  axiomatised by  $\{\varphi_i \mid i \in A\}$  for  $A \subseteq \omega$ , satisfy  $T_A = T_B \Leftrightarrow A = B$ . Such a string is called strongly independent (cf. [Gabbay, 1981, p. 73 ff.]).

There are intermediate logics without the FMP (cf. [Gabbay, 1981, p. 103 ff]).

There exists a strictly increasing chain of intermediate logics [Jankov, 1968; Fine, 1970].

There are exactly eight intermediate logics with the interpolation theorem [Maximova, 1977]. For more information cf. [Rautenberg, 1979, p. 288 ff].

### 6.8 Extensions of **IQC**

The study of intermediate predicate logics has not yet made advances comparable to those in propositional logic. We refer the reader to, e.g. [Ono, 1973].

The best known extension of **IQC** is the *logic of constant domains* introduced by [Grzegorzcyk, 1964], and axiomatised and shown complete by [Görnemann, 1971].

Put  $\mathbf{CD} = \mathbf{IQC} + \forall x(\varphi \vee \psi(x)) \rightarrow (\varphi \vee \forall x\psi(x))$ .

**THEOREM.** [Görnemann] **CD** is complete for the class of Kripke models with constant domain.

**Proof.** See [Gabbay, 1981, p. 50 ff]. ■

**CD** has somewhat unpleasant features as it is not closed under relativisation, i.e. if  $\tau(x)$  is some suitable formula then we may have  $\mathbf{CD} \vdash \sigma$  but  $\mathbf{CD} \not\vdash \sigma^{\tau(x)}$ , where  $\sigma^{\tau(x)}$  is the sentence obtained by relativizing all quantifiers. The reason being that although the domain is fixed, predicates need not be constant in Kripke models for **CD**.

The difference between **IQC** and **CD** disappears when we restrict ourselves to formula without  $\forall$  [Fitting, 1969] or without  $\forall$  and  $\exists$  [Gabbay, 1981].

Another noteworthy principle is the *double negation shift* DNS:

$$\forall x \neg \neg \varphi(x) \rightarrow \neg \neg \forall x \varphi(x).$$

Put  $\mathbf{MH} := \mathbf{IQC} + \mathbf{DNS}$ .

**MH** turns out to be complete for Kripke models with the property that each node is below some maximal node [Gabbay, 1981, p. 57 ff]. Keeping the proof of Glivenko's theorem in mind, it is not surprising that it holds for **MH**.

Actually **MH** is the smallest such extension of **IQC** [Gabbay, 1981, p. 14].

To finish this section, let us mention a rather different enterprise. Ono and Komori [1985] studied intuitionistic propositional calculus in the Gentzen sequent formalisation, without the contraction rule. They generalise Kripke models to monoid-Kripke models and establish the completeness theorem. Ono [1985] extends this study to predicate calculus.

The study of logics without structural rules is the subject of Girard's *linear logic*, cf. [Girard *et al.*, 1989].

## 7 FIRST-ORDER THEORIES

A number of basic notions of intuitionistic mathematics can faithfully be studied in the framework of intuitionistic first-order logic. Although the



situation is similar to that in classical logic, there is one disturbing aspect peculiar to intuitionistic logic (or rather its semantics): the absence of natural (or standard) models. Let us compare the state of affairs with classical logic. For theories in **CQC** we have not only the traditional notion of model, as presented by Tarski (cf. [van Dalen, 1997]), but also the notion of Boolean-valued model [Rasiowa and Sikorski, 1963]. The reader who is not familiar with the theory of Boolean valued models, may think of a topological model over a discrete space, i.e. with  $\mathcal{O}(X) = \mathcal{P}(X)$ . Or he may think of Heyting algebras with the extra condition  $\neg\neg x = x$  for all  $x$  (or  $\neg x \sqcup x = T$ ).

The truth values  $\llbracket \varphi \rrbracket$  are simply elements of a Boolean algebra  $B$ . There is among the Boolean algebras a canonical one that is contained in all Boolean algebras, the two-element algebra  $2 = \{0, 1\}$ , with operations given by the traditional truth tables. Now there is for each Boolean-valued model a (truth preserving homomorphism onto a Boolean-valued model over 2, i.e. an ordinary model. Hence truth in all Boolean-valued models is equivalent to truth in all ordinary models. So the notion of truth according to ordinary model theory coincides with that of Boolean-valued model theory (cf. [Rasiowa and Sikorski, 1963, p. 295]). The ordinary models can thus be considered as the real (or standard) models among the Boolean-valued ones. This relation does not exist for intuitionistic semantics (say Heyting-valued, to take the most general one). For although 2 is contained in (and can be obtained as homomorphic image of) all Heyting-algebras, truth in all Heyting value model is certainly not the same as truth in all 2-valued (i.e. classical) models.

The fact that for intuitionistic first-order theories there does not exist a canonical model notion in the various semantics that we have exhibited is one that we have to accept, unpleasant as it may be.

Philosophically speaking there are two ways to open to use— (1) look for a codification of the Brouwer–Heyting–Kreisel notion of proof- interpretation, (2) give up the notion of ‘standard’ truth, or intended model.

We will discuss the problem of semantics later, but not without pointing out that the absence of a standard model for arithmetic in any of the semantics introduced earlier, is rather embarrassing (see below).

We will discuss a number of basic first-order theories below. The most fundamental is the theory of equality.

### 7.1 *The Theory of Equality* **EQ**

The axioms for **EQ** are the familiar ones:

the universal closures of

1.  $x = x$
2.  $x = y \rightarrow y = x$

$$3. x = y \wedge y = z \rightarrow x = z$$

(whenever we list axioms we will tacitly presuppose the clause ‘the universal closure of’).

As in classical logic one shows by induction on  $\varphi$   $x = y \rightarrow (\varphi(x) \rightarrow \varphi(y))$ .

Let us consider a Kripke model for **EQ**. We will denote the binary relation that interprets  $=$  in  $\alpha$  by  $\approx_\alpha$ . One easily sees that  $\approx_\alpha$  is an equivalence relation in (the domain of) each node. In general, however,  $\approx_\alpha$  is not the identity. For, suppose that  $\approx_\alpha$  were the identity in each node, then if for  $a, b \in \alpha$   $a \not\approx_\alpha b$ , we see that for all  $\beta \geq \alpha$ ,  $a \not\approx_\beta b$  in  $\beta$ . Hence  $\alpha \Vdash a = b \vee a \neq b$ . Conversely, if  $\mathcal{K} \Vdash \forall xy(x = y \vee x \neq y)$ , then we can construct from  $\mathcal{K}$  a Kripke model  $\mathcal{K}'$  with  $\approx$  the identity in each node.

For, since  $\approx_\alpha$  is an equivalence relation we can form equivalence classes  $[a]_\alpha$  for each  $a$  in  $\alpha$ .

Define  $\alpha \Vdash' a = b := [a]_\alpha = [b]_\alpha (\Leftrightarrow a \approx_\alpha b)$ .

Claim:  $\alpha \Vdash \varphi \Leftrightarrow \alpha \Vdash' \varphi$  for all  $\alpha \in K$ .

**Proof.** Induction on  $\varphi$ . The definition of  $\Vdash'$  takes care of the atomic  $\varphi$ .  $\wedge$  and  $\vee$  are immediate,  $\alpha \not\Vdash \varphi \rightarrow \psi \Leftrightarrow \exists \beta \geq \alpha, \beta \Vdash \varphi$  and  $\beta \not\Vdash \psi \Leftrightarrow$  (induction hypothesis)  $\exists \beta \geq \alpha, \beta \Vdash' \varphi$  and  $\beta \not\Vdash' \psi \Leftrightarrow \alpha \not\Vdash' \varphi \rightarrow \psi$ .

$\alpha \Vdash \forall x\varphi(x) \Leftrightarrow \forall a \in \alpha, \alpha \Vdash \varphi(\bar{a}) \Leftrightarrow \forall [a] \in \alpha, \alpha \Vdash' \varphi(\bar{a}) \Leftrightarrow \alpha \Vdash' \forall x\varphi(x)$  (where  $\bar{a}$  is the name of  $a$  in  $\mathcal{K}$  and of  $[a]$  in  $\mathcal{K}'$ ). A similar argument handles  $\exists$ . A slight boost of the argument yields the same result for arbitrary languages.

Summing up: **IQC** with decidable equality is complete for normal Kripke models (i.e. with  $=$  interpreted by real equality).

The theory **EQ** is of interest since the basic theories of real life depend heavily on equality.

In general  $T^c$  will denote the classical theory  $T + \varphi \vee \neg\varphi$ ,  $T^d$  will denote the theory  $T +$  decidability for atoms. We will keep superscripts for ‘logical’ and prefixes for ‘mathematical’ variants of theories.

The following facts have been proved:

$$\mathbf{EQ}^d = \mathbf{EQ}^c$$

decidable

$\mathbf{EQ}^d$	+	[Lifschitz, 1969]
$\mathbf{EQ}^s$	-	[Lifschitz, 1969]
$\mathbf{EQ}$	-	[Lifschitz, 1969],

where  $\mathbf{EQ}^s$  is the theory of stable inequality:  $\forall xy(\neg\neg x = y \rightarrow x = y)$ . ■

## 7.2 The Theory of Apartness, **AP**

For practical purposes one needs in intuitionistic mathematics a strong inequality relation. For example, in the theory of the reals one needs a prop-

erty like ‘ $x$  has a positive distance to 0 ( $\exists k(|x| > 2^{-k})$ ), to make sure that  $x$  has an inverse. A mere inequality would not do.

The positive inequality relation was introduced by Brouwer in 1918 and axiomatized by Heyting.

*Notation  $x\#y$  read  $x$  is apart from  $y$ .*

**AP** has the axioms of **EQ** plus the following ones:

$$\begin{aligned} \neg x\#y &\leftrightarrow x = y \\ x\#y &\rightarrow x\#z \vee y\#z. \end{aligned}$$

One easily derives the following:

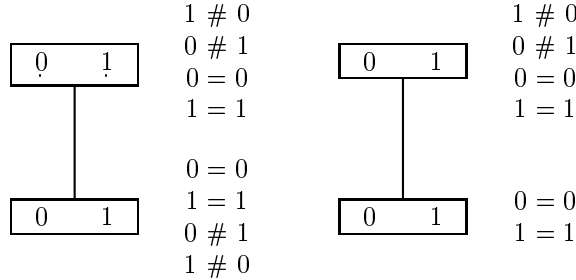
**FACT 37.** The following are derivable in **AP**.

$$\begin{aligned} x\#y &\rightarrow y\#x \\ x\#y &\rightarrow x \neq y \\ \neg\neg x = y &\rightarrow x = y. \end{aligned}$$

In particular **AP** has a stable equality. Most theories that occur in basic mathematics have an apartness relation. Combinatory logic, however, does not allow an apartness relation since its equality is not stable.

A theory with decidable equality trivially has an apartness relation, namely the inequality. One and the same structure may, however, carry more apartness relations.

**EXAMPLE**



The above models carry the same, decidable, equality, but distinct apartness relation.

The apartness relation influences the equality relation, the question is does it stop at the stability axiom or does it carry stronger conditions? The answer is provided in [van Dalen and Statman, 1979] where the axiomatisation of the equality fragment of **AP** is studied.

Consider the following sequence of inequalities  $\neq_n$

$$\begin{aligned} x \neq_0 y &:= \neg x = y \\ x \neq_{n+1} y &:= \forall z(x \neq_n z \vee y \neq_n z). \end{aligned}$$

For each  $n$  we formulate a stability axiom

$$S_n := \neg x \neq_n y \rightarrow x = y.$$

FACT.

$$\begin{array}{ll} \mathbf{EQ} \vdash S_n \rightarrow S_m & \text{for } n > m \\ \mathbf{AP} \vdash S_n & \text{for all } n \\ \mathbf{AP} \vdash x \# y \rightarrow x \neq_n y & \text{for all } n. \end{array}$$

Consider the  $\omega$ -stable theory of equality  $\mathbf{EQ}_\omega^s = \mathbf{EQ} + \{S_n \mid n \in \omega\}$

$\mathbf{EQ}_\omega^s$  turns out to be the equality fragment of  $\mathbf{AP}$ .

THEOREM.

1.  $\mathbf{AP}$  is conservative over  $\mathbf{EQ}_\omega^s$ .
2.  $\mathbf{EQ}_\omega^s$  is not finitely axiomatisable.

Van Dalen and Statman proved the theorem by means of a normal form theorem for  $\mathbf{AP}$ . There is however a short and elegant proof by Smoryński using model theory [Smoryński, 1977], that we will reproduce here.

**Proof of 1.** Suppose  $\mathbf{EQ}_\omega^s \not\vdash \varphi$ . Consider the Kripke model  $\mathcal{K}$  obtained in the model existence lemma.

Define  $\Gamma \Vdash a \# b := \Gamma \Vdash a \neq_n b$  for all  $n$ .

Claim  $\#$  is an apartness relation. We will only consider  $\forall xy(\neg x \# y \rightarrow x = y)$  (everything else is trivial).

Suppose the bottom node,  $\Gamma_0$ , does not force it, then  $\Gamma_1 \Vdash \neg a \# b$  and  $\Gamma_1 \not\vdash a = b$  for some  $\Gamma_1$ . Since  $\mathcal{K}$  is a model of  $\mathbf{EQ}_\omega^s$  we have  $\Gamma_1 \not\vdash \neg a \neq_n b$  for all  $n$ .

Now  $\Gamma_1 \cup \{a \neq_n b \mid n \in \omega\}$  is consistent, for else  $\Gamma_1 \cup \{a \neq_n b \mid n \in \omega\} \vdash \perp$  and hence  $\Gamma_1 \vdash \neg a \neq_m b$ , i.e.  $\Gamma_1 \Vdash \neg a \neq_m b$ , for some  $m$ . Therefore there exists a prime theory  $\Delta \supseteq \Gamma_1$  with  $a \neq_n b \in \Delta$  for all  $n$ , so  $\Delta \Vdash a \# b$ . Contradiction.

Hence  $\Gamma_0 \Vdash \forall xy(\neg x \# y \rightarrow x = y)$ .

2. is shown by constructing suitable Kripke models ■

As a corollary we obtain the undefinability of  $\#$  in terms of  $=$ . For, if  $\mathbf{AP} \vdash x \# y \leftrightarrow \varphi(x, y)$  for a suitable equality formula  $\varphi$ , then we would have a finite axiomatisation of  $\mathbf{EQ}_\omega^s$  (note that the above example also establishes the same fact).

Observe that we could accept the apartness relation as basic and define equality by  $x = y := \neg x \# y$  if we replace  $\neg x \# y \leftrightarrow x = y$  by  $\neg x \# x$  and  $x \# y \rightarrow y \# x$ .

*Further facts:* **AP** is undecidable [Smoryński, 1973a], [Gabbay, 1981, p. 258]. There are Kripke models of  $\mathbf{EQ}_\omega^s$  that do not carry any apartness relation at all [van Dalen and Statman, 1979]. For a treatment of apartness in a sequent calculus setting, see [Negri and von Plato, 2001].

### 7.3 The Theory of Order, **LO**

In classical logic linear order is singled out from the partial orders by requiring any two elements to be comparable, i.e.  $x < y \vee x = y \vee y < x$ . This axiom would be excessively strong in an intuitionistic context, since not even the reals would be ordered. Therefore Heyting proposed another axiom, that we shall adopt.

The language of **LO** contains the predicate symbols  $<$  and  $=$ . The axioms of **LO** are those of **EQ**, plus

$$\begin{aligned} x < y \wedge y < z &\rightarrow x < z \\ x = y &\leftrightarrow \neg x < y \wedge \neg y < x \\ x < y &\rightarrow z < y \vee x < z. \end{aligned}$$

It is a simple exercise in logic to show the following:

*Fact*

1.  $\neg x < x$
2.  $x = y \wedge x < z \rightarrow y < z$
3.  $x < y \vee y < x$  is an apartness relation.

Heyting called a relation satisfying the axioms of **LO** a *pseudo-order* relation and a relation satisfying, moreover,  $x < y \vee x = y \vee y < x$  an order relation. Since, however, the first kind of relation turns out to be the more important and the more common of the two, we have adopted the present terminology.

Since

$$\mathbf{LO} + \forall xy(x < y \vee x = y \vee y < x) \vdash (x < y \vee \neg x < y) \wedge (x = y \vee \neg x = y)$$

we call this system *decidable linear order* **LO<sup>d</sup>**.

Conversely, the decidability of  $<$  and  $=$  implies the comparability of  $x$  and  $y$ .

The theory of *dense linear ordering*, **DLO**, is obtained by adding

$$\exists z(x < y \rightarrow x < z < y), \exists y(x < y), \exists y(y < x).$$

Variants of **DLO** are considered by Smoryński [1977].

Not only can we define a canonical apartness relation in **LO**, but also show that the theory containing apartness and order (**AP** + **LO** +  $x\#y \rightarrow x < y \vee y < x$ ) is conservative over **AP** (and hence over **EQ<sub>w</sub><sup>s</sup>**) [Smoryński, 1977; Hartog, 1978].

The theories **LO** and **DLO** are undecidable [Smoryński, 1977] cf. [Gabbay, 1981], but **DLO<sup>d</sup>** is decidable and coincides with its classical counterpart **DLO<sup>c</sup>**, [Smoryński, 1973a].

In [Gabbay, 1981] a number of refinements of the above results are treated.

#### 7.4 Logic with Operations

The set theoretical view of operations (functions) is that they are a special kind of relations, so we could do without the complications of introducing function symbols. However, the circumvention of function symbols is most unnatural, and, when we come to choice sequences, disastrous. The syntactic aspects of a first-order language with function symbols are strictly analogous to the classical ones.

We will therefore look into the semantic aspects. Consider a Kripke model  $\mathcal{K}$  with a functional relation  $R(x, y)$ , i.e.  $\mathcal{K} \Vdash \forall x \exists ! y R(x, y)$ . The properties of forcing tell us that for each  $\alpha$  we have that for each  $a \in D(\alpha)$  there is a unique  $b \in D(\alpha)$  such that  $\alpha \Vdash R(a, b)$ . That is  $R$  is a function on  $D(\alpha)$ . hence we define for each function symbol  $f$  an  $n$ -ary function  $f_\alpha : D(\alpha)^n \rightarrow D(\alpha)$ , for each  $\alpha$ .

The monotonicity condition is not quite obvious since elements of  $|D(\alpha)|$  are determined up to the relation  $\approx_\alpha$ .

The simplest solution is to put:  $\alpha \leq \beta \Rightarrow f_\alpha \subseteq f_\beta$ . There is another solution, however, that modifies the concept of Kripke model in the spirit of category theory, where one defines a Kripke model as a pre-sheaf over  $P$ , where  $P$  is a poset (cf. [Goldblatt, 1979, p. 256]).

Before formulating the modified notion of Kripke model, we recall the notion of *homomorphism* for (classical) structures.

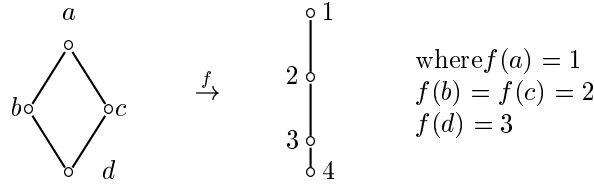
$f : A \rightarrow B$  is a homomorphism from structure **A** to **B** if  $f$  is a function from the universe of **A** into the universe of **B** such that  $f$  preserves all relations and functions, i.e.,

$$R^{\mathbf{A}}(a_1, \dots, a_n) \Rightarrow R^{\mathbf{B}}(f(a_1), \dots, f(a_n)) \text{ for all relations } R,$$

and

$$f(F^{\mathbf{A}}(a_1, \dots, a_n)) = F^{\mathbf{B}}(f(a_1), \dots, f(a_n)) \text{ for all functions } F.$$

EXAMPLE



$f$  is a homomorphism of the poset **A** into the poset **B**. We now come to our

**DEFINITION 38.** A Kripke model is a quintuple  $\mathcal{K} = \langle K, \leq, d, f, \Vdash \rangle$  where  $(K, \leq)$  is a poset,  $D$  assigns to each  $\alpha \in K$  a structure  $D(\alpha)$ ,  $f$  assigns to each pair  $\alpha, \beta$  with  $\alpha \leq \beta$  a homomorphism  $f_{\alpha\beta} : D(\alpha) \rightarrow D(\beta)$  such that  $f_{\alpha\alpha} = id_{D(\alpha)}$ , for all  $\alpha, f_{\beta\gamma} \circ f_{\alpha\beta} = f_{\alpha\gamma}$  for all  $\alpha \leq \beta \leq \gamma$ . The forcing relation  $\Vdash$  is defined as in Definition 13. Furthermore, equality is always interpreted as real identity.

We can always associate a model in the new sense to a model in the old sense by lumping together elements in equivalence classes under  $\approx_\alpha$ .

In dealing with concrete Kripke models we will act broad-mindedly and choose whichever notion is most convenient, or even use the old notion and think of a new one.

### 7.5 Heyting's Arithmetic, **HA**

The language of arithmetic contains,  $=, +, \cdot, S, 0, 1$  (and, when convenient, as many primitive recursive functions as we wish).

The axioms of **HA** are those of Peano's arithmetic plus the axioms of **EQ**.

1.  $x = y \leftrightarrow Sx = Sy$
2.  $\neg Sx = 0$
3.  $x + 0 = x$   
 $x + Sy = S(x + y)$
4.  $x \cdot 0 = 0$   
 $x \cdot Sy = x \cdot y + x$
5.  $\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(Sx)) \rightarrow \forall x\varphi(x)$ .

Number 5 is the schema of *mathematical* (or *complete*) *induction*. It can

also be presented in the form of a natural deduction rule

$$\frac{\begin{array}{c} [\varphi(x)] \\ \vdots \\ \varphi(0) \quad \varphi(Sx) \end{array}}{\varphi(x)}$$

It is a simple exercise to show the decidability of the equality relation.

**THEOREM 39.**

$$\begin{array}{l} \mathbf{HA} \vdash \forall xy(x = y \vee x \neq y) \\ \mathbf{HA} \vdash \forall x(x \neq 0 \Rightarrow \exists y(x = Sy)). \end{array}$$

For a number of formal proofs, see [Kleene, 1952].

Arithmetic has correctly received a considerable amount of attention. It is the theory of the hard core of intuitionistic mathematics, put forward by Brouwer in his *First Act of Intuitionism* (cf. [Brouwer, 1975, pp. 509], [Brouwer, 1981, p. 4], [Heyting, 1956, p. 13 ff]).

Apart from foundational motivations for studying **HA**, there is a pragmatic argument for investigating arithmetic. It is, so to speak, a showroom of metamathematical tools and results. We will only be able to discuss a minute part of the material that is available. The reader is referred to Troelstra [1973].

Since **HA** is a subsystem of **PA** (Peano's arithmetic) we cannot expect to find theorems contradicting the classical practice. We will have to look for metamathematical methods that capitalise on the constructive nature of intuitionistic logic.

**HA** has the properties EP and DP, that are popularly considered to be the hallmark of constructive theories. We will first show EP and return to the significance of EP and DP later.

**THEOREM 40.**

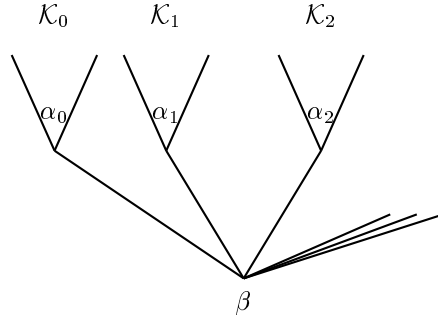
$$\begin{array}{ll} DP: & \mathbf{HA} \vdash \varphi \vee \psi \Rightarrow \mathbf{HA} \vdash \varphi \text{ or } \mathbf{HA} \vdash \psi \quad (\textit{disjunction property}) \\ EP: & \mathbf{HA} \vdash \exists x\varphi(x) \Rightarrow \mathbf{HA} \vdash \varphi(n) \text{ for some } n. \quad (\textit{existence property}) \end{array}$$

**Proof.** DP follows immediately from EP, since disjunction can be defined in terms of the existential quantifier:  $\mathbf{HA} \vdash (\varphi \vee \psi) \leftrightarrow \exists x((x = 0 \rightarrow \varphi) \wedge (x \neq 0 \rightarrow \psi))$ . therefore we will consider EP.

Let  $\mathbf{HA} \vdash \exists x\varphi(x)$ , but  $\mathbf{HA} \not\vdash \varphi(n)$  for all  $n$ .

Then, by the completeness theory, there are Kripke models  $\mathcal{K}_0, \mathcal{K}_1, \mathcal{K}_2, \dots$  such that  $\mathcal{K}_n \not\vdash \varphi(n)$ . We form a new Kripke model  $\mathcal{K}$ .





We take the disjoint union of the models  $\mathcal{K}_i$  and add an extra bottom node  $\beta$ .

The structure belonging to  $\beta$  is the standard model of (classical) arithmetic. Since  $\mathbb{N}$  is contained in all domains of the  $\mathcal{K}_i$ 's, the resulting model satisfies the conditions of the Kripke semantics. We will show that  $\mathcal{K} \Vdash \mathbf{HA}$ . The only non-trivial axiom is mathematical induction. So we must show  $\gamma \Vdash \psi(0), \gamma \Vdash \forall x(\psi(x) \rightarrow \psi(Sx)) \Rightarrow \gamma \Vdash \forall x\psi(x)$ , for all  $\gamma$ . For  $\gamma \neq \beta$  this is so by hypothesis, so consider  $\beta \Vdash \psi(0), \beta \Vdash \forall x(\psi(x) \rightarrow \psi(Sx))$ . We must show  $\forall \gamma \geq \beta \forall c \in D(\gamma), \gamma \Vdash \psi(c)$ . Again the only case that must be taken care of is  $\beta$  itself. So we must show  $\beta \Vdash \psi(n)$  for all  $n$ . but we know  $\beta \Vdash \psi(0)$  and  $\beta \Vdash \psi(n) \rightarrow \beta \Vdash \pi(Sn)$ . Hence, by induction in the metalanguage we get  $\beta \Vdash \psi(n)$  for all  $n$ .

Now  $\beta \Vdash \mathbf{HA}$ , so  $\beta \Vdash \exists x\varphi(x)$ , and hence  $\beta \Vdash \varphi(n)$  for some  $n$ . Contradiction with  $\mathcal{K}_n \not\Vdash \varphi(n)$ .

Therefore  $\mathbf{HA} \vdash \varphi(n)$  for some  $n$ . ■

Although EP seems to be stronger than DP, a result of Friedman shows this not to be the case for a large class of extensions of  $\mathbf{HA}$ .

**THEOREM 41** ([Friedman, 1975]). *For all RE extensions of  $\mathbf{HA}$  EP follows from DP.*

It seems attractive to consider EP as the characteristic of constructivity; if we can show the existence of an object with a property  $\varphi$ , then we can effectively indicate such an object. This is the constructive counterpart of the classical notion of 'pure existence'.

Kreisel has shown, however, that the possession of EP is neither necessary nor sufficient for constructive theories. The following example (due to Kreisel, cf. [Troelstra, 1973, p. 91]) may illustrate the matter.

Let  $\text{Prf}$  be the proof predicate of  $\mathbf{HA}$  (cf. [Kleene, 1952, p. 254]). Define  $\varphi(x) := \text{Prf}(x, \ulcorner 0 = 1 \urcorner) \vee \forall y \neg \text{Prf}(y, \ulcorner 0 = 1 \urcorner)$ . As  $\mathbf{HA}$  is consistent on the intended interpretation,  $\forall y \neg \text{Prf}(y, \ulcorner 0 = 1 \urcorner)$  is true, so evidently  $\exists x\varphi(x)$  is true. Moreover,  $\neg \text{Prf}(n, \ulcorner 0 = 1 \urcorner)$  is true for each  $n$ , and ( $\text{Prf}$  being primitive recursive) provable. Hence, for any  $n$   $\mathbf{HA} \vdash \varphi(n) \leftrightarrow \forall y \neg \text{Prf}(y, \ulcorner 0 =$

$1^\top$ ). By definition  $\mathbf{HA} \vdash \exists x\varphi(x) \leftrightarrow [\exists y\text{Prf}(y, \ulcorner 0 = 1^\top \urcorner) \vee \forall y\neg\text{Prf}(y, \ulcorner 0 = 1^\top \urcorner)]$ . Now  $\mathbf{HA} \vdash \exists x\varphi(x) \rightarrow \varphi(n)$  yields  $\mathbf{HA} \vdash [\exists y\text{Prf}(y, \ulcorner 0 = 1^\top \urcorner) \vee \forall y\neg\text{Prf}(y, \ulcorner 0 = 1^\top \urcorner)] \rightarrow \forall y\neg\text{Prf}(y, \ulcorner 0 = 1^\top \urcorner)$ , hence  $\mathbf{HA} \vdash \exists y\text{Prf}(y, \ulcorner 0 = 1^\top \urcorner) \rightarrow \forall y\neg\text{Prf}(y, \ulcorner 0 = 1^\top \urcorner)$ , and so  $\mathbf{HA} \vdash \forall y\neg\text{Prf}(y, \ulcorner 0 = 1^\top \urcorner)$ , or  $\mathbf{HA}$  proves its own consistency, contradicting Gödel's second theorem.

Note that  $T = \mathbf{HA} + \exists x\varphi(x)$  is an intuitionistically true theory and  $T \vdash \exists\varphi(x)$ , but by the above argument  $\not\vdash \varphi(n)$  for all  $n$ . So  $T$  does not have EP.

The proof of EP above was itself not constructive, we have used a proof by contradiction. So we cannot actually exhibit the promised number  $n$ .

There are various proofs that do provide the required instances. For example, by means of the Kleene slash [Troelstra, 1973, p. 177],  $q$ -realisability [Troelstra, 1973, p. 189 ff], and normalisation in Gentzen systems [Minc, 1974]. An interesting feature is the stability of the instantiation member in various methods. That is, quite different techniques for converting a provable  $\exists$ -statement into its instantiation yield the same number (cf. [Stein, 1980]).

It is to be noted that all proofs of DP (or EP) for  $\mathbf{HA}$  go essentially beyond the means of  $\mathbf{HA}$ . Actually one can make this precise (Myhill) in the following form: let  $T$  be an r.e. extension of  $\mathbf{HA}$  then there are sentences  $\varphi$  and  $\psi$  such that if  $T \vdash \text{Pr}(\varphi \vee \psi) \rightarrow \text{Pr}(\varphi) \vee \text{Pr}(\psi)$ , then  $T \vdash \text{Pr}(\ulcorner 0 = 1^\top \urcorner)$  where  $\text{Pr}(x)$  is the provability predicate for  $T$ .

In words, the price for 'provable DP' is that  $T$  proves its own inconsistency (cf. [Leivant, 1985]).

### Closure under rules

For a given derivation rule  $\frac{\varphi_1, \dots, \varphi_n}{\psi} R$  we automatically get that provability of the premises yields provability of the conclusion. We say that a theory is closed under a rule  $R$  if  $T \vdash \varphi_1, \dots, T \vdash \varphi_n \Rightarrow T \vdash \psi$ . Intuitionistic systems tend to be closed under various rules that are themselves not correct. We will list a few cases below. Consider the following principle.

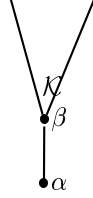
#### Markov's Principle, MP

$$\forall x(\varphi(x) \vee \neg\varphi(x)) \wedge \neg\neg\exists x\varphi(x) \rightarrow \exists x\varphi(x).$$

MP plays an important role in metamathematics. It naturally turns up in certain interpretations. Markov postulated it in the context of recursion theory in the form 'if it is impossible that a Turing machine does not halt, then it must halt', in the formalism of recursion theory:  $\neg\neg\exists zT(e, n, z) \rightarrow \exists zT(e, n, z)$  (cf. [Troelstra, 1973]). Thus Markov's formulation can be taken to deal with a primitive recursive  $\varphi(x)$ .

**THEOREM 42.** *MP is not derivable in  $\mathbf{HA}$ .*

**Proof.**[Smoryński] Let  $\varphi(x)$  be a primitive recursive formula such that  $\exists x\varphi(x)$  is independent of **HA** (e.g.  $\neg\text{Con}_{\mathbf{HA}}$ , the inconsistency of **HA**). Let  $\mathcal{K}$  be a Kripke model of **HA** +  $\exists x\varphi(x)$ . We put an extra node  $a$  at the bottom of  $\mathcal{K}$ , with  $D(\alpha)$  the standard model for **PA**.



Suppose **HA**  $\vdash \neg\neg\exists x\varphi(x) \rightarrow \exists x\varphi(x)$ . (\*)

Since the new model is a model of **HA** (cf. the proof of Theorem 40), we have  $\alpha \Vdash \neg\neg\exists x\varphi(x) \rightarrow \exists x\varphi(x)$ . But evidently  $\alpha \Vdash \neg\neg\exists x\varphi(x)$ , therefore  $\alpha \Vdash \exists x\varphi(x)$  and hence  $\alpha \Vdash \varphi(n)$  for some  $n$ .  $\varphi(x)$  being primitive recursive and  $\varphi(n)$  being true, a theorem from arithmetic tells us that **HA**  $\vdash \varphi(n)$ , so **HA**  $\vdash \exists x\varphi(x)$ . This contradicts the independence of  $\exists x\varphi(x)$ . Therefore (\*) is false. ■

Next we will show that **HA** is closed under Markov's rule.

**THEOREM 43.**

$$\begin{aligned} & \mathbf{HA} \vdash \forall x(\varphi(x) \vee \neg\varphi(x)), \mathbf{HA} \vdash \neg\neg\exists x\varphi(x) \Rightarrow \\ & \mathbf{HA} \vdash \exists x\varphi(x), \text{ for } FV(\varphi) = \{x\}. \end{aligned}$$

**Proof.** From **HA**  $\vdash \neg\neg\exists x\varphi(x)$ , we conclude **PA**  $\vdash \exists x\varphi(x)$  and so  $\exists x\varphi(x)$  is true in the standard model. Therefore,  $\varphi(n)$  is true for some  $n$ . Now using **HA**  $\vdash \varphi(n) \vee \neg\varphi(n)$  and DP we immediately get **HA**  $\vdash \varphi(n)$ , and thus **HA**  $\vdash \exists x\varphi(x)$ . ■

Our next principle is the

*Independence of Premise Principle, IP*

$$(\neg\varphi \rightarrow \exists x\psi(x)) \rightarrow \exists x(\neg\varphi \rightarrow \psi(x)).$$

The heuristic argument against IP is as follows:  $\neg\varphi \rightarrow \exists x\psi(x)$  may be seen to hold by constructing an instance  $n$  that depends on the proof of  $\neg\varphi$ . In  $\exists x(\neg\varphi \rightarrow \psi(x))$ , however, we are required to construct the instance  $n$  beforehand. This evidently is a stronger requirement. Formal independence proofs are given in [Troelstra, 1973, pp. 179, 369].

**THEOREM 44.**

$$\begin{aligned} & \mathbf{HA} \vdash \neg\varphi \rightarrow \exists x\psi(x) \Rightarrow \mathbf{HA} \vdash \exists x(\neg\varphi \rightarrow \psi(x)) \\ & (FV(\psi) = \{x\}). \end{aligned}$$

**Proof.** See below. ■

The case of Church's Thesis will be considered in Section 8.

We return to the closure under Markov's Rule, to demonstrate an extremely elegant proof by H. Friedman [1977].

First we introduce the *Friedman translation*  $\varphi \rightarrow \varphi^\rho$ : replace in  $\varphi$  each atomic subformula  $\psi$  by  $\psi \vee \rho$  (where  $\rho$  is a formula of **HA**). The translation has the following properties.

LEMMA 45.

1.  $\Gamma \vdash \varphi \Rightarrow \Gamma^\rho \vdash \varphi^\rho$  and  $\rho \vdash \varphi^\rho$ .
2. **HA**  $\vdash \varphi \rightarrow$  **HA**  $\vdash \varphi^\rho$ .
3. For any term  $t$ , **HA**  $\vdash \neg\neg\exists x(t(x, y) = 0) \Rightarrow$  **HA**  $\vdash \exists x(t(x, y) = 0)$ .

**Proof.** (1) and (2) are easily shown by a suitable induction.

(3) **HA**  $\vdash (\exists x t(x, y) = 0 \rightarrow \perp) \rightarrow \perp$ .

We apply the Friedman Translation with respect to

$$\begin{aligned} \rho := \exists x(t(x, y) = 0), \text{ then } ((\exists x t(x, y) = 0 \rightarrow \perp) \rightarrow \perp)^\rho = \\ [\exists x(t(x, y) = 0 \vee \exists x(t(x, y) = 0)) \rightarrow \perp \vee \exists x(t(x, y) = 0)] \\ \rightarrow (\perp \vee \exists x(t(x, y) = 0)). \end{aligned}$$

The latter formula is equivalent to  $\exists x(t(x, y) = 0)$ . Now apply (2). ■

So for the special case of  $t(x, y) = 0$ , closure under Markov's Rule has been established (i.e. in particular for primitive recursive functions  $f(x, y)$ ).

The general closure result is obtained by an application of closure under Church's Rule (cf. Section 8), i.e. if **HA**  $\vdash \forall x \exists y \varphi(x, y)$ , then **HA**  $\vdash \forall x \varphi(x, \{e\}x)$  for some  $e$  (index of total recursive function).

One easily derives **HA**  $\vdash \varphi(x, y) \vee \neg \varphi(x, y) \Rightarrow$  **HA**  $\vdash \varphi(x, y) \leftrightarrow \{e\}(x, y) = 0$ , for a suitable index  $e$ . We can replace  $\{e\}(x, y) = 0$  by  $\exists z(T(e, x, y, z) \wedge U(z) = 0)$  (cf. van Dalen's Algorithms chapter in Volume 1 of this *Handbook*).

The matrix of the latter expression is primitive recursive, so we may conservatively extend **HA** by adding a symbol  $f$  for its characteristic function. Hence we get **HA'**  $\vdash \varphi(x, y) \leftrightarrow \exists z(f(x, y, z) = 0)$ , where **HA'** is the extension by  $f$  and its defining equations. Now we may apply Lemma 45(3): **HA**  $\vdash \neg\neg\exists x \varphi(x, y) \Rightarrow$  **HA'**  $\vdash \neg\neg\exists x z(f(x, y, z) = 0) \rightarrow$  (Lemma 45 carries over to **HA'**) **HA'**  $\vdash \exists x z(f(x, y, z) = 0) \Rightarrow$  **HA**  $\vdash \exists x \varphi(x, y)$ .

Observe that the above argument yields closure under Markov's Rule for formulas with parameters.

We now apply the Friedman translation to the rule of independence of premises (A. Visser). For convenience we write ' $\vdash$ ' for '**HA**  $\vdash$ '.

Let  $\vdash \neg\varphi \rightarrow \exists x\psi(x)$ . We apply the Friedman translation with respect to  $\neg\neg\varphi$ . By Lemma 45(2)  $\vdash (\neg\varphi)^{\neg\neg\varphi} \rightarrow (\exists x\psi(x))^{\neg\neg\varphi} \dots (1)$ .

Observe that  $\vdash \neg\rho \rightarrow (\sigma^\rho \leftrightarrow \sigma)$  for any  $\sigma$  and  $\rho \dots (2)$ , as one can easily show by induction on  $\sigma$ .

Therefore also  $\vdash \sigma^\rho \rightarrow (\neg\rho \rightarrow \sigma) \dots (3)$ .

From(1) we get  $\vdash (\neg\varphi)^{\neg\neg\varphi} \rightarrow \exists x(\psi(x)^{\neg\neg\varphi})$ , and an application of (3) yields  $\vdash (\neg\varphi)^{\neg\neg\varphi} \rightarrow \exists x(\neg\varphi \rightarrow \psi(x))$ .  $(\neg\varphi)^{\neg\neg\varphi} = (\varphi \rightarrow \perp)^{\neg\neg\varphi} = \varphi^{\neg\neg\varphi} \rightarrow \neg\neg\varphi$ . Now apply (3) with  $\sigma = \varphi$  and  $\rho = \neg\neg\varphi$ , then  $\varphi^{\neg\neg\varphi} \rightarrow (\neg\varphi \rightarrow \varphi)$ , hence  $\varphi^{\neg\neg\varphi} \rightarrow \neg\neg\varphi$ . Hence  $\vdash \exists x(\neg\varphi \rightarrow \psi(x))$ . Friedman's translation is closely related to a straightforward translation of intuitionistic into minimal logic, cf. [Leivant, 1985] for details and also for syntactic criteria for closure under Markov's rule.

Closure under Markov's rule is exactly what one needs for identifying provably recursive functions in classical and intuitionistic arithmetic. Using the notion of Ch. 4 of Vol. 1 of this *Handbook*, we can say that the recursive function with index  $e$  is provably recursive in a theory  $S$  if  $S \vdash \forall x\exists yT(e, x, y)$  (for each input  $x$  the computation provably halts). Closure under Markov's rule tells us that **PA** and **HA** have exactly the same provably recursive functions (Kreisel). In other words, by restricting our arguments to intuitionistic logic we do not lose any recursive functions. Friedman extended this result to classical and intuitionistic set theory ZF ([Friedman, 1977], cf. also [Leivant, 1985]).

## 8 RELATION WITH OTHER LOGICS

First we consider a sub-logic of intuitionistic logic. Minimal logic was proposed by Johansson in reaction to the role of negation, in particular the *Ex falso sequitur quodlibet* rule (our *falsum rule*). His critique resulted in a rejection of the rule ' $\perp$ '. As a result, in his system of *minimal logic*,  $\perp$  cannot properly be distinguished from other atoms. This is reflected in the Kripke semantics for minimal logic.

**DEFINITION 46.** A Kripke model for **MQC** is obtained from Definitions 10 and 13 by deleting the condition on  $\perp$  (i.e.  $\alpha \Vdash \perp$  is allowed).

By a proof that is completely similar to that of Lemma 14 we get

**THEOREM 47 (Completeness for MQC).**

$$\Gamma \Vdash_{\mathbf{MQC}} \varphi \Leftrightarrow \Gamma \Vdash \varphi$$

(where  $\Vdash$  is understood in the sense of Definition 46).

It now follows immediately that **MQC** is a proper subsystem of **IQC** (similarly for **MPC** and **IPC**), for **MQC**  $\not\Vdash \perp \rightarrow \varphi$ . Consider the one point model in which  $\perp$  is forced.

Although minimal logic is strictly weaker than intuitionistic logic, they are in a sense of the same strength. To be precise, each can faithfully be interpreted in the other.

DEFINITION 48. The translations  $*$  and  $\dagger$  are defined by

$$\begin{aligned} \varphi^* &:= \varphi \vee \perp \text{ for atomic } \varphi \\ (\varphi \wedge \psi)^* &:= \varphi^* \wedge \psi^* \\ (\varphi \vee \psi)^* &:= \varphi^* \vee \psi^* \\ (\varphi \rightarrow \psi)^* &:= \varphi^* \rightarrow \psi^* \\ (\forall x\varphi)^* &:= \forall x\varphi^* \\ (\exists x\varphi)^* &:= \exists x\varphi^* \\ \\ \varphi^\dagger &:= \varphi[p/\perp]. \end{aligned}$$

where  $p$  is a propositional letter not occurring in  $\varphi$ .

Observe that the translation  $\dagger$  eliminates  $\perp$ , so for  $\varphi^\dagger$  we cannot use the falsum rule in **IQC**. That makes it plausible that  $\varphi^\dagger$  behaves in **IQC** as  $\varphi$  does in **MQC**.

THEOREM 49.

1. **IQC**  $\vdash \varphi \Leftrightarrow$  **MQC**  $\vdash \varphi^*$
2. **MQC**  $\vdash \varphi \Leftrightarrow$  **IQC**  $\vdash \varphi^\dagger$ ,

**Proof.** (1)  $\Leftarrow$  is trivial.

For  $\Rightarrow$  use induction on the derivation of  $\Gamma \vdash \varphi$  and observe

$$\vdash \Gamma \rightarrow \varphi \Leftrightarrow \Gamma \vdash \varphi.$$

(2)  $\perp$  behaves in the semantics for minimal logic like any atom, so validity of  $\varphi$  in all ‘minimal’ Kripke models is equivalent to validity of  $\varphi^\dagger$  in all ‘intuitionistic’ Kripke models. Alternatively, a simple proof theoretic argument based on the normal form theorem will do.  $\blacksquare$

Minimal logic enjoys most metalogical properties that can be expected, e.g. there is a normal form theorem for natural deduction derivations, normal derivations have the subformula property, etc. Its propositional calculus, **MPC**, is decidable (use Theorem 49). For more information the reader is referred to [Johansson, 1936; Prawitz, 1965; Prawitz and Malmnäs, 1968].

Our next goal is to investigate the relation between classical logic and intuitionistic logic. The first results antedate Heyting’s formalisation. Kolmogorov already in 1925 established a translation procedure [Kolmogorov, 1925], and in [Glivenko, 1929] a similar result is to be found. The next to investigate the relation between classical and intuitionistic logic (in the wider

context of arithmetic) were Gödel and Gentzen ([Gödel, 1933; Gentzen, 1933]). for more information on translations the reader is referred to [Leivant, 1985], [Troelstra and van Dalen, 1988].

We first present the result of Glivenko:

**THEOREM 50.**

$$\mathbf{CPC} \vdash \varphi \Leftrightarrow \mathbf{IPC} \vdash \neg\neg\varphi.$$

**Proof.** Recall that **IPC** is complete for finite Kripke models. Since in each maximal node all tautologies are forced, we see that  $\neg\neg\varphi$  is valid in all finite Kripke models if  $\varphi$  is a classical tautology. This shows  $\Rightarrow$ . The converse is trivial. ■

**COROLLARY 51.**

$$\mathbf{CPC} \vdash \neg\varphi \Leftrightarrow \mathbf{IPC} \vdash \neg\varphi.$$

There are a number of translations from **IQC** into **CQC** that have roughly similar properties. The main feature is the elimination of  $\forall$  and  $\exists$ . Let us call a formula *negative* if it does not contain  $\forall$  and  $\exists$  and if all its atoms are negated. The translation below assigns to each formula a negative formula.

Negative formulas have the following convenient property

**LEMMA 52.**  $\mathbf{IQC} \vdash \varphi \Leftrightarrow \neg\neg\varphi$  for *negative*  $\varphi$ .

**Proof.** An exercise in plain old logic. ■

**DEFINITION 53.** The translation  $^\circ$  is given by

$$\begin{aligned} \varphi^\circ &:= \neg\neg\varphi \text{ for atomic } \varphi \\ (\varphi \wedge \psi)^\circ &:= \varphi^\circ \wedge \psi^\circ \\ (\varphi \rightarrow \psi)^\circ &:= \varphi^\circ \rightarrow \psi^\circ \\ (\varphi \vee \psi)^\circ &:= \neg(\neg\varphi^\circ \wedge \psi^\circ) \\ (\forall x\varphi)^\circ &:= \forall x\varphi^\circ \\ (\exists x\varphi)^\circ &:= \neg\forall x\neg\varphi^\circ \end{aligned}$$

**THEOREM 54.**

1.  $\mathbf{CQC} \vdash \varphi \Leftrightarrow \mathbf{IQC} \vdash \varphi^\circ$
2.  $\mathbf{CQC} \vdash \varphi \Leftrightarrow \varphi^\circ$ .

**Proof.** (2) is routine. For (1) we consider instead  $\Gamma \vdash^c \varphi$  and  $\Gamma^\circ \vdash^i \varphi^\circ$  (where  $\vdash^c, \vdash^i$  stand for classical and intuitionistic derivability, and  $\Gamma^\circ$  is the set of translated  $\psi$ 's from  $\Gamma$ ). Proof by induction on the derivation of  $\Gamma \vdash \varphi$ , use Lemma 52. ■

Since negative formulas are invariant under the translation we get  
COROLLARY 55.

1.  $\mathbf{CQC} \vdash \varphi \Leftrightarrow \mathbf{IQC} \vdash \varphi$  for negative  $\varphi$ .
2. For theories with decidable atoms (i.e.  $T \vdash \varphi \vee \neg\varphi$  for atomic  $\varphi$ ) we have  $T \vdash^c \varphi \Leftrightarrow T^\circ \vdash^i \varphi$  for negative  $\varphi$ .

The latter is the case for arithmetic,  $\mathbf{HA}$ .

The above result tells us that  $\mathbf{PA}$  and  $\mathbf{HA}$  are relatively consistent, so intuitionistic arithmetic is, from a foundational point of view, just as much in need of a consistency proof as  $\mathbf{PA}$ .

There are some special results in the area, we list some.

FACTS 56.

1.  $\mathbf{CQC} \vdash \varphi \Leftrightarrow \mathbf{IQC} \vdash \varphi$  for  $\varphi$  a negation of a prenex formula (Kreisel).
2.  $\mathbf{CQC} \vdash \varphi \Leftrightarrow \mathbf{IQC} \vdash \neg\neg\varphi$ , and  
 $\mathbf{CPC} \vdash \neg\varphi \Leftrightarrow \mathbf{IQC} \vdash \neg\varphi$ , for  $\varphi$  without  $\forall$ . [Fitting, 1969, p. 52].
3. If  $\forall$  does not occur negatively in  $\varphi$  then  $\mathbf{CQC} \vdash \neg\varphi \Leftrightarrow \mathbf{IQC} \vdash \neg\varphi$  [Smoryński, 1973].
4. If  $\varphi$  is a  $\Pi_2^0$ -sentence (i.e. of  $\forall\exists$  form), then  $\mathbf{PA} \vdash \varphi \Leftrightarrow \mathbf{HA} \vdash \varphi$ . (Kreisel, cf. [Troelstra, 1973, Ch. 3, S. 8], lemma 46.)

#### *Intuitionistic logic seen from the modal viewpoint*

As we have sketched earlier, intuitionistic logic has certain strong intensional aspects, in particular the meaning of the connectives—expressed in terms of proofs and construction, or of knowledge—has an intensional ring. This has been observed by Gödel, who proposed a translation of intuitionistic logic into modal logic [Gödel, 1932].

The ‘necessity’ operator could be read here as ‘I have a proof’ or ‘I know that’.

DEFINITION 57. The translation  $^m$  is defined by

$$\begin{aligned} \varphi^m &:= \Box\varphi \text{ for atomic } \varphi \\ (\varphi \vee \psi)^m &:= \varphi^m \vee \psi^m \\ (\varphi \wedge \psi)^m &:= \varphi^m \wedge \psi^m \\ (\varphi \rightarrow \psi)^m &:= \Box(\varphi^m \rightarrow \psi^m) \end{aligned}$$

We will establish the relation between  $\mathbf{S4}$  and  $\mathbf{IPC}$ . Observe that in  $\mathbf{S4}$

$$\Box\perp \leftrightarrow \perp, \text{ so } (\neg\varphi)^m \leftrightarrow \Box\neg\varphi^m.$$



We borrow from modal logic the fact that **S4** is complete for Kripke models  $\langle K, R, \Vdash \rangle$  with a reflexive, transitive  $R$  (cf. [Hughes and Cresswell, 1968; Schütte, 1968] and Chapter II.1 of the *Handbook*).

The forcing notion is defined by

$$\begin{aligned} \alpha &\not\Vdash \perp \\ \alpha &\Vdash \varphi \wedge \psi \Leftrightarrow \alpha \Vdash \varphi \text{ and } \alpha \Vdash \psi \\ \alpha &\Vdash \varphi \vee \psi \Leftrightarrow \alpha \Vdash \varphi \text{ or } \alpha \Vdash \psi \\ \alpha &\Vdash \varphi \rightarrow \psi \Leftrightarrow \alpha \not\Vdash \varphi \text{ or } \alpha \Vdash \psi \\ \alpha &\Vdash \Box\varphi \Leftrightarrow \text{for all } \beta \text{ with } \alpha R\beta, \beta \Vdash \varphi. \end{aligned}$$

Observe that the propositional fragment is classical. Further, an intuitionistic Kripke model may be viewed as a modal Kripke model (not always conversely).

**LEMMA 58.** *Let  $\langle K, \leq, \Vdash \rangle$  be an intuitionistic Kripke model. Define a modal Kripke model with the same underlying poset and the same forcing for atoms denoted by  $\Vdash^m$ . Then for all  $\varphi, \alpha \Vdash \varphi \Leftrightarrow \alpha \Vdash^m \varphi^m$ ,*

**Proof.** Induction on  $\varphi$ . ■

**THEOREM 59.**  $\mathbf{IPC} \vdash \varphi \Leftrightarrow \mathbf{S4} \vdash \varphi^m$ .

**Proof.**  $\Leftarrow$ . Let  $\mathbf{IPC} \not\vdash \varphi$ , then there is a Kripke model with  $\alpha \not\Vdash \varphi$  for the bottom node  $\alpha$ . Now apply Lemma 58, then  $\mathbf{S4} \not\vdash \varphi^m$ .

$\Rightarrow$ . Let  $\mathbf{S4} \not\vdash \varphi^m$ , then there is a Kripke model such that  $\alpha_0 \not\Vdash^m \varphi^m$  for the bottom node  $\alpha$ : we turn this model into an intuitionistic Kripke model by first collapsing the model, i.e. we consider the equivalence relation  $\alpha \sim \alpha' := \alpha \leq \alpha' \wedge \alpha' \leq \alpha$  and introduce a new underlying set of nodes  $\alpha / \sim$ . For  $\alpha / \sim$  we put  $\alpha / \sim \Vdash \varphi := \alpha \Vdash^m \varphi^m$ , for atomic  $\varphi$ . This relation is obviously well-defined, so is the forcing relation for all formulas. An argument similar to that of Lemma 58 establishes  $\alpha / \sim \Vdash \varphi \Leftrightarrow \alpha \Vdash^m \varphi^m$ . we now conclude  $\alpha_0 / \sim \not\Vdash \varphi$ , so  $\mathbf{IPC} \not\vdash \varphi$ . ■

Artemov has picked up Gödel's thread and designed logic which incorporates both 'proof' and 'modality', [Artemov, 2001].

### 8.1 Strong Negation

Intuitionistic negation does not conform to the classical laws of double negation, De Morgan, etc. This is mainly so because negation is a rather weak connective. Think of its interpretations in a Kripke model: it is not decided on the spot. Or one may think of inequality versus apartness, 'being unequal' carries so much less information than 'being apart'. Could we possibly strengthen negation, so that the classical rules would be obeyed? As a matter of fact, this is what Nelson [1949] and Markov [1950] have done.

The new connective can more or less be viewed as an attempt to save the classical laws by brute force. Let us write ‘ $\sim \varphi$ ’ for the strong negation of  $\varphi$ .

The axioms for the logic with strong negation are those of **IQC** plus the following

$$\begin{aligned}
& \sim (\varphi \rightarrow \psi) \leftrightarrow \varphi \wedge \sim \psi \\
& \sim (\varphi \wedge \psi) \leftrightarrow \sim \varphi \vee \sim \psi \\
& \sim (\varphi \vee \psi) \leftrightarrow \sim \varphi \wedge \sim \psi \\
& \sim \varphi \wedge \varphi \rightarrow \psi \\
& \sim \exists x \varphi(x) \leftrightarrow \forall x \sim \varphi(x) \\
& \sim \forall x \varphi(x) \leftrightarrow \exists x \sim \varphi(x) \\
& \sim \neg \varphi \leftrightarrow \varphi, \sim \sim \varphi \leftrightarrow \varphi, \sim \varphi \rightarrow \neg \varphi.
\end{aligned}$$

One obtains a Kripke model for a logic with strong negation by incorporating a strong falsity which is verified at the spot.

**DEFINITION 60.** A Kripke model is a quadruple  $\langle K, \leq, D, i \rangle$  where  $K, \leq$  and  $D$  are as in Definition 10,  $i$  is the interpretation map which assigns  $-1, 0, 1$  to atoms and nodes such that  $\alpha \leq \beta, i(\varphi, \alpha) \neq 0 \Rightarrow i(\varphi, \alpha) = i(\varphi, \beta)$ .

We define  $[\varphi]_\alpha$  for formulas  $\varphi$  and nodes  $\alpha$ :

$$[\varphi]_\alpha := i(\varphi, \alpha) \text{ for atomic } \varphi \text{ with parameters in } D(\alpha),$$

where  $[\perp]_\alpha \neq 1$

$$[\varphi \wedge \psi]_\alpha := \min([\varphi]_\alpha, [\psi]_\alpha)$$

$$[\varphi \vee \psi]_\alpha := \max([\varphi]_\alpha, [\psi]_\alpha)$$

$$[\varphi \rightarrow \psi]_\alpha := \begin{cases} 1 & \text{if } \forall \beta \geq \alpha, [\varphi]_\beta = 1 \Rightarrow [\psi]_\beta = 1 \\ -1 & \text{if } [\varphi]_\alpha = 1 \text{ and } [\psi]_\alpha = -1 \\ 0 & \text{otherwise} \end{cases}$$

$$[\sim \varphi]_\alpha = \begin{cases} 1 & \text{if } [\varphi]_\alpha = -1 \\ -1 & \text{if } [\varphi]_\alpha = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$[\neg \varphi]_\alpha = \begin{cases} 1 & \text{if } \forall \beta \geq \alpha, [\varphi]_\beta \neq 1 \\ -1 & \text{if } [\varphi]_\alpha = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$[\forall x \varphi(x)]_\alpha = \begin{cases} 1 & \text{if } \forall \beta \geq \alpha, \forall b \in D(\beta) [\varphi(b)]_\beta = 1 \\ -1 & \text{if } \exists a \in D(\alpha) [\varphi(a)]_\alpha = -1 \\ 0 & \text{otherwise} \end{cases}$$

$$[\exists x \varphi(x)]_\alpha = \begin{cases} 1 & \text{if } \exists a \in D(\alpha), [\varphi(a)]_\alpha = 1 \\ -1 & \text{if } \forall \beta \geq \alpha, \forall b \in D(\beta) [\varphi(b)]_\beta = -1 \\ 0 & \text{otherwise.} \end{cases}$$

It is a matter of routine to show this semantics sound for logic with strong negation,  $\mathbf{IQC}^{sn}$ , in the sense that  $\mathbf{IQC}^{sn} \vdash \varphi \Rightarrow [\varphi]_\alpha = 1$  in all Kripke models  $\mathcal{K}$  and nodes  $\alpha \in K$ .

There are alternative ways to present the same model. One can assign to each node two sets of atomic statements:  $\alpha^+ = \{\varphi \mid i(\varphi, \alpha) = 1\}$  and  $\alpha^- = \{\varphi \mid i(\varphi, \alpha) = -1\}$ , and extend those sets as shown above to the strongly verified and strongly falsified sentences. Some authors write  $\alpha \Vdash \varphi$  for  $[\varphi]_\alpha = 1$  and  $\alpha \dashv\vdash \varphi$  for  $[\varphi]_\alpha = -1$ . These notations are purely a matter of convenience.

The completeness of  $\mathbf{IQC}^{sn}$  can be shown by the usual Henkin- or tableaux-technique (cf. [Thomason, 1969]) but also by a reduction to  $\mathbf{IQC}$ . We will indicate the steps.

(1) Observe that all strong negations can be driven in, so each  $\varphi$  is provably equivalent to a  $\varphi^*$  with all strong negations in front of atoms.

(2) We want to consider strongly negated atoms as atoms in their own right, so we double the language by adding a predicate  $\hat{P}$  for each predicate  $P$ . Indicate the new atoms by  $\hat{\varphi}$ . At the very least the strongly negated atoms should imply the negated atoms. Put  $\Gamma = \{\hat{\varphi} \rightarrow \neg\varphi \mid \varphi \text{ atomic sentence}\}$ , and let  $\bar{\sigma}$  be the formula one obtains by replacing  $\sim\sigma$  (atomic  $\varphi$ ) by  $\hat{\varphi}$  in  $\sigma^*$ .

*Claim:*  $\Gamma \vdash_{\mathbf{IQC}} (\sim\sigma) \rightarrow \neg\bar{\sigma}$  for all  $\sigma$ .

Prove this by induction on  $\sigma$ .

(3)  $\Gamma$  allows us to reduce  $\mathbf{IQC}^{sn}$  to  $\mathbf{IQC}$  in the following sense:  $\mathbf{IQC}^{sn} \vdash \sigma \Leftrightarrow \mathbf{IQC} + \Gamma \vdash \bar{\sigma}$ .

*Proof:* Induction on the proof length (or on the derivation in natural deduction).

(4) We can now apply the completeness theorem for  $\mathbf{IQC}$ . Let  $\mathbf{IQC}^{sn} \not\vdash \sigma$  then  $\mathbf{IQC} + \Gamma \not\vdash \bar{\sigma}$ , so there is an ordinary Kripke model  $\mathcal{K}$  in which all axioms of  $\Gamma$  are valid, but not so  $\bar{\sigma}$ . Turn  $\mathcal{K}$  into a strong negation model  $\mathcal{K}'$ , putting  $\alpha \dashv\vdash \varphi$  if  $\alpha \Vdash \hat{\varphi}$  for atomic sentences  $\varphi$ . Now one shows that  $\mathcal{K}'$  is a model of  $\mathbf{IQC}^{sn}$ , but not of  $\sigma$ .

Since an ordinary model is trivially a strong negation model, it is immediately seen that  $\mathbf{IQC}^{sn}$  is conservative over  $\mathbf{IQC}$ .

$\mathbf{IQC}^{sn}$  has some unusual properties, e.g.  $\vdash \varphi \leftrightarrow \psi$  does not imply  $\vdash \sim\varphi \leftrightarrow \sim\psi$  (consider  $\neg\varphi \leftrightarrow \varphi \rightarrow \perp$ ). For more information, cf. [Gabbay, 1981, p. 124 ff], [Rasiowa, 1974, Ch. XII] and [Rautenberg, 1979, p. 305 ff].

## 8.2 The Connections with $\lambda$ -Calculus and Combinatory Logic

Already in 1958 Curry pointed out that there is a remarkable correspondence between the implication fragment of  $\mathbf{IPC}$  and combinatory logic,  $\mathbf{CL}$ .

In particular the axioms correspond to the **CL** axioms as follows:

$$\begin{array}{lll}
\varphi \rightarrow \varphi & - Ix & = x \\
\varphi \rightarrow (\psi \rightarrow \varphi) & - Kxy & = x \\
(\sigma \rightarrow (\varphi \rightarrow \psi)) \rightarrow ((\sigma \rightarrow \varphi) \rightarrow (\sigma \rightarrow \psi)) & - Sxyz & = xz(yz).
\end{array}$$

Howard extended the correspondence in his ‘Formulas as types’ paper (1969, published in [Howard, 1980]). Once such a correspondence exists one is almost forced to look at the reduction processes in **CL** or  $\lambda$ -calculus and in natural deduction systems (cf. [Prawitz, 1971]). In Pottinger [1976] the isomorphism between natural deduction derivations and  $\lambda$ -terms has been exploited to obtain alternative proofs of the normalisation theorem for **IPC**.

In Martin-Löf’s type theory the parallelism between types and formulas is a key feature. For more information the reader is referred to Martin-Löf [1977; 1984] and Troelstra and van Dalen [1988].

## 9 THE ALGORITHMIC TRADITION

Intuitionistic logic was intended to codify constructive reasoning. The proof-interpretation expresses the meaning of the logical constants in terms of constructions. It seems plausible to try to delimit the class of constructions involved. Stephen Kleene conjectured in 1940 that in particular for a statement of the form  $\forall x \exists y \varphi(x, y)$  provability in **HA** should entail the existence of a recursive function  $f$  that acts as a choice function:  $\forall x \varphi(x, f(x))$  (cf. [Kleene, 1973]). This led Kleene to the notion of statements as ‘incomplete communications’, taking his cue from Hermann Weyl, see [van Dalen, 1995] e.g.  $\exists x \varphi(x)$  is an incomplete communication of a fuller statement giving an object  $x$  such that  $\varphi(x)$ . Likewise the other composite statements can be considered as incomplete statements, to be supplemented by extra information.

The result was the so-called *1945-realizability* or *recursive realizability*, a notion that we will formulate in the framework of **HA**. The sentence  $\varphi$  is realized by the number  $n$ ,  $nr\varphi$ , must, be thought of as  $n$  codes ‘the necessary information to establish  $\varphi$ ’.

**DEFINITION 61.**  $x r \varphi$  is a formula of **HA** with at most one free variable  $x$ , associated to the sentence  $\varphi$  (we use notation from Ch. 4 of Vol. 1 of this *Handbook*).

$$\begin{array}{ll}
x r \varphi & := \varphi \text{ for atomic } \varphi \\
x r (\varphi \wedge \psi) & := (x)_0 r \varphi \wedge (x)_1 r \psi \\
x r (\varphi \vee \psi) & := ((x)_0 = 0 \rightarrow (x)_1 r \varphi) \wedge ((x)_0 \neq 0 \rightarrow (x)_1 r \psi) \\
x r (\varphi \rightarrow \psi) & := \forall y (y r \varphi \rightarrow \{x\}y \downarrow \wedge \{x\}y r \psi) \\
x r \exists y \varphi(y) & := (x)_1 r \varphi((x)_0) \\
x r \forall y \varphi(y) & := \forall y (\{x\}y \downarrow \wedge \{x\}y r \varphi(y))
\end{array}$$

*Explanation:* The first clause tells us that any number realizes a true atomic sentence. However, no number realizes a false atomic sentence. The second clause is self-evident. The clause for the disjunction exhibits the effective nature of the disjunction. We can effectively test if  $(x)_0 = 0$  or  $(x)_0 \neq 0$ . Hence, the ‘realizer’ of a disjunction contains enough information to indicate the desired disjunct. The implication clause shows a resemblance to the proof interpretation of  $\rightarrow$ ,  $x$  is the index of a partial recursive function that transforms any realizer of  $\varphi$  into a realizer of  $\psi$ . In the case of  $\forall$  a similar resemblance can be observed. The clause for  $\exists$  tells us that a realizer of  $\exists y\varphi(y)$  contains the required instance and the information that realizes it.

Note that  $xr\varphi$  is a formula of **HA**, so it makes sense to ask for the truth of an instance  $nr\varphi$ , or its derivability in **HA**.

EXAMPLES 62.

1.  $x r (2 = 1 + 1) :\Leftrightarrow 2 = 1 + 1 (\leftrightarrow \top)$ .
2.  $x r \forall z \exists y (z = y) :\Leftrightarrow$   
 $\forall z (\{x\}z \downarrow \wedge \{x\}z r \exists y (z = y)) :\Leftrightarrow$   
 $\forall z (\{x\}z \downarrow \wedge (\{x\}z)_1 r (z = (\{x\}z)_0)) :\Leftrightarrow$   
 $\forall x (\{x\}z \downarrow \wedge z = (\{x\}z)_0)$ .

So if we take the index  $e$  of the identity function  $z \mapsto z$ , then  $(\{x\}z)_0 = \{e\}z$  and we can put  $\{x\}z = \langle \{e\}z, 0 \rangle$ . This is a (total) recursive function, so it has an index, say  $e_0$ . The number  $e_0$  realizes  $\forall z \exists y (z = y)$ .

Kleene’s realizability can be considered as an interpretation of **HA** in **HA** bringing out the constructive character of **HA**. This interpretation is sound in the following sense:

**THEOREM 63.**  $\mathbf{HA} \vdash \varphi \Rightarrow \mathbf{HA} \vdash n r \varphi$  for some  $n$ .

The proof is mainly a matter of perseverance (cf. [Kleene, 1952, p. 504], [Troelstra, 1973, p. 189]).

A consequence of this theorem is the fact that (assuming the consistency of **HA**) a realizable sentence  $\varphi$  is consistent with **HA**. For suppose that  $nr\varphi$  and  $\mathbf{HA} + \varphi$  is inconsistent, then  $\mathbf{HA} \vdash \neg\varphi$ , and hence  $\mathbf{HA} \vdash mr(\neg\varphi)$  for some  $m$ . But  $mr(\neg\varphi)$  is equivalent to  $\forall y (yr\varphi \rightarrow \{m\}y \downarrow \wedge \{m\}yr\perp)$ , and since  $\perp$  is not realisable, neither is  $\varphi$ . Contradiction.

The most striking application of this procedure for establishing consistency is:

**THEOREM 64.** *Church’s Thesis is consistent with HA.*

**Proof.** We have in mind a special form of Church’s Thesis, namely one that can be formulated in **HA**. We choose the following form:

$$\text{CT}_0 \quad \forall x \exists y \varphi(x, y) \rightarrow \exists z \forall x (\{z\}x \downarrow \wedge \varphi(x, \{z\}x)).$$

Observe that we can avoid the abbreviation  $\{z\}x \downarrow: \forall x \exists y \varphi(x, y) \rightarrow \exists z \forall x \exists u (T(z, x, u) \wedge \varphi(x, Uu))$  where  $T$  is Kleene's  $T$  predicate (cf. van Dalen's Algorithms Chapter in Volume 1 of this *Handbook*), and  $U$  is the output-extraction function.

For convenience we suppose that  $\varphi$  has only the variables  $x$  and  $y$  free.

We will need the following notation: if  $t$  is a term for a partial recursive function, then  $\Lambda x.t$  is the index of the partial recursive function given by  $t$  depending on  $x$  (if there are more variables we consider them as parameters; strictly speaking the notation is based on the  $S_n^m$ -theorem, Cf. Vol. 1 of this *Handbook*, p. 275, or [Kleene, 1952, p. 344]). For example,  $\Lambda x.x + y$  is the index of the unary function that adds  $y$ .

We will sketch the proof in such a way that the reader, if he wishes to do so, can provide the full details himself.

Let  $u r \forall x \exists y \varphi(x, y)$ , then  $\forall x (\{u\}x \downarrow \wedge \{u\}x r \exists y \varphi(x, y))$ , i.e.

$$\forall x (\{u\}x \downarrow \wedge (\{u\}x)_1 r \varphi(x, (\{u\}x)_0)) \dots (0).$$

Put  $t := \{u\}x$ , and  $a = \Lambda x.(t)_0, b = \mu w T(a, x, w), \delta(u) = \langle a, \Lambda x.\langle b, \langle 0, (t)_1 \rangle \rangle \rangle$ .

*Claim:*  $\delta(u) r \exists z \forall x \exists v (T(z, x, v) \wedge \varphi(x, Uv)) \dots (1)$ .

We carry out the steps as given in the definition.

$$\Lambda x.\langle b, \langle 0, (t)_1 \rangle \rangle r \forall x \exists v (T(a, x, v) \wedge \varphi(x, Uv)) \dots (2).$$

$$\langle b, \langle 0, (t)_1 \rangle \rangle r \exists v (T(a, x, v) \wedge \varphi(x, Uv)) \dots (3).$$

$$\langle 0, (t)_1 \rangle r T(a, x, b) \wedge \varphi(x, Ub) \dots (4).$$

or

$$T(a, x, b) \wedge (t)_1 r \varphi(x, Ub) \dots (5).$$

Now observe that by the definition of  $a$  and  $b, T(a, x, b)$  is true for all  $x$ . So  $0$  realizes it (where for convenience  $T(a, x, b)$  has been taken to be atomic; this is achieved by a simple conservative extension of **HA**). Furthermore,  $Ub$  is the output of  $\{a\}$  on input  $x$ , which is  $(t)_0$ , so  $(t)_1 r \varphi(x, Ub)$  can be read as  $(\{u\}x)_1 r \varphi(x, (\{u\}x)_0)$ . This holds by (0). The passage from (0) to (1) tells us that  $\Lambda u.\delta(u) r \text{CT}_0$ . ■

Almost the same argument establishes  $\text{ECT}_0$  (see below) [Troelstra, 1973, p. 195].

Troelstra has investigated the theory of the realizable sentences of arithmetic. It turns out that this fragment has a simple axiomatization (cf. [Troelstra, 1998, p. 416]).

$$\begin{aligned} \mathbf{HA} + \text{ECT}_0 &\vdash \varphi \leftrightarrow \exists x (x r \varphi) \\ \mathbf{HA} + \text{ECT}_0 &\vdash \varphi \leftrightarrow \mathbf{HA} \vdash \exists x (x r \varphi), \end{aligned}$$

where  $\text{ECT}_0$  is the Extended Thesis of Church:

$$\forall x(\varphi \rightarrow \exists y\psi xy) \rightarrow \exists u\forall x(\varphi \rightarrow (\{u\}x \downarrow \wedge \psi(x, \{u\}x)))$$

for *almost negative*  $\varphi$  (i.e.  $\varphi$  does not contain  $\forall$ , and  $\exists$  only in front of atoms).

$\mathbf{HA} + \text{CT}_0$  has been studied in detail by David McCarty, for his results see [McCarty, 1988]. Perhaps the most striking fact established by him is the categoricity of the theory:  $\mathbf{HA} + \text{CT}_0$  has no non-standard models.

Since Kleene's pioneering papers there has been a proliferation of notions. The reader is referred to [Troelstra, 1973] and [Troelstra, 1998] for the major notions in the context of arithmetic. There are also extensions to higher theories (e.g. set theory- like ones) (cf. [Feferman, 1979], [Beeson, 1985]).

In the fifties Gödel proposed a new interpretation of  $\mathbf{HA}$  (and extensions) based on functionals of all finite types (cf. [Gödel, 1958; Kreisel, 1959; Troelstra, 1973; Avigad and Feferman, 1998]). The basic idea is to reduce the logical complexity of sentences at the cost of increasing the types of the objects. Kreisel proposed the notion of 'modified realizability' (cf. Troelstra [1973; 1998]); Kleene transferred realizability to analysis by means of 'continuous function application'; in the context of first-order logic we mention Läuchli's 'abstract realizability'. A systematic and unifying treatment of various realizabilities has been given (cf. [Stein, 1980]).

The above-mentioned interpretations have led to a wealth of proof theoretic results, such as conservative extensions, and closure under rules. the reader is referred to [Troelstra, 1973; Troelstra, 1998] for detailed information.

The Russian school of A. A. Markov has made the algorithmic tradition the guideline for its actual mathematical practice. Its members consider mathematics as dealing with concrete, constructive objects. In particular they adhere to Church's thesis, so that, e.g. real numbers in their approach are given by recursive Cauchy sequences (hence the name 'recursive analysis'). Following Markov, they accept the principle  $\neg\neg\exists x\varphi(x) \rightarrow \exists x\varphi(x)$  for primitive recursive  $\varphi(x)$ —*Markov's Principle*. For a survey, cf. [Demuth and Kučera, 1979].

For a long time the 'algorithmic' interpretations have withstood attempts of unifying treatment together with the semantic interpretations. Recently, however, the framework of topos theory has provided a more semantic treatment of, e.g. realisability interpretations. In particular work of Hyland, Johnstone and Pitts [1980] on tripos theory and Hyland [1982] on the effective topos has provided a semantical home for the above kind of interpretations.

## 10 SECOND-ORDER LOGIC

Whereas first-order intuitionistic logic and its prominent theories, such as arithmetic, are just subtheories of the corresponding classical ones, the notions of second-order logic seem to dictate their own laws in the light of intuitionistic conceptions.

Traditionally, second-order logic is concerned with individuals, sets (and relations) and the only non-logical principle that is considered is the so-called comprehension axiom. Most studies are centered around second-order arithmetic, and extensions of it.

We will first discuss second-order logic.

The language of intuitionistic second-order logic  $\mathbf{IQC}^2$  contains variables and constants for

$$\begin{array}{ll} \text{individuals} & - \quad x_0, x_1, x_2, \dots, \quad c_0, c_1, c_2, c_3, \dots \\ n\text{-ary relations} & - \quad X_0^n, X_1^n, X_2^n, \dots, \quad C_0^n, C_1^n, C_2^n, \dots, \end{array}$$

where  $n \geq 0$ .

0-ary variables (constants) are called *propositional variables* (constants), 1-ary variables (constants) are called *set (species) variables* (constants). The atoms of  $\mathbf{IQC}^2$  are of the form  $X^0, C^0$  for 0-ary second-order terms, or  $X^n(t_1, \dots, t_n), C^n(t_1, \dots, t_n)$  for  $n$ -ary second-order terms  $X^n, C^n$  and first-order terms  $t_1, \dots, t_n$  (i.e. individual variables or constants).

In classical logic one thinks of 0-ary terms as denoting the truth values ‘true’, ‘false’. In our case we may think of truth values in a Heyting-algebra.

Formulas are defined as usual by means of the connectives  $\wedge, \vee, \rightarrow, \perp, \forall x, \forall X^n, \exists x, \exists X^n$ .

The rules of derivation (in Natural Deduction) are extended by the following quantifier rules:

$$\begin{array}{ccc} \forall^2 I \frac{\varphi}{\forall X^n \varphi} & & \forall^2 E \frac{\forall X^n \varphi}{\varphi^*} \\ & & \exists X^n \varphi \quad [\varphi] \\ \exists^2 I \frac{\varphi^*}{\exists X^n \varphi} & & \exists^2 E \frac{\begin{array}{c} \vdots \\ \psi \end{array}}{\psi} \end{array}$$

where  $\varphi^*$  is obtained from  $\varphi$  by replacing each occurrence of  $X^n(t_1, \dots, t_n)$  by  $\sigma(t_1, \dots, t_n)$ , for a certain  $\sigma$ , such that no free variable among the  $t_i$  becomes bound after substitutions.

Observe that  $\exists^2 I$  takes the place of the traditional *Comprehension Principle* (cf. [van Dalen, 1997, Chapter 4])

$$\exists X^n \forall x_1, \dots, x_n [\varphi(x_1, \dots, x_n) \leftrightarrow X^n(x_1, \dots, x_n)].$$



The surprise of second-order logic is the fact that the usual connectives are definable in terms of  $\forall$  and  $\rightarrow$ , this in sharp contrast to **IQC** (Prawitz). Given the rules for  $\forall$  and  $\rightarrow$  we can define the connectives as follows.

**DEFINITION 65.**

1.  $\perp := \forall X^0. X^0$
2.  $\varphi \wedge \psi := \forall X^0[(\varphi \rightarrow (\psi \rightarrow X^0)) \rightarrow X^0]$
3.  $\varphi \vee \psi := \forall X^0[(\varphi \rightarrow X^0) \rightarrow ((\psi \rightarrow X^0) \rightarrow X^0)]$
4.  $\exists x\varphi := \forall X^0[\forall x(\varphi \rightarrow X^0) \rightarrow X^0]$
5.  $\exists X^n\varphi := \forall X^0[\forall X^n(\varphi \rightarrow X^0) \rightarrow X^0]$ .

To be precise: given the rules for  $\forall$  and  $\rightarrow$  we can prove the rules for the defined connectives (cf. [Prawitz, 1965, p. 67], [van Dalen, 1997, p. 152]). For proof theoretical purposes the reduction of the number of connectives turns out to be an asset (cf. [Tait, 1975; Prawitz, 1971]),

The semantics for second-order logic are relatively straightforward generalizations of the existing semantics for first-order logics (cf. [Prawitz, 1970; Takahashi, 1970; Fourman and Scott, 1979]).

### 10.1 *Second-order Arithmetic, HAS*

The simplest formalisations of **HAS** (Heyting's second-order arithmetic with set variables), is obtained by adding the axioms for **HA** to second-order logic (in an extended language containing the obligatory operations and relations for arithmetic). Observe that, as a schema, the induction axiom is defined for the full language. The traditional issue in second-order arithmetic concerns the Comprehension Principle, CA. Should it have the full strength or should it be restricted to the predicative case? This topic has never been really central in intuitionistic considerations on higher-order objects. There certainly is not much to go on in Brouwer's writings. If we embrace the viewpoint that a set  $X$  is given when we know what it means to prove  $n \in X$ , then it is still not obvious to decide between the predicative and the impredicative viewpoint. Since the matter of predicativity is an issue in its own right, we bypass the topic.

Even at a quite low level sets of natural numbers turn out to be rather elusive. If we consider

$$\{n \mid \text{the } n^{\text{th}} \text{ decimal of } \pi \text{ is preceded by 20 nines}\}$$

then we do not know whether it is empty or not. So, even sets that have simple definitions may be rather wild (although not surprisingly so, as recursion theory has already shown us).

The universe of sets differs in an essential way from the universe of natural numbers. Whereas the latter are discretely given and completely determined with respect to each other, the first are pretty undetermined in the extensional sense, i.e. considered as being determined by their elements. This undeterminedness is brought out in the following *uniformity principle*, formulated by Troelstra

$$UP \quad \forall X \exists x \varphi(X, x) \rightarrow \exists x \forall X \varphi(X, x).$$

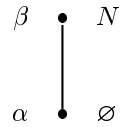
*In words:* if for each set  $X$  one can find a natural number  $x$  such that  $\varphi(X, x)$  then there is already one number  $x_0$  that satisfies  $\varphi(X, x_0)$  for *all*  $X$ . Surprising as this may seem, the almost immediate counter-examples from classical logic are seen not to work. For example, consider  $\forall X (X = \emptyset \vee X \neq \emptyset)$ , which can be written as  $\forall X \exists x ((x = 0 \rightarrow X = \emptyset) \wedge (x \neq 0 \rightarrow X \neq \emptyset))$ . Classically, this statement is true, but intuitionistically is in general not decidable whether a set is empty, cf. the set defined above. The uniformity principle is consistent with **HAS** + AC-NS, where the axiom or choice from number to species reads

$$AC-NS \quad \forall x \exists X \varphi(X, x) \rightarrow \exists Y \forall x \varphi((Y)_x, x)$$

(where  $y \in (Y)_x \Leftrightarrow \langle x, y \rangle \in Y$ ) [Troelstra, 1973a; van Dalen, 1974].

**HAS** has been studied via Kripke semantics in [Jongh and Smoryński, 1976]. They interpreted the first-order part as usual and took for sets of natural numbers growing families of sets (just like unary predicates in an ordinary Kripke model).

EXAMPLE 66.



Take in  $\alpha$  and  $\beta$  the standard model of (classical) arithmetic and let  $S_\alpha = \emptyset, S_\beta = N$ . Then  $\alpha \Vdash \neg \neg \forall x (x \in S)$ , i.e.  $\neg \neg S = N$ , but  $\alpha \not\Vdash \forall x (x \in S)$  and even  $\alpha \not\Vdash \exists x (x \in S)$ .

A number of proof theoretic results are obtained by semantic means, e.g. **HAS** has the disjunction and the existence property, but also the existence property for  $\exists X$ :

$$\mathbf{HAS} \vdash \exists X \varphi(X) \Rightarrow \mathbf{HAS} \vdash \varphi(\{x \mid \sigma(x)\}), \text{ for a suitable } \sigma(x),$$

i.e. if ‘there (provably) exists a set’  $X$ , then ‘there already exists a definable set’. We list a few closure properties.

1. **HAS** is closed under the Uniformity Rule, UR<sup>c</sup>  
 $\mathbf{HAS} \vdash \forall X \exists \varphi(X, x) \Rightarrow \mathbf{HAS} \vdash \exists x \forall X \varphi(X, x)$   
 (where  $FV(\varphi) = \{X, x\}$ ).
2. **HAS** is closed under Markov's Rule, MR  
 $\mathbf{HAS} \vdash \forall x(\varphi(x) \vee \neg\varphi(x)) \wedge \neg\neg\exists x\varphi(x) \Rightarrow \mathbf{HAS} \vdash \exists x\varphi(x)$ .
3. **HAS** is closed under Church's Rule, CR  
 $\mathbf{HAS} \vdash \forall x \exists y \varphi(x, y) \Rightarrow \mathbf{HAS} \vdash \exists e \forall x \varphi(x, \{e\}x)$ .
4. **HAS** is closed under the Rule of Choice, RC-NS  
 $\mathbf{HAS} \vdash \forall x \exists X \varphi(X, x) \Rightarrow \mathbf{HAS} \vdash \exists X \forall x \varphi((X)_x, x)$ .

Intuitionistic second-order arithmetic has been extensively studied by proof-theoretical means, e.g. [Martin-Löf, 1971; Prawitz, 1971; Girard, 1971] and Troelstra [1973; 1973a].

## 10.2 Choice Sequences

Whereas in classical mathematics one can define functions in terms of sets and vice versa, we here treat functions and sets more or less independently. Philosophically speaking this is rather obvious; the two notions are radically different. A set (say of natural numbers) is given to us as a *property* of natural numbers (cf. Brouwer [1918; 1981a]), whereas a function (say from natural numbers to natural numbers) is given as a process of assigning values to arguments. Interdefinability of these notions would be an unexpected coincidence. One can, of course, consider a function as a set of pairs, conversely one cannot, in general, give a set by a characteristic function. For let  $f : \mathbb{N} \rightarrow \{0, 1\}$  and  $n \in A \Leftrightarrow f(n) = 1$ , then it follows from  $\forall m(m = 1 \vee m \neq 1)$  that  $\forall n(n \in A \vee n \notin A)$ , i.e.  $A$  is *decidable* (mind you, not recursive, but decidable in the sense that ‘ $n$  belongs to  $A$  or does not belong to  $A$ ’). Since there is in intuitionistic mathematics an abundance of undecidable sets, we must conclude that the characteristic function approach to sets does not work.

For the sake of perspicuity we will in the following restrict ourselves to functions from  $\mathbb{N}$  to  $\mathbb{N}$ .

The nineteenth century had already brought us the immense progress of widening the function concept, in the form of ‘a function is a law that assigns a natural number to each natural number’, thus doing away with conditions of analyticity, etc. However, the discussions at the beginning of this century made it clear that on a reasonable reading of ‘law’, one would end up with a countable universe of functions, with all its mathematical drawbacks. Or, even worse, with the definability paradoxes (Richard and Berry).

In order to overcome these difficulties Brouwer introduced in 1918 a more liberal notion of function, which identified functions with choice processes. To quote from Brouwer [1981a]: “*Admitting two ways of creating new mathematical entities: firstly in the shape of more or less freely proceeding infinite sequences of mathematical entities previously acquired; . . .*”

These sequences were introduced in 1918 as *choice sequences* (Wahlfolgen). In a later stage Brouwer spoke of *arrows*. One has to think an idealised mathematician who at consecutive stages chooses natural numbers. This (mental) choice process may be highly involved, e.g. the subject may be in the course of the process put all kinds of restrictions on future choices. He may, for example, at a certain stage give up all freedom and follow a given law, or he may decide at the beginning that he will never completely give up his freedom of choice.

The matter of higher-order restrictions on future choices (i.e. restrictions on restrictions, etc.) has sparked some debate. Indeed, Brouwer himself has questioned their usefulness (cf. Brouwer [1981a, p.13]; [1975, p. 511]).

Once choice sequences were introduced, Brouwer was faced with the non-trivial problem of how to exploit them in mathematics. Put otherwise, what properties can one extract from the basic conception of a choice process? In the very first paper on the subject Brouwer laid down the following continuity principles: *A law that assigns to each choice sequence  $\xi$  a natural number  $n$  must completely determine  $n$  after a finite initial segment of  $\xi$  has been determined* (cf. Brouwer [1918, p. 13]; [1975, p. 160]).

The matter of establishing the basic properties of choice sequences calls for ‘informal rigour’ (a term introduced by Kreisel [1967], referring to a precise non-formal analysis of certain conceptually given concepts, leading to more or less basic axioms (principles)). A general analysis of this kind is not within the scope of the present chapter. The reader is referred to [Troelstra, 1977; Dummett, 1977] and [van Atten and van Dalen, forthcoming]. Without going into all details we will indicate a language for a theory of choice sequences (also called *intuitionistic analysis*).

Add to the language of arithmetic, function variables  $\xi_1, \xi_2, \xi_3, \dots$  and suitable function constants (e.g. the primitive recursive functions). The result is a two sorted language. We add all axioms of **IQC** for both sorts. In general one adds the (rather weak) comprehension principle

$$\forall x \exists! y \varphi(x, y) \rightarrow \exists \xi \forall x \varphi(x, \xi(x)).$$

What more is to be added depends on the notion under consideration. The comprehension principle is, e.g., correct for general choice sequences, but not in general for lawlike or lawless sequences. We will use roman symbols  $f, g, h, \dots$  for lawlike sequences (i.e. choice sequences given by a law). One may use various sorts of choice sequences in one and the same context (cf. [Kreisel and Troelstra, 1970; Troelstra, 1977]). We will, however, consider

a particularly perspicuous kind of choice sequence, introduced by Kreisel. The choice sequences we have in mind constitute, so to speak, a limiting case where no restrictions whatsoever will be placed on future choices. The resulting notion is that of ‘*lawless sequence*’. In what way does the lawlessness of a sequence manifest itself? Let us do a thought experiment: we make successive choices  $\xi(0), \xi(1), \xi(2), \dots$ , such that at each stage there is a complete freedom for the next choice (think of the throws of a die, where there is however an overall restriction to numbers  $\leq 6$ ), we add the successive values and we find that the sum of a certain initial segment is a prime number. For example,  $\xi(0) = 4, \xi(1) = 2, \xi(2) = 2, \xi(3) = 0, \xi(4) = 1, \xi(5) = 8, \xi(6) = 2, \dots$ , and  $4 + 2 + 2 + 0 + 1 + 8 = 17$ , which is a prime. So for this  $\xi$  we have established ‘There is an initial segment of  $\xi$  such that its sum is a prime number’—abbreviated by  $\varphi(\xi)$ .

It is, however, immediately clear that any lawless sequence  $\eta$  that starts with the same initial segment  $\langle 4, 2, 2, 0, 1, 8 \rangle$  also satisfies  $\varphi$ . This is an instance of the general principle of *open data*:

$$\varphi(\xi) \rightarrow \exists x \forall \eta (\bar{\xi}x = \bar{\eta}x \rightarrow \varphi(\eta)),$$

where  $\bar{\xi}x = \langle \xi(0), \xi(1), \dots, \xi(x-1) \rangle$ , i.e. the coded (cf. Algorithms Chapter, volume 1 of this *Handbook*) initial segment of length  $x$  of  $\xi$ . *In words*: if  $\varphi$  holds for the lawless sequence  $\xi$  then there is an initial segment of  $\xi$  such that all lawless continuations of it also satisfy  $\varphi$ . Or less precise,  $\xi$  having the property  $\varphi$  is determined by a suitable initial fragment.

The principle can be justified as follows: the idealized mathematician establishes  $\varphi(\xi)$  after a finite number of values of  $\xi$  has been chosen, because at any time that is all the available information on  $\xi$  he has. but, therefore, the continuation of this particular initial segment is irrelevant, i.e. all continuations  $\eta$  also have the property  $\varphi$ .

It is quite often helpful to think of a choice sequence (function) as a path in the full tree of all finite sequences of natural numbers. So suppose  $\varphi(\xi)$  holds on the basis of the information of segment  $\langle 0, 2, 3, 0, 1 \rangle$ , then  $\varphi$  holds for all lawless sequences (paths) that pass through the node  $\langle 0, 2, 3, 0, 1 \rangle$  in the tree. but in the tree topology this is a (basic) open. Hence if  $\varphi$  holds for  $\xi$ , it holds for an open neighbourhood of  $\xi$  (this explains the name ‘open data’).

There are two more basic principles:

$$\text{LS1} \quad \forall x \exists \xi (\xi \in x),$$

where  $x$  is a coded initial segment and  $\xi \in x$  stands for  $\xi(0) = (x)_0, \dots, \xi(k-1) = (x)_{k-1}$ , where  $k$  is the length of  $x$ . In words each initial segment can be extended into a lawless sequence.

This principle is harder to justify, if no restrictions at all are allowed, how does one make certain that the first  $k$  choices can be made to conform to

a given segment? It is best to view this principle as slightly modifying the original notion: we allow at the beginning the specification of an arbitrary finite initial segment. In this way all finite sequences actually occur as initial segments.

If we make no assumptions about initial segments then the sequences offer a less satisfactory mathematical theory. Troelstra [1983] introduced this weaker notion (without LS1) under the name of *proto-lawless* sequences.

Finally, if the idealized mathematician considers two (mental) lawless choice processes, then he knows if the processes are identical or not, so we have

$$\text{LS2 } \forall \xi \eta (\xi \equiv \eta \vee \neg \xi \equiv \eta),$$

where  $\equiv$  is the intensional identity between sequences, considered as mental choice processes.

The principle of open data is formulated as

$$\text{LS3 } \varphi(\xi) \rightarrow \exists x (\xi \in x \wedge \forall \eta \in x \varphi(\eta)).$$

Actually, Kreisel's notion of 'lawlessness' requires also a certain independence of sequences, so that the sequences are also lawless with respect to each other. An example: say we generate a lawless sequence  $\xi(0), \xi(1), \xi(3), \dots$  and we drop the first value, is the remaining sequence  $\xi(1), \xi(2), \xi(3), \dots$  lawless? Individually viewed, yes, but in conjunction with the original sequence, no.

This leads us to extend LS3 as follows:

$$\begin{aligned} \text{LS3}_n \quad & \varphi(\xi, \xi_1, \dots, \xi_n) \wedge \not\equiv (\xi, \xi_1, \dots, \xi_n) \\ & \exists x (\xi \in x \wedge \forall \eta \in x (\not\equiv (\eta, \xi_1, \dots, \xi_n) \rightarrow \varphi(\eta, \xi_1, \dots, \xi_n))), \end{aligned}$$

where  $\rightarrow \equiv (\xi, \xi_1, \dots, \xi_n)$  stands for  $\bigwedge_{i=1}^n \xi \not\equiv \xi_i$ .

Without the extra clause  $\not\equiv (\xi, \xi_1, \dots, \xi_n)$  the principle is false. For example, consider  $\varphi(\xi, \xi) := \xi \equiv \xi$ . Suppose we could apply LS3<sub>1</sub>, then

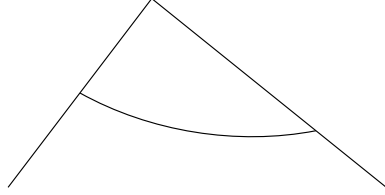
$$\xi \equiv \xi \rightarrow \exists x (\xi \in x \wedge \forall \eta \in x (\xi \equiv \eta)),$$

i.e. there is an initial segment of  $\xi$  such that every extension of it coincides with  $\xi$ . This plainly contradicts LS1.

From these principles we can already derive a number of unusual results.

**THEOREM 67.**

1.  $\xi \equiv \eta \leftrightarrow \forall x (\xi x = \eta x)$
2.  $\forall \xi \neg \neg \exists x (\xi x = 0)$
3.  $\forall \xi \neg \exists \eta \forall x (\xi(x+1) = \eta(x))$



4.  $\neg\exists\xi(\xi = f)$ , for a lawlike function  $f$ .

**Proof.** (a)  $\rightarrow$  is trivial. For  $\leftarrow$  we can use a proof by contradiction by LS2. So let  $\xi \not\equiv \eta$ , we may apply LS3<sub>1</sub> to  $\varphi(\xi, \eta) := \forall y(\xi(y) = \eta(y))$ :

$$\exists x(\xi \in x \wedge \forall \zeta \in x(\zeta \not\equiv \eta \rightarrow \forall y(\zeta(y) = \eta(y))).$$

Let, therefore, the initial segment  $\langle \xi(0), \dots, \xi(k) \rangle$  be such that for  $\zeta \not\equiv \eta$  and  $\zeta(0) = \xi(0), \dots, \zeta(k) = \xi(k)$  it follows that  $\forall y(\zeta(y) = \eta(y))$ . Since  $\zeta(k+1)$  can be chosen freely, we choose  $\zeta(k+1) = \eta(k+1) + 1$  (i.e. we apply LS2 to  $\langle \xi(0), \dots, \xi(k), \eta(k+1) + 1 \rangle$  to obtain a  $\zeta$ ). But this contradicts  $\forall y(\zeta(y) = \eta(y))$ . Hence  $\xi \equiv \eta$ .

(b) Suppose  $\neg\exists x(\xi x = 0)$ , or by logic,  $\forall x(\xi x \neq 0)$ . Apply LS3 and add a zero to the initial segment of  $\xi$  that exists according to LS3.

(c) Apply LS3<sub>1</sub> to  $\forall x(\xi(x+1) = \eta(x))$ .

(d)  $\xi = f$  is an abbreviation for  $\forall x(\xi(x) = f(x))$  (*extensional equality*). Apply LS3 to the latter formula. ■

Could we do better than (b) and even show  $\forall \xi \exists x(\xi(x) = 0)$ ? The answer is no, but we need a strengthening of the system to show this.

We have already defined what a bar is (cf. p. 25). Now we can formalise it in analysis:  $B$  is a bar (in the tree of all finite sequences) if  $\forall \xi \exists x(\xi \bar{x} \in B)$ . Such a bar is a denumerable set of sequences. Can we present such a bar by a convenient function? The technique is not difficult, we consider a function  $e : \mathbb{N} \rightarrow \mathbb{N}$ , such that  $e(x) = 0$  if  $x$  is a (coded) sequence above the bar, and  $e(x) > 0$  for  $x$  on or below the bar.

More formally:

$$e \in K_0 := \forall \xi \exists x(e(\bar{\xi}x) > 0) \wedge \forall xy(e(x) > 0 \wedge y \geq x \rightarrow e(x) = e(y))$$

(where  $y \geq x$  stands for ‘the sequence  $y$  extends  $x$ ’).  $K_0$  is the class of neighbourhood functions (or moduli of continuity, also called *Brouwer operations*). Note that the part of the tree above  $B$  is a well-founded tree. Kreisel and Troelstra have considered  $K_0$  as an inductively defined class of lawlike functions (cf. [Troelstra and van Dalen, 1988, p. 223 ff.]).

The neighbourhood functions have been introduced with respect to lawless sequences, i.e. if  $e$  gives a bar  $B$  then each lawless sequence  $\xi$  hits the

bar. It is however plausible to widen the scope of such  $e$ , such that *each* sequence (not necessarily lawless) hits the bar. This *extension principle* which states that for each  $e \in K_0$  we have  $\forall \alpha \exists x (e(\bar{\alpha}x) \neq 0)$  (where  $\alpha$  ranges over *all* sequences), is required for certain applications of the theory. A justification of the extension principle by means of an abstraction operator is put forward by Troelstra [1977, p. 20].

Here we will for simplicity use  $K_0$  as given above. We will now formulate a strong continuity principle: if  $\forall \xi \exists x \varphi(\xi, x)$ , then  $x$  can be found for a given  $\xi$  by a neighbourhood function  $e$  from  $K_0$  as follows: look for the first initial segment  $\xi y$  such that  $e(\xi y) = k + 1 > 0$  and put  $x = k$ . Let us agree to write  $e(\xi) = k$  when we follow the above procedure, then we get the following principle:

$$\text{LS4} \quad \forall \xi \exists x \varphi(\xi, x) \rightarrow \exists e \in K_0 \forall \xi \varphi(\xi, e(\xi)).$$

By the extension principle  $e$  operates on all possible sequences. We will use this to show  $\neg \forall \xi \exists x (\xi(x) = 0)$ . Suppose  $\forall \xi \exists x (\xi(x) = 0)$ , then by LS4  $\forall \xi (\xi(e(\xi)) = 0)$ , i.e.  $e$  picks a zero of  $\xi$ . Now consider  $f$  such that  $\forall x (f(x) = 1)$ . Determine  $e(f)$ , say  $k$ .  $f$  ‘hits’ the bar determined by  $e$  in a node  $m$ , which hence is an initial segment of  $f$ . Now extend this segment  $m$  with enough 1’s such that the total length of the resulting  $n$  exceeds  $k$ . By LS1 there is a lawless sequence  $\xi$  with initial segment  $n$ . By definition, however,  $e(\xi) = e(f)$  and  $0 = \xi(e(\xi)) = f(e(f)) = 1$ . Contradiction. Hence  $\neg \forall \xi \exists x (\xi(x) = 0)$ .

The above lines may serve to illustrate the highly unusual character of lawless sequences and the extraordinary richness of the intuitionistic universe of functions. Kreisel and Troelstra have established for a certain system elimination theorems, i.e. translations that eliminate the choice sequences (cf. [Kreisel and Troelstra, 1970]; [Troelstra and van Dalen, 1988, 12.3.1]). This may, with due caution, be viewed as evidence for the viewpoint that choice sequences are only a *façon de parler*.

Note however that the evidence is rather incomplete in the sense that the theorems range over a few formal theories. Moreover such a viewpoint would violently conflict with the ontological status of the mentally generated objects of intuitionism.

Choice sequence have the didactic disadvantage that one cannot show an isolated copy, unless it happens to be given by a law. This situation changed when Joan Moschovakis adapted a topological model of Scott for the reals to choice sequences. A similar interpretation was presented by the author in the framework of Beth models. Since the latter approach allows one a nice visualisation we will sketch it here (cf. [van Dalen, 1978]).

In order to facilitate the presentation, we consider models with the universal tree (the tree of all finite sequences of natural numbers) as underlying poset. We will also denote the finite sequences  $(n_0, \dots, n_{k-1})$  by  $\vec{n}$ .



Whereas in a Kripke model the condition  $\forall x \exists! y \xi(x)y$  forces us to interpret a sequence (function) in each node as a total function, in a Beth model  $\vec{n} \Vdash \forall x \exists! y \xi(x) = y$  only tells us that eventually on a bar  $B$ , the outputs  $y$  for an input  $x$  will be determined, so the natural interpretation of a sequence  $\xi$  is a growing family  $\xi^{\vec{n}}$  of partial functions with the property that along a path the union of all these  $\xi^{\vec{n}}$ 's yield a total function. A concept that confirms rather well to the heuristic notion of 'choices being made in time'. In the nodes the choice function is only partially determined, but the whole model allows us to view the choice sequences, as it were from a higher viewpoint, as completed.

So we take a Beth model of arithmetic (containing only standard numbers) and consider as the universe of choice sequences *all* such growing families of partial functions.

EXAMPLES

(1) Define  $\xi^{\vec{n}} := \vec{n}$  for each  $\vec{n}$ , i.e. in each finite sequence the partial function is just this sequence. Evidently the conditions are satisfied.

(2) Define  $\xi^{(\cdot)} = (\cdot)$  and  $\xi^{(i)} = \lambda x. i$ , i.e. at the bottom node we take the empty function, and at its immediate successors  $\langle i \rangle$  we take the constant functions with value  $i$ . Observe that  $\langle \cdot \rangle \Vdash \exists x \forall y (\xi(y) = x)$  (a simple exercise in Beth semantics), so the model tells us that the sequence is constant, although externally it is *not*.

The particular model with underlying tree of all finite sequences of natural numbers validates a list of principles:

$$\text{AC - NF} \quad \forall x \exists \xi \varphi(x, \xi) \rightarrow \exists \eta \forall x \varphi(x, (\eta)_x),$$

where  $(\eta)_x(y) = \eta(\langle x, y \rangle)$ , *the axiom of choice from numbers to functions*.

$$\text{SC!} \quad \forall \xi \exists! x \varphi(x, \xi) \rightarrow \exists \tau \in K_0 \forall \xi \varphi(\tau(\xi), \xi),$$

*the strong continuity principle with uniqueness restriction*.

$\text{BI}_M$  *monotone bar induction*, a principle that can be considered as an intuitionistic version of induction over well-founded relations, or of transfinite induction (cf. [Troelstra and van Dalen, 1988; Kleene and Vesley, 1965]).

$$\text{KS} \quad \exists \xi (\varphi \leftrightarrow \exists x \xi(x) \neq 0),$$

*Kripke's Schema*, a principle which will be discussed in the next section.

The validity of Kripke's Schema is fairly simple to establish. For the remaining principles we refer to [van Dalen, 1978]. The idea is to go up in the tree and in each node to check if  $\varphi$  has been forced. So we define  $\xi^{\vec{n}}$  to be a finite sequence of the same length as  $\vec{n}$ . Let  $\vec{n}$  have length  $k$ , then we put  $\xi^{\vec{n}}(k) = 0$  if  $\vec{n} \not\Vdash \varphi$  and  $\xi^{\vec{n}}(k) = 1$  if  $\vec{n} \Vdash \varphi$ . this defines a proper choice sequence. Clearly, for any  $\vec{n}$ , if  $\vec{n} \Vdash \varphi$  then  $\vec{n} \Vdash \exists x \xi(x) \neq 0$ .

Conversely, suppose that  $\vec{n} \Vdash \exists x \xi(x) \neq 0$ , then there is a bar  $B$  for  $\vec{n}$  such for each  $\vec{m} \in B, \vec{m} \Vdash \xi(n) \neq 0$  for some  $n$ . This implies  $\xi^{\vec{m}}(n) = 1$ , which by definition means that  $\vec{m} \Vdash \varphi$ . So  $\varphi$  is forced on a bar for  $\vec{n}$ . Hence  $\vec{n} \Vdash \varphi$  (cf. Lemma 11). Therefore  $\langle \cdot \rangle \Vdash \exists \xi(\varphi \leftrightarrow \exists x \xi(x) \neq 0)$ .

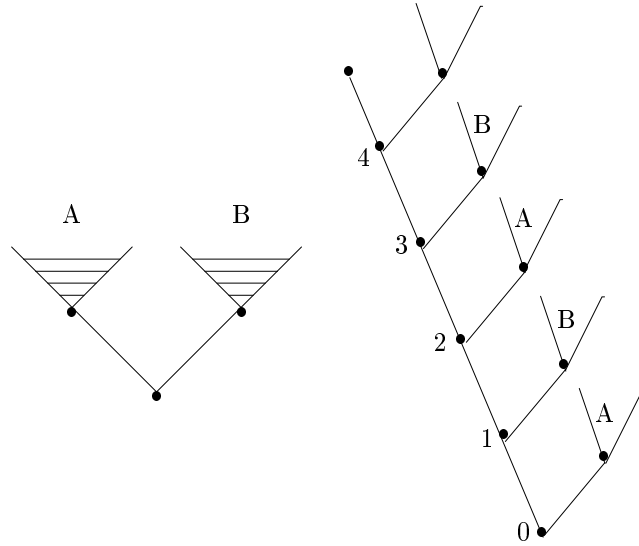
Although Kripke semantics has shown itself superior to Beth semantics in many respects, the latter is the more convenient one to treat functions. For, in the Kripke model, a function must in each world be interpreted by a total function (just evaluate  $\alpha \Vdash \forall x \exists y (f(x) = y)$ ); in a Beth model however one can ‘postpone’ the assignment of outputs to inputs, and this allows for a particularly simple model for analysis. In order to use Kripke models for dealing with analysis one has to exploit the expanding of the domains, and this calls, in the case of arithmetic, immediately for non-standard numbers. Hardly natural!

The above model for analysis has the drawback that its first-order theory is classical, i.e. each classically true sentence of first-order arithmetic holds in the model. Therefore the model cannot play the role of ‘standard model’ of analysis. The model shows, however, that it is consistent to put an intuitionistic second-order theory on top of a classical first-order theory. The problem of the ‘standard model’ occurs already for first order arithmetic. In **HA** one can show not only  $\forall xy(x = y \vee x \neq y)$ , but also  $m = n \Leftrightarrow \mathbf{HA} \vdash \vec{m} = \vec{n}$ . So in a topological model for arithmetic, with only standard numbers we have  $\llbracket t = s \rrbracket = X$  or  $\emptyset$ , i.e. atoms take only the values  $\top$  or  $\perp$ . Now a simple induction shows that  $\llbracket \varphi \rrbracket$  takes the values  $\top$  or  $\perp$  for all sentences  $\varphi$ . So we get full true, classical arithmetic. Therefore, in order to obtain intuitionistic features, say the failure of the principle of the excluded middle, we have to assume the presence of non-standard numbers. This all points towards serious limitations of the present semantic treatment of intuitionistic theories.

### 10.3 The Disjunction Property for Analysis

We have seen that Kripke models may be ‘put together’ as a means for proving metalogical results, e.g. the disjunction property (cf. p. 44). In view of the usefulness of Beth models for interpreting analysis, it would be convenient to have a similar operation in Beth semantics.

Roughly speaking, one takes the disjoint union of two Kripke models and adds one bottom node (left-hand figure). The domain of the bottom node is contained in all domains of the Kripke models **A** and **B**. In the case of Beth models one would place the models **A** and **B** alternately on top of the linearly ordered set of natural numbers (right-hand figure). Here we run into difficulties; what should the domain be in the nodes  $0, 1, 2, \dots$ ? The semantics does not allow for non-constant domains, so we reach a dead-end. Now our generalised semantics comes in handy, if we allow expanding domains then we can use this ‘gluing’ technique. Let us outline how to



obtain DP.

Suppose  $T \vdash \varphi \vee \psi$  and  $T \not\vdash \varphi, T \not\vdash \psi$ , and  $\mathbf{B} \not\vdash \psi$ . Find a domain that is contained in the bottom nodes of  $\mathbf{A}$  and  $\mathbf{B}$  (this is not always possible, it depends on the class of models of  $T!$ ), and place this in the nodes  $0, 1, 2, \dots$  of the co-called *spine*.

Now  $0 \Vdash \varphi \vee \psi$ , so there is a bar  $B$  such that for each  $\beta \in B, \beta \Vdash \varphi$  or  $\beta \Vdash \psi$ . The bar intersects the spine, say in  $n$ . If  $n \Vdash \varphi$  then there is a copy of  $\mathbf{A}$  above  $n$  in which  $\varphi$  is forced. Contradiction. Similarly for  $n \Vdash \psi$ . Hence  $T \vdash \varphi$  or  $T \vdash \psi$ .

The above gluing construction is particularly fruitful for analysis, (cf. Dalen [1984; 1986]). It yields simple proofs of the disjunction and existence property for various systems of analysis. The main problem is the definition of the universe of sequences in the resulting model. One assigns to the nodes  $n$  on the spine, the set of finite sequences of length  $n$ . The sequences (in the sense of the model) in node  $n$  are then all possible extensions of these sequences to partial functions in higher nodes.

J. Moschovakis had already established DP and EP for some systems by proof theoretical means in [Moschovakis, 1967].

The reason for dwelling on arithmetic and its extensions, in particular analysis, is that this hard core of mathematical logic is the ultimate testing ground for logical methods. Analysis is extremely important because it illustrates the typical consequences of intuitionism. In analysis one can most clearly see the conflict between the classical and the intuitionistic approaches. Brouwer used the following theorem to illustrate the typical

properties of intuitionism:

*Each function from the closed interval  $[0,1]$  to  $\mathbb{R}$  is uniformly continuous [1923].*

From the intuitionistic principles we can derive similar results. The continuity principle SC! tells us that each function from sequences to numbers is continuous. however, classically one can define the function  $F$  such that

$$F(\xi) = \begin{cases} 0 & \text{if } \xi \text{ contains a } 0 \\ 1 & \text{otherwise.} \end{cases}$$

This  $F$  cannot be continuous, because continuity would mean that the occurrence of a 0 in a sequence  $\xi$  can be predicted on the basis of an initial segment of  $\xi$ . *Quod non.*

The conflict with classical mathematics is here particularly striking. In general, analysis and higher-order systems are the perfect grounds for demonstrating the proper character of intuitionism; it distinguishes itself from narrow constructivism and finitism by its embracing abstract notions.

A topic that has been omitted altogether is *second-order propositional logic*. Whereas in classical logic this is a subject of great dullness (think of quantifications over a set of two truth-values), it is not so in the intuitionistic version. In contrast to the classical system, the intuitionistic one is undecidable (cf. [Gabbay, 1981] for a systematic treatment of this topic).

#### 10.4 Remarks on the Axiom of Choice

In classical mathematics the axiom of choice used to be considered as something that, if it should hold at all, should hold globally (this is not to say that no refinements of AC have been considered, but that the foundational evidence seems to point that way (cf. [Shoenfield, 1967, p. 253]). In intuitionism there is fairly solid evidence for the validity of the principle of countable choice; let  $\forall x \exists y \varphi(x, y)$  be given for, say  $x$  and  $y$  ranging over natural numbers, then we have a proof of  $\forall x \exists y \varphi(x, y)$ , i.e. a construction that provides for each  $x \in \mathbb{N}$  a proof of  $\exists y \varphi(x, y)$ . this, in turn, means that we have a construction that yields a  $y$  and a proof of  $\varphi(x, y)$ . So we have a construction that for each  $x$  yields a  $y$  such that  $\varphi(x, y)$  holds (i.e. has a proof). This construction provides a choice function  $f$ , such that  $\forall x \varphi(x, f(x))$  holds. So AC-NN is valid on an intuitionistic interpretation of the logic. (In Martin-Löf's type theory is is actually provable.)

Let us now consider AC-RN. The following is true: for each real number  $x$  there is a natural number  $y$  such that  $x < y$  (recall that  $x$  is given by a Cauchy sequence). If there were a choice function  $f$  such that  $x < f(x)$ , then—by Brouwer's theorem (cf. [Heyting, 1956, p. 46], [Brouwer, 1981a, p. 80])— $f$  has to be continuous. But a continuous function from  $\mathbb{R}$  to  $\mathbb{N}$  is constant. Contradiction.

An inspection of the proof, as given for AC-NN, will show what has gone wrong.

Our assumption reads in full:  $\forall x \in \mathbb{R}, \exists y \in \mathbb{N}(x < y)$ , so the proof interpretation tells us that for each choice sequence  $x$  of rational numbers  $\pi(x)$  is a proof of ‘ $x$  is a Cauchy sequence  $\rightarrow \exists y \in \mathbb{N}(x < y)$ ’, i.e.  $\pi(x)$  applied to a proof of ‘ $x$  is a Cauchy sequence’ yields a proof of

$$\exists y \in \mathbb{N}(x < y).$$

Now, finishing the argument, we find a choice function that depends on  $x$  and on a proof that  $x$  is a Cauchy sequence. But now we see that  $f$  is not extensional, i.e.  $x_1 = x_2 \rightarrow f(x_1, \pi_1) = f(x_2, \pi_2)$  fails, where  $\pi_i$  is a proof that  $x_i$  is a Cauchy sequence. Therefore the continuity theorem was not applicable. The moral of this digression is that one has to spell out the assumption of AC in full. In case of AC-NN we are on safe ground because a natural number by virtue of its mode of generation carries its own proof that it is a natural number.

The general axiom of choice is intuitionistically out of the question, as Diaconescu (cf. [Goldblatt, 1979]) has shown that it implies the excluded third. The following simple argument, due to Goodman and Myhill, proves Diaconescu’s result.

Let  $\varphi$  be any statement. Form the sets

$$\begin{aligned} A &:= \{n \in \mathbb{N} \mid n = 0 \vee (n = 1 \wedge \varphi)\}, \\ B &:= \{n \in \mathbb{N} \mid n = 1 \vee (n = 0 \wedge \varphi)\}. \end{aligned}$$

We have  $\forall X \in \{A, B\} \exists y \in \mathbb{N}(y, X)$ .

AC would supply us with a function  $f$  such that  $\forall X \in \{A, B\} (f(X) \in X)$ . Since  $f(X)$  is a natural number, we get  $f(A) = f(B) \vee f(A) \neq f(B)$ . If  $f(A) = f(B)$ , then  $\varphi$  holds, and if  $f(A) \neq f(B)$ , then  $\neg\varphi$  holds. For suppose  $\varphi$ , then  $A = B$  (extensionally) so  $f(A) = f(B)$ . Contradiction. So the validity of AC for this particularly simple case implies  $\varphi \vee \neg\varphi$ .

## 11 THE CREATING SUBJECT

In Brouwer’s writings some explicit references to the agent of mathematical activity occur (1948) (cf. [Brouwer, 1975, p. 478]). The basic ideas were already present in his lectures in the late 1920s, cf. [Brouwer, 1992]; the publication was, however, postponed until after the second world war. In due time this practice has become known under the name ‘theory of the creating subject’. Brouwer introduced the creating subject for the purpose of establishing some stronger results in the area of so-called negative predicates. In particular he showed that inequality on the reals is strictly weaker than apartness:  $\neg\forall xy(x \neq y \rightarrow x \# y)$ .

Kreisel, Kripke and others have analyzed the principles involved in those proofs (cf. [Kreisel, 1967]).

The creating subject is assumed to operate in linear time of order-type  $\omega$ . It ‘experiences the truth’ of statements  $\varphi$  at stages  $0, 1, 2, \dots$ . The exact nature of ‘experiencing the truth’ is, of course, left open. One may think of ‘proving’, ‘observing’ or ‘knowing’, etc. By suitably idealising the creating subject we may assume that:

1. it retains truths that have been experienced;
2. at each stage it knows  $\varphi$  or it does not know  $\varphi$ . that is, ‘knowing  $\varphi$  at stage  $n$ ’ is *decidable*;
3.  $\varphi$  holds if it has been ‘experienced’, by the creating subject.

This is in perfect accordance with the intuitionistic dogma that mathematics has its seat in the human mind, and that the only way to establish something is to have a mental ‘proof’ or ‘experience’ for it.

The converse can be defended under the purely solipsistic view that what is the case solely depends on mental experience of the (unique) creating subject. Then, if  $\varphi$  holds it follows that the creating subject has come to know it at some stage. If one allows for an intersubjective viewpoint, then the matter is less clear. The statement  $\varphi$  may hold without the creating subject (one of many) having established it. In this case it seems plausible that it is impossible that the creating subject will ever experience  $\varphi$ .

The theory of the creating subject has been formalized by Kreisel [1967], in a theory containing at least (a fragment of) arithmetic, and a tensed modal operator  $\Box_x$ , to be read as ‘the creating subject knows (has evidence, a proof for, . . . etc.) at time  $x$ ’.

The principles under (1), (2) and (3) can now be formulated as

1.  $\Box_x \varphi \rightarrow \Box_{x+y} \varphi$
2.  $\Box_x \varphi \vee \neg \Box_x \varphi$
3.  $\varphi \leftrightarrow \exists x \Box_x \varphi$ .

In the intersubjective case (3) splits into the following parts

$$\Box_x \varphi \rightarrow \varphi \text{ and } \varphi \rightarrow \neg \neg \exists x \Box_x \varphi.$$

For the applications that Brouwer had in mind the weaker reading suffices. The justification for the solipsistic version, however, is more convincing. In principle there is no objection to iterate the operator  $\Box$ , and in Brouwer’s consistent view that reflection on one’s own mental activity is possible, or even necessary it seems quite correct to do so. However, all problems that arise in and around predicativity, reappear here as well.

In the following pages we will look at the full theory as given by (1), (2) and (3).

### 11.1 Kripke's Schema

If we have function variables available, then we can eliminate the modal operator and retain all its benefits by keeping track of the knowledge of the creating subject by means of a function that registrates if the creating subject knows  $\varphi$  at stage  $x$ .

Define

$$\xi(x) = \begin{cases} 0 & \text{if } \neg \Box_x \varphi \\ 1 & \text{if } \Box_x \varphi, \end{cases}$$

then by (3)  $\varphi \rightarrow \exists x \Box_x \varphi$ , so  $\varphi \rightarrow \exists x \xi(x) \neq 0$ . conversely  $\exists x \xi(x) \neq 0 \rightarrow \exists x \Box_x \varphi$  and hence  $\exists x \xi(x) \neq 0 \rightarrow \varphi$ . This proves *Kripke's Schema*.

KS

$$\exists \xi (\varphi \leftrightarrow \exists x \xi(x) \neq 0).$$

By a similar argument one obtains the weak Kripke's Schema in the inter-subjective case

KS<sup>-</sup>

$$\exists \xi ((\exists x \xi(x) \neq 0 \rightarrow \varphi) \wedge (\neg \varphi \rightarrow \forall x \xi(x) = 0)).$$

Kripke's Schema is used, for instance, for the construction of certain strong counterexamples (cf. [Hull, 1969]). We list some of them:

The statements refer to the intuitionistic reals.

1.  $\neg \forall xy ((\neg x < y \wedge x \neq y) \rightarrow y < x)$
2.  $\neg \forall xy (x \neq y \rightarrow x \# y)$
3.  $\neg \forall xy (x \neq y \rightarrow \neg x < y \vee \neg y < x)$
4. not every bounded set without points of accumulation is bounded in number (refutation of the Bolzano–Weierstrauss theorem).

Note that these results are *strong* in comparison to the older results which only yielded 'we cannot prove that ...'. The first of these strong counterexamples was presented by Brouwer in 1949.

Myhill has shown that KS is inconsistent with the continuity principle for functions, which is a generalisation of SC:

$$\forall \xi \exists \eta \varphi(\xi, \eta) \rightarrow \exists F \forall \xi \varphi(\xi, F(\xi)),$$

where  $F$  is a continuous function (cf. [Troelstra, 1969]). KS is however consistent with SC (Kroll (cf. [Grayson, 1981; Scowcroft, 1999])).

### 11.2 Kripke's Schema and the Continuum

Brouwer's strong refutations already show that the creating subject refines our insight into the structure of the continuum. Further results in this area have been obtained in [van Dalen 1999A]. Brouwer had already shown that the continuum is *indecomposable*, in the sense that if  $\mathbb{R} = A \cup B$  and  $A \cap B = \emptyset$ , then  $A = \mathbb{R}$  or  $A = \emptyset$ . On the basis of KS it can be shown that all negative, dense subsets of  $\mathbb{R}$  are likewise indecomposable (where  $X \subseteq \mathbb{R}$  is negative if  $\neg\neg x \in X \rightarrow x \in X$ ). Hence, e.g., the irrationals and the not-not-rationals are indecomposable. These subsets are therefore connected in the topological sense, and they have dimension 1. This is in sharp contrast to the classical theory, where the irrationals are zero-dimensional.

Kripke's schema also allows us to show a kind of converse to Brouwer's indecomposability theorem:  $\text{KS} + \mathbb{R}$  is indecomposable  $\Rightarrow$  there are no discontinuous functions on  $\mathbb{R}$ . We will, by way of illustration, sketch the proof: let  $f : \mathbb{R} \rightarrow \mathbb{R}$  be discontinuous. It is no restriction to assume that  $f(0) = 0$  and that  $f$  is discontinuous in 0. So there is a  $k$  and there are  $x_n$  such that  $|x_n| < 2^{-n}$  and  $|f(x_k)| > 2^{-k}$ . Now consider the statement  $r \in \mathbb{Q}$  for an  $r \in \mathbb{R}$ . We apply KS to  $r \in \mathbb{Q} \vee r \notin \mathbb{Q}$ :

$\exists \xi (\exists x \xi(x) \neq 0 \leftrightarrow r \in \mathbb{Q} \vee r \notin \mathbb{Q})$ . For convenience we assume that  $\xi$  is positive at most once, with value 1.

Define  $a_n = \begin{cases} x_n & \text{if } \forall k \leq n (\xi(k) = 0) \\ x_p & \text{if } p \leq n \text{ and } \xi(p) = 1 \end{cases}$

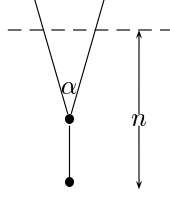
Clearly  $(a_n)$  converges, say  $\lim a_n = a$ . Now  $|f(a_n)| < 2^{-k}$  or  $|f(a_n)| > 0$ , hence  $a \neq x_n$  for all  $n$ , or  $a \neq 0$ . The first is impossible, since it would imply  $\neg(r \in \mathbb{Q} \vee r \notin \mathbb{Q})$ . The latter implies  $\exists n \xi(n) = 1$ , and hence  $r \in \mathbb{Q} \vee r \notin \mathbb{Q}$ . Since this holds for arbitrary  $r$ , we have got a decomposition of  $\mathbb{R}$ . Contradiction. Therefore there are no discontinuous functions on  $\mathbb{R}$ .

Although the theory of the creating subject has a richer language it is actually conservative over the theory with Kripke's Schema ([van Dalen, 1978]).

### 11.3 The Interpretation of the Creating Subject

We have already seen how to validate KS in the Beth model for analysis. A slight adaptation will provide an interpretation of the tensed modal 'knowledge' operator. We define  $\vec{n} \Vdash \Box_k \varphi$  if for all  $\vec{m}$  on the bar for  $\vec{n}$  of nodes of length  $k$ ,  $\vec{m} \Vdash \varphi$  (note that this bar may be below  $\vec{n}$ ). We'll check the axioms (2) and (3).





(2) For a given  $\varphi$  we have for each  $\vec{m}$  of length  $k$ ,  $\vec{m} \Vdash \varphi$  or  $\vec{m} \not\Vdash \varphi$ . For such a  $\vec{m}$  we can conclude  $\forall \vec{p} \geq \vec{m}, \vec{p} \not\Vdash \Box_n \varphi$  from  $\vec{m} \not\Vdash \varphi$ , i.e.  $\vec{m} \not\Vdash \varphi \Rightarrow \vec{m} \Vdash \neg \Box_k \varphi$ . So for each  $\vec{m}$  on the bar of nodes of length  $k$  we have  $\vec{m} \Vdash \Box_k \varphi$  or  $\vec{m} \Vdash \neg \Box_k \varphi$ . Hence  $\langle \rangle \Vdash \Box_k \varphi \vee \neg \Box_k \varphi$ .

(3) If  $\vec{n} \Vdash \varphi$  and with  $\text{lth}(\vec{n}) = k$ , then  $\vec{n} \Vdash \Box_k \varphi$  and hence  $\vec{n} \Vdash \exists x \Box_x \varphi$ . Conversely, if  $\vec{n} \Vdash \exists x \Box_x \varphi$  then there is bar  $B$  for  $\vec{n}$  such that for each  $\vec{m} \in B$  there is an  $k(\vec{m})$  with  $\vec{m} \Vdash \Box_k \varphi$ . Applying the above definition and Lemma 11 we conclude  $\vec{m} \Vdash \varphi$  for each  $\vec{m} \in B$ .

Applying Lemma 11 once more we get  $\vec{n} \Vdash \varphi$ .

#### 11.4 Kripke's Schema and a Representation of Sets of Natural Numbers

Although we cannot use characteristic functions to represent sets, we can use Kripke's Schema to obtain a substitute.

Let  $X$  be a set of natural numbers, then by KS

$$\forall x \exists \xi [x \in X \leftrightarrow \exists y (\xi y = 0)].$$

(switching = and  $\neq$  is a harmless act). Applying the axiom of choice from numbers to functions, AC-NF, we get

$$\exists \eta \forall x [x \in X \leftrightarrow \exists y (\eta \langle x, y \rangle = 0)].$$

So each set  $X$  can be represented by a sequence. This allows for a translation of second-order arithmetic into analysis with KS.

Using this representation, there is a simple argument that deduces the Uniformity Principle,

$$\forall X \exists x \varphi(X, x) \rightarrow \exists x \forall X \varphi(X, x),$$

from the Weak Continuity Principle,

$$\forall \xi \exists x \varphi(\xi, x) \rightarrow \forall \xi \exists xy \forall \eta (\xi y = \bar{\eta} y \rightarrow \varphi(\eta, x)),$$

in the presence of KS, cf. [van Dalen, 1977].

The theory of the creating subject has remained controversial until this day. The introduction of an element of subjectivity runs counter to the tradition of the exact sciences. It is, however, an unavoidable step in representing some of Brouwer's arguments.

Evidently the theory, as it stands, is far from complete. Questions concerning the number of conclusions per step, or regarding disjunction (e.g. is  $\Box_n(\varphi \vee \psi) \rightarrow \Box_n\varphi \vee \Box_n\psi$  a valid principle?), remain unsettled, and evidence seems to be scarce. Dummett has investigated the subject [Dummett, 1977] and Posy has applied the theory in a case study of Brouwer's paper on virtual order [Posy, 1980], cf. also [Posy, 1976].

## 12 THE LOGIC OF EXISTENCE

For the practice of classical logic and mathematics it suffices to consider only total operations. For example, consider the inverse-operation,  $a^{-1}$ . For real numbers  $a^{-1}$  exists if  $a \neq 0$ , so we cannot apply the classical trick of defining  $a^{-1} := 0$  for the remaining  $a$ 's. This should leave us with a partial function, since  $\neq$  is not a decidable relation on  $\mathbb{R}$ .

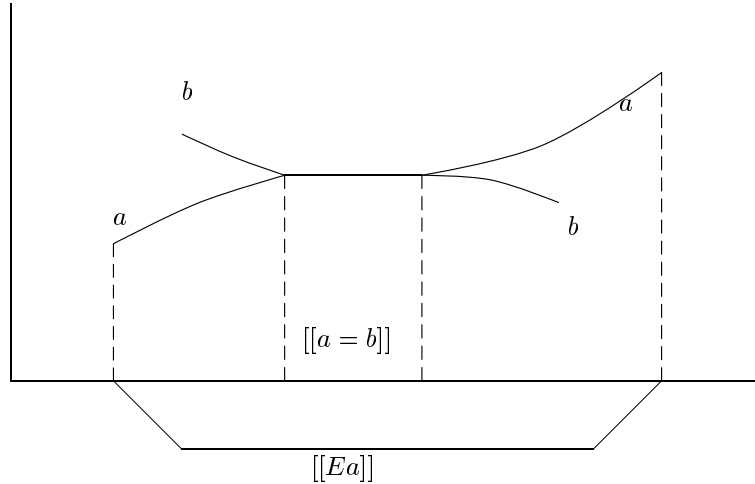
$$a^{-1} = \begin{cases} 1/a & \text{if } a \neq 0 \\ 0 & \text{if } a = 0 \end{cases}$$

(note that it could not possibly be total, (for then it had to be uniformly continuous on  $[0,1]$ ). One could avoid the problem by only discussing multiplication, but that would be a sin against the time-honoured practice of mathematics. So we would prefer to have  $a^{-1}$ , even if it means allowing partially defined terms and problems of existence.

Traditionally the matter is dealt with in free logic (cf. Bencivenga's chapter on Free Logics in this *Handbook*); we will, however briefly discuss existence here since it comes up naturally in intuitionistic logic, and since it has surprising semantic aspects.

The semantic aspects of partial elements can conveniently be demonstrated in any of the models introduced earlier. We will first consider Kripke models. Elements occur in certain domains and not in others, so they have a natural mode of existence in Kripke models. We define  $\alpha \Vdash Ea$  iff  $a \in D(\alpha)$ .  $E$  behaves as an ordinary predicate and we can handle it as usual. We may explicitly introduce the *extent* of  $a$  as follows.  $\llbracket Ea \rrbracket = \{\alpha \mid a \in D(\alpha)\}$ , clearly  $\llbracket Ea \rrbracket$  is an open set in the canonical topology.  $\llbracket Ea \rrbracket$  is that part of the underlying topological space where  $a$  exists; this explains the name *partial element*:  $\llbracket Ea \rrbracket$  need not be all of the space. Since in Beth models elements exist always (likewise in topological models), one has to consider the general models of Section 3 in order to introduce partial elements.

There is a kind of paradigm for the semantics of partial elements: sheaves over topological spaces. Without going into technical details we will sketch



the approach so that the reader can form an impression. Consider two topological spaces  $X$  and  $Y$  and continuous maps into  $Y$  defined on open subsets of  $X$ . The reader may take  $\mathbb{R}$  for  $X$  and  $Y$ . We define for such a map (*section*)  $a : \llbracket Ea \rrbracket = \text{domain } a$ .

In order to get a reasonable realistic theory we also want to interpret equality of partial elements. A natural choice is  $\llbracket a = b \rrbracket = \text{Int}\{t \in X \mid a(t) = b(t)\}$ .

Note that  $\llbracket a = a \rrbracket = \llbracket Ea \rrbracket$  and also  $\llbracket a = b \rrbracket \subseteq \llbracket Ea \rrbracket \cap \llbracket Eb \rrbracket$ . Using our knowledge of the topological interpretation (i.e. the interpretation of the connectives), we see that  $a = b \rightarrow Ea \wedge Eb$  is true. Equality satisfies the laws  $a = b \leftrightarrow b = a$  and  $a = b \wedge b = c \rightarrow a = c$ , since  $\llbracket a = b \rrbracket = \llbracket b = a \rrbracket$  and  $\llbracket a = b \rrbracket \cap \llbracket b = c \rrbracket \subseteq \llbracket a = c \rrbracket$ , but not  $a = a$ . For  $\llbracket a = a \rrbracket \neq X$ , in general, we see that the presence of partial elements affects the theory of identity (cf. [Scott, 1979]).

Of course propositional logic is not affected by the introduction of partial elements. It is predicate logic that requires attention. Turning Quine's dictum 'existence = being quantified over' around, we stipulate that one can only quantify over existing elements.

For existential quantification this makes sense,  $\exists x\varphi(x)$  means that there *exists* an element that satisfies  $\varphi$ . Adding 'but it need not exist' would be plain cheating. For universal quantification we read  $\forall x\varphi(x)$  as 'for any  $a$  picked from the domain  $\varphi(a)$  holds', which commits us to existing elements (note that in classical logic  $\exists$  decides the matter for  $\forall$ ).

The above is reflected in the axioms and rules of quantification.

$$\begin{array}{c}
[Ex] \\
\vdots \\
\forall I \frac{\varphi(x)}{\forall x\varphi(x)} \\
\exists I \frac{\varphi(t) \quad Et}{\exists x\varphi(x)}
\end{array}
\qquad
\begin{array}{c}
\forall E \frac{\forall x\varphi(x) \quad Et}{\varphi(t)} \\
\qquad \qquad \qquad [\varphi(x), Ex] \\
\qquad \qquad \qquad \vdots \\
\exists E \frac{\exists x\varphi(x) \quad \sigma}{\sigma}
\end{array}$$

(with the obvious restrictions).

In a Hilbert-type system we retain the rules  $\forall I$  and  $\exists E$  in the form

$$\frac{\psi \wedge Ex \rightarrow \varphi(x)}{\psi \rightarrow \forall x\varphi(x)} \qquad \frac{\varphi(x) \wedge Ex \rightarrow \sigma}{\exists x\varphi(x) \rightarrow \sigma}$$

and add the axioms

$$\forall x\varphi(x) \wedge Et \rightarrow \varphi(t) \quad \varphi(t) \wedge E(t) \rightarrow \exists x\varphi(x).$$

As sketched above one also has to revise the identity rules. There are two possible notions of identity, a strong one, where one requires both elements to exist, and a weaker one, where one automatically equates elements there where they do not exist.

The above equality,  $=$ , is the strong one. The weaker one can be defined by  $a \equiv b := Ea \vee Eb \rightarrow a = b$ .

The notions are interdefinable as is shown by the following fact

$$a = b \leftrightarrow a \equiv b \wedge Ea \wedge Eb.$$

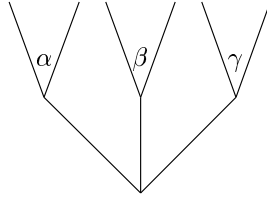
this provides us with the following axioms

$$\begin{array}{l}
x = x \leftrightarrow Ex \\
x = y \rightarrow y = x \\
x = y \wedge y = z \rightarrow x = z.
\end{array}$$

One has to select carefully the right formulation in cases involving equivalence or existence. For example,  $x \equiv x \rightarrow Ex$  is false, but  $\forall x(x \equiv x \rightarrow Ex)$  is correct, for it is equivalent on logical grounds to  $\forall x(Ex \rightarrow (x \equiv x \rightarrow Ex))$ .

The theory of partial elements is the ideal setting for the introduction of a description operator. For  $Ix.\varphi(x)$  is just a term, it has no existential import; it has to satisfy a certain formula when and where it uniquely exists. For instance, it is axiomatized by

$$\forall y[y = Ix.\varphi(x) \leftrightarrow \forall x(\varphi(x) \leftrightarrow x = y)].$$



In our model we just look for those nodes where a unique element is forced to satisfy  $\varphi(x)$  and we put them together to one partial element. Let  $\alpha \Vdash \varphi(0) \wedge \exists! x \varphi(x)$ ,  $\beta \Vdash \varphi(1) \wedge \exists! \varphi(x)$ ,  $\gamma \Vdash \varphi(0) \wedge \varphi(1)$ , then  $Ix.\varphi(x)$  is interpreted in the model as being 0 in  $\alpha$  and 1 in  $\beta$  and undefined (non-existent) in  $\gamma$ . So  $Ix.\varphi(x)$  is locally equal to given elements. This being ‘locally’ equal to something, or ‘locally’ true, etc. is a characteristic consequence of the forcing conditions of Beth semantics (cf. Section 3 above).

In a theory with identity we want to be able to replace equals by equals. Should one restrict this to the strong equality? For general (extensional) formulas this seems too restrictive, even weak equality would preserve properties, so we formulate the axiom as

$$x \equiv y \wedge \varphi(x) \rightarrow \varphi(y).$$

Following Scott [1979] we call ‘=’ identity and ‘ $\equiv$ ’ equivalence. The reader is referred to this basic paper, and to [Troelstra and van Dalen, 1988, p. 50], for more information on existence and partial elements. We add one more remark: if the theory has function symbols, then the following can be said: if one gets an output, then there must have been an input, or more formally  $Ef(x) \rightarrow Ex$ . Reversing the arrow we get the condition for a total function:  $Ex \rightarrow Ef(x)$ .

One should observe that quantifiers affect existence; they are not neutral as one would maybe expect. the interpretation of ‘ $\forall x$  \_\_\_\_\_’ is ‘for all  $x$  that exist \_\_\_\_\_’. One can actually prove  $\forall x \varphi(x) \leftrightarrow \forall x (Ex \rightarrow \varphi(x))$ .

We now return to the model of continuous maps from  $\mathbb{R}$  to  $\mathbb{R}$ .

We can operate on these functions in a pointwise manner, e.g. add multiply etc. By definition we have

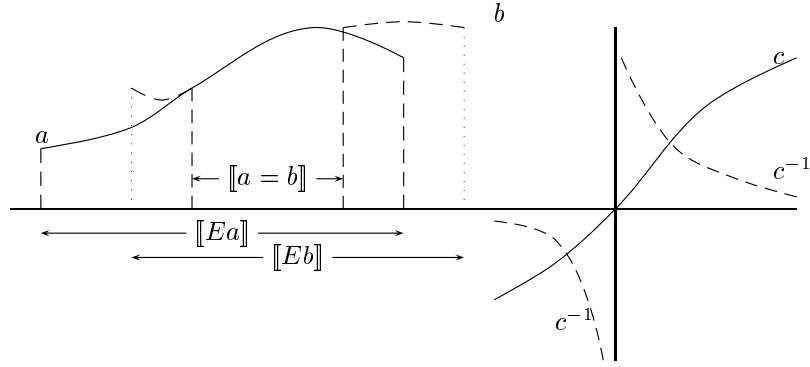
$$[f \equiv g] := \text{Int}\{x \in \mathbb{R} \mid f(x) = g(x) \vee f(x) \text{ and } g(x) \text{ are undefined}\}.$$

For quantification we put:

$$\begin{aligned} [\forall x \varphi(x)] &= \text{Int} \cap_f [Ef \rightarrow \varphi(f)] \\ [\exists x \varphi(x)] &= \cup_f [Ef \wedge \varphi(f)]. \end{aligned}$$

Now it is a matter of simple verification to check the axioms listed above.

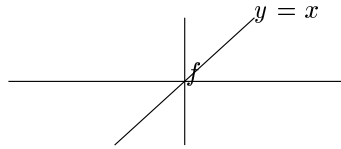
Let us look at the description operator in this model.



The elements of the model have the convenient property, peculiar to sheaves, that they can be glued together if they coincide on overlapping domains (elementary analysis). Moreover, one can always restrict a function to a smaller open domain. We use these properties to interpret  $Ix.\varphi(x)$  as the union of all continuous functions restricted to the part where they uniquely satisfy  $\varphi(x)$ , more formal  $\cup_f f \upharpoonright \llbracket \forall x(\varphi(x) \leftrightarrow f = x) \rrbracket$ , where  $f \upharpoonright U$  stands for the subfunction of  $f$  obtained by restriction of  $f$  to  $U$ .

Applying this to the inverse, we put  $h^{-1} = Ix.(xh = 1)$ . Putting together all small functions, that locally act as in inverses, we obtain a function defined on the subdomain of  $h$  obtained by leaving out the zero's of  $h$ .

Note that the model once more demonstrates the necessity of strengthening the equality relation for the existence of inverses. Note that  $\llbracket f \neq 0 \rrbracket = \text{Int}\llbracket f = 0 \rrbracket^c = \text{Int}(\text{Int}\{x \mid f(x) = 0\})^c = \mathbb{R}$  but  $\llbracket \exists x(xf = 1) \rrbracket = \mathbb{R} \setminus \{0\}$ . So  $\llbracket f \neq 0 \rightarrow \exists x(xf = 1) \rrbracket \neq \mathbb{R}$ .



Therefore we use the apartness relation, #,

$$\llbracket f \# g \rrbracket := \{x \mid f(x) \neq g(x) \text{ and } f(x) \text{ and } g(x) \text{ are defined}\}.$$

Now we get  $\llbracket f \# 0 \rightarrow \exists x(xf = 1) \rrbracket = \mathbb{R}$  (observe that this amounts to  $\llbracket f \# 0 \rrbracket \subseteq \llbracket \exists x(xf = 1) \rrbracket$ ).

After Fourman had developed the *sheaf interpretation* for the case of topological spaces, Fourman and Scott generalised the approach to sheaves over complete Heyting algebra's (so-called  $\Omega$ -sets), this approach is to be found in their paper of 1979. At that time there had already been done

a great deal in categorical logic, after the pioneering work of W. Lawvere. This topic has reached a size and technical refinement that places it utterly beyond the present book. For an introduction the reader is referred to Goldblatt's book [1979], Fourman and Scott [1979], Grayson [1984], [MacLane and Moerdijk, 1992; McLarty, 1992].

The generalisations of sheaves over topological spaces (in particular over *sites*) have provided models for various kinds of choice sequences (cf. [Hoeven and Moerdijk, 1984]).

### *Recommended Reading*

Beeson [1985], Bishop and Bridges [1985], Brouwer [1975; 1981], van Dalen [1973; 1999b], Dummett [1977], Fraenkel *et al.* [1973], Goldblatt [1979], Heyting [1956], Troelstra [1969; 1977] and Troelstra and van Dalen [1988], *Philosophica Mathematica* **6**, 1998.

*University of Utrecht.*

## BIBLIOGRAPHY

- [Aczel, 1968] P. Aczel. Saturated intuitionistic theories. In [Schmidt, Schütte and Thiele, 1968, pp. 1–11].
- [Artemov, 2001] S. Artemov. Explicit provability: the intended semantics for intuitionistic and modal logic. *Bull. Symb. Logic*, **7**, 1–36, 2001.
- [van Atten and van Dalen, forthcoming] M. van Atten and D. van Dalen. Arguments for Brouwer's continuity principle. Forthcoming.
- [Avigad and Feferman, 1998] Avigad and S. Feferman. Gödel's functional ('Dialectica') interpretation. In *Handbook of Proof Theory*, S. R. Buss, ed. pp. 337–406. Elsevier, Amsterdam, 1998.
- [Barendregt, 1984] H.P. Barendregt. *The Lambda Calculus. Its Syntax and Semantics*. North-Holland, Amsterdam, 1984, 2nd reprint edition in paperback, 1997.
- [Beeson, 1979] M. Beeson. A theory of constructions and proofs. Preprint No 134. Dept of Maths, Utrecht University, 1979.
- [Beeson, 1985] M. Beeson. *Foundations of Constructive Mathematics. Metamathematical Studies*. Springer Verlag, Berlin, 1985.
- [Bishop and Bridges, 1985] E. Bishop and D. Bridges. *Constructive Analysis*, Springer, Berlin, 1985.
- [Brouwer, 1907] L.E.J. Brouwer. *Over de Grondslagen der Wiskunde*. Thesis, Amsterdam. Translation 'On the foundations of mathematics', in [Brouwer, 1975, pp. 11–101]. New edition in [Brouwer, 1981].
- [Brouwer, 1908] L.E.J. Brouwer. De onbetrouwbaarheid der logische principes. *Tijdschrift voor wijsbegeerte*, **2**, 152–158. Translation 'The unreliability of the logical principles' in [1975, pp. 107–111]. Also in [1981].
- [Brouwer, 1918] L.E.J. Brouwer. Begründung der Mengenlehre unabhängig vom logischen Satz vom ausgeschlossenen Dritten. I. *Koninklijke Nederlandse Akademie van Wetenschappen Verhandelingen le Sectie 12*, no 5, 43 p. Also in [Brouwer, 1975, pp. 150–190].
- [Brouwer, 1975] L.E.J. Brouwer. *Collected Works*, I. A. Heyting, ed. North Holland, Amsterdam, 1975.

- [Brouwer, 1981] L.E.J. Brouwer. *Over de Grondslagen der Wiskunde*, Aangevuld met Ongepubliceerde Fragmenten, Correspondentie met D.J. Korteweg, Recensies door G. Mannoury, etc. D. van Dalen, ed. Mathematisch Centrum, varia 1, Amsterdam, 1981.
- [Brouwer, 1981a] L.E.J. Brouwer. *Brouwer's Cambridge Lectures on Intuitionism*. D. van Dalen, ed. Cambridge University Press, Cambridge, 1981.
- [Brouwer, 1992] L.E.J. Brouwer. *Intuitionismus*, D. van Dalen, ed. BI-Wissenschaftsverlag, Mannheim, 1992.
- [Burgess, 1981] J.P. Burgess. The completeness of intuitionistic propositional calculus for its intended interpretation. *Notre Dame J. Formal Logic*, **22**, 17–28, 1981.
- [Chang and Keisler, 1973] C.C. Chang and H.J. Keisler. *Model Theory*. North-Holland, Amsterdam, 1973.
- [Curry and Feys, 1958] H.B. Curry and R. Feys. *Combinatory Logic I*. North-Holland, Amsterdam, 1958.
- [van Dalen, 1973] D. van Dalen. Lectures on intuitionism. In [Mathias and Rogers, 1973, pp. 1–94].
- [van Dalen, 1974] D. van Dalen. A model for HAS. A topological interpretation of the theory of species of natural numbers. *Fund Math*, **82**, 167–174, 1974.
- [van Dalen, 1977] D. van Dalen. The use of Kripke's Schema as a reduction principle. *Journal of Symbolic Logic*, **42**, 238–240, 1977.
- [van Dalen, 1978] D. van Dalen. An interpretation of intuitionistic analysis. *Ann. Math. Logic*, **13**, 1–43, 1978.
- [van Dalen, 1984] D. van Dalen. How to glue analysis models. *Journal of Symbolic Logic*, **49**, 1339–1349, 1984.
- [van Dalen, 1986] D. van Dalen. Glueing of analysis models in an intuitionistic setting. *Studia Logica*, **45**, 181–186, 1986. To appear.
- [van Dalen, 1993] D. van Dalen. The continuum and first-order logic. *JSL*, **57**, 1417–1424, 1993.
- [van Dalen, 1995] D. van Dalen. Hermann Weyl's intuitionistic mathematics. *Bull. Symb. Logic*, **4**, 145–169, 1995.
- [van Dalen, 1997] D. van Dalen. *Logic and Structure*, 3rd edn. Springer Verlag, Berlin, 1997.
- [van Dalen, 1999a] D. van Dalen. From Brouwerian counterexamples to the creating subject. *Studia Logica*, **62**, 305–314, 1999.
- [van Dalen, 1999b] D. van Dalen. *Mystic, Geometer and Intuitionist. The Life of L.E.J. Brouwer. Volume 1: The Dawning Revolution*, Oxford University Press, 1999.
- [van Dalen, 2000] D. van Dalen. Development of Brouwer's Intuitionism. In *proof Theory. History and Philosophical Significance*, V. F. Hendricks, S. A. Pedersen and K. F. Jørgensen, eds. pp. 117–152. Kluwer, Dordrecht, 2000.
- [van Dalen and Statman, 1979] D. van Dalen and R. Statman. Equality in the presence of apartness. In [Hintikka, Niiniluoto and Saarinen, 1979, pp. 95–118].
- [Davis, 1965] M. Davis, ed. *The Undecidable. Basic Papers on Undecidable Propositions, Unsolvable Problems and Computable Functions*. Raven Press, New York, 1965.
- [Demuth and Kučera, 1979] O. Demuth and A. Kučera. Remarks on constructive mathematical analysis. In *Logic Colloquium '78*, M. Boffa, D. van Dalen and K. McAloon, eds. pp. 81–130. North Holland, Amsterdam, 1979.
- [Dummett, 1973] M. Dummett. The philosophical basis of intuitionistic logic. In *Logic Colloquium '73*, H. E. Rose and J. C. Shepherdson, eds. pp. 5–40, North-Holland, Amsterdam, 1973. Also in M. Dummett, *Truth and Other Enigmas*, pp. 215–247, Duckworth, London, 1978.
- [Dummett, 1977] M. Dummett. *Elements of Intuitionism*. Oxford University Press, Oxford, 1977.
- [Dyson and Kreisel, 1961] V.H. Dyson and G. Kreisel. *Analysis of Beth's Semantic Construction of Intuitionistic Logic*. Technical Report no. 3. Appl. math. and statistics lab. Stanford University, 65 pp. 1961.
- [Feferman, 1979] S. Feferman. Constructive theories of functions and classes. In *Logic Colloquium '78*, M. Boffa, D. van Dalen and K. McAloon, eds. pp. 159–224. North Holland, Amsterdam, 1979.



- [Fenstad, 1971] J.E. Fenstad, ed. *Proceedings of the Second Scandinavian Logic Symposium*, North-Holland, Amsterdam, 1971.
- [Fine, 1970] K. Fine. An intermediate logic without the finite model property. Unpublished, 1970.
- [Fitting, 1969] M.C. Fitting. *Intuitionistic Logic, Model Theory and Forcing*. North-Holland, Amsterdam, 1969.
- [Fourman, 1982] M. Fourman. Notions of choice sequence. In [Troelstra and van Dalen, 1982, pp. 91–106].
- [Fourman and Scott, 1979] M.P. Fourman and D. S. Scott. Sheaves and logic. In [Fourman, Mulvey and Scott, 1979, pp. 302–401].
- [Fourman, Mulvey and Scott, 1979] M.P. Fourman, C.J. Mulvey and D.S. Scott. *Applications of Sheaves. Proceedings Durham 1977*. Lecture Notes No 753, Springer Verlag, Berlin, 1979.
- [Fraenkel *et al.*, 1973] A. Fraenkel, Y. Bar-Hillel, A. Levy and D. van Dalen. *Foundations of Set Theory*. North-Holland, Amsterdam, 1973.
- [Friedman, 1975] H. Friedman. The disjunction property implies the numerical existence property. *Proc Nat Acad. Sci.*, **72**, 2977–2878, 1975.
- [Friedman, 1977] H. Friedman. The intuitionistic completeness of intuitionistic logic under Tarskian semantics. Abstract, SUNY at Buffalo, 1977.
- [Friedman, 1977a] H. Friedman. New and old results on completeness of **HPC**. Abstract, SUNY at Buffalo, 1977.
- [Friedman, 1977b] H. Friedman. Classically and intuitionistically provably recursive functions. In *Higher Set Theory*, pp. 21–27, Springer Verlag, Berlin, 1977.
- [Gabbay, 1970] D.M. Gabbay. Decidability of the Kripke–Putnam system. *Journal of Symbolic Logic*, **35**, 431–437, 1970.
- [Gabbay, 1971] D.M. Gabbay. Semantic proof of the Craig Interpolation theorem for intuitionistic logic and extensions I, II. In *Logic Colloquium 69*, (eds R.O. Gandy and C.M.E. Yates), pp. 391–410. North-Holland, Amsterdam, 1971.
- [Gabbay, 1977] D.M. Gabbay. On some new intuitionistic propositional connectives. *Studia Logica*, **36**, 127–139, 1977.
- [Gabbay, 1981] D.M. Gabbay. *Semantical Investigations in Heyting's Intuitionistic Logic*. D. Reidel, Dordrecht, 1981.
- [Gabbay and de Jongh, 1974] D.M. Gabbay and D.H. de Jongh. Sequences of decidable finitely axiomatisable intermediate logics with the disjunction property. *Journal of Symbolic Logic*, **39**, 67–79, 1974.
- [Gallier, 1995] J. Gallier. On the correspondence between proofs and  $\lambda$ -terms. In *The Curry–Howard Isomorphism*, (ed. Ph. de Groote), pp. 55–138. Cahiers de centre de Logique, 8. Academia Louvain-la-Neuve, 1995.
- [Gentzen, 1933] G. Gentzen. Über das Verhältnis zwischen intuitionistischer und klassischer Arithmetik. English translation in [Szabo, 1969, pp. 53–67].
- [Gentzen, 1935] G. Gentzen. Untersuchungen über das logische Schliessen I, II. *Math. Zeitschrift*, **39**, 176–210, 405–409, 1935.
- [Girard, 1971] J.Y. Girard. Une extension de l'interprétation de Gödel à l'analyse et son application à l'élimination des coupures dans l'analyse et la théorie des types. In [Fenstad, 1971, pp 63–92].
- [Girard *et al.*, 1989] J.Y. Girard, P. Taylor and Y. Lafont. *Proofs and Types*. Cambridge University Press, Cambridge, 1989.
- [Glivenko, 1929] V. Glivenko. Sur quelques points de la logique de M. Brouwer. *Académie Royale de Belgique, Bull. de la classe des sciences*, (5), VI. 15, pp. 183–188, 1929.
- [Goad, 1978] C.A. Goad. Monadic infinitary propositional logic: a special operator. *Repts Math. Logic*, **10**, 43–50, 1978.
- [Gödel, 1932] K. Gödel. Zum intuitionistischen Aussagenkalkül. *Akademie der Wissenschaften in Wien. Math. naturwiss. Klasse. Anzeiger*, **69**, 65–66, 1932. Also in *Ergebnisse eines Math. Koll.*, **4**, 42.
- [Gödel, 1933] K. Gödel. Zur intuitionistischen Arithmetik und Zahlentheorie. *Ergebnisse eines mathematischen Kolloquiums*, **4**, 34–38, 1933. English translation in [Davis, 1965, pp. 75–81]. Cf. *Journal of Symbolic Logic*, **31**, 484–494, 1966.

- [Gödel, 1958] K. Gödel. Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. *Dialectica*, **12**, 280–287. English translation in *Journal of Philosophical Logic*, **9**, 133–142, 1980.
- [Goldblatt, 1979] R. Goldblatt. *Topoi. The Categorical analysis of Logic*. North Holland, Amsterdam, 1979. Revised edition, 1984.
- [Görnemann, 1971] S. Görnemann. A logic stronger than intuitionism. *Journal of Symbolic Logic*, **36**, 249–261, 1971.
- [Grayson, 1981] R. Grayson. Concepts of general topology in constructive mathematics and in sheaves. *Ann. Math. Logic*, **20**, 1–41, 1981.
- [Grayson, 1984] R. Grayson. Heyting-valued semantics. In *Logic Colloquium 1982*, G. Lolli, G. Longo and A. Marcja, eds. pp. 181–208. North Holland, Amsterdam, 1984.
- [Grzegorzczak, 1964] A. Grzegorzczak. A philosophically plausible interpretation of intuitionistic logic. *Indagationes Mathematicae*, **26**, 569–601, 1964.
- [Harrop, 1958] R. Harrop. On the existence of finite models and decision procedures for propositional calculi. *Proc. Camb. Phil. Soc.*, **54**, 1–13, 1958.
- [Hartog, 1978] W. den Hartog. *A Proof Theoretic Study. The Theory of Pseudo-order as a Conservative Extension of the Theory of Apartness*. Dept of Math. Rijksuniversiteit Utrecht, Preprint no. 77, 1978.
- [Heijenoort, 1967] J. van Heijenoort. *From Frege to Gödel, a Source Book in Mathematical Logic 1879–1931*. Harvard University Press, Cambridge, MA, 1967.
- [Heyting, 1930] A. Heyting. Die formalen Regeln der intuitionistischen Logik. *Sitzungsberichte der preussischen Akademie von Wissenschaften*, pp. 42–56, 1930. ‘Die formalen Regeln der intuitionistischen Mathematik’, *Ibid.* pp. 57–71, 158–169. Also (partly) in *Two Decades of Mathematics in the Netherlands*, Amsterdam, 1978.
- [Heyting, 1934] A. Heyting. *Mathematische Grundlagenforschung. Intuitionismus Beweisstheorie*. Springer, Berlin, 1934.
- [Heyting, 1956] A. Heyting. *Intuitionism. An Introduction*. North Holland, Amsterdam, 1956.
- [Hintikka, Niiniluoto and Saarinen, 1979] J. Hintikka, I. Niiniluoto and E. Saarinen, eds. *Essays on Mathematical and Philosophical logic*, D. Reidel, Dordrecht, 1979.
- [Hoeven and Moerdijk, 1984] G. van Hoeven and I. Moerdijk. Sheaf models for choice sequences. *Annals of Pure and Applied Logic*, **27**, 63–107, 1984.
- [Hosoi, 1967] T. Hosoi. On intermediate logics, I. *J. Fac. Sci. Univ. of Tokyo*, **14**, 293–312, 1967.
- [Howard, 1980] W. Howard. The formulae-as-types notion of construction. In [Seldin and Hindley, 1980, pp. 479–490].
- [Hughes and Cresswell, 1968] G.E. Hughes and M.J. Cresswell. *An Introduction to Modal Logic*. Methuen, London, 1968.
- [Hull, 1969] R.C. Hull. Counterexamples in intuitionistic analysis using Kripke’s schema. *Z. Math. Logik und Grundlagen der Math.*, **15**, 241–246, 1969.
- [Hyland, 1982] M. Hyland. The effective topos. In [Troelstra and van Dalen, 1982, pp. 165–216].
- [Hyland, Johnstone and Pitts, 1980] M. Hyland, P.T. Johnstone and A.M. Pitts. Triples theory. *Math. Proc. Camb. Phil. Soc.*, **88**, 205–232, 1980.
- [Jankov, 1968] V.A. Jankov. Constructing a sequence of strongly independent super-intuitionistic propositional calculi. *Soviet Math. Dok.*, **9**, 806–807, 1968.
- [Jaskowski, 1936] S. Jaskowski. Recherches sur le système de la logique intuitioniste. *Actes du Congrès Intern. de Phil. Scientifique. VI. Phil des mathématiques, Act. Sc. et Ind 393*. pp. 58–61. Paris, 1936.
- [Johansson, 1936] I. Johansson. Der Minimalkalkül, ein reduzierter intuitionistischer Formalismus. *Compositio Math.*, **4**, 119–136, 1936.
- [Johnstone, 1982] P. Johnstone. *Stone Spaces*. Cambridge University Press, Cambridge, 1982.
- [Jongh, 1980] D.H. de Jongh. A class of intuitionistic connectives. In *The Kleene Symposium*, J. Barwise, H. J. Keisler and K. Kunen, eds. pp. 103–112. North Holland, Amsterdam, 1980.
- [Jongh and Smoryński, 1976] D.H. de Jongh and C. Smoryński. Kripke models and the intuitionistic theory of species. *Ann. Math. Logic*, **9**, 157–186, 1976.

- [Kino, Myhill and Vesley, 1970] A. Kino, J. Myhill and R.E. Vesley. *Intuitionism and Proof Theory. Proceedings of the Summer Conference at Buffalo, New York*, 1968. North Holland, Amsterdam, 1970.
- [Kleene, 1952] S.C. Kleene. *Introduction to Meta-mathematics*. North Holland, Amsterdam, 1952.
- [Kleene, 1973] S.C. Kleene. Realisability: A retrospective survey. In [Mathias and Rogers, 1973], pp. 95–112].
- [Kleene and Vesley, 1965] S.C. Kleene and R.E. Vesley. *The Foundations of Intuitionistic Mathematics, especially in Relation to Recursive Functions*. North Holland, Amsterdam, 1965.
- [Klop, 1980] J.W. Klop. *Combinatory Reduction Systems*. Thesis, Rijks Universiteit Utrecht. Also MC Tract 127. Math. Centre Amsterdam, 1980.
- [Kolmogorov, 1925] A.N. Kolmogorov. On the principle of the excluded middle, Russian. *Matematičeski Sbornik*, **32**, 646–667, 1925. English translation in [Heijenoort, 1967], pp. 414–437].
- [Kreisel, 1958] G. Kreisel. A remark on free choice sequences and the topological completeness proofs. *J. Symbolic Logic*, **23**, 369–388, 1958.
- [Kreisel, 1959] G. Kreisel. Interpretations of analysis by means of constructive functionals of finite type. In *Constructivity in Mathematics*, A. Heyting, ed. pp. 101–128. North Holland, Amsterdam, 1959.
- [Kreisel, 1962] G. Kreisel. On weak completeness of intuitionistic predicate logic. *JSL*, **27**, 139–158, 1962.
- [Kreisel, 1965] G. Kreisel. Mathematical Logic. In *Lectures on Modern Mathematics III*, T. L. Saaty, ed. pp. 95–195. Wiley & Sons, New York, 1965.
- [Kreisel, 1967] G. Kreisel. Informal rigour and completeness proofs. In *Problems in the Philosophy of Mathematics*, I. Lakatos, ed. North Holland, Amsterdam, 1967.
- [Kreisel, 1970] G. Kreisel. Church's Thesis, a kind of reducibility axiom of constructive mathematics. In [Kino, Myhill and Vesley, 1970], pp. 121–150].
- [Kreisel and Putnam, 1957] G. Kreisel and H. Putnam. Eine Unableitbarkeitsbeweismethode für den intuitionistischen Aussagenkalkül. *Archiv. f. math. Logik*, **3**, 74–78, 1957.
- [Kreisel and Troelstra, 1970] G. Kreisel and A.S. Troelstra. Formal systems for some branches of intuitionistic analysis. *Ann. Math. Logic*, **1**, 229–387, 1970.
- [Kripke, 1965] S. Kripke. Semantical analysis of intuitionistic logic I. In *Formal Systems and Recursive Functions*, J. Crossley and M. Dummett, eds. pp. 92–129. North Holland, Amsterdam, 1965.
- [Leivant, 1976] D. Leivant. Failure of completeness properties of intuitionistic predicate logic for constructive models. *Ann. Sc. Univ. Clermont. Sér Math.*, **13**, 93–107, 1976.
- [Leivant, 1985] D. Leivant. Syntactic translations and provably recursive functions. *Journal of Symbolic Logic*, **50**, 682–688, 1985.
- [Lemmon and Scott, 1966] E. Lemmon and D.S. Scott. Intensional logics. Published in 1977 as *An Introduction to Modal Logic*, K. Segerberg, ed. Blackwell, Oxford, 1966.
- [Lifschitz, 1969] V.A. Lifschitz. Problems of decidability for some constructive theories of equalities. In *Studies in Constructive Mathematics and Mathematical Logic*, A.O. Slisenko, ed. Consultants Bureau, New York, 1969.
- [MacLane and Moerdijk, 1992] S. MacLane and I. Moerdijk. *Sheaves in Geometry and Logic*, Springer, Berlin, 1992.
- [McCarty, 1988] D.C. McCarty. Constructive validity is nonarithmetic. *JSL*, **53**, 1036–1041, 1988.
- [McKinsey, 1939] J.C.C. McKinsey. Proof of the independence of the primitive symbols of Heyting's calculus of propositions. *JSL*, **4**, 155–158, 1939.
- [McLarty, 1992] C. McLarty. *Elementary Categories, Elementary Toposes*. Oxford University Press, 1992.
- [Markov, 1950] A.A. Markov. Konstruktivnaja logika. *Usp. Mat. Nauk*, **5**, 187–188, 1950.
- [Martin-Löf, 1971] P. Martin-Löf. Hauptsatz for the theory of species. In [Fenstad, 1971].

- [Martin-Löf, 1977] P. Martin-Löf. Constructive mathematics and computer programming. In *Logic, Methodology and the Philosophy of Science VI*, L.J. Cohen, J. Los, H. Pfeiffer and K.P. Podewski, eds. pp. 153–179. North Holland, Amsterdam, 1982.
- [Martin-Löf, 1984] P. Martin-Löf. *Intuitionistic Type Theory*. Notes by G. Sambin of a series of lectures given in Padova, June 1982. Bibliopolis, Naples, 1984.
- [Mathias and Rogers, 1973] A.R.D. Mathias and H. Rogers, Jr. *Cambridge Summer School in Mathematical Logic*, Springer Lecture Notes, Vol. 337, Berlin, 1973.
- [Maximova, 1977] L.L. Maximova Craig interpolation theorem and amalgamable varieties. *Soviet Math. Dok.*, **18**, 1550–1553, 1977.
- [Minc, 1966] G. Minc. Skolem's method of elimination of positive quantifiers in sequential calculi. *Dokl. Akad. Nauk SSR*, **169**, 861–864, 1966.
- [Minc, 1974] G. Minc. On E-theorems. (In Russian). Investigation in constructive mathematics VI. *Zapisky Nauk. Sem. Leningrad Steklov Inst.*, **40**, 110–118, 1974.
- [Moschovakis, 1967] J. Moschovakis. Disjunction and existence in formalised intuitionistic analysis. In *Sets, Models and Recursion Theory*, J. Crossley, ed. pp. 309–31. North Holland, Amsterdam, 1967.
- [Moschovakis, 1973] J. Moschovakis. A topological interpretation for second-order intuitionistic arithmetic. *Comp. Math.*, **26**, 261–276, 1973.
- [Moschovakis, 1980] J. Moschovakis. A disjunctive decomposition theorem for classical theories. In *Constructive Mathematics*, F. Richman, ed. pp. 250–259. Springer Verlag, Berlin, 1980.
- [Negri and von Plato, 2001] S. Negri and J. von Plato. *Structural Proof Theory*, Cambridge University Press, Cambridge, 2001.
- [Nelson, 1949] D. Nelson. Constructible falsity. *Journal of Symbolic Logic*, **14**, 16–26, 1949.
- [Nishimura, 1966] T. Nishimura. On formulas of one variable in intuitionistic propositional calculus. *Journal of Symbolic Logic*, **25**, 327–331, 1966.
- [Ono, 1972] H. Ono. Some results on intermediate logics. *Publ. RIMS*, pp. 117–130, Kyoto University 9, 1972.
- [Ono, 1973] H. Ono. A study of intermediate predicate logics. *Publ. RIMS*, pp. 619–649. Kyoto University 8, 1973.
- [Ono, 1985] H. Ono. Semantical analysis of predicate logics without the contraction rule. *Studia Logica*, **44**, 187–196, 1985.
- [Ono and Komori, 1985] H. Ono and Y. Komori. Logics without the contraction rule. *Journal of Symbolic Logic*, **50**, 169–201, 1985.
- [Posy, 1976] C.J. Posy. Varieties of indeterminacy in the theory of general choice sequences. *J. Phil. Logic*, **5**, 91–132, 1976.
- [Posy, 1977] C.J. Posy. The theory of empirical sequences. *J. Phil. Logic*, **6**, 47–81, 1977.
- [Posy, 1980] C.J. Posy. On Brouwer's definition of unextendable order. *History and Phil. Logic*, **1**, 139–149, 1980.
- [Pottinger, 1976] G. Pottinger. A new way of normalising intuitionist propositional logic. *Studia Logica*, **35**, 387–408, 1976.
- [Prawitz, 1965] D. Prawitz. *Natural Deduction. A Proof Theoretical Study*. Alqvist & Wiksell, Stockholm, 1965.
- [Prawitz, 1970] D. Prawitz. Some results for intuitionistic logic with second-order quantification rules. In [Kino, Myhill and Vesley, 1970, pp. 259–270].
- [Prawitz, 1971] D. Prawitz. Ideas and results in proof theory. In [Fenstad, 1971].
- [Prawitz, 1977] D. Prawitz. Meanings and proofs: on the conflict between classical and intuitionistic logic. *Theoria*, **43**, 2–40, 1977.
- [Prawitz, 1979] D. Prawitz. Proofs and the meaning and the completeness of the logical constants. In [Hintikka, Niiniluoto and Saarinen, 1979, pp. 25–40].
- [Prawitz and Malmnäs, 1968] D. Prawitz and P. Malmnäs. A survey of some connections between classical, intuitionistic and minimal logic. In [Schmidt, Schütte and Thiele, 1968, pp. 215–229].
- [Rasiowa, 1974] H. Rasiowa. *An algebraic Approach to Non-Classical Logics*. North Holland, Amsterdam, 1974.
- [Rasiowa and Sikorski, 1963] H. Rasiowa and R. Sikorski. *The Mathematics of Metamathematics*, Panstwowe Wydawnictwo Naukowe Warszawa, 1963.

- [Rauszer, 1980] C. Rauszer. An algebraic and Kripke-style approach to a certain extension of intuitionistic logic. *Dissertationes Mathematicae*, **CLXVII**, Warszawa, 67 pp. 1980.
- [Rautenberg, 1979] W. Rautenberg. *Klasische und nichtklassische Aussagenlogik*, Vieweg & Sohn, Braunschweig/Wiesbaden, 1979.
- [Renardel de Lavalette, 1981] G.R. Renardel de Lavalette. The interpolation theorem in fragments of logic. *Indag. Math.*, **43**, 71–86, 1981.
- [Schmidt, Schütte and Thiele, 1968] H. A. Schmidt, K. Schütte and H.J. Thiele, eds. *Contributions to Mathematical Logic*. North Holland, Amsterdam, 1968.
- [Schroeder-Heister, 1984] P. Schroeder-Heister. A natural extension of natural deduction. *Journal of Symbolic Logic*, **49**, 1284–1300, 1984.
- [Schütte, 1962] K. Schütte. Der Interpolationssatz der intuitionistischen Prädikatenlogik. *Math. Ann.*, **148**, 192–200, 1962.
- [Schütte, 1968] K. Schütte. *Vollständige Systeme modaler und intuitionistischer Logik*, Ergebnisse der Mathematik und ihrer Grenzgebiete, **42**, Springer Verlag, Berlin, 1968.
- [Scott, 1968] D.S. Scott. Extending the topological interpretation to intuitionistic analysis II. In [Kino, Myhill and Vesley, 1970, pp. 235–255].
- [Scott, 1979] D.S. Scott. Identity and existence in intuitionistic logic. In [Fourman, Mulvey and Scott, 1979, pp. 660–696].
- [Scowcroft, 1999] P. Scowcroft. Some purely topological models for intuitionistic analysis. *Ann. Pure Appl. Logic*, **98**, 173–216, 1999.
- [Segerberg, 1968] K. Segerberg. Propositional logics related to Heyting's and Johansson's. *Theoria*, **34**, 26–61, 1968.
- [Seldin and Hindley, 1980] J.P. Seldin and J.R. Hindley. *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, Academic Press, London, 1980.
- [Shoenfield, 1967] J.R. Shoenfield. *Mathematical Logic*, Addison Wesley, Reading, MA, 1967. Re-issued in paperback by A.K. Peters, 2001.
- [Smoryński, 1973] C. Smoryński. *Investigation of Intuitionistic Formal Systems by Means of Kripke Models*. Diss. Univ. of Illinois, 1973.
- [Smoryński, 1973a] C. Smoryński. Elementary intuitionistic theories. *Journal of Symbolic Logic*, **38**, 102–134, 1973.
- [Smoryński, 1977] C. Smoryński. On axiomatising fragments. *Journal of Symbolic Logic*, **42**, 530–544, 1977.
- [Smoryński, 1978] C. Smoryński. The axiomatisation problem for fragments. *Ann. Math. Logic*, **14**, 193–221, 1978.
- [Smoryński, 1982] C. Smoryński. Non-standard models and constructivity. In [Troelstra and van Dalen, 1982, pp. 459–464].
- [Sommaruga, 2000] G. Sommaruga. *History and Philosophy of Constructive Type theory*. Kluwer, Dordrecht, 2000.
- [Stein, 1980] M. Stein. Interpretations of Heyting's arithmetic. An analysis by means of a language with set symbols. *Ann. Math. Logic*, **19**, 1–31, 1980.
- [Stein, 1981] M. Stein. A general theorem on existence theorems. *Zeit. Math. Logik Grundl. Math.*, **27**, 435–452, 1981.
- [Sundholm, 1983] G. Sundholm. Constructions, proofs and the meanings of the logical constants. *J. Phil. Logic*, **12**, 151–172, 1983.
- [Swart, 1976] H.C.M. de Swart. Another intuitionistic completeness proof. *Journal of Symbolic Logic*, **41**, 644–662, 1976.
- [Szabo, 1969] M.E. Szabo, ed. *The Collected Papers of Gerhard Gentzen*. North Holland, Amsterdam, 1969.
- [Tait, 1975] N. Tait. A realisability interpretation of the theory of species. In *Logic Colloquium*, R. Parikh, ed. pp. 240–251. Springer Verlag, Berlin, 1975.
- [Takahashi, 1970] M. Takahashi. Cut-elimination theorem and Brouwerian-valued models for intuitionistic type theory. *Comment. Math. Univ. St Pauli*, **19**, 55–72, 1970.
- [Thomason, 1969] R. H. Thomason. A semantical study of constructible falsity. *Z. Math. Logik und Grundlagen der Mathematik*, **15**, 247–257, 1969.
- [Troelstra, 1969] A.S. Troelstra. *Principles of Intuitionism*. Springer Verlag, Berlin, 1969.

- [Troelstra, 1973] A.S. Troelstra. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Springer Verlag, 1973.
- [Troelstra, 1973a] A.S. Troelstra. Notes on intuitionistic second-order arithmetic. In [Mathias and Rogers, 1973, pp. 171–205].
- [Troelstra, 1977] A.S. Troelstra. *Choice Sequences. A Chapter of Intuitionistic Mathematics*. Oxford University Press, Oxford, 1977.
- [Troelstra, 1978] A.S. Troelstra. Commentary on Heyting [1930]. In *Two Decades of Mathematics in the Netherlands*, Part 5, Math. Centre, Amsterdam, 1978.
- [Troelstra, 1979] A.S. Troelstra. *A Supplement to Choice Sequences*. Report 79-04 of the Math Institute University of Amsterdam. 22 pp. 1979.
- [Troelstra, 1980] A.S. Troelstra. The interplay between logic and mathematics: intuitionism. In *Modern Logic—A Survey*. E. Agazzi, ed. pp. 197–221. D. Reidel, Dordrecht, 1980.
- [Troelstra, 1981] A.S. Troelstra. On a second-order propositional operator in intuitionistic logic. *Studia Logica*, **40**, 113–140, 1981.
- [Troelstra, 1981a] A.S. Troelstra. Arend Heyting and his contribution to intuitionism. *Nieuw Archief voor Wiskunde*, **24**, pp. 1–23, 1981.
- [Troelstra, 1983] A.S. Troelstra. Analysing choice sequences. *Journal of Philosophical Logic*, **12**, 197–260, 1983.
- [Troelstra, 1998] A.S. Troelstra. Realizability. In *Handbook of Proof Theory*, (ed. S.R.Buss). pp. 407–474. Elsevier, Amsterdam, 1998.
- [Troelstra and van Dalen, 1982] A. Troelstra and D. van Dalen, eds. *The L.E.J. Brouwer Centenary Symposium*, North Holland, Amsterdam, 1982.
- [Troelstra and van Dalen, 1988] A. Troelstra and D. van Dalen. *Constructivism in Mathematics*, I, II, Elsevier, Amsterdam, 1988.
- [Troelstra and Schwichtenberg, 1996] A. Troelstra and H. Schwichtenberg. *Basic proof Theory*. Cambridge University Press, Cambridge, 1996.
- [Veldman, 1976] W. Veldman. An intuitionistic completeness theorem for intuitionistic predicate logic. *J. Symbolic Logic*, **41**, 159–166, 1976.
- [Veldman and Janssen, 1990] W. Veldman and M. Janssen. Some observations on intuitionistically elementary properties of linear orderings. *Arch. Math. Logic*, **29**, 171–187, 1990.
- [Veldman and Waaldijk, 1996] W. Veldman and F. Waaldijk. Some elementary results in intuitionistic model theory. *JSL*, **61**, 745–767, 1996.
- [Visser, 1982] A. Visser. On the completeness principle: a study of provability in Heyting's Arithmetic. *Annals Math. Logic*, **22**, 263–295, 1982.
- [Zucker and Tragesser, 1978] J. T. Zucker and R. S. Tragesser. The adequacy problem for inferential logic. *J. Philosophical Logic*, **7**, 501–516, 1978.

WALTER FELSCHER

## DIALOGUES AS A FOUNDATION FOR INTUITIONISTIC LOGIC

### SUMMARY OF CONTENTS

The principal content of this article is a (new) foundation for intuitionistic logic, based on an analysis of argumentative processes as codified in the concepts of a *dialogue* and a *strategy* for dialogues. This work is presented in Section 3. A general historical introduction is given in Section 2. Since already there the reader will need to know exactly what a dialogue and a strategy shall be, these basic concepts are defined in the (purely technical) Section 1.

#### 1 BASIC CONCEPTS: DIALOGUES AND STRATEGIES

I consider a first-order language, built with variables  $x, y, \dots$  and terms  $t$ ; formulas shall be constructed from atomic formulas with the propositional connectives  $\wedge, \vee, \rightarrow, \neg$  and the quantifiers  $\forall, \exists$ ; I shall also consider  $\vee, \wedge_1, \wedge_2, \exists$  as *special symbols* in their own right. By an *expression* I understand either a term or a formula or a special symbol. I introduce two further symbols  $P$  and  $Q$ ; taking two new (and disjoint) copies of the set of expressions, I form for every expression  $e$  two new expressions  $Pe$  and  $Qe$ , the *P-signed* and the *Q-signed version* of the expression  $e$ .

The symbols  $P, Q$  shall symbolise two persons engaged in an argument or in a dialogue; I shall use  $X, Y$  as variables for  $P, Q$  and shall assume  $X \neq Y$ . An *argumentation form* is a schematic presentation of an argument, concerning a logically composite assertion; it describes how a composite assertion made by  $C$  may be *attacked* by  $Y$  and how, if possible, this attack may be *answered* by  $X$ . As the logical form of the composite assertion shall completely determine the argument, each of the four propositional connectives and each of the two quantifiers determines an argumentation form:

$\wedge$ :	assertion:	$X w_1 \wedge w_2$	
	attack:	$Y \wedge_i$	(i.e., $Y$ chooses $i = 1$ or $i = 2$ )
	answer:	$X w_i$	
$\vee$ :	assertion:	$X w_1 \vee w_2$	
	attack:	$Y \vee$	
	answer:	$X w_i$	(i.e., $X$ chooses $i = 1$ or $i = 2$ )
$\rightarrow$ :	assertion:	$X w_1 \rightarrow w_2$	
	attack:	$Y w_1$	
	answer:	$X w_2$	
$\neg$ :	assertion:	$x \neg w$	
	attack:	$Y w$	
	answer:	<i>no answer possible</i>	
$\forall$ :	assertion:	$X \forall x w$	
	attack:	$Y t$	(i.e., $Y$ chooses the term $t$ )
	answer:	$X w(t)$	
$\exists$ :	assertion:	$X \exists x w$	
	attack:	$Y \exists$	
	answer:	$X w(t)$	(i.e., $X$ chooses the term $t$ ).

In the last two answers I have written  $w(t)$  for the substitution instance obtained from  $w$  if the term  $t$  is substituted for the variable  $x$ .

A dialogue shall be a (finite or infinite) sequence  $\delta$  of statements, i.e., signed expressions, stated alternately by  $P$  and  $Q$  and progressing in accordance with the argumentation forms; I shall consider only such dialogues which are begun by  $P$ . Since it is necessary to distinguish carefully between attacks, answers and the assertions they refer to, I shall introduce besides  $\delta$  an accompanying sequence  $\eta$  of references, and there I shall use the symbols  $A$  for *attack* and  $D$  for *answer (defense)*. For notational convenience, I shall assume that a natural number *is* the set of all smaller natural numbers (whence 0 is the first natural number), and a *sequence* shall always be a function, defined on either a natural number or on the set  $\omega$  of all natural numbers. The precise definition then reads as follows:

A *dialogue*  $\delta, \eta$  consists of two sequences such that

$\delta$  is a sequence of signed expressions,

$\eta$  is a function defined on the *positive* members of  $\text{def}(\delta)$ , and if  $n$  in  $\text{def}(\eta)$  is an ordered pair  $[m, Z]$  such that  $m$  is a natural number less than  $n$  and  $Z$  is either  $A$  or  $D$  ,

satisfying the properties (D00)–(D02):



- (D00)  $\delta(n)$  is  $P$ -signed if  $n$  is even and  $Q$ -signed if  $n$  is odd;  $\delta(0)$  is a composite formula.
- (D01) If  $\eta(n) = [m, A]$  then  $\delta(m)$  is a composite formula and  $\delta(n)$  is attack upon  $\delta(m)$  according to the appropriate argumentation form.
- (D02) If  $\eta(p) = [n, D]$  then  $\eta(n) = [m, A]$  and  $\delta(p)$  is the answer to the attack  $\delta(n)$  according to the appropriate argumentation form.

The signed formulas occurring as values of  $\delta$  are called the *assertions* of the dialogue while the remaining values of  $\delta$  are *symbolic statements* or, more correctly, *symbolic attacks*. The numbers in  $\text{def}(\delta)$  are called the *positions* or *places* of the dialogue. If  $Pv$  is the assertion  $\delta(0)$ , the dialogue is said to be a dialogue *for* the formula  $v$  (or, sometimes, for  $Pv$ ).

Assume now that a particular class  $H$  of dialogues is given, defined maybe by additional conditions, which has the property that, for every position  $n$  of an  $H$ -dialogue  $\delta, \eta$ , the *restrictions* of  $\delta, \eta$  to positions  $i$  such that  $i \leq n$  form an  $H$ -dialogue again. Assume further that a subclass of  $H$  has been defined, consisting of certain *finite*  $H$ -dialogues which then are said to be the  $H$ -dialogues *won* by  $P$ . Let  $v$  be a composite formula; to say that  $P$  has an  $H$ -strategy shall mean that  $P$  is in possession of a system of information, consisting of possible choices of  $P$ -statements in dialogues, such that every  $H$ -dialogue for  $v$  is won by  $P$  if only  $P$  chooses, after every statement made by  $Q$ , its own statement from this system of information. In order to formulate a more precise definition, recall that a *tree*  $S$  is a partially ordered set of elements called *nodes* with the following properties: there exists a *largest* element  $e_S$  (the top node), and for every node  $e$  the number  $\|e\|$  of nodes  $f$  such that  $e \leq f < e_S$  is *finite*; every node except  $e_S$  has exactly one *upper neighbour* but may have arbitrarily many *lower neighbours* (i.e., the tree is branching downwards). A *path* in  $S$  is a linearly ordered subset of nodes which, together with each of its elements  $e$ , contains all the preceding nodes  $f$  with  $e \leq f$ ; a *branch* is a path which is maximal. If  $A$  is a branch of  $S$ , let  $\alpha_A$  be the unique order-preserving bijection which maps either a natural number or all of  $\omega$  onto  $A$ , i.e.  $\|\alpha_A(i)\| = i$  holds for every node  $\alpha_A(i)$  in  $A$ . Consider now a tree  $S$  and functions  $\delta, \eta$  where  $\delta$  is defined on all nodes of  $S$  and  $\eta$  on the nodes different from  $e_S$ ; for every branch  $A$  define  $\delta_A = \delta \cdot \alpha_A, \eta_A = \eta \cdot \alpha_A$ . The triplet  $S, \delta, \eta$  then is an  $H$ -strategy for  $v$  if

- (S0) For every branch  $A$  of  $S$  the pair  $\delta_A, \eta_A$  is an  $H$ -dialogue for  $v$  which is won by  $P$ .
- (S1) For every node  $e$  of  $S$  the following is the case. If  $\|e\|$  is odd then  $S$  does not branch at  $e$ . If  $\|e\|$  is even then  $e$  has as many lower neighbours as  $Q$  has possibilities to extend, by adding a new position, to an  $H$ -dialogue the (restricted) dialogue leading to  $e$ ,

and  $\delta, \eta$  assign these lower neighbours the values which realise these possibilities.

The general definitions having been established, particular classes of dialogues can be introduced. To do so, I shall need the following terminology. Let  $\delta, \eta$  be a dialogue, and let  $\delta(n)$  be one of its attacks. The attack  $\delta(n)$  will be said to be *open at a position  $k$*  with  $n < k$  if there is no position  $n'$  with  $n < n' \leq k$  which carries an answer  $\delta(n')$  to that attack. In particular, an attack upon a formula  $X \neg v$  remains open at all later places. A *D-dialogue* shall be a dialogue  $\delta, \eta$  satisfying the following properties (D10)–(D13) :

- (D10)  $P$  may assert an atomic formula only after it has been asserted by  $Q$  before: if  $\delta(n) = Pa$  and  $a$  is atomic then there exists  $m$  such that  $m < n$  and  $\delta(m) = Qa$ .
- (D11) If, at a position  $p-1$ , there are several open attacks suitable to be answered at  $p$ , then only the *latest* of them may be answered at  $p$  : if  $\eta(p) = [n, D]$  and if  $n < n' < p, n'-n \equiv 0 \pmod{2}$ ,  $\eta(n') = [m', A]$  then there exists  $p'$  such that  $n' < p' < p, \eta(p') = [n', D]$ .
- (D12) An attack may be answered at most once: for every  $n$  there exists at most one  $p$  such that  $\eta(p) = [n, D]$ .
- (D13) A  $P$ -formula may be attacked at most once: if  $m$  is even then there exists at most one  $n$  such that  $\eta(n) = [m, A]$ .

A  $D$ -dialogue is said to be *won by  $P$*  if it is finite, ends with an even position and if the rules do not permit  $Q$  to continue with another attack or answer. In that case the last position carries an atomic formula asserted by  $P$ .

The importance of  $D$ -dialogues rests in the fact that the formulas for which there exist  $D$ -strategies are precisely those provable in intuitionistic logic. This follows from the following, stronger

**EQUIVALENCE THEOREM.** *There exist recursive algorithms which, for every formula  $v$ , transform a proof of the sequent  $\Rightarrow v$  in Gentzen's calculus LJ (for intuitionistic logic) into a  $D$ -strategy — and vice versa.*

Contrary to first appearances, a proof of this theorem is by no mean obvious; it cannot be pursued here and may be found in Felscher [1981; 1985].

An *E-dialogue* shall be a  $D$ -dialogue satisfying the additional condition that  $Q$  can react only upon the immediately preceding utterance of  $P$ :

- (E) For every  $n$  in  $\text{def}(\delta)$ : if  $n$  is odd then  $\delta(n)$  is either attack upon  $\delta(n-1)$  or answer to  $\delta(n-1)$ .

An  $E$ -dialogue is said to be *won by  $P$*  if, again, it is finite, ends with an even position and if now the rules for  $E$ -dialogues do not permit  $Q$  to continue

with either an attack or an answer. There will be occasion to refer to the following result which is auxiliary to the proof of the Equivalence Theorem.

**EXTENSION LEMMA.** *There is a recursive algorithm by which every  $E$ -strategy can be embedded into a  $D$ -strategy.*

It follows from this lemma that the Equivalence theorem holds also for  $E$ -strategies in place of  $D$ -strategies.

Readers not familiar with the use of dialogues may appreciate the following *examples* in which  $a, b, \dots$  are assumed to be atomic formulas.

- (1a)
- |    |  |                          |
|----|--|--------------------------|
| 0. | $P(a \wedge b) \rightarrow (a \wedge b)$ |                          |
| 1. | $Q(a \wedge b)$                          | $[0, A]$                 |
| 2. | $P \wedge_1$                             | $[1, A]$                 |
| 3. | $Qa$                                     | $[2, D]$                 |
| 4. | $P \wedge_2$                             | $[1, A]$                 |
| 5. | $Qb$                                     | $[4, D]$                 |
| 6. | $P(a \wedge b)$                          | $[1, D]$                 |
| 7. | $Q \wedge_1$                             | $[6, Q]$                 |
| 8. | $Pa$                                     | $[7, D]$                 |
|    |  | 7. $Q \wedge_2$ $[6, Q]$ |
|    |  | 8. $Pb$ $[7, D]$         |

- (1b)
- |    |  |                          |
|----|--|--------------------------|
| 0. | $P(a \wedge b) \rightarrow (a \wedge b)$ |                          |
| 1. | $Q(a \wedge b)$                          | $[0, A]$                 |
| 2. | $P(a \wedge b)$                          | $[1, D]$                 |
| 3. | $Q \wedge_1$                             | $[2, A]$                 |
| 4. | $P \wedge_1$                             | $[1, A]$                 |
| 5. | $Qa$                                     | $[4, D]$                 |
| 6. | $Pa$                                     | $[3, D]$                 |
|    |  | 3. $Q \wedge_2$ $[2, A]$ |
|    |  | 4. $P \wedge_2$ $[1, A]$ |
|    |  | 5. $Qb$ $[4, D]$         |
|    |  | 6. $Pb$ $[3, D]$         |

Here we have two different  $D$ -strategies for the same formula.

- (2a)
- |    |                               |          |
|----|-------------------------------|----------|
| 0. | $P(a \rightarrow \neg\neg a)$ |          |
| 1. | $Qa$                          | $[0, A]$ |
| 2. | $P\neg\neg a$                 | $[1, D]$ |
| 3. | $Q\neg a$                     | $[2, A]$ |
| 4. | $Pa$                          | $[3, A]$ |

- (2b)
- |    |                               |          |
|----|-------------------------------|----------|
| 0. | $P(\neg\neg a \rightarrow a)$ |          |
| 1. | $Q\neg\neg a$                 | $[0, A]$ |
| 2. | $P\neg a$                     | $[1, A]$ |
| 3. | $Qa$                          | $[3, A]$ |

The first example is a  $D$ -strategy. In the second example,  $P$  cannot win if  $(D11)$  shall not be violated.

- (3)
- |    |                                      |          |
|----|--------------------------------------|----------|
| 0. | $P((a \wedge \neg a) \rightarrow b)$ |          |
| 1. | $Q(a \wedge \neg a)$                 | $[0, A]$ |
| 2. | $P \wedge_1$                         | $[1, A]$ |
| 3. | $Qa$                                 | $[2, D]$ |
| 4. | $P \wedge_2$                         | $[1, A]$ |
| 5. | $Q \neg a$                           | $[4, D]$ |
| 6. | $Pa$                                 | $[5, A]$ |

This is a  $D$ -strategy. The same reasoning holds for  $P \neg(a \wedge \neg a)$ .

- (4)
- |    |  |          |
|----|--|----------|
| 0. | $P((a \rightarrow a) \rightarrow b) \rightarrow b$ |          |
| 1. | $Q(a \rightarrow a) \rightarrow b$                 | $[0, A]$ |
| 2. | $P(a \rightarrow a)$                               | $[1, A]$ |
| 3. | $Qb$   | $[2, D]$ |
| 4. | $Pb$   | $[1, D]$ |
| 5. | $Qa$   | $[2, A]$ |
| 6. | $Pa$   | $[5, D]$ |
- |    |      |          |
|----|------|----------|
| 3. | $Qa$ | $[2, A]$ |
| 4. | $Pa$ | $[3, D]$ |
| 5. | $Qb$ | $[2, D]$ |
| 6. | $Pb$ | $[1, D]$ |

This is a  $D$ -strategy. If we omit positions 5 and 6, we still obtain an  $E$ -strategy.

- (5a)
- |    |  |          |
|----|--|----------|
| 0. | $P((a \rightarrow b) \rightarrow a) \rightarrow a$ |          |
| 1. | $Q(a \rightarrow b) \rightarrow a$                 | $[0, A]$ |
| 2. | $P(a \rightarrow b)$                               | $[1, A]$ |
| 3. | $Qa$   | $[2, D]$ |
| 4. | $Pa$   | $[1, D]$ |
- |    |      |          |
|----|------|----------|
| 3. | $Qa$ | $[2, A]$ |
|----|------|----------|

The left  $E$ -dialogue is won by  $P$  but not the right one. There is no strategy as long as  $(D11)$  shall not be violated.

- (5b)
- |    |  |          |
|----|--|----------|
| 0. | $P \neg \neg(((a \rightarrow b) \rightarrow a) \rightarrow a)$ |          |
| 1. | $Q \neg(((a \rightarrow b) \rightarrow a)$                     | $[0, A]$ |
| 2. | $P((a \rightarrow b) \rightarrow a) \rightarrow a)$            | $[1, A]$ |
| 3. | $Q(a \rightarrow b) \rightarrow a$                             | $[2, A]$ |
| 4. | $P(a \rightarrow b)$   | $[3, A]$ |
- |    |      |          |
|----|------|----------|
| 5. | $Qa$ | $[4, D]$ |
| 6. | $Pa$ | $[3, D]$ |
- |    |  |          |
|----|--|----------|
| 5. | $Qa$   | $[4, A]$ |
| 6. | $P((a \rightarrow b) \rightarrow a) \rightarrow a$ | $[1, A]$ |
| 7. | $Q(a \rightarrow b) \rightarrow a$                 | $[6, A]$ |
| 8. | $Pa$   | $[7, D]$ |

This an  $E$ -strategy and is easily extended to a  $D$ -strategy.

- (6)
- |     |  |        |   |       |                              |
|-----|--|--------|---|-------|------------------------------|
| 0.  | $P((a \rightarrow b) \rightarrow (a \rightarrow c)) \rightarrow (a \rightarrow (b \rightarrow c))$ |        |   |       | 0                            |
| 1.  | $Q(a \rightarrow b) \rightarrow (a \rightarrow c)$   |        |   | [0,A] | 1                            |
| 2.  | $P(a \rightarrow (b \rightarrow c))$   |        |   | [1,D] | 0                            |
| 3.  | $Qa$   |        |   | [2,A] | 1                            |
| 4.  | $P(b \rightarrow c)$   |        |   | [3,D] | 0                            |
| 5.  | $Qb$   |        |   | [4,A] | 1                            |
| 6.  | $P(a \rightarrow b)$   |        |   | [1,A] | 2                            |
|     |  |        |   |       |                              |
| 7.  | $Qa$   | [6,A]  | 3 | 7.    | $Q(a \rightarrow c)$ [6,D] 1 |
| 8.  | $Pb$   | [7,D]  | 2 | 8.    | $Pa$ [7,A] 2                 |
| 9.  | $Q(a \rightarrow c)$   | [6,D]  | 1 | 9.    | $Qc$ [8,D] 1                 |
| 10. | $Pa$   | [9,A]  | 2 | 10.   | $Pc$ [5,D] 0                 |
| 11. | $Qc$   | [10,D] | 1 | 11.   | $Qa$ [6,A] 3                 |
| 12. | $Pc$   | [5,D]  | 0 | 12.   | $Pb$ [11,D] 2                |

This is again a  $D$ -strategy. If we omit positions 9–12 on the left branch and positions 11–12 on the right branch then we obtain an  $E$ -strategy. The numbers appearing to the right of the values of  $\eta$  are the *orders* of the respective assertions as they will be defined in Section 3.3.

- (7) Let  $d_0$  be the formula  $f \wedge \neg f$  for some (atomic)  $f$ .

- |    |   |       |   |       |                |
|----|---|-------|---|-------|----------------|
| 0. | $P(a \wedge ((b \rightarrow a) \rightarrow d_0)) \rightarrow c$ |       |   |       | 0              |
| 1. | $Q(a \wedge ((b \rightarrow a) \rightarrow d_0))$               |       |   | [0,A] | 1              |
| 2. | $P \wedge_1$  |       |   | [1,A] |                |
| 3. | $Qa$  |       |   | [2,D] | 1              |
| 4. | $P \wedge_2$  |       |   | [1,A] |                |
| 5. | $Q(b \rightarrow a) \rightarrow d_0$                            |       |   | [4,D] | 1              |
| 6. | $P(b \rightarrow a)$  |       |   | [5,A] | 2              |
|    |   |       |   |       |                |
| 7. | $Qd_0$  | [6,D] | 1 | 7.    | $Qb$ [6,A] 3   |
|    |   |       |   | 8.    | $Pa$ [7,D] 2   |
|    |   |       |   | 9.    | $Qd_0$ [6,D] 1 |

This can be completed so as to become a  $D$ -strategy;  $Qd_0$  can be handled as in example (3), and on the left branch we then will have to add the steps occurring in the right branch as positions 7 and 8.

- (8) Let  $d_0$  be as in (7) and define recursively for  $i = 0, 1, \dots$

$$e_i = a \wedge ((b \rightarrow a) \rightarrow d_i) \rightarrow c,$$

$$d_{i+1} = e_i \rightarrow d_0.$$

Example (7) then gives a  $d$ -strategy for  $e_0$  and, moreover, shows that there is a  $D$ -strategy for each  $e_{i+1}$  with  $Qd_i$  appearing in the positions 7 and 9 respectively. The  $D$ -strategy for  $e_i$  then contains assertions of orders up to  $i+3$ .

- (9)
- |     |  |       |
|-----|--|-------|
| 0.  | $P((a \rightarrow b) \rightarrow v) \rightarrow (c \rightarrow (b \rightarrow a))$ |       |
| 1.  | $Q(a \rightarrow b) \rightarrow v$   | [0,A] |
| 2.  | $P(c \rightarrow (b \rightarrow a))$   | [1,D] |
| 3.  | $Qc$   | [2,A] |
| 4.  | $P(a \rightarrow b)$   | [1,A] |
| 5.  | $Qa$   | [4,A] |
| 6.  | $P(b \rightarrow a)$   | [3,D] |
| 7.  | $Qb$   | [6,A] |
| 8.  | $Pa$   | [7,D] |
| 9.  | $Qv$   | [4,D] |
| 10. | $Pb$   | [5,D] |

The left-dialogue satisfies (D10), (D12), (D13), and observing these rules  $Q$  has no possibility to continue it. The rule (D11) is violated at place 6. If the formula  $v$  is chosen suitably then the dialogues can be extended so as to obtain a strategy, e.g., if  $v$  is  $a \wedge b$  or  $c \rightarrow a$  or  $(a \rightarrow d) \rightarrow d$ . However, for these choices of  $v$  already a  $D$ -strategy for the initial contention can be found.

- (10)
- |    |   |       |
|----|---|-------|
| 0. | $P(\neg(a \rightarrow b)) \rightarrow (a \vee d)$ |       |
| 1. | $Q\neg(a \rightarrow b)$                          | [0,A] |
| 2. | $P(a \vee d)$                                     | [1,D] |
| 3. | $Q\vee$   | [2,A] |
| 4. | $P(a \rightarrow b)$                              | [1,A] |
| 5. | $Qa$  | [4,A] |
| 6. | $Pa$  | [3,D] |
| 7. | $Qc$  | [4,D] |

Also this dialogue satisfies (D10), (D12), (D13) and violates (D11).

- (11)
- |    |  |       |
|----|--|-------|
| 0. | $P(a \wedge ((\neg a \rightarrow b) \rightarrow c)) \rightarrow c$ |       |
| 1. | $Q(a \wedge ((\neg a \rightarrow b) \rightarrow c))$               | [0,A] |
| 2. | $P\wedge_1$  | [1,A] |
| 3. | $Qa$   | [2,D] |
| 4. | $P\wedge_2$  | [1,A] |
| 5. | $Q(\neg a \rightarrow b) \rightarrow c$                            | [4,D] |
| 6. | $P(\neg a \rightarrow b)$  | [5,A] |

7. $Q\neg a$ [6,A] 8. $Pa$ [7,A] 9. $Qc$ [6,D] 10. $Pc$ [1,D]	7. $Qc$ [6,D] 8. $Pc$ [1,D] 9. $Q\neg a$ [6,A] 10. $Pa$ [9,A]
--	--

In the left dialogue,  $Q$  violates (D11) at place 9. If this place and the following place 10 are omitted, a  $D$ -strategy remains.

(12)

0. $P(\neg(b \rightarrow (c \vee d))) \rightarrow (a \rightarrow b)$	
1. $Q\neg(b \rightarrow (c \vee d))$	[0,A]
2. $Pa \rightarrow b$	[1,D]
3. $Qa$	[2,A]
4. $P(b \rightarrow (c \vee d))$	[1,A]
5. $Qb$	[4,A]
6. $Pc \vee d$	[5,D]
7. $Q\vee$	[6,A]
8. $Pb$	[3,D]

$P$  violates (D11) at place 8.

## 2 THE LITERATURE ON DIALOGUES

### 2.0

It was P. Lorenzen who, in addresses in 1958 and 1959, published as Lorenzen [1960; 1961], proposed the idea that an autonomous foundation of intuitionistic logic should be based on the concepts of a dialogue and of a strategy for dialogues. Emphasizing the autonomy of such a foundational approach, Lorenzen preferred to speak of a *constructive* or *effective* logic and avoided the more familiar name of *intuitionistic* logic. While the first descriptions of dialogues seemed to use only the properties named here (D00)–(D02), it soon became clear that additional rules would be required if *only* intuitionistically provable formulas should be those which could be secured by strategies for dialogues. Such additional rules were formulated by Lorenz [1961] who defined (among other types) the kind of dialogues called  $D$ -dialogues here; they appear in Lorenzen [1962; 1967] and in Kamalah and Lorenzen [1967]. While these presentations attempted to arrive at an appropriate definition for a dialogue by specializing the general notion, a different approach was taken in [Lorenz, 1973] and [Lorenzen and Schwemmer, 1973] where there is considered at first a very narrow type of dialogue (permitting both  $P$  and  $Q$  to react only upon the immediately preceding step) which then is liberalised to types of dialogues which are, essentially, the  $E$ -dialogues of Section 1.

Lorenzen's basic idea is indeed a very attractive one. However, although described in a variety of articles and books, its presentations have been marred by a recurrence of ambiguous definitions and incomplete, if not erroneous proofs. And all these elaborations so far have suffered from two major defects.

### 2.1

The first defect is of a technical-mathematical character. If a new development of intuitionistic logic, based on the concepts of dialogues and strategies, shall be given, then one expects an equivalence theorem to be established which states that provability by strategies coincides with provability by one of the known calculi for intuitionistic logic. The proof of such a theorem remained missing for many years.

A first attempt to prove an equivalence theorem was made in Lorenz's dissertation [Lorenz, 1961]; it was repeated in [Lorenz, 1968]. A certain part of Lorenz's dissertation was corrected in [Stegmüller, 1964]; some claims made in other parts were proved while other ones were refuted in the Diplomarbeit of W. Kindt [1970]. Kindt's refutations were acknowledged in footnote 12 of [Lorenz, 1968] where it is said that a correction of the erroneous statements in [Lorenz, 1961] would require "ein paar detaillierte technische Vorbereitungen" (cf. also a similar remark in footnote No. 16); unfortunately, these few, detailed technical preparations have never been presented to the public, and the gaps in Lorenz's attempt still appear to be unfilled. (It is somewhat distressing that in the presumably authoritative collection of [Lorenzen and Lorenz, 1978] the article [Lorenz, 1968] is simply reprinted together with its footnotes; the part of [Lorenz, 1961] to which footnote No.12 refers has been omitted altogether.)

A first correct proof of an equivalence theorem was given by Kindt [1972]; however, the dialogues studied by Kindt are *not*  $D$ -dialogues but employ instead of  $(D11)$  a different rule. In [Lorenzen and Schwemmer, 1973, pp. 59 and 71] it is observed that the  $E$ -strategies (in the sense of Section 1) considered there give rise to a calculus of 'Dialogstellungen' which (at least in the propositional case) may be transformed into a calculus of Beth-tableaux such that provability by  $E$ -strategies implies intuitionistic provability. A new, and simpler approach to an equivalence theorem for  $D$ -dialogues was developed by Haas [1980] and it seems that the technical gaps contained in this work are only minor and can actually be filled. An attempt to prove an equivalence theorem for  $E$ -dialogues is contained in [Mayer, 1981] and Dr E. C. W. Krabbe informs me that an equivalence theorem for  $E$ -dialogues is contained in [Krabbe, 1982] and in [Barth and Krabbe, 1982].

The equivalence theorem stated in Section 1 was presented in [Felscher, 1981] and in a revised form in [Felscher, 1985].



## 2.2

The second defect from which the elaborations of Lorenzen's idea have suffered concerns the matter of foundations. The rules (D00)–(D01) are just precise descriptions of the intention that dialogues should proceed through applications of the argumentation forms. But when additional rules (the 'Dialog-Rahmenregeln' in the terminology of the Lorenzen school) had to be imposed, the question arose whether such rules could be explained as natural codifications of principles of argumentation. Such principles, if foundationally sound, would have to be based on an analysis of the use of dialogues as a means to establish a systematical, indisputable and convincing conduction of formal arguments.

When Lorenzen [1960; 1961] wrote about 'Sprachspiele', i.e., language games, he used this word in the sense of the ancients'  $\alpha\gamma\omega\nu$ , referring to a regulated (linguistic) process, the rules of which were to be governed by an insight which, although not further explained, was clearly assumed to be present. This attitude changed with [Lorenz, 1961] who, attempting a mathematical formalisation, began to make use of the concepts of a mathematical discipline known as the Theory of Games. Games there are mathematical objects, describing procedures as varied as whist and rummy at the one end and the games invented by warriors and economists at the other end. The rules then may be arbitrary: what matters is that they are adhered to; and the convention that a game is won because the other player can't draw any more may be brought about by rather odd rules of the game (such as, e.g., the categorical application of an equaliser). Matters were not improved when Lorenz [1961] observed that a change of dialogue rules would give rise to a type of dialogue the strategies for which would prove precisely the *classically* provable formulas. For this situation made it perfectly clear that the mathematical arbitrariness of the Theory of Games, being a tool to describe formally such different ways of reasoning as are classical logic and intuitionistic logic, could not possibly produce a philosophical foundation for either one of them. The mathematical apparatus for the Theory of Games was used heavily in the mathematical work of Kindt [1970; 1972]. On the other hand, phrases referring to 'dialogue games' have spread through a certain kind of literature where a mathematical terminology is borrowed in order to give at least the appearance of conceptual precision.

It appears that Lorenzen himself did not follow the fashion of a game-theoretical reduction. However the foundational discussions presented, e.g. in [Lorenzen and Schwemmer, 1973] are not based on an argumentative analysis; in particular, atomic statements and their negations are discussed with respect to a semantical distinction of true and false, and the difference between classical and intuitionistic logic is made to depend on the decidability ('Wahrheits-Definitheit': definiteness with respect to truth) of atomic statements. As crown's evidence for the generally unsatisfactory state of

a foundational discussion I may quote Kambartel [1979] who, in the very conclusion of this article, writes in respect to the problem of justification:

... Rechtfertigungsproblem. Dieses besteht darin, die *schematischen* Dialogspiele selbst von einem *argumentativen* Gebrauch der logischen Partikeln her zu begründen. Die dialogische Logik hat dieses Problem bisher dadurch überspielt, dass sie die Dialogspiele methodisch als 'erste' Festlegung des Gebrauchs der logischen Partikeln behandelt. Das dabei vernachlässigte, über die schematische Ebene hinausführende Rechtfertigungsproblem schlägt dann spätestens in der Rahmenregeldiskussion wieder durch. In der Tat werden dort 'Rechtfertigungen' für die Wahl solcher Regeln, z.B. neuerdings immament schematisch oder, wie zunächst geschehen, im eher intuitiven Rückgriff auf halb-schematisch analysierte Beispiele, beigebracht. Weder 'technische' Kriterien noch die Verallgemeinerung von Beispielen stellen aber bereits einen im engeren Sinne *normativen* Zugang zur Logik dar ...

### 2.3

Lorenzen's argumentation forms have also been put to use by K. J. J. Hintikka, but this with quite different intentions. Hintikka, beginning with [Hintikka, 1968], developed what he calls a *game-theoretical semantics*, and a more recent series of articles on this topic have been collected in [Saarienen, 1979]. A *semantical game* in Hintikka's terminology may indeed be viewed as a dialogue in the sense of Section 1 although Hintikka restricts his attention to the single argumentation forms and nowhere cares to formulate game rules proper (such that the implied reference to mathematical games remains but an incantation). The point, however, is that Hintikka is concerned with a linguistical analysis of natural languages and *not* with a foundation of (classical or intuitionistic) logic. For this purpose, argumentation forms and dialogues are used as tools for the semantical evaluation of logically composite expressions (which may be more complex than first-order logic would permit to express) in domains governed by classical logic. There is, therefore, no connection of Hintikka's work with that discussed in the present article.

## 3 FOUNDATIONS OF DIALOGUES

In this Section I shall develop an argumentative foundation for the use of particular types of dialogues, the *D*-dialogues, as a basis for intuitionistic logic. That such a foundation is wanted was outlined in subsection 2.2.

### 3.1 *The Argumentative Interpretation of Logical Operators*

There exists a well known provability interpretation of the logical operations (connectives and quantifiers) which then may be considered as being represented by Gentzen's calculi of proofs, i.e., natural deduction and the sequent calculus LJ; for details, cf. van Dalen's article in this volume. In the same manner, the argumentation forms may be viewed as expressing an *argumentative interpretation* of the logical operations, and as far as the positive connectives and the quantifiers are concerned this interpretation is obvious enough. Concerning implication,  $Y$  attacks  $Xw_1 \rightarrow w_2$  by offering  $Yw_1$  as an *admission* (or *local hypothesis*) and  $X$  may react by either answering with  $Xw_2$  or attacking  $Yw_1$  (provided  $w_1$  is composite). Concerning negation, the situation is the same as in the case of the provability interpretation: if external, semantical references to truth and falsity shall be avoided, we must enrich the basic concept of provability by adding either refutability or absurdity as a primitive notion. Since we are aiming for intuitionistic logic, we introduce a constant  $\Delta$  symbolising *absurdity* and then understand  $\neg w$  as an abbreviation of  $w \rightarrow \Delta$ . The principle of *ex absurdo quodlibet* takes as its first form that he,  $X$ , who is forced to assert  $\Delta$  then must concede, without further argument, any assertion made by  $Y$ . A speaker professing  $\Delta$  thus brings himself into a position precluding any further debate, and so we may just as well omit this fatal step and state, as the second form of *ex absurdo quodlibet*, that  $\Delta$  must not be asserted. This then explains why an attack  $Yw$  upon  $Xw \rightarrow \Delta$ , i.e.,  $X\neg w$ , cannot be answered.

### 3.2 *Basic Principles for Dialogues*

Gentzen's calculi of proofs are easily explained in that they represent the weakest consequence relation for which the provability interpretation is valid. The connection between dialogues and the argumentative interpretation of logical operations is (not only more complicated but also) located on a different level: it is not the dialogues but the *strategies* for dialogues which will correspond to proofs. I thus formulate the *basic purpose* for the use of dialogues:

- (A<sub>0</sub>) Logically provable assertions shall be those which, for *purely formal* reasons, can be upheld by a strategy covering every dialogue chosen by  $Q$ .

The dialogue rules (D00)–(D02), describing the use of argumentation forms, simply produce the (linguistic) material of the dialogue which then will have to be organised by the dialogue rules proper. Extending the intentions expressed in the formulation of the argumentation forms, I formulate the *argumentative intent* in the pursuit of a single dialogue:

- (A<sub>1</sub>) A dialogue is, on the part of  $Q$  an attempt to put into doubt (to refute) the initial assertion made by  $P$  ; it is, on the part of  $P$ , an attempt to uphold this assertion, and if  $P$  succeeds in doing so this will mean that  $P$  wins the dialogue.

In the light of these intentions we now will have to clarify

- (b<sub>0</sub>) how to determine the dialogue rules proper,  
 (b<sub>1</sub>) the notion that  $P$  wins a dialogue.

It must be emphasised that the concepts occurring here *cannot* be studied separately but must be analysed *simultaneously* and in constant regard of the purpose (A<sub>0</sub>).

### 3.3 Dependence, Positive Dependence and Order

The notions to be discussed in this subsection are auxiliary. Let  $\delta, \eta$  be a dialogue. I shall say that a statement *depends directly* on an earlier statement if it is either an attack upon or an answer to that statement; I define *dependence* to be the transitive and reflexive relation generated by direct dependence. Dependence, therefore, is an order relation, contained in the linear order given by  $\delta$  ; since every statement, different from the initial one, depends directly on *exactly one* earlier statement, it follows that dependence defines the ordering of a *tree* on the set of all statements, i.e., on  $\text{im}(\delta)$  with  $\delta(0)$  as its top node. I define a *chain* to be a sequence of statements in  $\text{im}(\delta)$  such that each of its members, except the first one, depends directly on its predecessor in that sequence; every chain  $\kappa$  thus arises from a path in the dependence tree by removing the nodes above  $\kappa(0)$  from the path. Every chain is a subsequence of  $\delta$  and  $\delta$  itself is pieced together from various chains, some of which may only have one member.

While dependence is a relation defined between arbitrary statements, a second relation will be defined only between assertions of a dialogue  $\delta, \eta$ . An assertion  $Xv$  is an *immediate positive dependent* of an earlier assertion  $Xw$  if it is an answer to an attack upon  $Xw$  ; I define *positive dependence* to be the transitive and reflexive relation generated by immediate positive dependence.

The relation of positive dependence leads to the following classification of assertions. The initial assertion and its positive dependence shall be *of order 0* ; if  $Xv \rightarrow w$  or  $X\neg v$  is *of order n* then an attack  $Yv$  shall be *of order n+1* , and so shall be the positive dependence of this attack. It follows that the  $P$ -assertions are exactly those of even order and the  $Q$ -assertions are exactly those of odd order.

### 3.4 Contentions and Hypotheses

When asserting the initial statement of a dialogue,  $P$  contends it to be defensible. In this sense, the initial assertion is *contended*, and in the same manner assertions of order 0 are contended. Assertions of order 1, if they do not arise as positive dependents of earlier ones, are attacks  $Qw_1$  upon assertions  $Pw_1 \rightarrow w_2$  or  $P\neg w_1$ ; hence they are (global) *hypotheses* offered by  $Q$ , and then also their positive dependents are (particularisations of) hypotheses. Assertions of order 2, if they do not arise as positive dependents of earlier ones, are attacks  $Pw_2$  upon hypotheses  $Qw_1 \rightarrow w_2$  or  $Q\neg w_1$ ; here  $P$  takes up the hypotheses by admitting  $Pw_1$  as a (higher order) contention. Consequently, also these assertions are contended by  $P$ , and so are their positive dependents. Repeating this argument, it follows that *all*  $P$ -assertions are contended and that *all*  $Q$ -assertions are hypothetical. It will be advisable to observe the distinction made between global hypotheses in a dialogue as discussed here, and local hypotheses occurring as admissions in instances of argumentation forms.

Applying the argumentation forms, assertions of logically composite formulas are dissolved into assertions of lesser complexity: contentions are *upheld* and hypotheses are *developed*. Obviously, contentions  $Pw_1 \wedge w_2$ ,  $Pw_1 \vee w_2$ ,  $P\forall xw$ ,  $P\exists xw$  are upheld by holding up the immediate positive dependents (as chosen by  $P$  or prescribed by  $Q$ ), and the same holds for the development of the analogous hypotheses. Consider now contentions  $Pw_1 \rightarrow w_2$  where we include the case  $P\neg w_1$  by writing it as  $Pw_1 \rightarrow \Delta$ ; when attacked by  $Qw_1$  then  $P$  may either uphold the answer  $Pw_2$  or attack  $Qw_1$  in order to force  $Q$  into a further development of this hypothesis. Similarly, if a hypothesis  $Qw_2 \rightarrow w_2$  is attacked by  $Pw_1$  then  $Q$  may develop it into the answer  $Qw_2$  or  $Q$  may attack the contention  $Pw_1$ .

The process of dissolving logically composite assertions comes to an end once atomic assertions have been reached. Atomic formulas asserted by  $Q$  are hypotheses, intended by  $Q$  as describing particular situations which serve to refute  $P$ 's contention to have a defensible initial assertion; they do *not* need further justification. Atomic formulas asserted by  $P$ , however, remain contentions *in need* of justification. Securing them by material insight, such as, e.g., illumination or revelation, is unacceptable (anyway and in particular) if *purely formal* defensibility has been claimed by  $P$ . There remains, therefore, only one possibility for  $P$  to assert an atomic formula  $Pa$  for purely formal reasons:

- (c<sub>0</sub>)  $P$  may assert  $Pa$  only if  $Q$  admits  $Qa$  as a hypothesis *relevant* to the position of  $Pa$ .

For in that case  $P$  confronts  $Q$  with its own hypothesis to which  $Q$  cannot possibly object (in [Krabbe, 1982] this principle is mentioned with the ap-

appropriate name of *ipse dixisti*). This principle ( $c_0$ ) is purely descriptive in that it refers to a given, completed dialogue; it does not provide the means in order to enforce that, already during the performance of the dialogue, assertions  $Pa$  are made *only* in observance of ( $c_0$ ). Making reference to the linear structure of a dialogue, we therefore strengthen ( $c_0$ ) to

( $c_1$ )  $P$  may assert  $Pa$  only *after*  $Q$  has admitted  $Qa$  at an earlier, relevant position.

It appears that we now would have to clarify the notion of *relevance*, and a first attempt to do so would consist in producing a definition which describes, for every position of a dialogue, the set of hypotheses relevant for this position. I shall not proceed in this manner; rather, I shall introduce additional restrictive rules for dialogues with the effect that the family of *all* hypotheses occurring in a dialogue becomes *coherent* in the sense that its members, being admitted simultaneously, *do not* create distinctions of relevance: each of them is relevant for all atomic contentions asserted afterwards. The additional rules will, obviously, restrict the amount of information analysed in a dialogue. But the principal objects for us are strategies, not methods for winning a single dialogue, and no information will be lost if it only remains available within the system of dialogues belonging to a strategy.

### 3.5 How to Win a Dialogue

$P$  wins a dialogue if it succeeds in holding up its contentions. During the course of a dialogue, the initial contention is dissolved into more and more specialised subcontentions, and this specialisation comes to an end with the contention of atomic formulas as regulated by ( $c_1$ ). It is implicit in this conception that no composite contention is accepted as being upheld without further dissolution, for  $Q$  may always challenge it with an attack. Consider now a dialogue containing an atomic contention  $Pa$  which, for the moment, we assume as being of order 0. We then find a unique sequence of contentions, beginning with the initial contention  $Pv$  and ending with  $Pa$ , each of which (except the first one) is an immediate positive dependent of its predecessor. Consequently,  $Pa$  is the final step of a process by which  $Pv$  is narrowed down to more special contentions — and in view of ( $c_1$ ) this final step may be asserted for purely formal reasons. Of course, other sequences of specialisations of  $Pv$ , ending with other atomic formulas, may be possible, but since we are considering provability by strategies, these other possibilities are covered by other dialogues which a strategy will have to take into account. We thus arrive at

(WA) The initial contention is considered as having been upheld successfully if  $P$  has asserted an atomic contention of order 0.

Obviously, the same reasoning can be used to say that a contention of order  $n$  can be considered as having been upheld successfully if  $P$  has narrowed it down to an atomic contention of order  $n$  — but this observation remains without consequences.

There is, however, another way to consider the initial contention as having been upheld successfully; it rests on the principle of *ex absurdo quodlibet* applied to dialogues (and not only to argumentation forms). As a simplest situation, consider a hypothesis  $Qw_1$  of order 1, arising as an attack upon  $Pw_1 \rightarrow w_2$  of order 0, and assume that, after a sequence of positive dependents of  $Qw_1$ , the only possibility left to  $Q$  would be the assertion of an absurdity  $Q\Delta$ . In that case, the hypothesis  $Qw_1$  has been developed (and we may assume: by  $P$ 's prodding) into an absurdity and, therefore, has itself been shown as untenable. But as this hypothesis had been granted in the attack  $Qw_1$ , we conclude by *ex absurdo quodlibet* that  $Pw_1 \rightarrow w_2$  can be upheld *without* any further argument. More generally, a hypothesis  $Qr_1$  of order  $n+1$ ,  $n > 0$ , arising as an attack upon  $r_1 \rightarrow r_2$  of order  $n$ , is itself a development of the earlier hypothesis  $Qs_1 \rightarrow s_2$  of order  $n-1$  which gave rise to an attack  $Ps_1$  of which  $Pr_1 \rightarrow r_2$  is a positive dependent. Consequently, if  $Qr_1$  can be shown as leading to an absurdity, then also  $Qs_1 \rightarrow s_2$  must be considered as leading to an absurdity. Descending from  $n-1$  to 1, we conclude that already the first hypothesis  $Qw_1$  of order 1, giving rise to the higher order hypotheses resulting in  $Qs_1 \rightarrow s_2$ , leads to an absurdity, and thus again  $Pw_1 \rightarrow w_2$  can be upheld without further argument. We thus arrive at

(*WB*) The initial contention is considered as having been upheld successfully if ( $P$  has not asserted an atomic contention of order 0 but)  $Q$  has been brought into a position where its only possibility to continue would be the assertion of an absurdity.

Of course, at the present stage of our discussion no reason is visible why  $Q$  may become so restricted in its possibilities as is supposed in (*WB*); this will become clear after the following sections. What *can* be said already here is that if a dialogue is won by  $P$  then its last position is even (i.e., the last move is  $P$ 's), and if the initial contention does not contain a negation then the dialogue can be won only according to (*WA*).

### 3.6 Ramifications

I shall now discuss how to avoid the distinctions of relevance as they appear in condition ( $c_0$ ). Let us begin by considering more closely the situation described there: let  $\delta, \eta$  be a dialogue with positions  $j, k$  such that  $j < k$ ,  $\delta(j) = Qa$ ,  $\delta(k) = Pa$  where  $a$  is atomic. Since both these assertions depend on  $\delta(0)$ , there exist chains  $\kappa_0, \kappa_1$  from  $\delta(0)$  to  $\delta(j)$  and to  $\delta(k)$ , and it follows from  $j < k$  that  $\kappa_1$  cannot be an initial part of  $\kappa_0$ . Consequently,

either  $\kappa_0$  is an initial part of  $\kappa_1$  or  $\kappa_0, \kappa_1$  ramify at a position preceding  $\delta(j)$ . The first case is unproblematical. For let  $n$  be the last argument of  $\kappa_0$  such that  $\kappa_0(n) = \delta(j)$ ; being an atomic hypothesis,  $\kappa_0(n)$  must be an attack upon  $\kappa_0(n-1) = Pa \rightarrow w$  which then is answered by  $\kappa_1(n+1) = Pw$ , for this is the only way in which the atomic hypothesis  $Qa$  can have dependents in  $\kappa_1$ . Thus  $Pa$  is  $Pw$  or a dependent of  $Pw$ , and  $Qa$  must certainly be considered as relevant for  $Pa$ . In the second case, however, the ramification may cause distinctions of relevance.

I shall say that there is a *ramification at  $n$*  for two chains  $\kappa_0, \kappa_1$  of a dialogue if both chains coincide for all  $i$  such that  $i \leq n$ , if further both chains are defined (at least) for  $n+1$  and if  $\kappa_0(n+1)$  is different from  $\kappa_1(n+1)$ , i.e., if  $\kappa_0(n+1), \kappa_1(n+1)$  are stated at different positions of the dialogue. A ramification then arises in one of the three following ways:

- (1)  $\kappa_0(n+1), \kappa_1(n+1)$  are different attacks upon  $\kappa_0(n)$ ;
- (2)  $\kappa_0(n+1), \kappa_1(n+1)$  are different answers to the attack  $\kappa_0(n)$  upon  $\kappa_0(n-1)$ ;
- (3)  $\kappa_0(n)$  is an attack  $Yw_1$  upon  $\kappa_0(n-1) = Xw_1 \rightarrow w_2$ ,  $\kappa_i(n+1)$  is an attack upon  $\kappa_0(n)$  and  $\kappa_{1-i}(n+1)$  is the answer  $Xw_2$  to  $\kappa_0(n)$ ,  $i = 0, 1$ .

The attacks in (1) shall be called *distinct* if not only their positions but also the statements made by them are different; otherwise they are only *repeated attacks*. In the same way, I shall speak of *distinct* and of *repeated* answers. In this subsection I shall be concerned with ramifications of the first two types.

Distinct *attacks* are possible upon assertions  $Xw_1 \wedge w_2, X\forall xw$ . If  $P$  contends, say,  $Pw_1 \wedge w_2$  then  $P$  certainly should be able to contend both  $Pw_1$  and  $Pw_2$ . But a hypothesis arising during the analysis of  $Pw_1$  (e.g., if  $w_1$  is  $a \rightarrow a$ ) will *not* be relevant during the analysis of  $Pw_2$  (e.g., if  $w_2$  is  $b \rightarrow a$ ) and vice versa: such hypotheses are admitted for *one* but not for the *other* subcontention. Distinct attacks upon contentions, therefore, do cause distinctions of relevance. In a strategy, however, the possibility of distinct attacks by  $Q$  is already taken into consideration in that it leads to different dialogues which all have to be won by  $P$ . Consequently, there will be no loss of information if such distinct attacks are excluded from every single dialogue. On the other hand, a hypothesis  $Qw_1 \wedge w_2$  is present already *before* any ramification caused by different attacks and it should remain *before* in effect with its *complete content* also after the ramification. If, for instance, during the development of  $Qw_1$  the moves of  $P$  lead  $Q$  to admit  $Qa$  then this hypothesis should be considered as coherent with  $Qw_1$  as well as with  $Qw_1 \wedge w_2$  and also with  $Qw_2$ :  $Qa$  should be relevant also for any development of  $Qw_2$ .

Distinct *answers* can be given to attacks upon assertions  $Xw_1 \vee w_2, X\exists xw$ . If  $P$  contends, say,  $Pw_1 \vee w_2$  then it will have to contend only *one* of



$Pw_1, Pw_2$ . Again now, hypotheses arising during the analysis of one of these answers (e.g., if  $w_1$  is  $\neg a$ ) will not be relevant during the analysis of the other (e.g., if  $w_2$  is  $a$ ). Distinct answers to attacks upon contentions, therefore, do cause distinctions of relevance. However, if  $P$  has answered an attack with, say,  $Pw_1$  then a second, later answer with  $Pw_2$  would be useful only if  $Pw_1$  could not be upheld successfully and if  $P$  now would try a second attempt. But in a strategy we can demand that  $P$  knows what it is doing and does not proceed by trial and error, and there will be no loss of information if we exclude distinct answers to attacks upon contentions. On the other hand, if  $Q$  were to answer an attack upon, say  $Qw_1 \vee w_2$  with both  $Qw_1$  and  $Qw_2$  then it would grant not only the content of  $Qw_1 \vee w_2$  but actually that of  $Qw_1 \wedge w_2$ . In a strategy, however,  $P$  has to provide dialogues for all possible answers, and there will be no loss of information if we also exclude distinct answers to attacks upon hypotheses.

In order to discuss repeated attacks and repeated answers, we may now assume that the two chains  $\kappa_0, \kappa_1$  ramifying at  $n$ , are chosen as being maximal, i.e., as branches of the dependence tree. I now define inductively the notion of *corresponding couples*:  $\kappa_0(0), \kappa_1(0)$  form a corresponding couple; if  $\kappa_0(i), \kappa_1(i)$  form a corresponding couple then  $\kappa_0(i+1), \kappa_1(i+1)$  shall form a corresponding couple if they either appear at the same position of the dialogue or if they are (at least) identical as signed expressions *and also* are identical in their *mode* within the dialogue, i.e., as attacks or answers referring to  $\kappa_0(i), \kappa_1(i)$ . Let us assume now that the ramification at  $n$  arises under repetitions. Then  $\kappa_0(n+1), \kappa_1(n+1)$  still form a corresponding couple, and we may look for all corresponding couples  $\kappa_0(n+i), \kappa_1(n+i)$ . If these couples exhaust already one of the two chains, say  $\kappa_0$ , then that part of  $\kappa_0$  which starts at  $n+1$  does not contain any information which is not available in the corresponding part of  $\kappa_1$ . No information will be lost if that part of  $\kappa_0$  (or the corresponding part of  $\kappa_1$ ) is omitted from the dialogue, and so we may exclude the repetition at  $n+1$  which gave rise to the twofold presence of that part.

It remains to consider the case that there exists a common argument  $m$  of  $\kappa_0, \kappa_1$ ,  $n+1 < m$ , such that  $\kappa_0(m), \kappa_1(m)$  is not a corresponding couple. Let  $m$  be minimal for this property and observe that either  $P$  or  $Q$  acts at  $m$  in both of  $\kappa_0, \kappa_1$ . If  $P$  attacks at  $m$ , say  $Qw_1 \wedge w_2$  with  $Pw_1$  in the one and with  $Pw_2$  in the other chain, then these distinct attacks could have been carried out *without* the repetitive ramification at  $n$  (and with an acceptable ramification at the position of  $Qw_1 \wedge w_2$  instead). If  $P$  answers at  $m$ , say  $Qv$  with  $Pw_1$  and  $Pw_2$ , then the repetitive ramification at  $n$  just hides the fact that there is actually a ramification caused by distinct answers. The same types of ramifications occur if  $Q$  in both  $\kappa_0, \kappa_1$  attacks  $m$  or answers at  $m$ . Finally, if either  $P$  or  $Q$  attacks  $m$  in the one and answers  $m$  in the other chain then the repetitive ramification at  $n$  delays an actual ramification of type (3). Again now, *either* this ramification of type (3) could have been

carried out already at  $n$  or there are restrictions on the execution of such ramifications (as they actually will be discussed in the next section) and the repetitive ramification at  $n$  is employed in order to circumvent these restrictions.

Consequently, in *all* cases repetitive ramifications either can be avoided immediately or lead to distinctions of relevance of the sort discussed already for distinct ramifications. In any case, therefore, they may be excluded without loss of information.

The conditions effecting the exclusion of different answers and of different attacks upon contentions are precisely (D12) and (D13).

### 3.7 *Nested Attacks and Nested Answers*

There are good reasons why, in a dialogue, a certain answer upon an attack by  $Q$  is not stated immediately but only after some delay. The necessity to observe ( $c_1$ ) will cause such situations if

a contention  $Pw_1 \rightarrow w_2$  has been attacked and the admitted hypothesis  $Qw_1$  needs further elaboration in order to permit  $P$  either to assert  $Pw_2$  or to force  $Q$  into an absurdity,  
 or a contention, itself stated already under a hypothesis, has been attacked and now the earlier hypothesis needs further elaboration.

Delayed answers, therefore, cannot be excluded. It is such delays which cause a nesting of attacks and, thereby, a nesting of answers.

Let  $\delta, \eta$  be a dialogue and let  $\delta(m), \delta(n)$  be assertions such that  $m-n \equiv 0 \pmod{2}$ , let  $\delta(i)$  be an attack upon  $\delta(m)$  and let  $\delta(j)$  be an attack upon  $\delta(n)$  (whence also  $i-j \equiv 0 \pmod{2}$ ). I shall call  $\delta(j), \delta(i)$  a pair of *nested attacks* if  $j < i$  and if the attack  $\delta(j)$  is still open at  $i$ ; in that case  $\delta(i)$  is the *inner* attack and  $\delta(j)$  is the *outer* attack. I shall speak of *H-nested attacks* if the inner attack is a hypothesis; in that case, both  $m, n$  are even,  $\delta(m)$  is of the form  $Pw_1 \rightarrow w_2$  or  $P\neg w_1$  and  $\delta(i)$  is  $Qw_1$ .

Consider a pair of *H-nested attacks* as above and let  $Ps$  be the contention which  $P$  would have to state in order to answer the outer attack. The question then arises whether the hypothesis  $Qw_1$  can be considered as relevant for  $Ps$  and its dependents, since  $P$  contended the assertion  $\delta(n)$  *before*  $Qw_1$  has been admitted. For instance, if  $\delta(n)$  is  $Pr \rightarrow s$  then, by contending it,  $P$  states that it is prepared to hold up  $Ps$  under the hypothesis  $Qr$  — an additional hypothesis such as  $Qw_1$  is, at this stage, not available. Obviously, hypotheses available at the position  $n$  may be used as well, and we also can assume that distinctions of relevance have already been excluded at least as far as the positions  $n, m$ . The complete content now of  $\delta(m)$ , i.e., of  $Pw_1 \rightarrow w_2$  or  $P\neg w_1$  is, again, that  $P$  is prepared to hold up  $Pw_2$

(or, for that matter, even  $P\Delta$ ) under the hypothesis  $Qw_1$ . Therefore, the hypothesis  $Qw_1$  should be counted as being relevant if  $P$  actually fulfils the obligation to assert  $PW_2$  (or  $P\Delta \dots$ ):  $Qw_1$  together with  $Pw_2$  preserves the complete content of  $\delta(m)$ . We thus arrive at the principle

- ( $d_0$ ) For every pair of  $H$ -nested attacks: the hypothesis granted by the inner attack may be used as relevant in order to contend the answer to the outer attack *provided*  $P$  also contends the answer to the inner attack.

The dialogue won by  $P$  in example (9) satisfies ( $d_0$ ) for the  $H$ -nested attacks with  $n = 1$ ,  $m = 4$ ,  $j = 3$ ,  $i = 5$ ; the hypothesis  $Qw_1 = Qa$  is used in order to contend the answer  $\delta(6)$  to  $\delta(j)$  at 8, and conversely also this answer itself is employed in order to develop the hypothesis  $Qb$  which is needed when contending the answer  $\delta(10)$  to  $\delta(i)$ .

The principle ( $d_0$ ) is purely descriptive; it does not provide the means in order to enforce that only the situation described as desirable occurs. If the inner attack is upon  $P\neg w_1$  then  $P$  will never be able to fulfil the obligation expressed in ( $d_0$ ), and if we wish to avoid an additional label declaring  $Qw_1$  to be irrelevant for  $Ps$  then another formulation is wanted. But also if  $\delta(m)$  is  $Pw_1 \rightarrow w_2$  with  $w_2 \neq \Delta$ , a different formulation would be useful: the example (10) shows a pair of  $H$ -nested attacks violating ( $d_0$ ), but  $P$  has won the dialogue in accordance with ( $WA$ ) and  $Q$ , if it is left to respect ( $D12$ ), ( $D13$ ), cannot continue: there just is no position left for  $P$  to state the answer to the inner attack. We thus formulate a *rule* which forces  $P$  to contend this answer *before* it makes use of the hypothesis:

- ( $d_1$ ) For every pair of  $H$ -nested attacks: the inner attack must have been answered before the outer attack may be answered.

In this manner, we now have also a well-defined nesting of answers.

Consider now a pair of nested attacks upon contentions which is not  $H$ -nested; this means that the inner attack is symbolic. The statement of such an attack does not create any hypotheses possibly needed for the contention of the answer  $Ps$  to the outer attack. Still, there may be various reasons for  $P$  *not* to answer the outer attack  $\delta(j)$  immediately at  $j + 1$ , but to delay this answer to a position following that of the inner attack: actually the observance of ( $d_1$ ), together with that of ( $c_1$ ) may be one such reason. The example (12) communicated to me by Dr. E. C. W. Krabbe, shows a dialogue satisfying ( $D10$ ), ( $D12$ ), ( $D13$ ) and ( $d_1$ ), but the outer attack at 3 is answered at 8 while the inner attack at 7 remains open. Given the outer attack 3, also the hypothesis  $\delta(5)$  arises as an inner attack, and observance of ( $d_1$ ) forces  $P$  to state the contention  $\delta(6)$  before  $\delta(5)$  can be used in order to state the answer  $Pb$  to the outer attack. In avoiding the answer to the inner attack at 7,  $P$  now *fails* to uphold the contention  $\delta(6)$ . Thus ( $d_1$ ) has

been observed, but the promise to uphold  $\delta(6)$  is given lip service only. It follows from this example that, in order to observe the full meaning of  $(d_0)$  we have to strengthen the rule  $(d_1)$  to

$(d_2)$  For every pair of nested attacks upon contentions: the inner attack must have been answered before the outer attack may be answered.

As for nested attacks upon hypotheses, it follows from the definition of a strategy that, if  $P$  has a strategy at all, then  $P$  *a fortiori* has a strategy respecting

$(d'_2)$  For every pair of nested attacks upon hypotheses: the inner attack must have been answered before the outer attack may be answered.

There is reason to conjecture that also the converse implication holds, and a proof should be related to that of the Extension Lemma mentioned in Section 1. The idea of such a proof is illustrated by example (11): violating  $(D11)$ ,  $Q$  may try to withhold a certain hypothesis (e.g., absurdity), but  $P$  knows from the strategy how the answer to the outer attack had to be treated if it was stated immediately after this attack.

### 3.8 *D-Dialogues*

It was the purpose of the last two subsections to look for restrictions on dialogues which would permit us to avoid distinctions of relevance. We thus arrived at the rules  $(D12)$ ,  $(D13)$  in subsection 3.6 and at rule  $(D11)$  which is the conjunction of  $(d_1)$  and  $(d'_2)$  in subsection 3.7. Having imposed these rules, let us look once more at the family of hypotheses occurring in a dialogue. Clearly, every positive dependent of a hypothesis  $Qw$  is nothing but a specification or an instantiation of  $Qw$ ; and dependents of higher order, coming from intermediary contents, keep this character as well. Passing through a chain of dependents, we see that the hypotheses occurring there form a coherent set, i.e., a set of simultaneously admitted assumptions without distinctions of relevance. Different chains of dependents are joined together with the help of delayed attacks and answers, and the three types of ramifications which thus may arise were described at the beginning of subsection 3.6. The presence of our rules now insures that all those ramifications are excluded which would cause distinctions of relevance. What remains permitted are ramifications of type (3) with properly nested answers and ramifications of type (1) caused by different attacks upon a hypothesis. If  $P$  attacks a hypothesis  $Qw$  a first time, it forces from  $Q$  a certain system of specifications of the hypothesis  $Qw$  asserted by  $Q$  in the beginning; if  $P$  attacks  $Qw$  a second time then it may obtain a different system of specifications which, nevertheless, still is a system of specifications of this *same* hypothesis  $Qw$ : once  $Q$  has admitted  $Qw$  then it must

bear the consequences of being forced into specifications, and the specified hypotheses brought forward under the different answers given by  $Q$  to the *same* attack of  $P$  upon  $Qw$  express, when considered simultaneously, *more* than is contained in  $Qw$  alone, and thus they *do* create distinctions of relevance. It must be emphasised that this difference in the effects of attacks and answers is fundamental.

We thus find that, in the presence of  $(D11)$ ,  $(D12)$ ,  $(D13)$ , the set of *all* hypotheses occurring in a dialogue is coherent in the sense that no distinctions of relevance appear. Consequently, the condition  $(c_1)$  becomes  $(D10)$ , and our dialogues are  $D$ -dialogues.

At this point, a methodological observation appears to be appropriate. In the preceding two subsections I have presented arguments resulting in the introduction of additional rules with the purpose to avoid distinctions of relevance ( $(D12)$ ,  $(D13)$ ,  $(d_0)$ ). In subsection 3.6 the exclusion of certain moves in a dialogue was supported with the observation that no information will be lost if the possibilities, excluded from single dialogues, remain present in strategies; in subsection 3.7 the introduction of  $(d_0)$  was explained with the necessity to preserve the complete content. It should be noticed very clearly that these argumentations are based on an *informal understanding of purposes*; they are *not* justifications based on mathematical theorems. As a matter of fact, as long as we abstain from a formal definition of relevance, we cannot even formulate a theorem saying that strategies for dialogues *with* precautions on relevance (?) prove the same formulas as do strategies for dialogues with  $(D12)$ ,  $(D13)$ ,  $(d_0)$ .

### 3.9 How to Win a $D$ -dialogue

I have defined in subsection 3.5 what it means that  $P$  wins a dialogue. On the other hand, there is the purely formalist definition, taken from the literature and mentioned in Section 1, that a  $D$ -dialogue is won by  $P$  if  $Q$  has no way to continue. It remains to be shown that both notions are equivalent with respect to strategies.

If a  $D$ -dialogue has been won according to  $(WB)$  then it also has been won in the formal sense. If a  $D$ -dialogue has been won according to  $(WA)$  then it contains an atomic contention  $Pa$  of order 0 which, therefore, must be an answer. Consequently,  $Q$  cannot continue the dialogue at this position by referring to  $Pa$ . We now can show:

If for some formula  $v$ ,  $P$  has a strategy to win with  $(WA)$ ,  $(WB)$  all  $D$ -dialogues for  $v$  then  $P$  also has a strategy to win in the formal sense all  $D$ -dialogues for  $v$ . For consider a strategy the branches of which are won with  $(WA)$ ,  $(WB)$ . These branches could not be continued by  $Q$  if  $Q$  would have to respect the rules for  $E$ -dialogues. We thus obtain an  $E$ -strategy if, where necessary, we cut off ends of branches when  $Q$  begins to violate the

rule (*E*). It then follows from the Extension Lemma that this *E*-strategy may be extended again to a *D*-strategy.

For the converse implication, it can even be shown that a *D*-dialogue  $\delta, \eta$  which is won in the formal sense is won also according to (*WA*), (*WB*). As this is clear if, during the dialogue, *P* has stated an atomic contention of order 0, we may assume now that *P* has *not* stated an atomic contention of order 0; (*WB*) then will hold if we can prove that *Q* could continue if it were to state an absurdity: there still should exist a (last) open attack of the form  $P\neg w$ . Before continuing the proof, it will be useful to prepare some auxiliary notions.

Let  $\delta, \eta$  be a dialogue; let  $n$  be an odd position and assume that *Q* has already stated  $\delta(n)$ . Then  $\xi_0(n)$  shall be the number of contentions which, after this statement, still may be attacked by *Q* at later positions, and  $\xi_1(n)$  shall be the number of contentions which, being attacks, still may be answered by *Q* at later positions — and here I expressly *include* the unspeakable answers  $Q\Delta$  to attacks  $Pw$  upon some  $Q\neg w$ . The number  $\xi(n) = \xi_0(n) + \xi_1(n)$  is called the *degree* (of freedom) of  $n$ . The following characterisation will be useful:

$\xi(n)$  is the difference  $a(n) - d(n)$  where  $a(n)$  is the number of attacks upon hypotheses  $Qw_1 \rightarrow w_2$ ,  $w_1$  not atomic, which are contended before  $n$ , and  $d(n)$  is the number of atomic *P*-answers contended before  $n$ . This follows from the following observations in which  $2i+1$  is a position of  $\delta, \eta$ . If  $\delta(2i)$  is an attack then  $\xi(2i-1) \leq \xi(2i+1) \leq \xi(wi-1) + 1$ , and the right inequality becomes equality if, and only if,  $\delta(2i)$  is an attack upon a hypothesis  $Qw_1 \rightarrow w_2$  such that  $w_1$  is not atomic (whereas  $w_2$  may be  $\Delta$ ). If  $\delta(2i)$  is an answer which is not atomic then  $\xi(2i-1) = \xi(2i+1)$ . If  $\delta(2i)$  is an answer which is atomic then  $\xi(2i-1)$  must be positive (for otherwise *Q* could not act at  $2(i+1)$ ), and  $\xi(2i+1) = \xi(2i-1) - 1$ .

Consider now a *D*-dialogue. It follows from (*D12*) (*D13*) that every assertion has *at most one* atomic immediate positive dependent; consequently, every assertion has *at most one* atomic positive dependent. Let now  $Pa$  be an atomic contention of *positive* order; there then exists a first (highest) contention  $Pw_a$  of which  $Pa$  is a positive dependent, and  $Pa, Pw_a$  have the same order. By the preceding remark,  $Pw_a$  is unique, and as it has no positive predecessor, it must be an attack upon a hypothesis  $Qw_a \rightarrow u_a$  (where  $u_a$  may be  $\Delta$ ).

I now resume the proof where it was interrupted. Let  $2i$  be the last position of the dialogue. Then  $\delta(2i)$  cannot be a composite formula (since that could be attacked by *Q*) nor can it be a symbolic attack (since that could be answered); thus it must be an atomic contention  $Pa$ , necessarily of positive order. Let  $Pw_a$  be the unique, highest contention determined by  $Pa$  as above. If  $Pw_a$  is  $Pa$  then  $u_a$  must be  $\Delta$  (for otherwise *Q* could answer), and thus (*WB*) holds. Assume now that  $Pw_a$  is different from  $Pa$ ; then  $w_a$  is not atomic. It now will be sufficient to show that  $\xi(wi-1)$  is positive

— because the possibilities numbered by  $\xi(i-1)$  cannot comprise attacks or answers which actually could be stated (for in that case  $Q$  could still continue), and thus there must be at least one open attack to be answered only by absurdity. I thus have to prove that the difference  $a(2i-1)-d(2i-1)$  is positive. Let  $Pb$  be an atomic answer which contributes to  $d-(2i-1)$ ; it is of positive order and thus determines the unique attack  $Pw_b$ . Since  $Pb$  is an answer and  $Pw_b$  is an attack,  $Pb$  must be a proper dependent of  $Pw_b$ . Thus  $w_b$  is not atomic and, therefore, the attack  $Pw_b$  contributes to  $a(2i-1)$ . It follows that the map  $\varphi$  sending  $b$  into  $w_b$  is an injection of the set of contentions contributing to  $d(2i-1)$  into the set of contentions contributing to  $a(2i-1)$ . Since  $w_a$  is not atomic, the latter set contains  $Pw_a$ ; the former set, however, does not contain  $Pa$ . Consequently,  $a(w_i-1)$  is strictly larger than  $d(2i-1)$ .

### 3.10 Intuitionistic versus Classical Logic

As was mentioned in Section 2, Lorenz [1961] has observed that a change in the rules for  $D$ -dialogues produces a class of dialogues which I shall call  $C$ -dialogues, such that the formulas provable by  $C$ -strategies are precisely the classical provable formulas. The change leading from  $D$ -dialogues to  $C$ -dialogues consists in

cancelling (D11) and (D12) for  $P$ , but leaving them in effect for  $Q$ .

(If I understand Lorenz's and Lorenzen's writings correctly then they seem to demand the cancellation for  $P$  of (D12) only; the examples (2b) and (5a) show that this would not suffice.) It is not hard to see that  $C$ -strategies prove *only* classically provable formulas. For the case of *propositional* logic, the converse implication (i.e., every classically provable formula can be proved by a  $C$ -strategy) can be seen as follows. Observe first that an intuitionistically provable formula, being provable by a  $D$ -strategy, is trivially provable by a  $C$ -strategy. It is well known that if  $w$  is a classically provable formula then  $\neg\neg w$  is intuitionistically provable; assume now that  $w$  is not intuitionistically provable. Every  $D$ -dialogue  $\delta, \eta$  for  $\neg\neg w$ , won by  $P$ , begins with attacks  $\delta(1) = Q\neg w, \delta(2) = Pw$ , and since we assume that the part beginning at position 2 is *not* a  $D$ -dialogue for  $w$  won by  $P$ , there must be positions below 2 at which  $P$  attacks  $\delta(1)$  again. If we compare the branches in the dependence tree and look for the first positions at which they differ, we will find, refining the discussion in subsection 3.6, that this happens at contentions which could be obtained *without* the repetitive first part of the branch if repeated answers or answers in disregard of (D11) were permitted to  $P$ . Permitting such answers in  $C$ -dialogues, it can be shown that a  $D$ -strategy for  $\neg\neg w$  can be rebuilt into a  $C$ -strategy for  $w$ .

The mathematical fact that  $C$ -strategies can be used for classical logic is, in principle, not surprising; other proof-theoretical systems, e.g., those of Hilbert-type, can also be used for many varieties of logics. Rather than leading to amazement over the universal applicability of a mathematical tool (trees and strategies), this situation should teach us to emphasise the fundamental differences between intuitionistic and classical logic.

For the provability interpretation, as represented by Gentzen's calculi, Curry [1963, p. 260] has attempted a provability *explanation* of classical negation with help of his concept of *complete absurdity*, but this hardly will be considered to be a conceptual *foundation*. For the argumentative approach presented here, classical logic *cannot* be given a foundation by simply changing formal details of a foundation for intuitionistic logic. If we want to explain the rules governing classical negation then there appears to be no way to avoid the semantical notions of *true* and *false*: without these notions we cannot explain why distinctions of relevance may be discarded as it is done when  $P$  is permitted to repeat answers and to disregard ( $D11$ ). Thus, for classical logic, the entire conceptual frame employed for the foundation of intuitionistic strategies, has to be abandoned: there is no use for contentions and hypotheses, for defendability by purely formal reasons and for considerations of relevance. What *is* required, is a completely different conceptual framework, based on the notions of true and false and on the distribution of truth-values under logical operations. The foundation of classical logic within such a framework is well known, and the elegant formulation of classical tableaux due to Smullyan [1968] may easily be read as to depict a dialogue-strategy leading to a failure of the attempt to falsify a formula. Again, the argumentative explanation of winning a dialogue according to  $(A)$ ,  $(WB)$  is only formally related to the closure of branches in Smullyan's tableaux which *always* means the advent of absurdity.

#### 4 APPENDIX: CONCEPTS CONNECTED WITH THE EQUIVALENCE THEOREM

The equivalence theorem, formulated in Section 1, states the existence of certain transformations between strategies and proofs in the calculus LJ; the proof of this theorem cannot be presented here. It may, however, be instructive for the reader to become familiar with some concepts which originally were developed for this proof. For details which have to be suppressed here I refer to Felscher [1981; 1985].

The reader will have noticed that among the examples, listed at the end of Section 1, there is none which treats a formula with quantifiers. But this is no serious omission since the argumentation forms for quantifiers are, so to speak, the direct generalisations of the forms for conjunction and disjunction to the infinite case. For strategies, however, this treatment of quantifiers has



the effect that there may occur infinite ramifications: if  $S, \delta, \eta$  is a strategy and if a node  $e$  carries as  $\delta(e)$  either a formula  $P\forall xw$  or an attack  $P\exists$  upon a formula  $Q\exists xw$  then the tree  $S$  has an infinite ramification at  $e$  — every term  $t$  determines a lower neighbour of  $e$ , carrying either an attack  $Qt$  or an answer  $Qw(t)$ . Although the branches of  $S$  must be finite (as follows from (S0)), the strategy itself is an infinite object. It is obvious that this is a clear disadvantage of strategies as compared to the more usual notions of proof. I now shall abstract a *finite* object from a strategy, its skeleton.

Let  $H$  be a class of dialogues as in Section 1. An  $H$ -skeleton for a formula  $v$  is a triplet  $S, \delta, \eta$  with the same properties as an  $H$ -strategy for  $v$  *except* that in (S1) certain nodes  $e$  are excepted and, instead, are covered by

- (S1<sub>e</sub>) If  $\delta(e)$  is  $P\forall xw$  then only one lower neighbour of  $e$  carries an attack upon  $\delta(e)$ , and this attack is  $Qy$  where  $y$  is a variable not occurring free in any *expression*  $\delta(h)$  with  $h \geq e$ . If  $\delta(e)$  is an attack  $P\exists$  upon  $Q\exists xw$  then only one lower neighbour of  $e$  carries an answer, and this answer is  $Qw(y)$  where  $y$  is a variable not occurring free in any *expression*  $\delta(h)$  with  $h \geq e$ .

As is usual, the variable  $y$  will be called the *eigenvariable* in these situations. It is clear that every  $H$ -strategy contains various  $H$ -skeletons, and it is not hard to see that, conversely, every  $H$ -skeleton can be extended to an  $H$ -strategy. This observation has the important consequence that it suffices to consider  $H$ -skeletons which, having finite trees, are more easily handled in induction proofs. For instance, the Extension Lemma of Section 1 is proved in the form that every  $E$ -skeleton can be extended to a  $D$ -skeleton.

Unfortunately,  $E$ -skeletons still have certain undesirable properties. Consider the example of a formula  $\exists xa \rightarrow \exists xa$  where  $a$  is atomic; there are two  $E$ -strategies, viz.

0. $P\exists xa \rightarrow \exists xa$		0. $P\exists xa \rightarrow \exists xa$	
1. $Q\exists xa$	[0,Q]	1. $Q\exists xa$	[0,A]
2. $P\exists$	[1,A]	2. $P\exists xa$	[1,D]
3. $Qa(y)$	[2,D]	3. $Q\exists$	[2,A]
4. $P\exists xa$	[1,D]	4. $P\exists$	[1,A]
5. $Q\exists$	[4,A]	5. $Qa(y)$	[4,D]
6. $Pa(y)$	[5,D]	6. $Pa(y)$	[3,D]

In the right skeleton, the attack at 3 is answered with the substitution term  $y$  at 6; this answer must be delayed because the choice of the substitution term depends on the eigenvariable  $y$  appearing at 5. There are no phenomena of an analogous type in, say, the sequent calculus; in Lorenzen and Schwemmer [1973] and in Haas [1980], where an informal use of  $E$ -skeletons is made, the possibility that this situation might occur has been overlooked.

In order to circumvene this difficulty, I introduce the concepts of a *formal dialogue* and of a *formal strategy*, making use of the *formal argumentation forms* for  $\forall$  and  $\exists$ :

$Q\forall$ : assertion: $Q\forall xw$ attack: $Pt$ answer: $Qw(t)$	$P\forall$ : assertion: $P\forall xw$ attack: $Qy$ ( <i>eigenvariable</i> ) answer: $Pw(y)$
$Q\exists$ : assertion: $Q\exists xw$ attack: $P\exists$ answer: $Qw(y)$ ( <i>eigenvariable</i> )	$P\exists$ : assertion: $P\exists xw$ attack: $Qt$ answer: $Pw(t)$ .

I define a *formal E-dialogue* in exactly the same way in which I defined an *E-dialogue*, only now in (D01), (D02) the formal argumentation forms are used for quantifiers *and* the eigenvariable condition is imposed at the position indicated. The adjective *formal* then refers to the fact that, contrary to the intuitive understanding, in the attack  $Qt$  the term  $t$  is stated already by  $Q$ ; eigenvariables chosen at a later position then *must* respect these expressions  $Qt$ . I define a *formal E-strategy* in the same way in which I defined an *E-strategy*, but now with formal dialogues instead *and* with the following changes: there is *only one* possibility for  $Q$  taken into account for

answering an attack  $P\exists$     (case  $Q\exists$ ),  
 making an attack  $Qy$         (case  $P\forall$ ),  
 making an attack  $Qt$         (case  $P\exists$ ).

It then is obvious that every formal *E-strategy* can be transformed into an *E-skeleton*; it can be shown that, conversely, every *E-skeleton* can be transformed into a formal *E-strategy*.

It is the formal *E-strategies* which can be set into correspondence with LJ-proofs.

It follows from these observations that the disadvantage of dialogues consisting in

1. the treatment of quantifiers as infinite conjunctions and disjunctions disregarding Frege's discovery of finitary quantifier rules made possible by the use of free variables, and
2. the ensuing appearance of infinite strategies

is only apparent. It arose because we wanted to use the *same* argumentation forms (concerning quantifiers) for *both*  $P$  and  $Q$ ; it could have been avoided if, from the outset, we would have studied strategies instead of dialogues. This illustrates once more the difficulty, mentioned at the beginning of Section 3.2, that it is not the dialogues but the strategies which correspond to proofs: working with dialogues, we have to describe *in advance* the branches

of strategies which themselves are to be defined only with the help of these dialogues.

It also should be observed that, in contrast to Gentzen's calculi for provability, strategies and dialogues do *not* appear as natural representations of the relation of *provability from hypotheses* but only as those of the relation of *absolute provability*. Of course, if  $M$  is a finite set of sentences and  $m$  is a conjunction of these sentences then, for every sentence  $w$ , the sequents

$$M \Rightarrow w \quad \text{and} \quad \Rightarrow m \rightarrow w$$

are simultaneously derivable in LJ, and this permits us to reduce the provability of  $w$  from the hypotheses  $M$  to the absolute provability of  $m \rightarrow w$ . It also is obvious that a dialogue, discussing the derivability of  $M \Rightarrow w$ , should begin with an initial list of the  $Q$ -formulas determined by  $M$ , followed (or preceded) by the  $P$ -formula  $Pw$ . But no general rule on how to proceed from this initial list can be stated as long as we want to keep the alternation between  $P$  and  $Q$  during the progress of our dialogue. If  $a$  is atomic, a sequent such as  $a \Rightarrow a \vee w$  produces the initial list  $Qa, Pa \vee w$  which *must* be followed by an attack of  $Q$ ; on the other hand, a sequent such as  $a \wedge w \Rightarrow a$  produces the initial list  $Qa \wedge w, Pa$  which *must* be followed by an attack of  $P$ . Certainly, regulations circumventing these difficulties may be formulated, but apparently only at the cost of a loss in intuitive appeal.

*Obernau/Neckar*

## BIBLIOGRAPHY

- [Barth and Krabbe, 1982] E. M. Barth and E. C. W. Krabbe. *From Axiom to Dialogue*, De Gruyter, Berlin, 1982.
- [Curry, 1963] H. B. Curry. *Foundations of Mathematical Logic*. McGraw-Hill, New York, 1963.
- [Ehrensberger and Zinn, 1997] J. Ehrensberger and C. Zinn. DiaLog — a system for dialogue logic. In *CADE - 13, Conference on Automated Deduction*, Townsville, North Queensland, Australia. Pp. 446–460. Lecture Notes in Artificial Intelligence, Vol. 1249, Springer-Verlag, 1997.
- [Felscher, 1981] W. Felscher. Intuitionistic tableaux and dialogues. Prepared notes, distributed at the conference *The Present State of the Problem of Foundation of Mathematics*, Firenze, June 1981.
- [Felscher, 1985] W. Felscher. Dialogues, strategies and intuitionistic provability. *Annals of Pure and Applied Logic*, **28**, 217–254, 1985.
- [Haas, 1980] G. Haas. Hypothesendialoge, konstruktiver Sequenzenkalkül und die Rechtfertigung von Dialograhmenregeln. In *Theorie des wissenschaftlichen Argumentierens*, C.F. Gethmann, ed. pp. 136–161. Suhrkamp, Frankfurt, 1980.
- [Hintikka, 1968] K. J. J. Hintikka. Language-games for quantifiers. In *Studies in Logical Theory*, pp. 46–72. American Philosophical Quarterly Monograph Series 2, Blackwell, Oxford, 1968.
- [Kambartel, 1979] F. Kambartel. Überlegungen zum pragmatischen und argumentativen Fundament der Logik. In *Konstruktionen versus Positionen*, K. Lorenz, ed. pp. 216–228. de Gruyter, Berlin, 1979.

- [Kamlah and Lorenzen, 1967] W. Kamlah and P. Lorenzen. *Logische Propädeutik*. Bibliograph.Institut, Mannheim, 1967.
- [Kindt, 1970] W. Kindt. *Dialogspiele*. Diplomarbeit, Math. Institut Universität Freiburg, 1970.
- [Kindt, 1972] W. Kindt. Eine abstrakte Theorie von Dialogspielen. Dissertation, Universität Freiburg, 1972.
- [Krabbe, 1982] E.C. W. Krabbe. Studies in dialogical logic. Dissertation, Rijksuniversiteit Groningen, 1982.
- [Lorenz, 1961] K. Lorenz. Arithmetik und Logik als Spiele. Dissertation, Universität Kiel, 1961. Partially reprinted in [Lorenzen and Lorenz, 1978].
- [Lorenz, 1968] K. Lorenz. Dialogspiele als semantische Grundlage von Logikkalkülen. *Archiv Math. Logik Grundlagenforsch*, **11**, 32–55, 73–100, 1968. Reprinted in [Lorenzen and Lorenz, 1978].
- [Lorenz, 1973] K. Lorenz. Die dialogische Rechtfertigung der effektiven Logik. In *Zum normativen Fundament der Wissenschaft*, F. Kambartel and J. Mittelstraß, eds, pp. 250–280. Athenäum, Frankfurt, 1973. Reprinted in [Lorenzen and Lorenz, 1978].
- [Lorenzen, 1960] P. Lorenzen. Logik und Agon. In *Atti Congr. Internat. de Filosofia*, Vol. 4, Sansoni, Firenze, pp. 187–194. Reprinted in [Lorenzen and Lorenz, 1978].
- [Lorenzen, 1961] P. Lorenzen Ein dialogisches Konstruktivitätskriterium. In *Infinitistic Methods*, Proceed. Symp. Foundations of Math, PWN, Warszawa, pp. 193–200, 1961. Reprinted in [Lorenzen and Lorenz, 1978].
- [Lorenzen, 1962] P. Lorenzen. *Metamathematik* Bibliograph.Institut, Mannheim, 1962.
- [Lorenzen, 1967] P. Lorenzen. *Formale Logik*, 2nd ed., de Gruyter, Berlin, 1967.
- [Lorenzen and Lorenz, 1978] P. Lorenzen and K. Lorenz. *Dialogische Logik*, Wissenschaftl. Buchgesellschaft, Darmstadt, 1978.
- [Lorenzen and Schwemmer, 1973] P. Lorenzen and D. Schwemmer. *Konstruktive Logik, Ethik und Wissenschaftstheorie*, Bibliograph.Institut, Mannheim, 1973.
- [Mayer, 1981] G. Mayer. Die Logik im deutschen Konstruktivismus. Dissertation, Universität München, 1981.
- [Saarinen, 1979] E. Saarinen, ed. *Game-Theoretical Semantics*, D. Reidel, Dordrecht, 1979.
- [Smullyan, 1968] R. M. Smullyan. *First Order Logic*. Springer-Verlag, Heidelberg, 1968.
- [Stegmüller, 1964] W. Stegmüller. Remarks on the completeness of logical systems relative to the validity concepts of P. Lorenzen and K. Lorenz, *Notre Dame Journal of Formal Logic*, **5**, 81–112, 1964.

#### EDITOR'S NOTE

The dialogue system of this chapter has recently been implemented as a theorem prover **DIALOG** [1] and **COLOSSEUM** [2]. **DIALOG**, written in Lisp, offers a rule language for redefining the rules of the game. It also supports automatic and interactive proving and has a user-friendly interface. Efficiency, however, has not been the major concern, and therefore, **DIALOG** may serve more as a tool for teaching or for experimenting with the dialogue rules.

**COLOSSEUM** is a no-frills re-implementation of Dialogue Games in Prolog. Its dialogue rules are hardwired for intuitionistic first order predicate logic (as specified by Felscher above). **COLOSSEUM** allows automatic proving only, but it is much faster than **DIALOG** and is web accessible.<sup>1</sup>

<sup>1</sup><http://www8.informatik.uni-erlangen.de/IMMD8/staff/Zinn/Dialogue/Colosseum.html>

## BIBLIOGRAPHY

- [1] J. Ehrensberger and C. Zinn. A system for dialogue logic. In *14th. Int'l Conf. on Automated Deduction (CADE-14)*, number 1249 in LNAI. Springer, 1997.
- [2] C. Zinn. COLOSSEUM – An Automated Theorem Prover for Intuitionistic Predicate Logic based on Dialogue Games. In *Automated Reasoning with Analytic Tableaux and Related Methods (TABLEAUX): Position Papers*. Institute for Programming & Logics, University at Albany - SUNY, TR 99-1, 1999.



## FREE LOGICS

### I: Introduction

#### 1 WHAT ARE FREE LOGICS?

Some theorems of **CQC=**, such as those of the form

$$(1) \quad \exists x(x = \tau)$$

and

$$(2) \quad \varphi[\tau/x] \rightarrow \exists x\varphi,$$

are often accused of introducing into that theory—and thus into the very core of ‘our logic’—undesired ‘existential commitments’. However, the mere derivability of these sequences of symbols can hardly accomplish such a major feat by itself, and even when the theory is supplied with the usual ‘referential’ semantics, metaphysics is still far from being determined one way or another. 1 and 2 certainly require—by way of this semantics—that every singular term of the language receive an interpretation in the domain of quantification, but so what? The formal instrument does not specify the metaphysical counterpart of the relation between a symbol and its interpretation, nor does it tell you which things can or cannot belong to a domain of quantification. The formal instrument is neutral with respect to all these questions, and thus *by itself* cannot introduce any metaphysical commitments, existential or otherwise.

Things get more complicated when one takes into account the ideology most commonly associated with **CQC=** and its referential semantics. Then it becomes very ‘natural’ to think of a singular term as *denoting* its interpretation, hence to read the semantical requirement evoked by 1 and 2 as the requirement that every singular term denote. Even more importantly, if one agrees with Quine that ‘to be is to be a value of a bound variable’<sup>1</sup>—that is, if one assigns ‘existential import’ to quantifiers—the domain of quantification becomes the set of all and only those objects which exist in a given (possible) situation, and the above requirement is drastically strengthened, to the demand that every singular term denote *an existing object*. Now the ontological commitments are certainly apparent, and someone is bound to react to them in the name of logic’s ‘purity’.

---

<sup>1</sup>See for example [Quine, 1939]. In what follows, we will sometimes refer to this statement as *Quine’s dictum*.

Free logics<sup>2</sup> result from this reaction. However, since what they are a reaction to is a very delicate combination of many factors—a certain philosophical understanding of a certain formal interpretation of a certain formal system—it is difficult to say exactly what they are and how far they extend. To say—as is often said—that they are ‘logics free of existence assumptions with respect to their singular terms’ is too vague to be of much help, and also somewhat inaccurate from a historical point of view. For *every* formal system and *every* formal semantics can be free in this sense, given a suitable ideology, but this much tolerance was certainly not in the minds of the people who created free logics.<sup>3</sup> They wanted to *reform* classical logic, and *substitute* for it a *better* instrument, they thought that both the usual formal systems and the usual formal semantics were faulty in important ways, and it is only fair to define free logics so as to make sense of the precise task that they set for themselves.

On the other hand, it would not do to identify free logics with a certain class of theorems. For one thing, there is no *one* such class (as the expression ‘free logics’ should make clear),<sup>4</sup> and there is even some debate as to whether free logics result from restricting or rather extending classical logic.<sup>5</sup> But more importantly, we suggested above that all these modifications—whether restrictions or extensions—would make no sense (and in particular would not be legitimately referred to as free logics) if not in the context of certain interpretations of the formal systems, and of a certain understanding of these interpretations.

And finally, it would be totally unsatisfactory to define free logics in terms of a given semantics, or even a given *class* of semantics. For not only is a formal semantics (as well as a formal system) not enough to characterise the present enterprise in the absence of some ‘intuitive reading’ of it, but also the choice of a semantics is probably the most important question in this area, and we have to be careful not to prejudge such a fundamental issue by a biased *definition*.

Keeping all these reservations in mind will inevitably result in a less than straightforward characterisation of our subject, but the complications we will have to go through will prove instructive. For in this subject more than in others, logic, philosophy of logic and philosophy in general (especially metaphysics) are intertwined in a very delicate way, and it does not hurt if this delicate relation is emphasised right from the beginning.

In conclusion, I propose the following definition. A free logic is a formal system of quantification theory, with or without identity, which allows for some singular terms in some circumstances to be thought of as denoting no

---

<sup>2</sup>This expression was first used by Karel Lambert in 1960.

<sup>3</sup>See for example [Leonard, 1956] and [Lambert, 1967].

<sup>4</sup>Thus ‘free logics’ is the correct expression to refer to the whole subject, but ‘free logic’ is also very common.

<sup>5</sup>In this regard, see van Fraassen’s position sketched in Section 11.



existing object, and in which quantifiers are invariably thought of as having existential import.

A few comments and clarifications are in order. First of all, a terminological matter. The expression ‘thought of’, which occurs twice in the definition, must be regarded as inclusive of both the formal interpretation of the system and the intuitive (or philosophical) reading of this interpretation. When the formal semantics is missing (as was the case in free logics for several years), this ‘thinking of’ reduces entirely to its intuitive component.

Secondly, the definition requires that there be in the language of a free logic expressions construed as singular terms. A language containing no individual constants or descriptions and allowing individual variables to occur only bound in well-formed formulas (and there are languages of this sort for **CQC**, for example some of Quine’s) would hardly satisfy the present requirement.

Thirdly, the definition does not exclude the possibility that every singular term denotes in every circumstance, only that every singular term denotes *an existing object* in every circumstance. There are philosophers (Meinongians for example) who think that there are non-existing objects, and that singular terms may well denote them: the definition is neutral with respect to such views. However, to avoid awkwardness, usually I will refer to singular terms not denoting an existent simply as *non-denoting*.

Fourthly, the definition is concerned not with whether there *actually* are non-denoting singular terms, but only with whether there *may be*. A free logic is after all a *logic*; hence all that it can reasonably care for is logical possibility. When a logic acknowledges the possibility of non-denoting singular terms, we will say that it *allows for* non-denoting singular terms.

Fifthly, not every logic allowing for non-denoting singular terms is a free logic by our definition. In particular, all attempts at saving the formal system (and the formal semantics) of classical logic by some substitutional or Meinongian reading of the quantifiers are ruled out. On the other hand, it is perfectly possible to *add* to a free logic substitutional or Meinongian quantifiers, thus extending its expressive power.

Finally, even though referential semantics played a major role in the discussion above, the definition does not mention this semantics. The reason is that the existential import of quantifiers, and even the distinction between denoting and non-denoting singular terms, can be effectively mimicked in some non-referential semantics (for example, in Leblanc’s truth-value semantics),<sup>6</sup> even if the best way to understand what is going on in these semantics is still to compare them with their referential analogues. Thus the three factors to whose combination a free logic is a reaction come to play different roles in its definition: a free logic is the result of a modification of

---

<sup>6</sup>On this and other alternatives to the standard referential approach, see the chapter by Leblanc in Volume 2 of the present 2nd edition of this *Handbook*.

the *formal system* of **CQC** (or **CQC=**), motivated by a certain *intuitive reading* of it, which is best understood (at least so far) in the context of the usual referential *interpretation* of that system.

## 2 WHY FREE LOGICS?

The most general answer to this question has already been suggested in the discussion preceding my definition of a free logic. Though vigorously attacked from some quarters, the neopositivistic suspicion towards metaphysics is still highly influential in contemporary logic. Whether they regard metaphysics as sheer ‘nonsense’ or as a set of ‘synthetic’ statements to be neatly distinguished from the ‘analytic’ ones constituting their discipline, many logicians like logic to be metaphysically ‘pure’, or not to carry any metaphysical ‘baggage’—as the many debates in the area of quantified modal logic show sometimes quite dramatically. To apply such a general motivation to the present case, it is enough to regard even the simplest existential statements as metaphysical in nature.

However, this motivation by itself does not go very far towards motivating anything close to free logics. As we will see in the next section, classical logic has its own ways of dealing with these matters, and certainly many classical logicians would not accept without a fight the claim—presupposed by the alleged ‘justification’ of free logics suggested above—that classical logic is in any sense existentially committed or metaphysically ‘impure’. To get closer to the justification we are looking for, we need to weaken that claim as follows. Classical logic (if filtered through the usual interpretation, and the usual reading of this interpretation) does not allow for non-denoting singular terms. To be sure, this logic can be *used* in such a way as to avoid any philosophical commitments or any problems resulting from the limitation in question, but this requires the adoption of convoluted and *ad hoc* procedures of translation from natural language into the formal language and back (in a word, of a number of epicycles). Free logics, on the other hand, represent a much more straightforward and direct approach to the same problems: they make the translations easier, they allow expressions of natural language to be taken more often at face value, and they require fewer *ad hoc* assumptions.

This justification is certainly better than the first one, but still, it does not entirely fulfil its purpose. For it does not take into account the fact that the classical logician can shape his philosophy of language so as to make it fit his logic perfectly (and make his logic the most ‘natural’ thing in the world): Russell’s position—to be mentioned briefly in the next section—is in this respect typical. And this makes it clear once and for all that the adoption of some specific view in the philosophy of language is an essential step towards the justification of free (and perhaps all) logics.

There is a whole spectrum of such views that would do the job nicely, ranging from an extremely ‘metaphysical’ one to an extremely ‘pragmatic’ one. For the sake of illustration, let me briefly discuss these two extremes.

The ‘metaphysical’ extreme states simply that in natural language *there are* non-denoting singular terms. A singular term is an expression that *purports to* denote a single object, and many a singular term fails to achieve this purpose. Nonetheless, they are still singular terms: ‘Pegasus’ is as much a singular term as ‘Caesar’ or ‘3’, and ‘the winged horse’ or ‘the round square’ are as much singular terms as ‘the President of the USA in 2000’. Hence no formal system can give a faithful representation of the structure of natural language (and so be reasonably applied to it) if it does not allow for non-denoting singular terms.

The ‘pragmatic’ extreme, on the other hand, regards the *real existence* of non-denoting singular terms in natural language as totally irrelevant. Whether there are or there aren’t any, there are contexts in which some people use expressions *as* singular terms without assuming that they denote anything, or maybe even in the process of wondering whether they denote or not. For example, an attempt by a person to prove that God exists—or that ‘God’ denotes—might be conceived as a case in point. Whether these people are right or not, a logic allowing for non-denoting singular terms would also allow for a more direct and faithful representation (and evaluation) of their reasoning in those contexts. So this logic would be an instrument of wider and simpler applicability than classical logic, and would not prejudge important issues which it is inappropriate for logic to decide.

Of course, the classical logician can be expected to have responses to these motivations. It is certainly not news that in philosophy, or anywhere else, you can’t get something valuable for nothing. In the present case, this suggests that you need a position in between the two above extremes to transform the fear of metaphysical commitment so well entrenched in most contemporary logicians into a defence of free logics.

### 3 CLASSICAL LOGIC AND NON-DENOTING SINGULAR TERMS

As suggested earlier, the classical logician is not *forced* to modify his formal instrument by the mere presence in natural language of expressions like ‘Pegasus’ or ‘the round square’. He has at his disposal several techniques for dealing with alleged non-denoting singular terms within his own framework. Since all these techniques are treated extensively in other parts of the *Handbook*,<sup>7</sup> I will limit myself here to little more than listing them.

In the first section, I pointed out that the problem free logicians see in classical logic (and try to solve with their logics) is the following: classical

---

<sup>7</sup>In particular, in Hodges’ chapter in Volume 1 and Salmon’s chapter “Reference and Information Contents: Names and Descriptions” in a later volume.

logic makes it impossible to combine the presence of non-denoting singular terms with an ‘existential’ reading of quantifiers. A classical logician willing to avoid this problem, then, has two main options available: he can deny existential import to quantifiers, or exclude the possibility of non-denoting singular terms.

If he wants to go the first way, he will find two basic suggestions in the literature. One is to drop the referential scheme of interpretation altogether, and go back to the old substitutional scheme, quite popular in the days before Tarski’s systematisation of formal semantics. The other is to remain within the referential framework, but admitting non-existing objects (as well as existing ones) in the range of quantifiers.<sup>8</sup>

If he wants to go the second way, he will again have a choice between two alternatives: Russell’s theory of descriptions and Frege–Carnap’s chosen object theory. Within the first alternative, he will rule out non-denoting singular terms by simply denying the status of singular terms to all those expressions of natural language (that is, definite descriptions and ‘grammatically proper names’) that can ever be non-denoting, and retaining it only for those other expressions (that is, demonstratives) that look absolutely ‘secure’ from a denotational point of view. Within the second alternative, his strategy will be more subtle. For Frege never really denied (as Russell did—at least as far as logical form was concerned) that *there are* in natural language non-denoting singular terms, but claimed that their presence constitutes a *defect*, to be repaired in a ‘logically perfect’ language (see [Frege, 1892]). Thus, whereas Russell’s proposal extends very naturally to a complex philosophical position, which includes (at least) metaphysical and epistemological themes, Frege’s qualifies as an intrinsically pragmatic one, in whose favour nothing can be said better than Carnap’s words in [Carnap, 1947]: ‘there is no theoretical issue of right or wrong between the various conceptions, but only the practical question of the comparative convenience of different methods’ (p. 33).

#### 4 INCLUSIVE LOGICS

Chronologically, some of the first instances of a revisionary attitude about the existential ‘commitments’ of classical logic can be found in what Quine called *inclusive* logics, that is, logics allowing the domain of quantification to be empty. To dispel what seems to be a quite common misunderstanding, it needs to be pointed out once and for all that inclusive logics and free logics *are two different subjects*. A logic can be free without being inclusive, and can be inclusive without being free. However, it is also convenient to treat

---

<sup>8</sup>The first suggestion is usually associated with Leśniewski, the second one with Meinong. For more recent formulations, see in the first case Lejewski [1954; 1958] and [Luschei, 1962], in the second Parsons [1980] and Routley [1980].

the two subjects together. For, on the one hand, the problems they face are strictly connected, and on the other, as for example [Belnap, 1960] has pointed out, it is quite natural to require inclusiveness of a free logic and *vice versa*.

The first inclusive logic was developed (twenty-five years before the first free logics) by [Jaskowski, 1934]. Jaskowski's is a natural deduction system, which, in contrast with most other such systems, allows for two different kinds of assumptions (or 'suppositions'). One can assume formulas (which one indicates by prefixing the formula with the metalinguistic symbol  $S$ ), and one can assume singular terms (which one indicates by prefixing the term with the metalinguistic symbol  $T$ ). The way the assumption of terms works is made clear by the quantificational rules of the system, which are given below.

1. *Supposition of a term*: at any point in a deduction it is possible to introduce an assumption of the form  $T\tau$ , where  $\tau$  is a new term.
2. *Universal Instantiation*:  $\varphi[\tau/x]$  follows from  $\forall x\varphi$  and  $T\tau$ .
3. *Universal Generalisation*: if  $\varphi$  follows from  $T\tau$  then it is possible to deduce  $\forall\tau\varphi$ , and this conclusion does not depend on the assumption  $T\tau$  (which is thus 'discharged').<sup>9</sup>

To explain how these rules allow the domain to be empty (by disallowing proofs of formulas which would exclude this possibility), it is best to use an example. Consider then

$$(3) \quad \forall x\varphi \rightarrow \exists x\varphi,$$

a typical instance of an 'exclusive' formula and a theorem of classical logic, and try to prove it in Jaskowski's system. A reasonable way to go about this is to assume the antecedent of 3 and the negation of its consequent, that is, to start out with

$$(4) \quad S\forall x\varphi$$

$$(5) \quad S\forall x\neg\varphi.$$

However, given the particular form of (2) above, nothing follows from 4 or 5 without also supposing a singular term. Let us do so, and continue with

$$(6) \quad T\tau.$$

Now from 4 and 6 we get

---

<sup>9</sup>The fact that the Universal Generalisation can be given in this form depends on specific features of Jaskowski's system: in particular, on the fact that his only terms are variables.

$$(7) \quad \varphi[\tau/x]$$

and from 5 and 6 we get

$$(8) \quad \neg\varphi[\tau/x]$$

which of course contradict each other. So the assumptions are not consistent, but the key point here is that there are not two but *three* assumptions, and in particular 4 and 5 can still be perfectly consistent if nothing like 6 is accepted (which is exactly what one would find most natural in the case of the empty domain).<sup>10</sup> Thus the attempted proof of 3 is blocked.

Jaskowski considers this quantificational system very briefly, almost as an appendix to a paper mostly devoted to propositional logic. Possibly for this reason, the system has a number of unnecessary limitations, and the consequences of removing them are not explored. If they had been explored, the system might have turned out to be the first *free* logic as well as the first inclusive logic. To understand what I mean, consider that in the system in question (i) open formulas are not provable, (ii) there are no individual constants, and (iii) the metalinguistic symbol  $T$  has no object-language counterpart. If (iii) and either (i) or (ii) were dropped (and, say,  $T^*$  were the object-language counterpart of  $T$ ), rules (1)–(3) of p. 153 would immediately yield (in conjunction with the propositional rules) theorems like

$$(9) \quad (\forall x\varphi \wedge T^*\tau) \rightarrow \varphi[\tau/x]$$

$$(10) \quad \forall xT^*x,$$

while at the same time blocking the proof of formulas like

$$(11) \quad \forall x\varphi \rightarrow \varphi[\tau/x],$$

and as we will see these are the key features of most free logics.<sup>11</sup>

When something like the above happens, and the solution of a problem can be found almost automatically by solving *another* problem, one naturally is led to suspect that there exists something more than a *coincidence*, that there is indeed a real *connection* between the two problems. In retrospect, it is not difficult to see what the connection is. Free logics are logics allowing for non-denoting singular terms, and of course if the domain is empty then *all* singular terms are non-denoting; hence if an inclusive logic

<sup>10</sup>For in this domain there are no objects, hence nothing to talk about by using singular terms.

<sup>11</sup>This of course when  $T^*$  is read as a substitute of the more common  $E!$ . Furthermore, notice that removal of (iii) is not critical to generate a free logic: if (iii) is not dropped (but either (i) or (ii) is) what we obtain is a 'pure' free logic, that is, a free logic without existence or identity. Indeed, making the existence symbol metalinguistic is one way of constructing a natural deduction or Gentzen formulation of such a pure free logic.

allows for any singular terms at all, it must allow for non-denoting singular terms, and thus be free as well. In light of this consideration, it is easy to see that the only way Jaskowski's logic (or any inclusive logic for that matter) could avoid being free was by refusing to admit any singular terms (which is the philosophical meaning of limitations (i) and (ii) above). And one might expect that, simply by developing their instrument a little further, inclusive logicians would have finally 'reached' free logics in a very natural way. However, this is not what happened, and the reason is interesting.

As we will argue later at great length, the fundamental problem to be solved in the development of free logics is a semantical one: the problem of assigning reasonable truth-conditions to sentences containing non-denoting singular terms. Inclusive logicians went very close to hitting this problem when they considered dropping some of Jaskowski's limitations. Thus Mostowski [1951], when constructing an inclusive logic contravening (i) above, had to decide what to do with open formulas in the empty domain. In a language without individual constants (as his was), free variables are the only possible place-holders for singular terms, hence Mostowski's problem was at least in part a special case of the fundamental problem of free logics. But there was at the time no awareness of this, so he simply treated free variables in analogy with (universally) bound ones, and he made all open formulas true in the empty domain.

The system resulting from this choice had a surprising anomaly: *modus ponens* was not truth- or validity-preserving in it, as the following example illustrates.

$$(12) \quad \frac{\varphi(x) \quad \varphi(x) \rightarrow \exists y\varphi(y)}{\exists y\varphi(y)}$$

The presence of this anomaly could have worked as a stimulus towards more satisfactory solutions, if the general problem lingering in the background had been perceived. Since it was not, subsequent authors such as Hailperin [1953] and Quine [1954] regarded the anomaly as a mere nuisance, and preferred to avoid the question entirely by returning to Jaskowski's practice of excluding open theorems, thus contributing in a decisive way to sealing off what could otherwise have been a promising line of enquiry.

## II: Proof-Theory

### 5 AXIOMATIC SYSTEMS

We saw in the last section how inclusive logicians avoided the crucial (semantical) problem of free logics. It might be surprising to find out that

even most free logicians basically side-stepped this problem for ten years (at least in their published works), by limiting themselves to a purely proof-theoretical development of their logics. As a result, the history of free logics can be neatly divided into two (partly overlapping) periods: the first one mostly devoted to proof-theory and the second one mostly devoted to semantics. It is natural, then, in accounting for the subject, to follow the same pattern, and here we will do just that. In the present part we will discuss the formal *systems* of free logic, as elaborated largely between 1956 and 1967,<sup>12</sup> and in the next one the *interpretations* of these systems, whose development took off only beginning in 1966. Within each part, however, we will make no attempt at preserving any chronological order, but will be guided entirely by considerations of systematicity.

Every axiomatic formulation of **CQC** contains as a primitive assumption either the so-called *Law of Specification*

$$(13) \quad \forall x\varphi \rightarrow \varphi[\tau/x]$$

or some other principle or rule deductively equivalent to it. (For definiteness, we will refer from now on to a system containing 13 as a primitive assumption.)<sup>13</sup> Furthermore, all the theorems of **CQC** – that free logicians find questionable (including 1 and 2) are proved by making a substantial use of 13. It is natural to conclude, then, that the first step in the construction of an axiom system for free logic is going to be dropping 13.

When this is done, the remaining axioms permit the proof of the following weakened form of 13 (that we might call Restricted (Law of) Specification):

$$(14) \quad (\forall x\varphi \wedge \exists x(x = \tau)) \rightarrow \varphi[\tau/x]$$

Far from representing a problem for the free logician, however, this result is most welcome to him; for Restricted Specification (in contrast with Specification proper) is a law that makes perfectly good sense even in the presence of non-denoting singular terms (and existentially loaded quantifiers).

To understand why this is so, consider that the supplementary condition required in 14 to instantiate the universal quantification  $\forall x\varphi$  with respect to the singular term  $\tau$  can be legitimately read as stating that  $\tau$  denotes a value of a bound variable, or more simply (*via* Quine's *dictum*) that  $\tau$  is denoting. Thus on the one hand 14 says nothing (and in particular nothing

<sup>12</sup>For these systems, see Leonard [1956]; Leblanc and Hailperin [1959]; Hintikka [1959a] and Lambert [1963; 1967].

<sup>13</sup>Also, we will refer to a language without function symbols and with  $\forall$  as the only primitive quantifier. As a consequence of the latter, the counterpart of 13 in terms of  $\exists$  (that is, the *Law of Particularisation*

13\*  $\varphi[\tau/x] \rightarrow \exists x\varphi$ )

will not occur among the primitive assumptions. And finally, let me notice once and for all that here we will try to give a *uniform* treatment of the various free logics, disregarding notational and stylistic differences among their authors.



questionable) about non-denoting singular terms, and on the other, though it cannot be used to justify the dubious inference from

(15) Nothing (existent) is a winged horse

to

(16) Pegasus is not a winged horse,

it *can* be used to justify the perfectly legitimate one from 15 *and*

(17) Secretariat exists

to

(18) Secretariat is not a winged horse.

Beginning with a seminal paper by Leonard [1956], that practically inaugurated the subject, free logicians have insisted that two of their most important tasks are (a) making explicit the existential assumptions that are tacit in classical logic (and that only can justify—in their opinion—the presence there of ‘laws’ like 13), and (b) discriminating between the cases in which these assumptions are relevant and the cases in which they are not. 14 is a good example of how these two tasks can be successfully performed: on the one hand, the assumption that (the singular term)  $\tau$  be denoting—taken for granted by classical logic—is here expressed by

(19)  $\exists x(x = \tau)$

and on the other the relevance of this assumption is signalled by its very presence, thus distinguishing the case of 14 from, say, that of

(20)  $\varphi(\tau) \rightarrow \neg\neg\varphi(\tau)$ ,

which is also a theorem of both classical and free logic and in which no supplementary existential condition is given (or needed).

All of the above, however, is made possible by the fact that **CQC** = is a logic *with identity*, for the identity symbol plays a vital role in expressing existence in 19 and substitutivity of identicals a vital role in proving 14. What would happen if the starting point were an axiom system for **CQC**, that is, for classical logic *without* identity?

We can approach this problem in stages. First of all, notice that if indeed 19 expresses an existential commitment to the denotational character of  $\tau$ , it seems legitimate to use it as *definiens* for a new *existence* symbol, say in the following way:

(21)  $E!\tau =_{\text{df}} \exists x(x = \tau)$ , where  $x$  is alphabetically the first variable distinct from  $\tau$ .

By using this abbreviation, 14 could be rephrased as

$$(22) (\forall x\varphi \wedge E!\tau) \rightarrow \varphi[\tau/x],$$

thus making the meaning of the extra assumption even more explicit.

In the system resulting from **CQC** by dropping 13, neither 14 nor its definitional abbreviation 22 can be proved; yet, on the other hand, something like 14 or 22 is certainly *needed*. For, as already noted, the procedures of classical logic (and in particular, Universal Instantiation), though based on tacit existential assumptions, are of course unquestionable when these assumptions are *true*.

The simplest way of reintroducing the legitimate cases of instantiation after dropping 13 from **CQC** would be to add  $E!$  to the set of primitive symbols, and 22 to the set of axiom-schemata. There is however a more ingenious way, which makes use of *neither* the existence *nor* the identity symbol, and is due to [Lambert, 1963].

To understand this alternative, it is enough to take a closer look at 14. What this ‘law’ says is that if something is a value of a bound variable then it has all the properties (expressible in the language and) shared by *all* such values. This conditional statement, however, could be reformulated in universal terms: *every* value of a bound variable has all the properties (expressible in the language and) shared by all such values. And this reformulation in turn suggests

$$(23) \forall y(\forall x\varphi \rightarrow \varphi[y/x])$$

as a possible replacement for 14 or 22.

I have now developed the core of a ‘pure’ free logic **FQC**, of a free logic with existence **FQCE!**, of a free logic with identity **FQC=**, and of course of a free logic with existence *and* identity **FQCE! =**. Before presenting their final formulations, however, two further problems must be mentioned.

First of all, consider the system obtained from **CQC=** by substituting 23 for 13. In this system

$$(24) \forall x\exists y(y = x)$$

is provable, which seems to be a perfectly reasonable result. For every value of a bound variable is certainly also a value of any other bound variable. However, as shown by Bencivenga [1978a; 1980a], this very natural result is *not* provable in the system obtained from **CQC=** by simply dropping 13, not is its counterpart in terms of the existence symbol

$$(25) \forall xE!x$$

---

<sup>14</sup>This existence symbol was first used by Russell, but only with descriptions. It was [Leonard, 1956] who generalised its application to all singular terms.

provable in the system obtained from **CQC=** by substituting 22 for 13 (which again is not good news, given the evident connection between 25 and that ‘existential import’ of quantifiers that we regarded as a defining feature of free logics).

Secondly, it long remained an open problem in pure free logic whether

$$(26) \quad \forall x \forall y \varphi \rightarrow \forall y \forall x \varphi$$

is provable in the system obtained from **CQC** by substituting 23 for 13. Fine [1983] solved this problem in the negative, showing the independence of 26 from the system in question.

In conclusion, then, let us agree on what follows. **FQC** is obtained from **CQC** by substituting 23 and 26 for 13. **FQC=** is obtained from **CQC=** by substituting 23 for 13. **FQCE!** and **FQCE! =** are obtained from **CQC** and **CQC=**, respectively, by substituting 22 and 25 for 13.

Two final remarks. First, all of the above are in a sense *minimal* systems of free logic: a few stronger systems will be considered in the part on semantics. Second, it will also become clearer in the part on semantics that all these systems are *inclusive* as well as free: once again, it is the strict connection between the two sets of problems that allows us to automatically solve the one while addressing the other.

## 6 NON-AXIOMATIC SYSTEMS

Something must be said about natural deduction and Gentzen formulations of free logics. Indeed, the first two formal systems for free logics—those by Leblanc and Hailperin [1959] and Hintikka [1959a]—were natural deduction systems, which however did not receive much currency in the literature. As to Gentzen systems for free logics, they can be found in [Routley, 1966; Trew, 1970; Bencivenga, 1980b]. Here in formulating both kinds of systems we will take for granted standard rules for connectives and identity (as well as, in the case of Gentzen systems, standard axioms), and we will make a substantial use of the existence symbol in the quantificational rules. Systems for pure free logic (or free logic with identity but not existence) may be obtained by using the same rules but making ‘**E!**’ into a metalinguistic symbol, and thus accepting as theorems only formulas not containing it.<sup>15</sup>

With all these qualifications, a natural deduction system for free logic can be characterised by the following four rules.

---

<sup>15</sup>This is the strategy suggested in note 11. In the case of a free logic with identity but without existence, it would also be possible to have  $\exists x(x = \tau)$  do the job of  $E!\tau$ , but this would have the ‘unnatural’ consequence of making quantification theory dependent on identity theory.

(27) Introduction rule for  $\forall$  :

$$\frac{\begin{array}{c} \{E!a\} \\ \vdots \\ \varphi[a/x] \end{array}}{\forall x\varphi}$$

where  $a$  is a new individual constant not occurring in  $\varphi$ .

(28) Elimination rule for  $\forall$  :

$$\frac{\forall x\varphi \quad E!a}{\varphi[a/x]}$$

(29) Introduction rule for  $\exists$  :

$$\frac{\varphi[a/x] \quad E!a}{\exists x\varphi}$$

(30) Elimination rule for  $\exists$  :

$$\frac{\begin{array}{c} \{\varphi[a/x]\} \\ \{E!a\} \\ \vdots \\ \psi \end{array}}{\exists x\varphi \quad \psi}$$

where  $a$  is a new individual constant not occurring in  $\varphi$  or  $\psi$ .

On the other hand, a Gentzen system for free logic can be characterised by the following four rules.

(31) Introduction of  $\forall$  in the antecedent:

$$\frac{\Gamma, \varphi[a/x] \vdash \Delta \quad \Gamma' \vdash \Delta', E!a}{\Gamma, \Gamma', \forall x\varphi \vdash \Delta, \Delta'}$$

(32) Introduction of  $\forall$  in the succedent :

$$\frac{\Gamma, E!a \vdash \Delta, \varphi[a/x]}{\Gamma \vdash \Delta, \forall x\varphi}$$

where  $a$  does not occur in  $\Gamma, \Delta$  or  $\varphi$ .

(33) Introduction of  $\exists$  in the antecedent :

$$\frac{\Gamma, E!a, \varphi[a/x] \vdash \Delta}{\Gamma, \exists x\varphi \vdash \Delta}$$

where  $a$  does not occur in  $\Gamma, \Delta$  or  $\varphi$ .

(34) Introduction of  $\exists$  in the succedent :

$$\frac{\Gamma \vdash \Delta, \varphi[a/x] \quad \Gamma' \vdash \Delta', E!a}{\Gamma, \Gamma' \vdash \Delta, \Delta', \exists x\varphi}$$

### III: Semantics

#### 7 THE PROBLEM

Consider a simple subject-predicate sentence, say,

(35) Socrates is a man.

How is a truth-value to be assigned to 35 according to the usual referential semantics for classical logic (briefly, classical semantics)?

Very simply put, the answer is as follows. First of all, we establish a domain of quantification (which, given our adoption here of Quine's *dic-tum*, can be identified with the set of existing things). Then we look for the denotation of the singular term 'Socrates' and for the extension of the general term (or predicate) 'being a man' in that domain. And finally, we pronounce 35 true if that denotation is a member of that extension, and false otherwise.

There is more to this procedure than meets the eye. Indeed, it is impossible to set up (in a reasonable way) the conditions at which a given sentence is true without having some *theory of truth*, and the procedure in question is based on one such theory, that is, on what is usually called the *correspondence* theory of truth. 35 is (say) true—according to this theory—because it corresponds to *reality*, and it would be false if it did not. More generally, 35 is true in a given state of affairs (or 'possible world') if it corresponds to reality *there*, and false otherwise. If we were doing propositional logic, this correspondence between (atomic) sentences and reality would be the bottom line, but at the level of analysis of quantification theory, that is, when sentences are analysed into (singular and general) terms, the correspondence in question is to be reduced to some more *basic* correspondences: the ones between singular terms and the objects constituting extensions. If in general the correspondence theory wants to establish the truth of a sentence in terms of a fit between what the sentence says and the way the world is, then such basic correspondences represent at this level of analysis the points at which the fit must be sought. To go back to our example once more, 35 is true just in case the object corresponding to 'Socrates' is a member of the set corresponding to 'being a man'.

But then of course basic correspondences are the key to the whole matter. *Once* we have the basic correspondences relative to some sentence we can

determine whether the sentence corresponds to reality or not, but not before. In the world in which we live, we know that 35 is true and

(36) Plato is a table

is false, but this is because we know *who* Socrates and Plato are, and *which* things are men and tables. Probably we would not know if we did not know who Socrates is, and certainly we would be in big trouble if there were no Socrates.

This kind of trouble is exactly what awaits us when we introduce non-denoting singular terms into the picture. Non-denoting singular terms denote nothing existent. Of course, they could denote something else, and in what follows we will consider some such position, but this is *one* possibility among many, and we must also take into serious account the possibility that they denote nothing at all. And taking this possibility seriously means considering situations in which some of the basic correspondences required by classical semantics *are simply not there*.

This is more than an epistemological problem. Consider for example

(37) Secretariat is white

and

(38) Pegasus is white,

and suppose that Secretariat be taken to a remote planet, where its colour could not be ascertained. Also, to simplify things, suppose that none of the fictional writings about Pegasus said anything about its colour. Still, there would be a fundamental difference between 37 and 38. For the colour of Secretariat could not be ascertained *in fact*, due to the practical limitations of human beings, but could be ascertained *in principle*, by somebody able to overcome those practical limitations, whereas in the case of Pegasus the thing would be impossible in principle, too: since Pegasus is nowhere to be seen, no matter how our powers were to improve, they would not influence our ability (or rather, inability) to verify its colour.

So it is not a matter of what we know, but of what we think truth is. Under the circumstances imagined above, it looks like it's not the case that Pegasus is white. It is the case that it is *not* white? (Or—which is the same—it is *false* that it is white?) Maybe, but if it is so, it must be for (at least partly) different reasons than (say) in the case of 36, and we need our theory of truth to tell us exactly what the analogies and the differences are between the two cases. The correspondence theory by itself cannot tell us this, because its verdicts are based on data—the basic correspondences—that here are not always available. Perhaps all we need is a small clause taking explicit care of such 'exceptions', but still we need something, we need some way of deciding when sentences containing non-denoting singular

terms are true, and why. This is the main question to be faced in the course of constructing a semantics for free logics—and in my opinion in free logics in general. It is an important question because any answer to it is inevitably going to provide an alternative to, or at least a generalisation of, the correspondence theory of truth. It is a delicate question because the correspondence theory is an old and venerable one, and challenging it represents a true act of ‘revolutionary science’. In the rest of the present part, I will give an account of this revolution.

## 8 OUTER DOMAINS

Given the way in which we set up *the* problem of ‘free’ semantics (that is, accommodating for the presence of gaps in the basic correspondences), the easiest way to ‘solve’ this problem consists simply in *avoiding* any such gaps. This is substantially the way most classical logicians operate, either by assigning arbitrary denotations to (previously) non-denoting singular terms (*à la* Frege–Carnap) or by excluding (*à la* Russell) such (alleged) terms from the class of things in need of a direct semantical counterpart. However, there is a way of going in this direction without ending up in classical logic: all that we have to do is to acknowledge that ‘Pegasus’ or ‘the present King of France’ have a semantical counterpart (or a denotation) just as much as ‘Bill Clinton’ or ‘the present President of France’ do, only that such counterparts (or denotations) *are not members of the domain of quantification*, or, to put it more bluntly, *do not exist*.

Even if some suggestions of this kind are much older,<sup>16</sup> the first such proposal that appeared in print was contained in the review by Church [1965] of Lambert [1963]. The purpose of the review was a critical one: Church indeed meant to show that the whole enterprise of free logic was of very little philosophical significance. Actually however (and a little ironically), the main result it achieved was that of sketching one of the very first semantical treatments of the subject, and one that was going to have a lot of success in the next few years.

Briefly, the substance of Church’s contribution was as follows. Let  $S$  be any set, and let a classical interpretation of individual and predicate constants be defined on  $S$ . Let  $P$  be any monadic predicate, and let two new quantifiers be defined, to be read ‘for every  $x$ , if  $x$  is  $P$  then . . .’ and ‘there is an  $x$  such that  $x$  is  $P$  and . . .’. Church suggested (without actually proving it, but the claim was indeed true, and was proved later)<sup>17</sup> that the set of theorems of Lambert’s axiomatic system would coincide with the set

<sup>16</sup>For example [Leblanc and Thomason, 1968] mention a suggestion of outer domains made (to Leblanc) by Joseph Ullian in 1962, and apparently both Belnap and Lambert had outer domain semantics very early (but never published their results).

<sup>17</sup>A sketch of the proof is given in Section 11.

of (classically) logical truths containing only the new quantifiers. As Meyer and Lambert [1968] put it, free logic was then just a ‘simple exercise in a theory of restricted quantification’.

Shortly after Church’s ‘proposal’, at least three major attempts were under way at constructing free semantics along the lines implicitly (and unwillingly) suggested by it. None of them made explicit reference to Church, and quite possibly they were all totally independent of his review, but there is a factual, objective sense in which they were all developing the suggestions contained in it, and emphasising different aspects of them.

The system which went the closest to reproducing Church’s intuitions was the ‘logic of possible and actual objects’ proposed by Cocchiarella [1966]. Semantically, the basic unit of this logic (a *Cocchiarella structure*) can be conceived of as an ordered triple  $\langle A, A', I \rangle$ , where  $A$  is as usual a non-empty set and  $I$  is a (total) function interpreting individual and predicate constants on  $A$ . The new character in this story is  $A'$ , which is just any (possibly empty) subset of  $A$ .  $A$  is the range of quantifiers, but not of quantifiers having existential import: rather, its members are to be construed intuitively as ‘possible objects’.  $A'$ , on the other hand, is the range of *another* pair of quantifiers, which *do* have existential import. If we adopt the usual symbols for the ‘existentially committed’ quantifiers and for example  $\bigwedge$  and  $\bigvee$  for the more general ones, it is easy to see that  $\forall x\varphi$  can be true in a Cocchiarella structure while  $\varphi[\tau/x]$  is not (indeed, even while  $\exists x\varphi$  is not, if  $A'$  is empty), hence that Specification fails for the restricted quantifiers. On the other hand, this principle does hold for the unrestricted quantifiers, which suggests that a formal system for the logic in question can be obtained simply by pairing a classical logic for  $\bigwedge$  and  $\bigvee$  with a free logic for  $\forall$  and  $\exists$  and adding the schema

$$(39) \quad \bigwedge x\varphi \rightarrow \forall x\varphi,$$

which supplies the connection between the two sets of quantifiers.

Due to the presence of two sets of quantifiers and of principles like 39, Cocchiarella’s logic of possible and actual objects is in fact more than a minimal free logic in the sense of Part II, but by dropping the unrestricted quantifiers from the language and all the theorems containing them from the formal system, we would obtain exactly a minimal free logic in that sense. On the other hand, if we were to do this then the larger set  $A$  would not be the range of any quantifiers but would only be providing denotations for the individual constants not interpreted in  $A'$ . It might be natural then to represent the situation in a slightly different way and, instead of insisting on the set of existents being a subset of a larger set of possibles, focus on the distinction between existents and *non*-existents (that is, in terms of a Cocchiarella structure, between  $A'$  and  $A - A'$ ). And this in turn would bring us immediately to the variant of the present approach proposed by Leblanc and Thomason [1968]. Since this variant is probably the most



popular, we will give here a slightly more detailed account of it than we did of Cocchiarella's (or than we will do of Scott's). The reader can easily accommodate our remarks to the other variants.

A *Leblanc–Thomason* (or, more simply *LT*) structure is again an ordered triple  $\langle A, A', I \rangle$ , where however  $A$  and  $A'$  are two *disjoint* sets (called the *inner domain* and the *outer domain*, respectively) such that *their union* is non-empty (this union, of course corresponds to the set of possibles in a Cocchiarella structure, and it still plays an important role in the present context, though in a way it shifted to the background).  $I$  is a (total) function interpreting individual and predicate constants on  $A \cup A'$ . An LT structure is *null* if its inner domain is empty, and *non-null* otherwise. In a non-null LT structure, an *assignment* is a (total) function from the set of variables *to the inner domain*. Satisfaction is then defined as usual, but the fact that variables can only get values in the inner domain makes of this domain the range of quantifiers.

Leblanc and Thomason's semantics is inclusive as well as free, as is shown by the presence of null LT structures. Thus the problem arises once again of what to do in those structures with open formulas. However, this is not a problem that *we* need consider. Simplifying on Leblanc–Thomason's own treatment, we can agree to adopt the Jaskowski–Hailperin–Quine suggestion of accepting only closed theorems, and thus leave open formulas simply *uninterpreted* in null LT structures.<sup>18</sup> In contrast with the above authors, this won't produce any limitation in our expressive powers, because we already have individual constants as place-holders for singular terms, and individual constants behave in null LT structures just as they behave (when they are non-denoting) in the non-null ones.<sup>19</sup> Besides, we need not deal with open formulas as a preliminary step for evaluating quantified sentences in null LT structures, since we can agree once and for all that for all such structures  $\mathfrak{A}$  and all sentences  $\forall x\varphi$

$$(40) \mathfrak{A} \models \forall x\varphi.$$

An analogous attitude will be adopted (without further mention) with respect to all the alternative semantics to be presented here.

We have thus considered two variants of what we will call in general the *outer domain* approach to free semantics. As suggested above, they emphasise different aspects of this approach (and of the original suggestions by Church). Cocchiarella makes the most of the notion of restricted quantification, whereas Leblanc and Thomason make the most of the presence of two kinds of *denotata*. A third aspect of this approach is its similarity to Frege–Carnap's classical device; in both cases indeed the problem of non-denoting singular terms is solved by making them *denoting* (in a sense). It

<sup>18</sup>As to Leblanc and Thomason themselves, they preferred to follow Mostowski in weakening *modus ponens*.

<sup>19</sup>That is, in both cases they denote members of the outer domain.

is not surprising then that there be a further variant of the approach in question which makes the most of this similarity. Such a variant is due to Scott [1967].

Scott's theory (as on the other hand both Frege's and Carnap's) is actually a theory of (definite) descriptions, but the general semantical strategy it embodies makes perfectly good sense at the level of unanalysed singular terms, too. Very simply put, this strategy is as follows. Associate with each domain of quantification an entity not belonging to it, say the entity  $*$ . Since  $*$  is outside the range of quantifiers, by Quine's *dictum* it does not exist, but still it can be assigned as a semantical value to singular terms; hence  $\{*\}$  works practically as an outer domain. At the same time however, since this outer domain is a singleton,  $*$  works also like a Carnapian chosen object, in that all the (originally) non-denoting singular terms have it as their common semantical counterpart (or 'denotation').

Because of this last feature of Scott's semantics,

$$(41) (\neg E!\tau \wedge \neg E!\tau') \rightarrow \tau = \tau'$$

is logically true in it. Since of course 41 is not provable in the minimal free logics of Part II, they should be strengthened somewhat to generate a formal system adequate to the semantics in question. The simplest way to do this consists in adding 41 itself as a further axiom-schema. A more elaborate alternative would require the addition of a new symbol (for example, ' $*$ ') to the language, of a clause fixing its interpretation on the 'non-existent object' to the semantics, and of the schema

$$(42) \neg E!\tau \rightarrow \tau = *$$

to the deductive apparatus. Actually, if we were dealing with descriptions, this more elaborate alternative might turn out to be the simpler of the two, because in description theory ' $*$ ' could be introduced by definition, say by

$$(43) * =_{df} \iota x(x \neq x).$$

Our presentation of the outer domain approach to free semantics ends here, and we can conclude the present section with a brief appraisal of this approach. Such remarks will be of a general nature, and will leave aside the special developments recommended by Scott, which will be the subject of further discussion in the section on descriptions.

First of all, then, the positive side. Outer domain semantics is simple, and we know why: filling all the gaps left by non-denoting singular terms in the basic correspondences allows one to stick to the standard evaluation procedures, thus generating a feeling of familiarity for the whole enterprise. The problem of non-denoting singular terms is not so much solved as it is *dissolved*. The reason why 'Pegasus is white' is (say) true is not at all

different from the reason why ‘Secretariat is white’ would be: ‘Pegasus is white’ is true because the (non-existent) object Pegasus falls within the (non-existent part of the) extension of ‘white’.

Furthermore, outer domain semantics is formally very convenient. Semantical completeness is provable here in its stronger form: not only the set of logically true sentences but also the set of valid arguments is recursively enumerable, and thus the whole logic is under complete (proof-theoretical) control. Once again, the reason is not hard to find, even though probably it will be fully appreciated only later, when this semantics is contrasted with some of its alternatives.<sup>20</sup> The fact is that the semantics in question is *bivalent*: every sentence, whether or not it contains non-denoting singular terms, is either true or false (in any structure).

Whereas all the positive comments on this semantics have to do with practical or technical matters, all the negative ones have to do with philosophical matters. The first (and most common) of these comments is probably best put in the form of a question: what exactly is the status of the members of the outer domain? The most natural answer to this question is ‘non-existent objects’, but such an answer generates trouble.

It is not that non-existent objects are not philosophically ‘respectable’. On the contrary, they are quite popular in philosophy today, probably more than they ever were after Russell’s alleged ‘refutation’ of Meinong. Scholars of Meinongian inclination, for example Parsons [1980], have questioned the validity of that refutation, and constructed ingenious philosophical theories of non-existent objects. To be sure, such objects are difficult to deal with, mostly because—as noted by [Quine, 1948]—their identity conditions are far from clear, but to say that they are difficult is not to say—as Quine concluded a bit too hastily—that they are ‘well-nigh incorrigible’. After all, if we are not ready yet to give a satisfactory account of non-existent objects—and chances are that soon this will no longer be true, if indeed it is true now—the problem might be with us and our philosophy, rather than with the objects themselves.

All of this is very good, but unfortunately it does not even get close to removing the trouble we mentioned above. For the difficulty with using non-existent objects to construct a semantics for free logics is not that we are not ready to accept them or to account for them, but that accepting them and accounting for them should have little to do with one’s *logic*, and should depend instead on one’s *metaphysical* position—at least according to a quite common conception of logic and metaphysics.<sup>21</sup> And if this conception is

---

<sup>20</sup>See in this connection the end of Section 10.

<sup>21</sup>Given that metaphysics is often defined as the study of what *is* (insofar as it is), it may sound awkward to say that this discipline should also be concerned with what *is not*. The awkwardness however is reduced when we consider that most supporters of non-existent objects ascribe to them some sort of (watered-down) ‘being’ (often discriminating between ‘to be’ and the stronger ‘to exist’).

correct, then outer domain semantics is in conflict with what we regarded as the most basic motivation for having a free logic in the first place!

Another criticism of outer domain semantics has to do less with non-existent objects in themselves than with the particular use the semantics in question makes of them. For many supporters of such objects (including Meinong)<sup>22</sup> have held that at least some of them are ‘incomplete’, that is, such that for some property  $P$ , they have neither  $P$  nor not- $P$ , and indeed, there seems to be something to the claim that, if for example none of the stories about Pegasus says anything about its length, then (say) the sentence

(44) Pegasus is six feet long

is neither true nor false, but simply indeterminate. The present semantics, however, allows for no such ‘truth-value gaps’, and we must be careful not to introduce them too hastily into the picture. For if we decided to simply leave some members of the outer domain ‘undefined’ with respect to some predicate constant  $P$ , this would determine the immediate collapse of such logical laws as

(45)  $\varphi \vee \neg\varphi$ ,

and with them of most of **CPC**.

Though certainly quite serious, this criticism of outer domain semantics is not as damaging as the first one was. Truth-value gaps cannot be introduced too hastily in this semantics, but can be introduced after all. In a later section, we will mention a compromise between the outer domain and the supervaluational approach which saves much of the spirit of both while allowing (as supervaluations do) for truth-value gaps and (in a sense) ‘incomplete’ objects.

A third negative comment on outer domain semantics is even more dependent than the previous one on the particular ways this semantics has been formulated so far, and furthermore is itself grounded on a debatable philosophical position. It is just that some people find it objectionable that there be ‘genuine’ relations between existent and non-existent objects, and the semantics in question (in its usual formulations) seems to allow for such relations.

What I mean by ‘genuine’ deserves some words of explanation. If we admit non-existent objects, there are inevitably going to be some relations between them and the existent ones, because it is simply true that, say,

(46) I am thinking of Pegasus.

Relations such as the one expressed by 46, however, are of a very special kind; without entering into any detail, we can qualify them as ‘intentional’ or in some sense ‘modal’, and contrast them with such ‘purely descriptive’ (or ‘genuine’) relations as the one expressed by

---

<sup>22</sup>See [Findlay, 1963, p. 57].

(47) Peter is taller than Mary.

Now according to some philosophers of logic the predicate constants  $P, Q, R, \dots$  of quantification theory stand for genuine (nonintentional) relations, hence if you think that no such relations hold between existents and non-existents (and share this position in the philosophy of logic), you might be embarrassed by the fact that in the usual formulations of outer domain semantics an ordered pair  $\langle o_1, o_2 \rangle$  can fall into the extension of (say)  $P$  even when  $o_1$  is a member of the inner domain and  $o_2$  a member of the outer domain (or vice versa).

It would be possible to reformulate outer domain semantics so as to avoid this (for some people) unwelcome feature of it. However, this has not been done yet, and the present context is certainly not the right place to do it. Let me just notice in closing that at some point the problems raised by these reformulations will become interwoven with the problems raised by our second criticism of outer domain semantics. For there are several ways to go after claiming that the existent object  $a$  cannot hold the genuine relation  $P$  to the non-existent object  $b$ , and one of them is to say that

(48)  $Pab$

is an indeterminate sentence.

## 9 CONVENTIONS

The positions that we will consider in the present section are quite disparate. What holds them together (in my opinion at least) is the fact that they determine truth-values for sentences containing non-denoting singular terms pretty much by *fiat*, and that whatever discussions or justifications they offer of their choices are more or less of an ‘external’ nature, that is, have mostly to do with the practical consequences of these choices or with how much they ‘fit’ with other (already accepted) linguistic theories.<sup>23</sup> For this reason, I found it suggestive to group them around the word ‘convention’.

The most typical ‘conventional’ positions can be described very easily. Their basic semantical unit is a ‘partial’ structure  $\langle A, I \rangle$ , where  $A$  is the usual domain of quantification and  $I$  interprets (on  $A$ ) all the predicate constants and *some* (possibly all, possibly even none) of the individual constants. Truth-values for atomic formulas not containing non-interpreted constants are determined as usual, whereas all the atomic formulas containing such constants have the same truth-value, true or false as the case

---

<sup>23</sup>For example, [Burge, 1974] takes it as a crucial argument in favour of his approach that it allows him to save a Tarskian theory of truth. Of course, even authors going in different directions do sometimes offer ‘external’ justifications, but nowhere seem such justifications as crucial as in the semantics discussed in the present section.

may be. Borrowing (and adapting) some terminology by Lambert [1981], we can thus distinguish (on obvious grounds) between *positive* and *negative* conventional semantics.

The truth-values of complex formulas are also determined as usual. In particular, assignments are defined as (total) functions from the set of variables to the domain, and the satisfaction-condition for quantified formulas is the standard

$$(49) \mathfrak{A} \models \forall x\varphi[f] \text{ if and only if for every member } \alpha \text{ of the domain of } \mathfrak{A}, \mathfrak{A} \models \varphi[f, \alpha/x].$$

Completeness and the other usual metatheoretical results are not difficult to establish for conventional semantics: once more, bivalence makes things relatively easy. Indeed, there is in general not much to be said about the technicalities of these semantics; hence we might turn right away to some considerations for and against accepting them.

Given the present state of the literature, positive conventional semantics are little more than a theoretical possibility. Of course, it is a possibility of which most scholars in the field are aware, but nonetheless it has not become yet the core of a full-fledged semantical approach. In such a situation, we cannot expect to find a great deal of (published) discussion on the semantics in question; hence most of this discussion we will have to supply on our own.

An argument which could be given in favour of positive conventional semantics is that they make it very easy to validate the schema

$$(50) \tau = \tau,$$

which is usually regarded as expressing a logical law. However, this does not mean that such logics allow for a standard treatment of identity: they create problems with respect to the substitutivity of identicals. For consider a structure  $\langle A, I \rangle$  such that  $I(a)$  is not defined,  $I(b)$  is defined and  $I(b) \notin I(P)$ . In this structure,

$$(51) a = b$$

and

$$(52) Pa$$

are both true, but

$$(53) Pb$$

is false. On the other hand, the situation is not so desperate as it might seem. As we will see, many free semantics are forced to treat identity in some special way, and in particular to require explicitly that a sentence of the form 51 be false when exactly one of  $a$  and  $b$  is denoting. A similar special provision is all that we would need here.

Another criticism of positive conventional semantics could be that they, too, allow for ‘genuine’ relations between existents and non-existents. Of course, this criticism should be slightly reformulated, since the semantics in question do not literally allow for any non-existents at all, but only for the truth of (atomic) *sentences* containing both denoting and non-denoting singular *terms*. And of course, once reformulated, the criticism might be easily answered in a number of different ways, by adopting some more ‘special provisions’.

Turning now to negative conventional semantics, we must notice first of all that they have had much more success than their positive counterparts. Indeed, the first semantical account of a free logic ever published, the one by Schock [1964; 1968], was a negative conventional semantics, and so was one of the latest ones, by [Burge, 1974]. If we add that Russell’s classical description theory has a lot in common—in the results if not in the methods or the motivations—with these semantics, and that even authors going in a different direction—such as Scott [1967]—tend to agree with them when it comes to determining the truth-values of sentences,<sup>24</sup> we will have an idea of the persistent attraction of the approach in question.

What are the reasons for this attraction? Schock [1968] expresses his motivation as follows: ‘The application of a predicate to various terms holds just when the denotations of the terms stand in the relation denoted by the predicate; if not all of the terms denote, then their denotations cannot stand in the relation and the application does not hold’ (p. 21). In other words, since there is no denotation of ‘Pegasus’, the denotation of ‘Pegasus’ cannot stand in any relation with (say) the denotation of ‘Bellerophon’; hence

(54) Pegasus is loved by Bellerophon

is false.

As it is, this is not much of an argument. For it basically reduces to the circular claim that 54 is false because

(55) The denotation of ‘Pegasus’ stands in the relation of being loved by with the denotation of ‘Bellerophon’

is—where the falsity of 55 is as much in need of a justification as that of 54. One might try to shore it up by pointing out that—in the usual set-theoretical terms of classical semantics—the set of ordered pairs corresponding to the relation of being loved by is not going to contain any member corresponding to Pegasus and Bellerophon. But one might answer that classical set theory—just as classical semantics—is not prepared to deal with non-denoting singular terms, and that things could be different

---

<sup>24</sup>See the schema (I3) on p. 188 of [Scott, 1967], which Scott considers *very reasonable* when giving axioms for a theory (not however for pure logic).

in a ‘free set theory’ devised for this purpose.<sup>25</sup> The negative conventional semanticist might then reply that his approach does not require any such new technical instrument, that it allows one to preserve most of the classical framework, and that these ‘conservative’ features are very important from a pragmatic point of view.

With this appeal to conservatism we have struck a key note. For what exactly is so great about preserving the classical framework? And why should pragmatic arguments be so crucial in choosing a free semantics? Of course, one may think that pragmatic arguments are *always* crucial, or even that they are *the only* arguments one can give in favour of any theory, and that the traditional framework is *always* to be preserved whenever it is possible. This is a general position in the philosophy of logic (and of science), and I have nothing much to say about it—except that it does not seem to be shared by many free logicians. What I think deserves some comment is the opinion somebody might have that *in this particular case, because of the particular nature of the problem*, pragmatic arguments are more important than usual. More precisely, I refer to the opinion that, since non-denoting singular terms are basically ‘don’t cares’, the only criteria to use in assessing an attitude towards them are whether the attitude is simple, efficient, and does not require a vast revision of our conceptual framework. I think that this opinion may be very dangerous for free logic as a whole.<sup>26</sup> For after all, what is simpler and more conservative in this case than just sticking to *classical* logic, supplemented by some of the policies mentioned in Section 3? Thus, omitting any further comment on the specifics of negative conventional semantics,<sup>27</sup> I will conclude the present section with some remarks on why free logicians may think that non-denoting singular terms are *not* ‘don’t cares’, and why they might want something more than an efficient way to accommodate them in the classical framework.

It is not that free logicians are interested in non-denoting singular terms *in themselves*; it is not that they have some kind of perverse attraction for what does not exist. However, they are not bound to the realm of existents either; they do not share that ‘prejudice in favour of the actual’<sup>28</sup> which is so common (and possibly healthy) in other branches of knowledge. A scientific truth is true (at least in part) because of the way the world is, and given a sufficiently wide conception of the ‘world’ the same might be said of many philosophical truths, but a logical truth should be independent of any such factual matters, and in particular of what exists and what does not exist.

<sup>25</sup>Such free set theories have been developed in [Scott, 1967] and [Bencivenga, 1976]. To my knowledge, however, they have never been used in formulating semantics.

<sup>26</sup>Of course, so may be the more general position mentioned above. but that is also too general to be discussed here.

<sup>27</sup>But let me stop a minute to notice that—as far as identity is concerned—these semantics are in a sense in a dual position with respect to their positive counterparts: they easily validate substitutivity of identicals but invalidate self-identity.

<sup>28</sup>The expression is Meinong’s. See [Meinong, 1904].



A logical truth should depend only on (logical) form, and there seems to be no plausible (non *ad hoc*) ground for distinguishing between the form of ‘Pegasus is white’ and that of ‘Secretariat is white’. Of course, we know that somewhere there is a difference, because Secretariat exists and Pegasus does not, but invoking that piece of information, and discriminating on that score between the procedures involved in evaluating the two sentences, would be contaminating logic with mere contingencies.

In conclusion, non-denoting singular terms represent an important challenge for logic in general. For this reason, even though the free logician is not going to forget considerations of simplicity and theoretical conservatism, he may think that it is more crucial to rethink the whole subject, no matter how complicated and revisionary this process is going to be.

## 10 SUPERVALUATIONS AND BEYOND

The most organic attempt to date at rethinking the whole subject of truth theory in view of the presence of non-denoting singular terms was initiated in 1966 by two seminal papers by van Fraassen [1966a; 1966b], and pursued by van Fraassen himself and several other authors, including Skyrms [1968], Meyer and Lambert [1968], Woodruff [1971] and Bencivenga [1980b; 1981]. In the present section we will study this approach, which from its most characteristic technical instrument may be called the *supervaluational* approach.

The starting point of our analysis is once more conventions. According to van Fraassen [1966a], the truth-value of a sentence like

(56) Pegasus has a white hind leg,

or even the fact that this sentence has a truth-value, is ultimately to be established on the ground of some convention. This convention, however, belongs to the *philosophy of language*, and should receive *there* whatever justification it is going to receive. *Logic*, on the other hand, has nothing to do with any such conventions and justifications: the set of logical truths should be absolutely independent of the philosophy of language we decide to adopt. In particular, there will be conventions assigning True to 56 and conventions assigning False to it, but logic should not be committed to any of them. At the very most, we can think of logic as committed to the *logical product* of all possible conventions, to what all these conventions have in common, to what is going to be true (or false) *no matter what convention we adopt*.

This notion of the logical product of all possible conventions leads very naturally to the idea of a supervaluation, in the following way. Let a partial structure  $\mathfrak{A} = \langle A, I \rangle$  be given, and suppose that  $I(a)$  is not defined,  $I(b)$  is defined and  $I(b) \in I(P)$ . Application of the standard evaluation procedures establishes the truth of sentences like

(57)  $Pb$

(58)  $Pb \vee \neg Pb$

(59)  $\exists xPx$

as well as the falsity of

(60)  $\neg Pb$

(61)  $Pb \wedge \neg Pb$

(62)  $\forall x\neg Px,$

but determines no truth-value at all for

(63)  $Pa$

(64)  $\neg Pa$

(65)  $Pa \vee Pb$

(66)  $Pa \wedge \neg Pb$

(67)  $Pa \vee \neg Pa$

(68)  $Pa \wedge \neg Pa.$

Of course, 63–68 might receive any combination of truth-values on the ground of some convention or other, but it seems reasonable to restrict our attention to those conventions that are *classical* in the following sense: they assign truth-values to atomic formulas containing non-denoting singular terms in some way that it is not our present concern to examine (indeed, that we could for our present purposes regard as totally *arbitrary*), but then they proceed to evaluate complex formulas *in the standard way*.

The combination of any such classical convention and the information supplied by the partial structure will determine a valuation of all the sentences of the language. Let us agree to call any such valuation a *classical valuation* (on  $\mathfrak{A}$ ). Of course, all classical valuations will agree on all the sentences (like 57–62) that contain no non-denoting singular terms, but the interesting thing is that they will also agree on many sentences which *do* contain non-denoting singular terms. Thus for example 63 will receive the value True in some classical valuations and the value False in some others, but *every* classical valuation will verify 65 and 67 and falsify 66 and 68. In other words, there will be cases—and many of them—in which the logical product of all classical valuations will be non-empty, and as such informative, beyond what is determined by the partial structure. As the fate of

67 and 68 suggests, this supplementary information is enough to extend to non-denoting singular terms all of **CPC**.

The essence of van Fraassen's approach consists simply in *using* this supplementary information. More precisely, the *supervaluation*  $W_{\mathfrak{A}}$  for a partial structure  $\mathfrak{A}$  is characterised by him as the (partial) valuation which assigns True to the sentences that are true in all classical valuations on  $\mathfrak{A}$ , False to the sentences that are false in all classical valuations on  $\mathfrak{A}$ , and no truth-value at all to the remaining sentences. Then, in at least one of the alternatives he contemplates,<sup>29</sup> supervaluations are to constitute the basic (or *admissible*) valuations of free semantics, and all the other semantical notions are defined in their terms.

The fact that supervaluations preserve all of **CPC** without espousing any specific convention or admitting non-existent objects is certainly remarkable, but it is also important to point out that when we move beyond propositional logic supervaluations create serious problems.

The most apparent of these problems concern identity. When  $a$  is non-denoting, nothing so far prevents a classical valuation from falsifying

$$(69) \quad a = a,$$

and when both  $a$  and  $b$  are non-denoting, nothing so far prevents a classical valuation from verifying 51 and 52 and falsifying 53 on p. 170, thus invalidating substitutivity of identicals.

Further (and more subtle) problems concern quantification. To understand them, we must first of all ask ourselves how the present approach can be extended to deal with variables and open formulas. A natural way would seem to be the following. Define a convention for a partial structure  $\mathfrak{A}$  as a binary function from the set of atomic formulas and the set of assignments for  $\mathfrak{A}$  to  $\{T, F\}$  (that is, as a function assigning (arbitrary) truth-values to atomic formulas *relative to* assignments). Then let  $\mathfrak{A}$ , any convention and any assignment determine an *auxiliary classical valuation*, by using standard evaluation techniques for atomic formulas not containing non-denoting singular terms and for complex formulas, and relying on the convention for atomic formulas containing non-denoting singular terms. Point out that in the case of *sentences* assignments make no difference, and define on this ground the notion of a classical valuation on  $\mathfrak{A}$ , as determined only by  $\mathfrak{A}$  and a convention. Finally, define the supervaluation for  $\mathfrak{A}$  as the logical product of all classical valuations on  $\mathfrak{A}$ . All of this sounds very good, but unfortunately it does not work: for it may well be that in the case of sentences assignments *do* make a difference. Indeed, nothing so far prevents a convention from assigning True to some atomic sentence

---

<sup>29</sup>He also considers an alternative in which classical valuations themselves are the admissible valuations, but such an alternative is far less interesting or philosophically defensible; hence we will totally disregard it here.

(containing a non-denoting singular term) relative to some assignment and False to the same sentence relative to a different assignment. Hence the ‘definition’ suggested above of a classical valuation is not legitimate.

Van Fraassen’s solution of these problems is disappointing. Very simply put, he adds to the definition of a convention a number of *ad hoc* clauses which rule out—by *fiat*—all the possibilities contemplated above. More precisely, a convention  $K$  is to be defined in such a way that

1.  $K(\tau = \tau, f) = T$ ;
2.  $K(\varphi[\tau/x], f) = K(\varphi[\tau'/x], f)$  if either  $\tau[f] = \tau'[f]$  or  $K(\tau = \tau', f) = T$ ;
3.  $K(\varphi, f) = K(\varphi, f')$  if  $f(x) = f'(x)$  for every variable  $x$  occurring (free) in  $\varphi$ .

For analogous reasons, it is also required that

4.  $K(\tau = \tau', f) = F$  if exactly one of  $\tau[f]$  and  $\tau'[f]$  is defined.<sup>30</sup>

These additional clauses simplify the technical developments, and make some of the desired metatheoretical results easily available, but certainly don’t go in the direction of providing a satisfactory philosophical motivation for the resulting semantics. Indeed, they rather weaken whatever motivation there was after our first introduction of the supervaluational approach. For remember, the crucial point there was the neat separation promised by supervaluations between logic and philosophy of language, and the fact that they were supposed to be independent of specific conventions, and committed only to the logical product of *all* conventions. Now it would be hard to hold this point of view—in presence of so many restrictions on what counts as a convention. Even the fact that we should limit ourselves to *classical* conventions (or valuations) might begin to look suspicious, and the whole enterprise appear dangerously close to a gigantic circle. To put it bluntly, it seems that van Fraassen can assign truth-values to sentences containing non-denoting singular terms only to the extent to which he is *not* independent of a conventional attitude.

In my opinion, these shortcomings of supervaluational semantics are due less to the general idea of a supervaluation than to a failure on van Fraassen’s part to get deeper into its analysis. Indeed, I think that supervaluations come very close to providing that generalisation of the correspondence theory of truth that we judged necessary for a reasonable treatment of non-denoting singular terms. To justify this claim, it will be convenient to have a fresh look at the whole thing.

<sup>30</sup>Van Fraassen’s language does not contain  $E!$ . If it did, it would probably be necessary to add one more clause:

(e)  $k(E!\tau, f) = T$  if and only if  $\tau[f]$  is defined.

Let me begin by asking a direct question. Why is a sentence like 67 always true in supervaluational semantics, even when its only atomic component 63 is truth-valueless? One way to answer this question could be the following: even though 63 has no truth-value, if it *did* have a truth-value, *any* truth-value, 67 would be true. But what is required for 63 to have a truth-value? The semantics itself gives the answer: 63 has a truth-value if and only if *a* is denoting. Hence the answer to our original question may be rewritten as follows: 67 is true even when *a* is non-denoting (and 63 truth-valueless) because if *a were* denoting then it would be true.

This answer constitutes the core of a new theory of truth, which for the sake of a label we might call the *counterfactual* theory of truth. This theory substantially agrees with the correspondence theory on all sentences not containing non-denoting singular terms, but develops in an original way beyond that scope. Its most basic principle may be formulated as follows: a sentence containing non-denoting singular terms is true (false) if and only if it would be true (false) in case these terms *were* denoting, *no matter what their denotations were*. According to this principle, not only is 67 always (hence logically) true and 68 logically false, but also 69 is logically true and substitutivity of identicals is truth-preserving, and all of this as a consequence not of the adoption of *ad hoc* clauses but of the use of normal evaluation procedures. Also, it will be useful to point out right away that accepting the principle in question does not commit one in any way to outer domains or non-existent objects. For in outer domain semantics non-denoting singular terms simply ‘denote’ non-existents, whereas in the present approach these terms denote nothing, and we only take the liberty of considering *alternative* situations (or ‘possible worlds’) in which they denote, and of making their behaviour there relevant for the evaluation of sentences containing them in the situations (or worlds) in which they do not denote. We will see that compromises are possible between the counterfactual theory of truth and outer domain semantics, but such compromises are not inevitable.

Supervaluational semantics—as developed by van Fraassen—*suggests* the counterfactual theory, but does not explicitly *espouse* it. More precisely, this semantics does not get to the point of assigning a truth-value (or no truth-value) to a sentence containing non-denoting singular terms by considering situations in which these terms are denoting. Rather, it considers situations in which the atomic formulas containing these terms receive truth-values. In a way, it is as though supervaluational semantics were developing the suggestions leading to the counterfactual theory *only at a propositional level of logical analysis*. From this limitation springs in my opinion all the talk about conventions, for it seems that only on a conventional basis we can assign (what look like) arbitrary truth-values to unanalysed atomic formulas. And from the same source springs also the necessity of adding *ad hoc* clauses to the definition of a convention; for from a purely propositional

point of view there is simply no reason why a sentence like (69) should not be false, or more generally why the assignment of truth-values should fit the non-propositional logical structure of sentences.

The message sent by these considerations is quite clear: what we have to do to remove all that sounds *ad hoc* and circular in the semantics of supervaluations is carry the approach expressed by this semantics to a more specifically quantificational level. Several authors have received this message, and several semantics have been developed along these lines, but all of them had to face, and solve one way or another, a serious problem, whose realisation might well have been the main reason for van Fraassen's adoption of a 'propositional' treatment of non-denoting singular terms.

The problem is as follows. Suppose that supervaluational semantics be developed at a quantificational level in what looks like the most natural way. Given a partial structure  $\mathfrak{A}$  and a sentence  $\varphi$  containing singular terms that are non-denoting in  $\mathfrak{A}$ , one considers all *extensions* of  $\mathfrak{A}$  which make those terms denoting, and pronounces  $\varphi$  true (in  $\mathfrak{A}$ ) if it is true in all such extensions, false if it is false in all such extensions, and truth-valueless otherwise. Now consider the sentence

$$(70) \quad \forall xPx \rightarrow Pa,$$

and suppose that  $a$  be non-denoting (in some structure  $\mathfrak{A}$ ).

70 is an instance of Specification, and we know that rejecting Specification is the most distinctive feature of a free logic from a proof-theoretical point of view. In particular, 70 is not provable in any free logic unless a special clause is added to it which makes sure that  $a$  is denoting; hence we would expect that when  $a$  is *non*-denoting a free semantics had a way of invalidating 70. But this is simply not the case in the semantics sketched above. For 70 is certainly true in all extensions of  $\mathfrak{A}$  in which  $a$  is denoting, and so it is true in  $\mathfrak{A}$  itself. This argument can be easily generalised to any other instance of Specification, and the conclusion is startling: the most natural 'quantificational' development of supervaluational semantics leads not to free but to *classical* logic!

There are in the literature at least four different ways of addressing this problem.<sup>31</sup> The simplest one is advocated by Woodruff [1971], and consists substantially in mixing the supervaluational approach with the outer domain approach. To get the compromise in question we need to qualify the counterfactual theory of truth in the following way: a sentence containing non-denoting singular terms is true (false) if and only if it would be true (false) in case these terms were denoting, no matter what their denotations were but provided that they were *non-existent objects*. Thus all

---

<sup>31</sup>Except for the last one, the semantics to be discussed below do not present themselves explicitly as 'ways of addressing this problem'. But this is a good way of perceiving the substance of their contribution.

extensions of a partial structure  $\mathfrak{A}$  are to be conceived as all possible ways of adding an outer domain to it, and 70 is easily invalidated. The resulting semantics has a few advantages over the straight outer domain approach, especially because it does not assign a truth-value to *every* sentence containing non-denoting singular terms, and thus in a sense can accommodate for ‘incomplete’ objects, but it shares the most substantial problem of that approach, that is, the metaphysical commitment to non-existents.

A more sophisticated variant of this strategy was proposed (earlier) by Meyer and Lambert [1968]. It still consists substantially in allowing for outer domains, but these domains are thought of as constituted by *words*, not by objects. More precisely, non-denoting singular terms are thought of as themselves contained in what Meyer and Lambert call the *semantical*—not outer—domain of a *nominal interpretation*, and then predicates are distributed in all possible ways over these new ‘entities’ to form the *logical points* over the nominal interpretation. A sentence is true (false) in a nominal interpretation if and only if it is true (false) in all logical points over it, and true (false) in the underlying *real interpretation* (which corresponds to a partial structure) if and only if it is true (false) in all the nominal interpretations which ‘complete’ it.

This approach is certainly suggestive, but insufficiently motivated, and it needs further elaboration before becoming really practicable. The main problem with non-denoting singular terms is that of explaining *why* a sentence like

(71) Pegasus is a horse

has whatever truth-value it has (if any), but just on this question the authors become elusive. 71 is true in a logical point—they say—not because the object Pegasus is a horse there, but because there the word ‘Pegasus’ is a *horse-word*. Again, this is suggestive, but what exactly is implied by being a horse-word? And how are horse-words to be identified if not in terms of the truth of sentences of the form

(72)  $\tau$  is a horse?

Unless we answer these questions (and the authors don’t), the whole strategy might look circular, and haunted by the ghost of an ultimately ‘conventional’ attitude.

The third attempt at developing the suggestions contained in supervaluational semantics at a ‘deeper’ level of analysis is due to [Skyrms, 1968], and can be described as resulting from *two* distinct applications of those suggestions. Straight supervaluational technique (that is, assignment of arbitrary truth-values to atomic sentences containing non-denoting singular terms, and subsequent construction of the logical product of all the valuations so obtained) is used with truth-functional compounds, whereas with

*atomic* sentences (and in particular identities) containing non-denoting singular terms Skyrms substantially adopts the counterfactual theory of truth, by constructing the logical product of (the truth-values the sentences in question have in) all the *extensions* of the original structure that assign denotations to the (originally) non-denoting singular terms. Quantified sentences are treated in yet a third way, that is, *in the standard way*:  $\forall x\varphi$  is true (or, in Skyrms' terminology, *holds*) in a structure  $\mathfrak{A}$  just in case  $\varphi$  holds in  $\mathfrak{A}$  for every assignment.

Skyrms' motivations are expressed very clearly. Frege seems to have thought—he says—that all sentences containing non-denoting singular terms should be truth-valueless, but supervaluations add 'an Aristotelian notion of Redemption to the Fregean notion of Sin', in that '*if the logical structure is such that every way of filling up the "holes" makes it true (false), then the sentence is true (false) regardless of the holes*' (his italics). 'Van Fraassen', he continues quite correctly, 'applies this idea only to the extent to which logical structure is determined by the sentential connectives', but 'identity is also a logical constant, and I suggest that we apply this idea to identity statements' (p. 479). Unfortunately, Skyrms stops short of noticing that quantifiers, too, are logical constants, and thus should also contribute to 'determining the logical structure'. As a result, the supervaluational idea is not applied to quantification, and very little of quantified logic is 'redeemed'. In particular, when  $a$  and  $b$  are non-denoting, not only 70 but also

$$(73) \quad \forall xRxa \rightarrow \forall yRya$$

$$(74) \quad \forall x(Qxa \rightarrow Rxa) \rightarrow (\forall xQxa \rightarrow \forall xRxa)$$

$$(75) \quad Pa \rightarrow \forall xPa$$

$$(76) \quad a = b \rightarrow (\forall xRxa \rightarrow \forall xRxb)$$

$$(77) \quad (\forall xPx \wedge \exists x(x = a)) \rightarrow Pa$$

are truth-valueless.

Skyrms does not propose a formal system adequate to his semantics, nor did anybody else, and in fact David Kaplan has apparently proved that the semantics in question is not recursively axiomatisable. On the other hand, the approach advocated by Bencivenga [1980b; 1981] falls well within the mainstream of the 'standard' free logics we presented in Part II.

To get immediately to the core of Bencivenga's semantics, let us concentrate on the solution he offers for the problem connected with 70. Once more, consider a structure  $\mathfrak{A}$  in which  $a$  is non-denoting. Suppose that the antecedent of 70 be true in  $\mathfrak{A}$ , and consider an extension  $\mathfrak{A}'$  of  $\mathfrak{A}$  which assigns a denotation to  $a$ . Of course, 70 is true in  $\mathfrak{A}'$ : if for example we



assume that its consequent be false there, its antecedent will be false, too. The situation can be depicted as follows:

	$\forall xPx$	$Pa$
$\mathfrak{A}$	$T$	$-$
$\mathfrak{A}'$	$F$	$F$

Now Bencivenga points out that, since we are trying to evaluate 70 *in*  $\mathfrak{A}$ , the truth-values this sentence has *in other structures* (such as  $\mathfrak{A}'$ ) are really of no independent interest. The only reason why we refer to  $\mathfrak{A}'$  and other extensions is that  $\mathfrak{A}$  gives no information about the consequent of 70, and we hope that this lack of information can be remedied by extending  $\mathfrak{A}$ . Thus in the case of 69, in which, too,  $\mathfrak{A}$  gives no direct information, we are able by extending it to determine the value True (and save the ‘logical law’ of self-identity). However, we must not forget the *purely instrumental* character of the extensions in question, and in particular must not let them prevail over the information  $\mathfrak{A}$  *already gives*. What this means—in terms of the above diagram—is that it is perfectly legitimate to take into account the truth-value assigned by  $\mathfrak{A}'$  to  $Pa$ , since  $\mathfrak{A}$  assigns no truth-value to it, but this truth-value should be combined—to the extent to which our evaluation procedure is relative to  $\mathfrak{A}$ —with the truth-value  $\mathfrak{A}$ —not  $\mathfrak{A}'$ —assigns to  $\forall xPx$ , since in this case  $\mathfrak{A}$  already gives a definite response, and one that  $\mathfrak{A}'$  does not ‘complete’, but simply contradicts. And of course if truth-values are combined in this way, 70 turns out to be false. In general, then, it is all right to extend  $\mathfrak{A}$  in all possible ways and to construct the logical product of all (the valuations relative to) such extensions, but in defining the valuations in question whatever information is provided by  $\mathfrak{A}$  must always weigh more than the information provided by the other (auxiliary) sources.

This discussion leads very naturally to the definition of a new technical instrument: the valuation  $V_{\mathfrak{A}'(\mathfrak{A})}^{**}$  for an extension (or more precisely, a ‘completion’)<sup>32</sup>  $\mathfrak{A}'$  of  $\mathfrak{A}$  *from the point of view of*  $\mathfrak{A}$ . Without entering into the details of this definition, we can say that  $V_{\mathfrak{A}'(\mathfrak{A})}^{**}$  is determined by  $\mathfrak{A}$  wherever  $\mathfrak{A}$  assigns definite truth-values, and is determined by  $\mathfrak{A}'$  elsewhere. The supervaluational instrument is then applied to all these  $V_{\mathfrak{A}'(\mathfrak{A})}^{**}$  (where  $\mathfrak{A}'$  is a completion of  $\mathfrak{A}$ ), and gives the final truth-values (or lack of truth-values) relative to  $\mathfrak{A}$ .

Bencivenga’s semantics does not show any of the asymmetries or oddities of Skyrms’. There are not three different ways of evaluating sentences, and the formal systems introduced in Part II are provably adequate to (suitable versions of) it. Furthermore, this semantics is not committed in any way to outer domains, for exactly the same reasons for which the counterfactual theory of truth in general is not. In this semantics, the truth-value of a

<sup>32</sup>A completion of  $\mathfrak{A}$  is an extension of  $\mathfrak{A}$  which assigns a denotation to *all* singular terms.

sentence in a given structure often depends on the truth-values some parts of the sentences have in other structures (in which the non-denoting singular terms occurring in the sentence are denoting), but this does not add to any structure any new category of objects, even less non-existent objects. There are different structures, and different sets of objects exist in them: this much seems pretty safe to say, and this is all that the semantics in question needs.

Of course, the price must be paid somewhere. Here the price of this metaphysical simplification is paid in terms of a number of *logical* complications, first and foremost the definition of  $V_{\mathfrak{A}'(\mathfrak{A})}^{**}$ . Some people (for example, [Posy, 1982]) have objected to this definition, mostly because the valuation in question does not correspond to *any* (single) structure: what is true (false) in it is not just what is true (false) in  $\mathfrak{A}$ , nor just what is true (false) in  $\mathfrak{A}'$ , but some combination of the two. To this objection one might answer that it seems to be a tendency of contemporary philosophical logic to regard structures as themselves constituting a *structure*, rather than just a set, that is, as bearing to one another *relations* that are semantically significant. Kripke's semantics for modal logic is a sign of this tendency, and Bencivenga's doubly determined valuations may be another (perhaps more radical) sign of it.

Before concluding the present section, something must be said about a few formal properties of supervaluations. Such properties have been proved within the context of van Fraassen's original semantics, but the proofs could be easily adapted to most of the variants we presented here.

Let us begin by considering the simple sentence

(78)  $Pa$ .

We have already mentioned (and used) the fact that in supervaluational semantics 78 is true in a structure  $\mathfrak{A}$  only if

(79)  $E!a$

is also true there. And we also noticed that 78 cannot even be *false* in  $\mathfrak{A}$  unless 79 is true, or, to put it otherwise, that 79 is a semantical consequence not only of 78 but also of

(80)  $\neg Pa$ .

There are important historical connections here. Frege [1892] and Strawson [1950; 1952] emphasised the role that relations of *presupposition* play in natural language. According to their characterisation, a sentence  $\varphi$  presupposes a sentence  $\psi$  just in case the truth of  $\psi$  is a necessary condition for  $\varphi$  to have any truth-value at all. Thus for example

(81) John stopped beating his wife

presupposes both

(82) John has a wife

and

(83) John used to beat his wife.

For, if either 82 or 83 were not true, 81 would be neither true nor false: it would simply represent a ‘spurious’ use of language.

Particularly important are the relations of *existential* presupposition. According to Frege and Strawson, a sentence like

(84) The present King of France is wise

does not *imply*

(85) The present King of France exists

(as Russell claimed), but rather presupposes it, and in general any simple sentence containing singular terms presupposes the existence of denotations for those terms.

Classical semantics cannot express any *non-trivial* relation of presupposition

(and in particular existential presupposition). For in classical semantics every sentence (in every situation) has a truth-value, and thus the only sentences that can be presupposed are the logically true ones. Strawson used this fact as evidence that formal logic is in general inadequate to deal with natural language. On the other hand—as is illustrated by the relations between 78 and 79—supervaluational semantics does allow for non-trivial relations of existential presupposition (since 79 is not logically true in it); hence this semantics constitutes an implicit answer to Strawson’s challenge.<sup>33</sup>

It is crucial to the above argument that supervaluational semantics is *non-bivalent*. There are less positive sides to this failure of bivalence. For example, van Fraassen proved that a suitable formal system of free logic is weakly complete with respect to his semantics, but proved also that this system is *not* strongly complete with respect to the same semantics.<sup>34</sup> The essence of the proof is as follows. Suppose the system *were* strongly complete. Then, since

(86)  $Pa \vDash E!a,$

---

<sup>33</sup>For some developments along these lines, see van Fraassen [1968; 1969]. Apparently, however, the discovery that supervaluations allow for non-trivial presuppositional relations is due to Lambert (see [van Fraassen, 1968, p. 151]).

<sup>34</sup>For the first result, see [van Fraassen, 1966a], for the second one see [van Fraassen, 1966b].

we would have

$$(87) Pa \vdash E!a,$$

and since the Deduction Theorem is provable for the system in question, we could also conclude that

$$(88) Pa \rightarrow E!a$$

is a theorem in it. But this is impossible, because 88 is not logically true and the system is provably sound.

Once again, it is the failure of bivalence that allows for this result. For when 78 is true, we know from 86 that 79—hence also 88—is true, and when 78 is false 88 is true on purely propositional grounds. But 78 can also be neither true nor false, and in that case 88 has no truth-value either—which explains why it is not logically true.

The result in question leaves two interesting problems open. On the one hand, there is the obvious problem of whether or not a *different* formal system could be strongly complete for supervaluational semantics—that is, whether or not the set of supervaluationally valid arguments is recursively enumerable. On the other hand, we know that in bivalent semantics weak completeness (for a given formal system) plus compactness gives strong completeness (for the same system), but there is no reason to think that this implication should hold when bivalence fails. In particular, the above argument against strong completeness makes no reference to infinite sets of sentences; hence it still leaves the possibility open that the semantics be compact. Whether or not it is, is our second problem.

Woodruff [1984] answered both questions in the negative. The set of supervaluationally valid arguments is not recursively enumerable, and supervaluational semantics is not compact. On the other hand, [Bencivenga, 1983] has shown that the quantifier-free fragment of the semantics *is* compact. This result is interesting because our argument against strong completeness does not depend on quantifiers either; hence in quantifier-free supervaluational semantics it is indeed the case that weak completeness plus compactness does *not* give strong completeness.

## IV: Extensions and Connections

### 11 FREE LOGIC AND CLASSICAL LOGIC

We already know that it is possible to deal with free logic as restricted quantification theory. We saw the semantical side of this when introducing the outer domain approach. A syntactical result along the same line was

proved by [Meyer and Lambert, 1968]. We will give now a brief sketch of their proof.

Let  $\mathbf{L}$  be a first-order language with the existence but without the identity symbol. Let  $\neg$  and  $\rightarrow$  be the only primitive connectives of  $\mathbf{L}$ , and  $\forall$  its only primitive quantifier. Let  $\mathbf{L}'$  be the result of adding a new monadic predicate constant  $Q$  to  $\mathbf{L}$ , and let a translation  $*$  of  $\mathbf{L}$  into  $\mathbf{L}'$  be defined as follows:

1.  $(P\tau_1 \dots \tau_n)^* = P\tau_1 \dots \tau_n$ ;
2.  $(E!\tau)^* = Q\tau$ ;
3.  $(\neg\varphi)^* = \neg(\varphi^*)$ ;
4.  $(\varphi \rightarrow \psi)^* = \varphi^* \rightarrow \psi^*$ ;
5.  $(\forall x\varphi)^* = \forall x(Qx \rightarrow \varphi^*)$ .

What Meyer and Lambert showed (in effect) is that a sentence  $\varphi$  of  $\mathbf{L}$  is a theorem of **FQCE!** if and only if  $\varphi^*$  is a theorem of classical logic (briefly, a *classical theorem*).

The ‘only if’ part of this biconditional is straightforward. One need only show that the translation of every axiom of **FQCE!** is a classical theorem, and that if  $\varphi^*$  and  $(\varphi \rightarrow \psi)^*$  are classical theorems, so is  $\psi^*$ .

The ‘if’ part is more complicated. The reason is obvious: classical logic is more powerful than free logic, hence it is not at all trivial that classical logic does not allow one to prove translations more than free logic allows to prove theorems. What we need here is a conservative extension result, showing that the more powerful deductive tools available in classical logic (in particular Specification) do not extend the class of provable sentences *of a certain form*.

First of all, then, we must give a clear *formulation* of the result we need. For this purpose, Meyer and Lambert construct, for every sentence  $\varphi^*$  of  $\mathbf{L}'$ , the sentence  $\varphi^{*1}$ , by substituting  $E!$  for  $Q$ . Given that

$$(89) \quad \forall x(E!x \rightarrow \psi) \leftrightarrow \forall x\psi$$

is provable in **FQCE!**,  $\varphi^{*1}$  is provably equivalent to  $\varphi$  in **FQCE!**. Furthermore, it is obvious that  $\varphi^{*1}$  is a classical theorem just in case  $\varphi^*$  is. In conclusion, our problem reduces to showing that  $\varphi^{*1}$  is a classical theorem only if it is provable in **FQCE!**—and this is the conservative extension result we need.

Of course, if  $\varphi^{*1}$  is a classical theorem, its proof (in classical logic) may well contain instances of Specification. However, Meyer and Lambert want to show that, given the particular form of  $\varphi^{*1}$ , all the instances of Specification that might be needed to prove it are also instances of the weaker schema

$$(90) \quad \forall x(E!x \rightarrow \psi) \rightarrow (E!\tau \rightarrow \psi[\tau/x]),$$

which is provable in **FQCE!**. And this would follow if it were possible to show that, every time a universal quantification  $\forall x\chi$  occurs in a proof of  $\varphi^{*1}$  (in classical logic),  $\chi$  is of the form  $E!x \rightarrow \psi$ .

That  $\varphi^{*1}$  itself has the above property is trivial, but in an axiomatic system this shows nothing about the structure of the sentences that can be used to prove  $\varphi^{*1}$ . However, a solution is readily at hand. It is enough to reformulate the problem within a Gentzen system for classical logic.<sup>35</sup> Since this system has the subformula property, we may be sure that if  $\vdash \varphi^{*1}$  is provable in it then universal quantifiers occur in the appropriate contexts *in the whole proof*, hence that the following Gentzen-variant of 90 is all that is ever applied in the proof:

$$\frac{\Gamma, E!\tau \rightarrow \psi[\tau/x] \vdash \Delta}{\Gamma, \forall x(E!x \rightarrow \psi) \vdash \Delta}.$$

This concludes Meyer and Lambert's argument.

An analogous (and simpler) result is available in the opposite direction. Let **L** be as before, except that it contains the identity but not the existence symbol. Consider the *exclusive* free logic **EFQC=** obtained by adding to **FQC=** the axiom-schema

$$(91) \quad \forall x\varphi \rightarrow \exists x\varphi,$$

and the translation  $^+$  of **L** into **L** defined as follows:

1.  $\varphi^+ = (\exists x(x = a_1) \wedge \dots \wedge \exists x(x = a_n)) \rightarrow \varphi$ , where  $a_1, \dots, a_n$  are all the individual constants occurring in  $\varphi$ .

It is easy to show that a sentence  $\varphi$  is a classical theorem if and only if  $\varphi^+$  is a theorem of **EFQC=**.

For more formal connections between free logics and classical logic, the reader may consult [Trew, 1970]. We prefer to close the present section by discussing an opinion that challenges the most common view of the relations between these two (kinds of) logics, a view that we have endorsed here.

It is quite natural to think of free logics as *alternatives* to classical logic.<sup>36</sup> After all, the people who created the subject were reacting against principles of classical logic (such as Specification) that they considered *wrong*. Van Fraassen [1969], however, would rather think of a free logic as an *extension* of a classical logic, obtained by adding to it a theory of singular terms that was simply not available in the classical framework.

<sup>35</sup>To be precise, Meyer and Lambert do not refer to a Gentzen system but to a variant of such a system proposed by Anderson and Belnap. But the essence of their argument is the same as given here.

<sup>36</sup>In the case of minimal free logics without  $E!$ , these alternatives qualify as *fragments* of classical logic.

The *rationale* of van Fraassen's position is as follows. We have already mentioned the fact that classical logic may be formalised so as to exclude both open theorems and individual constants. Quine [1940], for example, proceeds in this way. And Quine's system—which basically has no place for singular terms—is a *subsystem* of some free logics, for example of the pure exclusive free logic **EFQC**, which bears to **FQC** the same relation **EFQCE!** bears to **FQCE!**.<sup>37</sup>

Of course, there are formalisations of classical logic that do account for singular terms, for example by admitting individual constants, and they are subsystems of *no* free logic whatsoever, but in van Fraassen's opinion these formalisations were adopted *faute de mieux*. In absence of an adequate theory of singular terms, classical logicians extended to these terms the principles of their logic of bound variables (or 'bound' logic). The extension was faulty, but this fault did not touch the substantial validity of classical logic as a bound logic. Free logics on the other hand set things right, restricting classical logic to its proper scope and supplying a specific treatment of singular terms (indeed, several such treatments).

In assessing this argument, it is of fundamental importance to notice that it operates at three different levels. At bottom, there is the simple fact that free logics handle quantifiers and bound variables in the standard (referential) way. As we said a number of times, free logics confer existential import to quantifiers, and accept Quine's *dictum* that to be is to be a value of a bound variable. Next, there is the fact that *it is possible* to construe classical logic as a bound logic, and thus make it a subsystem of some free logic. But finally, there is also the suggestion that *it is better* to construe classical logic in this way, that such a construal is more likely to 'capture the spirit' of both classical and free logics, and that any other position on the matter would be adopted *faute de mieux*.

This last is basically a value judgement, and as such more prescriptive than descriptive in nature. Its supporters might claim that its adoption would allow one to maintain a conservative attitude with respect to logic, and possibly remove some psychological obstacles to accepting free logics. I would rather insist that such a claim of conservatism does very little historical justice to both classical and free logicians. For classical logicians—*pace* van Fraassen—did have their own views about singular terms, views that *they* at least considered adequate and that free logicians did very little to preserve.

---

<sup>37</sup>Of course, Quine's system is also a subsystem of **EFQCE!**, but its relation to **EFQC** is more interesting, because they have the same language.

## 12 DESCRIPTIONS

From the very beginning, free logicians were concerned with definite descriptions. There are at least two reasons for this interest. The first one is historical: Russell's description theory was one of the most important instruments in the hands of classical logicians to deal with (alleged) non-denoting singular terms, hence an important test for free logics was whether or not they were able to handle the same subject in a more satisfactory way. The second reason is theoretical: the necessity of a free logics is more apparent the more inevitable the presence of non-denoting singular terms seems to be, and certainly descriptions (if they are considered singular terms) make it very difficult to deny that there are non-denoting singular terms. You may think that 'Pegasus' does not denote, but no major problem would follow (apart of course from a conflict with your intuitions) if you were to decide instead that it does. On the other hand, if you decide that 'the winged horse' is denoting, this will appear to contradict the truth of

(92) No (existing) horse is winged,

and even worse consequences will follow if you decide that 'the round square' or 'the entity different from itself' are denoting.

The basic principles of Russell's description theory—as given for example in Whitehead and Russell [1910]—were the two definitions<sup>38</sup>

(93)  $E!ix\varphi =_{df} \exists y(\forall x(\varphi \leftrightarrow x = y))$

(94)  $\psi[ix\varphi/y] =_{df} \exists y(\forall x(\varphi \leftrightarrow x = y) \wedge \psi)$ .

However, free logicians usually regard definite descriptions as genuine singular terms; hence they are not interested in the elimination procedures connected with definitions like 93–94. Rather, they are interested in the acceptability of the corresponding biconditionals

(95)  $E!ix\varphi \leftrightarrow \exists y(\forall x(\varphi \leftrightarrow x = y))$

(96)  $\psi[ix\varphi/y] \leftrightarrow \exists y(\forall x(\varphi \leftrightarrow x = y) \wedge \psi)$ .

Now free logicians never questioned 95; they usually regarded its right-hand member as giving both a necessary and a sufficient condition for the existence of a denotation of  $ix\varphi$ . Similarly, one half of 96, that is,

(97)  $\exists y(\forall x(\varphi \leftrightarrow x = y) \wedge \psi) \rightarrow \psi[ix\varphi/y]$

<sup>38</sup>For precision's sake, it must be noted that the two Russellian definitions contained scope operators. But these operators play practically no role in free description theories (the only exception I know of is [Scales, 1969]): hence to simplify things we will disregard them here. Also, in this paragraph we will always assume that  $y$  is free in  $\psi$ .



is universally accepted in free logic: *denoting* definite descriptions appear to everybody to conform to Russell's analysis. What is in question is the other half of 96, that is,

$$(98) \psi[\iota x\varphi/y] \rightarrow \exists y(\forall x(\varphi \leftrightarrow x = y) \wedge \psi),$$

and the main reason why it is in question is that it implies

$$(99) \psi[\iota x\varphi/y] \rightarrow E!\iota x\varphi.$$

For 99 forces one to consider false most sentences containing non-denoting descriptions,<sup>39</sup> including such sentences as

$$(100) \iota x\varphi = \iota x\varphi,$$

which most free logicians regard as logically true.

The first free description theory was proposed by [Leonard, 1956], but in a second-order modal language—which explains why it did not generate much response in the literature. A more accessible suggestion came from [Hintikka, 1959b].

Hintikka's theory is based on a single principle, the biconditional

$$(101) \tau = \iota x\varphi \leftrightarrow (\varphi[\tau/x] \wedge \forall x(\varphi \rightarrow x = \tau)).$$

101 implies both 95 and 97, but it also has a number of unwelcome consequences. In particular, [Lambert, 1962] showed that

$$(102) \varphi[\iota x\varphi/x]$$

follows from 101 and 100, and some instances of 102, such as

$$(103) P(\iota x(Px \wedge \neg Px)) \wedge \neg P(\iota x(Px \wedge \neg Px)),$$

are contradictory sentences!

Lambert's own solution of this problem consists in weakening Hintikka's theory, by substituting

$$(104) \forall y(y = \iota x\varphi \leftrightarrow (\varphi[y/x] \wedge \forall x(\varphi \rightarrow x = y)))$$

for 101. Now in a free logic assuming 104 as an axiom-schema is equivalent to assuming

$$(105) \underline{E!\iota x\varphi \rightarrow (\tau = \iota x\varphi \leftrightarrow (\varphi[\tau/x] \wedge \forall x(\varphi \rightarrow x = \tau)))},$$

<sup>39</sup>Again, we must notice that unless we adopt scope operators or deny to descriptions the status of singular terms (and Russell did both) 99 leads to downright inconsistency. But the main point of the argument is independent of this, since it can be made in connection with such simple sentences as 100. So once more we need not enter into unnecessary complications.

which shows that Lambert's theory—to be called **FD**—has something specific to say about descriptions<sup>40</sup> only to the extent to which they are denoting. For this reason, several authors (including Lambert, van Fraassen and Scott)<sup>41</sup> have considered **FD** a *minimal* free description theory, the common core as it were of all such theories.

The intuitive idea behind this characterisation is that, again, everybody agrees on how to treat denoting descriptions, and **FD** says nothing (specific) beyond that. Disagreements will arise among free description theorists only with respect to non-denoting descriptions and in this area a large number of alternatives are possible, which in general require the addition of further schemata to **FD**. Lambert [1962] mentions one of these alternatives, that is, the theory (to be called **FD**<sub>1</sub>) which results from adding to **FD** the schema

$$(106) \quad \tau = \iota x(x = \tau),$$

and [Lambert, 1964] a different one, obtained by replacing 104 with

$$(107) \quad \iota x\varphi = \tau \leftrightarrow \forall y(\tau = y \leftrightarrow (\varphi[y/x] \wedge \forall x(\varphi \rightarrow x = y))),$$

from which however 104 is derivable.

This last theory—to be called **FD**'<sub>2</sub>—is an interesting one. For it turns out that it is equivalent to the theory **FD**<sub>2</sub> which is obtained by adding to **FQCE**! = 104 and the principle 41 on p. 166—that is, by combining minimal description theory with Scott's free logic.

Van Fraassen and Lambert [1967] make some interesting remarks about the philosophical significance of the differences between all these description theories (which remarks—in view of the above equivalence result—apply *mutatis mutandis* to Scott's free logic). **FD**<sub>2</sub> (or **FD**'<sub>2</sub>)—they say—may be the right theory for some specific (and limited) purposes. For example, in the course of reconstructing mathematics non-denoting descriptions may well be regarded as 'don't cares', and a compromise between free logic and the chosen object theory may be the most efficient way to handle them. On the other hand, if we are interested in natural language, **FD**<sub>2</sub> is going to be too strong. To give just one example, such a theory would allow us to derive

$$(108) \quad \text{John avoided the explosion of the White House in 1965,}$$

from

$$(109) \quad \text{John avoided the accident at the corner of High Street and Pleasant Street,}$$

---

<sup>40</sup>That is, something that does not follow simply from treating descriptions as genuine singular terms, and thus extending to them the laws of (free) quantification and identity theory.

<sup>41</sup>See [van Fraassen and Lambert, 1967; Scott, 1970; Lambert, 1972].

and the fact that there was no accident at that corner or any explosion of the White House in 1965. For these other, more ‘philosophical’ purposes, a weaker theory like **FD** or **FD**<sub>1</sub> might be preferred.

The ‘liberality’ of this position is certainly attractive; however, there are problems with it. On the one hand, if **FD**<sub>2</sub> or **FD**'<sub>2</sub> are recommended on pragmatic grounds, to people who don't care much about non-denoting descriptions, how can they be preferred to the original chosen object theory—which is certainly simpler and thus even more recommendable from a pragmatic point of view? On the other hand, if we reject **FD**<sub>2</sub> and move to weaker theories, how are we going to choose among them? The intuitive acceptability of ‘laws’ like 106 by itself won't do, for we need a way of checking our intuitions on the matter, and even more importantly we need some kind of evidence that we have found *all* the relevant laws.

The problem with van Fraassen and Lambert's approach is that they do not give a semantical analysis of their theories, except for the minimal **FD**. They do present semantics for all these theories, and completeness theorems for them, but such ‘semantics’ do little more than duplicating the theories, and the arbitrary selections that seem to be at their foundations. Thus for example the fundamental unit of the ‘semantics’ for **FD**<sub>1</sub>—the **FD**<sub>1</sub>-structure—is defined essentially as an **FD**-structure that verifies all instances of 106, and such an approach certainly says very little about *why* 106 is a logical law, and which other laws (if any) should be accepted. This leaves us with the semantics for **FD**, but **FD** is a very weak theory, too weak even for its author Lambert.<sup>42</sup>

The above is substantially the same criticism already raised against van Fraassen's semantics for (free) quantification and identity theory. Just as in that case (say) self-identity was validated by *fiat*, so it happens now for 106. Thus we may expect to find here the same kinds of developments of van Fraassen's approach that we found there. And indeed, at least one such development is available, by Bencivenga [1978b; 1980c].

Once again, Bencivenga's starting point is the counterfactual theory of truth: a sentence containing non-denoting singular terms is true (false) if and only if it would be true (false) in case these terms were denoting.

However, a major complication arises in applying this theory to descriptions, in that it is not always possible for a description to denote, or for a set of descriptions to denote simultaneously. Thus

$$(110) \quad \iota x(Px \wedge \neg Px)$$

will never have a denotation (if not in some variant of the chosen object theory, which Bencivenga is not willing to accept), and

$$(111) \quad \iota xPx$$

---

<sup>42</sup>In this regard, see the conclusion of [Lambert, 1962].

$$(112) \quad \iota x \neg Px$$

$$(113) \quad \iota x(x = x),$$

though being all ‘consistent’ descriptions (and having denotations *somewhere*) will never have denotations *together*.

With respect to sentences containing either 110 or all of 111–113, Bencivenga faces a choice: he can either make them all vacuously true (or perhaps false) or modify his approach, for example by requiring as an additional condition for the truth (or falsity) of sentences containing non-denoting descriptions that there be at least one structure in which all such descriptions are denoting. He chooses the second route, and this choice gives rise to further complications. For now every sentence containing (say) 110, even such a sentence as

$$(114) \quad \iota x(Px \wedge \neg Px) = \iota x(Px \wedge \neg Px),$$

whose logical truth seems not to depend on a logical analysis of descriptions, becomes ‘essentially truth-valueless’—that is, does never receive a truth-value.

Bencivenga’s assessment of the situation is that free quantification and identity theories, though successful in removing the existential assumptions of classical logic, still carry with them some weaker assumptions, of possibility of existence. Such assumptions, however, are contradicted by (some) descriptions, hence our quantificational logic should be modified if we want to allow for a natural extension of it to descriptions. Whether we will actually make the modification in question or instead worry about assumptions of possibility where they really matter (that is, in languages with descriptions) will ultimately be decided—Bencivenga thinks—on practical grounds, and certainly relevant to these practical considerations is his proof that the set of logically true sentences of his (possibility-free) semantics for descriptions is not recursively enumerable.

So much for the extensions to descriptions of the supervaluational approach. Analogous extensions of the outer domain approach and of the ‘conventional’ approach were proposed by [Grandy, 1972] and by [Burge, 1974], respectively. Since Grandy’s development is less immediate than Burge’s, and contains at least one new theoretical notion, we will conclude the present section by briefly describing it.

The novelty of Grandy’s semantics is a function  $\pi$ , defined on all subsets of the union  $A \cup A'$  of the inner and the outer domain of an LT-structure and with values in  $A \cup A'$ . By definition,  $\pi(S) \in A$  if and only if  $S \cap A = \pi(S)$ , that is, the value of  $\pi$  for a given subset  $S$  of  $A \cup A'$  ‘exists’ just in case it is the only existing member of  $S$ . In a *Grandy structure*, defined as an ordered 4-tuple  $\langle A, A', I, \pi \rangle$ , the denotation of a description  $\iota x \varphi$  is the value  $\pi$  has for the subset of  $A \cup A'$  constituted by all objects satisfying  $\varphi$ .

Grandy's semantics does not force one to identify all non-existents, as for example Scott's does. Thus, in intuitive terms, if  $\pi$  assigns different values to the set of winged horses and the set of golden mountains, the sentence

(115) The winged horse = the golden mountain

turns out false. On the other hand, however, if

(116)  $\varphi[\tau/x] \leftrightarrow \psi[\tau/x]$

is *logically true*, then in every Grandy structure  $\varphi$  and  $\psi$  are satisfied by the same objects; hence

(117)  $\imath x\varphi = \imath x\psi$

is logically true, too. This is certainly an asset of Grandy's approach: in it one can validate in a natural way such schemata as

(118)  $\imath x\varphi = \imath x(\varphi \wedge \psi)$ ,

which are certainly as 'intuitive' as (say) 106 was and which in van Fraassen-Lambert's framework would require the addition of further *ad hoc* clauses. From a proof-theoretical point of view, the approach in question is characterised by the rule

(119) 
$$\frac{\vdash \chi \rightarrow (\varphi[\tau/x] \leftrightarrow \psi[\tau/x])}{\vdash \chi \rightarrow \imath x\varphi = \imath x\psi}, \text{ if } \tau \text{ does not occur in } \chi,$$

which allows for a simple proof of 118 and the like.

#### ACKNOWLEDGEMENTS

Work for the completion of this paper was partly supported by a Faculty Fellowship of the School of Humanities, University of California at Irvine. Thanks are due to Nuel Belnap, Gerald Charlwood, Wilfrid Hodges, Karel Lambert and Brian Skyrms for comments on earlier drafts of the paper.

*University of California at Irvine*

#### BIBLIOGRAPHY

- [Barba., 1989] J. L. Barba. A modal version of free logic, *Topoi*, **9**, 131–135, 1989.
- [Belnap, 1960] N. D. Belnap, Jr. Review of [Hintikka, 1959a], *Journal of Symbolic Logic*, **25**, 88, 1960.
- [Bencivenga, 1976] E. Bencivenga. Set theory and free logic. *Journal of Philosophical Logic*, **5**, 1–15, 1976.
- [Bencivenga, 1978a] E. Bencivenga. A semantics for a weak free logic. *Notre Dame Journal of Formal Logic*, **19**, 646–652, 1978.

- [Bencivenga, 1978b] E. Bencivenga. Free semantics for indefinite descriptions. *Journal of Philosophical Logic*, **7**, 389–405, 1978.
- [Bencivenga, 1980a] E. Bencivenga. A weak free logic with the existence sign. *Notre Dame Journal of Formal Logic*, **231**, 572–576, 1980.
- [Bencivenga, 1980b] E. Bencivenga. *Una logica dei termini singolari*, Boringhieri, Torino, 1980.
- [Bencivenga, 1980c] E. Bencivenga. Free semantics for definite descriptions. *Logique et Analyse*, **23**, 393–405, 1980.
- [Bencivenga, 1981] E. Bencivenga. Free semantics. *Boston Studies in the Philosophy of Science*, **47**, 31–48, 1981.
- [Bencivenga, 1983] E. Bencivenga. Compactness of a supervaluational language. *Journal of Symbolic Logic*, **48**, 384–386, 1983.
- [Bencivenga, 1990] E. Bencivenga. Free from what? *Erkenntnis*, **33**, 9–21, 1990.
- [Bencivenga, Lambert and van Fraassen, 1991] E. Bencivenga, K. Lambert and B. C. van Fraassen. *Logic, Bivalence and Denotation*, 2nd edition, Ridgeview, Atascadero, (California), 1991.
- [Burge, 1974] T. Burge. Truth and singular terms. *Nous*, **8**, 309–325, 1974.
- [Carnap, 1947] R. Carnap. *Meaning and Necessity*. University of Chicago Press, Chicago, 1947.
- [Church, 1965] A. Church. Review of [Lambert, 1963]. *Journal of Symbolic Logic*, **30**, 103–104, 1965.
- [Cocchiarella, 1966] N. Cocchiarella. A logic of possible and actual objects. *Journal of Symbolic Logic*, **31**, 688, 1966.
- [Findlay, 1963] A. N. Findlay. *Meinong's Theory of Objects and Values*, Clarendon Press, Oxford, 1963.
- [Fine, 1983] K. Fine. The permutation principle in quantificational logic. *Journal of Philosophical Logic*, **12**, 33–37, 1983.
- [Frege, 1892] G. Frege. Über Sinn und Bedeutung. *Zeitschrift für Philosophie und philosophische Kritik*, **100**, 25–50, 1892.
- [Grandy, 1972] R. Grandy. A definition of truth for theories with intensional definite description operators. *Journal of Philosophical Logic*, **1**, 137–155, 1972.
- [Hailperin, 1953] T. Hailperin. Quantification theory and empty individual-domains. *Journal of Symbolic Logic*, **18**, 197–200, 1953.
- [Hintikka, 1959a] J. Hintikka. Existential presuppositions and existential commitments. *Journal of Philosophy*, **56**, 125–137, 1959.
- [Hintikka, 1959b] J. Hintikka. Towards a theory of definite descriptions. *Analysis*, **19**, 79–85, 1959.
- [Jaskowski, 1934] S. Jaskowski. On the rules of supposition in formal logic. *Studia Logica*, **1**, 5–32, 1934.
- [Lambert, 1962] K. Lambert. Notes on E! III: A theory of descriptions. *Philosophical Studies*, **13**, 51–59, 1962.
- [Lambert, 1963] K. Lambert. Existential import revisited. *Notre Dame Journal of Formal Logic*, **4**, 288–292, 1963.
- [Lambert, 1964] K. Lambert. Notes on E! IV: A reduction in free quantification theory with identity and descriptions. *Philosophical Studies*, **15**, 85–88, 1964.
- [Lambert, 1967] K. Lambert. Free logic and the concept of existence. *Notre Dame Journal of Formal Logic*, **8**, 133–144, 1967.
- [Lambert, 1972] K. Lambert. Notes on free description theories: some philosophical issues and consequences. *Journal of Philosophical Logic*, **1**, 184–191, 1972.
- [Lambert, 1981] K. Lambert. On the philosophical foundations of free logic. *Inquiry*, **24**, 147–203, 1981.
- [Lambert, 1991] K. Lambert. *Philosophical Applications of Free Logic*, Oxford University Press, New York, 1991.
- [Leblanc and Hailperin, 1959] H. Leblanc and T. Hailperin. Nondesignating singular terms. *Philosophical Review*, **68**, 239–243, 1959.
- [Leblanc and Thomason, 1968] H. Leblanc and R. H. Thomason. Completeness theorems for some presupposition-free logics. *Fundamenta Math.*, **62**, 125–26, 1968.

- [Lejewski, 1954] C. Lejewski. Logic and existence. *British J. Philosophy of Science*, **5**, 104–119, 1954.
- [Lejewski, 1958] C. Lejewski. On Lesniewski's ontology. *Ratio*, **1**, 150–176, 1958.
- [Leonard, 1956] H. S. Leonard. The logic of existence. *Philosophical Studies*, **7**, 49–64, 1956.
- [Luschei, 1962] E. C. Luschei. *The Logical Systems of Lesniewski*. North-Holland, Amsterdam, 1962.
- [Meinong, 1904] A. Meinong. Über Gegenstandstheorie. In *Untersuchungen zur Gegenstandstheorie und Psychologie*, Barth, Leipzig, 1904.
- [Meyer and Lambert, 1968] R. K. Meyer and K. Lambert. Universally free logic and standard quantification theory. *Journal of Symbolic Logic*, **33**, 8–26, 1968.
- [Mostowski, 1951] A. Mostowski. On the rules of proof in the pure functional calculus of the first order. *Journal of Symbolic Logic*, **16**, 107–111, 1951.
- [Parsons, 1980] T. Parsons. *Nonexistent Objects*, Yale University Press, New Haven, 1980.
- [Posy, 1982] C. Posy. A free IPC is a natural logic. *Topoi*, **1**, 30–43, 1982.
- [Quine, 1939] W. V. O. Quine. Designation and existence. *Journal of Philosophy*, **36**, 701–709, 1939.
- [Quine, 1940] W. V. O. Quine. *Mathematical Logic*, Harvard University Press, Cambridge, 1940.
- [Quine, 1948] W. V. O. Quine. On what there is. *Review of Metaphysics*, **2**, 21–38, 1948.
- [Quine, 1954] W. V. O. Quine. Quantification and the empty domain. *Journal of Symbolic Logic*, **19**, 177–179, 1954.
- [Routley, 1966] R. Routley. Some things do not exist. *Notre Dame Journal of Formal Logic*, **7**, 251–276, 1966.
- [Routley, 1980] R. Routley. *Exploring Meinong's Jungle and Beyond*. Australian National University, Canberra, 1980.
- [Scales, 1969] R. Scales. *Attribution and Reference*. PhD Thesis, University of California at Irvine, 1969.
- [Schock, 1964] R. Schock. Contributions to syntax, semantics, and the philosophy of science. *Notre Dame Journal of Formal Logic*, **5**, 241–289, 1964.
- [Schock, 1968] R. Schock. *Logics without Existence Assumptions*. Almqvist and Wiksell, Stockholm, 1968.
- [Scott, 1967] D. Scott. Existence and description in formal logic. In R. Schoenman, ed. *Bertrand Russell, Philosopher of the Century*, pp. 181–200. Allen and Unwin, London, 1967.
- [Scott, 1970] D. Scott. Advice in modal logic. In K. Lambert, ed. *Philosophical Problems in Logic*, pp. 143–173. D. Reidel, Dordrecht, 1970.
- [Skyrms, 1968] B. Skyrms. Supervaluations: identity, existence, and individual concepts. *Journal of Philosophy*, **69**, 477–482, 1969.
- [Strawson, 1950] P. F. Strawson. On referring. *Mind*, **59**, 320–344, 1950.
- [Strawson, 1952] P. F. Strawson. *Introduction to Logical Theory*. Methuen, London, 1952.
- [Trew, 1970] A. Trew. Nonstandard theories of quantification and identity. *Journal of Symbolic Logic*, **35**, 267–294, 1970.
- [van Fraassen, 1966a] B. C. van Fraassen. The completeness of free logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **12**, 219–234, 1966.
- [van Fraassen, 1966b] B. C. van Fraassen. Singular terms, truth-value gaps, and free logic. *Journal of Philosophy*, **67**, 481–495, 1966.
- [van Fraassen, 1968] B. C. van Fraassen. Presupposition, implication, and self-reference. *Journal of Philosophy*, **69**, 136–152, 1968.
- [van Fraassen, 1969] B. C. van Fraassen. Presuppositions, supervaluations, and free logic. In K. Lambert, ed. *The Logical Way of Doing Things*, pp. 67–91. Yale University Press, New Haven, 1969.
- [van Fraassen and Lambert, 1967] B. C. van Fraassen and K. Lambert. On free description theory. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **13**, 225–240, 1967.

- [Whitehead and Russell, 1910] A. N. Whitehead and B. Russell. *Principia Mathematica*, Vol. 1, Cambridge University Press, Cambridge, 1910.
- [Woodruff, 1971] P. W. Woodruff. Free logic, modality and truth. Unpublished manuscript, 1971.
- [Woodruff, 1984] P. W. Woodruff. On supervaluations in free logic. *Journal of Symbolic Logic*, **49**, 943–950, 1984.



## MORE FREE LOGIC

By a *free* logic is generally meant a variant of classical first-order logic in which constant terms may, under interpretation, fail to refer to individuals in the domain  $D$  over which the bound variables range, either because they do not refer at all or because they refer to individuals outside  $D$ . If  $D$  is identified with what is assumed by the given interpretation to exist, in accord with *Quine's dictum* that “to be is to be the value of a [bound] variable,”<sup>1</sup> then a free variation on classical semantics does not require that all constant terms refer to existents, and in this sense such terms lack existential import.

Classical semantics treats free variables like constants, at least in the quantifier clause of the valuation rules. When we stipulate that  $\exists xA$  is true iff  $A$  is true for some assignment of a value  $\alpha(x)$  in  $D$  to  $x$ , we are treating  $x$  at its free occurrences in  $A$  as a constant that refers to  $\alpha(x)$ . In free semantics, free variables are also generally treated like constants, which means that they need not be assigned values in  $D$ ; thus free variables and variable terms (such as  $x + y$  or  $1/x$ ) constructed from them also lack existential import. However, when reckoning the truth of  $\exists xA$  in terms of the truth of  $A$  for assignments of values to  $x$ , we consider only assignments  $\alpha$  for which  $\alpha(x) \in D$ . Thus, although neither constant nor variable terms need refer to individuals in  $D$ , free semantics honors Quine's dictum.<sup>2</sup>

In classical semantics, free variables have existential import because  $D$  is non-empty: there is always something in  $D$  for  $x$  to be assigned by  $\alpha$ . Variants of classical semantics in which this requirement is relaxed so that  $D$  may be empty are said to be *inclusive*. A semantics that is free and inclusive is said to be *universally* free: the range of the bound variables may be empty, and even if it is not, neither constant nor variable terms have existential import.

This survey of free logic will begin by considering its motivation, then move to reviewing various kinds of free semantics and the syntactic proof systems designed to capture the forthcoming notions of logical truth or logical consequence, and conclude by describing some applications of free logics, notably free description theory. As this summary may suggest, my emphasis throughout will be on semantics. The account is self-contained

---

<sup>1</sup>Quine [1948, p. 15]. That Quine means *bound* variables here is clear from his earlier statement [p. 13] that “a theory is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order for the affirmations made in the theory to be true.”

<sup>2</sup>Compare Bencivenga's [1986, p. 375] characterization: “A free logic is a formal system of quantification theory, with or without identity, which allows for some singular terms in some circumstances to be thought of as denoting no existing object, and in which quantifiers are invariably thought of as having existential import.”

and does not presuppose familiarity with [Bencivenga, 1986], reproduced in this volume. There is new material on motivation, applications, and neutral free semantics, while areas of overlap differ in detail and emphasis. Semantic options are laid out in greater detail, as are free description theories built upon them, but I pay less attention to the history of ideas.

## 1 QUICK REVIEW OF CLASSICAL FIRST-ORDER LOGIC

Partly to settle notation and terminology, and partly because free logics are variants of it, let us first quickly review classical first-order logic with identity, which we may take to be framed in formal first-order languages  $L$ .

The logical vocabulary of  $L$  includes the identity operator  $=$ , plus an adequate set of quantifiers and truth-functional operators. Let us assume they are the universal quantifier  $\forall$ , negation  $\neg$ , and material conditional  $\rightarrow$ . The non-logical vocabulary of  $L$  includes (individual) variables, plus perhaps constants, ( $k$ -place) function-names, and ( $k$ -place) predicates. We need not specify these symbols, which will vary with  $L$ ; its sentences are to represent the logical forms of certain sentences of natural language, and its non-logical vocabulary will be chosen accordingly. Many formulations of free logic employ a 1-place existence predicate  $E$  or  $E!$ , but such a predicate can generally be defined in terms of identity,<sup>3</sup> so we need not include it in the non-logical vocabulary. The proof method of  $L$  will require an unbounded list of variables or special constants.

After defining *terms* as (1) variables, (2) names, and (3) complex terms  $ft_1 \dots t_k$ , where the  $t_i$  are terms and  $f$  is a  $k$ -place function-name, the formation rules of  $L$  identify *formulae* as (4) subject-predicate formulae  $Pt_1 \dots t_k$ , where the  $t_i$  are terms and  $P$  is a  $k$ -place predicate, (5) identities  $s = t$ , where  $s$  and  $t$  are terms, (6) negations  $\neg A$ , where  $A$  is a formula, (7) conditionals  $(A \rightarrow B)$ , where  $A$  and  $B$  are formulae, and (8) universals  $\forall xA$ , where  $x$  is a variable and  $A$  is a formula.<sup>4</sup> Identities and subject-predicate formulae are *atomic*; atomic formulae and their negations are *elementary*.

Subsequently, I shall use the following syntactical variables, with or without subscript: for variables:  $x$ ,  $y$ , and  $z$ ; for constants:  $a$ ,  $b$ , and  $c$ ; for function-names:  $f$ ; for predicates:  $P$ ; for terms:  $s$  and  $t$ ; for formulae:  $A$ ,  $B$ , and  $C$ ; for sets of formulae:  $X$ . Conjunctions  $(A \& B)$ , disjunctions  $(A \vee B)$ , biconditionals  $(A \leftrightarrow B)$ , and existentials  $\exists xA$  may be defined as usual in terms of  $\neg$ ,  $\rightarrow$ , and  $\forall$ .  $s \neq t$  abbreviates  $\neg s = t$ . The outermost parentheses in conditionals, conjunctions, disjunctions, and biconditionals standing alone will be omitted.  $ft_1 \dots t_k$  and  $Pt_1 \dots t_k$  will be used with

<sup>3</sup>For exceptions, see [Garson, 1991], discussed below in Section 5.3, and [Gumb, 1998].

<sup>4</sup>To avoid the notational clutter that attends the use of single- and quasi-quotation, I shall generally follow Church [1956] in using symbols of  $L$  as names for themselves and juxtaposition for juxtaposition.

the assumption that  $f$  and  $P$  are  $k$ -place; where necessary, commas and parentheses will disambiguate expressions, as in  $Pf(x, y)$ , and may also be inserted to enhance readability, as in  $\exists x(x = fx)$ .

An occurrence of a term  $t$  in a formula  $A$  is *bound* in  $A$  provided it is an occurrence in a part  $\forall xB$  of  $A$ , where  $x$  occurs in  $t$ ; an occurrence of  $t$  in  $A$  is *free* if it is not bound. The *bound (free)* variables of  $A$  are those with a bound (free) occurrence in  $A$ . A *sentence* is a formula without free variables.  $A(x_1, \dots, x_k/t_1, \dots, t_k)$  is the result of simultaneously replacing the  $x_i$  at each free occurrence in  $A$  by  $t_i$ , having (if necessary) first made such occurrences *free for*  $t_i$  in  $A$ : if a free occurrence of  $x_i$  in  $A$  is in a part  $\forall yB$ , where  $y$  occurs in  $t_i$ , replace each occurrence of  $y$  in  $\forall yB$  by the first variable that occurs in neither  $A$  nor any of the  $t_j$ ; relabel the result ' $A$ ' and repeat until there are no such occurrences. I shall write  $A(x_1, \dots, x_k)$  for  $A$  and  $A(t_1, \dots, t_k)$  for  $A(x_1, \dots, x_k/t_1, \dots, t_k)$ .  $\exists!xA$  or  $\exists!xA(x)$  abbreviates  $\exists x\forall y(A(x/y) \leftrightarrow y = x)$ , where  $y$  is not  $x$ . In writing  $\exists x(x = t)$ , I assume that  $x$  does not occur in  $t$ . The *universal closure*  $\forall A$  of  $A$  is  $\forall x_1 \dots \forall x_k A$ , where the free variables of  $A$  are  $x_1, \dots, x_k$ .

An *interpretation*  $I$  of  $L$  is a pair  $\langle D, d \rangle$ , where  $D$  is a set and  $d$  is a denotation function defined on the constants, function-names, and predicates of  $L$ , such that:

- i1.  $D$  is non-empty;
- i2.  $d(a) \in D$ ;
- i3. If  $f$  is  $k$ -place,  $d(f)$  is a total  $k$ -ary function  $D \rightarrow D$ .
- i4. If  $P$  is  $k$ -place,  $d(P)$  is a  $k$ -ary relation in  $D$ .

An *assignment*  $\alpha$  is a function that assigns individuals  $\alpha(x)$  in  $D$  to the variables. An  $x$ -variant of  $\alpha$  is an assignment that differs from  $\alpha$  at most at  $x$ .

Under  $I$  and  $\alpha$ , terms refer to individuals of  $D$  according to the reference rules:

- r1.  $x$  refers to  $\alpha(x)$ .
- r2.  $a$  refers to  $d(a)$
- r3.  $ft_1 \dots t_k$  refers to  $d(f)(\alpha_1, \dots, \alpha_k)$ , if  $t_i$  refers to  $\alpha_i$ .

Under  $I$  and  $\alpha$ , formulae are true or false (and false if not true) according to the valuation rules:

- v1.  $Pt_1 \dots t_k$  is true iff  $\langle \alpha_1, \dots, \alpha_k \rangle \in I(P)$ , if  $t_i$  refers to  $\alpha_i$ .
- v2. If  $s$  refers to  $\alpha$  and  $t$  to  $\beta$ , then  $s = t$  is true iff  $\alpha$  is  $\beta$ .

- v3.  $\neg A$  is true iff  $A$  is false.
- v4.  $A \rightarrow B$  is false iff  $A$  is true and  $B$  is false.
- v5.  $\forall x A$  is false iff  $A$  is false for some  $x$ -variant of  $\alpha$ .<sup>5</sup>

Since the referents of terms without variables and the truth-values of sentences are independent of  $\alpha$ , I shall speak of referents and truth-values under  $I$  in such cases.

Logical relations and properties are defined as usual in terms of the totality of interpretations:  $A$  is a *logical consequence* of  $X$  ( $X \models A$ ) iff there is no interpretation and assignment under which all the  $X$ -formulae are true and  $A$  is false;  $X$  is *satisfiable* iff there is some interpretation and assignment under which all the  $X$ -formulae are true;  $A$  is *logically true (false)* iff  $A$  is true (false) under each interpretation and assignment;  $A$  and  $B$  are *logically equivalent* iff, under each interpretation and assignment,  $A$  is true iff  $B$  is true.  $A_1, \dots, A_k \models B$  means:  $\{A_1, \dots, A_k\} \models B$ .  $X, A \models B$  means:  $X \cup \{A\} \models B$ .  $X \not\models A$  means: not  $X \models A$ .

These definitions embody what Kleene [1967, p. 103] terms the *conditional* reading of free variables: free variables are treated by r1 as names of  $D$ -individuals. By contrast, the *generality* reading treats free variables as if they were universally quantified. It may be captured by stipulating that  $A$  is true (false) under  $I$  iff  $A$  is true (false) under  $I$  and  $\alpha$  for each  $\alpha$ . We can then drop “and assignment” from the above definitions. However, we end up with weaker notions of logical consequence and logical equivalence (and a stronger notion of satisfiability). For the logical consequence relation  $\models_g$ , we have  $X \models_g A$  iff  $\forall X \models \forall A$ , where  $\forall X = \{\forall B : B \in X\}$ , so that  $X \models_g A$  if  $X \models A$  but not conversely (e.g.,  $Px \models_g \forall x Px$ , but  $Px \not\models \forall x Px$ ). If  $X$  is a set of sentences, the two consequence relations coincide, since  $X \models A$  iff  $X \models \forall A$  and here we have  $\forall X = X$ .

From the semantic perspective assumed here, the aim of proof theory is to provide syntactic characterizations of logical properties and relations, which are defined in semantic terms. In particular, we want a syntactic notion of *proof from hypotheses* that captures the logical consequence relation:  $A$  is provable from hypotheses in  $X$  ( $X \vdash A$ ) iff  $A$  is a logical consequence of  $X$  ( $X \models A$ ), at least if  $X$  is a set of sentences. A proof system with this property is said to be *strongly complete*. A proof system in which  $A$  is

<sup>5</sup>Most presentations of free logic give a substitutional account of quantification, on which v5 would read instead:  $\forall x A(x)$  is false iff  $A(a)$  is false for some constant  $a$ . If  $\forall x$  is to have the force of ‘for all individuals  $x$ ’, every individual in  $D$  must be named by some constant or other. If  $D$  is uncountable, the terms and formulae of  $L$  will then be undecidable. This awkward result may be avoided by proving, *via* the Löwenheim-Skolem theorem, that interpretations may be restricted to countable universes without altering logical consequence relations, so that no more than a countable infinity of constants need be assumed. By contrast, the objectual account of quantification given in v5 does not require an elaborate justification.

provable (from no hypotheses) iff  $A$  is logically true ( $\vdash A$  iff  $\models A$ ) is *weakly complete*. If detachment or *modus ponens*

MP  $A, A \rightarrow B \vdash B$

holds, and the deduction theorem or conditional proof

CP  $X \vdash A \rightarrow B$  provided  $X, A \vdash B$

holds for sentences  $A$ , then weak completeness is equivalent to:  $X \vdash A$  iff  $X \models A$  for *finite* sets  $X$  of sentences.

Many strongly complete systems in a variety of styles are known for classical first-order semantics. It will be useful to give one that can be modified in simple ways to capture logical consequence for at least some free variations on classical semantics. The simplest proof systems to describe are Hilbert-style systems, which specify logical axioms and rules of inference, and define a proof of  $A$  from hypotheses  $X$  as a finite sequence  $\langle A_1, \dots, A_k \rangle$  such that  $A_k = A$  and each  $A_i$  is either a member of  $X$ , or a logical axiom, or is derived from previous formulae in the sequence by a rule of inference. Unlike natural deduction systems, in which some inference rules (such as CP) are conditional, those of a Hilbert-style system are (like MP) categorical.

Since the propositional part of the system does not matter here, we may adopt the simple inference rule

T  $A_1, \dots, A_k \vdash B$

if  $B$  is a tautological consequence of  $\{A_1, \dots, A_k\}$ , that is, there is no assignment of truth-values to universals and atomic formulae for which each  $A_i$  is true and  $B$  is false in virtue of rules v3 and v4.

The quantifier rule and axioms are as in [Church, 1956, p. 172]; the rule is generalization:

UG  $A \vdash \forall xA$

and the axiom schemas are distribution and specification:

A1  $\forall x(A \rightarrow B) \rightarrow (A \rightarrow \forall xB)$ , if  $x$  is not free in  $A$

A2  $\forall xA(x) \rightarrow A(t)$

Finally, we have the identity axiom schemas:

A3  $x = x$

A4  $x = y \rightarrow (A(x) \rightarrow A(y))$ , where  $A$  is atomic.

Let us agree that  $A(x)$  is  $A(z/x)$  and  $A(y)$  is  $A(z/y)$ , so that  $A(y)$  results from  $A(x)$  by replacing  $x$  at some occurrences by  $y$ .

Let us call this system CL. Like most Hilbert-style systems, CL is sound with respect to  $\models_g$  but not  $\models$ : if  $X \vdash A$ , then  $X \models_g A$  but not necessarily  $X \models A$  (consider UG). The limitation is immaterial if  $X$  consists of sentences, as it will if  $L$  is used to represent arguments in some natural language.

## 2 MOTIVATIONS FOR FREE LOGIC

The motivations for free logic may be grouped under four headings. (1) Classical first-order semantics embodies existence assumptions that can produce weird results and constrain what can be done to avoid them. Accordingly, (2) certain philosophical doctrines can be expressed in first-order languages, classically conceived, only with difficulty and in ways that will seem artificial. Then there are general considerations of logical form: (3) if logical form captures truth-conditions in the sense of determining correct truth-values in all possible situations, then logical semantics must be universally free. Finally, for those who remain unconvinced, there is a pragmatic argument: (4) the representation of logical moves in a classical system can often be considerably simplified if we *pretend* that certain expressions are terms that need not refer to any existent, either because they do not refer at all or because they refer to *pretend* objects.

### 2.1 Classical Existence Assumptions

The existence assumptions built into classical first-order semantics are implicit in i1–i3:  $D$  is non-empty, constants refer to individuals of  $D$ , and function-names refer to total functions  $D \rightarrow D$ . These assumptions constrain the meaning of logical forms. As implicit premises, they permit some surprising inferences. While such unwanted conclusions can be avoided, the ways of doing so, constrained as they are by these assumptions, may seem artificial and too complex.<sup>6</sup> Let us consider the existence assumptions in turn.

---

<sup>6</sup>Various free logics can be represented as classical first-order theories. Let  $L_f$  be the first-order language of a free system FL in which free logical truth has been characterized in terms of Hilbert-style provability from logical axioms by logical rules of inference:  $\models A$  iff  $\vdash_{FL} A$ ; and let  $L_c$  result from  $L_f$  by adding a 1-place predicate  $E$ . Trew [1970] shows the dedicated reader how to (1) translate sentences  $A$  of  $L_f$  into sentences  $tr(A)$  of  $L_c$  and, for each of a variety of systems FL, how to (2) write axioms  $Ax(FL)$  that classically constrain the interpretation of  $E$ , so that  $\vdash_{FL} A$  iff  $Ax(FL) \vdash_{CL} tr(A)$ , *i.e.*, iff  $tr(A)$  is classically provable from  $Ax(FL)$ .

- (a) Existential conclusions may be validly drawn from premises that are not overtly existential, because the assumption of a non-empty universe operates as an implicit existential premise. For example, ‘A person is good iff she loves everyone, so some person loves all good persons’ is valid when assigned the simple form

$$\frac{\forall x(Gx \leftrightarrow \forall yLxy)}{\exists x\forall y(Gy \rightarrow Lxy)}$$

For if there is a good person  $x$ , then  $x$  loves everyone, hence all good persons. If there is no good person, let  $x$  be anybody:  $x$  loves all good persons (at least on the standard no-counterexample interpretation of ‘all’), so someone loves all good persons. The crucial step is the passage from ‘ $\forall y(Gy \rightarrow Lxy)$  is true of any  $x$ ’ to ‘ $\forall y(Gy \rightarrow Lxy)$  is true of some  $x$ ’, a move licensed (albeit *sotto voce*) by the assumption that the range of  $x$  is non-empty.<sup>7</sup>

These awkward results can be sidestepped by complicating logical form: take the variables to range, not over persons but over some wider class, and relativize the quantifiers to the subclass of persons by introducing an appropriate 1-place predicate. The resulting argument

$$\frac{\forall x(Px \rightarrow (Gx \leftrightarrow \forall y(Py \rightarrow Lxy)))}{\exists x(Px \& \forall y(Py \rightarrow (Gy \rightarrow Lxy)))}$$

is invalid, since the premise is true and the conclusion false when  $P$  is assigned the empty extension, which of course is permitted in classical semantics.

Constants and function-names complicate the transformation. To preserve the validity of ‘Pope John-Paul II is nobody’s spouse, so he isn’t his own spouse’, we will need to add a premise  $Pj$  to do the work of i2 and a premise  $\forall x(Px \rightarrow PSx)$  to do the work of i3:

$$\frac{\begin{array}{c} \neg\exists x(Px \& j = Sx) \\ \forall x(Px \rightarrow PSx) \\ Pj \end{array}}{j \neq Sj}$$

<sup>7</sup>A more mathematical example is the generation of an empty set, apparently *ex nihilo*, in some ZF-formulations of pure set theory, *e.g.* [Shoenfield, 1967]:  $\exists y\forall z(z \notin y)$  follows from the subset axiom in the form  $\forall x\exists y\forall z(z \in y \leftrightarrow (z \in x \& A(z)))$ . For let  $x$  be a set; by the subset axiom, there is a set  $y$  whose members are the sets  $z$  such that  $z \in x$  and  $z \neq z$ ; since no set  $z$  is such that  $z \neq z$ , there are no such sets and  $y$  has no members; so there is a set with no members. What is disturbing here is that the subset axiom has an essentially *conditional* form: if  $x$  exists, so does any describable subset of  $x$ . How can it yield a *categorical* existence claim? The answer is that classical semantics guarantees that the range of  $x$  is non-empty, so some set will exist.

If  $P$  is assigned the empty extension, then  $Pj$  will be false, but for a reason that will strike some people as incorrect:  $j$  refers to something not in the extension of  $P$ . If there were no people, ‘Pope John-Paul II is a person’ might be false, but not because ‘Pope John-Paul II’ refers to something (zero, say) which is not in the now empty extension of ‘person’; ‘Pope John-Paul II’ doesn’t refer at all in this situation. At best, we have a semantic proxy for failure of reference that delivers the right truth-values — at best because one might want to hold that ‘Pope John-Paul II is a person’ is neither true nor false when ‘Pope John Paul II’ does not refer.

None of these manoeuvres will alter the validity of such arguments as ‘Everything is self-identical, so something is’

$$\frac{\forall x(x = x)}{\exists x(x = x)}$$

since the range of  $x$  is non-empty. And it might be objected that the account of logical possibility given by classical semantics is defective, for surely the range of the variables *could* be empty.

- (b) i2 requires that constants refer to something in the range of the variables, which permits quick proofs of the existence of God (or Grendel, or anything you like), since  $\exists x(x = g)$  is logically true if  $g$  refers to something in the range of  $x$ . If you feel bad about doing so little work for such large results, you can give a short CL-proof:

- |   |          |
|---|----------|
| 1. $x = x$                                    | A3       |
| 2. $\forall x(x = x)$                         | UG(1)    |
| 3. $\forall x(x = x) \rightarrow g = g$       | A2       |
| 4. $\forall x(x \neq g) \rightarrow g \neq g$ | A2       |
| 5. $\exists x(x = g)$                         | T(2,3,4) |

The instances of A2 are logically true because  $g$  must refer to something in the range of  $x$ .

The standard Russellian fix for these problems is to replace constants  $g$  that may not refer by predicate constructions: find a *singular* predicate  $G$  true of at most one thing, which you’d be willing to label  $g$  if there were such a thing (in the present case, perhaps Anselm’s ‘nothing greater than  $x$  can be or be conceived’ will do) and replace atomic parts  $A(g)$  by  $\exists y(\forall z(Gz \leftrightarrow y = z) \& A(y))$ , where  $y$  is not free in  $A$ . Then the existence claim  $\exists x(x = g)$  becomes  $\exists x\exists y(\forall z(Gz \leftrightarrow y = z) \& x = y)$ , which is obviously not logically true.



Alternatively, add  $\forall x\forall y((Gx \& Gy) \rightarrow x = y)$  as a premise (an axiom) to constrain interpretations of  $G$ , and replace atomic parts  $A(g)$  by  $\exists y(Gy \& A(y))$ . This will transform the existence claim  $\exists x(x = g)$  into  $\exists x\exists y(Gy \& x = y)$ , and the issue of its logical truth into that of the validity of

$$\frac{\forall x\forall y((Gx \& Gy) \rightarrow x = y)}{\exists x\exists y(Gy \& x = y)}$$

which is obviously invalid. This procedure will make atomic sentences  $A(g)$  with non-referring terms  $g$  false, since the predicate  $G$  will be true of nothing. If you think that some subject-predicate sentences with non-referring subjects — such as ‘Grendel was slain’ — are true, you can use the replacement  $\forall y(Gy \rightarrow A(y))$  for them, as Mendelson [1989, p. 613] observes. But something seems to be missing here. The reason for the truth of ‘Grendel was slain’ and the falsity of ‘Grendel was pink’ is really the same: whatever singular predicate we find for Grendel is true of nothing. Moreover, if you think that some subject-predicate sentences with non-referring subjects, such as van Fraassen’s [1966, p. 82] ‘Pegasus has a white hind leg’, are neither true nor false, you will not be happy with Russell’s way of dealing with them, since classical semantics is bivalent: any sentence is true or false.

- (c) i3 requires that functions be total, which validates such arguments as ‘Every spouse loves his or her spouse (and, of course, the spouse of one’s spouse is oneself), so nobody is unloved’, if it is given the simple form

$$\frac{\forall x\forall y(x = Sy \rightarrow (y = Sx \& LxSx))}{\neg\exists x\neg\exists yLyx}$$

For any person  $x$  is such that  $x$ ’s spouse loves the spouse of  $x$ ’s spouse, who of course is  $x$ , and everyone has a spouse, since  $S$  is total. In real life, not everyone has a spouse, but the straightforward way of saying this,  $\neg\forall x\exists y(y = Sx)$ , is logically false.

In mathematical applications of logic, it would be convenient to introduce notations for partial functions, *e.g.*, to define predecessor  $P$  in terms of successor  $S$  in the natural numbers by  $\forall x\forall y(Px = y \leftrightarrow Sy = x)$  or division  $/$  in terms of multiplication  $\times$  in the reals by  $\forall x\forall y\forall z(x/y = z \leftrightarrow z \times y = x)$ . However, such definitions will not do: given  $\neg\exists x(Sx = 0)$ , the first entails  $\neg\exists x(P0 = x)$ , which is logically false; given  $\forall x(x \times 0 = 0)$ , the second entails  $S0 = 0$ , which contradicts  $\neg\exists x(Sx = 0)$ . Nor can we simply exclude the troublesome arguments.  $\forall x(x \neq 0 \rightarrow \forall y(Px = y \leftrightarrow Sy = x))$  leaves  $P$  undefined at 0, and

$\forall x \forall y (y \neq 0 \rightarrow \forall z (x/y = z \leftrightarrow z \times y = x))$  leaves  $x/y$  undefined when the value of  $y$  is 0, whereas classical semantics requires that functions be everywhere defined.

These difficulties may be avoided by replacing function names with relational predicates for their graphs.

$$\frac{\forall x \forall y (S y x \rightarrow (S x y \& L x S x))}{\neg \exists x \neg \exists y L y x}$$

is invalid, and definitions like  $\forall x \forall y (P x y \leftrightarrow S y = x)$  or  $\forall x \forall y \forall z (D x y z \leftrightarrow z \times y = x)$  are fine. For some applications, however, we will need to add a premise (an axiom) to the effect that the defined predicate is functional, *e.g.*,  $\forall x \forall y \forall z ((S x y \& S x z) \rightarrow y = z)$ . And we lose the considerable advantages of functional notation.

Alternatively, we can represent a partial function  $f$  by a total function  $F$  that coincides with  $f$  where  $f$  is defined and is given some arbitrary value elsewhere. If the arbitrary value for the predecessor and division functions is 0, then their definitions may be given as  $(\forall x (x \neq 0 \rightarrow \forall y (P x = y \leftrightarrow S y = x)) \& P 0 = 0)$  and  $\forall x (\forall y (y \neq 0 \rightarrow \forall z (x/y = z \leftrightarrow z \times y = x)) \& x/0 = 0)$ . In the case of the spouse function, we might pick an unmarried person (the Pope, say) and extend the spouse function to unmarried people  $x$  by stipulating that the spouse of  $x$  is the Pope. If  $S$  now represents this function and  $p$  names the Pope, then  $S x \neq p$  will tell us that  $x$  has a spouse (that the partial spouse function is defined at  $x$ ), and we can recast the premise of the argument as  $\forall x (S x \neq p \rightarrow \forall y (x = S y \rightarrow (y = S x \& L x S x)))$ . The price is a violation of ordinary usage: ‘The Pope’s spouse is the Pope’ is not true. Since the Pope is unmarried, ‘The Pope’s spouse’ doesn’t refer to the Pope or to anyone else; it doesn’t refer to anything. Where the partial function  $f$  is onto, as is the predecessor function, we cannot identify the arguments at which it is defined in this way. No matter what individual  $i$  is picked for the arbitrary value,  $F x \neq i$  will be false for some  $x$  at which  $f$  is defined.

## 2.2 Logical Habitats for Philosophical Doctrines

If free semantics permits us to avoid strange results, it also permits us to state strange doctrines. Free semantics provides a more neutral logical setting than classical semantics for certain philosophical views. It permits distinctions upon which they depend to be made in a straightforward way and does not prejudice the case against them by rendering important claims logically false. Let us consider an assortment.

- a. Meinong claimed, notoriously, that there are non-existent objects, that things like the golden mountain may have *being* while lacking *existence*. The straightforward form of ‘*a* does not exist’ is  $\neg\exists x(a = x)$ , but this is logically false in classical semantics, since *a* must refer to something in the range of *x*. A device like Russell’s will avoid this, but at the expense of a complex form  $\neg\exists x\exists y(\forall z(Az \leftrightarrow z = y) \& y = x)$ , where *Az* is uniquely true of *a*.

We could instead take the variables to range over beings and introduce a 1-place existence predicate *E*; quantification over existents would be represented by relativizing quantifiers to *E*,  $\forall x(Ex \rightarrow A(x))$  indicating that *A(x)* is true of all existents,  $\exists x(Ex \& A(x))$  that it is true of some existent. Then  $\exists x\neg Ex$  will represent ‘there are non-existent objects.’ We can even let *Ex* abbreviate  $\exists y(y = x)$ , provided we follow Lejewski and give identity a non-standard meaning: an interpretation is  $\langle D, d, o \rangle$ , where  $\langle D, d \rangle$  is classical and  $o \in D$ , and  $s = t$  is true under *I* and  $\alpha$  iff *s* and *t* refer under *I* and  $\alpha$  to the same individual of *D* and this individual is not *o*. Relative to interpretations and valuations of this sort,  $\exists x\neg\exists y(y = x)$  is *logically* true.

These classical or quasi-classical approaches do not really honor Quine’s dictum. Here to *be* is to be the value of a variable, all right, but to *exist* is not. By contrast, a free semantics that permits constants to refer to individuals *outside* the range of *x* allows us to say simply that *a* does not exist without immediately contradicting ourselves and without abandoning Quine’s useful connection between quantification and existence.

If we wish to make the more general Meinongian claim that there are non-existent objects — if we wish to quantify over them — then it may be argued that we are really committed to objects with being and should simply treat them classically, delimiting the subclass of existents with *E*. Alternatively, we could add a special quantifier  $\exists_b y$ , meaning ‘there is a being *y* such that’ and write  $\exists_b y\neg\exists x(y = x)$ : to be is to be the value of a variable bound by  $\exists_b$ , to exist is to be the value of a variable bound by  $\exists$ . For free semantics of this sort — and an argument that the semantics of *any* modal or tense logic can be built up from it — see [Cocchiarella, 1991]. More modestly, we could limit what does not exist to a single Lejewskian object, named by *o*, and then *define*  $\exists_b xA(x)$  as  $A(o) \vee \exists xA(x)$ . For this approach — and a proof that it is deductively equivalent to Lejewski’s — see [Lambert and Scharle, 1967].

- b. The truth of ‘Fred exists, but might not have’ is typically explained by gesturing toward a world that is possible relative to ours at which ‘Fred does not exist’, *i.e.*,  $\neg\exists x(x = f)$ , is true. Accordingly, at such a

world,  $f$  cannot refer to something in the range of  $x$ : Fred is not among the existents of that world. Standard Kripke-style modal semantics is free.

- c. Intuitionists and constructivists influenced by them reject non-constructive proofs in mathematics. A proof of  $\exists xA(x)$  is, for them, nothing more nor less than a proof of  $A(t)$  for some  $t$ , and such a proof is not secured merely by showing that a contradiction can be obtained from the assumption that  $\forall x\neg A(x)$ . Now if we have not established that  $f$  is defined at  $s$ , we can hardly claim to have a proof of  $fs = fs$  and therefore a proof of  $\exists x(x = fs)$ . Yet standard formulations of intuitionist logic follow classical logic in regarding  $t = t$  as a logical axiom and licensing the inference to  $\exists x(x = t)$ . Accordingly, we must either banish partial functions from intuitionist logic, thereby limiting its reach in mathematics and ignoring the views of patriarchs like Brouwer, or further modify the logic, this time in the direction of free logic. For free variations on Kripke-style semantics for intuitionistic logic, see [Posy, 1982].
- d. Evans [1979] has noted that the standard examples of contingent *a priori* truths presuppose a free semantics. If we stipulate that ‘Julius’ refers to whoever (uniquely) invented the zipper, then ‘if someone (uniquely) invented the zipper, Julius did’ appears to be (1) true, (2) *a priori*, but (3) contingent. It is true because if someone (uniquely) invented the zipper, that person is Julius because ‘Julius’ refers to whoever (uniquely) invented the zipper. It is *a priori* because we need only understand the stipulation to see that it is true. It is contingent because there is a possible world in which it is false, given that someone actually did (uniquely) invent the zipper: in the actual world Julius (uniquely) invented the zipper, but he might not have: at a possible world in which someone else (uniquely) invented the zipper, ‘Julius did’ is false. A free semantics is presupposed because in classical semantics we cannot introduce a constant like  $j$  with a defining axiom  $\forall x(x = j \leftrightarrow A(x))$  without first establishing that  $\exists!x A(x)$ . Otherwise, we could stipulate that  $\forall x(x = j \leftrightarrow x \neq x)$  and end up with the logically false  $j \neq j$ . In this case,  $A(x)$  is  $\forall y(Zy \leftrightarrow y = x)$ , representing ‘ $x$  (uniquely) invented the zipper’. Thus, the candidate sentence  $(\exists!xZx \rightarrow \forall y(Zy \leftrightarrow y = j))$  — if someone (uniquely) invented the zipper, Julius did — is not well-formed unless  $\exists!xZx$  is true, so it cannot be known *a priori* to be true. If, however, we allow non-referring names with the understanding that atomic constructions involving them are false,  $\forall x(x = j \leftrightarrow \forall y(Zy \leftrightarrow y = x))$  will be true whether or not  $\exists!xZx$  is true; and if  $\exists!xZx$  is true,  $j$  must refer and  $\forall y(Zy \leftrightarrow y = j)$  will also be true.

- e. Aristotle held that true predication requires a subject: the truth of ‘Socrates is well’ presupposes the existence of Socrates: ‘Socrates’ must refer. However, the falsity of ‘Socrates is well’ does not: ‘Socrates is well’ is false if Socrates exists but illness is attributable to him; it is also false if Socrates does not exist. ‘Socrates is ill’ and ‘Socrates is not well’ do not give the same information: the former attributes illness to Socrates, the latter merely denies that wellness is attributable to him: he may be ill, or he may not exist at all.

While classical semantics can handle contrary predications like ‘John is rich’ and ‘John is poor’ simply by not identifying the extension of ‘rich’ with the complement of the extension of ‘poor’, it does not allow ‘Socrates is well’ and ‘Socrates is ill’ to be false if ‘Socrates’ does not refer: all names refer. A free semantics in which non-referring subject-terms render subject-predicate sentences false embodies Aristotle’s view in a natural way. Scales [1969] has extended it by allowing complex predicates  $\lambda xA(x)$  to be formed from open sentences  $A(x) : \lambda xA(x)t$  is true iff  $t$  refers and  $A(t)$  is true. Since  $A(t)$  may be true when  $t$  does not refer,  $\lambda xA(x)t$  and  $A(t)$  may differ in truth-value. For example,  $\neg Wp$  representing ‘Pegasus is not winged’ is true, while  $(\lambda x\neg Wx)p$ , representing ‘Pegasus is wingless’ is false, if ‘Pegasus’ does not refer.

- f. Russell and Meinong go further than Aristotle, holding that ascribing to a sentence a subject-predicate *form* requires that the subject-term refer,

whether the sentence is true or false, a doctrine now embodied in classical semantics. As Lambert [1986, p. 276] notes, Meinong held that ‘the golden mountain is golden’ is a predication, so there must *be* a golden mountain; Russell couldn’t swallow the conclusion, and so denied the sentence a subject-predicate form.

If the Russell–Meinong view of predication is accepted, ‘exists’ cannot be a predicate. One argument, extracted from Mendelson [1989, p. 609], is this: (1) If ‘exists’ is a predicate, then singular existentials of the form ‘ $s$  exists’ are subject-predicate sentences with subject  $s$ . (2) In a subject-predicate sentence, “the subject stands for something and the predicate says something about that for which the subject stands.” So (3) if ‘exists’ is a predicate, ‘ $s$  exists’ is trivially true. But (4) some singular existentials (such as ‘Neptune exists’) are not trivially true, and others (such as ‘Vulcan exists’) are simply false. So (5) ‘exists’ is not a predicate. Since (2) is enshrined in classical semantics, it will be difficult to treat ‘exists’ as a predicate  $E$  in the classical setting — unless we abandon (1) or Quine’s dictum.<sup>8</sup> For the

<sup>8</sup>Mendelson, under the spell of (2), does abandon (1): he construes “atomic-looking”

extension of  $E$  will be the range of the bound variables and any term  $s$  must refer to something in this range, so  $Es$  will indeed be trivially true.

- g. An alternative to both Aristotle and Russell-Meinong holds that ‘The present king of France is bald’ is a *subject-predicate sentence* whose truth *or falsity* requires the existence of the present king of France, or the truth of ‘The present king of France exists.’ ‘The present king of France is bald’ then *presupposes* ‘The present king of France exists’ in Strawson’s [1952, p. 175] sense: a statement  $S$  presupposes a statement  $S'$  iff “the truth of  $S'$  is a precondition for the truth-or-falsity of  $S$ .” Presupposition is an interesting relation only if  $S$  can be neither true nor false, since otherwise  $S$  presupposes  $S'$  iff  $S'$  is necessarily true. So a formal treatment will require giving up bivalence. If classical semantics is assumed, then  $Bk$  does indeed presuppose  $\exists x(x = k)$ , for the latter is logically true. However, as just observed in Section 2.2f,  $\exists x(x = k)$  cannot be an adequate representation of the contingent truth ‘The present king of France exists’.

The natural semantic setting for presupposition is a free semantics in which (1) subject-predicate sentences with non-referring subjects are neither true nor false, (2) existence claims of the form  $\exists x(x = t)$  are false if  $t$  does not refer, (3)  $\neg A$  is neither true nor false iff  $A$  is neither true nor false, and (4)  $X \models A$  iff  $A$  is true whenever all the  $X$ -sentences are true. Presupposition may then be characterized as

$$A \text{ presupposes } B \text{ iff } A \models B \text{ and } \neg A \models B$$

and  $Bk$  presupposes  $\exists x(x = k)$ , because both  $Bk \models \exists x(x = k)$  and  $\neg Bk \models \exists x(x = k)$ , although  $\not\models \exists x(x = k)$ . Free semantics of this “neutral” kind are discussed in Section 3.6. For a different supervaluational treatment of presupposition, see [van Fraassen, 1968].

- h. Mereology conceives of individuals as wholes with parts. For the mereologist, ‘The rivers of Canada are numerous’ does not claim that a particular set — the set of Canadian rivers — has many members, but that a particular whole — the mereological sum of the Canadian rivers — has many parts. Wholes whose parts are spatio-temporal individuals are also spatio-temporal, though perhaps spatio-temporally discontinuous; by contrast, sets of spatio-temporal individuals are abstract objects. Since every whole is a part of itself, there cannot be

---

sentences  $Ps$  as either  $\exists x(Sx \& Px)$  or  $\forall x(Sx \rightarrow Px)$ , where  $S$  is a singular predicate. However, his non-standard semantics generates the truth-values that  $Ps$  would have if it were a subject-predicate sentence whose subject may refer to individuals outside the range of the bound variables.

a whole with no parts (though there may be wholes — true atoms — with no proper parts). While set theory can embrace the empty set, mereology does not recognize an empty whole.

In thinking rigorously and systematically about parts and wholes, it is very convenient to employ operators such as *binary sum* ( $x + y$  is the whole whose parts  $z$  are such that  $z$  is part of  $x$  or  $z$  is part of  $y$ ), *generalized sum* ( $\sigma x A(x)$  is the whole whose parts  $z$  are such that there is a  $y$  such that  $A(y)$  of which  $z$  is part), *binary product* ( $x \times y$  is the whole whose parts  $z$  are such that  $z$  is part of  $x$  and  $z$  is part of  $y$ ), etc. But such operators will not be everywhere defined, if there is no empty whole: both  $\sigma x(x \neq x)$  and  $x \times y$ , where  $x$  and  $y$  have no common part, would be the empty whole, if there were such a thing, but there is not, so both  $\sigma x(x \neq x)$  and  $x \times y$ , where  $x$  and  $y$  have no common part, are undefined. If mereological theories are to be framed as first-order theories, classical semantics is an unwelcome constraint: the mereologist must either do without these operators and conduct all logical business in terms of cumbersome predicates, or he must hold his nose and introduce a constant for something he denies exists, *viz.* the empty whole. A free semantics that permits partial operators is much more congenial. For a discussion of free mereological theories, see [Simons, 1991].

### 2.3 Logical Form

I take the present view of logical form to have these elements:

- a. Sentences-in-context have a semantic structure or logical form: their truth-values (truth, falsity, or lack thereof) reduce, *via* recursive semantic rules, to the semantic values of their unstructured parts or elements (*e.g.*, a subject-predicate sentence is true iff the referent of the subject term belongs to the extension of the predicate).
- b. An interpretation assigns appropriate semantic values to such elements (*e.g.*, extensions to predicates).
- c. Any logically possible situation is represented by some interpretation; in particular, the actual situation is represented by an interpretation, so that the actual truth or falsity of a sentence reduces to the actual semantic values of its elements.
- d. The logical properties and relations of sentences-in-context are determined by their semantic structures and the totality of interpretations *via* semantic definitions of these properties and relations (*e.g.*, logical truth is truth under every interpretation); such definitions provide the

basis for judging the soundness and adequacy of proof systems that give purely syntactic accounts of proof and proof from hypotheses.

Granting this, it is easy to argue for *inclusive* semantics: bound variables may range over the empty domain, for surely it is logically possible that there be nothing at all. To be sure, once we have investigated this case, we might want to heed Quine's advice [1953, p. 161] and put it aside for pragmatic reasons if, by including it, we would "cut ourselves off from laws applicable to all other cases". Or we might include it by using those laws but performing an extra check on results: existentials will be false and universals true in the empty domain, and the other sentences will be truth-functional compounds of these if we have followed Quine and purged the language of all singular terms but variables. Still, what justifies the truth-value assignments is a look at this case and thinking through the application of general semantic rules to it. Note that the no-counterexample interpretation is required if universals are to be true and that vacuous quantifiers are not idle in the empty domain ( $\exists y\forall xPx$  is false while  $\forall xPx$  is true).<sup>9</sup>

We may also argue for *free* semantics as follows: (1) In accord with Quine's dictum, the range of bound variables in a given interpretation is restricted to individuals that exist in the possible situation represented by the interpretation. (2) There are sentences in which expressions that look for all the world like names do not, in actual use, refer to actual individuals. So (3) if these expressions are treated as names, no interpretation that represents the actual situation can assign them referents in the range of the bound variables. (4) What is permitted in interpretations that represent the actual situation must also be permitted in interpretations that represent possible situations. So (5) if these expressions are treated as names, interpretations in general need not assign them referents in the range of the bound variables.

(1) is true of classical semantics, whether we identify an interpretation with (i) a meaning function that associates with each element an appropriate semantic value in the *actual* world, or (ii) a possible world, at which the meanings of elements in the actual world determine their semantic values, or (iii) a meaning function at a possible world.<sup>10</sup> Sentences of type (2) include:

---

<sup>9</sup>For discussion of systems that treat vacuous quantifiers in the empty domain differently, see [Lin, 1983].

<sup>10</sup>The main defect of (i), aside from having to represent reasoning about hypothetical situations indirectly, is that the valid arguments depend upon how many individuals there are in the actual world. If the number is finite —  $k$ , say — any argument with a premise stating that there are more than  $k$  individuals will be valid, since the premises are not satisfiable. The main defect of (ii), aside from issues of epistemic access to possible worlds, is that arguments like 'Some even number has an irrational square root, 2 is an even number, so 2 has an irrational square root' will be valid, since the conclusion is presumably true at any possible world. Hence, (iii).



- s1. If *Vulcan* exists and its orbit lies within that of Mercury, it's going to be very hard to observe.
- s2. *Vulcan* doesn't exist.
- s3. *Zeus* is not Allah.
- s4. The ancient Greeks worshipped *Zeus*.
- s5. *Pegasus* is a winged horse.
- s6. Bosch painted a picture of *hell*.
- s7. Mary admires *Eustacia Vye*.

The simple thought behind (4) is that the actual is a special case of the possible. A name like 'Hillary Clinton' happens to refer to a particular individual of the actual world; but the name could have been attached to some other individual of this world or to an individual of some other possible world. If we are prepared to regard an expression like 'Eustacia Vye' as a name that does not in fact refer to an individual of this world (but to the heroine of Thomas Hardy's novel, *The Return of the Native*), then we should allow that in a possible situation, however conceived, it need not refer to an individual that exists in that situation.

(5), the conclusion of the argument, is conditional, and we may avoid adopting free semantics by refusing to admit that such expressions are, despite appearances, singular terms. The cost of such denial is dealing with them in some other way, and those we have seen above are awkward. What basis is there for holding that 'Mary admires Hillary Clinton' is a relational subject-predicate construction, while 'Mary admires Eustacia Vye' is not? The obvious difference is that 'Hillary Clinton' refers to a real person and 'Eustacia Vye' does not. But why should logical form depend upon *that*?

Lambert [1998, p. 157] and Kroon [1991, p. 21] suggest that logical form is independent of empirical fact, and hence independent of whether terms actually refer. The premise, however, seems too strong. We may agree that the logical form of a sentence cannot depend upon its truth or falsity, since its truth or falsity is determined by its logical form and the actual semantic values its elements, including the referents of its terms. But it does not follow that form cannot depend on whether expressions that we are tempted to classify as terms refer, particularly if we cannot figure out how to get semantic rules to deliver truth-values smoothly in such cases.

Indeed, it may be argued that logical form does depend to an extent on empirical fact, since it is sentences-in-use that have such forms. When Alice says, 'I'm hungry', and Bill adds, 'But I'm not', there is no contradiction: the form of what Alice says is represented by *Ha* and the form of what Bill says by  $\neg Hb$ . If logical forms reflect the different uses of indexicals

like ‘I’, as they must in one way or another, then why may they not reflect such empirical facts as whether “terms” refer? We might want to say that the logical form of ‘I’m hungry’ is the same in both cases, namely, the simple subject-predicate form  $Ps$ , and that *Ha* represents, in this form, the *propositional content* of what Alice says. But then it is representations of propositional content that will do the work that (a)–(d) require of logical form.

I prefer a more pragmatic argument: logical form should be reasonably accessible, and the closer it is to surface form the better, other things equal. We do not want to misconstrue logical form and have to reformulate reasoning when we discover our error. If logical forms are contingent upon whether certain expressions refer — a matter that may be very difficult to settle — then logic may not be very useful. We don’t want to have to revise reasoning about the unknown solution to some equation if we find out, as a result of that very reasoning (how else?), that there is no solution (or no unique solution). Moreover, we’d like to be able to conduct such reasoning using a term  $t$  for the solution, rather than (say) in Russell’s indirect and clumsy fashion.<sup>11</sup>

But perhaps we need not abandon the classical perspective, even if we admit that the italicized terms in s1-s7 are singular terms. Let me sketch two objections of this kind:

**OBJECTION 1.** Suppose we insist that sentences with terms that do not refer, or that seem to refer to things outside the range of the variables, are neither true nor false. Then applying classical laws will lead from truths to truths: we may apply existential generalization to ‘Mary admires Eustacia Vye’ to get ‘Mary admires someone’, but the premise is untrue, so the falsity of the conclusion (if it is false) does not invalidate the rule. The reason for including the actual situation (in which such terms as ‘Eustacia Vye’ do not refer to anyone) among the possible situations is to mesh logic with truth: the conclusion of a sound argument should be true. But if we insist that sentences with terms like ‘Eustacia Vye’ are truth-valueless, we need not worry that classical logic will lead us from true premises to untrue conclusions. Accordingly, such sentences can be set aside as ‘don’t cares’.<sup>12</sup>

There are three problems with this proposal. First, it is difficult to maintain that all such sentences are truth-valueless; indeed, anyone not bewitched by some theory will take most of s1–s7 to be *true*. Second, familiar rules of inference such as *addition*

---

<sup>11</sup>Note, however, that a free semantics that does not support extensionality in the sense that  $\forall x(A(x) \leftrightarrow B(x)) \models A(t) \leftrightarrow B(t)$  is probably not going to support this reasoning either, since it will typically involve moving from  $A(t)$  and  $\forall x(A(x) \leftrightarrow B(x))$  to  $B(t)$ . We shall probably have to conduct it under the additional assumption that  $t$  exists:  $\exists x(x = t)$ .

<sup>12</sup>This objection is suggested by van Fraassen [1966, Section 3].

ADD

 $A \vdash A \vee B$ 

will then not preserve truth, contrary to what is alleged: ‘Washington is the capitol of the USA so either Washington or Atlantis is the capitol of the USA’ has a true premise and a truth-valueless conclusion. Third, no justification has yet been given for the claim that terms which do not refer to things in the range of the bound variables render sentences truth-valueless; such a justification requires an extension of the ordinary semantic rules to this case.

OBJECTION 2. Let us concede that sentences with terms that do not refer to existents can be true or false; still, it does not follow that we must reject classical semantics for free semantics:

- a. Perhaps, as Stenlund [1973], Burge [1974], and Kroon [1991] suggest, some sentences of this kind can be handled classically, albeit by shifting from possible to imaginary or hypothetical worlds. In actual use, a sentence like ‘Pegasus is a winged horse’ invokes an implicit ‘in myth’ operator that, in effect, shifts attention from the actual situation to an imaginary one in which Pegasus exists (and indeed turns out to be a winged horse). That is, in actual use or context, the sentence is not about this world, but about another one, at which the normal reference conditions of classical semantics are fulfilled. Thus, classical semantics is all we need to understand why the reasoning of Sherlock Holmes in “Silver Blaze” about the missing horse is sound in the world of Doyle’s story: “. . . he must have gone to King’s Pyland or to Mapleton; he is not at King’s Pyland. Therefore he is at Mapleton.”<sup>13</sup>

However, this manoeuvre works only for sentences like s5. Consider s3: Zeus is not Allah, but *where?* Insofar as we can understand the truth of most of the sentences s1-s7 in terms of reference to imaginary individuals, it is reference *across*, not *within*, worlds, and that is not going to be accommodated by classical interpretations. All we need do to create problems for the recommended treatment of ‘Pegasus is a winged horse’ is to add ‘though such things do not exist’. If ‘Pegasus is a winged horse’ is true, so is ‘Pegasus is a winged horse, though such things do not exist’, but moving to a world of myth will render it self-contradictory. We can fix this by staying here in the actual world and understanding ‘Pegasus’ to refer to something in a world of myth, but that abandons classical semantics for some free variant of it.

- b. Perhaps what is problematic about sentences like s4, s6, and s7 is not failure of reference, but referential opacity, which we are not going to

---

<sup>13</sup>A. C. Doyle, “Silver Blaze”, in *The Complete Sherlock Holmes* (Garden City: Garden City Books, 1930), pp. 383–401 at p. 393.

be able to handle with extensional logical forms anyway. Contexts like ‘worshipped ...’, ‘picture of ...’ and ‘admires ...’, according to this view, are not referentially transparent: truth-value need not be preserved by replacing a referential term by a co-referring term. The suggestion is that if John worships God, it does not follow that John worships Satan, even if God and Satan turn out to be identical (or as identical as the members of the Trinity). And similarly for the others.

This objection does not help in the other cases. Nor does it seem very convincing. There is certainly a sense in which John does worship Satan if he worships God and God is Satan. And similar things may be said about the other contexts. A picture of Fred is a picture of the Grand Dragon of the Ku Klux Klan, if that’s who Fred is. ‘Admires’ is no more intensional than ‘loves’, the standard logic-text example of a relational predicate.

#### 2.4 *Derived Rules and Axioms*

Reflection on the cases discussed in Section 2.1 will suggest that we can work around the constraints of classical semantics in various ways without giving up its familiar simplicity. We can relativize the quantifiers to a predicate  $E$  that we interpret as true of existents; irreferential terms are those that refer to individuals of  $D$  that are not in  $d(E)$ , function-names  $f$  such that  $d(f)(\alpha) \notin d(E)$  for some  $\alpha \in d(E)$  represent functions  $d(E) \rightarrow d(E)$  that are not total. Or we can eliminate non-referring terms in Russell’s way. While such representations may not be perfect, they may be good enough for most purposes.

However, classical representations of irreferential names and partial functions can be cumbersome, and at some point those who work with them will want to develop some derived rules to facilitate reasoning and its formal or informal representation. Such rules will be those of a corresponding free logic. Development of various free logics can therefore provide a number of ‘off the shelf’ systems that can be applied in such cases. Let us consider two examples.

- a. Suppose that you believe Russell was correct in holding that descriptions like ‘the present King of France’ are not genuine singular terms and that what appear to be subject-predicate constructions like ‘the present King of France is bald’ are actually complex quantifier constructions. Still, working with these complex constructions is about as inviting as programming in machine language; you will want to develop some macros. So let’s *pretend* that ‘the present King of France’ is a term and ‘the present King of France is bald’ is a subject-predicate sentence whose truth-value is given by Russell’s quantifier construction, and attempt to develop a system of derived axioms and rules of

inference that enables you to treat descriptions *as if* they were terms. The obvious minimal constraint on such a system is that  $A$  should be provable iff the Russellian expansion of  $A$  is logically true. If we can pretend that descriptions are terms, we can also *pretend* that they can refer — or fail to refer, as the case may be — and ask what sort of logical semantics we may *pretend* underlies the system of derived axioms and rules. As Scales [1969] and Burge [1974] show, the result of this process is a free logic, whose underlying semantics may be sufficiently compelling to loosen your allegiance to Russell and to classical logic.

- b. Quine’s pure set theory NF [1969] is framed in a first-order language without identity, whose non-logical symbols are variables and the 2-place predicate  $\in$ . Identity (in the weak sense of indiscernibility) is introduced by definition:  $x = y$  abbreviates  $\forall z((z \in x \leftrightarrow z \in y) \& (x \in z \leftrightarrow y \in z))$ .<sup>14</sup> The only axioms are *extensionality*

$$\forall z(z \in x \leftrightarrow z \in y) \rightarrow x = y$$

and *restricted comprehension*

$$\exists x \forall y (y \in x \leftrightarrow A(y)),$$

where  $A$  is *stratified*: it is possible to replace the variables by numerals so that subject-predicate constructions  $x \in y$  become  $n \in n + 1$ . The restriction is designed to secure the safety of type theory without the pain. Quine’s view [1969, p. 16] is that “much... of what is commonly said of classes with the help of ‘ $\in$ ’ can be accounted for as a mere manner of speaking, involving no real reference to classes or any irreducible use of ‘ $\in$ ’.” Singular terms  $\{x : A(x)\}$  for classes — terms Quine calls *class abstracts* — are introduced by a pair of contextual definitions:  $y \in \{x : A(x)\}$  abbreviates  $A(y)$ , and  $\{x : A(x)\} \in \alpha$ , where  $\alpha$  is a class abstract, abbreviates  $\exists y(y = \{x : A(x)\} \& y \in \alpha)$ . If class abstracts are regarded as referring to *pretend* or *virtual* classes, then virtual classes which belong to virtual classes are real, though the virtual classes to which they belong need not be; sets are members of real classes. Identity may be extended to class abstracts by taking  $\alpha = \beta$  to abbreviate  $\forall x(x \in \alpha \leftrightarrow x \in \beta)$ , so that  $\alpha = \{x : A(x)\} \leftrightarrow \forall x(x \in \beta \leftrightarrow A(x))$  and  $\alpha = \{x : x \in \alpha\}$ .

Existential generalization fails for class abstracts:  $A(\alpha) \not\models \exists x A(x)$  for some  $A(\alpha)$ . In particular, let  $A(x)$  be  $x = \alpha$ , where  $\alpha$  is  $\{x : B(x)\}$ . Then the premise  $A(\alpha)$  is  $\alpha = \alpha$ , *i.e.*,  $\{x : B(x)\} = \{x : B(x)\}$ , *i.e.*,

---

<sup>14</sup>Of course, this will not force ‘=’ to be interpreted as numerical identity; no set of axioms constraining the interpretation of ‘=’ can do that, which is why identity is commonly regarded not as a predicate but as a logical operator.

$\forall y(y \in \{x : B(x)\} \leftrightarrow y \in \{x : B(x)\})$ , i.e.,  $\forall y(B(y) \leftrightarrow B(y))$ , which is logically true. The conclusion  $\exists xA(x)$  is  $\exists x(x = \alpha)$ , i.e.,  $\exists x\forall y(y \in x \leftrightarrow y \in \{x : B(x)\})$ , i.e.,  $\exists x\forall y(y \in x \leftrightarrow B(y))$ , which amounts to unrestricted comprehension, since  $B(y)$  can be any formula and the premise will be true. If  $B(y)$  is  $y \notin y$ , we obtain the logical falsehood  $\exists x\forall y(y \in x \leftrightarrow y \notin y)$  of Russell's paradox. However, the inference will be valid if  $\exists x(x = \alpha)$  is added as an extra premise:  $A(\alpha)$ ,  $\exists x(x = \alpha) \models \exists xA(x)$ . For then we can move from  $A(\alpha)$  to  $A(x)$  (in virtue of  $\models \alpha = \beta \rightarrow (A(\alpha) \leftrightarrow A(\beta))$  for variables or set abstracts  $\alpha, \beta$ ), from which of course  $\exists xA(x)$  follows. This is not a problem here because the added premise is just the conclusion when  $B(y)$  is  $y \notin y$ .

This is typical of free logics:  $A(t)$ ,  $\exists x(x = t) \models \exists xA(x)$ , but not necessarily  $A(t) \models \exists xA(x)$ . A natural question now is whether we can capture the logical moves involving class abstracts  $\alpha$  in a set of derived axioms and rules which treat them as genuine terms  $t$  and, if so, whether there is a natural free semantics that we might pretend underlies their use. A sufficiently natural semantics might blur the distinction between pretense and reality, especially in pure mathematics, where there does not seem to be much difference between pretending that mathematical objects exist and asserting that they do. For a restructuring of Quine's NF along these lines, see [Scott, 1967].

### 3 FREE SEMANTICS

Free departures from classical semantics are usually — though not always, as in [Farmer, 1995], [Feferman, 1995], and [Woodruff, 1984] — universally free. If we are going to permit terms that do not refer to individuals in the range of the bound variables, why not include the case where no term can refer to such an individual, simply because there are none? This is easy enough to do semantically, though the required adjustments to proof systems are a bit more trouble.

There are two ways in which terms may fail to refer to individuals in the range of the bound variables: either they refer to individuals outside this range, or they do not refer at all.

The first way leads to *outer domain* semantics, a straightforward bivalent modification of classical semantics in which a classical domain  $D$  is divided into a possibly empty inner domain  $D_i$ , over which the bound variables range, and an outer domain  $D_o$ . With the exception of the quantifier valuation clause  $v5$ , in which  $x$ -variants must now be understood to assign to  $x$  an individual of the inner domain, the interpretation and valuation clauses of classical semantics can be adopted without change.

The second way involves *partial* interpretations  $I = \langle D, d \rangle$ , where  $D$  may be empty and the denotation function  $d$  is partial on constants and assigns

partial functions  $D \rightarrow D$  to function-names. Assignment functions  $\alpha$  are also typically partial on variables. Here we face the problem of “assigning reasonable truth conditions to sentences containing non-denoting singular terms,” as Bencivenga [1986, p. 382] observes. This requires giving reasoned responses to the following questions:

- q1. If  $t$  does not refer, must  $ft$  also be non-referring? Does ‘Jack Aubrey’s sovereign’ refer to England’s George III, although ‘Jack Aubrey’ does not refer to a real person but to the fictitious captain of Patrick O’Brian’s sea novels?
- q2. Should  $d$  treat predicates as names of partial truth-valued functions on  $D$ ? Is ‘2 is green’ false (because ‘2’ refers to something that is not in the extension of ‘green’) or is it truth-valueless (because ‘2’ does not refer to something that is colored)?
- q3. If  $t$  does not refer, may  $Pt$  be true? If not, is it false or is it truth-valueless?
- q4. If  $t$  does not refer, is  $t = t$  true, false, or truth-valueless? If  $s$  refers and  $t$  does not, is  $s = t$  false or truth-valueless?
- q5. If formulae may lack truth-value, how are the classical truth-tables for the connectives to be extended? Shall we count  $A \rightarrow B$  true or truth-valueless if  $A$  is false and  $B$  is truth-valueless, or  $B$  is true and  $A$  is truth-valueless?
- q6. If formulae may lack truth-value, how are the quantifier clauses to be modified? Should  $\forall xA$  read ‘ $\forall xA$  is false if  $A$  is false for some  $x$ -variant of  $\alpha$ , and  $\forall xA$  is true otherwise’ or ‘ $\forall xA$  is false if  $A$  is false for some  $x$ -variant of  $\alpha$ , and  $\forall xA$  is true if  $A$  is true for each  $x$ -variant of  $\alpha$ ’ or ‘ $\forall xA$  is true if  $A$  is true for each  $x$ -variant of  $\alpha$ , and false otherwise’?
- q7. If formulae may lack truth-value, how are the definitions of logical properties and relations to be modified? Should logical consequence preserve truth? non-falsehood? both?

Applications may decide some of these questions. For example:

1. If we wish to allow for the non-strict functions and relations of computer science, then we will answer ‘No’ to q1 and ‘Yes’ to the first part of q3. Following Gumb and Lambert [1997], we might then implement such permissions by adding a virtual entity  $u$  (for ‘undefined’) to  $D$ :  $d$  will assign to a  $k$ -place function-name  $f$  a total  $k$ -ary function  $d(f) : D \cup \{u\} \rightarrow D \cup \{u\}$  and to a  $k$ -place predicate  $P$  a  $k$ -ary relation  $d(P)$  in  $D \cup \{u\}$  as its extension. If we want  $ft$  to be undefined, though  $t$  refers to  $\alpha$ , we will set  $d(f)(\alpha) = u$ ; if we want  $ft$

to be defined, though  $t$  does not refer, we will require  $d(f)(u) \in D$ ; if we want  $Pt$  to be true when  $t$  does not refer, we will put  $u$  in the extension  $d(P)$  of  $P$ .  $u$  is not to be regarded as a strange entity, but as a notational device for simplifying the statement of semantic rules, as in [Kroon, 1991].

2. If we think that ‘2 is green’ presupposes ‘2 is colored’ in the sense of Section 2.2g, we will answer ‘Yes’ to q2. We can then follow Smiley [1960] and Ebbinghaus [1969] and have  $I$  assign  $P$  both a domain of application  $D(P)$  in  $D$  and, within that domain, an extension  $d(P)$ . If we want  $Px$  to be truth-valueless when  $x$  is assigned  $\alpha$ , then we put  $\alpha$  outside  $D(P)$ ; if we want  $Px$  to be false when  $x$  is assigned  $\alpha$ , then we put  $\alpha$  in  $D(P) - d(P)$ .

The large decision is whether to answer q3 and q4 in a way that permits truth-valueless atomic formulae — and forces us to answer q5-q7. We can prune the choice tree considerably by opting for bivalence at the atomic level. However, such a decision needs to be rationalized. It will not do simply to argue that atomic formulae with non-referring terms should be false because (a) any atomic formula  $A(t)$  is a predication which is true just in case the  $A(x)$  is true of the referent of  $t$ , so that (b) where  $t$  fails to refer,  $A(t)$  is not true, so (c) where  $t$  fails to refer,  $A(t)$  must be false. For what underwrites the move from (b) to (c) is bivalence. However, there may be applications which call for such a ruling. For example, Farmer [1995, p. 281] claims that in the “traditional approach to partial functions” in mathematics, variables and constants *always* refer, functions may be partial and  $ft$  does not refer if  $t$  does not refer or  $d(f)$  is not defined at  $d(t)$ , while  $Pt$  is *false* if  $t$  does not refer.

The usual route to *true* atomic formulae with non-referring terms is *story semantics*, a non-referential variant of outer domain semantics: treat non-referring terms *as if* they referred to individuals in an outer domain, taking the formulae that are true under such a pretense to constitute a *story*  $S$  which supplements a partial referential interpretation  $I$  and assignment  $\alpha$ . Story semantics is equivalent to outer domain semantics in which the individuals of the outer domain  $D_o$  are treated as virtual or pretend objects. Story (or virtual outer domain) semantics seems to provide a natural way of dealing with sentences that are about fictional or mythical entities, at least if we do not want to follow Meinong in reifying them. Note, however, that we don’t yet have a justification for the bivalence that is built into this type of semantics, because *actual* stories or myths, unlike the stories of story semantics, are not complete. Nothing Doyle wrote decides the fictional truth-value of ‘Sherlock Holmes died before 1920,’ yet if  $Bh$  represents this sentence and  $h$  does not refer, any story  $S$  will include either  $Bh$  or  $\neg Bh$ .

If atomic formulae with non-referring terms are all to be false, then an excursion into stories is unnecessary: we can simply modify the valuation



rules v1 and v2 for atomic formulae to require as much. Note, however, that  $t = t$  will then be false, if  $t$  does not refer. Even stranger results await those who follow Frege [1892] and deny truth-values to atomic formulae with non-referring terms. For if  $Pt$  lacks truth-value, then  $\neg Pt$  and  $Pt \vee \neg Pt$  also appear to lack truth-value. Accordingly, some instances of  $A \vee \neg A$  are not logically true, if logical truth is truth under each partial interpretation and assignment. This is sufficiently disturbing to motivate a search for an respectable alternative that permits  $Pt$  to be truth-valueless while insuring that  $t = t$  and  $A \vee \neg A$  are logically true. The usual proposal is a *supervaluational* semantics, in which truth under a partial interpretation is understood as truth under all completions of it.

A free semantics in which *some* atomic formulae with terms that do not refer to individuals in the range of the bound variables are *true* is said to be *positive*.<sup>15</sup> If *all* such atomic formulae are *false* (*truth-valueless*), the semantics is said to be *negative* (*neutral*). Both outer domain and story semantics are positive in this sense, as is supervaluational semantics. However, the more significant divide is between bivalent and non-bivalent accounts. I shall first discuss bivalent free semantics, both positive and negative, then non-bivalent free semantics, including supervaluations.

### 3.1 Positive Bivalent Semantics: Outer Domains

Outer domain free semantics involves minimal change in classical semantics. An outer domain interpretation  $I = \langle D, d \rangle$  is classical, except that  $D$  is partitioned into an inner domain  $D_i$  and an outer domain  $D_o$ . Thus, it is altered to:

i<sub>o</sub>1.  $D$  is non-empty, and  $D = D_i + D_o$

No change is needed in the classical notion of an assignment. However, bound variables are to range over  $D_i$ , so the classical notion of an  $x$ -variant must be modified to require that  $x$  is assigned a value in  $D_i$ : an  $x$ -variant of  $\alpha$  is an assignment that differs from  $\alpha$  at most at  $x$  and assigns  $x$  a value in  $D_i$ . Note that  $\alpha$  now need not be an  $x$ -variant of  $\alpha$ . If  $D_i$  is empty,  $\alpha$  has no  $x$ -variants, so in this case universals are true and existentials are false. Evidently:

$$\begin{array}{l} \not\models \exists x(x = x) \\ Pt \not\models \exists xPx; \text{ but } A(t), \exists x(x = t) \models \exists xA(x) \\ \forall xPx \not\models Pt; \text{ but } \forall xA(x), \exists x(x = t) \models A(t) \\ Pt \not\models \exists x(x = t) \\ \models t = t \end{array}$$

---

<sup>15</sup>Here I follow the recent usage of Lambert [1997, p. 62]. Other meanings of 'positive' can be found in the free logic literature. Bencivenga [1986, p. 397] terms (conventional) semantics *positive* if *each* atomic formula containing a non-referring term is true, while Lambert [1991b, p. 344] uses 'positive' merely as a synonym for 'non-negative'.

If desired, an existence predicate  $E!$  may be introduced by taking  $E!t$  to abbreviate  $\exists x(x = t)$ . The extension  $d(E!)$  of  $E!$  under  $I$  will be  $D_i$ .

A Hilbert-style system PFL (for ‘positive free logic’) of axioms and rules<sup>16</sup> that is strongly complete relative to outer domain free semantics may be obtained from CL by replacing A1 with

$$\begin{aligned} & \forall x(A \rightarrow B) \rightarrow (\forall xA \rightarrow \forall xB) \\ & A \rightarrow \forall xA, \text{ if } x \text{ is not free in } A \end{aligned}$$

modifying A2 to

$$(\forall xA(x) \& \exists x(x = t)) \rightarrow A(t)$$

modifying A3 to

$$t = t$$

modifying A4 to

$$s = t \rightarrow (A(s) \rightarrow A(t)), \text{ if } A \text{ is atomic}$$

and adding

$$\forall x\exists y(y = x).$$

Lambert [1991a, p. 9] characterizes outer domain semantics as embodying a “Meinongian world picture”: the inner domain consists of existents, while beings that lack existence are relegated to the outer domain. This identification is a bit misleading, since Meinong held that non-existent beings are indeterminate with respect to certain properties — the golden mountain has no specific height — whereas the objects of an outer domain are determinate in virtue of  $i4$  and  $v1$ . A true Meinongian outer domain semantics would not be bivalent. Moreover, outer domains are sometimes taken to consist of pretend or virtual objects: that is, objects that we pretend exist so as to provide a referential semantics for terms that do not refer to existents. But Meinong did not regard having being as a matter of pretense.

### 3.2 Positive Bivalent Semantics: Stories

Story semantics can be regarded as a non-referential version of outer domain semantics. A story interpretation  $\langle I, S \rangle$  consists of a *partial* interpretation  $I$  that permits non-referring terms and a *story* (or *convention*)  $S$  that assigns truth-values to atomic formulae containing such terms.

<sup>16</sup>This formulation is based on [Meyer and Lambert, 1968]; see also [Lambert, 1997, p. 39]. Leblanc [1968] has shown how to derive  $(\forall xA \& E!t) \rightarrow A(t)$  from the other axioms by the rules of inference. For discussion of related systems, see [Bencivenga, 1986, Sections 5 and 6].

A partial interpretation  $I = \langle D, d \rangle$  is classical except that  $D$  may be empty,  $d$  is partial on constants, and  $d$  assigns function-names partial functions  $D \rightarrow D$ . In particular:

- $i_p1.$   $D$  is a possibly empty set.
  - $i_p2.$  If  $d$  is defined at  $a$ ,  $d(a) \in D$ .
  - $i_p3.$  If  $f$  is  $k$ -place,  $d(f)$  is a partial  $k$ -ary function  $D \rightarrow D$ .
- $i4$  remains unchanged. Assignments  $\alpha$  may be partial, but as in  $i_p2$  if  $\alpha$  is defined at  $x$ ,  $\alpha(x) \in D$ . The rules for reference under  $I$  and  $\alpha$  must be reformulated to take account of non-referring terms:
- $r_p1.$  If  $\alpha$  is defined at  $x$ , then  $x$  refers to  $\alpha(x)$ ; otherwise,  $x$  does not refer.
  - $r_p2.$  If  $d$  is defined at  $a$ , then  $a$  refers to  $d(a)$ ; otherwise,  $a$  does not refer.
  - $r_p3.$  If each  $t_i$  refers and  $d(f)$  is defined at  $\langle \alpha_1, \dots, \alpha_k \rangle$ , where  $t_i$  refers to  $\alpha_i$ , then  $ft_1 \dots t_k$  refers to  $d(f)(\alpha_1, \dots, \alpha_k)$ ; otherwise,  $ft_1 \dots t_k$  does not refer.

However, irreferential terms do not lead to truth-valueless formulae, since the story  $S$  supplies the missing truth-values for atomic formulae. A substitutional account of quantification would permit us simply to identify  $S$  with a set of atomic sentences satisfying certain conditions.<sup>17</sup> Objectual quantification requires a somewhat more complicated account, derived from Woodruff [1984]. Here a story is a *function*  $S$  from assignments  $\alpha$  to sets  $S(\alpha)$  of atomic formulae with non-referring terms satisfying the following conditions:

- s1. If  $t$  does not refer, then  $t = t \in S(\alpha)$ .
- s2. If just one of  $s$  and  $t$  refers, then  $s = t \notin S(\alpha)$ .
- s3. If neither  $s$  nor  $t$  refers and  $s = t \in S(\alpha)$ , then  $A(s) \in S(\alpha)$  iff  $A(t) \in S(\alpha)$ .
- s4. If both  $s$  and  $t$  refer to the same individual of  $D$ , then  $A(s) \in S(\alpha)$  iff  $A(t) \in S(\alpha)$ .
- s5. If  $\alpha$  and  $\beta$  agree on the free variables of  $A$ , then  $A \in S(\alpha)$  iff  $A \in S(\beta)$ .

Truth-values for atomic formulae under  $\langle I, S \rangle$  and  $\alpha$  are fixed by  $I$  and  $\alpha$  if all terms refer, and by  $S$  and  $\alpha$  otherwise:

- $v_s1.$  If each  $t_i$  refers, then  $Pt_1 \dots t_k$  is true iff  $\langle \alpha_1, \dots, \alpha_k \rangle \in d(P)$ , where  $t_i$  refers to  $\alpha_i$ ; otherwise,  $Pt_1 \dots t_k$  is true iff  $Pt_1 \dots t_k \in S(\alpha)$ .

<sup>17</sup>The conditions are s1–s4 with ' $S(\alpha)$ ' replaced by ' $S$ '.

- v<sub>s</sub>2. If both  $s$  and  $t$  refer, then  $s = t$  is true iff  $\alpha$  is  $\beta$ , where  $s$  refers to  $\alpha$  and  $t$  refers to  $\beta$ ; otherwise,  $s = t$  is true iff  $s = t \in S(\alpha)$ .

The classical valuation rules v3–v5 are unchanged, except that truth and falsity are now relative to  $\langle I, S \rangle$  and  $\alpha$ . Note that if  $t$  does not refer under  $I$  and  $\alpha$ , then  $\exists x(x = t)$  is false under  $\langle I, S \rangle$  and  $\alpha$ . If  $t$  does not refer because  $D$  is empty, there are no  $x$ -variants of  $\alpha$ , hence no  $x$ -variants  $\beta$  of  $\alpha$  for which  $x = t$  is true under  $\langle I, S \rangle$  and  $\beta$ ; so  $\exists x(x = t)$  is false under  $\langle I, S \rangle$  and  $\alpha$ . If  $D$  is not empty, let  $\beta$  be any  $x$ -variant of  $\alpha$ . Since  $x$  refers under  $I$  and  $\beta$  and  $t$  does not,  $x = t \notin S(\beta)$  and  $x = t$  is false under  $\langle I, S \rangle$  and  $\beta$ ; therefore,  $\exists x(x = t)$  is false under  $\langle I, S \rangle$  and  $\alpha$ .

It can be shown that to any outer domain interpretation  $\langle D_i + D_o, d \rangle$  and assignment  $\alpha$  there corresponds a story interpretation  $\langle D_i, d', S' \rangle$  and assignment  $\alpha'$  that preserves truth-values, and conversely. Thus, adopting story semantics does not require any change in PFL.

Story semantics is now somewhat unfashionable. Lambert [2001] regards his creation as an unsuccessful attempt to develop “a philosophically palatable semantics for positive free logic whose domain consists of a single set of (intuitively) existing objects, whose denotation function is partial, and whose truth definition makes no appeal to other worlds or ‘extensions’ of the domain.” It is unsuccessful, in his view, because a story is “simply a list of sentences governed by some logical laws, hence a story in a very Pickwickian sense indeed.” However, it seems no more Pickwickian than identifying *properties* of  $D$ -individuals with the subsets of  $D$ , as is standard in classical semantics.

Another objection is Bencivenga’s: without a semantic rationale, conditions s1–s5 are *ad hoc*, and the “logical laws” that they build into a story  $S$  are without foundation.<sup>18</sup> For example, if we are going to require that  $a = a \in S(\alpha)$  when  $a$  does not refer under  $I$ , why shouldn’t we also require that  $Pa \in S(\alpha)$  when  $\forall xPx$  is true, but  $a$  does not refer, under  $I$ ? Why can’t  $\forall xA(x) \rightarrow A(t)$  also claim the status of a logical law, contrary to the desires of free logicians? Perhaps this objection can be partially met by arguing that conditions s1–s5 capture the rules of language games about fictional or pretend entities.<sup>19</sup> However, as noted above, this justification will be incomplete unless we can argue that such language games commit

<sup>18</sup>Bencivenga [1986, p. 403], and [this volume, p. 176]. Woodruff [1984, p. 944] characterizes conditions like s1–s5 as “constraints designed to ensure that we get the right results (for instance, that the laws of identity continue to hold).”

<sup>19</sup>I confess that I do not find Walton’s use of this idea very illuminating. According to Walton [1990, p. 400], when Sally claims that Tom Sawyer attended his own funeral, she is claiming that *The Adventures of Tom Sawyer* is such that “to behave in a certain way, to engage in an act of pretense of a certain kind while participating in a game authorized for it, is fictionally to speak the truth.” Unless Sally is a very unusual person, this is false. As an account of when Sally’s assertions of ‘Tom Sawyer attended his own funeral’ are true, it is more promising, but obscure.

participants to bivalence, *i.e.*, to accepting  $A$  or  $\neg A$ , for any atomic sentence  $A$  with a non-referring term.

### 3.3 Negative Bivalent Semantics

Negative free semantics is story semantics without the story: an interpretation  $I$  is precisely as it is in story semantics, that is, it is a partial interpretation defined by  $i_p1$ – $i_p3$  and  $i4$ . The rules of reference  $r_p1$ – $r_p3$  are unchanged, but  $v_s1$  and  $v_s2$  are altered to declare that atomic formulae with non-referring terms are false:

- $v_r1$ . If each  $t_i$  refers, then  $Pt_1 \dots t_k$  is true if  $\langle \alpha_1, \dots, \alpha_k \rangle \in d(P)$ , where  $t_i$  refers to  $\alpha_i$ ; otherwise,  $Pt_1 \dots t_k$  is false.
- $v_r2$ . If  $s$  and  $t$  refer, then  $s = t$  is true if  $\alpha$  is  $\beta$ , where  $s$  refers to  $\alpha$  and  $t$  refers to  $\beta$ ; otherwise,  $s = t$  is false.

The subscript ‘ $r$ ’ is for ‘Russell’. In contrast to outer domain semantics, we have:

$$\begin{aligned}
 Pt & \models \exists x(x = t) \\
 & \not\models t = t \\
 \neg \exists x(x = t) & \models t \neq t.
 \end{aligned}$$

A Hilbert-style system NFL (for ‘negative free logic’) of logical axioms and rules<sup>20</sup> that is strongly complete with respect to this semantics may be obtained from PFL by altering the identity axiom schema  $t = t$  to

$$\forall x(x = x)$$

and adding

$$A(t) \rightarrow \exists x(x = t), \text{ if } A \text{ is atomic.}$$

A somewhat less free version of negative free semantics is employed by Farmer [1995] and Feferman [1995] to formalize reasoning about partial functions in mathematics. Since mathematical domains — natural numbers, sets, etc. — are assumed to be non-empty,  $i_p1$  is replaced by the classical  $i1$ . Since in mathematical practice variables and constants are assumed to refer, assignments are total and  $i_p2$  is replaced by the classical  $i2$ ; the classical reference rules  $r1$  and  $r2$  (resp.) replace  $r_p1$  and  $r_p2$  (resp.). For a Hilbert-style axiomatization LPT of this semantics, with extensions to a partial combinatory logic  $CL_p$  and a partial  $\lambda$ -calculus  $\lambda_p$ , see [Feferman, 1995]. For a type-theoretical extension LUTINS of this semantics that has been axiomatized to provide a basis for automated theorem proving, see

<sup>20</sup>See [Burge, 1974, p. 191] and [Lambert, 1997, p. 83].

[Farmer, 1995]. Following Beeson [1985, p. 98], Farmer and Feferman replace the existence predicate  $E!$  with  $\downarrow$  (read ‘is defined’);  $t \downarrow$  is equivalent to  $\exists x(x = t)$ .

Lambert [1991a, p. 9] characterizes negative free semantics as *Russellian*: there are no entities, Meinongian or virtual, beyond the existents, and non-referring terms render atomic formulae false just as they do for Russell. However, Russell’s truth-values result from treating non-referring terms as descriptive and atomic formulae containing them as abbreviations for more complex formulae that turn out to be false if the descriptions are empty. So his truth-values are rationalized by some analysis of constructions containing non-referring terms. That is not yet the case here. If we hold, with Burge [1974, p. 193] following Aristotle, that “true predications at the most basic level express comments on topics, or attributions of properties or relations to objects,” then we will agree that “lacking a topic or object, basic predications cannot be true.” But this will not get us all the way to negative free semantics unless we buy bivalence, for which Burge does not argue. The rest of the justification will probably have to be provided by particular applications, as when Farmer [1995, p. 282] notes that the “traditional approach to partial functions” in mathematics holds that “formulae are always true or false” and that “application of a predicate is false if any argument is undefined.”

### 3.4 *Intermission: Axiomatizing Equivalence and Implication*

Before turning to non-bivalent free semantics, let us take note of Lin’s [1983] study of equivalence and implication for various bivalent free semantics.

Some elementary logic texts, such as [Tidman and Kahane, 1999], present natural deduction systems that include replacement rules  $A[B] \vdash A[C]$ , where  $A[C]$  results from  $A[B]$  by replacing a part  $B$  by  $C$ . For each such rule, there is a decidable syntactic relation  $R$  such that (i)  $B R C$  or  $C R B$  and (ii)  $R$ -related formulae are logically equivalent. Examples are double negation, where  $\neg\neg B R B$ , and DeMorgan’s laws, where  $\neg(B \vee C) R \neg B \& \neg C$  and  $\neg(B \& C) R \neg B \vee \neg C$ . Any rule of this kind is *closed under ordinary replacement*: if  $B \vdash C$  is an instance, so is  $A[B] \vdash A[C]$ . Let us call such rules *replacement closed*.

Classical semantics supports ordinary replacement in the sense that if  $B$  is logically equivalent to  $C$ , then  $A[B]$  is logically equivalent to  $A[C]$ . So if  $B$  is provable from  $A$  by replacement closed rules,  $B$  is logically equivalent to  $A$ . For classical semantics and free variants that support ordinary replacement, it is natural to ask if the converse holds: is there a system  $S$  of replacement closed rules such that  $B$  is  $S$ -provable from  $A$  if  $B$  is logically equivalent to  $A$ ? Positive results are summarized in Chart *B* of [Lin, 1983, p. 86].

Textbook authors warn students not to apply implicational rules like ADD to parts of formulae, since  $\neg B \vdash \neg(B \vee C)$  is unsound. However,

applications of such rules to certain parts of formulae are sound. If a positive (negative) part  $B$  of a formula  $A[B^+](A[B^-])$  is characterized as in [Schütte, 1960, p. 11], then  $A[B^+] \vdash A[B \vee C^+]$  is sound because the truth (falsity) of a positive (negative) part renders a formula true. Moreover,  $A[B \vee C^-] \vdash A[B^-]$  is sound as well. So we may generalize addition to a pair of rules:  $A[B^+] \vdash A[B \vee C^+]$  and  $A[B \vee C^-] \vdash A[B^-]$ .

A natural question is whether implication can be characterized by a system of such paired rules R1 and R2, where R1 is  $A[B^+] \vdash A[C^+]$ , R2 is  $A[C^-] \vdash A[B^-]$ , and  $B$  bears some decidable syntactic relation  $R$  to  $C$ . Given Schütte's account of positive and negative parts, classical semantics supports polar replacement: if  $B \models C$ , then  $A[B^+] \models A[C^+]$  and  $A'[C^-] \models A'[B^-]$ . This suggests that the paired rules R1 and R2 should be *closed under polar replacement*: if  $B \vdash C$  is an instance of R1 (R2), then  $A[B^+] \vdash A[C^+]$  is an instance of R1 (R2) and  $A'[C^-] \vdash A'[B^-]$  is an instance of R2 (R1). For then, from a proof  $\langle B = B_1, \dots, B_k = C \rangle$  of  $C$  from  $B$ , we could obtain (1) a proof of  $A[C^+]$  from  $A[B^+]$  by  $B_i \rightarrow A[B_i^+]$ , and (2) a proof of  $A'[B^-]$  from  $A'[C^-]$  by  $B_i \rightarrow A'[B_i^-]$  and reversing the resulting sequence of formulae.

Schütte's notion of positive and negative part does not license closure under polar replacement, since a positive (negative) part of a negative part of  $A$  need not be a negative (positive) part of  $A$ . In the case of addition, for example,  $B \vee A \vdash (B \vee C) \vee A$  is an instance of R1, but  $\neg((B \vee C) \vee A) \vdash \neg(B \vee A)$  is not an instance of R2. However, there is another notion of positive and negative part that does support polar replacement and for which implicational replacement holds for classical semantics:  $B$  is a positive (negative) part of  $A$  iff  $B$  occurs within the scope of an even (odd) number of negations in  $A$ , where the quantifier is  $\forall$  and the connectives are  $\neg$ ,  $\&$ , and  $\vee$ , and  $\exists$  and  $\rightarrow$  are defined as usual in terms of them. It is this notion that Lin uses in defining closure under polar replacement. He then characterizes implication in classical semantics and various free (bivalent) variations by systems of implicational rules closed under polar replacement; results are summarized in Chart A of [Lin, 1983, p. 47].

### 3.5 Non-bivalent Semantics: Supervaluations

Frege [1892, p. 70] claims that “anyone who seriously took [‘Odysseus was set ashore at Ithaca while sound asleep’] to be true or false would ascribe to the name ‘Odysseus’ a reference.” His view is that subject-predicate sentences with non-referring terms are truth-valueless. If interpretations and assignments are partial, then Frege's view dictates that  $v_r 1$  be replaced by

$v_f 1$ . If each  $t_i$  refers, then  $Pt_1 \dots t_k$  is true if  $\langle \alpha_1, \dots, \alpha_k \rangle \in d(P)$  and  $Pt_1 \dots t_k$  is false if  $\langle \alpha_1, \dots, \alpha_k \rangle \notin d(P)$ , where  $t_i$  refers to  $\alpha_i$ ; otherwise,

$Pt_1 \dots t_k$  lacks truth-value.

If ‘=’ is regarded as a predicate — or as a binary truth-valued operator that has no output without a pair of inputs — then  $v_r2$  must be replaced by

$v_f2$ . If  $s$  and  $t$  refer, then  $s = t$  is true if  $\alpha$  is  $\beta$  and  $s = t$  is false if  $\alpha$  is not  $\beta$ , where  $s$  refers to  $\alpha$  and  $t$  refers to  $\beta$ ; otherwise,  $s = t$  lacks truth-value.

Once bivalence is lost at the atomic level, we must face questions q5–q7. There are two plausible answers to q5: either the *weak* or the *strong* truth-tables, as Kleene [1950, p. 334] calls them. The weak tables assign precisely the same truth-values as do the classical tables, leaving the compound truth-valueless in all other cases. Thus  $\neg A$  is truth-valueless when  $A$  is truth-valueless, and  $A \rightarrow B$  is truth-valueless when  $A$  or  $B$  (or both) is truth-valueless. The strong tables treat negation in the same way, but preserve certain features of the classical tables for other compounds:  $A \rightarrow B$  is true when  $A$  is false or  $B$  is true, regardless of whether the other constituent has a truth-value. Thus, treating  $A \vee B$  as  $\neg A \rightarrow B$  and  $A \& B$  as  $\neg(A \rightarrow \neg B)$ , disjunctions are true if at least one disjunct is true, and conjunctions are false if at least one conjunct is false. Neither of these answers to q5 will prevent such classical logical truths as  $Pt \vee \neg Pt$  and  $s = t \rightarrow (Ps \rightarrow Pt)$  from ending up with no truth-value when  $t$  and  $s$  do not refer. Note also that  $t = t$  will have no truth-value if  $t$  does not refer.

Supervaluational semantics is an attempt to avoid such alien results, while permitting some atomic formulae to lack truth-values. The basic idea is to consider *completions* of a partial interpretation  $I$  and to revise the valuation rules so that  $A$  is *supertrue* (*superfalse*) under  $I$  and  $\alpha$  if, for each completion  $I'$  of  $I$  and  $\alpha'$  of  $\alpha$ ,  $A$  is true (false) under  $I'$  and  $\alpha'$ , and  $A$  is *supervalueless* under  $I$  and  $\alpha$  otherwise. In van Fraassen’s [1966] original development of the idea, completions of  $I$  are achieved by adding stories  $S$ , which he terms *classical valuations over  $I$* : if  $I = \langle D, d \rangle$ , then  $I' = \langle I, S \rangle$ , where  $S$  is defined as in story semantics.

Since  $Pt \vee \neg Pt$ ,  $s = t \rightarrow (Ps \rightarrow Pt)$ , and  $t = t$  are true under any story interpretation  $\langle I, S \rangle$ , they are supertrue under any partial interpretation  $I$ , and hence logically true with respect to supervaluational semantics. More generally,  $A$  is logically true with respect to supervaluational semantics (in the sense of being supertrue under every partial interpretation  $I$ ) iff  $A$  is logically true with respect to story semantics. Accordingly, the Hilbert-style axiomatization PFL of story semantics is weakly complete with respect to supervaluational semantics.

However, it is not strongly complete:  $Pa \models_s \exists x(x = a)$  but  $Pa \not\models \exists x(x = a)$ , where  $X \models_s A$  iff  $A$  is supertrue under each partial interpretation  $I$  and assignment  $\alpha$  for which each  $B \in X$  is supertrue. Note first that if  $a$  does not refer under  $I$ ,  $Pa$  has no supervalue, since there are stories  $S$  and  $S'$



for which  $Pa \in S$  but  $Pa \notin S'$ .<sup>21</sup> Thus, if  $Pa$  is supertrue under  $I$ , then  $a$  must refer under  $I$ . So  $\exists x(x = a)$  is true under any  $I'$ , where  $I'$  is a completion of  $I$ , and therefore  $\exists x(x = a)$  is supertrue under  $I$ . Accordingly,  $Pa \models_s \exists x(x = a)$ . However,  $\not\models_s Pa \rightarrow \exists x(x = a)$ , since if  $a$  does not refer under  $I$ , then, as noted in Section 3.2,  $\exists x(x = a)$  is false under any  $\langle I, S \rangle$ . However, for some completions  $\langle I, S \rangle$  of  $I$ ,  $Pa$  is true under  $\langle I, S \rangle$  and for others, it is false; thus,  $Pa \rightarrow \exists x(x = a)$  is not supertrue under every interpretation  $I$ . Accordingly, by weak completeness,  $\not\vdash Pa \rightarrow \exists x(x = a)$ . But the deduction theorem holds for PFL, so  $Pa \not\vdash \exists x(x = a)$ .

Since PFL is strongly complete with respect to story semantics, we have  $Pa \models_s \exists x(x = a)$  but  $Pa \not\models \exists x(x = a)$ . This is analogous to the situation in classical semantics, where  $Px \models_g Pa$  but  $Px \not\models Pa$ . This suggests that, just as the generality interpretations of classical logic treat free variables as if they were universally quantified, supervaluations may also involve implicit quantification. Woodruff [1984] has shown that they do indeed, and that it is second-order.<sup>22</sup> Consider a subject-predicate formula  $Pxa$ , a partial interpretation  $I$ , where  $d(a)$  is undefined, and an assignment  $\alpha$ , for which  $\alpha(x) \in D$ . The story  $S$  in a completion  $I' = \langle I, S \rangle$  of  $I$  may be regarded as assigning an extension  $d(P_a)$  to a predicate  $P_a$  defined by  $P_a x \leftrightarrow Pax$ :  $\beta(x) \in d(P_a)$  iff  $Pax \in S(\beta)$ , where  $\beta$  is an  $x$ -variant of  $\alpha$ . Thus,  $Pxa$  is supertrue under  $I$  and  $\alpha$  iff for each  $S$ ,  $Pxa$  is true under  $\langle I, S \rangle$  and  $\alpha$  iff for each extension  $d(P_a)$ ,  $P_a x$  is true under  $I$  and  $\alpha$  iff  $\forall P_a P_a x$  is true under  $I$  and  $\alpha$ . This sketch of the argument assumes that  $a$  does not refer. Woodruff shows how to conditionalize such assumptions to obtain, for any  $A$ , a second-order normal form  $tr(A)$ , such that  $A$  is supertrue under  $I$  and  $\alpha$  iff  $tr(A)$  is supertrue under  $I$  and  $\alpha$ . Moreover,  $tr(A)$  is such that if  $A$  is supertrue under  $I$  and  $\alpha$ , a part of  $tr(A)$  of the form  $\forall P_1 \dots \forall P_k B$ , where  $B$  contains no constant that does not refer under  $I$  and  $\alpha$ , is true under  $I$  and  $\alpha$ .<sup>23</sup>

Woodruff goes on to establish that supervaluational semantics inherits the pathologies of classical second-order semantics. Compactness ( $X$  is satisfiable if every finite subset of  $X$  is satisfiable), the upward Löwenheim–Skolem theorem ( $X$  is satisfiable in  $\omega$  if there is some  $k$  such that  $X$  is satisfiable in  $\{0, \dots, k + j\}$  for each  $j$ ), and the downward Löwenheim–Skolem theorem ( $X$  is satisfiable in  $\omega$  if  $X$  is satisfiable in some larger set) all fail; and finite logical consequence is not recursively axiomatizable (there is no recursive set of axioms such that  $A_1, \dots, A_k \models_s B$  iff  $A_1, \dots, A_k \vdash B$ ).

<sup>21</sup>Since we are dealing with sentences here, I suppress mention of assignments.

<sup>22</sup>Note that supervaluations are also like generality interpretations in not treating connectives as (strict) truth-functions.  $Px \vee \neg Px$  can be true (supertrue) without either  $Px$  or  $\neg Px$  being true (supertrue).

<sup>23</sup>Woodruff's construction is carried out for languages without function-names. In addition, partial interpretations are partial only with respect to constants: domains are non-empty and assignment functions are total.

The question naturally arises whether there are constraints on stories that restore a first-order regime. That is, are there constraints  $\mathcal{C}$  such that if supertruth under  $I$  is understood as truth under each completion  $\langle I, S \rangle$  of  $I$  such that  $S$  satisfies  $\mathcal{C}$ , then the desirable properties of classical first-order semantics (compactness, the Löwenheim–Skolem theorems, and the recursive axiomatizability of logical consequence) are assured? Woodruff [1991] shows that what he terms ‘actualist’ constraints will do the trick. Constraints  $\mathcal{C}$  of this type, which require too much development to describe here, have the effect of making each formula  $A$  superequivalent to an *actualist* formula  $A'$ , in which every occurrence of a constant  $a$  is in a part of the form  $\exists x(x = a)$ ,  $\exists x(x = a) \& B$ , or  $\exists x(x = a) \rightarrow B$ . Superequivalence here means that  $A$  is supertrue under  $I$  iff  $A'$  is supertrue under  $I$ , where supertruth under  $I$  is now truth under  $\langle I, S \rangle$  for each  $S$  that satisfies  $\mathcal{C}$ . What Woodruff [1991, p. 227] terms “the first-order character of actualist semantics” then follows from the fact that actualist formulae are *stable*:  $A$  is true under  $\langle I, S \rangle$  iff  $A$  is true under  $\langle I, S' \rangle$ , for any stories  $S$  and  $S'$ .<sup>24</sup> Woodruff [1991, p. 225] suggests that “the text of some story, theory or myth” could function as an actualist constraint. Details, however, are left to the reader’s imagination; as he notes at the outset, his treatment is quite abstract.

The equivalence of story semantics and outer domain semantics will suggest another way to complete a partial interpretation  $I = \langle D, d \rangle$ : embed it in an outer domain interpretation  $I' = \langle D + D_o, d' \rangle$ , where:

- i<sub>e</sub>2.  $d'(a) = d(a)$  if  $d$  is defined at  $a$ , and  $d'(a) \in D_o$  otherwise.
- i<sub>e</sub>3.  $d'(f)(\alpha_1, \dots, \alpha_k) = d(f)(\alpha_1, \dots, \alpha_k)$  if  $d(f)$  is defined at  $\langle \alpha_1, \dots, \alpha_k \rangle$ , and  $d'(f)(\alpha_1, \dots, \alpha_k) \in D_o$  otherwise.
- i<sub>e</sub>4.  $d'(P)$  is the restriction of  $d(P)$  to  $D^k$ , if  $P$  is  $k$ -place.

Partial assignments  $\alpha$  are similarly completed by requiring that  $\alpha'(x) = \alpha(x)$  if  $\alpha$  is defined at  $x$  and  $\alpha'(x) \in D_o$  otherwise. We can then stipulate that  $A$  is supertrue (superfalse) under  $I$  and  $\alpha$  iff, for each completion  $I'$  of  $I$  and completion  $\alpha'$  of  $\alpha$ ,  $A$  is true (false) under  $I'$  and  $\alpha'$ .

Bencivenga [1980] develops an equivalent semantics that embeds partial interpretations in *classical* interpretations with non-standard valuation

<sup>24</sup>Instead of stories  $S$  over  $I$ , Woodruff speaks of *conventions*  $C$  over  $I$ , which he characterizes as consisting of (1) an equivalence relation  $\sim$  on the constants that do not refer under  $I$  and (2) an extension in  $D$  for each atomic formula whose terms are non-referring constants and, for some  $k \geq 0$ , the first  $k$  variables. (2) treats atomic formulae as predicates (sentences as 0-place predicates) and is subject to the constraints that (a)  $a = x$  and  $x = a$  are true of nothing in  $D$ , (b)  $a = b$  is true if  $a \sim b$ , and (c) if  $A$  is obtained from  $B$  by replacing constants by equivalent constants, then  $A$  and  $B$  are true of the same tuples of individuals of  $D$ . Evidently, to each story  $S$  corresponds a convention  $C$  that gives us the same information about the truth values of atomic formulae with non-referring terms, assuming that  $D$  is non-empty and  $\alpha(x) \in D$ , and conversely.

rules. A classical interpretation  $I' = \langle D', d' \rangle$  over a partial interpretation  $I = \langle D, d \rangle$  satisfies the following conditions:

- i<sub>c</sub>1.  $D'$  is non-empty and  $D \subset D'$ .
- i<sub>c</sub>2. If  $d$  is defined at  $a$ , then  $d'(a) = d(a)$ .
- i<sub>c</sub>3. If  $d(f)$  is defined at  $\langle \alpha_1, \dots, \alpha_k \rangle$ ,  
then  $d'(f)(\alpha_1, \dots, \alpha_k) = d(f)(\alpha_1, \dots, \alpha_k)$ .<sup>25</sup>
- i<sub>c</sub>4.  $d(P) \subset d'(P)$ .

Partial assignments  $\alpha$  are completed by stipulating that  $\alpha'(x) = \alpha(x)$  if  $\alpha$  is defined at  $x$ . Note that, save for i<sub>c</sub>1, these conditions are weaker than those on outer domain completions.

If supertruth is reckoned in terms of classical completions, then  $Pt \vee \neg Pt$ ,  $s = t \rightarrow (Ps \rightarrow Pt)$ , and  $t = t$  will be logically supertrue, but so will  $\exists x(x = a)$  and  $Pt \rightarrow \exists xPx$  — an unwelcome result in free logic. Bencivenga’s technical solution to this problem is essentially to modify the valuation rules v<sub>1</sub>, v<sub>2</sub>, and v<sub>5</sub> for classical interpretations over partial interpretations.<sup>26</sup> The notion of an  $x$ -variant in v<sub>5</sub> must be understood as in outer domain semantics:  $x$  must be assigned something in  $D$ . v<sub>1</sub> and v<sub>2</sub> become:

- v<sub>b</sub>1. If each  $t_i$  refers under  $I$  and  $\alpha$ , then  $Pt_1 \dots t_k$  is true under  $I'$  and  $\alpha'$  if  $\langle \alpha_1, \dots, \alpha_k \rangle \in d(P)$  and  $Pt_1 \dots t_k$  is false under  $I'$  and  $\alpha'$  if  $\langle \alpha_1, \dots, \alpha_k \rangle \notin d(P)$ , where the referent of  $t_i$  under  $I$  and  $\alpha$  is  $\alpha_i$ ; otherwise,  $Pt_1 \dots t_k$  is true under  $I'$  and  $\alpha'$  if  $\langle \beta_1, \dots, \beta_k \rangle \in d'(P)$  and  $Pt_1 \dots t_k$  is false under  $I'$  and  $\alpha'$  if  $\langle \beta_1, \dots, \beta_k \rangle \notin d'(P)$ , where  $\beta_i$  is the referent of  $t_i$  under  $I'$  and  $\alpha'$ .
- v<sub>b</sub>2. If just one of  $s$  and  $t$  refers under  $I$  and  $\alpha$ , then  $s = t$  is false under  $I'$  and  $\alpha'$ ; otherwise,  $s = t$  is true under  $I'$  and  $\alpha'$  if  $s$  and  $t$  refer under  $I'$  and  $\alpha'$  to the same individual, and  $s = t$  is false under  $I'$  and  $\alpha'$  if  $s$  and  $t$  refer under  $I'$  and  $\alpha'$  to different individuals.<sup>27</sup>

<sup>25</sup>Bencivenga’s formal language does not contain function-names, but presumably they would be handled in this way.

<sup>26</sup>The non-standard valuation rules capture valuation under  $I'$  “from the point of view of”  $I$ , as Bencivenga [1986, p. 409] and [this volume, p. 181], puts it. For his own somewhat different presentation of the rules, see [Bencivenga, 1980, pp. 101–103].

<sup>27</sup>The long-winded form of these rules permits their use in Bencivenga’s [1980b] free description theory, where such atomic sentences as  $P(\iota x(x \neq x))$  lack truth-value under  $I'$ . See Section 4.4 below.

v<sub>b</sub>2, like s<sub>2</sub> or i<sub>e</sub>2, implies that  $s = t$  is superfalse if  $s$  refers and  $t$  does not. On Frege’s view, reflected in v<sub>f</sub>2,  $s = t$  should lack truth-value in this case. For a variant of supervaluational semantics that aims to honor Frege’s position, see [Skyrms, 1968]. Skyrms account is not quite correct — subscripts on ‘ $G$ ’ must be reversed in (ii), lest  $s = t$  be supervalueless when  $s$  and  $t$  refer to different individuals (or, where  $s$  is not  $t$ , to the same individual) — and it has the strange consequence that  $\exists x(x = a)$  is supervalueless, not superfalse, when  $a$  does not refer.

These adjustments preserve the logical supertruth of  $Pt \vee \neg Pt$ ,  $t = t$ , and  $s = t \rightarrow (Ps \rightarrow Pt)$ . However, if  $a$  does not refer under  $I$ ,  $\exists x(x = a)$  is superfalse under  $I$  and  $\alpha$ . (Let  $I'$  and  $\alpha'$  classically complete  $I$  and  $\alpha$ . If  $D$  is empty, there are no  $x$ -variants of  $\alpha'$ , so  $\exists x(x = a)$  is false under  $I'$  and  $\alpha'$ . If  $D$  is non-empty and  $\beta'$  is any  $x$ -variant of  $\alpha'$ , then  $x = a$  is false under  $I'$  and  $\beta'$  by  $v_b2$ , so  $\exists x(x = a)$  is false under  $I'$  and  $\alpha'$ .) Moreover,  $Pt \rightarrow \exists xPx$  is not logically supertrue. (If  $t$  does not refer under  $I = \langle D, d \rangle$  and  $\alpha$ , but  $d(P)$  is empty, then  $Pt \rightarrow \exists xPx$  is supervalueless under  $I$  and  $\alpha$ . For  $\exists xPx$  is false under each classical completion  $I' = \langle D', d' \rangle$  of  $I$ , whereas  $d'(P)$  may be defined so as to include or to exclude the referent of  $t$  under  $I'$  and  $\alpha'$ .)

Supervaluations do turn out desired results, subject to the limitations revealed by Woodruff [1984]. Though certain sentences (such as  $Pa$  and  $t = s$ ) may be supervalueless, the classical laws that free logicians like (such as  $t = t$ ,  $Pt \vee \neg Pt$ , and  $(Pt \ \& \ \exists x(x = t)) \rightarrow \exists xPx$ ) are logically supertrue, while those they dislike (such as  $Pt \rightarrow \exists xPx$ ) are not. However, anyone who regards logical properties and relations as fundamentally semantic will regard such a justification of laws as circular. Bencivenga's appeal to classical completions with non-standard valuation rules is designed to provide a semantic rationale for supervaluations, which otherwise appear to be merely a "technical instrument".<sup>28</sup>

Bencivenga's case is as follows: (1) Where terms refer (as in 'Caesar wore a white tunic when he crossed the Rubicon'), truth or falsity may be identified with the outcome of an ideal *practical* experiment that compares what the sentence says with the way the world is. (2) In most cases where terms do not refer (as in 'Pegasus has a white hind leg'), such practical experiments are out of the question; but we may nonetheless identify truth or falsity with the outcome of *mental* experiments (represented by classical completions of partial interpretations) that assign such terms non-existent referents. (3) However, no mental experiment can override the facts, in the sense of altering the outcome of an ideal practical experiment (*e.g.*, that 'Pegasus exists' is false); hence, the non-standard valuation rules  $v_b1$  and  $v_b2$ . (4) Where all mental experiments (so constrained by the facts) agree on a truth-value for a sentence (as with 'Pegasus is Pegasus'), it is reasonable to assign it that value; where they disagree (as with 'Pegasus has a white hind leg'), it is reasonable to regard it as truth-valueless.<sup>29</sup> Together, these conditions give us what Bencivenga [1986, p. 406], and [this volume, p. 179], calls the *counterfactual* theory of truth: "a sentence containing non-denoting singular terms is true (false) if and only if it would be true (false) in case these terms were denoting, no matter what their denotations were but provided that they were *non-existent objects*".

<sup>28</sup>The phrase is Bencivenga's [1986, p. 405], and [this volume, p. 178].

<sup>29</sup>Bencivenga [1980a, p. 225].

Unfortunately, the counterfactual theory of truth seems merely to restate the difficulty. Why should truth, which is ordinarily regarded as *correspondence to fact*, be reckoned in terms of what is *contrary to fact*? Why should we reckon that ‘Pegasus is Pegasus’ *is* true because it *would be* true if, *contrary to fact*, ‘Pegasus’ did refer? We do sometimes decide what is the case by considering what would be the case if things were different, as when we apply the semantic definition of a valid argument. But usually this is not a good idea. The fact that Milosevic would agree to autonomy for Kosovo if he were reasonable does not, unfortunately, tell us that he will do so. Why is truth more like validity than Balkan politics? If partial interpretations merely reflected incomplete information about referents, lack of truth-value would represent ignorance of truth-value and supervaluations would make sense. I don’t know whether ‘The first person born in China in 1999 was a boy’ is true, but clearly ‘The first person born in China in 1999 is the first person born in China in 1999’ is true no matter who this person turns out to be. But lack of information about the referent of ‘Odysseus’ is not what leads Frege to deny a truth-value to ‘Odysseus was set ashore at Ithaca while sound asleep’. I think we know everything there is to know about the referent of ‘Odysseus’: there is no such thing. If supervaluations make sense in free logic, I believe we do not yet know why.

Before leaving supervaluational semantics, let us note a connection with Kripke-style modal semantics established by Barba [1989]. The introduction to Barba’s paper suggests that we will be shown how to translate sentences  $A$  of an ordinary first-order language  $L$  with identity into sentences  $tr(A)$  of the corresponding modal language  $L_{\Box}$  and how to associate with a partial interpretation  $I$  of  $L$  a modal interpretation  $I'$  of  $L_{\Box}$  so that  $A$  is supertrue (superfalse) under  $I$  iff  $tr(A)$  is true (false) under  $I'$ . But no such scheme is possible, since standard modal semantics is bivalent and supervaluational semantics is not. Instead, Barba shows how to associate with a partial interpretation  $I$  of  $L$  a class  $K_I$  of modal interpretations so that  $A$  is supertrue under  $I$  iff  $\Box \Diamond A$  is true under each interpretation in  $K_I$ . Modal interpretations here are non-standard in some respects. For example, they are partial:  $a$  need not refer at world  $w$  — but if it refers at  $w$  to  $\alpha$ ,  $\alpha$  exists in  $w$  and in every world  $w'$  accessible from  $w$ , and  $a$  refers at  $w'$  to  $\alpha$ . However, Barba’s [1989, p. 134] valuation rules  $V_L$  are bivalent: if  $a$  does not refer at  $w$ ,  $Pa$  is true (!) at  $w$ .

### 3.6 Non-bivalent Semantics: Neutral Free Semantics

Supervaluations are the last stop before neutral free semantics, where even  $t = t$  will lack truth-value if  $t$  does not refer, and lack of truth-value at the atomic level is inherited by at least some compounds, among them such classical logical truths as  $Pt \vee \neg Pt$  and  $s = t \rightarrow (Ps \rightarrow Pt)$ , when neither  $s$  nor  $t$  refers. This may not appear to be a very promising destination for the

logician: if logical truths are conceived as those sentences which are always true, there are not going to be many — or perhaps any — logical truths.

However, the situation may not be as bad as it appears to be, for two reasons. First, quantification may restore truth-values: if  $\exists x(x = y)$  is to express ‘ $y$  exists’, then  $\exists x(x = a)$  should be false when  $a$  does not refer under  $I$  and  $\alpha$ , although  $x = a$  will lack truth-value under  $I$  and any  $x$ -variant of  $\alpha$ . This can be achieved by understanding  $\exists xA$  to be true under  $I$  and  $\alpha$  if  $A$  is true under  $I$  and some  $x$ -variant of  $\alpha$  and to be false otherwise. Second, a weaker notion of logical truth —  $A$  is logically true iff  $A$  is never false — may serve as well in many applications.  $t = t$ ,  $Pt \vee \neg Pt$ , and  $s = t \rightarrow (Ps \rightarrow Pt)$  are logically true in this weaker sense, as are such laws of free logic as  $(A(t) \& \exists x(x = t)) \rightarrow \exists xA(x)$ , where  $x$  does not occur in  $t$ .

The underlying semantic rationale for neutral free semantics is Frege’s functional view of reference: predicates and ‘=’ name functions from individuals to truth-values. If functions are operations, as Frege seems to have thought, then the semantic rules governing subject-predicate and identity constructions are  $v_f1$  and  $v_f2$ , for where there is no input to an operation, there is no output either. The truth-functional connectives name truth-functions, so the same line of thought dictates the weak tables for them.<sup>30</sup>  $v_3$  and  $v_4$  become:

$v_f3$ .  $\neg A$  is true if  $A$  is false;  $\neg A$  is false if  $A$  is true;  $\neg A$  lacks truth-value if  $A$  lacks truth-value.

$v_f4$ .  $A \rightarrow B$  is false if  $A$  is true and  $B$  is false;  $A \rightarrow B$  is true if  $A$  is true and  $B$  is true, or  $A$  is false and  $B$  is true, or  $A$  is false and  $B$  is false;  $A \rightarrow B$  lacks truth-value if either  $A$  or  $B$  lacks truth-value.

In classical semantics,  $\exists$  and  $\forall$  may be regarded as naming functions from ‘propositional functions’ to truth-values. Under  $I$  and  $\alpha$ ,  $A(x)$  names the 1-ary propositional function  $\mathcal{A}: D \rightarrow \{T, F\}$  whose value at  $\alpha'(x)$ , where  $\alpha'$  is an  $x$ -variant of  $\alpha$ , is the truth-value of  $A(x)$  under  $I$  and  $\alpha'$ . Then  $\exists(\mathcal{A}) = T$  if  $\mathcal{A}(\alpha) = T$  for some  $\alpha \in D$  and  $\exists(\mathcal{A}) = F$  otherwise, while  $\forall(\mathcal{A}) = T$  if  $\mathcal{A}(\alpha) = T$  for each  $\alpha \in D$  and  $\forall(\mathcal{A}) = F$  otherwise. If these clauses are carried over to the present case, where  $\mathcal{A}$  may be a *partial* function  $D \rightarrow \{T, F\}$ , we have:

$v_f5$ .  $\exists xA$  is true if  $A$  is true for some  $x$ -variant of  $\alpha$ ; otherwise,  $\exists xA$  is false;  $\forall xA$  is true if  $A$  is true for each  $x$ -variant of  $\alpha$ ; otherwise,  $\forall xA$  is false.

Note that  $\exists xA$  and  $\neg\forall x\neg A$  are no longer equivalent in the sense of being true, false, or truth-valueless together. If  $a$  does not refer,  $\exists xPxa$  is false

<sup>30</sup>Woodruff [1970, p. 128] argues that the strong tables — which dictate replacing the second clause of  $v_f4$  by ‘ $A \rightarrow B$  is true if  $A$  is false or  $B$  is true’ and the third by ‘ $A \rightarrow B$  lacks truth-value otherwise’ — are required by Frege’s view that reference is a function of sense. For skepticism, see [Lehmann, 1994, p. 326].

(because  $Pxa$  lacks truth-value for any assignment to  $x$ ) but  $\neg\forall x\neg Pxa$  is true ( $\forall x\neg Pxa$  is false, since  $\neg Pxa$  is always truth-valueless).

The valuation rules  $v_f1$ – $v_f5$  are those of Lehmann [1994], except that he defines  $\forall xA$  as  $\neg\exists x\neg A$ , so that the universal clause of  $v_f5$  becomes ‘ $\forall xA$  is false if  $A$  is false for some  $x$ -variant of  $\alpha$ ; otherwise  $\forall xA$  is true’ and hence  $\forall xPxa$  is true (!) when  $a$  does not refer. Smiley’s [1960, p. 126] rules differ only in taking  $\exists xA(\forall xA)$  to lack truth-value if  $A$  lacks truth-value for some assignments to  $x$  but is otherwise false (true). Thus,  $\exists x(a = fx)$  will lack truth-value under  $\langle D, d \rangle$  if  $a$  refers,  $d(f)$  is partial, but  $d(a) \neq d(f)(\alpha)$  for any  $\alpha \in D$  at which  $d(f)$  is defined.

I have noted that logical truth may be understood in a strong or a weak sense. Similarly, logical consequence may be defined in a number of ways, depending upon whether we want valid inference (1) to lack counterexamples, (2a) to preserve truth, or (2b) to preserve non-falsehood:

$X \models_1 A$  iff there are no  $I$  and  $\alpha$  such that: each  $X$ -formula is true while  $A$  is false.  
 $X \models_{2a} A$  iff there are no  $I$  and  $\alpha$  such that: each  $X$ -formula is true while  $A$  is not true.  
 $X \models_{2b} A$  iff there are no  $I$  and  $\alpha$  such that:  $A$  is false while no  $X$ -formula is false.

$\models_1$  supports contraposition ( $\neg B \models \neg A$  provided  $A \models B$ ) but not transitivity ( $X \models A$  provided  $X' \models A$  and  $X \models B$  for each  $B \in X'$ ):  $\neg\exists x(x = a) \models_1 a = a$  and  $a = a \models_1 \exists x(x = a)$ , but  $\neg\exists x(x = a) \not\models_1 \exists x(x = a)$ . Both  $\models_{2a}$  and  $\models_{2b}$  support transitivity, but neither supports contraposition:  $Pa \models_{2a} \exists x(x = a)$  but  $\neg\exists x(x = a) \not\models_{2a} \neg Pa$ , while  $\neg\exists x(x = a) \models_{2b} \neg Pa$  but  $Pa \not\models_{2b} \exists x(x = a)$ . Both transitivity and contraposition can be had by combining (2a) and (2b), as in Blamey [1986, pp. 5 and 58], to require that valid inference preserve (3) both truth and non-falsity. That is,  $X \models_3 A$  iff  $X \models_{2a} A$  and  $X \models_{2b} A$ .  $\models_3$  is obviously stronger than  $\models_{2a}$  or  $\models_{2b}$ , each of which is stronger than  $\models_1$ . Note that  $A$  is strongly logically true iff  $\models_{2a} A$  and weakly logically true iff  $\models_1 A$  (or  $\models_{2b} A$ ).

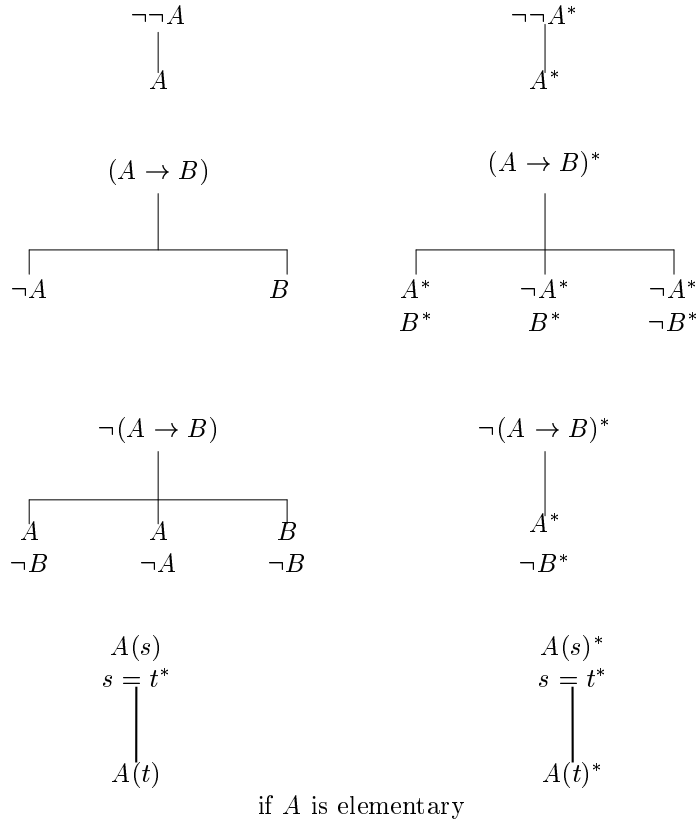
Each of these consequence relations can be expressed in terms of a notion of satisfiability, which in turn can be represented syntactically by a variant of Jeffrey’s [1991] tree method. Add a marker  $*$  to  $L$ , and call  $A^*$  a  $*$ -formula. Let  $Y$  range over sets of formulae and  $*$ -formulae.  $Y$  is  $*$ -satisfiable iff there is some  $I$  and  $\alpha$  for which each  $*$ -formula of  $Y$  is true and no formula of  $Y$  is false. The basic free consequence relations defined above may be expressed in terms of  $*$ -satisfiability as:

$X \models_1 A$  iff  $X^* \cup \{\neg A^*\}$  is not  $*$ -satisfiable  
 $X \models_{2a} A$  iff  $X^* \cup \{\neg A\}$  is not  $*$ -satisfiable  
 $X \models_{2b} A$  iff  $X \cup \{\neg A^*\}$  is not  $*$ -satisfiable

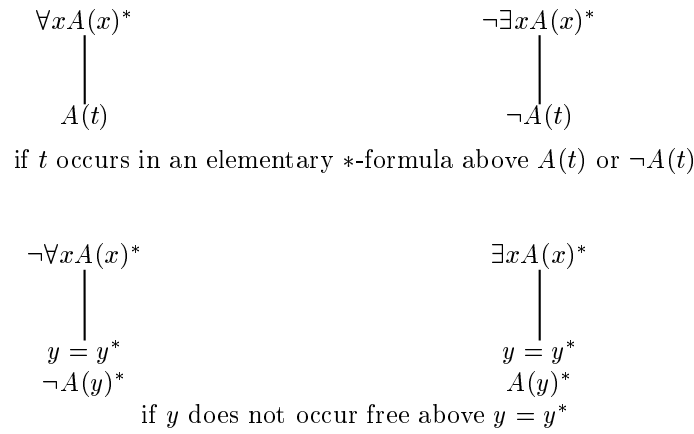
As in the classical case, a tree for finite  $Y$  is obtained by first listing the members of  $Y$  vertically and then extending this list downward in a branching array by application of reductive rules.<sup>31</sup> Here, each classical

<sup>31</sup>The tree method may be modified to accommodate infinite sets of formulae. For a sketch of the argument applied to the free case, see [Lehmann, 1994, Section 4].

rule for a connective is replaced by two rules: one governing formulae, the other governing \*-formulae.



Quantifier rules apply only to \*-formulae:





In addition, we need a rule for converting formulae into \*-formulae:



if  $A$  is (i) a quantified formula or the negation thereof; or (ii) an elementary formula, each term of which occurs in an elementary \*-formula<sup>32</sup> above  $A^*$ .

As the rule for  $\neg(A \rightarrow B)$  will suggest, branches are not closed when they contain  $A$  and  $\neg A$ , as they are in the classical case, because both may lack truth-value. Instead, at least one of  $A$  and  $\neg A$  must be a \*-formula. Similarly,  $t \neq t^*$ , not the classical  $t \neq t$ , closes a branch. A tree is *closed* if each of its branches is closed. Let  $Y \vdash$  iff some tree for some finite subset of  $Y$  is closed. It can be proved that  $Y \vdash$  iff  $Y$  is not \*-satisfiable; for details, albeit for a slightly different system of rules, see [Lehmann, 1994].

A Hilbert-style axiomatization of  $\models_{2a}$  or  $\models_{2b}$  would probably require introducing a non-Fregean connective  $t$ , as in [Smiley, 1960], [Woodruff, 1970], or [Robinson, 1974]:  $tA$  is true if  $A$  is true and is false otherwise.<sup>33</sup> Many classical tautologies are only weakly logically true, rules like ADD do not preserve truth, while rules like MP do not preserve non-falsehood. A Hilbert-style axiomatization of  $\models_1$  seems beyond reach, since  $\models_1$  is not transitive.

#### 4 FREE DESCRIPTION THEORIES

The requirement that singular terms denote something in the range of the variables constrains classical description theory, just as it constrains classical logic. In natural languages there are many singular terms that may be regarded as *descriptions* having the form ‘the (one and only)  $x$  such that  $\dots x \dots$ ’, where ‘ $\dots x \dots$ ’ is some condition on  $x$ : if ‘ $\dots x \dots$ ’ is true only of  $\alpha$ , then ‘the (one and only)  $x$  such that  $\dots x \dots$ ’ refers to  $\alpha$ . For example, ‘the least prime’ refers to 2 because ‘ $x$  is prime and no prime is less than  $x$ ’ is true only of 2. Indeed, one might want to maintain, as does Quine [1997, p. 103], that the “universal form of singular terms” is ‘the (one and only)  $x$  such that  $\dots x \dots$ ’ in the sense that *any* constant or variable singular term may be regarded as having this form. Names, such as ‘Socrates’, can be handled by introducing singular predicates, such as ‘ $x$  Socratizes’. Variable terms,

<sup>32</sup>Assuming  $\forall_f 1$ . If subject-predicate formulae with referring terms can lack truth-value, replace ‘elementary \*-formula’ here with ‘identity \*-formula or negated identity \*-formula’.

<sup>33</sup>If  $A^*$  is identified with  $tA$ , then  $X$  is \*-satisfiable iff some  $I$  and  $\alpha$  fails to falsify any member of  $X$ . Smiley [1960] does not give any proof method, and Woodruff’s [1970] natural deduction system is unsound.

such as ‘the least prime greater than  $y$ ’, may be construed as descriptions whose referents will generally depend upon the values of their variables.

We may represent descriptions by adding to a first-order language  $L$  the description operator  $\iota$  with formation rule: if  $A$  is a formula,  $\iota xA$  is a term. Let  $L_\iota$  be the resulting language. The corresponding semantic rule will specify that under  $I$  and  $\alpha$ ,

r4.  $\iota xA$  refers to whatever uniquely satisfies  $A$

where  $\alpha$  uniquely satisfies  $A$  iff (i)  $A$  is true under  $I$  and  $\alpha'$ , where  $\alpha'$  is the  $x$ -variant of  $\alpha$  for which  $\alpha'(x) = \alpha$ , and (ii)  $A$  is false under  $I$  and any other  $x$ -variant of  $\alpha$ . This understanding we might hope to capture proof-theoretically by adding to CL a schema sometimes termed ‘Lambert’s Law’:

LL  $\quad \forall y(y = \iota xA \leftrightarrow \forall x(A \leftrightarrow y = x))$

where  $y$  is not free in  $A$  (and a free occurrence of  $y$  in  $A$  is now any occurrence not in a part of  $A$  of the form  $\forall yB$  or  $\iota yB$ ).<sup>34</sup>

If  $A$  has no free variables other than  $x$ , then whatever  $\iota xA$  designates under  $I$  and  $\alpha$  will be an individual of  $D$  that is independent of  $\alpha$ , so we may treat  $\iota xA$  as a constant. If in addition  $y$  is free, then whatever  $\iota xA$  designates will be an individual of  $D$  that depends upon the value  $\alpha(y)$  of  $y$ , and we may regard  $\iota xA$  as giving the value of some function at  $y$ . Thus, having descriptions  $\iota xA$  available would permit defining constants  $c$  by  $c = \iota xA$  and function-names  $f$  by  $\forall y(fy = \iota xA)$ .

However, if  $\iota xA$  is to be a singular term, classical semantics demands that it refer to something in the range of the variables. There is no problem if  $\exists!x A$  is true. For then some individual  $\alpha$  of  $D$  will uniquely satisfy  $A$ , and  $\iota xA$  will refer to it by r4; such descriptions are said to be *proper*. But r4 tells us nothing about the referent of an *improper* description  $\iota xA$ . If  $\exists!x A$  is false, so that nothing uniquely satisfies  $A$ , we must nonetheless specify a referent in  $D$  for  $\iota xA$ . Moreover, LL is *false* if  $\exists!x A$  is false, since  $\forall y(y = \iota xA \leftrightarrow \forall x(A \leftrightarrow y = x)) \models \exists!x A$ : both  $\forall x(A \rightarrow x = \iota xA)$  and  $A(\iota xA)$  follow classically from LL, so  $\exists!x A$  follows as well.

Indeed, some instances of LL, as when  $A$  is  $x \neq x$  or  $Px \& \neg Px$ , are *logically false*. The corresponding instances of  $A(\iota xA)$  —  $\iota x(x \neq x) \neq \iota x(x \neq x)$  or  $P(\iota x(Px \& \neg Px)) \& \neg P(\iota x(Px \& \neg Px))$  — are sometimes called *Meinong’s paradox*, after Russell’s derivation of them from Meinong’s principle that ‘...  $x$  ...’ is true of the  $x$  such that ...  $x$  ..., a principle expressed by  $A(\iota xA(x))$ . Lambert [1991b; 1995] shows that Russell’s paradox may also be derived from LL and that ways of evading it in set theory parallel ways of evading Meinong’s paradox in description theory.

<sup>34</sup>Lambert [1991b; 1995] labels this system ‘NTDD’ (for ‘naive theory of definite descriptions’).

### 4.1 Classical Fixes

Accordingly, if we wish to take descriptive terms seriously within the classical framework, we must either modify the formation rule to exclude improper descriptions, or we must specify a referent in  $D$  for improper descriptions and modify LL.

The first approach is associated with Hilbert and Bernays. Recall that we may extend a first-order theory  $T$  to  $T'$  by adding a new constant  $c$  with the defining axiom  $\forall x(x = c \leftrightarrow A_c(x))$ , provided  $x$  is the only free variable of  $A_c(x)$  and  $\exists!xA_c(x)$  is a theorem of  $T$ . Similarly, we may introduce a new function name  $f$  with the defining axiom  $\forall x\forall y(y = fx \leftrightarrow A_f(x, y))$ , provided  $x$  and  $y$  are the only free variables of  $A_f$  and  $\forall x\exists!yA_f(x, y)$  is a theorem of  $T$ . Each addition is really only a notational change. We may eliminate  $c$  from a formula  $B'$  by replacing atomic parts  $Pc$  by  $\exists x(A_c(x) \& Px)$  and the resulting formula  $B$  will be a theorem of  $T$  iff  $B'$  is a theorem of  $T'$ . Similarly, we may eliminate  $f$  from a formula  $B'$  by replacing atomic parts  $Pft$  by  $\exists y(A_f(t, y) \& Py)$ , where  $y$  does not occur in  $t$  — and the resulting formula  $B$  will be a theorem of  $T$  iff  $B'$  is a theorem of  $T'$ .

Since we may think of  $c$  as  $\iota xA_c(x)$  and  $fx$  as  $\iota yA_f(x, y)$ , the conditions on defining  $c$  and  $f$  give the Hilbert-Bernays conditions for considering descriptive terms well-formed:  $\exists!xA_c(x)$  and  $\forall x\exists!yA_f(x, y)$ . Essentially, this amounts to saying that a descriptive term  $\iota xA$  is well-formed only under (consistent) assumptions  $X$  — the axioms of  $T$  — that entail  $\exists!xA$ , assumptions which accordingly function as additional premises in any argument involving  $\iota xA$ .

The Hilbert-Bernays approach has the awkward consequence of making the question of whether  $\iota xA$  is a term undecidable, since logical consequence is not decidable. Normally, of course, the syntactical categories of term, formula, and sentence are decidable, provided the basic categories of constant, variable, ( $k$ -place) function-name, and ( $k$ -place) predicate are decidable. But that is not the case here.

Instead of limiting attention to proper descriptions, we can instead follow Frege and stipulate a referent in  $D$  for the improper descriptions, modifying LL accordingly. To  $L$  we add a constant  $e$  for the designated ('empty') element. An interpretation of  $L$ , is just an interpretation  $\langle D, d \rangle$  of  $L$ , which accordingly will assign  $e$  a referent  $d(e)$  in  $D$ . The reference rule for  $\iota xA$  will now read:

r<sub>f</sub>4. If some individual  $\alpha$  of  $D$  uniquely satisfies  $A$ , then  $\iota xA$  refers to  $\alpha$ ;  
otherwise,  $\iota xA$  refers to  $d(e)$ .

Thus, all instances of the schema

$$F \ B(\iota xA) \leftrightarrow (\exists y(\forall x(A \leftrightarrow y = x) \& B(y)) \vee (\neg \exists y \forall x(A \leftrightarrow y = x) \& B(e)))$$

are logically true. To axiomatize logical truth, we may add

$$\text{LL}_f \quad \forall y(y = \iota x A \leftrightarrow ((\exists! x A \ \& \ A(y)) \vee (\neg \exists! x A \ \& \ y = e)))$$

in place of LL to CL.

Assigning improper descriptions an arbitrary referent is like arbitrarily completing a partial function, and here as there we must be prepared for some weird results: if some other constant  $c$  refers to  $d(e)$ , then  $c = \iota x(x \neq x)$ , will be true, though it will represent sentences like ‘Zero is the non-self-identical number’. Moreover,  $\exists y(y = \iota x(x \neq x))$  and  $\iota x(x = x) = \iota x(x \neq x)$  are logically true.

An alternative to treating descriptions  $\iota x A$  as genuine terms is to follow Russell and to regard formulae  $B(\iota x A)$  in which they appear as giving only the surface form of corresponding sentences of natural language, their logical form being obtained by a transformation of  $B(\iota x A)$  that eliminates descriptions. A sentence like ‘The present King of France is bald’ looks like a subject-predicate sentence with form  $B(\iota x K x)$ , where  $K x$  represents ‘ $x$  is King of France at present’, but  $\iota x K x$  is not a genuine singular term, according to Russell. Why? Because (1) genuine singular terms are meaningful, (2) the meaning of a meaningful singular term is its denotation, and (3) ‘the present King of France’ has no denotation. In Russell’s view, the logical form of ‘The present King of France is bald’ is not subject-predicate, but existential:  $\exists y(\forall x(K x \leftrightarrow y = x) \ \& \ B y)$ , which says that one and only one thing is King of France at present, and that thing is bald. For Russell, we have

$$\text{R} \quad P(\iota x A) \leftrightarrow \exists y(\forall x(A \leftrightarrow y = x) \ \& \ P y)$$

for predicates  $P$ , but not in general.

‘The present King of France is not bald’ looks like a negated subject-predicate sentence with form  $\neg B(\iota x K x)$ ; what is its logical form? If we think of  $\iota x K x$  as occurring in  $B(\iota x K x)$ , the form of  $\neg B(\iota x K x)$  will be  $\neg \exists y(\forall x(K x \leftrightarrow y = x) \ \& \ B y)$ ; if we think of  $\iota x K x$  as occurring in  $\neg B(\iota x K x)$ , the form of  $\neg B(\iota x K x)$  will be  $\exists y(\forall x(K x \leftrightarrow y = x) \ \& \ \neg B y)$ . In the former *narrow-scope* reading, the sentence is true: it is being read as ‘it’s not the case that: the present King of France is bald’. In the latter *wide-scope* reading, the sentence is false: it is being read as ‘the present King of France is non-bald’. If  $A$  is description-free, LL is logically true if  $\iota x A$  is taken to have narrowest *scope*:

$$\forall y(\exists z(\forall x(A \leftrightarrow z = x) \ \& \ y = z) \leftrightarrow \forall x(A \leftrightarrow y = x))$$

But LL is false if  $\iota x A$  is improper and is taken to have wider scope:

$$\begin{aligned} & \forall y \exists z(\forall x(A \leftrightarrow z = x) \ \& \ (y = z \leftrightarrow \forall x(A \leftrightarrow y = x))) \\ & \exists z(\forall x(A \leftrightarrow z = x) \ \& \ \forall y(y = z \leftrightarrow \forall x(A \leftrightarrow y = x))) \end{aligned}$$

Accordingly, the transformation of a formula  $B$  containing descriptions into a  $\iota$ -free formula  $tr(B)$  requires an indication of the scope of the descriptions in  $B$ . This may be done by using the notation  $[\iota xA]B$  to indicate that the scope of  $\iota xA$  in  $B$  is  $B$ . The wide scope reading of  $\iota xKx$  in  $\neg B(\iota xKx)$  would then be indicated by  $[\iota xKx]\neg B(\iota xKx)$  and the narrow scope reading by  $\neg[\iota xKx]B(\iota xKx)$ . More formally, we may add to the formation rules of  $L$  two clauses:

1. If  $B$  is a formula in which  $\iota xA$  appears,  $[\iota xA]B$  is a formula
2. If  $B$  is a formula in which (a) each description  $\iota xA$  that occurs in  $B$  occurs in a subformula  $[\iota xA]C$  of  $B$  and (b) any subformula  $[\iota xA]C$  of  $B$  is such that some occurrence of  $\iota xA$  in  $C$  is not in a subformula  $[\iota xA]C'$  of  $C$ , then  $B$  is a  $[\iota]$ -formula.

For the purposes of (2), subformulae of  $B$  include formulae that appear in descriptions in  $B$ . Scope indicators  $[\iota xA]$  are like quantifiers, and (2) is analogous to a clause defining sentences as formulae lacking free occurrences of variables. (a) rules out unscoped occurrences of descriptions, and (b) rules out vacuous scope indicators. Let this language be  $L_{[\iota]}$ .  $tr$  then maps  $[\iota]$ -formulae of  $L_{[\iota]}$  into  $\iota$ -free formulae of  $L$ :

- t1.  $tr(A) = A$  if  $A$  is not a  $[\iota]$ -formula.
- t2.  $tr$  commutes with connectives and quantifiers.
- t3.  $tr([\iota xA]B(\iota xA)) = \exists y(\forall x(tr(A) \leftrightarrow y = x) \& tr(B(y)))$ .

Despite its enormous influence, Russell's treatment of descriptive terms is ill-motivated and cumbersome. The denotative theory of meaning which led Russell to banish improper descriptions from the realm of terms has little plausibility, especially in view of the many failed attempts to capture intension in extension (*e.g.*, the meaning of a sentence is the proposition it expresses, and that is a *set of possible worlds*). Although scope indicators may be useful to disambiguate constructions like 'The present King of France isn't bald', they introduce a complication that in many cases is of no use. For example, 'Winston Churchill or the present King of France is bald' appears to be unambiguous in English, though Russell's formalism provides two readings:  $Bc \vee [\iota xKx]B(\iota xKx)$ , which is true, and  $[\iota xKx](Bc \vee B(\iota xKx))$ , which is false. There are no scope indicators in English, and assigning Russellian forms is an *ad hoc* business. The validity of 'The janitor is guilty, so the janitor or the accused is guilty' requires a narrow-scope construal of 'the accused'; the validity of 'The accused is not guilty, those who are not guilty are innocent, so the accused is innocent' requires a wide-scope construal of 'the accused' in the premise.

A treatment of descriptions that enables us to handle them as we would other terms, while avoiding the problems of the other two approaches is obviously worth pursuing. Insofar as these problems arise from the straightjacket of classical semantics, we may hope to avoid them either by permitting improper descriptions to refer to individuals of an outer domain or not to refer at all.

#### 4.2 Outer Domain Free Description Theory

An outer domain interpretation for descriptions may be obtained from an outer domain interpretation  $\langle D_i + D_o, d \rangle$  by adding a denotation function  $d'$  for descriptions:

i<sub>o</sub>5.  $d'(\iota xA) \in D_o$ .

$d'(\iota xA)$  will be the referent of  $\iota xA$ , if  $\iota xA$  turns out to be improper. Thus, r4 is rewritten:

r<sub>o</sub>4. If some individual  $\alpha$  of  $D_i$  uniquely satisfies  $A$ , then  $\iota xA$  refers to  $\alpha$ ; otherwise,  $\iota xA$  refers to  $d'(\iota xA)$ .

Relative to this semantics, LL is logically true. If LL is added to PFL, we obtain a complete axiomatization of this semantics. This free description theory is rather weak; let us call it mFD ('m' for 'minimal').<sup>35</sup>

mFD permits many individuals in  $D_o$  and places no restrictions on their assignment as referents to improper descriptions. The limiting case appears to be the one in which  $D_o$  consists of a single individual, which accordingly must be the referent of any improper description. If interpretations of  $L_i$  require as much, then

FD2  $(\neg\exists x(x = s) \& \neg\exists x(x = t)) \rightarrow s = t$

is logically true. If  $s$  and  $t$  do not refer to existents, *i.e.*, to individuals in the inner domain, then they must refer to individuals of the outer domain, but there is just one of these. This semantics may be axiomatized by adding FD2 to mFD. Alternatively, we could add  $\neg\exists y(y = \iota xA) \rightarrow \iota xA = \iota x(x \neq x)$ , where  $y$  is not free in  $\iota xA$ , as in [Scott, 1967, p. 35]. Let us term this theory MFD ('M' for 'maximal').

MFD has obvious affinities with the Fregean treatment of improper descriptions in classical semantics. The classical principle F holds with  $\iota x(x \neq x)$  in place of  $e$ , and we must identify the present King of France with the unicorn in the closet. However,  $\exists y(y = \iota x(x \neq x))$  and  $\iota x(x = x) = \iota x(x \neq x)$  are no longer logically true.

<sup>35</sup>This theory is called FD by Lambert and van Fraassen [1972, p. 160] and MFD by Lambert [1997, p. 118].

Lambert [1997, p. 118] notes that between mFD and MFD lie various free description theories, partially ordered by inclusion. One linear progression is:

$$T_1 \quad \text{mFD} + \iota x(x = t) = t$$

$$T_2 \quad \text{mFD} + (A(t) \ \& \ \forall x(A \rightarrow x = t)) \rightarrow A(\iota xA)$$

$$T_3 \quad \text{mFD} + (\neg \exists x(x = t) \ \& \ A(t)) \rightarrow A(\iota xA)$$

These theories tell us when  $A(\iota xA)$  is true. LL  $\models \exists y(y = \iota xA) \rightarrow A(\iota xA)$ , if  $y$  is not free in  $\iota xA$ , so  $\exists y(y = \iota xA) \rightarrow A(\iota xA)$  is a theorem of mFD and thus of  $T_1$ ,  $T_2$ ,  $T_3$ , and MFD:  $A(\iota xA)$  holds when  $\iota xA$  refers to an existent. If  $A$  is false of every individual, as when  $A$  is  $x \neq x$  or  $Px \ \& \ \neg Px$ , then of course  $A$  is not true of  $\iota xA$ . Otherwise, mFD is non-committal about whether  $A(\iota xA)$  is true, and theories  $T_1$ ,  $T_2$ , and  $T_3$  give us more information.

- $T_1$ . If  $A(x)$  is  $x = t$  and  $\iota xA$  refers to a non-existent, so does  $t$ .  $\iota x(x = t) = t$  identifies these nonexistents and thus makes  $A(\iota xA)$  true when  $A$  is  $t = x$ . Hence, the one and only thing that is Vulcan is Vulcan.
- $T_2$ . If  $t$  refers to a non-existent and  $A$  is true of it, then  $\forall x(A \rightarrow x = t)$  will hold only if  $A$  is not true of any existent. Thus  $(A(t) \ \& \ \forall x(A \rightarrow x = t)) \rightarrow A(\iota xA)$  tells us that  $A(\iota xA)$  provided  $A$  is true of some non-existent but not true of any existent. This will be the case if  $A$  is  $t = x$  and  $t$  does not refer, so  $T_2$  contains  $T_1$ . But it also makes *the* unicorn in the closet *a* unicorn in the closet, since  $Ux$  will not be true of any existent.
- $T_3$ .  $(\neg \exists x(x = t) \ \& \ A(t)) \rightarrow A(\iota xA)$  says that  $A$  is true of  $\iota xA$  provided (i)  $A$  is true of some non-existent. LL assures that  $A$  is true of  $\iota xA$  provided  $A$  is uniquely true of some existent, so the extra content of  $T_3$  over mFD is that  $A$  is true of  $\iota xA$  provided (i) and either (ii)  $A$  is not true of any existent or (iii)  $A$  is true of more than one existent. So the extra content over  $T_2$  is that  $A$  is true of  $\iota xA$  provided (i) and (iii). Hence, *the* lost treasure is *a* lost treasure, since there are mythical lost treasures and more than one real one. Models of MFD have just one non-existent, so if  $\iota xA$  refers to a non-existent,  $A(\iota xA)$  will be true provided  $A$  is true of some non-existent. However, MFD is stronger than  $T_3$ , since the latter permits more than one non-existent.

Another linear progression is:

$$T'_1 \quad \text{mFD} + \forall_c x(A(x) \leftrightarrow B(x)) \rightarrow \iota xA = \iota xB$$

$$T'_2 \quad \text{mFD} + \forall x(A(x) \leftrightarrow B(x)) \rightarrow \iota xA = \iota xB$$

$\forall_c$  is the *comprehensive* universal quantifier with the classical valuation rule v5, where  $D$  is  $D_i + D_o$  and  $x$ -variants may assign values in  $D_0$ .  $\forall_c x(A(x) \leftrightarrow B(x)) \rightarrow \iota xA = \iota xB$  says that  $\iota xA$  and  $\iota xB$  are co-referential provided  $A$  and  $B$  are co-extensive in the strong sense of being true of the same existents and non-existents. Thus, the present King of France needn't be the unicorn in the closet.  $\forall x(A(x) \leftrightarrow B(x)) \rightarrow \iota xA = \iota xB$  says that  $\iota xA$  and  $\iota xB$  are co-referential provided  $A$  and  $B$  are co-extensive in the weaker sense of being true of the same existents. Thus, the present King of France is the unicorn in the closet, though the lost treasure needn't be identified with either:  $T'_2$ , like  $T_2$ , is weaker than MFD. Note that  $\forall x(A(x) \leftrightarrow B(x)) \rightarrow \forall y(y = \iota xA \leftrightarrow y = \iota xB)$ , and hence  $\forall_c x(A(x) \leftrightarrow B(x)) \rightarrow \forall y(y = \iota xA \leftrightarrow y = \iota xB)$ , is a theorem of mFD.

Outer domain semantics for descriptions permits a formal representation of Anselm's ontological argument. Here individuals that exist *in re* belong to the inner domain, while individuals that exist *in intellectu* populate the outer domain. If  $Gx$  represents 'nothing greater than  $x$  can be conceived', then a simple version of the argument is:

$$\frac{\neg \exists y(y = \iota xGx) \rightarrow \neg G(\iota xGx) \quad G(\iota xGx)}{\exists y(y = \iota xGx)}$$

This argument is valid by *modus tollens*. In mFD, the premises are falsifiable, and both are required for validity. By contrast, as Mann [1967] has observed, a Russellian treatment of the descriptions collapses the argument into something trivial and question-begging. The antecedent of the first premise is equivalent to  $\exists x(x \neq x)$  if its form is taken to be  $[\iota xGx] \neg \exists y(y = \iota xGx)$  and to  $\neg \exists! xGx$  if its form is taken to be  $\neg [\iota xGx] \exists y(y = \iota xGx)$  or  $\neg \exists y[\iota xGx](y = \iota xGx)$ ; the consequent is equivalent to  $\exists x(Gx \& \neg Gx)$  if its form is taken to be  $[\iota xGx] \neg G(\iota xGx)$  and to  $\neg \exists! xGx$  if its form is taken to be  $\neg [\iota xGx] G(\iota xGx)$ . Hence the first premise is either logically true or logically false. The second premise is equivalent to  $\exists! xGx$ , as is the conclusion, read either as  $[\iota xGx] \exists y(y = \iota xGx)$  or as  $\exists y[\iota xGx](y = \iota xGx)$ . So the argument becomes:  $A, \exists! xGx / \exists! xGx$ , where  $A$  is logically true or logically false.



### 4.3 Russellian Free Description Theory

To obtain a ‘Russellian’ free description theory for  $L_i$ , we need only supplement negative free semantics with

- r<sub>p</sub>4. If some individual  $\alpha$  of  $D$  uniquely satisfies  $A$ , then  $\iota xA$  refers to  $\alpha$ ; otherwise,  $\iota xA$  does not refer.

While this treatment of descriptions differs from Russell’s in taking  $\iota xA$  to be a genuine term, it does make  $P(\iota xA)$  false if  $\iota xA$  is improper, just as Russell insisted: the Russellian biconditionals R are logically true. Every instance of LL is also logically true, and indeed logical truth relative to this semantics may be axiomatized by adding LL to NFL, as in [Burge, 1974]. Let us call this theory rFD.

Principles defining extensions of mFD generally do not carry over to rFD. Both  $(A(t) \& \forall x(A \rightarrow x = t)) \rightarrow A(\iota xA)$  and  $(\neg \exists x(x = t) \& A(t)) \rightarrow A(\iota xA)$  are OK: if  $\iota xA$  refers, then  $A(\iota xA)$ ; if  $\iota xA$  does not refer, then  $A(t)$  and  $A(\iota xA)$  will have the same truth-value if  $t$  does not refer, and  $t$  cannot refer if  $\neg \exists x(x = t)$  is true or  $(A(t) \& \forall x(A \rightarrow x = t))$  is true while  $\iota xA$  does not refer. The other principles are not OK. The ‘cancellation’ principle  $\iota x(x = t) = t$  is false when  $t$  does not refer. FD2 is not always true; indeed, some instances, such as  $\neg \exists y(y = \iota x(x \neq x)) \rightarrow \iota x(x \neq x) = \iota x(x \neq x)$ , are logically false, since its antecedent is logically true and its consequent is logically false.  $\forall x(A(x) \leftrightarrow B(x)) \rightarrow \forall y(y = \iota xA \leftrightarrow y = \iota xB)$ , which says that if  $A$  and  $B$  are co-extensive,  $\iota xA$  and  $\iota xB$  do not differ in denotation, is always true. But  $\forall x(A(x) \leftrightarrow B(x)) \rightarrow \iota xA = \iota xB$  is not always true; indeed,  $\forall x(x \neq x \leftrightarrow x \neq x) \rightarrow \iota x(x \neq x) = \iota x(x \neq x)$  is logically false.

rFD does not capture Russell’s scope distinctions. We have schema R

$$[\iota xA]P(\iota xA) \leftrightarrow \exists y(\forall x(A \leftrightarrow y = x) \& Py)$$

but not, for example,

$$[\iota xA]\neg P(\iota xA) \leftrightarrow \exists y(\forall x(A \leftrightarrow y = x) \& \neg Py)$$

In a sense, only descriptions with narrowest scope are treated as genuine singular terms.

By introducing machinery for forming *complex predicates*  $\lambda xB$  from formulae  $B$ , Scales [1969] is able to represent scoped descriptions as genuine singular terms satisfying the more general schema

$$S \quad \lambda yB(y)(\iota xA) \leftrightarrow \exists y(\forall x(A \leftrightarrow y = x) \& B(y))$$

Obtain  $L_\lambda$  from  $L_i$  by adding the predicate-forming operator  $\lambda$  with formation rule:  $\lambda x_1 \dots x_k A$  is a  $k$ -place (complex) predicate if the free variables of  $A$  are  $x_1, \dots, x_k$ . ‘Russellian’ interpretations of  $L_\lambda$  are obtained from those of  $L_i$  by stipulating that the extension  $d(\lambda x_1 \dots x_k A)$  of the complex

predicate  $\lambda x_1 \dots x_k A$  is the set of  $k$ -tuples  $\langle \alpha_1, \dots, \alpha_k \rangle$  of elements of  $D$  of which  $A$  is true, where  $A$  is true of  $\langle \alpha_1, \dots, \alpha_k \rangle$  iff  $A$  is true when  $x_i$  is assigned  $\alpha_i$ . Note that if  $t$  does not refer,  $\lambda x \neg Px(t)$  is false while  $\neg Pt$  is true.  $L_\lambda$  thus embodies Aristotle's view that truly attributing a property requires an existing subject, though truly denying an attribution does not; 'Pegasus is wingless' is false, while 'Pegasus does not have wings' is true.

Schema S holds because (1) if  $\iota x A$  refers to  $\alpha$ ,  $A$  is true only of  $\alpha$ , and  $\alpha$  is in the extension of  $\lambda y B(y)$  iff  $B(y)$  is true of  $\alpha$ , and (2) if  $\iota x A$  does not refer,  $\lambda y B(y)(\iota x A)$  is false and, because  $\exists! x A$  is false, so is the right side of S. Logical truth in  $L_\lambda$  may be axiomatized by adding LL and

$$\lambda x_1 \dots x_k A(t_1, \dots, t_k) \leftrightarrow (\exists x_1 (x_1 = t_1) \& \dots \& \exists x_k (x_k = t_k) \& A(t_1, \dots, t_k))$$

to NFL; see Scales [1969, p. 11] and Lambert [1997, p. 112].

Recall that for Russell,  $[\iota]$ -formulae  $A$  of  $L_{[\iota]}$  are abbreviations for  $\iota$ -free formulae  $tr(A)$  of  $L$ :  $tr(A)$  gives the Russellian meaning of  $A$ . We may also translate  $[\iota]$ -formulae  $A$  into formulae of  $L_\lambda$  by  $tr'$ , where  $tr'$  is defined by t1, t2 and

$$t3'. \quad tr'([\iota x A]B(\iota x A)) = \lambda y tr'(B(y))(\iota x tr'(A)).$$

Under any interpretation  $I$  of  $L_\lambda$  and assignment  $\alpha$ ,  $tr'(A)$  gets the same value as  $tr(A)$ : in the basis case where both  $A$  and  $B$  are  $\iota$ -free,  $tr(A)$  is the left side of schema  $S$ , while  $tr'(A)$  is the right side. Accordingly, scoped descriptions can be regarded as genuine singular terms without altering Russellian truth-values, provided their contexts are treated as complex predicates.

Kroon [1991, p. 24] has observed that "the Russellianizing of definite descriptions is a clumsy and unnatural business — far more clumsy and unnatural than its defenders seem to realize". The problematic constructions Kroon has in mind are those in which we refer to something by *describing* a description, as in 'The man denoted by the description John just used is bald.' To represent constructions of the general form 'what's denoted by the description that  $\phi$ s has  $P$ ' *à la* Russell, we would need a description predicate  $DES$  true of descriptions ' $\iota x \phi(x)$ ', a corresponding open-sentence predicate  $COS$  true of pairs  $\langle \iota x \phi(x), \phi(x) \rangle$ , and a satisfaction predicate  $SAT$  true of pairs  $\langle \alpha, \phi(x) \rangle$  iff  $\alpha$  satisfies ' $\phi(x)$ '. Then the ugly Russellian analysis would be:  $\exists x (\exists y (\forall z ((DES(z) \& \phi(x)) \leftrightarrow z = y) \& COS(y, x)) \& \exists w (\forall z (SAT(z, x) \leftrightarrow z = w) \& P(w)))$ . How much simpler it would be if we could write  $P(den(\iota x (DES(x) \& \phi(x))))$ , where  $den$  represents the denotation function.

Kroon develops a modified free Russellian semantics for such constructions. Imagine that  $L_\iota$  has been supplemented with vocabulary that permits naming its terms and formulae (so that  $DES$  may be defined) and let  $L_{\iota'}$

result from  $L_t$  by adding semantic predicates  $TRUE$  and  $DEN$  (so that den may be defined by  $den(x) = y \leftrightarrow DEN(x, y)$ ). Interpretations  $I$  of  $L_t$  are as in negative free semantics, except that Kroon assumes that every individual of  $D$  is named by some constant of  $L_t$  so that reference and valuation rules need be given only for constant terms and sentences.  $d(TRUE)$  and  $d(DEN)$  are defined by a fixed point construction over  $I$ . The extensions of  $I$  used in this construction assign semantic predicates  $P$  both a set  $d_t(P)$  of which they are true and a set  $d_f(P)$  of which they are false; the initial extension  $I_0$  of  $I$  makes  $TRUE$  false of every non-sentence and true of nothing and  $DEN$  false of every pair  $\langle \alpha, \beta \rangle$  where  $\alpha$  is not a constant term and true of nothing. If  $t$  refers to  $\alpha$  and  $\alpha$  is neither in  $d_t(TRUE)$  nor in  $d_f(TRUE)$ , then  $TRUE(t)$  lacks truth-value; if  $s$  refers to  $\alpha$  and  $t$  refers to  $\beta$  and  $\langle \alpha, \beta \rangle$  is neither in  $d_t(DEN)$  nor in  $d_f(DEN)$ , then  $DEN(s, t)$  lacks truth-value.  $r_p4$  is modified so that  $\iota xA$  is *undefined* if  $\exists!xA$  lacks truth-value. As usual in negative free semantics, subject-predicate and identity sentences containing non-referring terms are false; however, if no constituent term fails to refer and at least one is undefined, they lack truth-value. Strong tables are used for the connectives;  $\forall xA$  lacks truth-value if  $A(c)$  is not false for any  $c$  but lacks truth-value for some  $c$ .

#### 4.4 Non-bivalent Free Description Theories

Finally, it is possible to give  $L_t$  a supervaluational or a neutral free semantics, so that non-referring descriptions  $\iota xA$  generate truth-value gaps.

A slight obstacle to extending Bencivenga's [1980] supervaluational semantics to  $L_t$  is that some terms, such as  $\iota x(x \neq x)$ , are now not going to refer under any classical extension  $I'$  and  $\alpha'$  of a partial interpretation  $I$  and assignment  $\alpha$ . Thus, valuation rules v3-v5 need to be rewritten to allow for formulae that are neither true nor false under  $I'$  and  $\alpha'$ . Bencivenga [1980b, p. 396] specifies Kleene's strong tables for the connectives and stipulates that  $\forall xA$  is truth-valueless if  $A$  is never false but lacks truth-value for some assignment to  $x$ :

$\forall xA$  is true (false) under  $I'$  and  $\alpha'$  if  $A$  is true (false) under  $I'$  and  $\alpha'_x$  for each (some)  $x$ -variant  $\alpha'_x$  of  $\alpha$ ; otherwise,  $\forall xA$  lacks truth-value under  $I'$  and  $\alpha'$ .

The reference rule for descriptions will be

If some individual uniquely satisfies  $A$  under  $I$  and  $\alpha$ , then  $\iota xA$  refers to it under  $I'$  and  $\alpha'$ ; if no individual uniquely satisfies  $A$  under  $I$  and  $\alpha$  but some individual uniquely satisfies  $A$  under  $I'$  and  $\alpha'$ , then  $\iota xA$  refers to it under  $I'$  and  $\alpha'$ ; otherwise  $\iota xA$  does not refer under  $I'$  and  $\alpha'$ .

Supervaluational semantics for  $L$  renders  $t = t$  and  $Pt \vee \neg Pt$  logically true, but that is not the case for  $L_\iota$ . Since  $\iota x(x \neq x)$  does not refer under any classical extension  $I'$  of  $I$ , neither  $\iota x(x \neq x) = \iota x(x \neq x)$  nor  $P\iota x(x \neq x) \vee \neg P\iota x(x \neq x)$  is true or false under  $I'$ , so both are supervalueless under (any)  $I$ .  $\iota xPx = \iota xPx$  will also be supervalueless under  $I = \langle D, d \rangle$  if  $d(P)$  is empty (as when  $Px$  represents ‘ $x$  is a unicorn in the closet’), since  $\iota xPx$  will fail to refer under some extension  $I'$  of  $I$ . If you want ‘the unicorn in the closet is the unicorn in the closet’ to be true, then you can follow Bencivena [1980b, p. 398] and modify the notions of supertruth and superfalsity so that  $A$  is *supertrue* (*superfalse*) under  $I$  and  $\alpha$  iff  $A$  has a truth-value under some some classical extension  $I'$  and  $\alpha'$  of  $I$  and  $\alpha$ , and  $A$  is true (false) under each such extension. The semantic rationale for this manoeuvre, however, is unclear.

All instances of LL are logically supertrue for this semantics. Under  $I'$  and  $\alpha'$ ,  $\iota xA$  refers, if at all, to something in  $D$  that uniquely satisfies  $A$ , whereas if  $\iota xA$  does not refer, both  $y = \iota xA$  and  $\forall x(A \leftrightarrow y = x)$  are false of each individual of  $D$  in virtue of  $v_b2$ . The principles generating positive free description theories stronger than mFD are, in general, not logically supertrue. Counterexamples to  $\iota x(x = t) = t$ ,  $(A(t) \& \forall x(A \rightarrow x = t)) \rightarrow A(\iota xA)$ ,  $(\neg \exists x(x = t) \& A(t)) \rightarrow A(\iota xA)$ ,  $(\neg \exists x(x = t) \& \neg \exists x(x = s)) \rightarrow s = t$ , and  $\forall x(A \leftrightarrow B) \rightarrow (\iota xA = \iota xB)$  are provided by  $t = s = \iota x(x \neq x)$  and  $A = B = x \neq x$ . No axiomatization of logical supertruth is given, since Bencivena establishes that no axiomatization is possible.

Description theories that incorporate neutral free semantics have been developed by Stenlund [1973] and Robinson [1974]. Indeed, their systems probably deserve to be regarded as the first complete neutral free logics. Both theories essentially identify referenceless terms with improper descriptions: interpretations of  $L_\iota$  are classical for Robinson and Stenlund.<sup>36</sup> Thus, constants — and variables under assignment<sup>37</sup> — refer *via* i1–i2 and r1–r2 to individuals of  $D$ , and function-names designate total functions on  $D$  by i3.

Robinson’s treatment of free semantics leaves a good deal to the imagination — it must be inferred from his proof system.<sup>38</sup> Except as noted below, however, he appears to be committed to the same rules of reference

<sup>36</sup>Stenlund [1973, p. 63] permits  $D$  to be empty. However, this does not appear to be consistent with his Theorem 6.2.1 [p.66], which states that  $t \downarrow$  is provable iff  $t$  refers under each interpretation, and the fact that  $c \downarrow$  is an axiom [p.17].

<sup>37</sup>Both Stenlund and Robinson treat quantification substitutionally.

<sup>38</sup>For example, Robinson’s [1974, p. 498] rule viii allows us to derive  $\forall xA \rightarrow A(t)$  provided we have derived formulae, written below as  $\forall xA \downarrow$  and  $t \downarrow$ , to the effect that  $\forall xA$  has a truth-value and  $t$  refers. This seems to require the understanding of the quantifiers given below, since if  $\forall xA$  is false and  $t$  refers, we cannot be sure that  $A(t)$  — and therefore  $\forall xA \rightarrow A(t)$  — is not truth-valueless unless  $A(x)$  has a truth-value for any assignment to  $x$ .

and valuation that Stenlund [1973, p. 64] gives explicitly. The classical r3 is modified to:

If  $t_i$  refers to  $\alpha_i$ , then  $ft_1 \dots t_k$  refers to  $d(f)(\alpha_1, \dots \alpha_k)$ ; otherwise,  $ft_1 \dots t_k$  does not refer.

Descriptions are governed by r<sub>p</sub>4, and subject-predicate formulae, identities, and negations by the Fregean valuation rules v<sub>f</sub>1–v<sub>f</sub>3. Robinson appears to endorse the weak reading of  $\rightarrow$  embodied in v<sub>f</sub>4. Stenlund, however, amends it so that  $A \rightarrow B$  is *true* if  $A$  is false and  $B$  is truth-valueless; the weak table would render his system unsound, since  $\exists!x(x \neq x) \rightarrow (\iota x(x \neq x) = \iota x(x \neq x))$  is provable in it. Stenlund regards  $\forall xA$  as truth-valueless if  $A$  is truth-valueless for some assignment to  $x$ , as when  $A$  is  $P\iota y(f(y) = x)$  and  $d(f)$  is not 1-1. Thus, the Fregean valuation rule v<sub>f</sub>5 needs to be modified to:

$\forall xA$  is true if  $A$  is true for each  $x$ -variant of  $\alpha$ , and  $\forall xA$  is false if  $A$  is false for some  $x$ -variant of  $\alpha$  and not truth-valueless for any  $x$ -variant of  $\alpha$ ; otherwise,  $\forall xA$  is truth-valueless.

A peculiar consequence of this understanding of the quantifiers is that sentences like  $\neg\exists x(x = \iota y(y \neq y))$  are not true, but truth-valueless.

As one might expect of a neutral semantics, not all instances of LL are logically true in the sense of being true under every interpretation —  $\forall y(y = \iota x(x \neq x) \leftrightarrow \forall x(A \leftrightarrow y = x))$  is truth-valueless under any interpretation — though none are false under any interpretation. The same substitutions that generate supervalueless instances of the principles that extend mFD will generate truth-valueless instances of them here.

Stenlund supplies a natural deduction system of rules for this semantics, Robinson a Hilbert-style system. Both employ notation for indicating that terms refer and formulae have truth-value.<sup>39</sup> Let us use Beeson's [1985, p. 98] operator  $\downarrow$  for this purpose. Define a  $\downarrow$ -formula as  $e \downarrow$ , where  $e$  is a term or formula, and extend the valuation rules to  $\downarrow$ -formulae by:

$t \downarrow$  is true if  $t$  refers; otherwise,  $t \downarrow$  is false.  
 $A \downarrow$  is true if  $A$  is true or false; otherwise,  $A \downarrow$  is false.

Both systems have axioms and rules of three kinds. Those of the first kind, such as

$$\begin{array}{l} \vdash c \downarrow \\ t \downarrow \vdash ft \downarrow \\ \exists!xA \vdash \iota xA \downarrow \\ t \downarrow \vdash Pt \downarrow \\ s \downarrow t \downarrow \vdash s = t \downarrow \\ A \downarrow \vdash \neg A \downarrow \end{array}$$

<sup>39</sup>Stenlund uses  $t \in I$  for  $t \downarrow$  and  $A \in F$  for  $A \downarrow$ ; Robinson uses  $\triangleleft e$  for  $e \downarrow$ .

permit us to prove  $\downarrow$ -formulae. Rules of the second kind license moving from  $\downarrow$ -formulae to formulae, as by

$$t \downarrow \vdash t = t$$

More interesting examples are Robinson's

$$A_1 \downarrow, \dots, A_k \downarrow \vdash B(A_1, \dots, A_k),$$

provided  $B(A_1, \dots, A_k)$  is a classical tautology,<sup>40</sup> and Stenlund's rule of conditional proof:

$$X \vdash A \rightarrow B \text{ provided } X \vdash A \downarrow \text{ and } X, A \vdash B.$$

More familiar rules of the third kind permit deriving formulae from formulae; they include MP and

$$\exists! x A \vdash A(\iota x A).$$

Both Stenlund and Robinson provide completeness proofs. Robinson sketches a proof that  $K \vdash A$  iff  $K \models_{2a} A$ , where  $A$  is a formula or  $\downarrow$ -formula,  $K$  is a set of  $\iota$ -free sentences, and each non-logical symbol of  $A$  occurs somewhere in  $K$ . Stenlund claims only weak completeness ( $\vdash A$  iff  $\models_{2a} A$ ), though his proof may be generalizable to a result like Robinson's.

## 5 OTHER APPLICATIONS

### 5.1 Predication Again

Recall the Russell-Meinong view that predication presupposes a subject in the strong sense that a subject-predicate form cannot be ascribed to a sentence unless the subject exists. Quine [1960, p. 96] is more liberal: "Predication joins a general term and a singular term to form a sentence that is true or false according as the general term is true or false of the object, if any, to which the singular term refers."<sup>41</sup> This is really just a special case, since Quine counts any open sentence with purely referential occurrences of variables as a predicate: such open sentences are true or false *of* (tuples of) objects.

The generalization to open sentences obliterates Aristotle's distinction between 'Socrates is ill' and 'Socrates is not well', but Quine's account of predication is like Aristotle's in allowing for irreferential terms. Strangely enough, as Lambert [1986, p. 277] observes, Quine's preferred logical idiom has no singular terms at all, except for variables, which do not challenge

<sup>40</sup>This is presumably what Robinson [1974, p. 498] intends by rule vii; the paper is marred by an unusually large number of printing errors and omissions.

<sup>41</sup>See also [Quine, 1953, p. 163], where Quine argues that "the notion that ' $Fa$ ' and ' $\sim Fa$ ' implies 'a exists'" is rooted in the "familiar confusion" of meaning with denotation.

the Russell-Meinong view if their range is non-empty. The others, “a major source of theoretical confusion” in Quine’s view [1953, p. 167], are to be eliminated in favor of predicate constructions *à la* Russell. Lambert argues that Quine ought instead to have embraced a free semantics, and to a limited extent Quine [1997] has recently done so. In any case, his expressed view of predication needs a free setting, where its development is not entirely obvious.

In classical semantics, Quinean predications are extensional in two senses: (1)  $s = t \models A(s) \leftrightarrow A(t)$  and (2)  $\forall x(A(x) \leftrightarrow B(x)) \models A(t) \leftrightarrow B(t)$ . (1) says that if  $s$  and  $t$  are co-referring, then  $A(s)$  is true iff  $A(t)$  is true; it realizes Quine’s condition that the terms in predications occur purely referentially. (2) says that if  $A(x)$  and  $B(x)$  are co-extensive predicates, then  $A(t)$  is true iff  $B(t)$  is true. Free semantics will generally support (1), but not necessarily (2). ‘ $x$  rotates’, ‘ $x$  exists and  $x$  rotates’, and ‘if  $x$  exists, then  $x$  rotates’ are co-extensive predicates, but ‘Vulcan rotates’, ‘Vulcan exists and rotates’, and ‘if Vulcan exists, it rotates’ will not end up with the same truth-values in bivalent free semantics. If  $v$  does not refer in outer domain semantics, we can make  $Rv$  true or false, but  $\exists y(y = v) \& Rv$  is false and  $\exists y(y = v) \rightarrow Rv$  is true regardless. If  $v$  does not refer in negative free semantics,  $Rv$  is false, but  $\exists y(y = v) \& Rv$  is false and  $\exists y(y = v) \rightarrow Rv$  is true.

Extensionality of type (2) may be restored by following Scales [1969] and regarding predications  $A(t)$  as the result of applying a complex predicate  $\lambda xA$  to  $t$ . Recall that the extension  $d(\lambda xA)$  of  $\lambda xA$  in  $L_\lambda$  consists of those individuals of  $D$  (or of  $D_i$ , if we employ outer domain semantics) of which  $A(x)$  is true. Thus, we have  $\forall x(\lambda xA(x) \leftrightarrow \lambda xB(x)) \models \lambda xA(t) \leftrightarrow \lambda xB(t)$ . For discussion, see Lambert [1986; 1997a; 1998]. For a supervaluational treatment of  $L_\lambda$  that supports “general-term extensionality” of this kind, see Lambert and Bencivenga [1986].

On any of these free semantic treatments,  $L_\lambda$  embodies two kinds of predication: ordinary predication, which does not have existential import ( $Pt \not\models \exists x(x = t)$ ), and complex predication, which does ( $\lambda xPx(t) \models \exists x(x = t)$ ). Of course, we needn’t introduce complex predicates to achieve this; we could simply define two types of subject-predicate constructions in  $L$ , say,  $P(t)$  for ordinary predication and  $P[t]$  for predication with existential import. Lambert and Simons [1994] suggest that ordinary predication  $P(t)$  corresponds to *characterization*, while  $P[t]$  corresponds to *classification*. The latter (‘Catso is a tuxedo cat’) presupposes an individual to classify, the former (‘Catso is hungry’) traditionally does not.

## 5.2 Definitions

The fact that neither outer domain nor Russellian free semantics supports  $\forall x(A(x) \leftrightarrow B(x)) \models A(t) \leftrightarrow B(t)$  creates problems for introducing definitions in theories based on them. To take the simplest case, we may wish to

add a new predicate  $P$  to the language of  $T$  with a defining axiom

$$\text{Df} \quad \forall x(Px \leftrightarrow A(x))$$

and then regard  $Pt$  as an abbreviation of  $A(t)$ . But when  $t$  does not refer to an existent, this may not be possible. If  $t$  refers to something in the outer domain, Df does not tell us whether  $Pt \leftrightarrow A(t)$ , because the bound variables range only over the inner domain. If  $t$  does not refer in Russellian free semantics,  $Pt$  will be false but  $A(t)$  may be true, as when  $A$  is  $x \neq x$ .

If outer domains consist of some fixed finite number of individuals  $\alpha_1, \dots, \alpha_k$ , we can name them  $e_1, \dots, e_k$  and use Lambert and Scharle's [1967] trick, noted above in Section 2.2a, to extend quantification to the outer domain: let  $\forall_c x A$  abbreviate  $\forall x A \& A(e_1) \& \dots \& A(e_k)$ . If we then give what Gumb and Lambert [1997] call a "full explicit definition" of  $P$  by

$$\text{Df}_c \quad \forall_c x(Px \leftrightarrow A(x)),$$

we may regard  $Pt$  as an abbreviation of  $A(t)$ , since  $\forall_c x(Px \leftrightarrow A(x)) \models Pt \leftrightarrow A(t)$ . Gumb and Lambert develop this approach to definitions for outer domains with just one individual *err*, which could represent the 'error object' of certain programming languages. A proof of Beth's definability theorem is sketched.

Dwyer's [1988] approach is somewhat more general. In outer domain semantics, partial functions  $D_i \rightarrow D_i$  are represented by total functions  $D_i \rightarrow D_i + D_o$ : if  $\alpha \in D_i$  but  $d(f)(\alpha) \in D_o$ , then  $f$  represents a function that is undefined at  $\alpha$ . More precisely, from an outer domain interpretation  $\langle D_i + D_o, d \rangle$  we may extract a partial interpretation  $\langle D_i, d_p \rangle$ :

Let  $d_p(c) = d(c)$  if  $d(c) \in D_i$ ; otherwise,  $d_p$  is not defined at  $c$ .

If  $\alpha_i \in D_i$ , then let  $d_p(f)(\alpha_1, \dots, \alpha_k) = d(f)(\alpha_1, \dots, \alpha_k)$

if  $d(f)(\alpha_1, \dots, \alpha_k) \in D_i$ ; otherwise,  $d_p(f)$  is not defined at  $\langle \alpha_1, \dots, \alpha_k \rangle$ .

If  $\alpha_i \in D_i$ , let  $\langle \alpha_1, \dots, \alpha_k \rangle \in d_p(P)$  iff  $\langle \alpha_1, \dots, \alpha_k \rangle \in d(P)$ .

Outer domain interpretations that coincide when restricted to the inner domain generate the same partial interpretation  $I$ . The class  $C(I)$  of such "internally invariant" outer domain interpretations can be regarded as representing  $I$ . Dwyer exploits this connection to characterize definability for partial functions from an outer domain perspective.

The problem here again is that the classical conditions on definitions do not carry over. We cannot allow just any formula  $A(x)$  in definition Df of  $P$ , because not every open sentence is *stable* in the sense of being true of the same individuals of  $D_i$  as we move from one interpretation of  $C(I)$  to another. For example, let  $I$  be a partial interpretation in which  $D$  is the set of real numbers, so that under assignment neither  $x/0$  nor  $x + 1/0$  refers.



Under any assignment relative to  $I' \in C(I)$ , however,  $x/0$  and  $x + 1/0$  will designate “unreal” numbers  $\alpha$  and  $\beta$  in  $D_o$ , where  $\alpha$  and  $\beta$  may or may not coincide. Accordingly, as we move from one  $C(I)$ -interpretation to another,  $x/0 = x + 1/0$  will be true of different sets of reals; it is not stable. If  $A(x)$  is  $x/0 = x + 1/0$ , then Df does not characterize any property of reals.

Dwyer [1988, p. 31] develops a sufficient syntactic condition on stability (*viz.*, atomic constituents of  $A$  contain no more than one non-logical symbol), which is used in reformulating the classical conditions on admissible definitions. Free versions of Robinson’s joint consistency theorem, Craig’s interpolation lemma, and Beth’s definability theorem are proved to establish the adequacy of this account.

### 5.3 Modality

As noted at the end of Section 3.5, Barba [1989] has shown how to understand the supertruth of  $A$  in terms of something like the logical truth of  $\Box \Diamond A$ . Two additional connections between free and modal logic are described in this section.

- a. Garson [1991] develops a general system of quantified intensional logic based on free logic. By a general system, he means one (1) from which particular systems can be obtained by specifying (a) constraints on interpretations and (b) additional axioms or rules and (2) for which completeness can be established by a general proof — one easily modified to establish the completeness of these particular systems.

Let us assume a first order language  $L'$  without function-names, but with operators  $\Box$  (if  $A$  is a formula, so is  $\Box A$ ) and  $E!$  (if  $t$  is a term,  $E!t$  is a formula). QS-interpretations  $I = \langle W, w_0, R, D, E, d \rangle$  of  $L'$  are generalizations of Kripke interpretations. As usual,  $W$  is a set of possible worlds,  $w_0 \in W$  represents the actual world,  $R$  is a binary accessibility relation on  $W$ ,  $D$  is a (non-empty) set of possible individuals, and  $d$  assigns *intensions*  $d(P) : W \rightarrow \mathcal{P}(D^k)$  to  $k$ -place predicates  $P$ ,  $d(P)(w)$  being the extension of  $P$  at  $w$ . In Kripke semantics,  $E(w) \subset D$  is the set of individuals that exist in  $w$ ; in Garson’s generalization,  $E(w)$  is a set of *individual intensions*  $W \rightarrow D$ , and  $\alpha$  exists in  $w$  if  $\alpha = f(w)$  for some  $f \in E(w)$ . In Kripke semantics,  $d$  also assigns possible individuals  $d(a)$  to constants  $a$ ; in Garson’s version,  $d$  assigns individual intensions  $d(a)$  to constants  $a$ . Like constants, variables  $x$  are assigned individual intensions  $\alpha(x)$ ; an  $x$ -variant of  $\alpha$  at  $w$  is an assignment  $\beta$  that differs from  $\alpha$  at most at  $x$ , where  $\beta(x) \in E(w)$ . Kripkean interpretations and assignments, in which designation is rigid, are the special case where individual intensions are constant functions.

Under interpretation  $I$  and assignment  $\alpha$ :

- r<sub>i</sub>1.  $x$  refers at  $w$  to  $\alpha(x)(w)$ .
- r<sub>i</sub>2.  $a$  refers at  $w$  to  $d(a)(w)$ .
  
- v<sub>i</sub>1. If  $t_i$  refers at  $w$  to  $\alpha_i$ ,  $Pt_1 \dots t_k$  is true at  $w$  iff  $\langle \alpha_1, \dots, \alpha_k \rangle \in d(P)(w)$ .
- v<sub>i</sub>2. If  $t_i$  refers at  $w$  to  $\alpha_i$ , then  $t_1 = t_2$  is true at  $w$  iff  $\alpha_1 = \alpha_2$ .
- v<sub>i</sub>3.  $\neg A$  is true at  $w$  iff  $A$  is false at  $w$ .
- v<sub>i</sub>4.  $A \rightarrow B$  is false at  $w$  iff  $A$  is true at  $w$  and  $B$  is false at  $w$ .
- v<sub>i</sub>5.  $\forall x A$  is true at  $w$  iff  $A$  is true at  $w$  for each  $x$ -variant of  $\alpha$  at  $w$ .
- v<sub>i</sub>6.  $\Box A$  is true at  $w$  iff  $A$  is true at each  $w'$  such that  $wRw'$ .
- v<sub>i</sub>7.  $E!t$  is true at  $w$  iff the intension of  $t \in E(w)$ .
- v<sub>i</sub>8.  $A$  is true iff  $A$  is true at  $w_0$ .

As noted in Section 2.2b, Kripke-semantics is basically outer domain free semantics, with the individuals that exist in  $w$  constituting  $w$ 's inner domain, while the rest of  $D$  functions as  $w$ 's outer domain. In Kripke semantics,  $E!t$  can be defined by  $\exists x(x = t)$ , but not here: v<sub>i</sub>7 treats  $E!$  as a predicate of intensions. If  $R$  is universal,  $E!t$  is equivalent to  $\exists x \Box(x = t)$ ; but in general existence is not definable from identity.

For this semantics, Garson sketches a complete Hilbert-style system GS, consisting of (a) propositional modal axioms and rules appropriate to the accessibility relation  $R$ , (b) identity axioms and rules that we may (with the stipulation that  $E!t$  is *not* atomic) identify with the identity axioms of PFL, and (c) quantifier rules which are generalizations of those of free logic:<sup>42</sup>

$$\text{GUI} \quad G[\forall x A] \vdash G[E!t \rightarrow A(t)]$$

GUG

If  $\vdash G[E!t \rightarrow A(t)]$  and  $t$  does not occur in  $G[\forall x A]$ , then  $\vdash G[\forall x A]$

Here  $G[B]$  is any formula of one of the following forms

$$\frac{B}{A \rightarrow B} \quad \frac{\Box(A_1 \rightarrow \dots \Box(A_k \rightarrow B) \dots)}{A \rightarrow \Box(A_1 \rightarrow \dots \Box(A_k \rightarrow B) \dots)}$$

<sup>42</sup>This is apparently what Garson [1991, p. 134] intends by rules GUI and GUG, which are not clear as stated.

After sketching a Henkin-style completeness argument for GS, Garson illustrates the generality of QS+GS by showing how to obtain several familiar systems (including Kripke's) as special cases.

- b. Schweizer [1990] develops Skryms' [1978] defense of the metalinguistic reading of necessity claims against Montague's [1963] argument that such readings are incoherent. The metalinguistic reading treats necessity not as an operator  $\Box$  (for 'necessarily') applying to sentences, but as a predicate  $N$  (for 'is necessary') applying to names of sentences.  $M$ -interpretations  $I$  of such languages are pairs  $\langle I_0, C \rangle$ , where  $C$  is a set of interpretations and  $I_0 \in C$ .  $A$  is true under  $I$  iff  $A$  is true under  $I_0$ , and  $d(N)$  is such that if  $\ulcorner A \urcorner$  names  $A$ ,  $N\ulcorner A \urcorner$  is true under  $I' \in C$  iff  $A$  is true under each  $C$ -interpretation.

Gödel-numbering allows us to develop the metalinguistic interpretation of necessity in an extension  $T$  of formal arithmetic, whose language includes the predicate  $N$ . If  $\ulcorner A \urcorner$  is the numeral for  $g(A)$ , we want to read  $N\ulcorner A \urcorner$  as ' $\ulcorner A \urcorner$  is necessary'. To support this reading, Montague argues,  $T$  should be such that for any sentence  $A$ ,

- (1)  $\vdash_T A$ , then  $\vdash_T N\ulcorner A \urcorner$
- (2)  $\vdash_T N\ulcorner A \urcorner \rightarrow A$

Assuming  $T$ 's proof method is complete, (1) says that if  $\ulcorner A \urcorner$  is necessary in the sense of being true in every model of  $T$ , then it is provable that  $\ulcorner A \urcorner$  is necessary. (2) says that the standard modal principle 'if  $\ulcorner A \urcorner$  is necessary, then  $A$  is true' is provable. Now diagonalization gives us a sentence  $B$  such that

- (3)  $\vdash_T \neg N\ulcorner B \urcorner \leftrightarrow B$

(2) and (3) imply  $\vdash_T \neg N\ulcorner B \urcorner$ . But  $\vdash_T \neg N\ulcorner B \urcorner$  and (3) imply  $\vdash_T B$ , which with (1) implies  $\vdash_T N\ulcorner B \urcorner$ . So  $T$  is inconsistent.

Let  $L$  be a first-order language that includes the language of formal arithmetic, let  $L_\Box$  be the standard modal extension of  $L$ , and let  $L_N$  result from  $L$  by adding the necessity predicate  $N$ . Assuming a Gödel-numbering of  $L_N$ , we may translate  $L_\Box$ -formulae  $A$  into  $L_N$ -formulae  $tr(A)$  by:  $tr(\Box A) = N\ulcorner A \urcorner$ ;  $tr$  commutes with  $\neg$ ,  $\rightarrow$ , and  $\forall x$ ; and  $tr(A) = A$  for  $L$ -formulae  $A$ . If  $I = \langle W, w_0, R, D, E, d \rangle$  is a Kripke interpretation of  $L_\Box$  with universal accessibility relation  $R$ ,<sup>43</sup> Schweizer shows how to obtain an  $M$ -interpretation  $tr(I)$  of  $L_N$  so that for sentences  $A$ ,  $A$  is true under  $I$  iff  $tr(A)$  is true under  $tr(I)$ . The

---

<sup>43</sup>Schweizer [1990, p. 165] stipulates only that  $I$  is an S5 interpretation ( $R$  is an equivalence relation), but his construction assumes that  $R$  is universal. This is a stronger assumption:  $\Box(Pa \ \& \ \neg \forall x Px)$  can be true if  $R$  is an equivalence relation but not if  $R$  is universal, assuming as usual that  $D = \cup_{w \in W} E(w)$ .

construction evades Montague's problem because the Gödel-sentence  $B$  that issues from diagonalization is not  $tr(A)$  for any sentence  $A$ , whereas the modal axiom holds only in the form  $N^\top tr(A)^\top \rightarrow tr(A)$ .

Since  $tr(I)$  is an  $M$ -interpretation  $\langle I_0, C \rangle$ ,  $C$  is a set of interpretations such that  $N^\top A^\top$  is true under  $tr(I)$  iff  $A$  is true under each interpretation in  $C$ . The connection with free logic is that these  $C$ -interpretations are outer domain interpretations. At each  $w \in W$ ,  $I$  induces an outer domain interpretation  $I(w) = \langle D_i + D_o, d_w \rangle$  of  $L$ , where  $D_i = E(w)$ ,  $D_o = D - E(w)$ , and  $d_w$  is the restriction of  $d$  to  $w$ :  $d_w(a) = d(a)$  and  $d_w(P) = d(P)(w)$ . When  $d_w$  is appropriately extended to  $N$ , we can identify  $I_0$  with  $I(w_0)$  and  $C$  with  $\{I(w) | w \in W\}$ .<sup>44</sup> The extension of  $d_w$  is in stages corresponding to the number of nested occurrences of  $\Box$  in  $A$ : we put  $g(tr(A))$  in  $d_w(N)$  at stage  $k + 1$  provided  $tr(A)$  is true under  $I(w)$  at stage  $k$ . At stage 0,  $A$  is  $\Box$ -free, so  $tr(A) = A$ , which gets a truth-value under each outer domain interpretation  $I(w)$ .

Schweizer [1990, p. 170] states that "analogous equivalence results can be obtained for the other normal systems of quantified modal logic, by simply utilizing the relevant accessibility relation  $R \dots$  within the eligible set of models." Presumably, his suggestion is that  $M$ -interpretations  $I$  now be conceived as triples  $\langle I_0, C, R \rangle$ , where  $C$  is a set of interpretations,  $R$  is a binary relation on  $C$ , and  $I_0 \in C$ . Then  $A$  is true under  $I$  iff  $A$  is true under  $I_0$  and  $N^\top A^\top$  is true under  $I' \in C$  iff  $A$  is true under each  $I''$  such that  $I'R I''$ . The ordinary notion of a metalinguistic interpretation is then the special case where  $R$  is universal.

#### ACKNOWLEDGEMENTS

I thank Karel Lambert for directing my attention to many of the works cited in this survey.

*University of Connecticut, USA.*

#### BIBLIOGRAPHY

In cases where an essay has been reprinted, sometimes with cuts or other changes, page-reference citations are to the reprinted version.

- [Barba, 1989] J. Barba. A modal version of free logic, *Topoi*, **8**, 131–5, 1989.  
 [Bencivenga, 1980] E. Bencivenga. Free semantics. In *Italian Studies in the Philosophy of Science*, M. Dalla Chiara, ed. pp. 31–48. Reidel, Dordrecht, 1980. Reprinted in [Lambert, 1991, pp. 98–110].

<sup>44</sup>All of the  $a$ -variants of  $I(w)$ , which differ from  $I(w)$  at most at  $d_w(a) \in E(w)$ , must also be included in  $C$ .

- [Bencivenga, 1980a] E. Bencivenga. Truth, correspondence, and non-denoting singular terms. *Philosophia*, **9**, 219–229, 1980.
- [Bencivenga, 1980b] E. Bencivenga. Free semantics for definite descriptions, *Logique et analyse*, **23**, 393–405, 1980.
- [Bencivenga, 1986] E. Bencivenga. Free logics. In *Handbook of Philosophical Logic, Vol. III*, D. Gabbay and F. Guentner, eds. pp. 373–426. Reidel, Dordrecht, 1986. Reproduced in this volume.
- [Beeson, 1985] M. Beeson. *Foundations of Constructive Mathematics*. Springer-Verlag, Berlin, 1985.
- [Blamey, 1986] S. Blamey. Partial logic. In *Handbook of Philosophical Logic, Vol. III*, D. Gabbay and F. Guentner, eds. pp. 1–70. Reidel, Dordrecht, 1986.
- [Burge, 1974] T. Burge. Truth and singular terms. *Nous*, **8**, 309–325, 1974. Reprinted in [Lambert, 1991, pp. 189–204].
- [Church, 1956] A. Church. *Introduction to Mathematical Logic, Vol. I*. Princeton University Press, Princeton, 1956.
- [Cocchiarella, 1991] N. Cocchiarella. Quantification, time, and necessity. In [Lambert, 1991, pp. 242–256].
- [Dwyer, 1988] R. C. Dwyer. *Denoting and Defining: A Study in Free Logic*. University Microfilms International, Ann Arbor, 1988.
- [Ebbinghaus, 1969] H.-D. Ebbinghaus. Über eine Prädikatenlogik mit partiell definierten Prädikaten und Funktionen. *Archiv für mathematische Logik und Grundlagenforschung*, **12**, 39–53, 1969.
- [Evans, 1979] G. Evans. Reference and contingency. *The Monist*, **62**, 161–189, 1979.
- [Farmer, 1995] W. M. Farmer. Reasoning about partial functions with the aid of a computer. *Erkenntnis*, **43**, 279–294, 1995.
- [Feferman, 1995] S. Feferman. Definedness. *Erkenntnis*, **43**, 295–320, 1995.
- [Frege, 1892] G. Frege. On sense and reference. In *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black, eds. pp. 56–78. Basil Blackwell, Oxford, 1966.
- [Garson, 1991] J. Garson. Applications of free logic to quantified intensional logic. In [Lambert, 1991, pp. 111–142].
- [Gumb, 1998] R. D. Gumb. Does identity precede existence? Read at World Congress of Philosophy, Boston, 1998.
- [Gumb and Lambert, 1997] R. D. Gumb and K. Lambert. Definitions in nonstrict positive free logic. *Modern Logic*, **7**, 25–55, 1997.
- [Jeffrey, 1991] R. Jeffrey. *Formal Logic: Its Scope and Limits*. McGraw Hill, New York, 1991.
- [Kleene, 1950] S. C. Kleene. *Introduction to Metamathematics*. D. van Nostrand, Princeton, 1950.
- [Kleene, 1967] S. C. Kleene. *Mathematical Logic*. John Wiley & Sons, New York, 1967.
- [Kroon, 1991] F. W. Kroon. Denotation and description in free logic. *Theoria*, **57**, 17–41, 1991.
- [Lambert, 1986] K. Lambert. Predication and ontological commitment. In *Die Aufgaben der Philosophie in der Gegenwart: Aktien des 10. Internationalen Wittgenstein Symposiums*, W. Leinfellner and F. N. Wuketits, eds. pp. 281–287. Hölder-Pichler-Temsky, Wien, 1986. Reprinted in [Lambert, 1991, pp. 273–284].
- [Lambert, 1991] K. Lambert, ed. *Philosophical Applications of Free Logic*. Oxford University Press, New York, 1991.
- [Lambert, 1991a] K. Lambert. The nature of free logic. In [Lambert, 1991, pp. 3–14].
- [Lambert, 1991b] K. Lambert. A theory about logical theories of “expressions of the form ‘the so and so’, where ‘the’ is in the singular”. *Erkenntnis*, **35**, 337–346, 1991.
- [Lambert, 1995] K. Lambert. On the reduction of two paradoxes. In *Physik, Philosophie und die Einheit der Wissenschaften*, L. Krüger and B. Falkenburg, eds. pp. 21–32. Spektrum Akademischer Verlag, Heidelberg, 1995.
- [Lambert, 1997] K. Lambert. *Free Logics: Their Foundations, Character, and Some Applications Thereof*. Academia Verlag, Sankt Augustin, 1997.
- [Lambert, 1997a] K. Lambert. Nonextensionality. In *Das weite Spektrum der analytischen Philosophie*, W. Lenzen, ed. pp. 135–148. de Gruyter, Berlin, 1997.

- [Lambert, 1998] K. Lambert. Fixing Quine's theory of predication. *Dialectica*, **52**, 153–160, 1998.
- [Lambert, 2001] K. Lambert. From predication to programming. *Minds and Machines*, **11**, 2001.
- [Lambert and Bencivenga, 1986] K. Lambert and E. Bencivenga. A free logic with simple and complex predicates. *Notre Dame Journal of Formal Logic*, **27**, 247–256, 1986.
- [Lambert and Scharle, 1967] K. Lambert and T. Scharle. A translation theorem for two systems of free logic. *Logique et Analyse*, **39–40**, 328–341, 1967.
- [Lambert and Simons, 1994] K. Lambert and P. Simons. Characterizing and classifying: explicating a biological distinction. *The Monist*, **77**, 315–328, 1994.
- [Lambert and van Fraassen, 1972] K. Lambert and B. C. van Fraassen. *Derivation and Counterexample: An Introduction to Philosophical Logic*. Dickenson, Encino, 1972.
- [Leblanc, 1968] H. Leblanc. On Meyer and Lambert's quantificational calculus FQ. *Journal of Symbolic Logic*, **33**, 275–280, 1968.
- [Lehmann, 1994] S. Lehmann. Strict Fregean free logic. *Journal of Philosophical Logic*, **23**, 307–336, 1994.
- [Lin, 1983] Y. Lin. *Replacement-Closed Rules for Free and for Classical Logic*. University Microfilms International, Ann Arbor, 1983.
- [Mann, 1967] W. E. Mann. Definite descriptions and the ontological argument. *Theoria*, **30**, 211–229, 1967. Excerpted in [Lambert, 1991, pp. 257–272].
- [Mendelsohn, 1989] R. L. Mendelsohn. Objects and existence: reflections on free logic. *Notre Dame Journal of Formal Logic*, **30**, 604–623, 1989.
- [Meyer and Lambert, 1968] R. Meyer and K. Lambert. Universally free logic and standard quantification theory. *Journal of Symbolic Logic*, **33**, 8–26, 1968.
- [Montague, 1963] R. Montague. Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability. *Acta Philosophica Fennica*, **16**, 153–167, 1963. Reprinted in *Formal Philosophy: Selected Papers of Richard Montague*, R. H. Thomason, ed. pp. 286–302. Yale University Press, New Haven, 1978.
- [Posy, 1982] C. J. Posy. A free IPC is a natural logic: strong completeness for some intuitionistic free logics. *Topoi*, **1**, 30–43, 1982. Reprinted in [Lambert, 1991, pp. 49–81].
- [Quine, 1948] W. V. O. Quine. On what there is. *Review of Metaphysics*, **2**, 21–38, 1948. Reprinted in [Quine, 1963, pp. 1–19].
- [Quine, 1953] W. V. O. Quine. Meaning and existential inference. In [Quine, 1963, pp. 160–167].
- [Quine, 1960] W. V. O. Quine. *Word and Object*. MIT Press, Cambridge, 1960.
- [Quine, 1963] W. V. O. Quine. *From a Logical Point of View*. Harper Torchbooks, New York, 1963.
- [Quine, 1969] W. V. O. Quine. *Set Theory and its Logic*. Harvard University Press, Cambridge, 1969.
- [Quine, 1997] W. V. O. Quine. Free logic, description, and virtual classes. *Dialogue*, **36**, 101–108, 1997.
- [Robinson, 1974] A. Robinson. On constrained denotation. In *Nonstandard Analysis and Philosophy* (vol. 2 of *Selected papers of Abraham Robinson*, H. Keisler, et al., eds.), W. Luxemburg and S. Körner, eds. pp. 493–504. Yale University Press, New Haven, 1979.
- [Scales, 1969] R. D. Scales. *Attribution and Existence*. University Microfilms International, Ann Arbor, 1969.
- [Schütte, 1960] K. Schütte. *Beweistheorie*, Springer-Verlag, Berlin, 1960.
- [Schweizer, 1990] P. Schweizer. *A Metalinguistic Interpretation of Modality*, University Microfilms International, Ann Arbor, 1990.
- [Scott, 1967] D. Scott. Existence and description in formal logic. In *Bertrand Russell: Philosopher of the Century*, R. Schoenman, ed. pp. 181–200. Little, Brown and Company, Boston, 1967. Reprinted in [Lambert, 1991, pp. 28–48].
- [Shoenfield, 1967] J. R. Shoenfield. *Mathematical Logic*. Addison-Wesley, Reading, 1967. Re-issued in paperback by A. K. Peters, 2001.
- [Simons, 1991] P. M. Simons. Free part-whole theory. In [Lambert, 1991, pp. 285–305].

- [Skyrms, 1968] B. Skyrms. Supervaluations: identity, existence, and individual concepts. *Journal of Philosophy*, **65**, 477–482, 1968.
- [Skyrms, 1978] B. Skyrms. An immaculate conception of modality. *Journal of Philosophy*, **75**, 77–96, 1978.
- [Smiley, 1960] T. Smiley. Sense without denotation. *Analysis*, **20**, 125–135, 1960.
- [Stenlund, 1973] S. Stenlund. *The Logic of Description and Existence*. Filosofiska Studier Nr. 18, Uppsala Universitet, 1973.
- [Strawson, 1952] P. F. Strawson. *Introduction to Logical Theory*. Methuen, London, 1952.
- [Tidman and Kahane, 1999] P. Tidman and H. Kahane. *Logic and Philosophy: A Modern Introduction*. Wadsworth, Belmont, 1999.
- [Trew, 1970] A. Trew. Nonstandard theories of quantification and identity. *Journal of Symbolic Logic*, **35**, 267–294, 1970.
- [van Fraassen, 1966] B. C. van Fraassen. Singular terms, truth-value gaps, and free logic. *Journal of Philosophy*, **63**, 481–495, 1966. Reprinted in [Lambert, 1991, pp. 82–97].
- [van Fraassen, 1968] B. C. van Fraassen. Presupposition, implication, and self-reference. *Journal of Philosophy*, **65**, 136–152, 1968. Reprinted in [Lambert, 1991, pp. 205–221].
- [Walton, 1990] K. L. Walton. *Mimesis as Make-Believe: On the Foundations of the Representational Arts*. Harvard University Press, Cambridge, 1990.
- [Woodruff, 1970] P. W. Woodruff. Logic and truth-value gaps. In *Philosophical Problems in Logic*, K. Lambert, ed. pp. 121–142. Reidel, Dordrecht, 1970.
- [Woodruff, 1984] P. W. Woodruff. On supervaluations in free logic, *Journal of Symbolic Logic*, **49**, 943–950, 1984.
- [Woodruff, 1991] P. W. Woodruff. Actualism, free logic, and first-order supervaluations. In *Existence and Explanation*, W. Spohn, *et al.* eds. pp. 219–231. Kluwer Academic Publishers, Boston, 1991.





STEPHEN BLAMEY

## PARTIAL LOGIC

### INTRODUCTION

When I was originally asked to write about ‘partial logic’ for the first edition of the *Handbook*, I was a little puzzled: I was taken to be an expert in an apparently well defined subject area that I didn’t know existed. But it turned out to be the sort of thing I had written about in my D.Phil. thesis, so I had somewhere to start. Nowadays the label ‘partial logic’ is much more familiar, and a lot of work is being done in the area it covers. The bulk of my own work, though—most of it dating right back to thesis days—has not yet been published: I have been bewilderingly bad about this. In particular, the various promises made in the first edition about forthcoming work have still not been fulfilled. In spite of this, I have resisted the temptation just to shove in more material of my own for the second edition—except in small ways here and there. Additions are largely in response to what has newly appeared in print.

A wide range of work will be surveyed (much more now than in the first edition), but the backbone of this chapter is the development of what I call ‘simple partial logic’. It is against this backbone that other more sophisticated projects are discussed. Simple partial logic results from the simple-minded following through of the idea that classical logic may be loosened up to cater for non-denoting singular terms and neither-true-nor-false sentences—to cater for them in a uniform way as semantically ‘undefined’ items—and at the same time to cater for ‘partially defined’ functors: term-forming functors, predicates, and sentence connectives. These functors have to accommodate undefined arguments, but they may also produce undefined compounds even when all their arguments are fully defined. In particular, we shouldn’t ignore sentence connectives of this kind: once loosened up, classical propositional logic needs to be filled out with connectives such as *interjunction* and *transplication*. The uniformity behind all this comes from the idea of representing partial functions by monotonic functions—as explained in Section 1—and using monotonically representable partial functions to interpret functors of whatever logical category.

All sections have undergone some stylistic revision for the second edition, and most of them have been expanded. Note that Section 2 now has more subsections: there is a new introductory subsection, which means that subsections 2.1 to 2.5 have become subsections 2.2 to 2.6; and the old subsection 2.6 has split into three—2.7 to 2.9—so that the old 2.7 is now 2.10. Section 4 has been disrupted in a similar way: the old subsection 4.1 has split into 4.1 and 4.2; subsection 4.3 is new; and the old subsection 4.2 has

split into 4.4 and 4.5. There has been a more straightforward reorganization to Sections 6 and 7: a new subsection has been introduced as 6.3, which means that the old subsections 6.3 and 6.4 become 6.4 and 6.5; and the old subsection 7.2 has split into two: 7.2 and 7.3. The other Sections retain their original structure.

× × ×

**Notation for *interjunction*:**— In the first edition an interjunction sign was formed by juxtaposing two ‘×’s: ××. This was a pity, because it made the symbol a bit too flat. Interjunction is a squadging of conjunction and disjunction, and so the symbol for it should be a simultaneous occurrence of ‘∧’ and ‘∨’: ⋈. Sadly, the notation ‘××’ has found its way into the literature, and—much worse—this has sometimes become just two ‘x’s: xx. I urge anyone who wants to write an interjunction sign in the future to avoid ‘xx’ at all costs: ‘××’ is tolerable, but I recommend ‘⋈’.

## 1 A SKETCH OF SIMPLE PARTIAL LOGIC

### 1.1 *Classical Semantics as Partial Semantics*

In classical logic sentences are either true ( $\top$ ) or false ( $\perp$ ) and the interpretation of the standard sentence connectives can be given in the following way:

$$\begin{aligned} \neg\phi \text{ is } & \begin{cases} \top & \text{iff } \phi \text{ is } \perp \\ \perp & \text{iff } \phi \text{ is } \top, \end{cases} \\ \phi \wedge \psi \text{ is } & \begin{cases} \top & \text{iff } \phi \text{ is } \top \text{ and } \psi \text{ is } \top \\ \perp & \text{iff } \phi \text{ is } \perp \text{ or } \psi \text{ is } \perp, \end{cases} \\ \phi \vee \psi \text{ is } & \begin{cases} \top & \text{iff } \phi \text{ is } \top \text{ or } \psi \text{ is } \top \\ \perp & \text{iff } \phi \text{ is } \perp \text{ and } \psi \text{ is } \perp, \end{cases} \\ \phi \rightarrow \psi \text{ is } & \begin{cases} \top & \text{iff } \phi \text{ is } \perp \text{ or } \psi \text{ is } \top \\ \perp & \text{iff } \phi \text{ is } \top \text{ and } \psi \text{ is } \perp, \end{cases} \\ \phi \leftrightarrow \psi \text{ is } & \begin{cases} \top & \text{iff } (\phi \text{ is } \top \text{ and } \psi \text{ is } \top) \text{ or } (\phi \text{ is } \perp \text{ and } \psi \text{ is } \perp) \\ \perp & \text{iff } (\phi \text{ is } \top \text{ and } \psi \text{ is } \perp) \text{ or } (\phi \text{ is } \perp \text{ and } \psi \text{ is } \top). \end{cases} \end{aligned}$$

For simple partial logic we shall adopt precisely these classical  $\top/\perp$  conditions; only we give up the assumption that all sentences have to be classified either as  $\top$  or as  $\perp$ . This leaves room for the classification *neither- $\top$ -nor- $\perp$* . At present we are concerned merely to highlight a parallel with classical semantics, and under the parallel we can think of the third classification as a ‘truth-value gap’. This thought is taken a little further in Sections 1.2

and 3. But the point, if any, of seeing the third classification as different in philosophical kind from  $\top$  and  $\perp$  will of course depend on what particular motivation we consider for adopting the forms of partial logic. (See, especially, Sections 2 and 5.)

To interpret universal and existential quantifiers over a given domain  $D$ , we shall again exploit the fact that the classical interpretation leaves room for a gap between  $\top$  and  $\perp$  when we write out  $\top$ -conditions and  $\perp$ -conditions separately. Assuming that a language has—or can be extended so as to have—a name  $\bar{a}$  for each object  $a$  in  $D$ ,

$$\begin{aligned} \forall x\phi(x) \text{ is } & \begin{cases} \top & \text{iff } \phi(\bar{a}) \text{ is } \top \text{ for every } a \text{ in } D \\ \perp & \text{iff } \phi(\bar{a}) \text{ is } \perp \text{ for some } a \text{ in } D, \end{cases} \\ \exists x\phi(x) \text{ is } & \begin{cases} \top & \text{iff } \phi(\bar{a}) \text{ is } \top \text{ for some } a \text{ in } D \\ \perp & \text{iff } \phi(\bar{a}) \text{ is } \perp \text{ for every } a \text{ in } D. \end{cases} \end{aligned}$$

Most treatments of classical logic stipulate that the domain be non-empty. We shall not be so restrictive:  $D$  may be empty.

These  $\top/\perp$ -conditions for  $\forall x$  and  $\exists x$  of course presuppose a semantic account of predicate/singular-term composition. And this mode of composition deserves some attention, since it is the most familiar place to locate the cause of a sentence's being neither 'true' nor 'false'. It has been considered to give rise to a truth-value gap in two different ways: either (i) because a term  $t$  may lack a denotation and may, for this reason, make a sentence  $\phi(t)$  neither true nor false; or (ii) because a predicate  $\phi(x)$  may be only 'partially defined'—not either true or false of some object or objects—so that, if  $t$  denoted such an object,  $\phi(t)$  would be neither true nor false. We shall want to accommodate both these ideas in one uniform account of predicate/singular-term composition. Our approach will be sketched in Section 1.2, along with an approach to functors which form singular terms from singular terms.

But there is one particular atomic predicate to consider immediately: the identity predicate. Once again we can adopt classical  $\top$ -conditions and  $\perp$ -conditions *verbatim* for a sentence  $t_1 = t_2$ :

$$t_1 = t_2 \text{ is } \begin{cases} \top & \text{iff } t_1 \text{ and } t_2 \text{ denote the same thing} \\ \perp & \text{iff } t_1 \text{ and } t_2 \text{ denote different things.} \end{cases}$$

This means that if either  $t_1$  or  $t_2$  is non-denoting, then  $t_1 = t_2$  is neither  $\top$  nor  $\perp$ . Identity is an untypically straightforward case. At least, so it is if we restrict attention to a determinate relation over a discrete domain of objects—as we shall.

× × ×

Whatever general framework we set up for predicate/singular-term composition, our logic has so far been revealed as 'partial' only in the weak sense

that it accommodates value-gaps that might arise from the interpretation of non-logical terms or predicates. This is because the interpretation of classical logical vocabulary is classical. But there is a stronger sense of ‘partial logic’: a logic will be partial in the stronger sense if it provides the resources for explaining why a sentence may be neither  $\top$  nor  $\perp$  in terms of logical vocabulary—vocabulary, that is, with a fixed meaning in the logic. We should look for modes of logical composition whose interpretation can give rise to truth-value gaps, even when any classical sentence constructed out of the same non-logical vocabulary (with the same interpretation) would have to be either  $\top$  or  $\perp$ .

Assuming that we have worked out the general account of how non-denoting terms can give rise to truth-value gaps, a term-forming descriptions operator would be an example of gap-introducing logical vocabulary. This is because a term  $\iota x\phi(x)$  may turn out not to denote, even when  $\phi(x)$  is totally defined. Assuming that  $\phi(x)$  is in fact totally defined, then the denotation conditions for  $\iota x\phi(x)$  must be that if  $a$  is an object in the domain, then:

$$\iota x\phi(x) \text{ denotes } a \quad \text{iff} \quad \forall x[x = \bar{a} \leftrightarrow \phi(x)] \text{ is } \top,$$

where, as before,  $\bar{a}$  is a name—pre-existing or specially introduced—for  $a$ . In other words,  $\iota x\phi(x)$  denotes an object if and only if that object uniquely satisfies  $\phi(x)$  and is non-denoting if there is no such object. Of course, we also have to consider the case where  $\phi(x)$  is not totally defined, but the denotation conditions stated will continue to make sense. Furthermore, given the general constraint to emerge in Section 1.2, they will turn out to be the only possible ones for a determinate relation of identity over a discrete domain of objects (see Section 6.4).

These  $\iota$ -terms involve a rather complicated route to neither- $\top$ -nor- $\perp$  sentences. There is a much more straightforward, and no less interesting, kind of gap-introducing vocabulary: sentence connectives. Consider the following  $\top/\perp$ -conditions for the connectives  $\&$  and  $/$ , the first of which we shall call *interjunction* and the second *transplication*:

$$\begin{aligned} \phi \& \psi \text{ is } & \begin{cases} \top & \text{iff } \phi \text{ is } \top \text{ and } \psi \text{ is } \top \\ \perp & \text{iff } \phi \text{ is } \perp \text{ and } \psi \text{ is } \perp, \end{cases} \\ \phi / \psi \text{ is } & \begin{cases} \top & \text{iff } \phi \text{ is } \top \text{ and } \psi \text{ is } \top \\ \perp & \text{iff } \phi \text{ is } \top \text{ and } \psi \text{ is } \perp. \end{cases} \end{aligned}$$

Notice that  $\&$  has the  $\top$ -conditions of  $\wedge$  and the  $\perp$ -conditions of  $\vee$ , while  $/$  has the  $\top$ -conditions of  $\wedge$  but the  $\perp$ -conditions of  $\rightarrow$ . And so these connectives clearly meet our desideratum of introducing value gaps: we do not necessarily have to look to predicate/singular-term composition to find a logical explanation why a sentence may be neither  $\top$  nor  $\perp$ . The particular usefulness of  $\&$  and  $/$  will be touched upon in Section 2.2 and several later sections.

Among our logical vocabulary we shall also include a constantly true sentence  $\top$ , and a constantly false one  $\perp$ . Thus we are using ‘ $\top$ ’ and ‘ $\perp$ ’ both as truth-value labels and to stand for logical constants; and, in a similar way, we shall use ‘ $*$ ’ both to label the classification ‘neither- $\top$ -nor- $\perp$ ’ and to stand for a sentence which is logically neither  $\top$  nor  $\perp$ . There will also be a logically non-denoting singular-term, denoted by ‘ $\otimes$ ’—which will be used also to denote the classification ‘non-denoting’. In the presence of the term  $\otimes$ , we shall then be able to abandon  $\neg$ -terms without any loss in expressive power: this is explained in Section 6.4.

x x x

Finally, we must consider the relation of (logical) consequence. Our semantic definition of ‘ $\psi$  is a consequence of  $\phi$ ’ is, loosely stated, that

- (i) whenever  $\phi$  is  $\top$ ,  $\psi$  is  $\top$ , and (ii) whenever  $\psi$  is  $\perp$ ,  $\phi$  is  $\perp$ .

And so, yet again, we are using a definition which conjoins two formulations of the classical definition, one involving  $\top$  and the other  $\perp$ —formulations which are equivalent in total logic, but not in partial logic. To illustrate the idea, consider for the moment just a propositional calculus with formulae built up from atomic sentences using the connectives we have introduced. Then ‘interpretations’ will simply be partial assignments of  $\top$  and  $\perp$  to atomic sentences, and formulae may be evaluated according to our  $\top/\perp$ -clauses for the connectives. We shall use ‘ $\vDash$ ’ for the relation of logical consequence, and so  $\phi \vDash \psi$  if only if (i) and (ii) above both hold when ‘whenever’ is understood to mean ‘under any partial assignment under which’. (By ‘partial assignment’ I do not mean to exclude total assignments: here, as elsewhere, ‘partial’ means ‘not necessarily total’.)

The tendency among authors on partial logics of one sort or another is to take condition (i) on its own to define logical consequence; and sometimes (i) and (ii) are used to frame two separate notions—for example, in [Dunn 1975], [Hayes 1975] and, in disguised form, in [Woodruff 1970]. In [Cleave 1974], on the other hand, there is a (rather algebraic) version of our double-barrelled definition. And across the literature of the last twenty years the picture has not greatly changed. But perhaps making a choice between these alternatives is not such a fundamental matter. After all, we can define the two halves of our single notion:

$$\begin{aligned} \phi \vDash^{\top} \psi & \text{ iff } \phi \vDash * \vee \psi, \\ \phi \vDash^{\perp} \psi & \text{ iff } \phi \wedge * \vDash \psi. \end{aligned}$$

And, putting them back together again,

$$\phi \vDash \psi \text{ iff } \phi \vDash^{\top} \psi \text{ and } \phi \vDash^{\perp} \psi.$$

Or, if we invoke negation, either one of the halves on its own would do:

$$\phi \vDash \psi \text{ iff } \phi \vDash^{\top} \psi \text{ and } \neg \psi \vDash^{\top} \neg \phi \text{ iff } \phi \vDash^{\perp} \psi \text{ and } \neg \psi \vDash^{\perp} \neg \phi.$$

The issue might be set in a more interesting context if thought were given to the connection between these definitions and inferential practice; but this question goes far beyond our semantics-orientated essay.

To motivate working with the double-barrelled definition we can adduce some arguments from theoretical neatness. First, the law of contraposition holds:

$$\phi \vDash \psi \quad \text{iff} \quad \neg\psi \vDash \neg\phi.$$

Secondly, logical equivalence—a relation which must be taken to obtain between two formulae if and only if they take the same resultant classification under any interpretation—turns out as mutual consequence. Using ‘ $\simeq$ ’ for equivalence,

$$\phi \simeq \psi \quad \text{iff} \quad \phi \vDash \psi \text{ and } \psi \vDash \phi.$$

Thirdly, equivalence and consequence fit together with conjunction and disjunction in the natural (at least the classical) way:

$$\phi \simeq \psi \wedge \phi \quad \text{iff} \quad \psi \simeq \phi \vee \psi \quad \text{iff} \quad \phi \vDash \psi.$$

These properties of  $\vDash$  break down for  $\vDash^\top$  and for  $\vDash^\perp$ .

Neatness aside, some interesting differences between working with  $\vDash$  and working just with  $\vDash^\top$  (equally just with  $\vDash^\perp$ ) can be extracted from [Langholm 1988]. In particular, it emerges that in a first-order logic without non-denoting terms some interpolation results for  $\vDash^\top$  are much cheaper than corresponding results for  $\vDash$ . (On interpolation for  $\vDash$  in a full first-order language, see Sections 6.5, 7.2, and 7.3.)

In Section 6.5 we shall present a rigorous definition of (double-barrelled) consequence for first order languages, and there will be two generalisations. First, we shall be interested not *merely* in logical consequence, but in relations of consequence determined by a given range of interpretations—to match a proof theoretical notion of consequence in a given theory (presented in Section 7.1). Secondly, consequence will be defined between sets of formulae, rather than individual formulae: not only will several premises be allowed, but also several ‘conclusions’—to be understood disjunctively. This will match our sequent-style proof theory; and another advantage of the double-barrelled definition will then emerge: we shall be able to frame fewer and simpler rules, since sequent principles will be able to constrain the  $\top$ -conditions and  $\perp$ -conditions of logical vocabulary at one go.

There is, finally, a different kind of generalization to consider: more-than-two-place ‘consequence’ relations. For example,  $\vDash^\top$  and  $\vDash^\perp$  are combined into a four-place relation in [Langholm 1989, Fenstad 1997, Bochman 1998]. If, for simplicity’s sake, we restrict attention to single formulae rather than sets of formulae, then the relation—call it  $C$ —can be defined as follows:  $C(\phi_1, \psi_1, \phi_2, \psi_2)$  if and only if whenever  $\phi_1$  is  $\top$  and  $\psi_2$  is  $\perp$ , then either

$\psi_1$  is  $\top$  or  $\phi_2$  is  $\perp$ . Notice that we could define  $C$ , using negation, in terms of either  $\vDash^\top$  or  $\vDash^\perp$ :

$$C(\phi_1, \psi_1, \phi_2, \psi_2) \text{ iff } \phi_1 \wedge \neg\psi_2 \vDash^\top \psi_1 \vee \neg\phi_2 \text{ iff } \neg\psi_1 \wedge \phi_2 \vDash^\perp \neg\phi_1 \vee \psi_2.$$

Alternatively—and I have myself found this more useful to work with—we could adopt a four-place relation  $C'$  that just conditionalizes the two place  $\vDash$ :  $C'(\phi_1, \psi_1, \phi_2, \psi_2)$  if and only if whenever  $\phi_1$  is  $\top$  and  $\psi_1$  is  $\perp$ , then  $\phi_2 \vDash \psi_2$ . In terms of  $\vDash$  this relation could be defined as follows:

$$C'(\phi_1, \psi_1, \phi_2, \psi_2) \text{ iff } \phi_1 \wedge \neg\psi_1 \wedge \phi_2 \vDash \psi_2 \vee \neg\phi_1 \vee \psi_1.$$

In Section 7.1 we shall use the the proof-theoretical correlate of  $\vDash$  to define a three-place consequence relation along these lines—one that ignores the  $\psi_1$  argument place. Some of the quantifier and identity rules are most perspicuously presented in terms of this relation. (Compare the three- and four-place relations used for systems of modal logic in [Blamey and Humberstone 1991].)

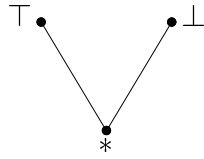
### 1.2 Partial Semantics as Monotonic Semantics

To interpret sentence connectives we have specified  $\top$ -conditions and  $\perp$ -conditions for formulae constructed by means of them:  $*$ -conditions then take care of themselves. Even so,  $*$  is a semantic classification, and the apparatus of 3-valued logic is at our disposal: our  $\top/\perp$ -conditions are summed up in the following matrices. (The constant sentences  $\top$ ,  $*$  and  $\perp$  can be thought of as 0-place connectives, but their matrices are trivial).

$\phi$	$\neg\phi$	$\phi$	$\psi$	$\phi \wedge \psi$	$\phi \vee \psi$	$\phi \times \psi$	$\phi \leftrightarrow \psi$	$\phi \rightarrow \psi$	$\phi / \psi$
$\top$	$\perp$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$	$\top$
$*$	$*$	$\top$	$*$	$*$	$\top$	$*$	$*$	$*$	$*$
$\perp$	$\top$	$\top$	$\perp$	$\perp$	$\top$	$*$	$\perp$	$\perp$	$\perp$
		$*$	$\top$	$*$	$\top$	$*$	$*$	$\top$	$*$
		$*$	$*$	$*$	$*$	$*$	$*$	$*$	$*$
		$*$	$\perp$	$\perp$	$*$	$*$	$*$	$*$	$*$
		$\perp$	$\top$	$\perp$	$\top$	$*$	$\perp$	$\top$	$*$
		$\perp$	$*$	$\perp$	$*$	$*$	$*$	$\top$	$*$
		$\perp$	$\perp$	$\perp$	$\perp$	$\perp$	$\top$	$\top$	$*$

Partial assignments of  $\top$  or  $\perp$  to atomic constituents can now be replaced by total assignments of  $\top$ ,  $*$  or  $\perp$ . And, if we take it that each assignment assigns a classification to all of a denumerable stock of atomic formulae, then everything will fit neatly into place when we just assign  $*$  to any vocabulary we are not interested in.

Let us now impose a simple ordering  $\sqsubseteq$  on  $\{\top, *, \perp\}$ :



$$x \sqsubseteq y \quad \text{iff} \quad \text{either } x = * \text{ or } x = y.$$

Equivalently:  $x \sqsubseteq y$  iff both, if  $x = \top$ , then  $y = \top$ , and, if  $x = \perp$ , then  $y = \perp$ . Then we can extend the use of ' $\sqsubseteq$ ' to define a 'degree-of-definedness' relation between assignments  $v$  and  $w$ :

$$v \sqsubseteq w \quad \text{iff} \quad v(p) \sqsubseteq w(p) \text{ for every atomic formula } p.$$

In other words,  $v \sqsubseteq w$  if and only if wherever  $v$  assigns the value  $\top$  or  $\perp$ ,  $w$  assigns that value also. If  $v(\phi)$  is the result of evaluating a formula  $\phi$  under  $v$ , it is then easy to deduce the following *monotonicity of evaluation*:

$$\text{if } v \sqsubseteq w, \text{ then } v(\phi) \sqsubseteq w(\phi), \text{ for every formula } \phi.$$

An intuitive way to think about this is that if a formula has taken on a value ( $\top$  or  $\perp$ ), then this value persists when any atomic gaps ( $*$ ) are filled in by a value ( $\top$  or  $\perp$ ) (cf. Lemma 3 in section 6.2).

Here we have a global monotonicity condition, but we might direct attention to individual formulae. If all atomic formulae occurring in  $\phi$  are among  $p_1, \dots, p_n$ , then we can specify a  $3^n$ -row matrix for  $\phi$ , which describes a function  $f$  from  $\{\top, *, \perp\}^n$  into  $\{\top, *, \perp\}$ , where  $f(x_1, \dots, x_n)$  is the classification of  $\phi$  under the assignment of  $x_i$  to  $p_i$ ,  $1 \leq i \leq n$ . And  $f$  will then be a monotonic function. That is to say

$$\text{if } x_i \sqsubseteq y_i \text{ for all } i, \text{ then } f(x_1, \dots, x_n) \sqsubseteq f(y_1, \dots, y_n).$$

Observe that this is equivalent to monotonicity in each coordinate separately.

What lies behind both forms of monotonicity is that the matrix for each sentence connective describes a monotonic function and that the class of monotonic functions is closed under composition. The question then arises: Is our logic expressively adequate for all monotonic functions? It is. In Section 4.1 we shall show that  $\neg$ ,  $\wedge$ ,  $\vee$ ,  $\boxtimes$ ,  $\top$ , and  $\perp$  form a neatly complete bunch of connectives.

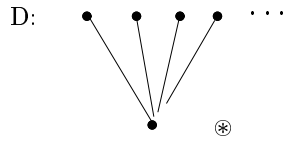
Our 'partial' propositional logic could, then, simply be seen as the total logic of 3-valued monotonic modes of sentence composition—modes  $\phi(p_1, \dots, p_n)$  that are interpreted by monotonic functions. The connection between the two ways of looking at it is made by the idea that monotonic functions from  $\{\top, *, \perp\}^n$  into  $\{\top, *, \perp\}$  can be taken to represent partial functions from  $\{\top, \perp\}^n$  into  $\{\top, \perp\}$ . Modes of composition in the logic



can then be taken to be interpreted by partial functions. On this understanding of the mathematical semantics,  $\top$  and  $\perp$  are obviously the only ‘truth values’ that there are:  $*$  plays a role merely in the representation of partial functions by monotonic total ones. Thus the idea that a sentence classified  $*$  suffers from a ‘truth-value gap’ is given immediate, but in itself uninteresting, sense.

× × ×

The use of monotonic functions to represent partial ones has nothing specifically to do with truth functions. Given any domain  $D$ , we can pick on an extraneous object  $\otimes$  and consider functions from  $(D \cup \{\otimes\})^n$  into  $D \cup \{\otimes\}$  which are monotonic—in exactly the same sense as before—with respect to an order relation  $\sqsubseteq$  given by:



$$x \sqsubseteq y \quad \text{iff} \quad \text{either } x = \otimes \text{ or } x = y.$$

Equivalently:  $x \sqsubseteq y$  if and only if, for any  $a \in D$ , if  $x = a$  then  $y = a$ . These functions can be taken to represent partial functions from  $D^n$  into  $D$ . And we can just as easily consider a range of different domains  $D_1, \dots, D_{n+1}$ , each fixed up with their own extraneous objects  $\otimes_1, \dots, \otimes_{n+1}$ , and represent a system of partial functions from  $D_1 \times \dots \times D_n$  into  $D_{n+1}$  by functions from  $(D_1 \cup \{\otimes_1\}) \times \dots \times (D_n \cup \{\otimes_n\})$  into  $D_{n+1} \cup \{\otimes_{n+1}\}$  which are monotonic with respect to the respective orderings. A simple example would be the system of partial  $n$ -place relations on a domain  $D$ , represented by monotonic functions from  $(D \cup \{\otimes\})^n$  into  $\{\top, *, \perp\}$ . If  $n = 1$ , these would be ‘partial subsets’ of  $D$ .

The functions represented are partial not only in that they may be undefined for some  $n$ -tuple of arguments, but also in that they allow for ‘empty argument places’:  $*$  and  $\otimes$  stand equally for the gap of an empty argument place and for the gap of no output value. This suggests that these partial functions might aptly be deployed to provide the uniform account of linguistic composition that we demanded in Section 1.1—to handle partially defined functors that may embrace non-denoting terms.

But what kind of sense does it make to say that monotonic functions represent partial ones? The notion of representation is itself unproblematic: it is just the same as when we say that ordinary total functions can be represented in set theory by sets of a certain kind. Still, when it is observed that an ‘empty argument place’ does not necessarily mean no output value

(consider for example the matrices for  $\wedge$  and  $\vee$ ), it may be objected that it is nonsense to talk of a *function* which can yield an output value from an incomplete, possibly total vacuous, array of input values. This thought, only thinly veiled in talk about functors, seems to have figured in some discussions of Frege, and we shall tackle it in this context in Section 3.2. A different—and opposite—reaction would be to question all the fuss about monotonicity: granted the idea of  $*$  and  $\otimes$  representing gaps in both input and output, why restrict the range of representing functions at all? In Section 2 we shall see how some specific applications for partial functions in semantics call for the monotonicity constraint, and a more general view will emerge when we discuss the first reaction. For the moment we can put the point intuitively: the output value, if any, of a monotonically representable partial function can be seen to depend, and depend only, on the input values in occupied argument places (and not on the gaps of empty ones), precisely because of the monotonicity condition that if a gap is ‘filled in’, then the output value remains fixed.

The degree-of-definedness ordering  $\sqsubseteq$  becomes more interesting than merely a gap versus an object when we push the idea of representing partial functions up to higher-level categories—to functions with systems of partial functions as their domain (and possibly also as their range). Consider, the simple example of the system of partial subsets of a domain  $D$ , represented by monotonic functions from  $D \cup \{\otimes\}$  into  $\{\top, *, \perp\}$ . Between two such functions  $f$  and  $g$  we can define  $f \sqsubseteq g$  to mean that  $f(x) \sqsubseteq g(x)$  for any  $x$  in  $D \cup \{\otimes\}$ . Then, to represent partial subsets of the system of partial subsets of  $D$ , we can use functions on the monotonic functions—functions  $F$  into  $\{\top, *, \perp\}$  which are themselves monotonic:

$$\text{if } f \sqsubseteq g, \text{ then } F(f) \sqsubseteq F(g).$$

Intuitively, the point of this higher-level monotonicity is that if  $F$  yields a value when applied to  $f$ , then this depends, and depends only, on the range of output values of  $f$ , not on its gaps. This means that if  $g$  behaves like  $f$  except possibly that it is more defined, then  $F$  must send  $g$  to the same value it sends  $f$  to.

A full hierarchy will emerge for higher-level categories of monotonically-representable partial functions, and a non-trivial study of its characteristics can be found in [Lepage 1992]. In [Muskens 1989] and in [Lapierre 1992], on the other hand, there are special hierarchies designed to interpret intensional partial logic. Muskens has a cunning reduction of functional application and abstraction to operations on partial relations, which are what his hierarchy is actually a hierarchy of. But Lepage and Lapierre adopt a more familiar style of reduction: they take hierarchies of just one-place functions as primitive. Nothing is lost, because a domain of partial functions from  $D_1 \times \dots \times D_n$  into  $D_{n+1}$  is isomorphic to, and can be modelled by, the domain of partial functions from  $D_1$  into the domain of partial functions from

$D_2$  into ... into the domain of partial functions from  $D_n$  into  $D_{n+1}$ . And so, in particular, if  $D_{n+1} = \{\top, \perp\}$ , then we have a modelling of partial  $n$ -place relations. In [Tichý 1982] it had been argued that such a reduction to one-place functions was possible only with domains of total functions, but Lepage exposes the error in Tichý's argument.

\* \* \*

To provide a semantics for first-order languages we need neither go very far up the hierarchy nor reduce all functions to one-place ones. Predicates will be interpreted by monotonically-representable partial sets and relations over a domain  $D$ . Similarly,  $n$ -place functors which form singular terms out of singular terms will be interpreted by monotonically representable partial functions from  $D^n$  into  $D$ . And in a model theory, conceived of as a theory developed in some standard set theory, we can expect to work with the representing monotonic functions. A model will directly assign such a function to unstructured predicate symbols and term-functor symbols, but we are no less interested in the complex predicates that arise as formulae  $\phi(x_1, \dots, x_n)$ , with free variables  $x_1, \dots, x_n$  signaling the argument places, and in the complex term-functors that arise as compound terms  $t(x_1, \dots, x_n)$ . If we take free variables to range over  $D \cup \{\otimes\}$ , are we guaranteed that these complex modes will be monotonic? We are, given that every unstructured functor—logical and non-logical alike—is interpreted *via* a monotonic function of the appropriate category, since combining monotonic functions invariably leads to a monotonic function. Straightforward functional composition lies behind all linguistic combinations except for the variable-binding quantifiers  $\forall$  and  $\exists$  (and also the variable-binding operator  $\iota$ , if we include it: see Section 6.4).

In the simplest case quantifiers are just second-level predicates, taking a one-place predicate  $\phi(x)$  to a sentence  $\forall x\phi(x)$  or  $\exists x\phi(x)$ . Disentangling them from the apparatus of variable-binding, it is easy to see that the  $\top/\perp$ -conditions we gave for  $\forall$  and  $\exists$  match an interpretation *via* monotonic second-level functions  $F_\forall$  and  $F_\exists$  on the domain of partial subsets of  $D$ :

$$F_\forall(f) = \begin{cases} \top & \text{iff } f(a) = \top \text{ for every } a \text{ in } D \\ \perp & \text{iff } f(a) = \perp \text{ for some } a \text{ in } D, \end{cases}$$

$$F_\exists(f) = \begin{cases} \top & \text{iff } f(a) = \top \text{ for some } a \text{ in } D \\ \perp & \text{iff } f(a) = \perp \text{ for every } a \text{ in } D. \end{cases}$$

But quantifiers play a general role in converting any  $(n+1)$ -place predicate  $\phi(x_1, \dots, x_i, \dots, x_{n+1})$ , into an  $n$ -place predicate  $\forall x_i\phi(x_1, \dots, x_i, \dots, x_{n+1})$  or  $\exists x_i\phi(x_1, \dots, x_i, \dots, x_{n+1})$ , and we have to check that monotonicity will always be preserved in this move. This is easy enough. Notice that variables bound by a quantifier will 'range over' just the domain of objects  $D$ —not, as free variables do, over the whole of  $D \cup \{\otimes\}$ .

Finally, what about the interpretation of singular terms?—‘closed’ terms, which contain no free variables? To fit in with the model-theoretic apparatus for functors, we should expect to be able to assign an object in the domain of quantification to a term to mean that the term denotes that object, and to assign  $\otimes$  to a non-denoting term. If we stipulate the classification of all unstructured singular terms in this way, the apparatus of monotonic functions will then yield an appropriate classification for compound closed terms.

The reader who is eager for formal details could now skip on to Section 6. But a few further remarks are prompted, if we want seriously to understand a term’s denoting an element of  $D$  in a way that matches the informal idea of a term’s standing for an object. A sharp contrast must be drawn with the assignment of  $\otimes$  to a term. For  $\otimes$  is not the nonsense of an object which doesn’t exist; nor is it a special object picked on (Frege-style) to be the actual denotation for terms that should really be non-denoting:  $\otimes$  has been introduced simply as part of the apparatus for representing partial functions. It does then make sense to see  $\otimes$  playing a derived model theoretic role as the *semantic classification* ‘non-denoting’, but it would be courting confusion if we then went on to think of the monotonic functions of the model theory just as functions on semantic classifications. The classification of a denoting term would then turn out to be the very object denoted, but to keep semantic levels straight, we should distinguish the object  $a$  that a term denotes from the classification ‘denoting- $a$ ’: such a classification is not an object in the domain and can be aligned with  $\otimes$ . Of course, objects and the corresponding classifications do correspond one-to-one, and so it is in fact open to us to adopt an alternative understanding of the semantics right from the start—as a semantics that operates throughout on classifications. And this could either be thought of as a total monotonic semantics on all classifications or as a partial semantics on the range of classifications ‘denoting-so-and-so’ (see Section 3).

Observe that a parallel finickiness over sentences and  $*$  would be called for only if the assignment of  $\top$  or  $\perp$  to a sentence were intended to be more than a model-theoretic device for classifying sentences—as it would, for example, according to Frege’s unified theory of reference, where the truth-values  $\top$  and  $\perp$  are seriously thought of as objects denoted by sentences. Otherwise, it is harmless to take the monotonic functions that represent partial ones simply as (total) functions on semantic classifications.

### 1.3 Comparisons with Supervaluations

The preceding remarks bring our partial logic very much in line with traditional truth-table approaches. The most notable difference is simply in the choice of connectives. We have the novelty of gap-introducing modes, such as interjunction, but we have not introduced any of the familiar gap-closing

vocabulary, which tends to have a metalinguistic flavour. There is no ‘it is true that...’ connective, for example, which is often introduced to turn gappy sentences into false ones. Nor can we define such a mode: it would not be monotonic. In Section 4 we take up the theme of non-classical vocabulary, but now we contrast simple partial logic with an altogether more sophisticated approach, viz. supervaluations. See [Van Fraassen 1966].

To illustrate the basic—but by no means the theoretically most general—idea, consider the question of evaluating a classical propositional formula under a given partial assignment of the truth values  $\top$  and  $\perp$  to atomic constituents. First we are to evaluate the formula in the ordinary classical way, under all total assignments which extend the partial assignment. Then the formula is taken to be  $\top$  if all these total assignments make it  $\top$ ;  $\perp$ , if they all make it  $\perp$ ; and  $*$  otherwise. In other words, using the definitions we have already introduced, the supervaluational evaluation  $v^s(\phi)$  of a formula  $\phi$  can be given by:

$$v^s(\phi) = \begin{cases} \top & \text{iff } w(\phi) = \top \text{ for all total } w \text{ such that } v \sqsubseteq w \\ \perp & \text{iff } w(\phi) = \perp \text{ for all total } w \text{ such that } v \sqsubseteq w. \end{cases}$$

It is easy to see that this scheme of evaluation yields global monotonicity of evaluation, just as well as simple partial logic (see Section 2.5):

$$\text{if } v \sqsubseteq w, \text{ then } v^s(\phi) \sqsubseteq w^s(\phi), \text{ for every (classical) formula } \phi.$$

However, since the basic evaluation of formulae is just classical, the idea of using monotonic functions to give the interpretation of sentence modes has no role to play. In simple partial logic the monotonicity of a mode  $\phi(p_1, \dots, p_n)$  can be stated in terms of a substitutivity condition: given any particular assignment  $v$ , and any formulae  $\psi_1, \dots, \psi_n, \chi_1, \dots, \chi_n$ ,

$$\text{if } v(\psi_i) \sqsubseteq v(\chi_i) \text{ for all } i, \text{ then } v(\phi(\psi_1, \dots, \psi_n)) \sqsubseteq v(\phi(\chi_1, \dots, \chi_n)).$$

But clearly there is nothing parallel for the supervaluational scheme. Say, for example, that  $v^s(p) = v(p) = *$  and  $v^s(q) = v(q) = \top$ , then  $v^s(p \vee \neg p) = \top$  but  $v^s(p \vee \neg q) = *$ .

This example points up in a particularly startling way the ‘intensional’ character of supervaluational semantics, which is a departure from the spirit of classical logic. It is, however, a price that supervaluation theorists are willing to pay in order to preserve what is considered to be a more important feature of classical logic, viz. the stock of classical tautologies. More exactly, it is considered important to be able to capture the ‘logical truths’ of classical logic—formulae true under any total assignment—as ‘logical truths’ of partial logic—formulae true under any partial assignment. The supervaluational scheme makes this work, because, if  $\phi$  is a classical formula, then  $\phi$  is a classical tautology if and only if  $v^s(\phi) = \top$  for any partial assignment  $v$ .

This contrasts markedly with our naive scheme of evaluation: logical truths of *any* kind are very thin on the ground. Indeed, only formulae containing some occurrence of one of the constant sentences  $\top$  or  $\perp$  can ever be true under all partial assignments.

But why should we be interested in logical truth? In [Thomason 1972, p. 231], where the author is arguing in favour of supervaluational techniques in spite of their intensionality, the suggestion seems to be that the truths of a logic are supposed to capture distinctions between good and bad reasoning. But why so? Can we not leave it to the laws of logical consequence—or perhaps to a more encompassing theory of logical relationships between formulae—to capture cannons of correct reasoning? Then we might still be in a good position to show that classical tautologies are indeed ‘preserved’ in partial logic. Consider, for example, the relation  $\vDash$  which we defined in Section 1.1 (or  $\vDash^\top$  would serve equally well). It is easy to check that, assuming  $\phi$  is a classical formula,  $\phi$  is a classical tautology if and only if

$$[p_1 \vee \neg p_1] \wedge \dots \wedge [p_n \vee \neg p_n] \vDash \phi,$$

where  $p_1, \dots, p_n$  are the atomic constituents of  $\phi$ . Does this not set classical tautologies in exactly their rightful place? The formula to the left of ‘ $\vDash$ ’ could never be  $\perp$ , but it is not trivially  $\top$ , as it would be under the supervaluational scheme: it is  $\top$  precisely when all the  $p_i$  are either  $\top$  or  $\perp$ .

Observe that it would be vain to expect the logic of monotonic matrices to capture even its own relation of logical consequence in terms of truth: there can be no mode of composition  $\phi(p, q)$  such that  $\psi \vDash \chi$  if and only if  $\phi(\psi, \chi)$  is logically true. For if there were, then  $\phi(*, *)$  would be  $\top$ , but  $\phi(\top, \perp)$  would not be, which violates monotonicity. And this has nothing specifically to do with our double-barrelled definition of  $\vDash$ : it is exactly the same with either  $\vDash^\top$  or  $\vDash^\perp$ . If we wanted to introduce some special conditional connective to play the role of  $\phi(\cdot, \cdot)$ , then either it would have to have a non-monotonic matrix (see Section 4.4), or else it would lead to an intensional semantics of the kind we discuss in Section 2.7. However, the exercise we have set ourselves is to use the framework of consequence to set up logic without any such connective.

It would be a mistake to suppose that the theory of supervaluations is not actually concerned with logical relations. On the contrary, there is much sophisticated work involved with comparing and contrasting relations of ‘implication’, ‘necessitation’, ‘presupposition’, etc., etc.—for example in [Van Fraassen 1967, Van Fraassen 1971]. But here the theory quickly becomes rather abstract and we lose sight of any particular formal language. In contrast, simple partial logic puts emphasis on a particular logical vocabulary, and this includes gap-introducing connectives such as interjunction and transplication. These connectives actually prove something of a nuisance to the supervaluational idea: the definition we gave

for  $v^s(\psi)$  continues to make sense when  $\&$  and  $/$  are allowed to occur in  $\psi$ , but the point of the exercise is rather spoilt, since there will be formulae of the overall form of classical tautologies which do not come out true. For example, if  $\phi$  is  $p \& \neg p$ , then there can be no  $v$ —not even a  $v$  which is already total—such that  $v^s(\phi \vee \neg\phi) = \top$ . In the face of this problem various supervaluational manoeuvres might be prompted: consider, for example, [Van Fraassen 1975], where Belnap’s connective of ‘conditional assertion’ (see Sections 2.3 and 4.5) is supervaluationalized.

The supervaluational evaluation of a formula  $\phi$  under an assignment  $v$  is a boosting-up of its simple evaluation, in that  $v(\phi) \sqsubseteq v^s(\phi)$ . The question then arises what other kinds of boost-up evaluation may be defined—in particular, what kinds  $k$  such that  $v(\phi) \sqsubseteq v^k(\phi) \sqsubseteq v^s(\phi)$ —and [Langholm 1988] experiments with various definitions. So long as we remain with propositional logic, these in fact turn out to yield the same result as supervaluational semantics, but corresponding definitions of the evaluation of first-order formulae in partial relational structures give rise to non-trivial differences. Aside from any intrinsic interest in varying the definition of evaluation, this proves to be a useful model-theoretic technique for investigating extensions of a classical language. However, Langholm’s partial relational structures do not capture the full semantics of monotonically-representable partial functions. And, as far as I know, it remains uninvestigated how his work fits in with the model theory we introduce in Section 6 and use in Section 7.

## 2 SOME MOTIVATIONS AND APPLICATIONS

### 2.1 *Varieties of Partiality*

In classical logic a sentence, or the assertion of a sentence in a particular context, is classified as either TRUE or FALSE: the classification is an assessment of propositional content against how things are—or maybe against a possible way for things to be. And the propositional content is fixed as what it is precisely by conditions for its assessment. Specifying such conditions is then a way of specifying meaning for a sentence, due account being taken, in one way or another, of contextual parameters. This, roughly, is the picture that standardly goes along with classical logic. What about partial logic?

Different concerns prompt different partial-logic pictures: these are not necessarily intended to supplant the classical picture, but may offer a modification of a part of it, or may simply offer something to complement it or to flesh it out in some way. Among the variety of motivations for adopting partial logic, some will wear on their sleeves a picture they fit, but others leave it a contentious matter what picture to fit them into. As an introduction to this variety, I want to draw two rough and ready distinctions to

be discerned between different accounts of the point of classifying sentences as  $\top$  ('true'), or  $\perp$  ('false'), or neither- $\top$ -nor- $\perp$ , rather than just TRUE or FALSE.

First, let us distinguish between a one-tier and a two-tier framework for assessment. The one-tier framework is something like this:

- (1) The classification 'neither- $\top$ -nor- $\perp$ ' is, like  $\top$  and  $\perp$ , *a way of assessing content expressed* in (the assertion of) a sentence (in a context)—a way of assessing it against how things actually are, or against a possible way for things to be.

This framework lends itself to a straightforward scheme of meaning-specification: a specification of content-fixing conditions for assessment as either  $\top$ , or  $\perp$ , or neither- $\top$ -nor- $\perp$ , will be a specification of meaning. But it leaves open how, as an assessment of content, to understand what 'neither- $\top$ -nor- $\perp$ ' means. In what sense, if any, is this a 'gap' rather than just a third truth value? How do the three classifications  $\top$ ,  $\perp$ , and neither- $\top$ -nor- $\perp$  mesh with the two classical truth values, if they mesh at all?—in other words, how, if at all, does content fixed by classification in partial logic mesh with classical propositional content?

The two-tier framework, on the other hand, does not leave these questions open:

- (2) The classification 'neither- $\top$ -nor- $\perp$ ' is a way of assessing (the assertion of) a sentence (in a context) *to signify that no content is expressed*—nothing to be either  $\top$  or  $\perp$ . Then  $\top$  and  $\perp$  may themselves just be taken to be the classical truth values TRUE and FALSE.

But in this framework for assessment the account of meaning-specification will be complicated. We seem to need both a specification of conditions for assessing when there is content, and a specification of content-fixing conditions (which will be classical TRUTH/FALSITY conditions). But how exactly these two tiers fit together, or whether they can somehow be wrapped up into one, is left open. The two-tier framework will suggest itself most obviously—though not exclusively—when things have to do with the contribution of a context in determining propositional content. For example, it might be said of an assertion of the sentence 'This is blue' that it is a precondition for there being any content to be either  $\top$  or  $\perp$  that there is something which, in the context of the assertion, can be understood to be what 'this' stands for.

The second distinction is between two different choices for what a sentence is to be assessed against. The contrast between a one-tier and a two-tier framework was formulated with the following 'global' kind of set-up in mind:



- (A) The assessment of (the assertion of) a sentence (in a context) as either  $\top$ , or  $\perp$ , or neither- $\top$ -nor- $\perp$ , is against (a formal representation of) the *whole* way things are, or a possible *whole* way for things to be.

But there may be reasons to invoke a ‘local’ kind of set-up:

- (B) The assessment of (the assertion of) a sentence (in a context) as either  $\top$ , or  $\perp$ , or neither- $\top$ -nor- $\perp$ , is against (a formal representation of) some *part of* the way things are, or some possible *part of* a way for things to be.

The wholeness of a global set-up is not meant to rule out relativity to a particular domain of discourse, or to the vocabulary of a particular language. For example, there would be nothing non-whole about the standard model for a first-order language of arithmetic. But in a local set-up we might be working with a mere ‘part’ of this model which, say, consisted just of the information that 10 to 31 are natural numbers and that  $10 < 30$  and  $11 < 29$ , but nothing more.

In a global set-up the classification neither- $\top$ -nor- $\perp$  will arise—whether in the one-tier or the two-tier framework—in virtue of some specific feature of a sentence, perhaps in conjunction with a feature of a particular context of assertion. But in a local set-up a different sort of explanation arises for the classification neither- $\top$ -nor- $\perp$ . The classifications  $\top$  and  $\perp$  may be thought of as ‘positive’ truth values that an assessment can determine, leaving ‘neither- $\top$ -nor- $\perp$ ’ to mean that no positive truth value is determined: a sentence may be neither  $\top$  nor  $\perp$  because the mere part against which it is assessed does not have enough in it to determine anything positive. Local set-ups, will not appear standing on their own: they will be constitutive of some wider semantic system which invokes assessment against partial states or stages of information in one way or another. And it will only be within the wider system that questions about propositional content and sentence meaning can be raised and answered.

Three different ways have emerged to understand ‘neither  $\top$  nor  $\perp$ ’, and there would be nothing but confusion if we tried to assimilate them. But in an overall semantic enterprise more than one of these ways may be in play at the same time—perhaps independently of one another, or perhaps interdependently: there will then be issues about criss-crossing or meshing. (And to complicate things further, our characterization of a one-tier framework describes a general kind of understanding of ‘neither  $\top$  nor  $\perp$ ’ of which there may be various instances.) Criss-crossing would arise, for example, if we were working with a notion of content determined by conditions for (global) assessment as either  $\top$  or  $\perp$  or neither  $\top$  nor  $\perp$ , but if we also wanted a classification for there being no content: then, presumably, sentences would have to be classified as either  $\top$  or  $\perp$  or neither  $\top$  nor  $\perp$  or neither  $\top$  nor  $\perp$  nor neither- $\top$ -nor- $\perp$ . An example of meshing, on the

other hand, will arise if the global assessment of sentences as either  $\top$  or  $\perp$  or neither  $\top$  nor  $\perp$  is to be explained as the outcome of a succession—or some more complicated structure—of set-ups for local assessment. We shall come across meshing of this sort in several places, and the question will arise whether the resulting global assessment is to be taken in a one-tier or a two-tier framework. Maybe, though, this distinction is not as cut and dried as my over-neat schematizing would suggest.

We shall be scratching only the surface of the possible complexity of things. The first few applications we consider are ones that assume global assessment, but a role for local set-ups will become increasingly more prominent as we move through the list. Some of the issues raised by the examples in this section will be discussed in subsequent sections; though the discussions still leave a lot of loose ends.

## 2.2 *Presupposition*

In the context of a logic which admits of sentences which are neither ‘true’ ( $\top$ ) nor ‘false’ ( $\perp$ ), the ‘presupposition’ of a sentence can simply be thought of as its ‘either- $\top$ -or- $\perp$ ’ conditions. Then, whether we are working with a one-tier or a two-tier framework in which to specify the overall  $\top/\perp$ -conditions of a sentence, its presupposition will be constitutive of these  $\top/\perp$ -conditions. Such a notion makes quite general sense, but the terminology is usually associated with a particular application: when triclassificatory logic is deployed in an account of a particular linguistic phenomenon called ‘presupposition’.

A paradigm example sentence would be one containing a definite description, such as

- (1) The present King of France is sane.

It might be said that if this sentence were used to make an assertion, then the existence of a (unique) present King of France is not thereby asserted as a straightforward ‘conjunctive constituent’—as it would be in an assertion of ‘There’s someone who (alone) is presently King of France and who is sane’—but figures in some other, subtler, way: it is presupposed. Theoretical approaches to the linguistic phenomenon vary widely: see Scott Soames’s chapter of the *Handbook*. But the kind of approach that partial logic has relevance to is that according to which the presupposition associated with (the assertion of) a sentence is to be captured semantically as a presupposition in the sense we began with. Of course, to explain what it is that is being captured in this way, we would still have to look to a wider theory of meaning—an issue we shall touch upon in some later sections.

Anyhow, if we wished to construe the description ‘The present King of France’ as a singular term, then we might be prompted to treat (1) along the lines introduced in Section 1.1. Such a treatment would make it a

case of a ‘truth-value gap’ caused by a denotationless term—an idea which authors on presupposition like to trace back to [Frege 1892] but associate more strongly with Strawson in his attack on Russell’s theory of descriptions: [Russell 1905, Russell 1959], [Strawson 1950, Strawson 1964].

This is an encounter we ought to consider. At a superficial level it may simply be seen as a debate between someone who is sensitive to presupposition, and therefore wants to say that a sentence such as (1) is neither true nor false (Strawson) and someone who takes a conservative line that classical logic is to apply and that the sentence is just plain false (Russell). However, there are deeper stands which confuse this simple contrast. According to Russell, definite descriptions are not properly construed as singular terms at all, but are to be defined away in terms of identity and the quantifiers  $\forall$  and  $\exists$ . Strawson, on the other hand, not only construes descriptions as singular terms but suggests a particular theory of reference for them according to which they function much like demonstratives: conditions to determine whether or not they have a denotation and, if so, what it is, cannot be schematized outside a theory about how they are used in particular contexts to refer to particular things. But then, with partial logic at hand, we might actually be prompted to side very much with Russell and against Strawson. Let us consider three progressive stages of becoming more Russellian and less Strawsonian.

First, we might agree to consider descriptions as singular terms, but abandon the Strawsonian account of reference. Partial logic provides a semantics for ‘logically pure’ terms  $\iota x\phi(x)$  whose denoting-conditions depend solely on the way  $\phi(x)$  determines its extension over a given domain of objects. Perhaps we could work with such a semantics? As a residue from the Strawsonian account, we should recognize that description terms call for a contextually determined restriction on the range of the bound variable; but contextual dependence of this sort is a quite general phenomenon, in no way specific to definite descriptions, and it might best be treated separately—in some suitably general account of such dependence.

The second stage away from Strawson towards Russell is the thought that perhaps we might not always want to construe definite descriptions as singular terms. They share many features with quantifier phrases of the form ‘every  $F$ ’, ‘most  $F$ ’, and so on. And it is perhaps a virtue of Russell’s analysis that it casts ‘the  $F$ ’ as a quantifier phrase along with these other forms: the Russellian formula  $\exists x[\forall y[x = y \leftrightarrow Fy] \wedge Gx]$ —or anything equivalent will do equally well—can be seen as an analysis of a scheme of complex quantification  $\iota x[Fx, Gx]$  for ‘the  $F$  is  $G$ ’, just as  $\forall x[Fx \rightarrow Gx]$  is the familiar analysis of a scheme  $\forall x[Fx, Gx]$  for ‘every  $F$  is  $G$ ’. This analysis imposes classical total  $\top/\perp$ -conditions on  $\iota x[Fx, Gx]$ , but why not impose presuppositional  $\top/\perp$ -conditions instead?

Universal quantification has now come into the picture, and so it is pertinent to observe that a sentence such as

(2) All Jack's children are bald.

provides another standard example of presupposition: (2) presupposes that Jack is not childless. Hence we should think of imposing presuppositional  $\top/\perp$ -conditions on  $\forall x[Fx, Gx]$  as well as on  $\iota x[Fx, Gx]$ . What we need for these schemes is something along the following lines:

$$\begin{aligned} \iota x[Fx, Gx] \text{ is } \top & \text{ iff } \text{there is just one } F, \text{ which is } G; \\ \iota x[Fx, Gx] \text{ is } \perp & \text{ iff } \text{there is just one } F, \text{ which is not } G. \\ \\ \forall x[Fx, Gx] \text{ is } \top & \text{ iff } \text{something is an } F \text{ and any } F \text{ is } G; \\ \forall x[Fx, Gx] \text{ is } \perp & \text{ iff } \text{something is an } F \text{ and some } F \text{ is not } G. \end{aligned}$$

These interpretation clauses remain rather informal, but it is easy enough to see that  $\iota x[Fx, Gx]$  will be neither  $\top$  nor  $\perp$  unless there is exactly one  $F$ , and  $\forall x[Fx, Gx]$  will be neither  $\top$  nor  $\perp$  unless there is at least one  $F$ . In [Thomason 1979] the presupposition of universal sentences is handled in this way, though definite descriptions remain singular terms; in [Keenan 1973], on the other hand, descriptions are handled with a scheme of quantification. Note that if  $G$  is a straightforward unstructured predicate, then the  $\top/\perp$ -conditions of  $\iota x[Fx, Gx]$  should turn out to match those of  $G \iota x Fx$ , but  $\iota x[Fx, \dots x \dots]$  promises greater scope for scope distinctions than the singular term  $\iota x Fx$  (see Section 6.4).

The third stage of Russellianization should now be obvious: why not provide an analysis for the scheme  $\iota x[Fx, Gx]$  in terms of identity and the quantifiers  $\forall$  and  $\exists$ ? This, of course, should be an analysis in partial logic, which captures the presuppositional  $\top/\perp$ -conditions. And, while we are about it, why not give an analysis of  $\forall x[Fx, Gx]$  as well? In Section 4.2 we shall show how interjunction and transpication may be used to do this.

If we work with connectives of this sort, perhaps we shall then have progressed some way towards the ideal expressed in [Thomason 1979] of a formal language 'rich enough that every genuine instance of presupposition is formalizable'? Various kinds of presuppositional idiom might be tackled, since with a simple semantics for languages enriched with  $\&$  or  $/$  we can produce formulae which actually *exhibit* non-trivial presuppositions in virtue of 'logical structure' of a very basic kind. This provides something to complement abstract theorising about *relations* of presupposition, such as what occurs in some of the literature on supervaluations, where there is a baroque formal semantics for no particular language at all. For we should, I think, object to the contrast made in [Van Fraassen 1971, p. 138]. According to van Fraassen some non-classical logics, such as modal logic, contain 'non-classical connectors', while others, such as the 'logic of presuppositions', are where 'one studies non-classical relations among (sets of) sentences'. No: the logic of presuppositions should be non-classical in the first sense.

Of course, it is easy enough in simple partial logic to define a formal relation of presupposing—if we want to. We can say that  $\phi$  (logically) presupposes  $\psi$  if and only if  $\phi$  is  $\top$  whenever  $\psi$  is either  $\top$  or  $\perp$ . And once we have interjunction and transplication in our language, then even this simple-minded definition becomes interesting—and even when we restrict attention to propositional logic: for example,  $\phi/\psi$  presupposes  $\phi$ , and  $\phi \times \psi$  presupposes  $\phi \leftrightarrow \psi$ . On the other hand, observe that we could use transplication to define presupposing in terms of equivalence, in a way that matches the use of conjunction in a definition of entailment:  $\phi$  logically presupposes  $\psi$  if and only if  $\phi \simeq \psi/\phi$ . But all this is of parenthetical interest only, since a formal relation of presupposing will have no essential role to play when a semantic theory is set up in our logic.

### 2.3 *Conditional Assertion*

Related to the idea of a truth-value gap for sentences whose presupposition fails to obtain is the thought that naturally occurring conditional sentences of the form ‘if  $\phi$ ,  $\psi$ ’ are neither true nor false when  $\phi$  is false. And in [Belnap 1970] a possible world semantics is developed for a connective ‘/’ of ‘conditional assertion’ according to which, if  $\phi$  is false, then  $\phi/\psi$  is neither true nor false because it makes no assertion, in a degrammatized (*sic*) sense of assertion. Otherwise  $\phi/\psi$  ‘asserts’ what  $\psi$  ‘asserts’ (unless  $\psi$  itself makes no assertion). In Section 4.5 we shall consider this semantics and contrast Belnap’s ‘/’ with transplication in simple partial logic. But observe straightaway that Belnap’s project is manifestly to provide a partial logic for what we called the two-tier framework for the assessment of sentences: ‘no assertion’ means no propositional content to be either true or false.

This prompts us to ask whether partial logic for presupposition should be understood in the same way. Well, any formal treatment of a Strawsonian context-involving account of presupposition would slip naturally enough into a two-tier framework (though Strawson himself might eschew a formal enterprise). But I want to suggest that such a framework would be less happy for the ‘logically pure’ treatment we outlined for the presuppositional schemes of quantification  $\exists x[Fx, Gx]$  and  $\forall x[Fx, Gx]$ —or indeed for description terms  $\iota xFx$ . For example, the presupposition of the sentence ‘All Jack’s children are bald’ is taken simply to be the condition that Jack has children: whether or not this presupposition obtains is an objective fact of the matter, and in an assertion of the sentence it may be contextually quite remote, so that it would be something of a mystery how it might be supposed to effect the question whether or not there is assessable content in the assertion. To elaborate the point, say I know Jack, and say it is taken to be ‘mutual knowledge’ between us that Jack is a father; and say you announce ‘All Jack’s children are bald’. Let us assume, furthermore, that only

yesterday you had seen all Jack's children, and they were as bald as coots. Even so, if they had subsequently taken a wonder drug and had in the meantime sprouted hair, then we would say that you had made a false assertion, viz. an assertion with false content. If, on the other hand, they had all been run over by a bus, would this mean that your assertion was stripped of any content? What would the difference be between the two cases to the success of your linguistic performance as an expression of content? In particular, what difference to my understanding of your performance?

Any attempt to explain a two-tier framework for presuppositional semantics would need to counter these reflections. At least so far as sentences like our example sentence are concerned, it would seem to make more sense to espouse a one-tier framework and to seek an account of 'true', 'false' and 'neither-true-nor-false' simply as three different ways of assessing the content of assertions that sentences can be used to make, whatever status the classification 'neither-true-nor-false' might then turn out to have (see Sections 2.4 and 5.2).

#### 2.4 *Sortal Incorrectness*

Some basic examples of 'category mismatch', or 'sortal incorrectness' motivate allowing predicate/singular-term composition to give rise to a truth-value gap in the second of the two ways mentioned in Section 1.1, viz. because the predicate is not considered to be either true or false of a given object. For example, we might want to say that

- (1) The moon is sane.

is neither true nor false, on the grounds that the moon is just not the kind of thing to be either sane or insane. A logically conservative response would be that this simply means the sentence is false—very obviously so. But there is a counter-response that appeals to the behaviour of negation. In the sentence

- (2) The moon is not sane.

the negation seems naturally to 'go with the predicate', just as much as it would have if we had had 'insane' in place of 'not sane'. If (1) is false, so should (2) be, and a certain tension then arises, since (1) seems to be the straightforward negation of (2).

Precisely this tension is familiar, of course, from logically conservative treatments of paradigm presuppositional sentences, according to which presupposition failure is a straightforward case of falsity. For example, both of the following sentences would be said to be false, yet one is the natural negation of the other:

- (3) The present King of France is sane.

(4) The present King of France is not sane.

There is room here for considerable discussion concerning negation and ambiguity, but the fact remains that on its most natural reading (4) both appears to play a role as the direct negation of (3) and yet fails to be true for precisely the same reasons as (3).

In partial logic there is no tension with negation, since failure-to-be-true is subdivided between the classifications  $\perp$  and  $*$ , and we have a mode of negation which switches  $\perp$  with truth ( $\top$ ) but leaves  $*$  fixed. And so, if (1) and (3) are cast as  $*$ , (2) and (4) fall into place. Indeed, a desire to do justice to the naturalness of natural negation might alone be sufficient to motivate the apparatus of ‘partial’ semantics. Then  $\top$  and  $\perp$  might be considered ‘proper truth values’, as opposed to the ‘gap’  $*$ , just because they are the classifications that negation switches about. Saying this does not in itself preclude regarding  $*$  as a case of falsity (see Section 5.2). In other words, we may have an application for partial logic in a one-tier framework, along with a clear answer to the question how the three sentence classifications mesh with the classical truth values TRUTH and FALSITY:  $\top$  coincides with TRUTH, while FALSITY spans both  $\perp$  and  $*$ .

However this may be, the idea of sortal incorrectness presents its own special issues, and in [Thomason 1972] the behaviour of negation is just one strand in a highly developed semantic theory. Thomason rejects three-entry matrices for giving the meaning of standard connectives and adopts a logical framework of a supervaluational kind. One reason for his doing is this is the thought that sentences of the form of classical tautologies ought to be true. In Section 1.3 we discussed—and found fault with—the general argument behind this thought; now we should consider the particular example sentence that is chosen to back up the argument. This is ‘What I am thinking of is shiny or not shiny’. Thomason points out that if we were using three-entry matrices, it would be necessary to find out what is being thought of before we can say whether or not the sentence is true. It would be true if I were thinking of an apple, say, but sortally incorrect, and hence neither true nor false, if I were thinking of the number 2: this is because on any matrix approach—at least, on any non-eccentric one— $\phi \vee \neg\phi$  would be  $*$  if  $\phi$  were  $*$ . However, it is not clear why this fact should constitute a special problem for matrices or provide any extra ammunition for the general argument, though it is presented as if it did. This is especially puzzling, given the way Thomason deploys the related sentence ‘What I am thinking of is shiny’ against a ‘syntactic’ account of sortal incorrectness, according to which sortally incorrect sentences are intrinsically ungrammatical. For he points out precisely that we cannot know just by looking at the sentence whether or not it is sortally incorrect: the answer depends on discovering what is being thought about. This is a neat argument, but it will be an

uncomfortable one if it is considered to be a problem when we cannot tell *a priori* the sortal correctness or incorrectness of a sentence.

### 2.5 *Semantic Paradox*

A partial-valued approach to the semantic paradoxes rivals the ‘orthodox’ Tarskian account of a hierarchy of languages, in which the semantical predicates of a given language can apply only to the language immediately preceding it in the hierarchy. On this account, a simple paradoxical sentence such as ‘This sentence is false’ would be ruled out as anomalous on the grounds that there can be no place for it in a hierarchy. But in [Kripke 1975] an argument is deployed against the Tarskian theory very similar to the one Thomason deploys against a syntactical account of sortal incorrectness. The point is that paradoxicality cannot be seen as an intrinsic anomaly of given sentences—or for that matter of given configurations of sentences—since even the most innocent of truth-assertions and falsity-assertions can, in unfavourable circumstances, turn out to be paradoxical: examples of this involve people talking about one another’s assertions.

A lot of work has recently been done on the paradoxes—and a lot of that involves partiality in one way or another: see Visser’s chapter in the *Handbook* (and see Section 2.10). Here I shall focus on Kripke. To replace a syntactical hierarchy of truth predicates in different languages, he proposed a single language containing its own partially defined truth predicate. This idea had previously occurred in various authors (see [Martin 1970]), but Kripke took up the formal challenge of addressing particular interpreted languages, such as arithmetic, which are sufficiently rich already to provide the kind of self-reference that leads to paradox. Briefly described, his procedure is to graft a predicate symbol  $T$  onto a language and then to expand its interpretation so as to make  $T$  a truth predicate. It is a truth predicate in the sense that for any sentence  $\sigma$  (of the expanded language), if  $\bar{\sigma}$  is a name in the language for  $\sigma$ , then

$$\begin{aligned} T\bar{\sigma} \text{ is true } (\top) & \text{ iff } \sigma \text{ is true } (\top); \\ T\bar{\sigma} \text{ is false } (\perp) & \text{ iff } \sigma \text{ is false } (\perp). \end{aligned}$$

We shall be able to define a ‘Liar sentence’  $\sigma$ , such that  $\sigma$  is true if and only if  $\neg T\bar{\sigma}$  is true, and such that  $\sigma$  is false if and only if  $\neg T\bar{\sigma}$  is false, but there is no contradiction:  $\sigma$  and  $\neg T\bar{\sigma}$  will both be neither true nor false. The construction of a model to interpret  $T$  depends on the monotonicity of evaluation that partial logic can provide (see Sections 1.2 and 6.2). Kripke considers a supervaluational scheme of evaluation, but seems to prefer simply partial logic (see Section 5.1).

The actual method of model construction is a transfinite induction similar to ones used, for example, in [Gilmore 1974], [Feferman 1975] and, most cunningly, in [Scott 1975]. And compare Aczel’s induction in the appendix



to [Aczel and Feferman 1980]. These references all have to do with systems of type-free class abstraction, where paradoxes are diffused by going undefined: in particular, Scott defines truth/falsity conditions appropriate to turn a model for the  $\lambda$ -calculus into a partial-valued language of classes. From a *set*-theoretical point of view all these systems pay a rather high price, viz. the loss of extensionality, but some work has also been done using partial logic to set up extensional set theories: see [Hinnion 1994].

Truth theories and set theories are the obvious lairs for paradox, but it lurks too in quotational logic—logic set up in a language with explicit devices for talking about itself. For example, a sentence such as

$$M = \text{“}\exists p[“p” = M \wedge \neg p]\text{”}$$

may be thrown up, where  $M$  is a sentence name, and  $p$  is a sentence variable. If  $p$  is taken to range over *all* sentences, and if our background logic is classical, then we have a version of the Liar. One strategy for avoiding trouble is to impose a ranking on sentence variables, and a quotational logic with such a ranking is investigated in [Wray 1987a]. But Wray ends with a proposal for adopting partial logic as the background logic, so that variable-ranking can safely be dropped. And this proposal is carried through in [Wray 1987b].

In his article Kripke criticised other authors who had wanted to defuse the paradoxes by going partial, on the grounds that they did not provide ‘genuine theories’—no ‘precise semantical formulation of a language at least rich enough to speak of its own elementary syntax’, and no ‘mathematical definition of truth’. However, there is a sense of ‘theory’ in which Kripke himself did not provide a theory: that is to say a formal theory in the language for which we have a ‘precise semantical formulation’ and a ‘mathematical definition of truth’. Kripke’s definition of truth is a metalinguistic model-theoretic construction and he left it at that. He provided no system in which a truth-language can express its own semantical principles, let alone any stock of basic ‘axioms’ to generate such principles. I want to suggest that the way to fill in this gap is to use the definition we shall give in Section 7 of what a ‘theory’ is in partial logic. It is not clear, though, what Kripke himself would make of the suggestion, since he claimed that his *logic* is utterly classical. We shall pursue this thought a little way in Section 5.1.

## 2.6 *Stage-by-stage Evaluation*

The bare existence of models for a semantically closed language is only half of Kripke’s story about truth: the construction he employs to demonstrate the existence of such models is associated with an intuitive picture of how sentences can be evaluated as true or as false. In terms of this picture an

account is given—along lines originally explored in [Herzberger 1970]—of ‘paradoxicality’ and related notions. The monotonicity of evaluation now comes to life as a persistence condition governing a procedure of evaluation which runs through stages of increasing information. At a given stage the truth predicate has been defined to a given extent and sentences can be evaluated at that stage in the ordinary way—according to simple partial logic or a supervaluational scheme. But this evaluation then determines the truth predicate for the next stage of evaluation. The truth predicate becomes more defined, and as it becomes more defined so more sentences become true or false, and the truth predicate becomes still more defined . . . and so on. Monotonicity ensures that once a sentence has taken on the value ‘true’ or ‘false’, and the interpretation of the truth predicate has been strengthened accordingly, then it can neither become undefined nor switch truth value at any later stage of evaluation.

Recall the distinction we drew in section 2.1 between a ‘local’ and a ‘global’ set-up for assessing sentences. It would not seem inappropriate to think of the evaluation of sentences at each particular stage of information as a local set-up. But the succession of stages leads up to a global set-up, viz. a stable model to interpret semantically closed partial languages: this model can be seen as the result of pursuing a stage-by-stage evaluation process until it settles down and no new true or false sentences are produced. By general principles governing the inductive definition behind this process it must settle down sooner or later, though in the case of interesting languages this will not be without transfinite leaps to limit-ordinal stages, where all previous truths and falsehoods are gathered up to define the new interpretation of the truth predicate. Assuming, then, that the model we end up with constitutes a global set-up, the question arises whether it provides a one-tier or a two-tier framework of assessment. ‘One-tier’ would seem to be the obvious answer, but this seems to conflict with some of Kripke’s own remarks, and we shall return to the question in Section 5.1.

× × ×

However this may be, another, and in some ways rather simpler, illustration of monotonicity as a constraint in the context of a stage-by-stage process evaluation is provided by the discussion of partial recursive predicates in [Kleene 1952, Section 64]. ‘Kleene’s strong matrices’ are introduced here—the same matrices that we presented in Section 1.2. A partial recursive predicate  $P(\vec{x})$  may be undefined for some  $n$ -tuple  $\vec{a}$  of numbers, and, accordingly, Kleene first offers the simple gloss ‘true’, ‘false’ and ‘undefined’ for the matrix entries  $\top$ ,  $\perp$  and  $*$  (for which he used ‘ $t$ ’, ‘ $f$ ’ and ‘ $u$ ’). These classifications are intended to apply to sentences built up out of partial recursive predicates, and the point of monotonic (which Kleene calls ‘regular’) matrices can be described in terms of the derived role sentence modes play as modes which compound predicates. For, if  $\phi(p_1, \dots, p_n)$  is

a monotonic mode of sentence composition and  $P_1(\vec{x}), \dots, P_n(\vec{x})$  are partial recursive predicates, then  $\phi(P_1(\vec{x}), \dots, P_n(\vec{x}))$  is partial recursive also; while, conversely, if  $\phi(p_1, \dots, p_n)$  is *not* monotonic, then we can find predicates  $P_1(\vec{x}), \dots, P_n(\vec{x})$  which are themselves partial recursive, but which are such that  $\phi(P_1(\vec{x}), \dots, P_n(\vec{x}))$  is not. (See Kleene's Theorems XX and XXI.)

Kleene explains and illustrates monotonicity in terms of a particular kind of algorithm for the interpretation of partial recursive predicates. For a given input  $\vec{a}$ , one of these algorithms will either yield the output 'true', or yield the output 'false', or else go on for ever. A second, 'computational', construal then emerges for the matrix entries: 'true', 'false' and 'unknown (or value immaterial)'. These are classifications for a sentence  $P(\vec{a})$  which can be applied at successive stages in pursuing the algorithm for  $P(\vec{x})$  with input  $\vec{a}$ . The matrix for a given connective,  $\vee$  say, reflects the way algorithms for predicates  $Q(\vec{x})$  and  $R(\vec{x})$  are to be combined to yield an algorithm for  $Q(\vec{x}) \vee R(\vec{x})$ . The classification  $*$  means 'unknown' because if the value  $\top$  or  $\perp$  has not been decided at a given stage, then we do not know what might or might not happen at a further stage. On the other hand, it can also be glossed 'value immaterial', since we may be able to determine the value  $\top$  or  $\perp$  for a compound sentence independently of some constituent sentence which remains  $*$ . For example,  $Q(\vec{a}) \vee R(\vec{a})$  can be evaluated as  $\top$  if  $R(\vec{a})$  has been decided as  $\top$ , even if  $Q(\vec{a})$  remains  $*$ .

The original objective construal of the matrix-entries now falls into place in the following way: 'true' applies to sentences which are decided as  $\top$  at some stage, 'false' to those which are decided as  $\perp$  at some stage, and 'undecided' to sentences which are never decided as either  $\top$  or  $\perp$  at any stage—in other words, which remain  $*$  for ever. Thus Kleene's algorithms can never actually *tell* us that a sentence  $P(\vec{a})$  is undefined. (And since, if  $P(\vec{x})$  is partial recursive, it is, in general, undecidable whether or not  $P(\vec{a})$  is defined, it would, in general, be vain to demand a different kind of algorithm which did tell us.) This explains why none but monotonic connectives are admissible: a resultant value  $\top$  or  $\perp$ , decided by a compound algorithm, is allowed to depend only on out-put values  $\top$  or  $\perp$  from constituent algorithms—never on the classification  $*$ . (See Sections 1.2 and 3.2).

Here we appear to have a paradigm for the use of monotonically representable partial truth-functions. But in [Haack 1974, Haack 1978] it is claimed that Kleene ought rather to have used a supervaluational scheme of evaluation—indeed that his own arguments dictate this. There is no space to do full justice to Haack's remarkable claim, but it would appear to depend primarily on two things. The first is that Kleene mentions a secondary application for his matrices—to sentences built up from total predicates of a kind which are decidable (by one of his algorithms) on *part* of their domain and have their extension over the rest of the domain given by a separate stipulation. It seems that this enables Haack to misunderstand Kleene's gloss for  $*$  as 'lack of information that a sentence is  $\top$  or is  $\perp$ ' to mean

lack of information which of either  $\top$  or  $\perp$  it is. Kleene does not mean this, however:  $*$  (under its computational construal) signifies lack of information whether a sentence is  $\top$  or  $\perp$  or  $*$  *for ever*. It is difficult to see what sense Haack can have made of Kleene's discussion of the 'law of the excluded fourth', which is required to advance from the computational to the objective construal.

Secondly, and connected in some not altogether clear way with the mistaken idea that all sentences under consideration are really either  $\top$  or  $\perp$ , there seems to be a confusion between the constraint of monotonicity (regularity) and a totally different point about the particular matrices chosen for classical connectives: that they are, in Kleene's words, 'uniquely determined as the strongest possible regular extensions of the classical 2-valued truth-tables'. For Haack never actually mentions the notion of regularity, but she interprets Kleene's explanatory discussion of the constraint as if it were some kind of direct argument for a *desideratum* that modes of composition be as strong as possible. In [Haack 1974] she reports on Kleene's illustrative discussion of  $\vee$  (which I sketched above), but she seems to get the point back-to-front. And, in conclusion, she is prepared to announce the 'underlying principle' to be that 'if  $F(A, B, \dots)$  would be  $\top$  ( $\perp$ ) whether  $A, B, \dots$  were true or false, then it is to be  $\top$  ( $\perp$ ) if  $A, B, \dots$  are  $*$ '. If Kleene's principle were something like this, then perhaps we *should* consider supervaluational semantics. But it isn't and we shouldn't.

### 2.7 Stages, States, and Exotic Connectives

Partial logic extends in various directions to more elaborate kinds of semantics than we shall be pursuing. In one direction the computational idea of a process of evaluation can actually be built into the interpretation of some of the logical connectives: consider for example the semantics in [Thomason 1969] for the theory of constructible falsity. This theory is a kind of two-sided intuitionism whose proper constructivist interpretation—handled in [Nelson 1949] and [Lopez-Escobar 1972]—would appeal to twin notions of 'provability' and 'refutability' in the way that intuitionists appeal just to provability. But for a model theory we can consider a two-sided version of Kripke's semantics for intuitionistic logic.

For simplicity of illustration let us consider just a propositional language. Models can then be taken to consist of a set  $\mathbf{V}$ , whose elements  $\alpha$ , are each associated with a partial assignment  $v_\alpha$  of  $\top$  and  $\perp$  to atomic sentences, and a reflexive transitive relation  $\leq$  on  $\mathbf{V}$ , which satisfies the condition that if  $\alpha \leq \beta$  then  $v_\alpha \sqsubseteq v_\beta$ . The elements of  $\mathbf{V}$  are to be thought of as stages of information; and the condition on  $\leq$  is meant to embody the idea that when  $\alpha \leq \beta$  then  $\beta$  has all the information at  $\alpha$  but possibly more besides. Formulae are then evaluated at stages in  $\mathbf{V}$ . For atomic sentences the persistence of truth value ( $\top$  or  $\perp$ ) through stages of increasing information

is constitutive of the model, and the guiding constraint on evaluation rules is that this persistence be extended to all formulae. In other words, our definition of  $v_\alpha(\phi)$  must be such that, for any  $\phi$  if  $\alpha \leq \beta$  then  $v_\alpha(\phi) \sqsubseteq v_\beta(\phi)$ .

The evaluation of negations, conjunctions and disjunctions, at a given stage, involves only the classification at that stage of their immediate constituents—according to the  $\top/\perp$ -conditions of simple partial logic. But the evaluation of conditionals involves constituent classifications at stages of further information. Thus we have a system of local set-ups for assessment with a special kind of interdependence between the set-ups: it resides in the actual assessment conditions of a logical connective. Thomason and Lopez-Escobar give the following  $\top/\perp$ -conditions:

$$\begin{aligned} v_\alpha(\phi \rightarrow \psi) = \top & \text{ iff for every } \beta \geq \alpha, \text{ if } v_\beta(\phi) = \top \text{ then } v_\beta(\psi) = \top; \\ v_\alpha(\phi \rightarrow \psi) = \perp & \text{ iff } v_\alpha(\psi) = \top \text{ and } v_\alpha(\phi) = \perp. \end{aligned}$$

Notice that in fact it is only the  $\top$ -conditions that appeal to further stages. But, in virtue of them,  $\rightarrow$  matches a truth-preservation consequence relation:  $\phi \rightarrow \psi$  is true at any  $\alpha$  in any model if and only if, in any model,  $\psi$  is true at any  $\alpha$  at which  $\phi$  is true. We can take this to mean that  $\phi \rightarrow \psi$  is logically true if and only if  $\psi$  is a (single-barrelled) logical consequence of  $\phi$ .

This is how the theory has grown up, but the  $\top$ -conditions for  $\rightarrow$  could easily be modified to match a double-barrelled notion of consequence—one which also requires preservation of falsity from conclusion to premiss. And we might also adopt stronger  $\perp$ -conditions which, like the  $\top$ -conditions, appeal to further stages of information, and which match the failure of consequence:

$$\begin{aligned} v_\alpha(\phi \rightarrow \psi) = \top & \text{ iff for every } \beta \geq \alpha \left\{ \begin{array}{l} \text{if } v_\beta(\phi) = \top \text{ then } v_\beta(\psi) = \top \\ \text{and} \\ \text{if } v_\beta(\psi) = \perp \text{ then } v_\beta(\phi) = \perp; \end{array} \right. \\ v_\alpha(\phi \rightarrow \psi) = \perp & \text{ iff for every } \beta \geq \alpha \left\{ \begin{array}{l} v_\beta(\phi) = \top \text{ and } v_\beta(\psi) \neq \top \\ \text{or} \\ v_\beta(\psi) = \perp \text{ and } v_\beta(\phi) \neq \perp. \end{array} \right. \end{aligned}$$

The full point of adopting this strong interpretation of  $\rightarrow$  only emerges if we consider setting up non-logical theories in this sort of language:  $\phi \rightarrow \psi$  will be true in all models of a theory if and only if  $\psi$  follows from  $\phi$  in the theory; and  $\phi \rightarrow \psi$  will be false in all models of a theory if and only if  $\psi$ 's following from  $\phi$  is inconsistent with the theory. The details of this would take us too far afield, but see Sections 6.5 and 7.1 for non-logical theories in simple partial logic—and for an indication how to spell out theory-relative notions of ‘following-from’ and ‘(in)consistency’. Anyhow, in the framework of this kind of model there are various ways of ringing the changes on the

interpretation of particular connectives, and obviously a variety of different connectives could be introduced.

A similar framework is provided by the ‘data semantics’ of [Veltman 1981]: ‘data sets’ play the role of stages of information, and an increase-of-information ordering is given simply by the relation  $\sqsubseteq$  between these sets. In this framework Veltman interprets a pair of operators for ‘it may be that’ and ‘it must be that’. But analogous operators can be introduced into the two-sided Kripke models we have set up: let us write ‘ $\diamond$ ’ and ‘ $\square$ ’. For  $\diamond$  the  $\top/\perp$ -conditions will be that

$$\begin{aligned} v_\alpha(\diamond\phi) = \top & \text{ iff for some } \beta \geq \alpha, v_\beta(\phi) = \top; \\ v_\alpha(\diamond\phi) = \perp & \text{ iff for every } \beta \geq \alpha, v_\beta(\phi) = \perp; \end{aligned}$$

and  $\square$  is dual to  $\diamond$ :  $\square\phi$  is equivalent to  $\neg\diamond\neg\phi$ . In [Turner 1984] and [Wansing 1995] the consistency operator  $M$  of [Gabbay 1982] is translated into a partial-logic setting by giving it precisely the interpretation we have given  $\diamond$ . But observe that we have now introduced a crucial departure from the original models: the general persistence condition—that if  $\alpha \leq \beta$  then  $v_\alpha(\phi) \sqsubseteq v_\beta(\phi)$ —has now broken down. It is scuppered by the  $\top$ -conditions for  $\diamond$  (and dually by the  $\perp$ -conditions for  $\square$ ).

The search for exciting new operators can be continued by observing that  $\diamond$  and  $\square$  are a special case of something more general: ‘dynamic’ operators  $\langle\phi\rangle$  and  $[\phi]$ , formed from a formula  $\phi$ . For  $\langle\phi\rangle$  the  $\top/\perp$ -conditions will be that

$$\begin{aligned} v_\alpha(\langle\phi\rangle\psi) = \top & \text{ iff for some } \beta \geq \alpha, v_\beta(\phi) = \top \text{ and } v_\beta(\psi) = \top; \\ v_\alpha(\langle\phi\rangle\psi) = \perp & \text{ iff for every } \beta \geq \alpha, \text{ if } v_\beta(\phi) = \top \text{ then } v_\beta(\psi) = \perp. \end{aligned}$$

And, again,  $[\phi]$  is dual to  $\langle\phi\rangle$ :  $[\phi]\psi$  is equivalent to  $\neg\langle\phi\rangle\neg\psi$ . The formulae  $\langle\phi\rangle\psi$  and  $[\phi]\psi$  could in fact be thought of as kinds of conditional—‘if  $\phi$ , then it may be that  $\psi$ ’ and ‘if  $\phi$ , then it must be that  $\psi$ ’. (Notice that the  $\top$ -conditions of  $[\phi]\psi$ , though not the  $\perp$ -conditions, are exactly the same as those we originally gave for  $\phi \rightarrow \psi$ .) Anyhow,  $\diamond$  and  $\square$  can now be captured as  $\langle\top\rangle$  and  $[\top]$ .

In [Jaspars 1995] a logic is presented which not only contains these ‘upward-looking’ operators, but also a (mutually dual) pair of ‘downward-looking’ ones—let us write  $\langle\phi\rangle'$  and  $[\phi]'$ —whose  $\top$ -conditions and  $\perp$ -conditions at  $\alpha$  involve quantifying over  $\beta \leq \alpha$ . The  $\top/\perp$ -conditions for  $\langle\phi\rangle'$  are:

$$\begin{aligned} v_\alpha(\langle\phi\rangle'\psi) = \top & \text{ iff for some } \beta \leq \alpha, v_\beta(\phi) \neq \top \text{ and } v_\beta(\psi) = \top; \\ v_\alpha(\langle\phi\rangle'\psi) = \perp & \text{ iff for every } \beta \leq \alpha, \text{ if } v_\beta(\phi) \neq \top \text{ then } v_\beta(\psi) = \perp. \end{aligned}$$

Jaspars glosses  $\langle\phi\rangle'\psi$  as meaning ‘it is possible to retract  $\phi$  from the current state [of information] in such a way that  $\psi$  holds afterwards’. Now,

this takes us even further away from the original idea of a two-sided intuitionistic system than  $\diamond$  or  $\langle\phi\rangle$  does. Originally we were to think of the elements of  $\mathbf{V}$  as representing progressive stages in a process of discovery, for which the quasi-ordering  $\leq$  represented possible advances in information—indefeasible advances, which once achieved remained firm. The idea of ‘losing’ information had no role to play in interpreting the language, and the possibility that we might not only lose information, but subsequently ‘advance’ in a different and incompatible way, would have been in clear conflict with the intended interpretation of the model. But this possibility is now envisaged: we have variable states, not progressive stages, of information. [Wansing 1993] is a comprehensive essay investigating the ups and downs of all this; and [Wang and Mott 1998] provides a discussion of how quantifiers fit in.

Jaspars emphasizes the dynamic character of his semantics by defining two relations over the elements of  $\mathbf{V}$  which a formula determines as its ‘dynamic meaning’:

$$\begin{aligned} \alpha \llbracket \phi \rrbracket_{\top}^{\leq} \beta & \text{ iff } \alpha \leq \beta \text{ and } v_{\beta}(\phi) = \top; \\ \alpha \llbracket \phi \rrbracket_{\top}^{\geq} \beta & \text{ iff } \alpha \geq \beta \text{ and } v_{\beta}(\phi) \neq \top. \end{aligned}$$

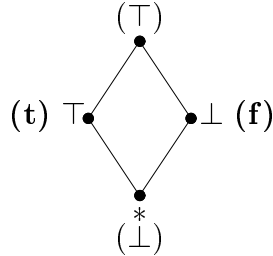
Thus  $\alpha \llbracket \phi \rrbracket_{\top}^{\leq} \beta$  ( $\alpha \llbracket \phi \rrbracket_{\top}^{\geq} \beta$ ) means that  $\beta$  is a possible way of extending (reducing)  $\alpha$  to include (remove) the information that  $\phi$  is true. The notation used here is mine; in particular, I have put in the subscript ‘ $\top$ ’ to point up the one-sidedness of these definitions: there is a complementary pair of relations, defined by replacing ‘ $\top$ ’ with ‘ $\perp$ ’.

These relations between states of information have been defined in terms of  $\leq$  and the evaluation of a formula at a state of information (which is itself defined in terms of  $\leq$ ). But an alternative strategy would be to take relations that determine dynamic meaning as semantically primitive—to define them directly, by recursion on the complexity of formulae. Definitions of this kind, giving an explicit ‘dynamic semantics’, are very popular nowadays: further examples appear at the end of Section 2.10 and in Section 4.3. In Section 4.3 there are also some general remarks on the very idea of a dynamic semantics.

## 2.8 Under-defined and Over-defined

Another way to extend simple partial logic is to consider more truth-value classifications than just  $\top$ ,  $*$  and  $\perp$ . In particular, if  $*$  means ‘neither  $\top$  nor  $\perp$ ’, what about a classification for ‘both  $\top$  and  $\perp$ ’? This might even make some sense in an application where ‘neither  $\top$  nor  $\perp$ ’ signifies a kind of undefinedness that is *under*definedness: there might then be a correlative notion of *over*definedness.

There is in any case an irresistible temptation to add a top element—never mind what it could mean—to the degree-of-definedness ordering on truth-value classifications. This yields a four-element lattice:



The labels in brackets are the ones used in [Scott 1973a]. Let us call this lattice  $D_0$ : the beauty of Scott's idea is that  $D_0$  can be naturally embedded into the domain  $D_1$  of monotonic functions from  $D_0$  into  $D_0$ —and this in a way which provides the basis for embedding  $D_1$  into *its* monotonic function space  $D_2$ , and so on. There is a sequence of nested domains, and a limit domain can be defined which constitutes a system of type-free functions closed under application and abstraction—a model for the  $\lambda$ -calculus.

But in fact a similar construction can be carried out if we start with our more modest *semi*-lattice of  $\top$ ,  $*$  and  $\perp$ —see [Barendregt 1984], for example—and so there is no special motivation here for adding 'over-defined' as a fourth truth-value classification. What a  $\lambda$ -calculus model of this sort provides is a kind of higher-order, but type-free, partial propositional logic: truth values and truth functions inhabit a single unified domain. Quantifiers, however, would seem to present something of a stumbling block in attempts to provide a full-blown type-free partial logic by means of this sort of construction. Application in a limit domain is, loosely speaking, defined in terms of approximations from preceding domains, and, even if we iterate the construction through transfinite stages, it is not clear how successive approximations could ever build up to any decent definition of quantification as a function both ranging over and contained in a limit domain. (The workable definitions I have discovered so far perhaps just about count as non-trivial, but they specify too weak a notion of quantification to be useful.) Furthermore, it does not seem that starting with the four-element lattice of truth-value classifications would offer any advantage. Intensional type-free logic, on the other hand, is much easier to obtain: consider [Scott 1975] and the other similar work mentioned in Section 2.5.

Anyhow, the idea that a sentence may be over-defined, in being both true and false, is one that paraconsistent logicians would like to make serious sense of: see Priest's chapter in the *Handbook*. But in this chapter we need only advert to places in work we have already mentioned where the



four-element lattice plays a role. First, then, it turns up in the type hierarchy of [Muskens 1989] (see Section 1.2). Secondly, the general framework set up in [Langholm 1988] allows both-truth-and-falsity as well as neither-truth-nor-falsity, though a definition of ‘coherence’ is immediately given to delineate those logics which run on just the three truth-value classifications  $\top$ ,  $\perp$  and  $*$  (see Section 1.3). And again in [Bochman 1998] partial logic turns out to be a special case in a more general four-valued framework (see Section 1.1). Compare, too, the work on the paradoxes in [Visser 1984].

### 2.9 *Non-deterministic Algorithms*

But four is still a small number: there are even more truth-value classifications in the ‘non-deterministic partial logic’ developed for the semantics of programming languages in [Päppinghaus and Wirsing 1981]. This logic is applicable to the evaluation of sentences under ‘non-deterministic algorithms’. The algorithms are ‘non-deterministic’ because at given stages in pursuing them a choice may be left of (finitely many) different ways to proceed. Assuming a particular choice is always made, then a sentence will either be evaluated as  $\top$  or as  $\perp$ , or else remain undefined ( $*$ ) (either because the procedure grinds to a conclusionless halt or because it goes on forever). But different choices might result in different resultant classifications. And so, for a given non-deterministic algorithm, there is a spread of alternative classifications. The seven values of Pappinghaus and Wirsing’s logic are the different possible spreads: the non-empty subsets of  $\{\top, *, \perp\}$ . The authors explain various constraints on the interpretation of modes of sentence composition and provide a stock of connectives which is expressively complete for modes meeting these constraints.

I am too out of touch properly to survey the role partial logic and its relatives have played in computer science. But I do know that in an extended version of [Blamey 1991], a degree-of-definedness ordering derived from partial logic is called in to handle divergence—along with non-determinism—in models for CSP processes.

### 2.10 *Situation Semantics*

‘Situation Semantics’ was introduced in [Barwise and Perry 1981a, Barwise and Perry 1981b] as a rival to the Fregean tradition in semantics according to which truth and truth conditions are central notions. Rather, it was argued, situations and truth-in-a-situation conditions are central. Quite an industry has subsequently developed, and there is now a chapter in the *Handbook* which is dedicated to the theory of situations. Here I shall restrict attention largely to the early foundational papers. In this work objects and relations are taken as metaphysically basic, and—suppressing complications to do with time and place—situations are then configurations of objects and re-

lations. They could be modelled over a given domain  $D$  of objects as partial functions from the set of all  $(n + 1)$ -tuples consisting of an  $n$ -place relation on  $D$  and  $n$  elements of  $D$  into the truth-values  $\top$  and  $\perp$ . (Empty argument places of the kind considered in Section 1.2 do not enter the picture here: we can take these functions, modelled set-theoretically, just as subsets of total functions.) Thus situations turn out to be a kind of partial model and provide a paradigm for the idea of a local set-up for the assessment of sentences. A simple sentence such as ‘John hits Mary’, for example, would be true (false) in a situation  $s$  if and only if  $s(\text{hits}, \text{John}, \text{Mary}) = \top$  ( $\perp$ ).

We might, then, think of the meaning of a sentence  $\phi$  as a predicate of situations—one which determines, as its truth-sided interpretation, the set ‘ $\llbracket\phi\rrbracket_{\top}$ ’ of situations in which it is true. (Barwise and Perry use ‘ $\llbracket\phi\rrbracket$ ’ for this set.) However, we can only think in this way once a number of parameters have been filled in. For the linguistic meaning of a sentence, just like that of any subsentential item, is given as a many-place relation with an array of argument places designed to reveal its sensitivity to both linguistic and non-linguistic context: and a great many of these argument places are for situations. For example, a definite description is evaluated for a denotation relative to a situation—a situation which can cross-refer in various ways with situation slots elsewhere in the architecture of a sentence, possibly, but not necessarily, to be ultimately determined by the context of utterance. Furthermore, situations are taken to be the very objects of perception in certain ‘naked infinitive’ constructions such as ‘Hilary sees Mary hit John’: roughly, this would be true in a situation in which Hilary sees a situation in which Mary hits John. Along these lines Barwise and Perry offer an account of the ‘logical transparency’ of such ‘... sees  $\phi$ ’ contexts, which contrasts with the opacity arising in sentences of the form ‘... sees *that*  $\phi$ ’.

Anyhow, if we ignore the internal structure of situations, then they can be thought of just as ‘partial possible worlds’—points with respect to which sentences are to be evaluated as  $\top$ ,  $\perp$  or neither  $\top$  nor  $\perp$ . This prompts comparison with other work: for example, in [Humberstone 1981] partial possible worlds are called ‘possibilities’ and are used to provide a semantics for traditional modal logic. (And see [Van Benthem and Van Eijck 1982, Fenstad *et al.* 1987, Van Benthem 1988] for more exploration of interconnections.)

× × ×

Classically propositions are often modelled as sets of possible worlds, but what happens if we are working with partial possible worlds or situations? In the early work we are considering Barwise and Perry suggest modelling propositions as sets of situations satisfying the coherence condition that if  $s \in P$  and  $s \subseteq s'$ , then  $s' \in P$ . And the interpretation sets  $\llbracket\phi\rrbracket_{\top}$  then turn out as propositions. This approach to propositions is later abandoned (see below), but it is worth pursuing a little way, if only as partial-

possible-world theory. In particular, the question arises how to define logical operations over propositions—operations to interpret modes of sentence composition. Conjunction and disjunction, obviously enough, turn out to be just intersection and union, so that  $\llbracket \phi \wedge \psi \rrbracket_{\top} = \llbracket \phi \rrbracket_{\top} \cap \llbracket \psi \rrbracket_{\top}$  and  $\llbracket \phi \vee \psi \rrbracket_{\top} = \llbracket \phi \rrbracket_{\top} \cup \llbracket \psi \rrbracket_{\top}$ . But what about negation? Barwise and Perry do not actually treat negation as a mode of sentence composition: it turns up in more complicated categories. Even so, given a proposition  $P$ , another proposition  $P^* = \{s^* \mid s \in P\}$  is determined, where  $s^*$  is the situation obtained from  $s$  by reversing the values  $\top$  and  $\perp$ . And for basic sentences  $\phi$ , such as ‘John hits Mary’,  $\llbracket \phi \rrbracket_{\top}^*$  turns out to be the set  $\llbracket \phi \rrbracket_{\perp}$  of situations in which  $\phi$  is false: this looks to be a likely candidate for  $\llbracket \neg \phi \rrbracket_{\top}$ .

But to cater for the negation of complex sentences, we had better modify our representation of propositions so that they have their negative side explicitly built in. If we take *pairs*  $\langle P, P^* \rangle$  of Barwise-and-Perry propositions to interpret sententially atomic items, then to interpret compound sentences we can use the following clauses:

$$\begin{aligned} \langle \llbracket \neg \phi \rrbracket_{\top}, \llbracket \neg \phi \rrbracket_{\perp} \rangle &= \langle \llbracket \phi \rrbracket_{\perp}, \llbracket \phi \rrbracket_{\top} \rangle, \\ \langle \llbracket \phi \wedge \psi \rrbracket_{\top}, \llbracket \phi \wedge \psi \rrbracket_{\perp} \rangle &= \langle \llbracket \phi \rrbracket_{\top} \cap \llbracket \psi \rrbracket_{\top}, \llbracket \phi \rrbracket_{\perp} \cup \llbracket \psi \rrbracket_{\perp} \rangle, \\ \langle \llbracket \phi \vee \psi \rrbracket_{\top}, \llbracket \phi \vee \psi \rrbracket_{\perp} \rangle &= \langle \llbracket \phi \rrbracket_{\top} \cup \llbracket \psi \rrbracket_{\top}, \llbracket \phi \rrbracket_{\perp} \cap \llbracket \psi \rrbracket_{\perp} \rangle. \end{aligned}$$

And we can add clauses for interjunction and transplication too:

$$\begin{aligned} \langle \llbracket \phi \bowtie \psi \rrbracket_{\top}, \llbracket \phi \bowtie \psi \rrbracket_{\perp} \rangle &= \langle \llbracket \phi \rrbracket_{\top} \cap \llbracket \psi \rrbracket_{\top}, \llbracket \phi \rrbracket_{\perp} \cap \llbracket \psi \rrbracket_{\perp} \rangle, \\ \langle \llbracket \phi / \psi \rrbracket_{\top}, \llbracket \phi / \psi \rrbracket_{\perp} \rangle &= \langle \llbracket \phi \rrbracket_{\top} \cap \llbracket \psi \rrbracket_{\top}, \llbracket \phi \rrbracket_{\top} \cap \llbracket \psi \rrbracket_{\perp} \rangle. \end{aligned}$$

These equations of course just model the  $\top/\perp$ -conditions proposed in Section 1.1.

We should (in parenthesis) observe that the same equations will serve if we are interested in capturing not the local assessment of a formula in a system of situations, partial possible worlds, or whatever, but rather the global assessment of a formula against complete possible worlds—against whole possible ways for things to be. If  $v_{\alpha}(p)$  is the (partial) evaluation of an atomic sentence  $p$  at a possible world  $\alpha$ , the following pair gives the interpretation of  $p$ :

$$\langle \{\alpha \mid v_{\alpha}(p) = \top\}, \{\alpha \mid v_{\alpha}(p) = \perp\} \rangle.$$

Then, given an arbitrary formula  $\phi$ , we may invoke the displayed equations to fix the sets of possible worlds  $\llbracket \phi \rrbracket_{\top}$ , in which  $\phi$  is  $\top$ , and  $\llbracket \phi \rrbracket_{\perp}$ , in which  $\phi$  is  $\perp$ . Assuming that this provides a one-tier framework of assessment—and it requires some ingenuity to see it providing anything else—the pair  $\langle \llbracket \phi \rrbracket_{\top}, \llbracket \phi \rrbracket_{\perp} \rangle$  then models the ‘partial proposition’ that  $\phi$  expresses.

Anyhow, in the framework of definitions of this kind, a natural version of our double-barrelled consequence relation would be that  $\phi \vDash \psi$  if and only

if both  $[\phi]_{\top} \subseteq [\psi]_{\top}$  and  $[\psi]_{\perp} \subseteq [\phi]_{\perp}$ . Barwise and Perry use just the first conjunct of this to define a notion of consequence (matching  $\models^{\top}$ )—and hence to define equivalence as bi-consequence. Thus defined, consequence and equivalence are stronger, and so more discriminating, than relations which the authors grudgingly label ‘logical’ and define as follows:  $\psi$  is a ‘logical consequence’ of (is ‘logically equivalent’ to)  $\phi$  if and only if, if  $s$  is any *total* situation, then  $s \in [\phi]_{\top}$  only if (if and only if)  $s \in [\psi]_{\top}$ . (Total situations are just situations that are total functions.) It is then one strand in their argument against the Fregean tradition that trouble results if ‘logical’ equivalence is expected to play a role which should rather be played by the more discriminating relation. This involves ringing the changes on the problem, if substitutively of ‘logical’ equivalents is allowed, of non-truth-functional modes of composition which create extensional contexts. The general aim here is to point up oddities which result from thinking directly in terms of truth values (and truth conditions), rather than situations (and truth-in-a-situation conditions). But it’s far from clear that this aim is met. The more discriminating relation of equivalence has nothing specially to do with the local set-ups of situation theory: it is available in *any* partial semantics. Oddities may equally well be avoided by going partial with a global set-up for assessment—and thinking directly in terms of the truth values  $\top$  and  $\perp$  (and  $\top/\perp$ -conditions).

x x x

In [Barwise and Etchemendy 1987] the apparatus of situations is invoked to address semantic paradox. This involves subjecting the notion of a proposition to some scrutiny, and we are offered two conceptions—‘Russellian’ and ‘Austinian’. Under either conception, the formal modelling of propositions is very different from the one presented above. First we have to have ‘states of affairs’: these are the basic constituents of situations, which, working with the definition we set out at the start, can be taken just to be the members of the sets representing the partial functions that model situations, viz.  $(n + 2)$ -tuples consisting of an  $n$ -place relation,  $n$  objects, and a truth value ( $\top$  or  $\perp$ ). Russellian propositions are then defined as constructs built up from states of affairs, in much the way formulae of a formal language are built up from atomic sentences. At bottom we have basic propositions which just are—or directly correspond to—individual states of affairs, and these will be true in a situation if and only if they are contained in it: truth conditions for arbitrary propositions can then be given by recursive clauses that follow their construction, in just the way that clauses are given for evaluating formulae. Austinian propositions, on the other hand, have a particular situation built in as a kind of contextual parameter—what the proposition is ‘about’. A construction from states of affairs gives a ‘proposition *type*’, which needs to be paired with a situation to model an actual

proposition. So an Austinian proposition contains within itself a situation with respect to which it is true or not.

To give an adequate perspective on the Liar sentence, Barwise and Etchemendy espouse Austinian propositions. The Liar sentence is to be taken in a situation-supplying context and will express a proposition about that situation. If  $s$  is the situation supplied and  $p_s$  is the proposition expressed, then  $p_s$  will be a constituent of itself: its proposition type will consist just of the state of affairs  $\langle T, p_s, \perp \rangle$ , where  $T$  is (an item to represent) the property of being true. (Aczel's theory of non-well-founded sets is invoked as the framework in which to define such self-reflexive propositions.) Thus  $p_s$  will be true if and only if  $\langle T, p_s, \perp \rangle \in s$ . But assuming that no situation can be unfaithful to semantic facts, so that  $\langle T, p_s, \perp \rangle \in s$  only if  $p_s$  is not true, it follows that  $p_s$  will not be true—in other words,  $\langle T, p_s, \perp \rangle \notin s$ . And, since  $s$  cannot be unfaithful to *this* fact,  $\langle T, p_s, \top \rangle \notin s$ .

But the modelling of propositions leaves no room for the conclusion that  $p_s$  is therefore neither true nor false: separate falsity conditions are not defined, and so, because  $p_s$  is not true, it's counted simply false. Rather than admitting a neither-true-nor-false proposition, we are invited to appreciate the inevitable partiality of the *situation*. This means we could always extend  $s$  to a situation  $s' = s \cup \{\langle T, p_s, \perp \rangle\}$ , which includes information about the proposition the Liar sentence expressed—though of course in a context supplying *this* situation the Liar sentence will express a different proposition  $p_{s'}$ , and  $\langle T, p_{s'}, \perp \rangle$  will not be contained in  $s'$ .

In [Groeneveld 1994] the idea that the Liar sentence actually drives us on from situation  $s$  to situation  $s'$  is taken up and built into a semantics for languages in which the Liar sentence can be formulated. Partial logic now comes back into the picture—a dynamic partial logic, for which a pair of relations  $[[\phi]]^+$  and  $[[\phi]]^-$  are defined between situations ('+' for  $\top$ , and '-' for  $\perp$ ). These may be glossed as follows:  $s[[\phi]]^+ s'$  if and only if ' $s'$  is the weakest extension of  $s$  that covers the information of  $\phi$ ';  $s[[\phi]]^- s'$  if and only if ' $s'$  is the weakest extension of  $s$  that rejects the information of  $\phi$ '.

### 3 FREGEAN THEMES

#### 3.1 Reference Failure

In Section 1.1 we announced that we should, in partial logic, be able to do justice to the idea that a sentence  $\phi(t)$  can be neither  $\top$  nor  $\perp$  because some constituent term  $t$  is non-denoting. This calls to mind Frege's theory of reference (*Bedeutung*), according to which the truth value 'true' or 'false' is the reference of a true or false sentence, just as the object denoted by a singular term is its reference, and according to which there is a general principle of reference failure that any compound expression lacks a reference

whenever any constituent expression lacks a reference. This principle would then explain particular claims that  $\phi(t)$  is neither  $\top$  or  $\perp$  ‘because’  $t$  is non-denoting. Of course, our partial logic does not obey this strict principle: if the range of interpretation for predicates  $\phi(x)$  is the system of monotonically representable partial subsets of a domain (see Section 1.2), then, since an empty argument place does not necessarily mean no output value,  $\phi(t)$  could be  $\top$  or  $\perp$  even if  $t$  is non-denoting. But can we argue that our semantics provides some other, subtler, general principle to give more than *ad hoc* content to particular claims that  $\phi(t)$  is neither  $\top$  nor  $\perp$  because  $t$  is non-denoting?

This question leads to thoughts that are in any case prompted if we pursue a Fregean parallel and think of  $\{\top, \perp\}$  as the range of reference for sentences and of a domain of objects, or indeed the corresponding classifications ‘denoting so-and-so’, as the range of reference for singular terms. And it is difficult to avoid the parallel. This is not because of any conception external to systematic semantics of what the ‘reference’ of a sentence or singular term is to consist in, but simply because it is a central strand in Frege’s theorising that compound reference be (functionally) dependent on constituent reference: the parallel points up precisely the dependence that must obtain according to the idea that modes of composition are interpreted by partial functions. But then there might seem to be a problem, since the dependence of reference on reference is supposed to be intimately connected with the strict Fregean principle of reference failure, which our logic does not obey. The connection is made (in rather different styles) in [Woodruff 1970, Dummett 1973, Haack 1974, Haack 1978], for example—and a host of more recent references could equally well be given. Haack even presents a deductive-looking argument to the effect that the principle actually follows from the idea of dependence. To defend our framework from the charge that its range of modes is too liberal for it to be understood as a semantics of partial functions, we have to argue that, on the contrary, the dependence of reference on reference does not in itself dictate the crude principle that a compound  $\theta(\alpha)$  lacks a reference whenever any constituent  $\alpha$  lacks a reference. Such an argument will be attempted in Section 3.2. It is not, of course, just a matter of predicate/singular-term composition: either  $\theta(\alpha)$  or  $\alpha$  could be either a singular term or a sentence. And at the end of Section 3.2 we shall generalize the question even further.

Frege himself regarded reference failure as a defect of ordinary language, and in his systematic logical language he went to great, and often artificial, lengths to avoid any kind of undefinedness arising. In [Frege 1891] the suggestion seems to be that logical laws could not be given otherwise. Perhaps this was because he tended to assimilate any kind of undefinedness into an intractable kind of ‘vagueness’, but it might anyway have seemed rather impractical to do with logic with so many gaps. In our semantics, however, there are not so many gaps. Moreover, what gaps there are will not ham-

per our formal development as they would have hampered Frege's, because we shall be presenting logic in terms of *consequence* rather than *truth* (see Sections 6.5 and 7.1.)

✕ ✕ ✕

However this might be, let us briefly consider some meta-semantical discussion of the Fregean idea that  $\phi(t)$  seriously 'lacks' a reference (truth value) when  $t$  'lacks' a reference (object denoted). [Dummett 1973, Chapters 10 and 12] approaches the matter by discerning different strands in Frege's notion of reference, and the possession of a 'semantic role' is taken to be the only strand in common between sentences and singular terms: the semantic role of an item turns out to be what we have been calling its 'semantic classification', though the *notion* of semantic role is anchored to more fundamental ideas (see Section 5.2). First, then, we should distinguish the realm of objects that can be denoted by terms from the realm of semantic roles, which includes the classification 'non-denoting'. Secondly, Dummett also insists on a distinction between the notion of 'truth-value' in the sense of semantic role, viz. classification or matrix entry in whatever semantics there is reason to adopt, and notions of truth and falsity applicable in the evaluation of what someone asserts using a sentence. Hence, no purchase is to be gained on the idea of sentences actually lacking a truth value by drawing a parallel with names lacking a bearer. Moreover, according to Dummett, whenever anyone ever asserts anything, one or other of the truth values in the second sense must apply (see Section 5.2). There is, though, room for the idea that a sentence may be neither 'true' nor 'false' if these labels apply to two, among more than two, semantic classifications. Dummett takes bearer-less names to be a paradigm source for the problems with negation that we discussed in Section 2.4, and, as we saw, these problems motivate a triclassificatory semantics.

According to Dummett we are concerned throughout with singular terms possessing a Fregean sense (*Sinn*), understood as a cognitive content which determines, but is independent of, the object, if any, denoted. In that case, there is no question of denotation failure in any way infecting what a sentence can express, and the right foundations for the use of partial logic to handle possibly-non-denoting singular terms will then be what in Section 2.1 we called a 'one-tier' framework for assessment. However, it would be contentious to assume that all singular terms can properly be treated in such a framework. It has been argued that the function of at least some singular terms is to introduce denoted objects so intimately into what their containing sentence is used to express that, should such a term in fact not denote, then nothing could have been expressed at all: there would be nothing to be either true or false. This is how we glossed 'neither  $\top$  nor  $\perp$ ' in the 'two-tier' framework, and it echoes the Strawsonian approach to the presupposition of definite descriptions, which we put on one side in Sec-

tions 2.2 and 2.3. But recent theorizing along these lines has become more concerned with demonstratives and proper names: some classic references are [Wiggins 1976], [McDowell 1977], and [Evans 1982], where it is argued that it is, in fact, an important strand in some of Frege's own thinking that bearerless proper names cannot be used to express anything—any Fregean thought. A rich debate has subsequently developed, encompassing both exegetical questions and the question what it is right to say: for example, [McDowell 1984, McDowell 1986, Bell 1990, Wiggins 1995, Sainsbury 1999, Wiggins 1999]. This is not the place to disentangle the debate, but we have to consider how it impinges on our theme about reference.

First, it would seem that at least pure description terms must fall within the scope of Dummett's account. (I eschew—though I cannot here provide a proper rebuttal of—the Russellian view that there should in principle be no such singular terms in a properly constituted logical language.) In that case, if we take descriptions as a paradigm for the singular terms that our logic is to accommodate, then it might be supposed that any problem about the dependence of reference on reference will have evaporated: surely we can simply extract from Dummett's account the picture of a total-valued semantics operating throughout on semantic classifications? But, even if we do this, the problem will reappear. Given our particular semantics, with monotonic functions interpreting modes of composition, we can ask what sense, if any, it makes to say of that semantics that it exhibits functional dependence just among the classifications  $\top$  and  $\perp$  and the classifications 'denoting-so-and-so'. This is precisely the question what sense it makes to say that monotonic functions represent partial ones. Itself the question remains internal to the mathematical semantics, but it becomes interesting in connection with at least some applications, if we want a general explanation behind the specific need for, or usefulness of, monotonic forms.

But what if we hanker after taking proper names as the paradigm for singular terms in partial logic? And what if we espouse the two-tier position that when a bearerless name makes a sentence neither true nor false, this is because there can be no Fregean thought expressed by the sentence? It might then be supposed that the kind of infection a bearerless name causes will be so radical that, however it occurs in a sentence, it must block the expression of a thought—so that the crudely Fregean principle will be inevitable. But I want tentatively to suggest that it is perhaps not so obviously inevitable. Central to arguments for the two-tier position is Frege's characterization of the sense of an expression as the 'mode of presentation' (*Art des Gegebenseins*) of a reference. It seems to follow from this that if there is no reference, then there can be no sense: there will be nothing for there to be any mode of presentation of. In particular, if a name has no bearer to be its reference, then it will have no mode of presentation of a bearer to be its sense. Now, suppose we espouse this characterization of sense, and accept the inference from it. Still, does it follow that a sentence containing



a bearerless name can express no thought? It might be supposed to follow, because a thought is the sense of a sentence and, as such, will be dependent on the sense of constituent expressions—in a way that somehow or other matches the dependence of the reference of the sentence (its truth value) on the reference of the constituents. But our thesis concerning reference is that such dependence does not entail the principle that a compound expression must lack a reference whenever any constituent does. If this is right, then a matching thesis concerning sense cannot be dismissed out of hand. There may be room for a sentence that can be used to express a thought even when a constituent name lacks a bearer. For there may be sense for the sentence even when there is no sense to the name.

This remains the mere mooting of a possibility: a thorough investigation is called for into the compositionality of sense, and this is no place for my inchoate thoughts on the matter.

### 3.2 *Functional Dependence*

The problem, recall, is to provide an account of functional dependence which makes sense of saying that the ‘reference’ if any, of a compound  $\theta(\alpha)$  depends on the ‘reference’, if any, of a constituent  $\alpha$ —an account which can explain why  $\theta(\alpha)$  may *sometimes* lack a reference *because*  $\alpha$  lacks a reference, but one which is not subject to the crude Fregean principle that it is *always* the case that

- (1) if  $\alpha$  lacks a reference, then  $\theta(\alpha)$  lacks a reference.

Here  $\theta(\ )$  is a functor, and for the moment we shall assume that both  $\alpha$  and  $\theta(\alpha)$  are either sentences or singular terms, though our remarks will be sufficiently general for it not to matter which. Frege himself wished actually to conflate these categories, but we will not be committed to that: indeed, we could envisage a many-sorted semantics with more than just two distinct domains of reference for basic, non-functor, categories. ([Wiggins 1984], for example, needs this.)

Now, when reference *failure* is *not* the issue, the principle that, with respect to given ranges of reference, ‘compound reference depends on constituent reference’ is familiar as an ‘extensionality’ condition—to pick out modes of composition as extensional predicates or truth-functional sentence functors, for example. Here the idea of dependence is actually being put to work, and what is important is not that each constituent reference has to pull its weight as something on which compound reference depends—a thought that would indeed suggest the crudely Fregean principle—but rather that compound reference depends *only* on constituent reference, not on anything else. This is often spelt out with the following substitutivity condition:

- (2) If  $\alpha$  and  $\beta$  have the same reference,  
then  $\theta(\alpha)$  and  $\theta(\beta)$  have the same reference.

As formulated, (2) presupposes that  $\alpha$ ,  $\beta$ ,  $\theta(\alpha)$  and  $\theta(\beta)$  each have a reference; but we are considering the possibility that expressions lack a reference, and the question naturally arises how (2) might be modified so as to allow for this. In fact is an appropriate answer to this question not precisely what we are looking for?

Presumably, then, we must adopt at least the following constraint on modes of composition:

- (3) If  $\alpha$  has a reference and  $\theta(\alpha)$  has a reference,  
then, if  $\beta$  has the reference of  $\alpha$ ,  $\theta(\beta)$  has the reference of  $\theta(\alpha)$ .

But what if  $\theta(\alpha)$  has a reference, though  $\alpha$  lacks one? We do not want to rule out this possibility, but, to preserve the idea of dependence, it must be constrained. An obvious thought is that if  $\theta(\alpha)$  has a reference even when  $\alpha$  lacks one, then  $\alpha$  must occur in  $\theta(\alpha)$  in a slot that happens to be irrelevant to determining the reference of  $\theta(\alpha)$ —given the reference of all other constituents. But in that case, *whatever*  $\beta$  we care to substitute for  $\alpha$ ,  $\theta(\beta)$  must have the reference of  $\theta(\alpha)$ . Hence for any  $\beta$  (given any  $\alpha$ ):

- (4) If  $\alpha$  lacks a reference but  $\theta(\alpha)$  has a reference,  
then  $\theta(\beta)$  has the reference of  $\theta(\alpha)$ .

And now, to replace (2), the conjunction of (3) and (4) can be logically manipulated into the following substitutivity condition:

- (5) If  $\beta$  has the reference, if any, of  $\alpha$ ,  
then  $\theta(\beta)$  has the reference, if any, of  $\theta(\alpha)$ .

Here, of course, we have to understand the antecedent in a way that makes it trivially true for any  $\alpha$  that lacks a reference.

We are now in a position to explain why it is sometimes apt to say that  $\theta(\alpha)$  lacks a reference ‘because’  $\alpha$  lacks one. For (4) yields a conditional form of Frege’s principle (1): (1) obtains when  $\alpha$ ’s slot in  $\theta(\alpha)$  *is* relevant to determining compound reference. It is a mark of relevance that there exist expressions  $\beta$  and  $\gamma$  such that  $\theta(\beta)$  and  $\theta(\gamma)$  take on a different reference, or such that one of them has a reference but not the other. And it follows from (4) that if such  $\beta$  and  $\gamma$  do exist, then condition (1) does obtain.

This discussion was originally prompted simply as a defence of our partial semantics against the strict Fregean principle (1). But the criterion of functional dependence embodied in condition (5) in fact does more: it dictates *precisely* a semantics of monotonically representable partial functions. Our semantics is not just not too liberal, but it is as liberal as it can be—given the criterion of dependence. To see this, consider a domain of reference  $D_1$  for constituent expressions  $\alpha$  and  $\beta$ , and a domain of reference  $D_2$  for compounds  $\theta(\alpha)$  and  $\theta(\beta)$ . Then (5) means precisely that

- (5') IF, for any  $a_1$  in  $D_1$ , if  $\alpha$  refers to  $a_1$ ,  $\beta$  refers to  $a_1$ ,  
 THEN, for any  $a_2$  in  $D_2$ , if  $\theta(\alpha)$  refers to  $a_2$ ,  $\theta(\beta)$  refers to  $a_2$ .

And so, if we assume that for any item in  $D_1$  there is—or can be introduced—an expression whose reference it is, then we may deduce from (5') that the interpretation of  $\theta(\ )$  can be given as a partial function from  $D_1$  into  $D_2$  of the kind that is representable by a monotonic function from the fixed-up domain  $D_1 \cup \{\otimes_1\}$  into the fixed-up domain  $D_2 \cup \{\otimes_2\}$ : recall Section 1.2. This deals with one-place modes of composition, but the idea generalizes easily enough to arbitrary  $n$ -place ones, since monotonicity coordinate by coordinate is equivalent to monotonicity across all coordinates.

It is interesting to contrast the discussion here with that in [Woodruff 1970, pp. 128-9], where the specific question is raised how to reconcile the use of Kleene's 'strong' matrices for  $\wedge$  and  $\vee$  (in other words the matrices we have adopted) with a generally Fregean way of thinking. Woodruff does not argue, as we have, that there is no trouble over the dependence of compound reference on constituent reference; rather, he argues that dependence may break down—for example when  $\phi \vee \psi$  is  $\top$  because  $\phi$  is  $\top$ , though  $\psi$  is  $*$ —but that this does not matter. The idea seems to be that, provided the constituent items of a sentence all have a sense, including ones without a reference, then we at least have a compound sense for the whole sentence, and this sense can be considered as determining a reference. However, according to our criterion of dependence, this detour through sense is unnecessary. And, to avoid entanglement with the debate that figured at the end of Section 3.1, the detour is in any case best not taken.

\* \* \*

So far we have been thinking of the function which interprets a functor simply as what exhibits dependence of compound reference on constituent reference, but, in Fregean theory, the interpreting functions are themselves the reference of functors, and compound reference 'depends' no less on this kind of reference than on the reference of a constituent singular term or sentence. What then of our monotonically representable partial functions? Can we see them as constituting a range of reference—or a range of 'partial reference'—which is subject to some suitable principle of dependence? It seems we can set them in this Fregean light by considering appropriate generalisations of principle (5) for higher-level functors  $\theta(\ )$  which take functors  $\alpha$  for arguments. If  $\theta(\ )$  is a simple second-level predicate, for example, (such as a first-order quantifier) the principle would be one which linguistically embodied the intuitive idea of dependence that we sketched in Section 1.2 in connection with partial subsets of the system of partial subsets of a given domain. But in fact we can cater for a complete hierarchy of functor categories—one which includes not only functors which take functors as arguments, but also (though this unFregean) functors which *make* functors.

There is no space to pursue these thoughts, but we should point out that it would be inadequate to think of the ‘partial reference’ of partial functors as a ‘partially specified’ (total) reference. This is the idea that [Dummett 1973, p. 170] would like to offer Frege, but it could not explain the subtlety of monotonically representable partial functions. The reason is that first-level functors accommodate empty argument places for reference-less terms in a way which is subject only to the constraint of principle (5). Full account has to be taken of this in our generalization of (5) to higher-level functors.

#### 4 NON-CLASSICAL CONNECTIVES

##### 4.1 *Interjunction and Transplication: Expressive Adequacy*

Let us begin with the proof of expressive adequacy. We argued in Section 1.2 that, since the matrices for the connectives of simple partial logic all describe monotonic functions, any propositional formula, however complex, must also have a matrix which describes a monotonic function. We now show that, conversely, given any monotonic function  $f$  from  $\{\top, *, \perp\}^n$  into  $\{\top, *, \perp\}$ , we can find a formula  $\phi_f(p_1, \dots, p_n) \text{---} \phi_f$  for short—whose matrix describes  $f$ : in other words,  $\phi_f$  will take the classification  $f(x_1, \dots, x_n)$  under the assignment of  $x_i$  to  $p_i$ . We shall use just  $\neg, \wedge, \vee, \wp, \top$  and  $\perp$  to define  $\phi_f$ .

The case when  $n = 0$  is easy: there are three 0-place functions, which are described by the trivial matrices for the logically constant sentences (or 0-place connectives)  $\top, *$  and  $\perp$ . And  $*$  can be defined away as  $\top \wp \perp$ . Otherwise, when  $n > 0$ , we can proceed as follows. First, for any  $n$ -tuple  $\vec{x} \in \{\top, *, \perp\}^n$  and any number  $i$  from 1 to  $n$ . Let the formulae  $\top(\vec{x}, i)$  and  $\perp(\vec{x}, i)$  be defined by cases—by cases within cases—as follows:

$$\begin{aligned} \top(\vec{x}, i) &= \left. \begin{array}{l} p_i \text{ if } x_i = \top \\ \neg p_i \text{ if } x_i = \perp \\ \top \text{ otherwise} \end{array} \right\} \dots \text{ if } f(\vec{x}) = \top, \\ &= \perp \dots \dots \dots \text{ otherwise;} \\ \\ \perp(\vec{x}, i) &= \left. \begin{array}{l} p_i \text{ if } x_i = \perp \\ \neg p_i \text{ if } x_i = \top \\ \perp \text{ otherwise} \end{array} \right\} \dots \text{ if } f(\vec{x}) = \perp, \\ &= \top \dots \dots \dots \text{ otherwise.} \end{aligned}$$

Then we can define  $\phi_f$  to be

$$\left[ \bigvee_{\vec{x} \in \{\top, *, \perp\}^n} \bigwedge_{1 \leq i \leq n} \top(\vec{x}, i) \right] \wp \left[ \bigwedge_{\vec{x} \in \{\top, *, \perp\}^n} \bigvee_{1 \leq i \leq n} \perp(\vec{x}, i) \right].$$

It is now not difficult to check that:

- (i) if the  $\left\{ \begin{array}{l} \text{left-hand} \\ \text{right-hand} \end{array} \right\}$  interjunct of  $\phi_f$  is  $\left\{ \begin{array}{l} \top \\ \perp \end{array} \right\}$  under the assignment of  $x_i$  to  $p_i$ , then  $f(x_n, \dots, x_n) = \left\{ \begin{array}{l} \top \\ \perp \end{array} \right\}$ ;
- (ii) if  $f(x_n, \dots, x_n) = \left\{ \begin{array}{l} \top \\ \perp \end{array} \right\}$ , then both interjuncts are  $\left\{ \begin{array}{l} \top \\ \perp \end{array} \right\}$  under the assignment of  $x_i$  to  $p_i$ .

Given the  $\top/\perp$ -conditions of  $\mathbb{X}$ , it follows from (i) and (ii) that the matrix of  $\phi_f$  does indeed describe the function  $f$ .

It also follows that the left-hand interjunct gives the  $\top$ -conditions of  $\phi_f$ , while the right-hand interjunct gives the  $\perp$ -conditions. And so these formulae provide interesting ‘normal forms’ for monotonic modes of sentence composition. In Section 6.3 we shall show that interjunctive normal forms of this kind exist in quantifier logic too. As specified  $\phi_f$  is likely to contain many otiose occurrences of  $\top$  and  $\perp$ , but there are obvious ways of obtaining a more economical formula.

We have shown that  $\{\neg, \wedge, \vee, \mathbb{X}, \top, \perp\}$  is a set of connectives adequate to express any monotonic function from  $\{\top, *, \perp\}^n$  into  $\{\perp, *, \top\}$ . The question now arises what other sets of connectives are expressively adequate. In particular, given the classical connectives (including  $\rightarrow$  and  $\leftrightarrow$ , which can be defined in terms of  $\neg, \wedge$  and  $\vee$  in the usual way), what are the variations on  $\mathbb{X}$ ? First, then, observe that transplication has equal expressive power. Not only is  $/$  definable in terms of  $\mathbb{X}$ , but also conversely:

$$\begin{aligned} \phi / \psi &\simeq [\phi \wedge \psi] \mathbb{X} [\phi \rightarrow \psi], \\ \phi \mathbb{X} \psi &\simeq [\phi \leftrightarrow \psi] / \phi \simeq [\phi \leftrightarrow \psi] / \psi. \end{aligned}$$

Or we could take the logically undefined sentence  $*$ . We observed above that  $*$  can be defined as  $\top \mathbb{X} \perp$ ; now observe that  $\mathbb{X}$  can be defined in terms of  $*$ :

$$\phi \mathbb{X} \psi \simeq [\phi \wedge *] \vee [\phi \wedge \psi] \vee [* \wedge \psi] \simeq [\phi \vee *] \wedge [\phi \vee \psi] \wedge [* \vee \psi].$$

Hence each of  $\mathbb{X}$ ,  $/$  and  $*$  has the same expressive power as either of the others.

But to give a more complete answer to our question, first consider the subclass of monotonic functions satisfying the following condition (a converse to the crude Fregean principle that we eschewed in Section 3):

$$\text{if } x_i \neq * \text{ for all } i, \text{ then } f(x_1, \dots, x_n) \neq *.$$

In [Van Benthem 1988] such functions are called ‘closed’. Thus the matrix of a formula  $\phi$  will describe a closed function if and only if, for all total assignments  $v$ , either  $v(\phi) = \top$  or  $v(\phi) = \perp$ ; and in [Langholm 1988] such formulae are called ‘determinable’. Clearly the matrix for any formula which

contains no connectives beyond  $\neg, \wedge, \vee, \top$  and  $\perp$  will describe a closed function, since closed functions are closed under composition; furthermore—and less trivially—any such function is described by the matrix of some such formula: in other words, a formula is determinable if and only if it equivalent to a classical formula. There are proofs of this—all different—in [Blamey 1980] and in the two works referred to above.

We are now in a position to provide a general answer to the adequacy question for monotonic modes of composition: the set  $\{\neg, \wedge, \vee, \top, \perp, \boxtimes\}$  is expressively adequate if and only if  $\boxtimes$  is a connective (of any arity) whose matrix describes a non-closed monotonic function. ‘Only if’ is immediate: compounding closed functions will never reach  $\boxtimes$ , for example. On the other hand, we can deduce ‘if’—the claim that *anything* monotonic and non-closed will do—from the fact that the constant sentence  $*$  will do. First,  $*$  itself is the one and only 0-place non-closed monotonic connective. Secondly, if  $n > 0$  and  $\boxtimes$  is an  $n$ -place connective whose matrix describes a non-closed monotonic function  $f_{\boxtimes}$ , then  $f_{\boxtimes}(x_1, \dots, x_n) = *$ , for some  $x_1, \dots, x_n$  such that either  $x_i = \top$  or  $x_i = \perp$  for each  $i$ . And so, together with the constant sentences  $\top$  and  $\perp$ ,  $\boxtimes$  will be sufficient to define  $*$ —and hence any monotonic mode.

For some particular applications of partial logic, the determinability of all formulae in the language may be a *desideratum*, so that non-closed connectives would be out of place. But in [Jaspars 1995] there is a more general claim, which, in the light of the discussion in Sections 1.2 and 3, would seem to be incorrect. He claims that it follows from the idea that being neither  $\top$  nor  $\perp$  means being ‘genuinely undefined’, rather than having a third truth value, that ‘whenever all the parts of some proposition have obtained a truth value, then the proposition ought to get a truth value as well’. However, without some question-begging assumption about the possible structure of propositions—or the sentences that express them—I cannot see why it follows. You might just as well say that it follows from the idea of a singular term’s being genuinely undefined, rather than denoting some specially introduced object, that whenever the constituent terms of a compound term are all defined, then the compound term must be too. But in that case ‘0<sup>-1</sup>’, for example, wouldn’t be undefined. No doubt Jaspars has particularly in mind the kind of undefinedness that arises from what in Section 2.1 I called a local set-up for assessment, so that being neither  $\top$  nor  $\perp$  means that so-and-so information is not sufficient to determine the value  $\top$  or  $\perp$ . But, even if so, this does not dictate any principle that information which *is* sufficient to determine a value for all constituents must *also* be sufficient to determine a value for the compound.

#### 4.2 Interjunction and Transplication: Logical Analysis

The two formulae given to define  $\bowtie$  in terms of  $*$  are each other's dual: and  $\bowtie$  is self-dual. This means that negation, when applied to an interjection, can be driven through to rest equally on both interjuncts. Applied to transplication, on the other hand, negation can be driven past the left-hand constituent—which we may call the *transplicator*—to rest on the right-hand constituent—which we may call the *transplicand*:

$$\neg[\phi \bowtie \psi] \simeq \neg\phi \bowtie \neg\psi, \quad \neg[\phi / \psi] \simeq \phi / \neg\psi.$$

If  $\top$  and  $\perp$  are thought of as the classifications which negation switches, then these equivalences reveal how it is that interjunction and transplication give rise to non-trivial either- $\top$ -or- $\perp$  conditions. Notice, then, that a transplicator can be taken to introduce a presupposition, in the sense that  $\phi$ 's being  $\top$  is a necessary condition for  $\phi / \psi$ 's being either  $\top$  or  $\perp$ . But interjunctions are more interesting:  $\phi \bowtie \psi$  can be thought of as expressing  $\phi$  and  $\psi$  'as standing or falling together', or—as the definition of  $\bowtie$  in terms of  $/$  makes explicit—under the presupposition that they are equivalent.

Recall that in Section 2.2 we gave informal  $\top / \perp$ -conditions for the schemes of presuppositional quantification  $\exists x[Fx, Gx]$  and  $\forall x[Fx, Gx]$ . We can now show how to capture these  $\top / \perp$ -conditions by analysis under interjunction and transplication. This is a project that could be generalized—see [Van Eijck 1995] and [Sandu 1998] for general frameworks in which to handle modes of quantification in partial logic—but  $\exists x[Fx, Gx]$  and  $\forall x[Fx, Gx]$  will do to illustrate the use of interjunction and transplication. For the moment we shall adopt the simplifying assumption that  $F$  and  $G$  are unstructured predicates, totally defined over a given domain: we can then assume that classical principles govern all classical-looking formulae.

First, then, the scheme  $\exists x[Fx, Gx]$ , for 'the  $F$  is  $G$ ', admits the following interjunctive analysis (where  $F!x$  abbreviates  $\forall y[x = y \leftrightarrow Fy]$ ):

$$\exists x[F!x \wedge Gx] \bowtie \forall x[F!x \rightarrow Gx].$$

Clearly the left-hand interjunct had the desired  $\top$ -conditions, and whenever it is in fact  $\top$ , the right-hand interjunct must also be  $\top$ ; similarly, the right-hand interjunct has the desired  $\perp$ -conditions and, whenever it is in fact  $\perp$ , the left-hand interjunct must also be  $\perp$ ; while the conditions under which the two interjuncts take on opposing truth-values are precisely the required  $*$ -conditions. Hence the interpretation of  $\bowtie$  guarantees that we have the right  $\top / \perp$ -conditions for  $\exists x[Fx, Gx]$ . Under presuppositional  $\top / \perp$ -conditions  $\neg \exists x[Fx, Gx]$  must be equivalent to  $\exists x[Fx, \neg Gx]$ : the scheme is self-dual. This is revealed by our analysis, since the negation of the formula above is equivalent to

$$\exists x[F!x \wedge \neg Gx] \bowtie \forall x[F!x \rightarrow \neg Gx].$$

To see this, first drive negation through onto the interjuncts, and thence onto  $Gx$ , and finally switch the interjuncts around.

This analysis of  $\downarrow x[Fx, Gx]$  is just the interjunction of formulae giving a classical Russellian analysis of  $\downarrow x[Fx, Gx]$  and of  $\neg \downarrow x[Fx, \neg Gx]$ . But there are other versions of classical analysis which contain  $\exists xF!x$  as a distinct conjunctive component. On a presuppositional interpretation this component is a presupposition, and the simple strategy of replacing conjunction by transplication yields the following formulae, either of which may serve to analyse  $\downarrow x[Fx, Gx]$ :

$$\begin{aligned} \exists xF!x / \forall x[Fx \rightarrow Gx], \\ \exists xF!x / \exists x[Fx \wedge Gx]. \end{aligned}$$

Notice that these formulae are equivalent because given that  $\exists xF!x$  is  $\top$  the  $\top/\perp$ -conditions of the two transpicands must coincide. Notice too that when we apply negation it slips past the transplicator onto the transpicand, and thence through onto  $Gx$ , to give

$$\begin{aligned} \exists xF!x / \exists x[Fx \wedge \neg Gx], \\ \exists xF!x / \forall x[Fx \rightarrow \neg Gx]. \end{aligned}$$

So again our analysis reveals that  $\neg \downarrow x[Fx, Gx]$  is equivalent to  $\downarrow x[Fx, \neg Gx]$ .

To provide a transpicative analysis for the scheme  $\forall x[Fx, Gx]$  of universal quantification, we can follow a similar pattern:

$$\exists xFx / \forall x[Fx \rightarrow Gx].$$

It is easy to check, given our simplifying assumption concerning  $F$  and  $G$ , that this formula captures the right presuppositional  $\top/\perp$ -conditions. And we should also consider a scheme  $\exists x[Fx, Gx]$ —to be dual to  $\forall x[Fx, Gx]$ , in having the same  $\perp/\top$ -conditions as  $\neg \forall x[Fx, \neg Gx]$ . The obvious analysis is:

$$\exists xFx / \exists x[Fx \wedge Gx].$$

We could use this to symbolize a sentence such as ‘Some of Jack’s children are bald’, which, no less than ‘All Jack’s children are bald’, carries the presupposition that Jack is not childless. I shall leave it as an exercise to provide an interjunctive analysis for  $\forall x[Fx, Gx]$  and for  $\exists x[Fx, Gx]$ .

We cannot, of course, rest with the assumption that  $F$  and  $G$  are unstructured and totally defined: if our schemes of analysis are any good, then they should continue to make appropriate sense when applied to arbitrary formulae  $\phi(x)$  and  $\psi(x)$  in place of  $Fx$  and  $Gx$ . And so we should consider what happens when one scheme of presuppositional quantification occurs embedded in another. Horrendously complicated formulae can arise if a number of quantifiers are analysed out together: in particular, occurrences of  $/$  or  $\times$  will be obscurely embedded not only within the scope of



sentence connectives (including other occurrences of themselves) but also within the scope of the quantifiers  $\forall$  and  $\exists$ . Yet it turns out that any formula, however complex, is in fact equivalent to one of the form  $\phi/\psi$  where  $\phi$  and  $\psi$  themselves contain no occurrence of either  $/$  or  $\&$ . Furthermore, we can specify rules systematically to transform an arbitrary formula into an equivalent formula of this form; and these rules can be framed so that the translocator  $\phi$  will capture the ‘overall presupposition’ of the formula:  $\phi$ ’s  $\top$ -conditions will be precisely the either  $\top$ -or- $\perp$  conditions of  $\phi/\psi$ —and hence too of the original formula. These transformation rules, which we shall present in Section 6.3, can be seen as a logician’s version of ‘projection rules’ for presupposition.

The examples presented here reveal only a small fraction of what interjunction and transplication have to offer in the analysis of presupposition: I hope there will very soon be a publication telling more of the story.

### 4.3 *Static versus Dynamic Semantics*

The idea of a ‘dynamic’ semantics that emerged rather abstractly at the end of Section 2.7, and turned up again in Section 2.10, has figured prominently in the linguistics literature: in particular, presupposition has been given a dynamic treatment. The questions therefore arise whether our use of transplication and interjunction in the analysis of presupposition can be captured in a dynamic semantics, and whether it has to be to provide an adequate foundation. The answers, I want to argue, are respectively ‘yes’ and ‘no’.

Approached dynamically, the meaning of a sentence is seen as captured by its potential to change contextual information states. These states might be taken to be cognitive states of an individual participant in linguistic exchange, or perhaps to be something more communal and complicated; and they might be represented in the form of a partial model of some kind, or as a set of total models or of possible worlds, or as structures that are formulae of some elaborate formal language, or whatever. The general idea can be traced back to work such as [Stalnaker 1972] and [Seuren 1976], and has been developed in [Kamp 1981, Kamp and Reyle 1993, Heim 1982, Seuren 1985, Veltman 1996], and so on. (See [Van Benthem 1991, Muskens *et al.* 1997], and so on, for illuminating surveys.) In such work the old-fashioned idea of giving meaning in terms of truth/falsity conditions is pushed aside—just as it is in situation semantics. Or, at least, it is pushed back, for we must come down to earth at some stage and actually *give* the meaning of the expressions of any particular language: this is the fundamental message of [Lewis 1972]. And presumably the way to come down to earth, *via* the dynamic apparatus, is to give conditions for the correctness of information states.

Anyhow, the presuppositional characteristics of a sentence seem always to be considered context-involving in some special way. But in [Blamey 1980] it was argued against [Karttunen 1973, Karttunen 1974] that a context-involving account of the meaning of presuppositional idioms was unnecessary and something of a distortion: contextual phenomena could best be accounted for on the basis of a semantical account—using the forms of partial logic—which was itself independent of a theory of context. A dynamic approach will not be set up in quite the same way as Karttunen’s, but can we make an analogous point? In [Beaver 1997] dynamic clauses are given to interpret a language with  $\neg$ ,  $\wedge$ , and  $/$  (though Beaver writes ‘ $\psi_\phi$ ’ for ‘ $\phi/\psi$ ’—notation which he adopts from the work in [Krahmer 1995]); and so let us consider his propositional semantics. We may describe the underlying models as consisting of a set  $\mathbf{V}$  of possible worlds  $\alpha$ , each  $\alpha$  determining a classical total assignment  $v_\alpha$  of  $\top$  or  $\perp$  to atomic sentences. States of information are then represented by sets of possible worlds (all those possible worlds compatible with the state of information represented), and to interpret a formula  $\phi$  there is a relation  $\sigma[[\phi]]\tau$  between states  $\sigma$  and  $\tau$ —glossed as meaning ‘it is possible to update  $\sigma$  with  $\phi$  to produce  $\tau$ ’. The definition of  $\sigma[[\phi]]\tau$  has the following dynamic clauses:

$$\begin{aligned} \sigma[[p]]\tau & \text{ iff } \tau = \sigma \cap \{\alpha \mid v_\alpha(p) = \top\}, \\ \sigma[[\neg\phi]]\tau & \text{ iff for some } v, \sigma[[\phi]]v \text{ and } \tau = \sigma \setminus v, \\ \sigma[[\phi \wedge \psi]]\tau & \text{ iff for some } v, \sigma[[\phi]]v \text{ and } v[[\psi]]\tau, \\ \sigma[[\phi / \psi]]\tau & \text{ iff } \sigma[[\phi]]\sigma \text{ and } \sigma[[\psi]]\tau. \end{aligned}$$

The question we should now ask is whether this definition for  $\sigma[[\phi]]\tau$  has to be taken as primitive, or whether the relation can be defined in terms of something which is static and arguably more basic.

Well, any formula  $\phi$  can obviously be evaluated in simple partial logic under a (total) assignment  $v_\alpha$  to yield a value  $v_\alpha(\phi)$ . And so if we define

$$[[\phi]]_\top = \{\alpha \mid v_\alpha(\phi) = \top\}, \quad [[\phi]]_\perp = \{\alpha \mid v_\alpha(\phi) = \perp\},$$

then  $\phi$ ’s content (in  $\mathbf{V}$ ) under partial semantics may be represented by the pair  $\langle [[\phi]]_\top, [[\phi]]_\perp \rangle$ . Alternatively, and equivalently, we could use the equations displayed in Section 2.10 directly to define content-evaluation for  $\phi$ . It then turns out that this content is sufficient to define the relation  $\sigma[[\phi]]\tau$ : a straightforward inductive argument shows that

$$\sigma[[\phi]]\tau \text{ iff } \tau = \sigma \cap [[\phi]]_\top = \sigma \setminus [[\phi]]_\perp.$$

(Hence, observe, the relation is actually a function, though not a total one.) The right-hand side is equivalent to saying that (i) for all  $\alpha \in \sigma$ , either  $v_\alpha(\phi) = \top$  or  $v_\alpha(\phi) = \perp$ , and (ii)  $\tau$  is got from  $\sigma$  by taking away all those  $\alpha$  such that  $v_\alpha(\phi) \neq \top$ —equivalently, given (i), such that  $v_\alpha(\phi) = \perp$ . This argument is essentially the same as the one presented in [Muskins *et al.* 1997] concerning dynamic clauses formulated in a slightly different way.

This brief commentary on Beaver’s apparatus falls short of a full justification for my answers to the opening questions, but it does show that a natural possible world semantics for presuppositional analysis in partial logic is sufficient to determine natural dynamic clauses. These clauses do not have to be taken as the foundation. It would, though, be more natural still if the world-relative assignments  $v_\alpha$  were not restricted to total ones: to function smoothly the atomic formulae of a logical syntax ought to be schematic for arbitrary sentences, and so not subject to any special semantic restriction.

#### 4.4 Non-Monotonic Matrices

Non-monotonic matrices provide the most obvious examples of what our languages cannot express. In [Woodruff 1970], for instance, there are several of the ‘metalinguistic’ sort of connective that we mentioned in Section 1.3. These are obtained by semantic descent from metalinguistic predicates or relations:

$\phi$	$\psi$	$\phi \cong \psi$	$\phi \Rightarrow \psi$	$\phi \mapsto \psi$
$\top$	$\top$	$\top$	$\top$	$\top$
$\top$	$*$	$\perp$	$\perp$	$\perp$
$\top$	$\perp$	$\perp$	$\perp$	$\perp$
$*$	$\top$	$\perp$	$\top$	$\top$
$*$	$*$	$\top$	$\top$	$\top$
$*$	$\perp$	$\perp$	$\top$	$\top$
$\perp$	$\top$	$\perp$	$\top$	$\top$
$\perp$	$*$	$\perp$	$\perp$	$\top$
$\perp$	$\perp$	$\top$	$\perp$	$\top$

$\phi$	$T\phi$	$F\phi$	$*\phi$	$+\phi$
$\top$	$\top$	$\perp$	$\top$	$\top$
$*$	$\perp$	$\perp$	$\top$	$\perp$
$\perp$	$\perp$	$\top$	$\perp$	$\top$

Thus  $\cong$ ,  $\Rightarrow$ , and  $\mapsto$  (for which Woodruff uses ‘ $\rightarrow$ ’) are obtained from relations of equivalence, presupposition and single-barrelled consequence (the relation  $\models^\top$  of Section 1.1) respectively. Woodruff comments that the ‘distinctive feature’ of these connectives is that they yield compounds which are defined even when every constituent is undefined. However, a mode  $t(p)$  which is just constantly  $\top$ , whatever the classification of  $p$ , would have this feature, and yet it is monotonic. From our point of view, ‘not monotonic’ is a more fundamental feature.

But is there any natural way of classifying more finely among additional connectives? It is a well-known result that the  $T$  connective, together with our  $\neg, \wedge, \vee$  and  $*$ , is expressively adequate for arbitrary matrices. And, given  $\neg, \wedge$ , and  $\vee$ , any of the other connectives listed above can define  $T$ . Hence together with monotonic modes they would each yield a full-blown 3-valued logic. This fact about Woodruff’s connectives is rather more interesting than

the simple fact that they are not monotonic, since it raises the question: are there non-monotonic connectives which would *not* provide a full-blown 3-valued logic if they were included with the monotonic modes? In other words: are there any logics whose expressive range is intermediate between the logic of monotonic matrices and the logic of arbitrary matrices? It turns out that there is precisely one.

To complement the relation  $\sqsubseteq$  on  $\{\top, *, \perp\}$  we can define a relation  $\square$ , which might be thought of as a relation of ‘compatibility’, in the following way:

$$x \square y \quad \text{iff} \quad \text{neither } (x = \top \text{ and } y = \perp) \text{ nor } (x = \perp \text{ and } y = \top).$$

This relation will be of interest in Sections 6 and 7, but in the present context it provides a characterization of the intermediate logic: it is the logic of those matrices which describe functions  $f$  that are ‘ $\square$ -preserving’ in the following sense:

$$\text{if } x_i \square y_i \text{ for all } i, \text{ then } f(x_1, \dots, x_n) \square f(y_1, \dots, y_n).$$

To see that  $\square$ -preserving logic fits in as we claim, notice first that monotonic functions are  $\square$ -preserving, though there are  $\square$ -preserving functions which are not monotonic: for example,  $f$  such that  $f(\top) = \top$ ,  $f(*) = \top$  and  $f(\perp) = *$ . And there are also functions which are not  $\square$ -preserving—including all the functions described by the matrices listed above. We now need two facts whose proofs are omitted, because they are tedious (though not difficult):

- (i) if we add to the monotonic sentence modes any non-monotonic  $\square$ -preserving mode, then we can express all  $\square$ -preserving functions.
- (ii) if we add to the monotonic sentence modes any non- $\square$ -preserving mode, then we can express all three-valued functions.

It is easy to check that the class of  $\square$ -preserving functions is closed under composition, and so it follows from (i) that the  $\square$ -preserving modes do indeed provide an intermediate logic. And then it follows from (ii) that this is the only one.

As a corollary of this argument we also have a general answer to the adequacy question for  $\square$ -preserving modes of composition:  $\{\neg, \wedge, \vee, \&, \top, \perp, \bowtie\}$  is expressively adequate if and only if  $\bowtie$  is a connective (of any arity) whose matrix describes a non-monotonic  $\square$ -preserving function.

#### 4.5 Two-Tier Semantics

We now turn to something more exotic, viz. the semantics of [Belnap 1970], which is intended to model a two-tier framework for assessment in which the classification  $*$  means ‘no assertion’: see Section 2.3 above. This is not to say it is intended to be a general modelling of any two-tier framework; nor is it plausibly taken as such: for example, it would not seem to be appropriate for developing any account of Fregean thoughts of the kind mooted at the end of Section 3.1. Anyhow, in Belnap’s semantics propositions are first modelled as sets of possible worlds in the usual classical way, so that a proposition is true at a world if and only if it contains that world, and then interpretation clauses are given which either assign a proposition to a formula at a world—for it to ‘assert’ at that world—or else leave a formula ‘unassertive’ at a world, with no proposition assigned to it. With this apparatus Belnap’s connective ‘/’ is interpreted by stipulating that at a world in which  $\phi$  asserts a false proposition  $\phi/\psi$  is unassertive, and at any other world  $\phi/\psi$  asserts what  $\psi$  asserts, unless  $\psi$  itself is unassertive, in which case  $\phi/\psi$  is again unassertive. Thus ‘/’ turns out very like transplication; though to match it up properly we should have to modify its interpretation so that  $\phi/\psi$  is unassertive not only when  $\phi$  is false, but also when  $\phi$  is unassertive.

This is a minor modification and would not disrupt Belnap’s idea. But we should stress that our (monotonic) interpretation of transplication in simple partial logic is in no way committed to further explication with Belnap’s apparatus. If we want to consider a possible-world semantics, then we have the alternative, and simpler, one-tier option of modelling propositions as ‘partial propositions’ of the kind first introduced in Section 2.10 and later invoked in Section 4.3—that is, as pairs of sets of possible worlds which just model our talk of  $\top/\perp$ -conditions. Any formula would then express a proposition at any world: either- $\top$ -or- $\perp$ -conditions would be constitutive of this proposition rather than being conditions for the existence of a proposition expressed.

The simpler one-tier option would certainly be more appropriate for a logic of presuppositional analysis: recall Section 2.3. But there is a further special point about the use of transplication in analysis which shows that Belnap’s interpretation for ‘/’ makes it crucially different. It would not just be a mistake to think that the role of a transplicator  $\phi$  in  $\phi/\psi$  is to determine whether or not anything is ‘asserted’, but it would be an even worse mistake to take  $\psi$  on its own to represent *what* is asserted, if anything is. We can think of the transpicand in an assertion-specifying role only if we take it filtered through the transplicator, so to speak. Recall that we used  $\forall x[Fx \rightarrow Gx]$  as a transpicand to analyse both  $\lambda x[Fx, Gx]$  and  $\forall x[Fx, Gx]$ : thus we may filter the same transpicand through different transplicators to get something entirely different. Furthermore, different transpicands may be filtered through the same transplicator to yield the same thing—the same

$\top/\perp$ -conditions. For if  $\exists xF!x$  is taken as the transplicator, then we saw that either  $\forall x[Fx \rightarrow Gx]$  or  $\exists x[Fx \wedge Gx]$  does equally well as a transpicand in an analysis of  $!x[Fx, Gx]$ —and there are plenty of other inequivalent formulae we could just as well have chosen:  $\forall x[F!x \rightarrow Gx]$  or  $\exists x[F!x \wedge Gx]$ , for example.

In [Beaver 1997] there is some ambivalence over a formula ‘ $\psi_\phi$ ’, which in Section 4.3 we assimilated to a transplication  $\phi/\psi$ . He glosses  $\psi_\phi$  as ‘the assertion of  $\psi$  carrying the presupposition that  $\phi$ ’, but this is ambiguous. Does it mean (i) the assertion of  $\psi$ , carrying the presupposition that  $\phi$ ; or (ii) the assertion of  $\psi$ -carrying-the-presupposition-that- $\phi$ ? The wording is more likely to convey reading (i), though apparently Beaver actually wants to leave both readings open. But as a gloss on our use of transplication only reading (ii) is admissible, where  $\psi$ -carrying-the-presupposition-that- $\phi$  is understood to mean  $\psi$ -filtered-through- $\phi$ , in the way that our examples of analysis illustrate. This is the content of any assertion that  $\phi/\psi$  represents: the presupposition that  $\phi$  is *constitutive* of this content, not a separate item just stuck on alongside.

This point about the undetachability of a transplicator could in fact be made independent of our espousal of a one-tier rather than a two-tier framework for presuppositional semantics. For *even if* we wanted to gloss the ‘neither- $\top$ -nor- $\perp$ ’ of presupposition failure to mean no assertion, what *is* asserted when  $\phi$  is true and  $\phi/\psi$  represents an assertion could not be specified by  $\psi$  on its own. If, as in Belnap’s semantics, classical propositions are the only candidates for the content of assertions, then, to put it in Belnap’s language, what  $\phi/\psi$  asserts when it asserts anything—that is, when  $\phi$  is true and  $\psi$  asserts something—cannot be what  $\psi$  asserts, but can only be the *conjunction* (intersection) of what  $\phi$  asserts and what  $\psi$  asserts. Indeed, it would be easy enough to revise Belnap’s clauses for ‘/’ along these lines. This is not a point against Belnap, of course; for recall that his ‘/’ is not intended for presuppositional analysis at all, but rather to construe conditionals.

Anyhow, as in an alternative to going to meet Belnap among the possible worlds, we could in fact unravel his semantics into simple  $\top/\perp$ -conditions. Clauses for evaluating a formula at a given world—clauses which make no appeal to any other world—are given in [Dunn 1975]. The following matrices for  $\wedge, \vee$ , and  $/$  then emerge:

$\phi$	$\psi$	$\phi \wedge \psi$	$\phi \vee \psi$	$\phi / \psi$
$\top$	$\top$	$\top$	$\top$	$\top$
$\top$	*	$\top$	$\top$	*
$\top$	$\perp$	$\perp$	$\top$	$\perp$
*	$\top$	$\top$	$\top$	$\top$
*	*	*	*	*
*	$\perp$	$\perp$	$\perp$	$\perp$
$\perp$	$\top$	$\perp$	$\top$	*
$\perp$	*	$\perp$	$\perp$	*
$\perp$	$\perp$	$\perp$	$\perp$	*

Thus, quite apart from ‘/’, the matrices for  $\wedge$  and  $\vee$  show a difference from simple partial logic: conjunction and disjunction are not monotonic (nor even  $\Box$ -preserving). This prompts a question: If we started out with our monotonic matrices for  $\wedge$  and  $\vee$ , then could we sensibly convert them into a Belnap-style two-tier semantics? This becomes a pertinent question in Section 5.1, where we shall address it.

× × ×

But first we should observe that the above non-monotonic, and *prima facie* rather odd, matrix for  $\vee$  also arises in [Ebbinghaus 1969], where a first-order semantics is offered to handle the kind of undefinedness that arises from natural modes of mathematical expression. Ebbinghaus presents his semantics by first giving clauses for when a formula is defined—in a given model—and then building truth conditions on top of this. The rules for disjunction are:

$$\begin{aligned} \phi \vee \psi \text{ is defined} & \text{ iff } \phi \text{ is defined or } \psi \text{ is defined,} \\ \phi \vee \psi \text{ is true} & \text{ iff } \phi \text{ is true or } \psi \text{ is true.} \end{aligned}$$

Hence, if \* means undefined,  $\top$  means true, and  $\perp$  means defined but not true, then Belnap’s matrix for  $\vee$  results. Negation is taken to work in the same way that it does in simple partial logic, and Ebbinghaus defines  $\Delta(\phi)$  as  $\phi \vee \neg\phi$ , to yield a sentence-mode expressing ‘ $\phi$  is defined’. Hence  $\Delta(\phi)$  yields \*, if  $\phi$  is \* (just as it would if we had defined it in simple partial logic). Contrast Woodruff’s  $+\phi$ .

The interpretation of the existential quantifier is analogous to disjunction:  $\exists x\phi(x)$  is taken to be defined just in case  $\phi(x)$  is defined for at least one element in the domain of quantification, and to be true just in case  $\phi(x)$  is true of at least one element. This interpretation is motivated by the desire to allow existential statements to come out false, even when the quantified predicate is undefined for some elements—and so not false of everything: for example, in the domain of rationals or reals,  $\exists x[x^{-1} = 0]$  is to be false, though  $0^{-1} = 0$  is undefined. Clearly this would not be possible

in monotonic logic. However, since (unlike Ebbinghaus) we envisage setting up all nonlogical theories directly in terms of consequence, we are not under the same pressure to assign such existential statements a truth value.

Disjunction and existential quantification thus turn out to be much ‘stronger’ than in simple partial logic. But conjunction and universal quantification are much ‘weaker’. For conjunction we have:

$$\begin{array}{ll} \phi \wedge \psi \text{ is defined} & \text{iff } \phi \text{ is defined and } \psi \text{ is defined,} \\ \phi \wedge \psi \text{ is true} & \text{iff } \phi \text{ is true and } \psi \text{ is true.} \end{array}$$

and so  $\phi \wedge \psi$  is undefined whenever either  $\phi$  or  $\psi$  is. Then  $\forall$  matches  $\wedge$  just as  $\exists$  matched  $\vee$ . These interpretations do not, therefore, yield the classical duality between  $\wedge$  and  $\vee$  and between  $\forall$  and  $\exists$ ; but they allow Ebbinghaus to frame neat rules for  $\Delta(\ )$  in a natural deduction system which is designed to axiomatize a truth-preservation notion of consequence.

This system falls squarely under the heading ‘partial logic’, but in much recent work there seems to be something of a division of interest. On the one hand, partial logicians tend to ignore undefined singular terms—perhaps because they are primarily concerned with partial states of information, or situations, or the like (see Sections 2.7 and 2.10); though this is certainly not a definitive reason for ignoring undefined terms. On the other hand, those setting up systems to accommodate undefined singular terms tend to prefer a logic which at the level of sentences is totally defined and two valued. See [Feferman 1995] for a magisterial exposition of doctrine—and for a survey of work; and for work specifically in the ‘free logic’ tradition, see Bencivenga’s chapter of the *Handbook*. But the system in [Lehmann 1994], for example, is an exception to the trend: it is a partial logic with undefined terms. This is work in the Fregean tradition, and I would want to take issue with it because it espouses the principle of functional dependence that in Section 3 I argued was unnecessarily crude.

## 5 PARTIAL LOGIC AS CLASSICAL LOGIC

### 5.1 *Partial Truth Languages*

A proper discussion of the idea of ‘alternative’ logics is far beyond the scope of this essay. But, *via* some themes we have touched upon already, we shall briefly puzzle over two particular accounts of how the triclassificatory semantics of partial logic can play a role which does *not*, in any interesting sense, give rise to an alternative to classical logic.

First consider [Kripke 1975] which we discussed in Sections 2.5 and 2.6. His remarks about logic are, in fact, rather sketchy and largely centred in footnotes, but nonetheless they are forcefully expressed. In footnote 18, for example, he claims that in adopting Kleene’s monotonic matrices for evaluating sentences he is doing no more than adopting ‘conventions for handling



sentences that do not express propositions' and that these conventions 'are not in any philosophically significant sense changes in logic'. For logic is supposed to apply primarily to propositions which are all either true or false. Kripke draws a parallel between handling possibly non-denoting (numerical) terms and handling sentences which are undefined (\*), and this parallel calls to mind our account (in Sections 1.2 and 3) of the partial-functional interpretation of functors. However, the parallel there was the Fregean one between objects denoted and the truth-values  $\top$  and  $\perp$ , whereas Kripke's parallel is between objects denoted by (numerical) terms and *propositions* expressed by sentences. And in the text he presents us with an explicitly two-tier picture of the meaning of a sentence: gapless truth conditions determine propositions, but sentences, which might turn out to be paradoxical and hence neither true nor false, are not directly interpreted by truth conditions, but by conditions for truth conditions.

Clearly these conditions must not only determine when a sentence expresses a proposition—has gapless truth conditions—but also *what* proposition a sentence expresses when it does express one. Kripke is vague at this point, but his picture of the interpretation of sentences looks to be of the same general kind that Belnap's semantics is intended to model. And so we return to the question raised in Section 4.5: can Kleene's monotonic matrices be made to fit with such a semantics? Kripke seems (in footnote 30) to suggest that they stand a better chance than a supervaluational scheme of evaluation. This is presumably because, according to this scheme, there would be the difficulty of sentences none of whose constituents expressed a proposition, but which are true, just because they are of the form of a tautology. The problem would be to say what proposition such a sentence expresses, in a way which does justice to ideas of compositionality whereby a compound proposition is in some sense determined by constituent propositions. However, even on the Kleene scheme we may have a sentence which is true even though one of its constituent sentences is neither true nor false, and so, according to Kripke, expresses no proposition: for example, something of the form  $\phi \vee \psi$ , where  $\phi$  is a straightforward truth and  $\psi$  is paradoxical. What proposition does  $\phi \vee \psi$  then express? And, in general, what are the rules which tell us what proposition a compound sentence expresses?

Let us assume we can make suitable sense of saying that propositions are closed under boolean operations (perhaps, but not necessarily, because we have modelled them as sets of possible worlds). And let us, by way of example, compare Belnap's and Kleene's matrices for disjunction:

$\vee$	$\top$	$*$	$\perp$
$\top$	$\top$	$\top$	$\top$
$*$	$\top$	$*$	$\perp$
$\perp$	$\top$	$\perp$	$\perp$

Belnap

$\vee$	$\top$	$*$	$\perp$
$\top$	$\top$	$\top$	$\top$
$*$	$\top$	$*$	$*$
$\perp$	$\top$	$*$	$\perp$

Kleene

The four corners of Belnap's matrix are accounted for by saying that if both disjuncts of a disjunction express (or 'assert') a proposition, then the disjunction expresses the corresponding disjunction of the propositions. If, on the other hand, neither disjunct expresses a proposition, then the disjunction expresses none: this explains the centre of the matrix. So far the two matrices coincide, but what happens when one disjunct expresses a proposition but not the other? The *prima facie* oddity of Belnap's matrix is explained by his stipulation that the disjunction expresses the same proposition as the proposition-expressing constituent. But what could Kripke say about Kleene's matrix? The only obvious course would be to make  $\vee$  the same kind of connective as Belnap's  $'/'$  of conditional assertion and to say that the existence of a proposition expressed by the disjunction depends on the *truth value* of the disjunct which expresses a proposition (the truth value of that proposition): if it is true, then this is the true proposition expressed, and if it is false, then no proposition is expressed.

It might, then, be possible to make sense of things along these lines, treating conjunction in a parallel way and, of course, extending it all to handle quantifiers. And some such elaboration of partial semantics would have to be given, if Kripke ever wants to set up logic for his truth languages so that it can be seen to apply to classical propositions that sentences might or might not express. But then we might ask what role these propositions would play in his account of truth and paradoxicality. We are invited to see the monotonicity-dependent construction of models in some way reflecting an intuitive evaluation process of sentences, in a progression of successive stages: as the process is pursued more sentences receive truth values. But we can hardly think of this process as evaluating sentences for the propositions, if any, they express. For, though monotonicity guarantees persistence of truth value, there would not be persistence of propositions. If, for example,  $\phi$  were true and  $\psi$  neither true nor false, but at some stage of evaluation  $\psi$  took on a truth value, then the proposition originally expressed by  $\phi \vee \psi$  would disappear as a disjunctive constituent of the later proposition. Or are classical propositions meant to be there from the start, in some sense, so that they can determine the process of evaluation? This is a picture it seems difficult to make sense of. So what theoretical role would classical propositions play? The oddity is that they seem to have no role.

But why should we envisage a two-tier semantics at all? The alternative is to give a direct account of meaning in terms of (partial)  $\top/\perp$ -conditions, so that sentences have 'partial propositions' as their meaning: see Section 2.10 above, and compare the remarks in Section 4.5. This would mesh naturally with Kripke's account of the stage-by-stage evaluation of sentences: as the evaluation progresses, so propositions become progressively 'more defined'. The idea of partial propositions is crying out for further elucidation, but if it can be provided, then we have the most straightforward way to gloss the formal construction of models for semantically closed languages. As

we explained in Section 2.6, a succession of partial, but progressively less partial, models culminates in a model which is still partial but which is stable: it throws up no new true or false sentences in terms of which to define (the truth predicate of) any less partial model. There are  $\top/\perp$ -conditions for all sentences in each model in the succession, and in the final stable model they give the final stable meaning of sentences of the language.

The natural logical apparatus to adopt would then be, or be something similar to, what we shall outline in Sections 6 and 7. And there is surely nothing to stop us interpreting this apparatus as delivering a logic that is essentially classical—richer than usual simply because it embodies rules for handling varieties of undefinedness. The presentation of partial logic in Section 1 was meant to reveal this interpretation as a coherent option.

## 5.2 *Natural Negation*

If we turn to Dummett's views on presupposition and the role a logic such as ours might play in providing a semantics, then the debate becomes a very different one. The idea that a sentence classified as  $*$  expresses no proposition, or that no assertion can be made using it, does not enter the picture at all. Thus Dummett's account is in what we have been calling a one-tier framework. But it does invoke two different aspects of meaning, and these give rise to two different levels of content. Sentences are semantically classified as  $\top$  or  $*$  or  $\perp$ , and there is a notion of the 'semantic content' of a sentence as its  $\top$ -versus- $*$ -versus- $\perp$  conditions; but assertions made using sentences are to be classified exhaustively into TRUE ones and FALSE ones, and the 'assertoric content' of a sentence matches TRUTH-versus-FALSITY conditions. Semantic classifications then divide into the 'designated', for sentences which can be used to make TRUE sentences, and the 'undesignated', for sentences which can be used to make FALSE ones. Presuppositional  $*$  will side with  $\perp$  as a case of FALSITY.

With this framework at hand, Dummett is polemical—for example in the introduction to [Dummett 1978]—against theorists who would deploy notions of 'truth' and 'falsity' matching the semantic classifications  $\top$  and  $\perp$  in a way which he reserves exclusively for TRUTH and FALSITY. For according to Dummett, so long as we concern ourselves with the linguistic activity of making assertions and with the meaning a sentence manifests in this linguistic practice, then a basic notion of objective truth and falsity leaves no room for anything but an exhaustive dichotomy into the TRUE and the FALSE. There is an exclusion clause for 'vagueness' and 'ambiguity'—which Dummett thinks of as cases where an assertion would have no fully determinate content (and which he supposes have nothing to do with presupposition)—but, otherwise, the way things are is either incorrectly ruled out by an assertion, in which case it is FALSE, or else it is not, in which case it is TRUE.

This thesis emerges in various places in [Dummett 1973], but is crispest in [Dummett 1959]. (Note that ‘anti-realist’ worries are not at issue here.)

Why then bother with a semantics that operates on the classifications  $\top$ ,  $*$ , and  $\perp$ ? The point, it is suggested, will simply be to obtain a smooth account of how sentences are composed from their constituents. To interpret modes of linguistic composition—not just sentence composition—a system of semantic classifications reveals how the meaning (semantic content) of a complex expression is determined by the meaning (semantic content) of its constituents; but the point of a systematic semantics of this sort is just to lead up in an appropriate way to a correct specification of TRUE-versus-FLASE conditions—assertoric content. It is here that the notion of ‘semantic role’, alluded to in Section 3.1, fits in: the classifications of a semantics capture one strand in the Fregean notion of reference because they play a role—a semantic role—in determining the TRUTH or FALSITY of (assertions made using) sentences. Thus the subtleties of a presuppositional semantics are taken to derive just from structural features we are prompted to discern in a language.

The most salient feature would seem to be negation. We saw in Section 2.4 that, to account for natural modes of negation as straight-forward sentence functors, we need to split non-truth (falsity) into  $\perp$ , which negation switches with  $\top$  (TRUTH), and  $*$ , which it leaves fixed. This is a standard example of Dummett’s to illustrate the role of triclassificatory semantics, and he uses it also to explain our naive inclination to apply the labels ‘true’, ‘false’ and ‘neither-true-nor-false’ directly to the evaluation of assertions themselves. For we are inclined, he suggests, to call the assertion of a sentence ‘false’ only if the assertion of the (natural) negation of that sentence would have been true (TRUE).

This seems to provide an explanation of the three-fold scheme of semantic classification—and hence of the phenomenon of presupposition—in terms of the TRUE/FALSE dichotomy and natural negation. But, as Dummett himself points out, natural negation is not a purely syntactical notion. Just consider the complex variety of forms: for example, ‘Some of Jack’s children are not bald’ is just as much a natural negation of ‘All Jack’s children are bald’ as ‘Not all Jack’s children are bald’ is. Hence natural negation is not identifiable as such in a meaning-independent way. Yet as natural speakers we do recognise it, and as theorists it is handy for us to do it justice. So, what *is* it? It is not unreasonable to call, in turn, for an explanation of this mode of sentence modification. Furthermore, why is natural negation *negation* at all? The classical truth values TRUE and FLASE are taken to be fundamental, but natural negation takes some FALSE sentences to ones that are again FALSE (when there is presupposition failure). At this point Dummett’s overall picture might leave us restless. For it does not seem to leave much room to answer these questions—or not without going round in a circle. For what can we say about natural negation other than that

it is a mode of sentence modification which is to be called in to spell out the way we talk about presupposition and its treatment in triclassificatory semantics?

To break out of the circle, we might be prompted to look to an account of presupposition in the theory of assertion—to mesh with the semantic notion cast in triclassificatory logic. And, whatever we think of the particular accounts on offer in the literature, there is surely *something* to be said along these lines. Dummett's response to this would probably be that we would just have decorated the circle with superficial aspects of meaning, unless it had been shown that presupposition can make a distinctive contribution to the cognitive adjustments that people undergo when they understand what is said to them; and that this could never be shown. Even so, in the work referred to at the end of Section 4.2 I'm foolhardy enough to attempt an account which is intended to provide more than superficial decoration.

## 6 FIRST-ORDER PARTIAL SEMANTICS

### 6.1 Languages and Models

In this section we outline a model-theoretic semantics to match the sketch of first-order partial logic given in Section 1. A few facts about the logic will emerge, and their proofs will be outlined in Section 7, after we have presented an axiomatization of logical laws. (I hope that a much fuller account of things will soon appear.) The languages we work with will contain no description terms, though Section 6.4 deals with how they would fit in.

Let us, then, take a language  $\mathcal{L}$  to consist of the following.

(a) Logical vocabulary:

- (1) sentence connectives  $\neg$ ,  $\wedge$ ,  $\vee$ ,  $\&$ ,  $\top$ , and  $\perp$ ,
- (2) quantifier symbols  $\forall$  and  $\exists$ ,
- (3) an identity predicate symbol  $=$ ,
- (4) a constant symbol  $\otimes$ ,
- (5) a set  $Var$  of denumerably many variables.

(b) Non-logical vocabulary:

- (1) a set  $Prd(\mathcal{L})$  of predicate symbols,
- (2) a set  $Fnc(\mathcal{L})$  of function symbols,
- (3) a set  $Cns(\mathcal{L})$  of constant symbols.

The elements  $P$  of  $Prd(\mathcal{L})$  and  $f$  of  $Fnc(\mathcal{L})$  are taken to come along with fixed numbers  $\lambda(P)$  and  $\mu(f)$  to give their number of argument places. Accordingly, a *model for*  $\mathcal{L}$  is to be a structure  $M$  consisting of

- (0) a set  $D_M$  (which does not have to be non-empty),
- (1) for each  $P \in Prd(\mathcal{L})$ ,  
a monotonic function  $P_M : (D_M \cup \{\otimes\})^{\lambda(P)} \rightarrow \{\top, *, \perp\}$ ,
- (2) for each  $f \in Fnc(\mathcal{L})$ ,  
a monotonic function  $f_M : (D_M \cup \{\otimes\})^{\mu(f)} \rightarrow D_M \cup \{\otimes\}$ ,
- (3) for each  $c \in Cns(\mathcal{L})$ ,  
an element  $c_M \in D_M \cup \{\otimes\}$ .

For assignments to variables we shall just use functions  $s : Var \rightarrow D_M \cup \{\otimes\}$ . Then, if we define the *terms* of a language  $\mathcal{L}$  in the usual inductive way, the classification  $M_s(t)$  of a term  $t$  under an assignment  $s$  is given as follows:

$$\begin{aligned}
 M_s(x) &= s(x), \text{ for all } x \in Var, \\
 M_s(\otimes) &= \otimes, \\
 M_s(c) &= c_M, \\
 M_s(ft_1 \cdots t_{\mu(f)}) &= f_M(M_s(t_1), \dots, M_s(t_{\mu(f)})).
 \end{aligned}$$

We can now build on this to define the *formulae* of  $L$  and their interpretation in a model. Formulae—like terms—are taken to be defined by functor-first construction throughout. But we shall be writing ‘ $\phi \wedge \psi$ ’, ‘ $c = d$ ’, etc., rather than ‘ $\wedge \phi \psi$ ’, ‘ $= cd$ ’, etc., and so be helping ourselves to brackets when necessary. This is just so much notation. And we can regard the following ‘definitions’ in the same light:

$$\begin{aligned}
 * &=_{df} \top \times \perp, \\
 \phi \rightarrow \psi &=_{df} \neg \phi \vee \psi, \\
 \phi \leftrightarrow \psi &=_{df} [\phi \rightarrow \psi] \wedge [\psi \rightarrow \phi], \\
 \phi / \psi &=_{df} [\phi \rightarrow \psi] \times [\phi \wedge \psi].
 \end{aligned}$$

Given an assignment  $s$ , a variable  $x$  and an element  $a$  in the fixed-up domain  $D_M \cup \{\otimes\}$  of a model  $M$ , let  $s(x|a)$  be the assignment such that  $s(x|a)(x) = a$  and  $s(x|a)(y) = s(y)$  if  $y$  is a variable distinct from  $x$ . Then the classification  $M_s(\phi)$  of a formula  $\phi$  under an assignment  $s$  can be specified as follows:

$$\begin{aligned}
M_s(\top) &= \top, \\
M_s(\perp) &= \perp, \\
M_s(t_1 = t_2) &= \begin{cases} \top & \text{iff } M_s(t_1), M_s(t_2) \in D_M, \text{ and } M_s(t_1) = M_s(t_2) \\ \perp & \text{iff } M_s(t_1), M_s(t_2) \in D_M, \text{ and } M_s(t_1) \neq M_s(t_2), \end{cases} \\
M_s(Pt_1 \dots t_{\lambda(P)}) &= \begin{cases} \top & \text{iff } P_M(M_s(t_1), \dots, t_{\lambda(P)}) = \top \\ \perp & \text{iff } P_M(M_s(t_1), \dots, t_{\lambda(P)}) = \perp, \end{cases} \\
M_s(\neg\phi) &= \begin{cases} \top & \text{iff } M_s(\phi) = \perp \\ \perp & \text{iff } M_s(\phi) = \top, \end{cases} \\
M_s(\phi \wedge \psi) &= \begin{cases} \top & \text{iff } M_s(\phi) = \top \text{ and } M_s(\psi) = \top \\ \perp & \text{iff } M_s(\phi) = \perp \text{ or } M_s(\psi) = \perp, \end{cases} \\
M_s(\phi \vee \psi) &= \begin{cases} \top & \text{iff } M_s(\phi) = \top \text{ or } M_s(\psi) = \top \\ \perp & \text{iff } M_s(\phi) = \perp \text{ and } M_s(\psi) = \perp, \end{cases} \\
M_s(\phi \otimes \psi) &= \begin{cases} \top & \text{iff } M_s(\phi) = \top \text{ and } M_s(\psi) = \top \\ \perp & \text{iff } M_s(\phi) = \perp \text{ and } M_s(\psi) = \perp, \end{cases} \\
M_s(\forall x\phi) &= \begin{cases} \top & \text{iff } M_{s(x|a)}(\phi) = \top, \text{ for every } a \in D_M \\ \perp & \text{iff } M_{s(x|a)}(\phi) = \perp, \text{ for some } a \in D_M, \end{cases} \\
M_s(\exists x\phi) &= \begin{cases} \top & \text{iff } M_{s(x|a)}(\phi) = \top, \text{ for some } a \in D_M \\ \perp & \text{iff } M_{s(x|a)}(\phi) = \perp, \text{ for every } a \in D_M. \end{cases}
\end{aligned}$$

These are the conditions for  $\top$  and  $\perp$ :  $M_s(\phi)$  is  $*$  if it is neither  $\top$  nor  $\perp$ . Observe how it is that variables have nothing more to do with  $\otimes$ , once they are bound by a quantifier.

The classification of a formula has been defined relative to an assignment, but we can neatly advance to a non-relative definition: let  $M(\phi)$  be  $M_s(\phi)$ , where  $s$  assigns  $\otimes$  to all variables. It will then follow (from Lemma 3) that  $M(\phi) = \top(\perp)$  if and only if  $M_s(\phi) = \top(\perp)$  for every assignment  $s$ . A *free* occurrence of a variable in a formula can be defined in the usual way, and sometimes we shall call free variables *parameters*. *Sentences* are parameter free formulae and, as we should expect, their classification is in any case quite independent of assignments. This is a corollary of the following standard semantical lemma:

LEMMA 1 (Relevant Variables).

- (1) If  $s_1(x) = s_2(x)$  for every  $x$  in  $t$ , then  $M_{s_1}(t) = M_{s_2}(t)$ .
- (2) If  $s_1(x) = s_2(x)$  for every  $x$  free in  $\phi$ , then  $M_{s_1}(\phi) = M_{s_2}(\phi)$ .

Let us use the notation ‘ $u(t/x)$ ’ for the term obtained from a term  $u$  by substituting  $t$  for  $x$  throughout. Similarly, let us use ‘ $\phi(t/x)$ ’ for the formula obtained from  $\phi$  by substituting  $t$  for all *free* occurrences of  $x$  in  $\phi$ . And we shall say that  $t$  is *substitutable for  $x$  in  $\phi$*  when no occurrence of a variable in  $t$  becomes a bound (i.e., not free) occurrence in  $\phi(t/x)$ . Then there is a second standard lemma:

LEMMA 2 (Substitution for Variables).

- (1)  $M_s(u(t/x)) = M_{s(x|M_s(t))}(u)$ .
- (2)  $M_s(\phi(t/x)) = M_{s(x|M_s(t))}(\phi)$ , *provided that  $t$  is substitutable for  $x$  in  $\phi$ .*

## 6.2 Monotonicity and Compatibility

Now for something more interesting: the monotonicity of evaluation (cf. Section 1.2). First we need to define a ‘degree-of-definedness’ relation,  $\sqsubseteq$ , between models for a given language  $\mathcal{L}$ : this consists in the appropriate ‘ $\sqsubseteq$ ’-relation holding between the respective interpretations of the vocabulary of  $\mathcal{L}$ . Writing it all out explicitly, in terms only of the basic relations on  $\{\top, *, \perp\}$  and on a fixed-up domain  $D \cup \{\otimes\}$ , we have:  $M \sqsubseteq N$  if and only if  $M$  and  $N$  have a common domain  $D$  and, for all  $P \in \text{Prd}(\mathcal{L})$ ,  $f \in \text{Fnc}(\mathcal{L})$  and  $c \in \text{Cns}(\mathcal{L})$ ,

- (1)  $P_M(\vec{a}) \sqsubseteq P_N(\vec{a})$ , for all  $\vec{a} \in (D \cup \{\otimes\})^{\lambda(P)}$ ,
- (2)  $f_M(\vec{a}) \sqsubseteq f_N(\vec{a})$ , for all  $\vec{a} \in (D \cup \{\otimes\})^{\mu(f)}$ ,
- (3)  $c_M \sqsubseteq c_N$ .

We also need to extend  $\sqsubseteq$ , in the natural way, to assignments:  $s_1 \sqsubseteq s_2$  iff  $s_1(x) \sqsubseteq s_2(x)$ , for all  $x \in \text{Var}$ . Then for terms as well as formulae:

LEMMA 3 (Monotonicity of Evaluation). *If  $M_1 \sqsubseteq M_2$  and  $s_1 \sqsubseteq s_2$ , then*

- (1)  $M_{1s_1}(t) \sqsubseteq M_{2s_2}(t)$ ,
- (2)  $M_{1s_1}(\phi) \sqsubseteq M_{2s_2}(\phi)$ .

The proof of this lemma is just a matter of checking—by induction on the complexity of terms and formulae.

To set alongside ‘degree-of-definedness’ there is also a ‘compatibility’ relation between models. In Section 4.4 we defined a relation  $\square$  on  $\{\top, *, \perp\}$ : neither  $\top \square \perp$  nor  $\perp \square \top$ , but otherwise  $\square$  holds. And, analogously, we can define  $\square$  on a fixed-up domain  $D \cup \{\otimes\}$  by:

$$a \square b \quad \text{iff} \quad a \text{ and } b \text{ are not distinct elements of } D.$$

Then to define compatibility between models:  $M \square N$  if and only if  $M$  and  $N$  have a common domain  $D$  and for all  $P \in \text{Prd}(L)$ ,  $f \in \text{Fnc}(L)$  and  $c \in \text{Cns}(L)$ ,



- (1)  $P_M(\vec{a}) \sqsubseteq P_N(\vec{a})$ , for all  $\vec{a} \in (D \cup \{\otimes\})^{\lambda(P)}$ ,
- (2)  $f_M(\vec{a}) \sqsubseteq f_N(\vec{a})$ , for all  $\vec{a} \in (D \cup \{\otimes\})^{\mu(f)}$ ,
- (3)  $c_M \sqsubseteq c_N$ .

And, as in the case of  $\sqsubseteq$ , a natural compatibility relation is induced between variable assignments:  $s_1 \sqsubseteq s_2$  iff  $s_1(x) \sqsubseteq s_2(x)$ , for all  $x \in Var$ . We could now prove a lemma parallel to Lemma 3, got by replacing ' $\sqsubseteq$ ' by ' $\sqsubset$ '; but this result will shortly be generalized, at least so far as formulae are concerned (part (2)), to something usefully stronger: Lemma 6.

Observe now that if  $M \sqsubseteq N$ , then we can coherently stick  $M$  and  $N$  together to define a model  $M \sqcup N$ , which is the least upper bound of  $M$  and  $N$  with respect to the  $\sqsubseteq$  ordering: if  $D$  is the common domain of  $M$  and  $N$ , then, the interpretation of  $P \in Prd(L)$ ,  $f \in Fnc(L)$  and  $c \in Cns(L)$ , is given by stipulating that,

- (1) for any  $\vec{a} \in (D \cup \{\otimes\})^{\lambda(P)}$ :
 
$$P_{M \sqcup N}(\vec{a}) = \begin{cases} \top & \text{iff either } P_M(\vec{a}) = \top \text{ or } P_N(\vec{a}) = \top \\ \perp & \text{iff either } P_M(\vec{a}) = \perp \text{ or } P_N(\vec{a}) = \perp, \end{cases}$$
- (2) for any  $\vec{a} \in (D \cup \{\otimes\})^{\mu(f)}$ , and any  $b \in D$ :
 
$$f_{M \sqcup N}(\vec{a}) = b \text{ iff either } f_M(\vec{a}) = b \text{ or } f_N(\vec{a}) = b,$$
- (3) for any  $b \in D$ :
 
$$c_{M \sqcup N} = b \text{ iff either } c_M = b \text{ or } c_N = b.$$

Similarly, if  $s_1$  and  $s_2$  are assignments  $D \cup \{\otimes\} \rightarrow Var$ , and if  $s_1 \sqsubseteq s_2$ , then an assignment  $s_1 \sqcup s_2$  is coherently defined by stipulating that for any  $x \in Var$ , and any  $a \in D$ ,  $s_1 \sqcup s_2(x) = a$  iff either  $s_1(x) = a$  or  $s_2(x) = a$ .

We shall also be interested in purely 'elementary' relations  $\sqsubseteq_e$  and  $\sqsubset_e$  between models—and also a relation of elementary equivalence  $\sim_e$ —which can indifferently be characterised either in terms of the classification of arbitrary formulae  $\phi$ , or sentences  $\phi$ , as follows:

$$\begin{aligned} M \sqsubseteq_e N & \text{ iff } M(\phi) \sqsubseteq N(\phi), \text{ for any } \phi, \\ M \sqsubset_e N & \text{ iff } M(\phi) \sqsubset N(\phi), \text{ for any } \phi, \\ M \sim_e N & \text{ iff } M(\phi) = N(\phi), \text{ for any } \phi. \end{aligned}$$

Notice that  $M \sim_e N$  if and only if  $M \sqsubseteq_e N$  and  $N \sqsubseteq_e M$ , just as  $M = N$  if and only if  $M \sqsubseteq N$  and  $N \sqsubseteq M$ .

Notice, too, that the relations  $\sqsubseteq$  and  $\sqsubset$ —and indeed the identity relation—can be characterized in terms of the evaluation of formulae:

LEMMA 4.

$$\begin{aligned} M \sqsubseteq N & \text{ iff } D_M = D_N \text{ and } M_s(\phi) \sqsubseteq N_s(\phi), \text{ for any } \phi \text{ and any } s, \\ M \sqsubset N & \text{ iff } D_M = D_N \text{ and } M_s(\phi) \sqsubset N_s(\phi), \text{ for any } \phi \text{ and any } s, \\ M = N & \text{ iff } D_M = D_N \text{ and } M_s(\phi) = N_s(\phi), \text{ for any } \phi \text{ and any } s. \end{aligned}$$

‘Only if’ follows trivially from Lemma 3 and the parallel result for  $\sqsupset$ ; ‘if’ can easily be checked by considering atomic formulae.

Relations of ‘degree-of-definedness’ and ‘compatibility’ also arise in a natural way between *formulae*. Let us restrict attention to ‘purely logical’ relations, defined by generalising over all the models for a given language; then, including also a relation  $\simeq$  of equivalence:

$$\begin{aligned}\phi \sqsubseteq \psi & \text{ iff } M_s(\phi) \sqsubseteq M_s(\psi), \text{ for any } M \text{ and any } s, \\ \phi \sqsupset \psi & \text{ iff } M_s(\phi) \sqsupset M_s(\psi), \text{ for any } M \text{ and any } s, \\ \phi \simeq \psi & \text{ iff } M_s(\phi) = M_s(\psi), \text{ for any } M \text{ and any } s.\end{aligned}$$

Notice that  $\phi \simeq \psi$  if and only if  $\phi \sqsubseteq \psi$  and  $\psi \sqsubseteq \phi$ .

The relation of compatibility between formulae gives rise to an interesting question. If  $\phi \sqsupset \psi$ , then  $\phi$  and  $\psi$  never take on conflicting truth values: can we then stick  $\phi$  and  $\psi$  together to yield a more defined formula  $\chi$  which takes the value  $\top$  or  $\perp$  whenever either one of  $\phi$  and  $\psi$  does? In other words, is there for compatible formulae any thing analogous to  $M \sqcup N$  for compatible models  $M$  and  $N$ ? Let us call  $\chi$  a *joint* for  $\phi$  and  $\psi$  if and only if, for any model  $M$  and assignment  $s$ ,

$$M_s(\chi) = \begin{cases} \top & \text{iff either } M_s(\phi) = \top \text{ or } M_s(\psi) = \top \\ \perp & \text{iff either } M_s(\phi) = \perp \text{ or } M_s(\psi) = \perp. \end{cases}$$

There is clearly no monotonic mode of sentence composition which we could use to compound  $\phi$  and  $\psi$  and thereby produce such a  $\chi$ , but in fact joints for compatible formulae always exist. In the restricted case of propositional logic this is an immediate corollary of ‘expressive adequacy’ (see Section 4.1 above), but it holds in quantifier logic too:

**THEOREM 5 (Compatibility Theorem).** *Any two logical compatible formulae have a joint.*

To prepare for our proof of this result in Section 7.3, we need two lemmas. The first is the promised generalization of the compatibility result parallel to Monotonicity of Evaluation (Lemma 3):

**LEMMA 6.** *If  $\phi \sqsupset \psi$ ,  $M_1 \sqsupset M_2$ , and  $s_1 \sqsupset s_2$ , then  $M_{1s_1}(\phi) \sqsupset M_{2s_2}(\psi)$ .*

To see this, consider  $M_1 \sqcup M_2$  and apply part (2) of Lemma 3. (Note that part (2) of Lemma 3 can itself be generalized along the lines of this lemma: replace ‘ $\sqsupset$ ’ by ‘ $\sqsubseteq$ ’.)

The second lemma could be thought of as saying that  $\phi$  and  $\psi$  have a ‘least upper bound’, viz. a joint, when and only when they have an ‘upper bound’. (Indeed, this makes quite literal sense if we think of the relation induced by  $\sqsubseteq$  on the Lindenbaum algebra of a language.)

**LEMMA 7.**  *$\phi$  and  $\psi$  have a joint if and only if there is a formula  $\lambda$  such that  $\phi \sqsubseteq \lambda$  and  $\psi \sqsubseteq \lambda$ .*

‘Only if’ is trivial. Conversely, given  $\lambda$ , the following formula is obviously a joint:  $[[\phi \vee \psi] \wedge \lambda] \bowtie [\lambda \vee [\psi \wedge \phi]]$ .

### 6.3 Interjunctive and Transplivative Normal Forms

In section 4.1 we promised normal forms in quantifier logic to match the in propositional normal forms that derive from our proof of expressive adequacy. Let us, then, say that a formula is in *interjunctive normal form* when it is an interjunction  $\psi \bowtie \chi$  such that neither  $\psi$  nor  $\chi$  contains any occurrence of  $\bowtie$  and such that, for any model  $M$  and any assignment  $s$ ,  $M_s(\psi \bowtie \chi) = \top$  if and only if  $M_s(\psi) = \top$ , and  $M_s(\psi \bowtie \chi) = \perp$  if and only if  $M_s(\chi) = \perp$ . Logical consequence has not yet been officially defined for our first-order languages, but from the outline in Section 1.1 it is easy to see that this condition will turn out equivalent to saying that  $\psi \vDash \chi$ . (The precise definition of  $\vDash$  is in section 6.5.) We can now show that an arbitrary formula  $\phi$  is logically equivalent to a formula in interjunctive normal form: in fact we can describe a procedure to transform  $\phi$  into normal form.

The procedure relies on the fact—easy to check—that our language admits ‘substitutivity of equivalents’: when a subformula is replaced by something equivalent, then the resulting formula is equivalent to the original one. This means we can first replace any atomic subformula  $\phi'$  of a formula  $\phi$  by  $\phi' \bowtie \phi'$ —which itself is clearly in normal form—and, since  $\phi' \simeq \phi' \bowtie \phi'$ , the resulting formula will be equivalent to  $\phi$ . Then we can progressively pull  $\bowtie$  out of the scope of the logical operators in  $\phi$ —both connectives and quantifiers—working up from those with narrowest scope to the one with widest scope. What makes this possible is that if  $\psi \bowtie \chi$  is in normal form, or if both  $\psi_1 \bowtie \chi_1$  and  $\psi_2 \bowtie \chi_2$  are in normal form, then the following equivalences hold, and the formula on the right of ‘ $\simeq$ ’ will again be in normal form:

$$\begin{aligned} \neg(\psi \bowtie \chi) &\simeq \neg\chi \bowtie \neg\psi \\ (\psi_1 \bowtie \chi_1) \wedge (\psi_2 \bowtie \chi_2) &\simeq (\psi_1 \wedge \psi_2) \bowtie (\chi_1 \wedge \chi_2) \\ (\psi_1 \bowtie \chi_1) \vee (\psi_2 \bowtie \chi_2) &\simeq (\psi_1 \vee \psi_2) \bowtie (\chi_1 \vee \chi_2) \\ (\psi_1 \bowtie \chi_1) \bowtie (\psi_2 \bowtie \chi_2) &\simeq (\psi_1 \wedge \psi_2) \bowtie (\chi_1 \vee \chi_2) \\ \forall x(\psi \bowtie \chi) &\simeq \forall x\psi \bowtie \forall x\chi \\ \exists x(\psi \bowtie \chi) &\simeq \exists x\psi \bowtie \forall x\chi. \end{aligned}$$

Thus we can pull  $\bowtie$  out of the scope of an operator by replacing a subformula of one of the forms displayed on the left by the equivalent formula on the right. At each stage equivalence to  $\phi$  is preserved; and at each stage the replacement subformula is in normal form: and so we end up with an equivalent formula in normal form.

The displayed equivalences do not of course hold unconditionally, except for the first. We could alternatively use ones that did, but the formulae

on the right would then be double the length. For example, to specify how to pull  $\bowtie$  out of the scope of a quantifier, when it governs an arbitrary interjunction, we need the following:

$$\begin{aligned}\forall x(\psi \bowtie \chi) &\simeq \forall x(\psi \wedge \chi) \bowtie \forall x(\psi \vee \chi) \\ \exists x(\psi \bowtie \chi) &\simeq \exists x(\psi \wedge \chi) \bowtie \exists x(\psi \vee \chi).\end{aligned}$$

Suitable equivalences for  $\wedge$ ,  $\vee$ , and  $\bowtie$  I leave as an exercise.

Let us now pretend that  $/$  is a primitive connective—and  $\rightarrow$  and  $\leftrightarrow$  as well. And let us say that a formula is in *transplicative normal form* when it is a transplicative  $\psi/\chi$  such that neither  $\psi$  nor  $\chi$  contains any occurrence of either  $/$  or  $\bowtie$  (so that there are only classical logical operators in  $\psi$  and  $\chi$ ) and such that, for any  $M$  and any  $s$ ,  $M_s(\psi) = \top$  if and only if either  $M_s(\psi/\chi) = \top$  or  $M_s(\psi/\chi) = \perp$ . Then if we have a procedure, along the lines of the one above, for transforming an arbitrary formula into an equivalent one in transplicative normal form, this will yield projection rules for presupposition of the kind we were interested in at the end of Section 4.2.

Such a procedure can be based on the following equivalences (which hold whether or not the constituents on the left are already in normal form):

$$\begin{aligned}\phi &\simeq (\phi \vee \neg\phi) / \phi \\ \neg(\psi / \chi) &\simeq \psi / \neg\chi \\ (\psi_1/\chi_1) \wedge (\psi_2/\chi_2) &\simeq ((\psi_1 \wedge \psi_2) \vee (\psi_1 \wedge \neg\chi_1) \vee (\psi_2 \wedge \neg\chi_2)) / (\chi_1 \wedge \chi_2) \\ (\psi_1/\chi_1) \vee (\psi_2/\chi_2) &\simeq ((\psi_1 \wedge \psi_2) \vee (\psi_1 \wedge \chi_1) \vee (\psi_2 \wedge \chi_2)) / (\chi_1 \vee \chi_2) \\ (\psi_1/\chi_1) \rightarrow (\psi_2/\chi_2) &\simeq ((\psi_1 \wedge \psi_2) \vee (\psi_1 \wedge \neg\chi_1) \vee (\psi_2 \wedge \chi_2)) / (\chi_1 \rightarrow \chi_2) \\ (\psi_1/\chi_1) \leftrightarrow (\psi_2/\chi_2) &\simeq (\psi_1 \wedge \psi_2) / (\chi_1 \leftrightarrow \chi_2) \\ (\psi_1/\chi_1) / (\psi_2/\chi_2) &\simeq (\psi_1 \wedge \psi_2 \wedge \chi_1) / \chi_2 \\ (\psi_1/\chi_1) \bowtie (\psi_2/\chi_2) &\simeq (\psi_1 \wedge \psi_2 \wedge (\chi_1 \leftrightarrow \chi_2)) / \chi_2 \\ \forall x(\psi / \chi) &\simeq (\forall x(\psi \wedge \chi) \vee \exists x(\psi \wedge \neg\chi)) / \forall x\chi \\ \exists x(\psi / \chi) &\simeq (\exists x(\psi \wedge \chi) \vee \forall x(\psi \wedge \neg\chi)) / \exists x\chi\end{aligned}$$

The first equivalence gives us a way to transform atomic subformulae, and the rest show to pull  $/$  out of the scope of any logical operator—including other occurrences of  $/$  itself.

If we have transformed a formula into transplicative normal form, then the resulting transplicator will be a summing up, in a  $/$ -and- $\bowtie$ -free formula, of any presupposition introduced into the original formula by  $/$  or by  $\bowtie$ . (Some horrendously complicated transplicators can arise, but obvious simplifications will be possible particular cases.) Furthermore, since the transplicative is also  $/$ -and- $\bowtie$ -free, we can see that a single occurrence of  $/$  is sufficient for representing the overall content—the  $\top/\perp$ -conditions—of the original formula.

But if projection rules are the only thing you want to get, then observe that the equivalences for  $/$  and  $\bowtie$  may be brought in line with the others:

$$\begin{aligned} (\psi_1/\chi_1) / (\psi_2/\chi_2) &\simeq (\psi_1 \wedge \psi_2 \wedge \chi_1) / (\chi_1/\chi_2) \\ (\psi_1/\chi_1) \bowtie (\psi_2/\chi_2) &\simeq (\psi_1 \wedge \psi_2 \wedge (\chi_1 \leftrightarrow \chi_2)) / (\chi_1 \bowtie \chi_2). \end{aligned}$$

A procedure based on these equivalences will transform a formula  $\phi$  into  $\psi/\phi$ , where  $\psi$  sums up the overall presupposition, as before, but  $\phi$  is left to stand.

On the other hand, we may want to pin down a  $/$ -and- $\bowtie$ -free transpicand more tightly. Observe that a formula  $\psi \bowtie \chi$  in interjunctive normal form will be equivalent to  $(\chi \rightarrow \psi) / \psi$  and to  $(\chi \rightarrow \psi) / \chi$ , which are in transpicative normal form. (We can make do with  $\chi \rightarrow \psi$ , rather than  $\chi \leftrightarrow \psi$ , because  $\psi \vDash \chi$ .) The transpicand  $\psi$  then fixes  $\top$ -conditions, while the transpicand  $\chi$  fixes  $\perp$ -conditions. I shall leave it as an exercise to formulate equivalences on which to base a procedure for transforming a formula directly into a transpicative normal form of each of these special kinds: the equivalences given for  $\wedge$ ,  $\vee$ ,  $\forall$ , and  $\exists$  can be kept, but the others need to be revised.

#### 6.4 A Parenthesis on Description Terms

If we expand our languages to contain a term-forming descriptions operator  $\iota$ , and if we consider its interpretation in the kind of model we are working with, then the denotation conditions sketched in Section 1.1 turn out in the following way: for any model  $M$ , and any assignment  $s$ , if  $a \in D_M$ , then

$$M_s(\iota x\phi) = a \quad \text{iff} \quad M_{s(y|a)}(\forall x[x = y \leftrightarrow \phi]) = \top.$$

And  $M_s(\iota x\phi) = \otimes$  if there is no such  $a$ . (We are here assuming that  $y$  is a variable distinct from  $x$  and extraneous to  $\phi$ .) These denotation conditions can be spelt out to mean that if  $a \in D_M$ , then

$$M_s(\iota x\phi) = a \quad \text{iff} \quad \begin{cases} M_{s(x|a)}(\phi) = \top, \text{ and} \\ M_{s(x|b)}(\phi) = \perp, \text{ for every } b \in D_M \text{ not identical to } a. \end{cases}$$

Hence, to be the denotation of  $\iota x\phi$ ,  $a$  has to be *determinately* ‘the unique  $x$  such that  $\phi$ ’:  $\phi$  must be false, not just not true, when any other object in  $D_M$  is assigned to  $x$ .

But do we have to work with such a stringent form of uniqueness? In the present context we do, on pain of violating monotonicity. Notice that, according to our definition,  $M_s(\iota x\phi)$  is an element of  $D_M$  only if  $M_{s(x|a)}(\phi)$  is either  $\top$  or  $\perp$  for any  $a$  in  $D_M$ . This guarantees monotonicity for  $\iota$ -terms. If, for  $M_s(\iota x\phi)$  to be an element  $a$  of  $D_M$ , we were to require only that  $M_{s(x|a)}(\phi) = \top$  and that  $M_{s(x|b)}(\phi) \neq \top$  for any  $b$  in  $D_M$  distinct from  $a$ , then there might be a model  $N$  such that  $M \sqsubseteq N$  and  $N_{s(x|b)}(\phi) = \top$  for

some such  $b$ , in which case  $N_s(\iota x\phi)$  could not be  $a$  and monotonicity would have been violated. (For example, take  $M$  and  $N$  to be models interpreting a predicate symbol  $P$  over the domain  $\{0, 1\}$ , where  $P_M(0) = P_N(0) = P_N(1) = \top$  and  $P_M(1) = *$ —fill in other details as you like—and consider  $\iota xPx$ .)

Notice, then, that according to our definitions  $\iota x\phi$  may be non-denoting for two different kinds of reason: either (i) because  $\phi$  is not sufficiently defined to determine a denotation, or (ii) because  $\phi$  is sufficiently highly defined to rule out there being one. Case (i) arises when the formula  $\exists y\forall x(x = y \leftrightarrow \phi)$  is  $*$ , and case (ii) when it is  $\perp$ . If we had a subtler theory of identity and of the interpretation of ‘singular terms’, then subtler interpretations for  $\iota x\phi$  would be available. But this leads far beyond the simple kind of model we are working with.

The literature on description terms is vast and varied, but two approaches which it is interesting to compare and contrast with the present one occur in [Smiley 1960] and [Scott 1967]. Smiley entertains ‘neither-true-nor-false’ sentences, but he is unconstrained by monotonicity; while Scott treats non-denoting terms in a logic which, at sentence-level, is classical and total. In [Czermak 1974], on the other hand, there is a theory more like the one here. But it should be emphasized that our definitions do not involve any special ideas concerning the interpretation of description terms: they merely follow a path which was pre-determined once we embarked on partial logic as the logic of monotonic modes of composition.

The standard semantical definitions and lemmas of Sections 6.1 and 6.2 all extend in the obvious way to languages which contain  $\iota$ —due account being taken of the fact that terms, as well as formulae, may now contain ‘bound’ variables. And so we have a framework in which to address the question whether, having introduced  $\iota$ -terms, we can after all ‘eliminate’ them without decreasing the expressive power they provide. But what does this mean? There are various degrees of eliminability that we should distinguish. In a weak sense,  $\iota$  would be eliminable provided that any formula were equivalent to an  $\iota$ -free one. In a stronger sense of eliminability there would be some procedure which we could apply to transform a formula into an equivalent  $\iota$ -free one. But we should really hope for something stronger still: to be in possession of a general scheme of *scope-free* elimination. And this is something we can indeed obtain.

To signal one or more occurrence in a formula of a term  $\iota x\phi$  (possibly ignoring other occurrences of  $\iota x\phi$ ) we can always pick on some extraneous variable  $y$  and describe the formula as  $\psi(\iota x\phi/y)$ . And so we can take our goal to be to define a scheme  $I(x, \phi, y, \psi)$  which does not involve  $\iota$  and which, for any  $\phi$  and  $\psi$ , will yield a formula equivalent to  $\psi(\iota x\phi/y)$ , provided only that  $\iota x\phi$  is ‘substitutable for  $y$  in  $\psi$ ’—i.e., that no free occurrence of a variable in  $\iota x\phi$  becomes a bound occurrence in  $\psi(\iota x\phi/y)$ . Then we may read the scheme  $I(x, \phi, y, \psi)$  as ‘the  $x$  such that  $\phi$  is a  $y$  such that

$\psi$ ', and it will provide for the 'scope-free' elimination of  $\iota$ -terms simply because  $\iota$ -languages admit 'substitutivity of equivalents': when a subformula is replaced by an equivalent one an equivalent formula results. The point is that to eliminate a term  $\iota x\phi$  from a formula we can apply the scheme to any subformula  $\psi(\iota x\phi/y)$  which binds no variables occurring free in  $\iota x\phi$ . Moreover, to transform a formula into an entirely  $\iota$ -free one, we can apply the scheme to  $\iota$ -terms in any order we like, and (variable-binding permitting) different occurrences of the same term can be eliminated all at once, or one at a time, or in any combination we choose. Such a scheme will then exhibit a semantical scope-freedom which exactly matches the scope-freedom possessed by an  $\iota$ -term in virtue of its syntactic category.

In Section 4.2 we presented a 'Russellian' analysis for a definite-description quantifier  $\iota x[\cdot, \cdot]$ , but any thought that this could serve as the required elimination scheme is soon dispelled. The  $\top/\perp$ -conditions for  $\iota x[\cdot, \cdot]$  certainly give definite descriptions a fair degree of semantical scope-freedom—in particular, freedom with respect to negation—but it is not thorough-going. For example, if  $\chi$  is  $\top$ , then  $\iota x[\phi, \psi] \vee \chi$  has to be  $\top$ , though  $\iota x[\phi, \psi \vee \chi]$  might be  $*$ . This is not a defect of our analysis for  $\iota x[\cdot, \cdot]$ , since scope sensitivity can be important if we are considering natural language description idioms, but we have to look elsewhere for a scheme to go proxy for definite descriptions that are construed as terms. In fact,  $\iota x[\cdot, \cdot]$  would not even serve to eliminate  $\iota$ -terms from atomic formulae. This is because our monotonicity constraint is sufficiently liberal to allow sentences  $Pt_1 \cdots \iota x\phi \cdots t_n$  which are  $\top$  or  $\perp$  even when  $\iota x\phi$  is  $\otimes$ , though  $\iota x\phi$  is  $\otimes$  only if  $\exists y\forall x[x = y \leftrightarrow \phi]$  is not  $\top$ , in which case  $\iota x[\phi, Pt_1 \cdots x \cdots t_n]$  must be  $*$ .

It is not surprising, given this last observation, that our scheme of elimination will involve the logically non-denoting term  $\otimes$ . Let us abbreviate the formula  $\forall x[x = y \leftrightarrow \phi]$  as  $\phi(x!y)$ , then we could use either of the following as definitions of  $I(x, \phi, y, \psi)$ :

$$\begin{aligned} & \exists y[\phi(x!y) \wedge \psi] \vee [\forall y[\phi(x!y) \rightarrow \psi] \wedge \psi(\otimes/y)], \\ & \forall y[\phi(x!y) \rightarrow \psi] \wedge [\exists y[\phi(x!y) \wedge \psi] \vee \psi(\otimes/y)]. \end{aligned}$$

To see that these formulae work, it is just a matter of checking  $\top/\perp$ -conditions (with the aid of an extended version of Lemma 2) to show that they are equivalent to  $\psi(\iota x\phi/y)$ —assuming, that is, that  $\iota x\phi$  is substitutable for  $y$  in  $\psi$ .

We have emphasized that an elimination scheme of this kind allows us to dispense with the syntax of description terms as terms without disrupting any of the characteristics they manifest as such. But in fact this could be achieved much more cheaply: simply introduce a primitive mode of complex quantification  $Dx[\cdot, \cdot]$  interpreted so that  $\text{---}Dx[\phi, \cdots x \cdots]\text{---}$  will always mimic  $\text{---}(\cdots \iota x\phi \cdots)\text{---}$ . Stating explicit  $\top/\perp$ -conditions for  $Dx[\cdot, \cdot]$  is routine. What we should now emphasize is that our definitions for a scheme

of elimination go a stage further than this: they show how a quantifier  $Dx[\cdot, \cdot]$  may be analysed in terms of simple and basic logical vocabulary. In other words, we can do for  $Dx[\cdot, \cdot]$  what in Section 4.2 we did for  $!x[\cdot, \cdot]$ .

In the basic languages presented in Section 6.1, the displayed elimination schemes can of course be viewed as definitions—explicit definitions for a complex quantifier or, ‘contextual definitions’ for an  $\iota$ -term. And so we have a sense in which  $\iota$  is definable in terms of  $\otimes$ . Conversely, if we have  $\iota$ , then  $\otimes$  can be defined directly—for example, as  $\iota x \perp$ . Hence the presence of either  $\otimes$  or  $\iota$  provides equivalent expressive resources in a first-order language subject to the kind of interpretation we are considering. However, we cannot dispense with  $\otimes$  in  $\iota$ -free languages without a decrease in expressive power: the atomic sentence  $P\otimes$ , for example, is equivalent to no  $\otimes$ -free formula. (To see this consider models  $M$  and  $N$  with the singleton domain  $\{0\}$  such that  $P_M(0) = P_M(\otimes) = P_N(0)$  and  $P_N(\otimes) = *$ : if  $s(x) = 0$ , for all  $x \in Var$ , then for any  $\otimes$ -free formula  $\phi$ ,  $M_s(\phi) = N_s(\phi)$ , though  $M_s(P\otimes) \neq N_s(P\otimes)$ .) In the presence of  $\otimes$ , on the other hand, other vocabulary distinctive to partial logic could be dispensed with: given our interpretation of  $=$ ,  $*$  could be defined as  $\otimes = \otimes$ , and hence—as we showed in Section 4.1— $\&$  (and  $/$ ) could also be defined.

Although  $\otimes$  is not *logically* eliminable, it remains a possibility that it is in some sense eliminable in particular non-logical theories set up in partial logic: we shall mention a theorem about this in Section 7.3.

### 6.5 Semantic Consequence

To provide for a suitably powerful notion of semantic consequence, conceived along the lines suggested in Section 1.1, our basic definition is of what it is for a model  $M$  for a language  $\mathcal{L}$ , together with an assignment  $s$ , to *reject* a pair  $\langle \Gamma, \Delta \rangle$  of sets of formulae of  $\mathcal{L}$ . We shall say that  $(M, s)$  rejects  $\langle \Gamma, \Delta \rangle$  if and only if

- either: (i)  $M_s(\phi) = \top$  for all  $\phi \in \Gamma$  and  $M_s(\psi) \neq \top$  for all  $\psi \in \Delta$ ,  
or: (ii)  $M_s(\phi) \neq \perp$  for all  $\phi \in \Gamma$  and  $M_s(\psi) = \perp$  for all  $\psi \in \Delta$ .

And let us say that  $M$  (on its own) *rejects*, or *is a counter model to*,  $\langle \Gamma, \Delta \rangle$  when there is an  $s$  such that  $(M, s)$  rejects  $\langle \Gamma, \Delta \rangle$ . Then, if  $\mathcal{M}$  is any class of models for  $\mathcal{L}$ ,  $\vDash_{\mathcal{M}}$ —consequence in  $\mathcal{M}$ —is defined by

$$\Gamma \vDash_{\mathcal{M}} \Delta \quad \text{iff} \quad \text{no model in } \mathcal{M} \text{ rejects } \langle \Gamma, \Delta \rangle.$$

When  $\mathcal{M}$  is the class of all models for a given language, we just write ‘ $\vDash$ ’: this is *logical consequence*. Following the common notational practice with turnstiles, we shall ignore squiggly brackets and the empty set, and replace union signs by commas: for example, ‘ $\vDash \top, \phi, \Delta$ ’ means that  $\emptyset \vDash \{\top, \phi\} \cup \Delta$ .



In Section 1.1 we remarked on single-barrelled relations of consequence. Note the way in which  $*$  may now be deployed to capture such relations:

$$\begin{aligned}\Gamma \vDash_{\mathcal{M}} *, \Delta & \text{ iff no model in } \mathcal{M} \text{ satisfies condition (i) above,} \\ \Gamma, * \vDash_{\mathcal{M}} \Delta & \text{ iff no model in } \mathcal{M} \text{ satisfies condition (ii) above.}\end{aligned}$$

And  $\Gamma \vDash_{\mathcal{M}} \Delta$  if and only if both  $\Gamma \vDash_{\mathcal{M}} *, \Delta$  and  $\Gamma, * \vDash_{\mathcal{M}} \Delta$ . In fact this biconditional is just an instance of a quite general principle: for *any* formula  $\phi$ ,  $\Gamma \vDash_{\mathcal{M}} \Delta$  if and only if both  $\Gamma \vDash_{\mathcal{M}} \phi, \Delta$  and  $\Gamma, \phi \vDash_{\mathcal{M}} \Delta$ .

In Section 7.1 we shall present logical laws using *sequents*: these will be understood to be pairs of *finite* sets, for which we use the special notation ' $\Gamma \succ \Delta$ ' instead of ' $\langle \Gamma, \Delta \rangle$ '. And we shall mention sequents in the same style that we state facts about consequence, writing ' $\succ, \top, \phi, \Delta$ ', for example, to stand for  $\emptyset \succ \{\top, \phi\} \cup \Delta$ . When  $M$  is not a counter model to  $\Gamma \succ \Delta$  we shall say that  $M$  is a *model* of  $\Gamma \succ \Delta$ , or that  $\Gamma \succ \Delta$  *holds in*  $M$ . More generally, if  $\Sigma$  is a set of sequents,  $M$  will be said to be a *model of*  $\Sigma$  if and only if  $M$  is a model of every sequent in  $\Sigma$ ; and ' $\mathcal{K}(\Sigma)$ ' will be the notation for the class of all such models. (Note: 'model for  $\mathcal{L}$ ', 'model of'  $\Sigma$ ).

A sequent  $\Gamma \succ \Delta$  embodies a principle of consequence— $\Delta$ 's following from  $\Gamma$ . It is a principle of *logical* consequence if  $\Gamma \vDash \Delta$ , in which case it holds in all models, but there are sequents which hold in some models but not in others; and there are also sequents, such as  $\emptyset \succ \emptyset$ , which hold in none. A set  $\Sigma$  of sequents then embodies a collection of such principles, and  $\vDash_{\mathcal{K}(\Sigma)}$  is the relation of consequence semantically determined by them:

$$\Gamma \vDash_{\mathcal{K}(\Sigma)} \Delta \text{ iff no model of } \Sigma \text{ rejects } \langle \Gamma, \Delta \rangle.$$

Observe, then, that  $\vDash_{\mathcal{K}(\emptyset)}$  is logical consequence; and that  $\vDash_{\mathcal{K}(\{\emptyset \succ \emptyset\})}$  is the universal relation between sets of formulae.

Clearly, if  $\Gamma \succ \Delta$  is contained in  $\Sigma$ , then  $\Gamma \vDash_{\mathcal{K}(\Sigma)} \Delta$ ; but the converse does not generally hold:  $\emptyset$  is an obvious counterexample. When it does hold—when  $\Sigma = \{\Gamma \succ \Delta \mid \Gamma \vDash_{\mathcal{K}(\Sigma)} \Delta\}$ —of sequents which is closed under the sequent principles it determines, and our proof theoretical apparatus will be designed to pick out precisely such sets of sequents as what 'theories' are in partial logic. Thus we shall be adopting an extensional notion of a theory, not involving any particular axiomatization. Pure logic, for a given language, will be one such theory, viz.  $\{\Gamma \succ \Delta \mid \Gamma \vDash \Delta\}$ .

But  $\vDash_{\mathcal{K}(\Sigma)}$  is a full-blown consequence relation between arbitrary (not necessarily finite) sets of formulae, and we should demand of our proof system that it yield consequence relations  $\vdash_{\Sigma}$  to match  $\vDash_{\mathcal{K}(\Sigma)}$ . We shall produce a suitable definition which is 'sound and complete' in that, for any  $\Gamma$  and  $\Delta$ ,

$$\Gamma \vDash_{\mathcal{K}(\Sigma)} \Delta \text{ iff } \Gamma \vdash_{\Sigma} \Delta.$$

But then we shall be able to show that the relation  $\vDash_{\mathcal{K}(\Sigma)}$  does not actually go beyond the sequent principles—the finite principles of consequence—determined by  $\Sigma$ —in fact not beyond those determined by some finite subset of  $\Sigma$ . For the definition of  $\vdash_{\Sigma}$  will guarantee that  $\Gamma \vdash_{\Sigma} \Delta$  if and only if  $\Gamma_0 \vdash_{\Sigma_0} \Delta_0$  for some finite subsets  $\Gamma_0$  of  $\Gamma$ ,  $\Delta_0$  of  $\Delta$  and  $\Sigma_0$  of  $\Sigma$ ; so that  $\vDash_{\mathcal{K}(\Sigma)}$  too turns out to be finitary in this way. Contraposing, we could state the fact as a two-pronged form of compactness:

**THEOREM 8 (Compactness).** *There is a model of  $\Sigma$  which rejects  $\langle \Gamma, \Delta \rangle$  iff, for every finite subset  $\Gamma_0$  of  $\Gamma$ ,  $\Delta_0$  of  $\Delta$  and  $\Sigma_0$  of  $\Sigma$ , there is a model of  $\Sigma_0$  which rejects  $\langle \Gamma_0, \Delta_0 \rangle$ .*

Two complementary parallels with standard treatments of classical logic are now emerging, which pervade the development of partial logic. First, *pairs* of sets of formulae and their rejectability (by a model and an assignment) play a role which *single* sets of formulae and their satisfiability (by a model and an assignment) usually play in classical logic. Secondly, sets of *sequents* and their models play the part which sets of *sentences* and their models play in classical logic. But why should things turn out like this?

It has already been explained—in Section 1.3—that principles of logical consequence cannot be summed up in terms of the truth of sentences, but the irreducibility of consequence to truth extends further than this. For, given a sequent  $\Gamma \succ \Delta$ , it is not in general possible to find a sentence  $\sigma$  such that  $M$  is a model of  $\Gamma \succ \Delta$  if and only if  $M(\sigma) = T$ —equivalently, if and only if  $M$  is a model of  $\succ \sigma$ . (Moreover, if there is no sentence, then there is no formula of any kind to play this role; since, if there were a formula  $\phi$  then a suitable sentence could be obtained by substituting  $\otimes$  for all parameters in  $\phi$ .) This contrasts with classical logic, in which a sequent  $\Gamma \succ \Delta$  can always be summed up in the sentence  $\forall \vec{x}[\mathbb{M}\Gamma \rightarrow \mathbb{W}\Delta]$ , where  $\mathbb{M}\Gamma$  is the conjunction of elements of  $\Gamma$ ,  $\mathbb{W}\Delta$  is the disjunction of elements of  $\Delta$ , and  $\forall \vec{x}$  binds all free variables.

We can extend and strengthen this point about partial logic: given a set  $\Sigma$  of sequents it is not in general possible to find a corresponding set  $\Sigma'$  of sequents of the truth-expressing form  $\succ \sigma$  such that  $M$  is a model of  $\Sigma$  if and only if  $M$  is a model of  $\Sigma'$ . To see this observe that *if* we can find such a  $\Sigma'$ , then  $\mathcal{K}(\Sigma)$  satisfies the following closure condition—because  $\mathcal{K}(\Sigma')$  obviously does and  $\mathcal{K}(\Sigma) = \mathcal{K}(\Sigma')$ .

$$(\dagger) \quad \text{If } M \in \mathcal{K}(\Sigma) \text{ and } M \sqsubseteq_e N, \text{ then } N \in \mathcal{K}(\Sigma).$$

In fact we could use the Compactness Theorem to show that  $(\dagger)$  is a sufficient, as well as a necessary, condition for finding such a  $\Sigma'$ . But the present point depends on its being necessary: we just have to produce a  $\Sigma$  such that  $\mathcal{K}(\Sigma)$  does not satisfy  $(\dagger)$ . A simple example would be  $\{P \otimes \succ Q \otimes, *\}$ : checking this is essentially an exercise in propositional logic.

Although the principles of consequence that arbitrary sequents express cannot be reduced to the *truth* of sentences, still, can we at least make do with parameter-free sequents, which contain only sentences, not arbitrary formulae? No, we cannot. Let us argue in the same pattern as before: the following is obviously a necessary condition (and in fact also a sufficient condition) for there being a set  $\Sigma'$  of parameter free sequents such that  $\mathcal{K}(\Sigma) = \mathcal{K}(\Sigma')$ .

$$(\ddagger) \quad \text{If } M \in \mathcal{K}(\Sigma) \text{ and } M \sim_e N, \text{ then } N \in \mathcal{K}(\Sigma).$$

However,  $\{Px \succ Qx, *\}$ , for example, does not satisfy  $(\ddagger)$ —though it is more involved to check out this example than the previous one. This is perhaps a little surprising: it means that the relation  $\sim_e$  of ‘elementary equivalence’ between models is a strictly weaker relation than the relation of being a model of the same sequents.

Anyhow, let us return to the relation  $\vDash$  of logical consequence. This has been defined relative to a particular language  $\mathcal{L}$ , but, as in classical logic, it is in fact an absolute notion, in the sense that  $\Gamma \vDash \Delta$  in  $\mathcal{L}_1$  if and only if  $\Gamma \vDash \Delta$  in  $\mathcal{L}_2$ , whenever the formulae in  $\Gamma$  and  $\Delta$  are formulae of both  $\mathcal{L}_1$  and  $\mathcal{L}_2$ . In particular,  $\Gamma \vDash \Delta$  in any given language if and only if  $\Gamma \vDash \Delta$  in the language containing no non-logical vocabulary other than that occurring in  $\Gamma$  or  $\Delta$ . Observe too that the relations of equivalence ( $\simeq$ ), degree-of-definedness ( $\sqsubseteq$ ) and compatibility ( $\sqcap$ ), which we defined in Section 6.2, are absolute in this sense. These facts are easy to check, using the notion of the *reduct*  $M \upharpoonright \mathcal{L}'$  of a model  $M$  for  $\mathcal{L}$  to a smaller language  $\mathcal{L}'$ :  $M \upharpoonright \mathcal{L}'$  is the model for  $\mathcal{L}'$  which has the same domain as  $M$  and interprets the vocabulary of  $\mathcal{L}'$  in the same way as  $M$ , just ignoring any vocabulary in  $\mathcal{L}$  but not in  $\mathcal{L}'$ . We shall use this definition later on, and we shall also talk of *expanding* a model  $M$  for  $\mathcal{L}$  to a model  $N$  for a bigger language  $\mathcal{L}^+$  when  $M = N \upharpoonright \mathcal{L}$ .

The absoluteness of  $\vDash$  means that we can state the following theorem without reference to any particular language (though its proof—in Section 7.3—will depend on being very finicky about languages).

**THEOREM 9 (Craig Interpolation).** *If  $\phi \vDash \psi$ , then  $\phi \vDash \lambda$  and  $\lambda \vDash \psi$  for some formula  $\lambda$  which contains no non-logical vocabulary which does not occur both in  $\phi$  and in  $\psi$ .*

It is noteworthy that there is an analogous result for degree-of-definedness: if  $\phi \sqsubseteq \psi$ , then  $\phi \sqsubseteq \lambda$  and  $\lambda \sqsubseteq \psi$  for an interpolant  $\lambda$  subject to the same constraint.

## 7 FIRST-ORDER PARTIAL THEORIES

## 7.1 Logical Laws

It will be neatest to take our logical laws as directly definitive of what a ‘theory’ is. The laws will be in the form of sequent axioms and sequent rules, and a *theory*, in a given language  $\mathcal{L}$ , is defined to be a set of sequents of  $\mathcal{L}$  which contains the sequent axioms and is closed under the sequent rules, in the sense that if the ‘premise(s)’ of a rule is (are) in the set then so is its ‘conclusion’. ‘Proofs’ then enter the picture in the following way. If, given a set  $\Sigma$  of sequents of  $\mathcal{L}$ , we define  $\bar{\Sigma}$  to be the intersection of all theories in  $\mathcal{L}$  which contain  $\Sigma$ , then  $\bar{\Sigma}$  will be a theory—the ‘smallest’ theory in  $\mathcal{L}$  containing  $\Sigma$ —and a sequent will be contained in  $\bar{\Sigma}$  if and only if there is a sequent proof of it from a finite subset of  $\Sigma$ . That things fit together in this way is just part of the general theory of inductive definitions (see for example [Aczel 1977]). We shall call  $\bar{\Sigma}$  the theory *axiomatised by*  $\Sigma$ ; and  $\Sigma$  will already be a theory if and only if  $\Sigma = \bar{\Sigma}$ . Pure logic, for a given language  $\mathcal{L}$ , then slots into place as the smallest theory in  $\mathcal{L}$ , viz.  $\bar{\emptyset}$ .

The first three laws are general principles of consequence, which we label after [Scott 1973b]: a basic axiom scheme (R), a (double) rule of thinning (M), and cut (T).

$$\begin{array}{l}
 \text{(R)} \qquad \qquad \phi \succ \phi \\
 \\
 \text{(M)} \qquad \frac{\Gamma \succ \Delta}{\Gamma \succ \phi, \Delta} \qquad \frac{\Gamma \succ \Delta}{\Gamma, \phi \succ \Delta} \\
 \\
 \text{(T)} \qquad \frac{\Gamma \succ \phi, \Delta \qquad \Gamma, \phi \succ \Delta}{\Gamma \succ \Delta}
 \end{array}$$

Clearly any instance of (R) will hold in any model, and if the ‘premise(s)’ of an instance of (M) or (T) hold in a model, then the ‘conclusion’ holds in that model. Hence individually these laws are ‘sound’. It will be left unsaid that all the remaining axioms and rules are individually sound in the same way: this can be checked using the definitions and lemmas of Section 6.1.

The next rule is a general rule of (S) of substitution. When  $\Theta$  is a set of formulae, we use ‘ $\Theta(t/x)$ ’ to stand for  $\{\theta(t/x) \mid \theta \in \Theta\}$ .

$$\text{(S)} \qquad \frac{\Gamma \succ \Delta}{\Gamma(t/x) \succ \Delta(t/x)}$$

This holds provided that the term  $t$  is substitutable for  $x$  in all the formulae in  $\Gamma$  and  $\Delta$  (see Section 6.1). In the presence of this rule we shall be able to specify the quantifier and identity laws with parameters, instead of using schematic letters for terms.

For propositional laws we can use the following. Double lines means the rule applies upwards as well as downwards, and ‘ $\neg\Theta$ ’ stands for  $\{\neg\theta \mid \theta \in \Theta\}$ .

$$\begin{array}{c}
\begin{array}{cc}
\triangleright \top & \perp \triangleright \\
\neg * \triangleright * & * \triangleright \neg * \\
\phi, \neg\phi \triangleright * & * \triangleright \neg\phi, \phi \\
\frac{\neg\Gamma \triangleright \Delta}{\neg\Delta \triangleright \Gamma} & \frac{\Gamma \triangleright \neg\Delta}{\Delta \triangleright \neg\Gamma} \\
\frac{\Gamma, \phi, \psi \triangleright \Delta}{\Gamma, \phi \wedge \psi \triangleright \Delta} & \frac{\Gamma \triangleright \phi, \psi, \Delta}{\Gamma \triangleright \phi \vee \psi, \Delta} \\
\frac{\Gamma, \phi, \psi \triangleright *, \Delta}{\Gamma, \phi \bowtie \psi \triangleright *, \Delta} & \frac{\Gamma, * \triangleright \phi, \psi, \Delta}{\Gamma, * \triangleright \phi \bowtie \psi, \Delta}
\end{array}
\end{array}$$

Observe how  $*$  may be deployed to cancel one or the other half of our double-barrelled notion of consequence. Thus, in particular, the rules for interjunction match  $\bowtie$  with  $\wedge$  for  $\top$ -conditions and with  $\vee$  for  $\top$ -conditions.

From these laws we can immediately deduce some further fundamental principles (which could be swapped in various obvious ways to provide alternative sets of propositional laws):

$$\begin{array}{c}
\begin{array}{cc}
\phi \triangleright \neg\neg\phi & \neg\neg\phi \triangleright \phi \\
\phi, \neg\phi \triangleright \neg\psi, \psi \\
\frac{\Gamma \triangleright \Delta}{\neg\Delta \triangleright \neg\Gamma} \\
\begin{array}{cc}
\phi \wedge \psi \triangleright \phi & \phi \triangleright \phi \vee \psi \\
\phi \wedge \psi \triangleright \psi & \psi \triangleright \phi \vee \psi \\
\phi, \psi \triangleright \phi \wedge \psi & \phi \vee \psi \triangleright \phi, \psi \\
\phi \bowtie \psi \triangleright \phi, * & *, \phi \triangleright \phi \bowtie \psi \\
\phi \bowtie \psi \triangleright \psi, * & *, \psi \triangleright \phi \bowtie \psi \\
\phi, \psi \triangleright \phi \bowtie \psi & \phi \bowtie \psi \triangleright \phi, \psi
\end{array}
\end{array}
\end{array}$$

Let us now adopt the abbreviation ‘ $\Gamma \xrightarrow{\phi} \Delta$ ’ for ‘ $\Gamma, \phi \triangleright \neg\phi, \Delta$ ’. The

force of such sequents can be expressed informally as ‘when  $\phi$  is true, then  $\Delta$  follows from  $\Gamma$ ’: recall the discussion at the end of Section 1.1. Then for quantifiers we can use the following up-and-down rules, subject to the proviso that  $x$  does not occur free in  $\Gamma$  or in  $\Delta$ :

$$\frac{\Gamma \succ^{x=x} \phi, \Delta}{\Gamma \succ \forall x \phi, \Delta} \qquad \frac{\Gamma, \phi \succ^{x=x} \Delta}{\Gamma, \exists x \phi \succ \Delta}$$

The proviso is only of importance for the downward rules, but given (S) its presence does not hamper the upward ones, which are equivalent to the following axioms:

$$\forall x \phi \succ^{x=x} \phi \qquad \phi \succ^{x=x} \exists x \phi.$$

Notice how  $x = x$  is here playing the role of an ‘existence predicate’.

Of course,  $x = x$  can never actually be false, and so we include the following axiom:

$$* \succ x = x.$$

And to capture the determinateness of identity:

$$x = x, y = y \succ x = y, \neg x = y.$$

For the substitutivity of identicals we adopt the following scheme, which means that whenever  $x = y$  is true, then occurrences of  $x$  and  $y$  can be shuffled around in a formula in any way you like:

$$\phi(x/u, y/v) \succ^{x=y} \phi(y/u, x/v).$$

However a further substitutivity principle is required to govern non-denoting terms:

$$\phi(x/z) \succ x = x, \phi(y/z).$$

Since parameters are schematic for terms, the force of this is that a non-denoting term can be replaced by any term without affecting the truth value of a formula, if it already has one.

If we were envisaging subtler theories of identity these laws would need to be modified, but in the present context they capture our semantics of monotonic composition, once we include an axiom for the logically non-denoting term:

$$x = \otimes \succ *.$$

There is room for variation in the choice of primitive laws for identity; but let us adopt these. We can then go on to derive a characteristic principle for  $\otimes$ , whose effect is that if a formula is true (or false), then it remains so on making any substitution for an occurrence of  $\otimes$ :

$$\phi(\otimes/x) \succ \phi(y/x), *.$$

And other basic laws are easily obtained; for example, the symmetry of identity and distinctness:

$$x = y \succ y = x,$$

and the transitivity of identity:

$$x = y, y = z \succ x = z, *.$$

Observe that  $*$  cannot be taken away here: if  $y$  assigned no object, then neither of the left-hand formulae can be false, even if  $x = z$  is. However, we can easily derive a general principle to handle distinctness as well as identity:

$$x = y, y = z \succ \overline{y=y} \overline{x = z}.$$

The laws we have given provide the definition of a *theory* (in  $\mathcal{L}$ ) and of the *theory*  $\overline{\Sigma}$  (in *la*) *axiomatised by*  $\Sigma$ , in the way explained at the outset. Furthermore between arbitrary sets  $\Gamma$  and  $\Delta$  of formulae of a language  $\mathcal{L}$  we can define the consequence relation  $\vdash_{\Sigma}$ , demanded in Section 6.5, by stipulating that  $\Gamma \vdash_{\Sigma} \Delta$  if and only if, for some finite subsets  $\Gamma_0$  of  $\Gamma$  and  $\Delta_0$  of  $\Delta$ ,  $\Gamma_0 \succ \Delta_0 \in \overline{\Sigma}$ . This will be if and only if there is a proof of  $\Gamma_0 \succ \Delta_0$  from some finite subset  $\Sigma_0$  of  $\Sigma$ —hence if and only if  $\Gamma_0 \succ \Delta_0 \in \overline{\Sigma_0}$ . Thus  $\vdash_{\Sigma}$  turns out to be finitary in the way announced in Section 6.5. Note that, although the definitions of  $\overline{\Sigma}$  and  $\vdash_{\Sigma}$  are relative to a particular language  $\mathcal{L}$ , a given set  $\Sigma$  of sequents will always be a set of sequents of (infinitely) many different languages. This means that, on its own, our notation is radically ambiguous, and we need to be careful when more than one language is in play.

Since our laws are individually sound, it is easy to check that no model of  $\Sigma$  can be a counter model to any sequent in  $\overline{\Sigma}$ : in other words, not just is it the case that  $\mathcal{K}(\overline{\Sigma}) \subseteq \mathcal{K}(\Sigma)$ , but  $\mathcal{K}(\overline{\Sigma}) = \mathcal{K}(\Sigma)$ . And the following theorem, which makes reference to arbitrary sets  $\Gamma$  and  $\Delta$ , is a trivial extension of this fact:

**THEOREM 10 (Soundness).** *If  $\Gamma \vdash_{\Sigma} \Delta$  then  $\Gamma \vDash_{\mathcal{K}(\Sigma)} \Delta$ .*

The converse, guaranteeing that  $\vdash_{\Sigma}$  coincides with the semantically defined relation  $\vDash_{\mathcal{K}(\Sigma)}$ , is rather more difficult to establish:

**THEOREM 11 (Completeness).** *If  $\Gamma \vDash_{\mathcal{K}(\Sigma)} \Delta$  then  $\Gamma \vdash_{\Sigma} \Delta$ .*

We shall turn our attention to the proof of completeness in Sections 7.2 and 7.3.

It is easy to see that  $\overline{\overline{\Sigma}} = \overline{\Sigma}$ , and so  $\vdash_{\overline{\Sigma}}$  is the same relation as  $\vdash_{\Sigma}$ . Also, given soundness,  $\vDash_{\mathcal{K}(\overline{\Sigma})}$  is the same relation as  $\vDash_{\mathcal{K}(\Sigma)}$ . Hence we would lose nothing by stating Theorem 11 with  $\Sigma$  restricted to theories. We *would* lose something if we restricted  $\Gamma$  and  $\Delta$  to finite sets, viz. being able to deduce as a corollary the full version of compactness stated in Theorem 8. But note

that when we do consider just finite sets of formulae, then Theorems 10 and 11 may be wrapped up together into the following equation:  $\bar{\Sigma} = \{\Gamma \succ \Delta \mid \Gamma \vDash_{\mathcal{K}(\Sigma)} \Delta\}$ .

These remarks put us in a position to convert the discussion in Section 6.5 of the conditions labelled (†) and (‡) into facts about theories. We may deduce from that discussion that a theory  $\Sigma$  is axiomatizable by sequents of the form  $\succ \sigma$  if and only if  $\mathcal{K}(\Sigma)$  is closed under the relation  $\sqsubseteq_e$ , in the sense of condition (†), and that a theory  $\Sigma$  is axiomatizable by parameter-free sequents if and only if  $\mathcal{K}(\Sigma)$  is closed under the relation  $\sim_e$ , in the sense of condition (‡). And there are various other results along these lines: necessary and sufficient conditions for a theory's being axiomatizable by sequents of a given kind are provided by specifying closure conditions on the class of its models.

In connection with soundness and completeness we should also think about 'consistency'. We have no use for a notion of the consistency of a set of formulae, but it makes sense to ask about the consistency of a set of sequents. Let us say that  $\Sigma$  is *consistent* if and only if  $\emptyset \succ \emptyset \notin \bar{\Sigma}$ . And '*inconsistent*' will just mean not consistent. Hence we may also define relational notions:  $\Sigma_1$  is *(in)consistent with*  $\Sigma_2$  if and only if  $\Sigma_1 \cup \Sigma_2$  is (in)consistent (which in turn makes sense of the words ' $\psi$ 's following from  $\phi$  is inconsistent with ...', used in Section 2.7: this means that  $\{\phi \succ \psi\}$  is inconsistent with ...). By rule (M), it follows that  $\Sigma$  is consistent if and only if  $\bar{\Sigma}$  does not contain all sequents (of the language in question). It also follows, by Theorems 10 and 11, that  $\Sigma$  is consistent if and only if  $\Sigma$  has a model, since the statement that  $\Sigma$  is *inconsistent* if and only if  $\Sigma$  has *no* model is just the special case of soundness and completeness when  $\Gamma$  and  $\Delta$  are both empty. On the other hand, the special case of Theorems 10 and 11 when  $\Sigma$  is empty gives the soundness and completeness of an axiomatization of the relation  $\vdash_{\emptyset}$  of logical consequence (for which we shall just write ' $\vdash$ '). Happily the *theory* thus axiomatised, viz.  $\bar{\emptyset}$ , turns out to be consistent, according to our definition, since there will be models of  $\emptyset$ —and hence too of  $\bar{\emptyset}$ —in great abundance.

It is noteworthy that to axiomatize pure logic we could abandon the system presented here and instead use a cut-free sequent calculus that has 'introduction rules' only. (See Sundholm's chapter in Volume 2 of the second edition of this *Handbook*.) One way to proceed would be to have 'negative' rules as well as 'positive' rules—the negative rules for vocabulary in the immediate scope of negation. In [Cleave 1974] there are rules of this kind that we could use for classical vocabulary, but to handle interjunction we need to include the following three-premise rules.



$$\begin{array}{c}
\frac{\Gamma, \phi, * \succ \Delta \quad \Gamma, \phi, \psi \succ \Delta \quad \Gamma, *, \psi \succ \Delta}{\Gamma, \phi \mathbb{X} \psi \succ \Delta} \\
\\
\frac{\Gamma \succ \phi, *, \Delta \quad \Gamma \succ \phi, \psi, \Delta \quad \Gamma \succ *, \psi, \Delta}{\Gamma \succ \phi \mathbb{X} \psi, \Delta} \\
\\
\frac{\Gamma, \neg\phi, * \succ \Delta \quad \Gamma, \neg\phi, \neg\psi \succ \Delta \quad \Gamma, *, \neg\psi \succ \Delta}{\Gamma, \neg[\phi \mathbb{X} \psi] \succ \Delta} \\
\\
\frac{\Gamma \succ \neg\phi, *, \Delta \quad \Gamma \succ \neg\phi, \neg\psi, \Delta \quad \Gamma \succ *, \neg\psi, \Delta}{\Gamma \succ \neg[\phi \mathbb{X} \psi], \Delta}
\end{array}$$

## 7.2 Model-Existence Theorems

Wrapping Soundness and Completeness up together, contraposing, and spelling out ‘ $\Gamma \not\vdash_{\mathcal{K}(\Sigma)} \Delta$ ’ we have that

$$\Gamma \not\vdash_{\Sigma} \Delta \quad \text{iff} \quad \text{there is a model of } \Sigma \text{ which rejects } \langle \Gamma, \Delta \rangle.$$

(The line through the turnstiles signifies negation.) We could then establish completeness (‘only if’) by adopting a Henkin-style strategy to boost up any pair  $\langle \Gamma, \Delta \rangle$  such that  $\Gamma \not\vdash_{\Sigma} \Delta$  to an exhaustive pair  $\langle \Gamma^+, \Delta^+ \rangle$  of sets of sentences of an extended language, from which we could then read off a model rejecting  $\langle \Gamma, \Delta \rangle$ . But this strategy can be elaborated to yield much more powerful model-existence results: kinds of interpolation theorem. We can then go on to deduce the Completeness Theorem and a lot more besides—facts both about pure logic and about non-logical theories.

To introduce the idea, consider the following set up:— $\Sigma_1$  is a set of sequents of a language  $\mathcal{L}_1$ , and  $\Gamma_1$  and  $\Delta_1$  are sets of formulae of  $\mathcal{L}_1$ ;  $\Sigma_2$  is a set of sequents of a language  $\mathcal{L}_2$ , and  $\Gamma_2$  and  $\Delta_2$  are sets of formulae of  $\mathcal{L}_2$ ; and  $\Lambda$  is a set of formulae common to both  $\mathcal{L}_1$  and  $\mathcal{L}_2$ . We can then ask:

$$\text{Is there a } \lambda \in \Lambda \text{ such that } \Gamma_1 \vdash_{\Sigma_1} \lambda, \Delta_1 \text{ and } \Gamma_2, \lambda \vdash_{\Sigma_2} \Delta_2?$$

(We may suppose that  $\vdash_{\Sigma_1}$  is defined relative to  $\mathcal{L}_1$  and  $\vdash_{\Sigma_2}$  relative to  $\mathcal{L}_2$ .) Notice that, provided  $\Lambda$  is non-empty, this is a generalization of the question ‘Is it the case that  $\Gamma \vdash_{\Sigma} \Delta$ ?’. For if  $\Sigma = \Sigma_1 = \Sigma_2$ ,  $\Gamma = \Gamma_1 = \Gamma_2$ , and  $\Delta = \Delta_1 = \Delta_2$ , then, by rules (M) and (T), the two questions must have the same answer.

And our interpolation theorems may be seen as generalizations of the Completeness Theorem, because they state that the answer ‘no’ to certain questions of the displayed form entails the existence of a pair of models  $M_1$  of  $\Sigma_1$  and  $M_2$  of  $\Sigma_2$  such that  $M_1$  rejects  $\langle \Gamma_1, \Delta_1 \rangle$ ,  $M_2$  rejects  $\langle \Gamma_2, \Delta_2 \rangle$  and

$M_1$  and  $M_2$  are related in a particular specified way: different ways for  $M_1$  and  $M_2$  to be related correspond to different assumptions about  $\Lambda$ . We also have corresponding generalizations of the Soundness Theorem, since the non-existence of an interpolant will be necessary as well as sufficient for the existence of a suitably related pair of models. But necessity is not as interesting as sufficiency; it gives us nothing new: it will always be immediately deducible from soundness.

To give a taste for all this, I shall develop a little way the case where, in the set up described,  $\Lambda$  is the set of all formulae of some sublanguage  $\mathcal{L}$  of  $\mathcal{L}_1$  and of  $\mathcal{L}_2$ . This is a simple and straightforward case, but even so we shall be able to deduce quite a lot from it.

First, to specify appropriate relationships between models, we need a generalization of the relations  $\sqsubseteq$  and  $\square$  defined in Section 6.2: if  $M_1$  is a model for  $\mathcal{L}_1$  and  $M_2$  is a model for  $\mathcal{L}_2$ , then there are relations of degree-of-definedness ( $\sqsubseteq_{\mathcal{L}}$ ) and of compatibility ( $\square_{\mathcal{L}}$ ) *relative to the vocabulary of a common sublanguage  $\mathcal{L}$* . With the notion of a reduct at hand (see Section 6.5), we can define the relations like this:

$$\begin{aligned} M_1 \sqsubseteq_{\mathcal{L}} M_2 & \text{ iff } M_1 \upharpoonright \mathcal{L} \sqsubseteq M_2 \upharpoonright \mathcal{L}, \\ M_1 \square_{\mathcal{L}} M_2 & \text{ iff } M_1 \upharpoonright \mathcal{L} \square M_2 \upharpoonright \mathcal{L}. \end{aligned}$$

Next observe that the claim that an interpolant exists can be analysed as the conjunction of three separate interpolant-existence claims:

LEMMA 12 (Combination Lemma).

*There is a  $\lambda \in \Lambda$  such that  $\Gamma_1 \vdash_{\Sigma_1} \lambda, \Delta_1$  and  $\Gamma_2, \lambda \vdash_{\Sigma_2} \Delta_2$*

*iff the following all hold:*

- (1) *there is a  $\lambda_1 \in \Lambda$  such that  $\Gamma_1 \vdash_{\Sigma_1} *, \lambda_1, \Delta_1$  and  $\Gamma_2, \lambda_1 \vdash_{\Sigma_2} *, \Delta_2$ ;*
- (2) *there is a  $\lambda_2 \in \Lambda$  such that  $\Gamma_1, * \vdash_{\Sigma_1} \lambda_2, \Delta_1$  and  $\Gamma_2, \lambda_2, * \vdash_{\Sigma_2} \Delta_2$ ;*
- (3) *there is a  $\lambda_3 \in \Lambda$  such that  $\Gamma_1 \vdash_{\Sigma_1} *, \lambda_3, \Delta_1$  and  $\Gamma_2, \lambda_3, * \vdash_{\Sigma_2} \Delta_2$ .*

‘Only if’ is trivial: put  $\lambda_1 = \lambda_2 = \lambda_3 = \lambda$ . For ‘if’ it is straightforward to check that we may take  $\lambda = [[\lambda_1 \wedge \lambda_3] \vee \lambda_2] \wp [\lambda_1 \wedge [\lambda_3 \vee \lambda_2]]$ .

We shall sketch a proof of a model-existence result that is in fact split up into three parallel theorems, corresponding to the three cases above: Theorem 13. But the Combination Lemma will show how they can be combined into one: Theorem 14. So there are two theorems to state. The assumptions common to both are that  $\mathcal{L}$  is a sublanguage of  $\mathcal{L}_1$  and of  $\mathcal{L}_2$ , and  $\Lambda$  is the set of all formulae of  $\mathcal{L}$ ; that  $\Gamma_1$  and  $\Delta_1$  are sets of formulae and  $\Sigma_1$  a set of sequents of a language  $\mathcal{L}_1$ ; and that  $\Gamma_2$  and  $\Delta_2$  are sets of formulae and  $\Sigma_2$  is a set of sequents of a language  $\mathcal{L}_2$ .

THEOREM 13 (Interpolant-Excluding Model Pairs: split-up version).

In each of the three cases

$$(1) * \in \Delta_1 \cap \Delta_2, \quad (2) * \in \Gamma_1 \cap \Gamma_2, \quad (3) * \in \Delta_1 \cap \Gamma_2,$$

there is no  $\lambda \in \Lambda$  such that  $\Gamma_1 \vdash_{\Sigma_1} \lambda, \Delta_1$  and  $\Gamma_2, \lambda \vdash_{\Sigma_2} \Delta_2$

iff

there are models  $M_1$  of  $\Sigma_1$  and  $M_2$  of  $\Sigma_2$ , with a common domain and assignments  $s_1$  and  $s_2$  such that  $(M_1, s_1)$  rejects  $\langle \Gamma_1, \Delta_1 \rangle$ ,  $(M_2, s_2)$  rejects  $\langle \Gamma_2, \Delta_2 \rangle$ , and

in case (1),  $M_1 \sqsubseteq_{\mathcal{L}} M_2$  and  $s_1 \sqsubseteq s_2$ ;

in case (2),  $M_2 \sqsubseteq_{\mathcal{L}} M_1$  and  $s_2 \sqsubseteq s_1$ ;

in case (3),  $M_1 \sqsubset_{\mathcal{L}} M_2$  and  $s_1 \sqsubset s_2$ .

THEOREM 14 (Interpolant-Excluding Model Pairs: combined version).

There is no  $\lambda \in \Lambda$  such that  $\Gamma_1 \vdash_{\Sigma_1} \lambda, \Delta_1$  and  $\Gamma_2, \lambda \vdash_{\Sigma_2} \Delta_2$

iff

there are models  $M_1$  of  $\Sigma_1$  and  $M_2$  of  $\Sigma_2$ , with a common domain and assignments  $s_1$  and  $s_2$  such that at least one of the following holds:

$$(1) \quad M_1 \sqsubseteq_{\mathcal{L}} M_2, \quad s_1 \sqsubseteq s_2, \quad \text{and} \quad \begin{cases} (M_1, s_1) \text{ rejects } \langle \Gamma_1, \{*\} \cup \Delta_1 \rangle, \\ (M_2, s_2) \text{ rejects } \langle \Gamma_2, \{*\} \cup \Delta_2 \rangle; \end{cases}$$

$$(2) \quad M_2 \sqsubseteq_{\mathcal{L}} M_1, \quad s_2 \sqsubseteq s_1, \quad \text{and} \quad \begin{cases} (M_1, s_1) \text{ rejects } \langle \Gamma_1 \cup \{*\}, \Delta_1 \rangle, \\ (M_2, s_2) \text{ rejects } \langle \Gamma_2 \cup \{*\}, \Delta_2 \rangle; \end{cases}$$

$$(3) \quad M_1 \sqsubset_{\mathcal{L}} M_2, \quad s_1 \sqsubset s_2, \quad \text{and} \quad \begin{cases} (M_1, s_1) \text{ rejects } \langle \Gamma_1, \{*\} \cup \Delta_1 \rangle, \\ (M_2, s_2) \text{ rejects } \langle \Gamma_2 \cup \{*\}, \Delta_2 \rangle. \end{cases}$$

It is now easy to see that the split-up version together with the Combination Lemma entails the combined version; and it is easy to check directly—from basic definitions—that the combined version entails the split-up version. Some applications can appeal directly to just one of the three cases of the split-up version, but most will invoke the combined one.

Now we sketch a proof—in its barest outlines—of Theorem 13. ‘If’ follows easily from soundness in each of the three cases. ‘Only if’ is non-trivial, but the main construction is the same in each case: distinguishing between them comes only at the very end.

First, then, take two disjoint sets  $C$  and  $D$  of new constants, where  $C$  is denumerable, and the cardinality of  $D$  is the maximum of the cardinalities

of the the two languages  $\mathcal{L}_1$  and  $\mathcal{L}_2$ ; and take some one-one function  $\pi$  from  $Var$  onto  $C$ . Now let  $\mathcal{L}_1^+$  and  $\mathcal{L}_2^+$  be the extensions of  $\mathcal{L}_1$  and  $\mathcal{L}_2$  got by taking  $C \cup D$  as additional constants; and let  $\Lambda^+$  be the set of all sentences obtained from a formula in  $\Lambda$  by making any substitution of constants from  $C \cup D$  for the parameters (so the sentences in  $\Lambda^+$  will be common to both  $\mathcal{L}_1^+$  and  $\mathcal{L}_2^+$ ). And finally, some notation: if  $\theta$  is a formula,  $\pi\theta$  is the formula obtained by substituting  $\pi(x)$  for all free occurrences of  $x$ ; and if  $\Theta$  is a set of formulae,  $\pi\Theta = \{\pi\theta \mid \theta \in \Theta\}$ .

Assuming that there is no  $\lambda \in \Lambda$  such that  $\Gamma_1 \vdash_{\Sigma_1} \lambda, \Delta_1$  and  $\Gamma_2, \lambda \vdash_{\Sigma_2} \Delta_2$ , it is now fairly easy to deduce that

there is no  $\lambda \in \Lambda^+$  such that  $\pi\Gamma_1 \vdash_{\Sigma_1} \lambda, \pi\Delta_1$  and  $\pi\Gamma_2, \lambda \vdash_{\Sigma_2} \pi\Delta_2$ ,

where  $\vdash_{\Sigma_1}$  and  $\vdash_{\Sigma_2}$  are now defined relative to the extended languages  $\mathcal{L}_1^+$  and  $\mathcal{L}_2^+$ , rather than  $\mathcal{L}_1$  and  $\mathcal{L}_2$ . The hard work is then to provide a construction that achieves the following. First,  $\pi\Gamma_1$ ,  $\pi\Delta_1$ ,  $\pi\Gamma_2$ , and  $\pi\Delta_2$  are extended to sets  $\Gamma_1^+$ ,  $\Delta_1^+$ ,  $\Gamma_2^+$ , and  $\Delta_2^+$  of sentences such that  $\Gamma_1^+ \cup \Delta_1^+$  exhausts all the sentences of  $\mathcal{L}_1$ ,  $\Gamma_2^+ \cup \Delta_2^+$  exhausts all the sentences of  $\mathcal{L}_2$ , and

there is no  $\lambda \in \Lambda^+$  such that  $\Gamma_1^+ \vdash_{\Sigma_1} \lambda, \Delta_1^+$  and  $\Gamma_2^+, \lambda \vdash_{\Sigma_2} \Delta_2^+$ .

(Notice that, since  $\perp \in \Lambda^+$ ,  $\Gamma_1^+ \not\vdash_{\Sigma_1} \Delta_1^+$ ; and, since  $\top \in \Lambda^+$ ,  $\Gamma_2^+, \not\vdash_{\Sigma_2} \Delta_2^+$ : thus  $\Gamma_1^+ \cap \Delta_1^+ = \Gamma_2^+ \cap \Delta_2^+ = \emptyset$ .) Secondly, the construction defines a subset  $D_0$  of  $D$  such that

for all  $d \in D_0$ ,  $d = d \in \Gamma_1^+ \cap \Gamma_2^+$  and  $\neg d = d \in \Delta_1^+ \cap \Delta_2^+$ ;  
 if  $\exists x\phi \in \Gamma_1^+$ , then  $\phi(d/x) \in \Gamma_1^+$ , for some  $d \in D_0$ ,  
 if  $\exists x\phi \in \Gamma_2^+$ , then  $\phi(d/x) \in \Gamma_2^+$ , for some  $d \in D_0$ ,  
 if  $\forall x\phi \in \Delta_1^+$ , then  $\phi(d/x) \in \Delta_1^+$ , for some  $d \in D_0$ ,  
 if  $\forall x\phi \in \Delta_2^+$ , then  $\phi(d/x) \in \Delta_2^+$ , for some  $d \in D_0$ .

(Thus quantifiers will be ‘witnessed’ by elements of  $D_0$ —which the first condition will guarantee are ‘defined’.)

Now we define relations  $\sim_1$  and  $\sim_2$  over  $D_0$  as follows:

$$\begin{aligned} d \sim_1 e & \text{ iff } d = e \in \Gamma_1^+ \text{ and } \neg d = e \in \Delta_1^+, \\ d \sim_2 e & \text{ iff } d = e \in \Gamma_2^+ \text{ and } \neg d = e \in \Delta_2^+. \end{aligned}$$

These turn out to be equivalence relations, and we can use them to factor out  $D_0$  to provide domains for models  $M_1^+$  for  $\mathcal{L}_1^+$  and  $M_2^+$  for  $\mathcal{L}_2^+$ , such that  $M_1^+$  is a model of  $\Sigma_1$  that rejects  $\langle \Gamma_1^+, \Delta_1^+ \rangle$ , and  $M_2^+$  is a model of  $\Sigma_2$  that rejects  $\langle \Gamma_2^+, \Delta_2^+ \rangle$ : the models can be defined in terms of  $\langle \Gamma_1^+, \Delta_1^+ \rangle$  and  $\langle \Gamma_2^+, \Delta_2^+ \rangle$  in much the same way that a classical model is defined from a consistent and complete set of sentences in a standard Henkin-style completeness proof.

But, by axiom (R),  $\Gamma_1^+ \cap \Lambda^+ \cap \Delta_2^+ = \emptyset$ , from which we can deduce that  $\sim_1$  and  $\sim_2$  are in fact the same relation, so that  $M_1^+$  and  $M_2^+$  have a common domain. Their reducts  $M_1$  and  $M_2$  to the original languages  $\mathcal{L}_1$  and  $\mathcal{L}_2$  then turn out to be models of  $\Sigma_1$  and of  $\Sigma_2$  such that  $(M_1, s_1)$  rejects  $\langle \Gamma_1, \Delta_1 \rangle$  and  $(M_2, s_2)$  rejects  $\langle \Gamma_2, \Delta_2 \rangle$ , where  $s_1$  and  $s_2$  are defined by putting  $s_1(x) = M_1^+(\pi(x))$  and  $s_2(x) = M_2^+(\pi(x))$ .

Finally, to deduce the relationship between  $M_1$  and  $M_2$ , and between  $s_1$  and  $s_2$ —which is peculiar to each of the three cases—we again make use of the fact that  $\Gamma_1^+ \cap \Lambda^+ \cap \Delta_2^+ = \emptyset$ . This guarantees the following facts:

- in case (1),  $M_1^+(\lambda) \sqsubseteq M_2^+(\lambda)$  for any  $\lambda \in \Lambda^+$ ;
- in case (2),  $M_2^+(\lambda) \sqsubseteq M_1^+(\lambda)$  for any  $\lambda \in \Lambda^+$ ;
- in case (3),  $M_1^+(\lambda) \sqsupset M_2^+(\lambda)$  for any  $\lambda \in \Lambda^+$ .

Hence, first, we can deduce that

- in case (1),  $M_{1s}(\lambda) \sqsubseteq M_{2s}(\lambda)$  for any  $\lambda \in \Lambda$  and any  $s$ ;
- in case (2),  $M_{2s}(\lambda) \sqsubseteq M_{1s}(\lambda)$  for any  $\lambda \in \Lambda$  and any  $s$ ;
- in case (3),  $M_{1s}(\lambda) \sqsupset M_{2s}(\lambda)$  for any  $\lambda \in \Lambda$  and any  $s$ .

But  $\Lambda$  contains all formulae of  $\mathcal{L}$ . And, for any  $\lambda \in \Lambda$ ,  $M_{1s}(\lambda) = (M_1 \upharpoonright \mathcal{L})_s(\lambda)$  and  $M_{2s}(\lambda) = (M_2 \upharpoonright \mathcal{L})_s(\lambda)$ . It therefore follows from Lemma 4 that the displayed conditions are equivalent, respectively, to

- (1)  $M_1 \sqsubseteq_{\mathcal{L}} M_2$ ,
- (2)  $M_2 \sqsubseteq_{\mathcal{L}} M_1$ ,
- (3)  $M_1 \sqsupset_{\mathcal{L}} M_2$ .

Secondly, since, for any variable  $x$  and any  $d \in D_0$ ,  $\pi(x) = d$  and  $\neg\pi(x) = d$  are both in  $\Lambda^+$ , we can also deduce—from the facts about  $M_1^+$  and  $M_2^+$ —that

- (1)  $s_1 \sqsubseteq s_2$ ,
- (2)  $s_2 \sqsubseteq s_1$ ,
- (3)  $s_1 \sqsupset s_2$ .

### 7.3 Some Proofs

The Completeness Theorem (Theorem 11) can now immediately be established: we shall argue by contraposition. Assume, then, that  $\Gamma \not\vdash_{\Sigma} \Delta$ . By rule (T), it follows that there can be no formula  $\lambda$  such that  $\Gamma \vdash_{\Sigma} \lambda, \Delta$  and  $\Gamma, \lambda \vdash_{\Sigma} \Delta$ . And so to show that some model of  $\Sigma$  rejects  $\langle \Gamma, \Delta \rangle$  we may apply Theorem 14, taking each of  $\mathcal{L}_1$ ,  $\mathcal{L}_2$ , and  $\mathcal{L}$  to be whatever language we're working with, and taking  $\Sigma_1 = \Sigma_2 = \Sigma$ ,  $\Gamma_1 = \Gamma_2 = \Gamma$ , and  $\Delta_1 = \Delta_2 = \Delta$ . This guarantees models  $M_1$  and  $M_2$ , with assignments  $s_1$  and  $s_2$ , which satisfy at least one of the three conditions specified. But each of these conditions obviously entails that both  $M_1$  and  $M_2$  reject  $\langle \Gamma, \Delta \rangle$ —which is over-kill: pick either one.

To establish the Compatibility Theorem (Theorem 5), we can appeal directly to case (3) of Theorem 13. Assume that  $\phi \sqsupset \psi$ , and—aiming for a

contradiction—assume that there is no joint for formulae  $\phi$  and  $\psi$ . Then, by Lemma 7, there is no lambda such that both  $\phi \sqsubseteq \lambda$  and  $\psi \sqsubseteq \lambda$ . But this is equivalent to the absence of any  $\lambda$  such that

$$\phi \vee \psi \vDash \lambda, * \quad \text{and} \quad *, \lambda \vDash \phi \wedge \psi.$$

By Soundness we can replace  $\vDash$  by  $\vdash$ , and then we have something in the right form to apply Theorem 13, case (3). Since we are working with pure logic in a single language, we take each of  $\mathcal{L}_1$ ,  $\mathcal{L}_2$ , and  $\mathcal{L}$  to be this language—so that  $\sqsubseteq_{\mathcal{L}}$  will just be  $\sqsubseteq$ —and we take  $\Sigma_1 = \Sigma_2 = \emptyset$ . Then we take  $\Gamma_1 = \{\phi \vee \psi\}$ ,  $\Gamma_2 = \{*\}$ ,  $\Delta_1 = \{*\}$ ,  $\Delta_2 = \{\phi \wedge \psi\}$ . This guarantees models  $M_1$  and  $M_2$ , with assignments  $s_1$  and  $s_2$ , such that  $(M_1, s_1)$  rejects  $\langle \Gamma_1, \Delta_1 \rangle$ ,  $(M_2, s_2)$  rejects  $\langle \Gamma_2, \Delta_2 \rangle$ ,  $M_1 \sqsubseteq M_2$ , and  $s_1 \sqsubseteq s_2$ . But the rejections mean that

$$(M_{1s_1}(\phi) = \top \text{ or } M_{1s_1}(\psi) = \top) \quad \text{and} \quad (M_{2s_2}(\phi) = \perp \text{ or } M_{2s_2}(\psi) = \perp)$$

Distributing ‘and’ across ‘or’ there are then four possibilities, each of which, by Lemma 6, contradicts the assumption that  $\phi \sqsubseteq \psi$ .

To establish Craig Interpolation (Theorem 9) we now make use of the fact that in Theorem 14  $\mathcal{L}_1$  and  $\mathcal{L}_2$  might be different languages. Given formulae  $\phi$  and  $\psi$ , let  $\mathcal{L}_1$  be the language whose non-logical vocabulary is precisely that occurring in  $\phi$ , let  $\mathcal{L}_2$  be the language whose non-logical vocabulary is precisely that occurring in  $\psi$ , and let  $\mathcal{L}$  be the language whose non-logical vocabulary is precisely that common to both  $\mathcal{L}_1$  and  $\mathcal{L}_2$ . Assume now that there is no Craig interpolant for formulae  $\phi$  and  $\psi$ : we have to show that  $\phi \not\sqsubseteq \psi$ . But, by Soundness, the absence of a Craig interpolant means that there is no formula  $\lambda$  of  $\mathcal{L}$  such that  $\phi \vdash \lambda$  and  $\lambda \vdash \psi$ . And so we may apply Theorem 14 taking  $\Sigma_1 = \Sigma_2 = \emptyset$ ,  $\Gamma_1 = \{\phi\}$ ,  $\Gamma_2 = \emptyset$ ,  $\Delta_1 = \emptyset$ ,  $\Delta_2 = \{\psi\}$ . This guarantees models  $M_1$  for  $\mathcal{L}_1$  and  $M_2$  for  $\mathcal{L}_2$ , along with assignments  $s_1$  and  $s_2$ , such that at least one of three possible conditions obtains. We shall consider each in turn.

In case (1),  $M_1 \sqsubseteq_{\mathcal{L}} M_2$ ,  $s_1 \sqsubseteq s_2$ ,  $(M_1, s_1)$  rejects  $\langle \{\phi\}, \{*\} \rangle$ , and  $(M_2, s_2)$  rejects  $\langle \emptyset, \{*, \psi\} \rangle$ . But now let  $M$  be an expansion of  $M_2$  which gives vocabulary in  $\mathcal{L}_1$  but not in  $\mathcal{L}_2$  the interpretation that  $M_1$  gives it. Then  $M_1 \sqsubseteq M \upharpoonright \mathcal{L}_1$ . Thus, by Monotonicity of Evaluation (Lemma 3), and since  $M \upharpoonright \mathcal{L}_1$  treats formulae of  $\mathcal{L}_1$  in the same way as  $M$ , it follows that

$$M_{1s_1}(\phi) \sqsubseteq (M \upharpoonright \mathcal{L}_1)_{s_1}(\phi) = M_{s_1}(\phi) \sqsubseteq M_{s_2}(\phi).$$

But  $(M_1, s_1)$ ’s rejecting  $\langle \{\phi\}, \{*\} \rangle$  means that  $M_{1s_1}(\phi) = \top$ , from which it follows that  $M_{s_2}(\phi) = \top$ . On the other hand,  $(M_2, s_2)$ ’s rejecting  $\langle \emptyset, \{*, \psi\} \rangle$  means that  $M_{2s_2}(\psi) \neq \top$ , from which it follows that  $M_{s_2}(\psi) \neq \top$ . Hence  $(M, s_2)$  rejects  $\langle \{\phi\}, \{\psi\} \rangle$ —showing that  $\phi \not\sqsubseteq \psi$ .

In case (2) we can argue in an exactly parallel way.

In case (3),  $M_1 \sqsubset_{\mathcal{L}} M_2$ ,  $s_1 \sqsubset s_2$ ,  $(M_1, s_1)$  rejects  $\langle\{\phi\}, \{*\}\rangle$ , and  $(M_2, s_2)$  rejects  $\langle\{*\}, \{\psi\}\rangle$ . But now let  $M_1^+$  be an expansion of  $M_1$  which gives vocabulary in  $\mathcal{L}_2$  but not in  $\mathcal{L}_1$  the interpretation that  $M_2$  gives it; and let  $M_2^+$  be an expansion of  $M_2$  which gives vocabulary in  $\mathcal{L}_1$  but not in  $\mathcal{L}_2$  in the interpretation that  $M_1$  gives it. Clearly  $M_2^+ \sqsubset M_1^+$ , and if  $M = M_2^+ \sqcup M_1^+$  and  $s = s_2 \sqcup s_1$ , then, by Monotonicity of Evaluation,

$$M_{1s_1}(\phi) = M_{1^+s_1}(\phi) \sqsubseteq M_s(\phi) \quad \text{and} \quad M_{2s_2}(\psi) = M_{2^+s_2}(\psi) \sqsubseteq M_s(\psi).$$

But the rejections mean, respectively, that  $M_{1s_1}(\phi) = \top$  and  $M_{2s_2}(\psi) = \perp$ . It follows that  $M_s(\phi) = \top$  and  $M_s(\psi) = \perp$ . Hence  $(M, s)$  rejects  $\langle\{\phi\}, \{\psi\}\rangle$ —again showing that  $\phi \not\equiv \psi$ .

Finally we shall use the Interpolant-Excluding Model Pairs Theorem to prove a result, which has not been mentioned before, about non-logical theories: a model-theoretic criterion for when a piece of non-logical vocabulary is definable in a theory  $\Sigma$ . First we need a relation  $\simeq_{\Sigma}$  of *equivalence in  $\Sigma$* —or  $\Sigma$ -*equivalence*. Now that we have soundness and completeness in place, we can indifferently define this relation either in terms of  $\vdash_{\Sigma}$  or in terms of the models of  $\Sigma$ :

$$\begin{aligned} \phi \simeq_{\Sigma} \psi & \text{ iff } \phi \vdash_{\Sigma} \psi \text{ and } \psi \vdash_{\Sigma} \phi, \\ & \text{ iff } M_s(\phi) = M_s(\psi), \text{ for any } M \in \mathcal{K}(\Sigma) \text{ and any } s. \end{aligned}$$

Then let us say that (i) a predicate symbol  $P$ , (ii) a function symbol  $f$ , (iii) a constant symbol  $c$ , is (*explicitly*) *definable in  $\Sigma$*  if and only if there is a formula  $\phi$  that does not contain (i)  $P$ , (ii)  $f$ , (iii)  $c$ , such that

$$(i) \ Px_1 \dots x_{\lambda(P)} \simeq_{\Sigma} \phi, \quad (ii) \ y = fx_1 \dots x_{\lambda(f)} \simeq_{\Sigma} \phi, \quad (iii) \ y = c \simeq_{\Sigma} \phi,$$

(where the displayed variables are assumed to be distinct from one another).

The definability theorem takes exactly the same form for each of these three cases, and so we can state it schematically for an item  $\alpha$  of non-logical vocabulary. Say that  $\mathcal{L}$  is the language of the theory  $\Sigma$ , and let  $\mathcal{L}_{\alpha}$  be the language got from  $\mathcal{L}$  by dropping  $\alpha$ , then

**THEOREM 15.**

$\alpha$  is definable in  $\Sigma$  iff, for any models  $M$  and  $N$  of  $\Sigma$ ,

$$(a) \text{ if } M \sqsubseteq_{\mathcal{L}_{\alpha}} N, \text{ then } M \sqsubseteq N, \quad \text{and} \quad (b) \text{ if } M \sqsubset_{\mathcal{L}_{\alpha}} N, \text{ then } M \sqsubset N.$$

In other words, it is necessary and sufficient for the definability of  $\alpha$  that given a pair of models of  $\Sigma$ , if (a) the relation  $\sqsubseteq$ , or (b) the relation  $\sqsubset$ , obtains between the interpretations of vocabulary other than  $\alpha$ , then it also obtains between the interpretations of  $\alpha$ . It is easy enough to check ‘only if’ directly. To establish ‘if’, we can argue by contraposition and invoke

Theorem 14. I shall sketch the case where  $\alpha$  is a predicate letter  $P$ : the other cases are not too different.

Assume, then, that  $P$  is not definable in  $\Sigma$ . This means that there is no formula  $\lambda$  of  $\mathcal{L}_\alpha$  such that  $Px_1 \dots x_{\lambda(P)} \vdash_\Sigma \lambda$  and  $\lambda \vdash_\Sigma Px_1 \dots x_{\lambda(P)}$ . Hence we can apply Theorem 14, taking both the  $\mathcal{L}_1$  and  $\mathcal{L}_2$  of that theorem to be the language  $\mathcal{L}$  of this one, and taking the  $\mathcal{L}$  of that theorem to be  $\mathcal{L}_\alpha$ . And we then take  $\Sigma_1 = \Sigma_2 = \Sigma$ ,  $\Gamma_1 = \{Px_1 \dots x_{\lambda(P)}\}$ ,  $\Delta_1 = \emptyset$ ,  $\Gamma_2 = \emptyset$ ,  $\Delta_2 = \{Px_1 \dots x_{\lambda(P)}\}$ . This guarantees models  $M_1$  and  $M_2$  of  $\Sigma$ , along with assignments  $s_1$  and  $s_2$ , such that at least one of three possible conditions obtains. We shall consider each in turn.

In case (1),  $M_1 \sqsubseteq_{\mathcal{L}_\alpha} M_2$ , but the rejection conditions, together with the fact that  $s_1 \sqsubseteq s_2$ , entail that  $M_1 \not\sqsubseteq M_2$ . For  $(M_1, s_1)$  rejects  $\langle \{Px_1 \dots x_{\lambda(P)}\}, \{*\} \rangle$ , so that  $M_{1s_1}(Px_1 \dots x_{\lambda(P)}) = \top$ , and therefore  $M_{1s_2}(Px_1 \dots x_{\lambda(P)}) = \top$ ; but  $(M_2, s_2)$  rejects  $\langle \emptyset, \{*, Px_1 \dots x_{\lambda(P)}\} \rangle$ , so that  $M_{2s_2}(Px_1 \dots x_{\lambda(P)}) \neq \top$ .

In case (2) we can argue in an exactly parallel way.

In case (3),  $M_1 \sqcap_{\mathcal{L}_\alpha} M_2$ , but the rejection conditions, together with the fact that  $s_1 \sqcap s_2$ , entail that  $M_1 \not\sqcap M_2$ . For  $(M_1, s_1)$  rejects  $\langle \{Px_1 \dots x_{\lambda(P)}\}, \{*\} \rangle$ , so that  $M_{1s_1}(Px_1 \dots x_{\lambda(P)}) = \top$ ; and  $(M_2, s_2)$  rejects  $\langle \{*\}, \{Px_1 \dots x_{\lambda(P)}\} \rangle$ , so that  $M_{2s_2}(Px_1 \dots x_{\lambda(P)}) = \perp$ : and therefore  $M_{1s}(Px_1 \dots x_{\lambda(P)}) = \top$  and  $M_{2s}(Px_1 \dots x_{\lambda(P)}) = \perp$ , where  $s = s_1 \sqcup s_2$ .

There are two noteworthy comments on this definability result. First, the condition on models of  $\Sigma$  is strictly stronger than the condition that whenever models *agree exactly* on vocabulary other than  $\alpha$ , then they also agree on  $\alpha$ . Secondly, it follows from the definability of  $\alpha$  in  $\Sigma$  that there will be a uniform procedure for transforming any formula into an  $\alpha$ -free  $\Sigma$ -equivalent one. In the case of a predicate symbol this is just a matter of making the obvious substitution. In the case of a definable function symbol  $f$ , on the other hand, there will be a scheme of elimination for terms  $ft_1 \dots t_{\mu(f)}$  that is scope-free in the same way that the description-scheme we specified in Section 6.4 is scope free. Given terms  $t_1, \dots, t_{\mu(f)}$ , we shall always be able to define  $f$  using a formula  $\phi$  that contains no variables occurring in  $t_1, \dots, t_{\mu(f)}$ :  $y = ft_1 \dots t_{\mu(f)} \simeq_\Sigma \phi$ . (We can always rewrite variables as required.) Then, by rule (S),

$$y = ft_1 \dots t_{\mu(f)} \simeq_\Sigma \phi(t_1/x_1) \dots (t_{\mu(f)}/x_{\mu(f)}) \quad (\phi(t_i/x_i) \text{ for short}).$$

It follows that, provided  $ft_1 \dots t_{\mu(f)}$  is substitutable for  $y$  in  $\psi$ ,  $\psi(ft_1 \dots t_{\mu(f)}/y)$  will be  $\Sigma$ -equivalent to each of the following:

$$\begin{aligned} & \exists y[\phi(t_i/x_i) \wedge \psi] \vee [\forall y[\phi(t_i/x_i) \rightarrow \psi] \wedge \psi(\otimes/y)]; \\ & \forall y[\phi(t_i/x_i) \rightarrow \psi] \wedge [\exists y[\phi(t_i/x_i) \wedge \psi] \vee \psi(\otimes/y)]. \end{aligned}$$

And a definable constant symbol can be handled in a parallel way—without any need to fuss about variables.



× × ×

Further model-theoretic results about non-logical theories can be derived from subtler versions of the Interpolant-Excluding Model Pairs Theorem(s). An example of this is the theorem we mentioned in Section 6.4 concerning the eliminability of  $\otimes$  in a theory  $\Sigma$ . By ‘eliminability’ let us agree to mean simply that any formula  $\phi$  is equivalent in  $\Sigma$  to some  $\otimes$ -free formula  $\psi$ :  $\phi \simeq_{\Sigma} \psi$ . And let us define a new degree-of-definedness relation  $\sqsubseteq_{\otimes}$  between models  $M$  and  $N$  by taking over the definition of ‘ $M \sqsubseteq N$ ’ given in Section 6.2, but restricting  $\vec{a}$ , in clauses (1) and (2), to  $D^{\lambda(P)}$  and to  $D^{\mu(f)}$ .  $D$  is the common domain of  $M$  and  $N$ , and so  $M \sqsubseteq_{\otimes} N$  if and only if  $N$  is more defined than  $M$  over objects in the domain. In general  $\sqsubseteq_{\otimes}$  is a strictly weaker relation than  $\sqsubseteq$ , but

**THEOREM 16.**  $\otimes$  is eliminable in a theory  $\Sigma$  if and only if, whenever  $M$  and  $N$  are non-empty models of  $\Sigma$  and  $M \sqsubseteq_{\otimes} N$ , then  $M \sqsubseteq N$ .

Another result about non-logical theories arises from further consideration of the Compatibility Theorem (Theorem 5). This theorem was a result about pure logic, but the question arises concerning an arbitrary theory  $\Sigma$  whether formulae that are *compatible in  $\Sigma$* —i.e. never take on conflicting truth values in models of  $\Sigma$ —have a *joint in the theory  $\Sigma$* —i.e. a formula with the  $\top/\perp$ -conditions of a joint in all models of  $\Sigma$ . The answer is ‘no’, but we can derive a model-theoretic criterion for when a theory is guaranteed joints for all compatible formulae. This result, however, requires more apparatus than we have developed—even to state, let alone to prove.

*St Edmund Hall, Oxford.*

## BIBLIOGRAPHY

- [Aczel 1977] P. Aczel. An Introduction to inductive definitions. In *Handbook of Mathematical Logic*, J. Barwise (ed.), pp. 739–782. North Holland, Amsterdam, 1977.
- [Aczel and Feferman 1980] P. Aczel and S. Feferman. Consistency of the Unrestricted Abstraction Principle using an Intensional Equivalence Operator. In *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, J. P. Seldin and J. R. Hindley, eds., Academic Press, London, 1980.
- [Barendregt 1984] H. P. Barendregt. *The Lambda Calculus*, North Holland, Amsterdam, 1984, 2nd reprint edition in paperback, 1997.
- [Barwise 1981] J. Barwise. Scenes and other situations. *J. Philosophy*, **78**, 369–397, 1981.
- [Barwise and Etchemendy 1987] J. Barwise and J. Etchemendy. *The Liar*, Oxford University Press, Oxford, 1987.
- [Barwise and Perry 1981a] J. Barwise and J. Perry. Situations and attitudes. *J. Philosophy*, **78**, 668–691, 1981.
- [Barwise and Perry 1981b] J. Barwise and J. Perry. Semantic innocence and uncompromising situations. In *Midwest Studies in Philosophy VI*, French *et al.* (eds), pp. 387–403. University of Minnesota Press, Minneapolis, 1981.
- [Barwise and Perry 1983] J. Barwise and J. Perry. *Situations and Attitudes*, MIT Press, Cambridge, MA, 1983.

- [Beaver 1997] D. I. Beaver. Presupposition. In *Handbook of Logic and Language*, J. van Benthem and A. ter Meulen, eds. pp. 939–1008. Elsevier, Amsterdam, 1997.
- [Bell 1990] D. Bell. How Russellian Was Frege? *Mind*, **99**, 267–277, 1990.
- [Belnap 1970] N. D. Belnap. Conditional assertion and restricted quantification. *Nous*, **4**, 1–13, 1970.
- [Blamey 1980] S. R. Blamey. *Partial-Valued Logic*, D.Phil. thesis, University of Oxford, 1980.
- [Blamey 1991] S. R. Blamey. The Soundness and Completeness of Axioms for CSP. In *Topology and Category Theory in Computer Science*, G. M. Reed, A. W. Roscoe and R. F. Wachter, eds., pp. 29–56, Oxford University Press, Oxford, 1991.
- [Blamey and Humberstone 1991] S. R. Blamey and L. Humberstone. A Perspective on Modal Sequent Logic. *Publications of the Research Institute for Mathematical Sciences*, **27**, Kyoto University, 763–782, 1991.
- [Bochman 1998] A. Bochman. Biconsequence Relations: A Four-Valued Formalism of Reasoning with Inconsistency and Incompleteness. *Notre Dame Journal of Formal Logic*, **39**, 47–73, 1998.
- [Cleave 1974] J. P. Cleave. Logical consequence in the logic of inexact predicates. *Z. Math. Logik Grundlagen Math*, **20**, 307–324, 1974.
- [Czermak 1974] J. Czermak. A logical calculus with descriptions. *J Philosophical Logic*, **3**, 211–228, 1974.
- [Dummett 1959] M. A. E. Dummett. Truth. *Proc. Aristotelian Soc.*, **59**, 141–162, 1959.
- [Dummett 1973] M. A. E. Dummett. *Frege*, Duckworth, London, 1973.
- [Dummett 1978] M. A. E. Dummett. *Truth and Other Enigmas*. Duckworth, London, 1978.
- [Dunn 1975] J. M. Dunn. Axiomatizing Belnap's conditional assertion. *J. Philosophical Logic*, **4**, 383–397, 1975.
- [Ebbinghaus 1969] H.-D. Ebbinghaus. Über eine Prädikatenlogik mit partiell definierten Prädikaten und Funktionen. *Arch. Math. Logik Grundlagenforschung*, **12**, 39–53, 1969.
- [Evans 1982] G. Evans. *The Varieties of Reference*. Oxford University Press, Oxford, 1982.
- [Feferman 1975] S. Feferman. : 1975, Non-extensional type-free theories of partial operations and classifications. In *Springer Lecture Notes in Mathematics* No.500. J. Diller and J. H. Muller, eds. pp. 73–118. Springer Verlag, 1975.
- [Feferman 1995] S. Feferman. Definedness. *Erkenntnis*, **43**, 295–320, 1995.
- [Fenstad 1997] J. E. Fenstad. Partiality. In *Handbook of Logic and Language*, J. van Benthem and A. ter Meulen, eds. pp. 649–682. Elsevier, Amsterdam, 1997.
- [Fenstad et al. 1987] J. E. Fenstad, P. K. Halvorsen, T. Langholm and J. van Benthem. *Situations, Language and Logic*, Reidel, Dordrecht, 1987.
- [Frege 1891] G. Frege. 1891, Funktion und Begriff. (Vortrag, gehalten in der Sitzung vom 9. Januar 1891 der Jenaischen Gesellschaft für Medizin und Naturwissenschaften.), Verlag H. Pohle, Jena, 1891. Tr. in *Translations from The Philosophical Writings of Gottlob Frege*. 2nd revised edn., P. Geach and M. Black, eds. Blackwell, Oxford, 1960.
- [Frege 1892] G. Frege. Über Sinn, und Bedeutung. *Zeitschrift für Philosophie und Philosophische Kritik*, pp. 25–50, 1892. Tr. in *Translations from The Philosophical Writings of Gottlob Frege*. 2nd revised edn., P. Geach and M. Black, eds. Blackwell, Oxford, 1960.
- [Gabbay 1982] D. Gabbay. Intuitionistic Basis for Non-Monotonic Logic. In *Proceedings of the 6th Conference on Automated Deduction*, Lecture Notes in CS 138, pp. 260–273, Springer-Verlag, Berlin, 1982.
- [Gilmore 1974] P. C. Gilmore. The consistency of partial set theory without extensionality. In *Axiomatic Set Theory: 1967 UCLA Symposium Proceedings of Symposium in Pure Mathematics*, Vol. 13, Part 1, T. Jech, ed. pp. 147–153. American Mathematical Society, 1974.
- [Groeneveld 1994] Dynamic Semantics and Circular Propositions. *Journal of Philosophical Logic*, **23**, 267–306, 1994.
- [Haack 1974] S. Haack. *Deviant Logic*, Cambridge University Press, Cambridge, 1974.

- [Haack 1978] S. Haack. *Philosophy of Logics*, Cambridge University Press, Cambridge, 1978.
- [Hayes 1975] P. Hayes. *Three-valued logic and Computer Science*, CSM-6, University of Essex, 1975.
- [Heim 1982] I. Heim. *The Semantics of Definite and Indefinite Noun Phrases*, PhD dissertation, University of Massachusetts, Amherst, 1982.
- [Herzberger 1970] H. G. Herzberger. Paradoxes of grounding in semantics. *J. Philosophy*, **67**, 145–167, 1970.
- [Hinnion 1994] R. Hinnion. Naive Set Theory with Extensionality in Partial Logic and in Paradoxical Logic. *Notre Dame Journal of Formal Logic*, **35**, 15–40, 1992.
- [Humberstone 1981] L. Humberstone. From worlds to possibilities. *J. Philosophical Logic*, **10**, 313–339, 1981.
- [Jaspars 1995] J. O. M. Jaspars. Partial Up and Down Logic. *Notre Dame Journal of Formal Logic*, **36**, 135–157, 1995.
- [Kamp 1981] H. Kamp. A Theory of Truth and Semantic Representation. In *Truth, Interpretation and Information*, J. Groenendijk et al., eds., pp. 1–41. Foris, Dordrecht, 1981.
- [Kamp and Reyle 1993] H. Kamp and U. Reyle. *From Discourse to Logic*, Kluwer, 1993.
- [Karttunen 1973] L. Karttunen. Presuppositions of Compound Sentences. *Linguistic Inquiry*, **4**, 167–193, 1973.
- [Karttunen 1974] L. Karttunen. Presuppositions and Linguistic Context. *Theoretical Linguistics*, **1**, 181–194, 1974.
- [Keenan 1973] E. L. Keenan. Presupposition in natural logic. *Monist*, **57**, 334–370, 1973.
- [Kleene 1952] S. C. Kleene *Introduction to Metamathematics*, North Holland, Amsterdam, 1952.
- [Krahmer 1995] E. Krahmer. *Discourse and Presupposition*, PhD dissertation, ITK/TILDIL Dissertation Series, University of Tilburg, 1995.
- [Kripke 1975] S. Kripke. Outline of a theory of truth. *J. Philosophy*, **72**, 690–716, 1975.
- [Langholm 1988] T. Langholm. *Partiality, Truth and Persistence*, CSLI Lecture Notes no. 15, CSLI, Stanford, 1988.
- [Langholm 1989] T. Langholm. *Algorithms for Partial Logic*, COSMOS Report no. 12, Department of Mathematics, University of Oslo, 1989.
- [Lapierre 1992] S. Lapierre. A Functional Partial Semantics for Intensional Logic. *Notre Dame Journal of Formal Logic*, **33**, 517–541, 1992.
- [Lehmann 1994] S. Lehmann. Strict Fregean Free Logic. *Journal of Philosophical Logic*, **23**, 307–336, 1994.
- [Lepage 1992] F. Lepage. Partial Functions in Type Theory. *Notre Dame Journal of Formal Logic*, **33**, 493–516, 1992.
- [Lewis 1972] D. Lewis. General Semantics. In *Semantics of Natural Language*, D. Davidson and G. Harman, eds., pp. 169–218, Reidel, Dordrecht, 1972.
- [Lopez-Escobar 1972] E. Lopez-Escobar. Refutability and elementary number theory. *Koninkl. Nederl. Akademie van Wetenschappen Proceedings, Series A*, **75**, 362–374, 1972. Also in *Indag. Math.*, **34**, 362–374.
- [McDowell 1984] J. McDowell. *De Re Senses*. In *Frege: Tradition and Influence*, C. Wright, ed., pp. 98–109, Blackwell, Oxford, 1984.
- [McDowell 1986] J. McDowell. Singular Thought and the Extent of Inner Space. In *Subject, Thought and Context*, J. McDowell and P. Pettit, eds., Oxford University Press, Oxford, 1986.
- [McDowell 1977] J. McDowell. On the sense and reference of a proper name. *Mind*, **86**, 362–374, 1977.
- [Martin 1970] R. L. Martin, ed. *The Paradox of the Liar*, Yale University Press, New Haven, 1970.
- [Muskens 1989] R. A. Muskens. *Meaning and Partiality*, Ph.D. Dissertation, University of Amsterdam, 1989.
- [Muskens et al. 1997] R. A. Muskens, J. van Benthem and A. Visser. Dynamics. In *Handbook of Logic and Language*, J. van Benthem and A. ter Meulen, eds. pp. 587–648. Elsevier, Amsterdam, 1997.
- [Nelson 1949] D. N. Nelson. Constructible falsity. *J. Symbolic Logic*, **14**, 16–26, 1949.

- [Päppinghaus and Wirsing 1981] P. Päppinghaus and N. Wirsing. *Nondeterministic Partial Logic: Isotonic and Guarded Truth-Functions*, Internal Report CSR-83-81, University of Edinburgh, 1981.
- [Russell 1905] B. Russell. On denoting. *Mind*, **14**, 479–493, 1905.
- [Russell 1959] B. Russell. Mr Strawson on referring. In *My Philosophical Development*, pp. 238–245. Allen and Unwin, London, 1959.
- [Sainsbury 1999] R. M. Sainsbury. Names, Fictional Names and Reality. *Proceedings of the Aristotelian Society*, Supp. Vol., **73**, 243–269, 1999.
- [Sandu 1998] G. Sandu. Partially Interpreted Relations and Partially Interpreted Quantifiers. *Journal of Philosophical Logic*, **27**, 587–601, 1998.
- [Scott 1967] D. S. Scott. Existence and description in formal logic. In *Bertrand Russell, Philosopher of the Century*, R. Schoenman, ed. pp. 181–200. Allen and Unwin, London, 1967.
- [Scott 1973a] D. S. Scott. Models of various type-free calculi. In *Logic, Methodology and Philosophy of Science IV*, P. Suppes *et al.*, eds. pp. 157–187. North Holland, Amsterdam, 1973.
- [Scott 1973b] D. S. Scott. Background to formalization. In *Truth, Modality and Syntax*, H. Leblanc, ed. pp. 244–273. North-Holland, Amsterdam, 1973.
- [Scott 1975] D. S. Scott. Combinators and classes. In  *$\lambda$ -Calculus and Computer Science*, C. Böhm, ed. pp. 1–26. Springer Verlag, Heidelberg, 1975.
- [Seuren 1976] P. Seuren. *Tussen Taal en Denken*, Oosthoek, Scheltema en Holkema, Utrecht, 1976.
- [Seuren 1985] P. Seuren. *Discourse Semantics*, Blackwell, Oxford, 1985.
- [Smiley 1960] T. J. Smiley. Sense without denotation. *Analysis*, **20**, 125–135, 1960.
- [Stalnaker 1972] R. Stalnaker. Pragmatics. In *Semantics of Natural Language*, D. Davidson and G. Harman, eds., pp. 380–397, Reidel, Dordrecht, 1972.
- [Strawson 1950] P. F. Strawson. On referring. *Mind*, **59**, 320–344, 1950.
- [Strawson 1964] P. F. Strawson. Identifying reference and truth values. *Theoria*, **30**, 96–118, 1964.
- [Thijssse 1992] E. G. C. Thijssse. *Partial Logic and Knowledge Representation*, PhD Thesis, University of Tilburg, 1992.
- [Thomason 1969] R. H. Thomason. R.H.: 1969, 'A semantical study of constructible falsity. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, **15**, 247–257, 1969.
- [Thomason 1972] R. H. Thomason. A semantic theory of sortal incorrectness. *J. Philosophical Logic*, **1**, 209–258, 1972.
- [Thomason 1979] S. K. Thomason. Truth-value gaps, many truth-values and possible worlds. In *Syntax and Semantics*, Vol.10. C. Oh and D. Dinneen, eds. pp. 357–369. Academic Press, New York, 1979.
- [Tichý 1982] P. Tichý. Foundations of Partial Type Theory. *Reports on Mathematical Logic*, **14**, 59–72, 1982.
- [Turner 1984] R. Turner. *Logics for Artificial Intelligence*, Ellis Horwood, Chichester, 1984.
- [Van Benthem 1988] J. F. A. K. van Benthem. *A Manual of Intensional Logic*, CSLI Lecture Notes No. 1, CLSI, Stanford, 1988.
- [Van Benthem 1991] J. F. A. K. van Benthem. Logic and the Flow of Information. In *Proceedings of the 9th International Congress of Logic, Methodology and Philosophy of Science (Uppsala, Sweden)*, D. Prawitz, B. Skirms and D. Westerstahl, eds., 1991.
- [Van Benthem and Van Eijck 1982] J. F. A. K. van Benthem and J. van Eijck. The dynamics of interpretation. *J. Semantics*, **1**, 1–20, 1982.
- [Van Eijck 1995] J. van Eijck. Quantifiers and Partiality. In *Quantifiers, Logic and Language*, Jaap van der Does and J. van Eijck, eds., Stanford University, 1995.
- [Van Fraassen 1966] B. C. van Fraassen. Singular terms, truth-value gaps and free logic. *J. Philosophy*, **63**, 481–495, 1966.
- [Van Fraassen 1967] B. C. van Fraassen. Presupposition, implication and self-reference. *J. Philosophy*, **65**, 136–152, 1967.
- [Van Fraassen 1971] B. C. van Fraassen. *Formal Semantics and Logic*, Macmillan, New York, 1971.

- [Van Fraassen 1975] B. C. van Fraassen. , B.C.: 1975, 'Incomplete assertion and Belnap connectives. In *Contemporary Research in Philosophical Logic and Linguistic Semantics*, D. Hockney *et al.*, eds. pp. 43–70. D. Reidel, Dordrecht, 1975.
- [Veltman 1981] F. Veltman. Data semantics. In *Formal Methods in the Study of Language*, J. Groenendijk and M. Stokhof, eds. pp. 541–565. Math. Centre Tract 136, Amsterdam, 1981.
- [Veltman 1996] F. Veltman. Defaults in Update Semantics. *Journal of Philosophical Logic*, **25**, 221–261, 1996.
- [Visser 1984] A. Visser. Four Valued Semantics and the Liar. *Journal of Philosophical Logic*, **13**, 181–212, 1984.
- [Wang and Mott 1998] X. Wang and P. Mott. A Variant of Thomason's First-Order Logic **CF** Based on Situations. *Notre Dame Journal of Formal Logic*, **39**, 74–93, 1998.
- [Wansing 1993] H. Wansing. *The Logic of Information Structures*, Springer, Heidelberg, 1993.
- [Wansing 1995] H. Wansing. Semantics-based Nonmonotonic Inference. *Notre Dame Journal of Formal Logic*, **36**, 44–54, 1995.
- [Wiggins 1976] David Wiggins. Frege's Problem of the Morning Star and the Evening Star. In *Studies in Frege*, M. Schirn, ed., Vol. II, pp. 221–255, Günthner Holzboog, Stuttgart-Bad Cannstatt, 1976.
- [Wiggins 1984] David Wiggins. The Sense and Reference of Predicates: A Running Repair to Frege's Doctrine and a Plea for the Copula. In *Frege: Tradition and Influence*, C. Wright, ed., pp. 126–143, Blackwell, Oxford, 1984.
- [Wiggins 1995] David Wiggins. The Kant–Frege–Russell View of Existence: Toward the Rehabilitation of the Second-Level View. In *Modality, Morality and Belief*, W. Sinnott-Armstrong, D. Raffman and N. Asher, eds., Cambridge University Press, Cambridge, 1995.
- [Wiggins 1999] David Wiggins. Names, Fictional Names and Reality. *Proceedings of the Aristotelian Society*, Supp. Vol., **73**, 271–286, 1999.
- [Woodruff 1970] P. Woodruff. Logic and truth value gaps. In *Philosophical Problems in Logic*, K. Lambert, ed. pp. 121–142. D. Reidel, Dordrecht, 1970.
- [Wray 1987a] D. O. Wray. Logic in Quotes. *Journal of Philosophical Logic*, **16**, 77–110, 1987.
- [Wray 1987b] D. O. Wray. Algebraic Quotational Logics. *Communication and Cognition*, **20**, 403–422, 1987.



## INDEX

- \*-satisfiable, 235
- Anselm, ontological argument, 244
- answer, 116
- apartness relation, 63
- argumentation form, 115
- Aristotle, on predication, 209, 226, 250
- attack, 116
- automated theorem proving, 225
- bar, 25
- Barba, on supervaluation and modality, 233
- Beeson, M., 7
- Behmann, 48
- being while lacking existence, 207
- Bencivenga, on story semantics, 224
- Bencivenga, on super valuation, 230
- Beth model, 29
- Beth, E. W., 22
- bivalence, 220
- Boolean algebra, 38
- Brouwer, L. E. J., 2, 95
- Brouwer, on partial functions, 208
- Burge, on descriptions, 217
- Burge, on fictional entitites, 215
- Burge, on predication, 226
- C*-dialogues, 139
- cancellation, 245
- characterization, 251
- choice sequences, 87
- Church's Thesis, 51, 81
- CL, 202
- class abstracts, 217
- classical description theory, 239
- classical existence assumptions, 202
- classical logic (CL), 202
- classification, 251
- closure under rules, 70
- combinatory logic, 18
- communication, 2
- compactness, 229
- complex predicate, 209, 245
- complex predicate, and extensionality, 251
- complex quantifiers, 279, 307, 313
- comprehension axiom, 217
- comprehensive quantifier, 207, 244, 252
- conditional reading of free variables, 200
- constant domain axiom, 35
- constructible falsity, 288
- contingent *a priori* truths, 208
- continuum, 100
- convention, 222, 230
- creating subject, 97
- Curry, H., 21
- Curry–Howard isomorphism, 19, 22
- D*-dialogue, 118
- Dalen, D. van, 25, 61, 63, 100
- De Morgan's Law, 35
- definability paradoxes, 87
- defining axiom, 252
- definite description(s), 278, 279, 294
- definite descriptions(s), 264
- definitions, 238, 251
- degree-of-definedness, 268, 270
- dense linear ordering, 65

- descriptions, 237
  - inproper, 238
  - proper, 238
- Diaconescu, 97
- Dialectica Interpretation, 5
- dialogue, 115
- disjunction property, 44, 68
- disjunction property for analysis, 94
- double negation principle, 35
- double negation shift, 35, 47
- Dummett's axiom, 35
- Dummett, M., 25, 51, 52, 102
- Dwyer, on definition, 252
- dynamic operators, 290
- dynamic semantics, 291, 297, 309
  
- E*-dialogue, 118
- Ebbinghaus, on attribution, 220
- Ehrenfeucht–Fraïssé games, 44
- elementary formula, 198
- elimination rules, 11
- elimination theorems, 92
- equivalence, axiomatisation of, 226
- error object, 252
- Evans, on contingent *a priori* truths, 208
- existence predicate, 198, 207, 209, 216, 254
- existence property, 45, 68
- expressive adequacy for monotonic truth functions, 268, 304
- extension principle, 92
- extensionality, 214, 251
- extensionality axiom, 217
  
- Farmer, on partial functions, 220, 225, 226
- FD, see mFD, 242
- FD2, see MFD, 242
- Feferman, on partial functions, 225
- Fine, K., 59
- finite Kripke models, 46
- finite model property, 55
  
- Firedman translation, 72
- Firedman, H., 73
- fixed point construction, 247
- formal argumentation forms, 142
- formal dialogue, 142
- formal strategy, 142
- formulas as types, 22
- Fourman, M., 106
- free description theory, neutral, 248
- free description theory, outer-domain, 242
- free description theory, Russellian, 245
- free description theory, supervaluational, 247
- free logic, 197
- Frege, on descriptions, 239
- Frege, on functions, 234
- Frege, on non-referring terms and bivalence, 227
- Friedman, H., 52, 69, 72
- functional dependence, 301
  
- Gödel sentence, 256
- Gödel, K., 74
- Görnemann, S., 60
- Gabbay, D. M., 47, 49, 53, 55, 96
- Gallier, J., 22
- Garson, on intensional logic, 253
- generality reading of free variables, 200
- Gentzen, G., 4, 74
- Girard, J. Y., 60
- Glivenko's theorem, 53
- Glivenko, V., 3, 75
- gluing, 94
- Goldblach's Conjecture, 3
- Goldblatt, R., 66
- Goodman, 97
- Grayson, R., 99
- Gumb, on definition, 252
  
- Harrop, R., 55
- Herbrand Theorem, 48



- Heyting algebras, 37  
 Heyting's arithmetic, 67  
 Heyting's second-order arithmetic, 85  
 Heyting, A., 4, 7, 22  
 Hilbert and Bernays, on descriptions, 239  
 Howard, W., 21  
  
 identity, axiomatisation of, 217  
 implication, axiomatisation of, 226  
 inclusive, 197, 212  
 independence of premiss principle, 35  
 intension, 253  
 intensional logic, 253  
 interjunction, 262, 264, 274, 280, 281, 295, 304, 307  
 interjunctive normal forms, 305  
 intermediate logic, 53  
 internal validity, 51  
 interpolation theorem, 59  
 introduction rules, 11  
 intuitionism, 208  
 intuitionistic logic, 4, 208  
**IPC**, 15  
**IQC**, 15  
  
 Jankov, V. A., 59  
 Jaskowski sequence, 47  
 Jeffrey trees, for neutral semantics, 235  
 Johansson, I., 74  
 Johnstone, P., 38  
 Jongh, D. de, 49  
  
 Kleene slash, 70  
 Kleene, S., 5  
 Kolmogorov, A. N., 3  
 Komori, Y., 60  
 Kreisel, G., 7, 22, 69, 73, 76, 97  
 Kreisel, H., 49  
 Kripke frame, 42  
 Kripke model, 29, 67  
  
 Kripke's schema, 93, 99, 100  
 Kripke, S., 22, 97  
 Kroll, 99  
 Kroon, on descriptions, 246  
 Kroon, on fictional entities, 215  
 Kroon, on logical form, 213  
  
 $\lambda$ -calculus, 18  
 Löwenheim–Skolem theorem, 229  
 $\lambda$ -calculus, 285, 292  
 Lambert's law, 238  
 Lambert, on definition, 252  
 Lambert, on logical form, 213  
 Lambert, on negative semantics, 226  
 Lambert, on outer domains and Meinong, 222  
 Lambert, on predication, 209  
 Lambert, on story semantics, 224  
 Lambert, on theories between mFD and MFD, 243  
 lattice of intermediate logics, 59  
 law of the excluded fourth, 288  
 lawles sequence, 89  
 Leblanc, on PFL, 222  
 Lehmann, on neutral semantics, 235  
 Leivant, D., 70  
 Lejewski, on identity, 207  
 Lemmon, E., 54  
 Lin, on equivalence and implication, 226  
 Lindenbaum algebra, 39  
 locally true, 105  
 logic of constant domains, 60  
 logical analysis in partial logic, 280, 307, 309, 311, 313  
 logical consequence, 200, 265, 266, 274, 296  
 logical consequence, in neutral semantics, 235  
 logical form, 211  
 logically neither true nor false sentence, 265

- logically neither true-nor-false sentence, 305
- logically non-denoting singular-term, 265
- logic of existence, 102
- Lorenzen, P., 123
- Malmnäs, P., 74
- Mann, on ontological argument, 244
- Markov's principle, 51, 70
- Markov, A. A., 83
- Martin-Löf type theories, 22
- Martin-Löf's type theory, 7
- Martin-Löf, P., 7, 22
- mathematical language, 3
- maximal free description theory (MFD), 242
- Maximova, L., 59
- McCarty, D., 83
- McCarty, D. C., 52
- McKinsey, J. C. C., 49
- meaning, as denotation, 240, 241
- Meinong's paradox, 238
- Meinong, on being vs. existence, 207
- Meinong, on bivalence, 222
- Meinong, on predication, 209
- Mendelson, on non-referring terms, 205
- mereology, 210
- Minc, G., 70
- minimal free description theory (mFD), 242
- minimal logic, 73
- modal semantics, 207, 208, 233, 253
- model existence lemma, 31
- monadic fragment, 48
- monotonically representable partial functions, 269–271, 287, 302
- monotonicity of evaluation, 268, 273, 284, 286
- Montague, on necessity, 255
- more-tahn-two-place 'consequence' relations, 266
- Moschovakis, J., 95
- Myhill, 97, 99
- naive theory of definite descriptions (NTDD), 238
- Natural Deduction, 10, 11
- natural deduction, 4
- natural negation, 282, 283
- necessity operator, 255
- necessity predicate, 255
- necessity, metalinguistic interpretation of, 255
- negative free logic (NFL), 225
- negative part of formula, 227
- negative semantics, 221, 225
- Negri, S., 18
- neighbourhood, 23
- neutral semantics, 221, 233
- new foundations (NF), 217
- Nishimura, T., 40
- non-deterministic algorithms, 293
- non-monotonic matrices, 311
- non-strict function, 219
- normal form theorem, 18
- normalisation theorem, 18
- objectual quantification, 200
- objectual quantification, in story semantics, 223
- Ono, H., 60
- ontological argument and Russellian descriptions, 244
- outer domain semantics, 218, 221
- partial element, 102
- partial interpretation, 218, 222
- partial interpretation, completion of, 228
- partial recursive predicates, 286
- path, 25
- PEM, 3

- Plato, J. von, 18  
 polar replacement, 227  
 positive free logic (PFL), 222  
 positive part of formula, 227  
 positive semantics, 221  
 possible worlds, 212  
 possible worlds and propositions,  
     241  
 Posy, C., 102  
 Prawitz, D., 17, 18, 49, 74, 85  
 predicate/singular-term composi-  
     tion, 263, 298  
 prenex fragment, 48  
 presupposition, 210, 220, 278, 281,  
     282, 307, 309, 311, 314  
 presuppositional analysis, 313  
 pretend objects, 202, 218, 224  
 principle of open data, 89  
 principle of the excluded third, 3  
 projection rules for presupposition,  
     309  
 proof interpretation, 6  
 proof-interpretation, 4  
 proof-terms, 19  
 propositional content, 214  
 provably recursive functions, 73  
  
 quantification, vacuous, 212  
 Quine's dictum, 197, 207  
 Quine, on classes, 217  
 Quine, on descriptions, 237  
 Quine, on eliminating singular terms,  
     251  
 Quine, on inclusive logic, 212  
 Quine, on predication, 250  
 Quine, on set theory, 217  
 quotational logic, 285  
  
 Rasiowa, H., 25, 38  
 Rautenberg, H., 53  
 Rautenberg, W., 59  
 realizability, 5  
 recursively axiomatisable, 229  
 reference failure, 297  
  
 referential opacity, 215  
 replacement, 226  
 Rieger–Nishimura lattice, 40  
 Robinson, on descriptions, 248  
 Russell's paradox, 218, 238  
 Russell, on descriptions, 204, 216,  
     240  
 Russell, on predication, 209  
  
 S5 semantics, 255  
 satisfiable, 200  
 Scales, on complex predicates, 209  
 Scales, on descriptions, 217, 245  
 Schütte, K., 25, 49  
 Schroeder-Heister, P., 18  
 Schweizer, on necessity, 255  
 Schwichtenberg, H., 49  
 scope  
     in natural language, 241  
     indicators of, 241  
     narrow, 240  
     of Russellian descriptions, 240  
     wide, 240  
 Scott, D., 54, 105, 106  
 Scowcroft, P., 99  
 second-order logic, 84  
 second-order quantification, and su-  
     pervaluation, 229  
 selective filtration, 55  
 semantic paradox, 284  
 sense, 299–301, 303  
 Sequent Calculus, 17  
 sheaf interpretation, 106  
 Sikorski, R., 25, 38  
 singular predicate, 204, 210, 237  
 situation semantics, 293  
 skeleton, 141  
 Skolem functions, 50  
 Skyrms, on necessity, 255  
 Skyrms, on supervaluation, 231  
 Smiley, on attribution, 220  
 Smiley, on quantification in netu-  
     ral semantics, 235

- Smoryński, C., 46, 50, 54, 55, 65, 71, 76
- sortal incorrectness, 282
- stable formula, 229
- stable open sentence, 252
- Statman, R., 63
- Stenlund, on descriptions, 248
- Stenlund, on fictional entities, 215
- story, 220, 222
  - actualist constraints on, 230
  - story interpretation, 222
  - story semantics, 220
  - story semantics and bivalence, 220
  - story semantics, equivalence to outer-domain semantics, 224
- strategy, 115, 117
- stratified formula, 217
- Strawson, on presupposition, 210
- strong completeness, 200
- strong continuity principle, 92, 93
- strong negation, 77
- strong normalisation theorem, 18
- strong tables, 228
- subformula property, 18
- substitutional quantification, 200
- Sundholm, G., 7
- superfalsity, 228
- supertruth, 228
- supervaluation, 221, 228
- supervaluations, 272, 283
- Swart, H. C. M. de, 52
- Tarski, A., 22, 61
- term-forming descriptions operator, 264
- theories in partial logic, 266
- theory of apartness, 62
- theory of equality, 61
- theory of order, 65
- topological interpretation, 4, 22
- topological space, 22
- topos theory, 83
- transplicand, 307, 313
- transplication, 264, 274, 280, 281, 295, 304, 307, 310, 313
- transplicator, 307, 313
- Trew, on free logics as first-order theories, 202
- Troelstra, A., 49, 82, 86
- truth connective, 237, 249
- truth, counterfactual theory of, 232
- uniformity principle, 86
- universal Beth model, 52
- universally free, 197, 218
- unsolved problem, 5
- Veldman, W., 44, 52
- virtual classes, 217
- Visser, A., 72
- Walton, on pretense, 224
- weak completeness, 201
- weak tables, 228
- Woodruff, on Frege, 234
- Woodruff, on supervaluation, 229

# Handbook of Philosophical Logic

2nd Edition

Volume 6

edited by Dov M. Gabbay and F. Guentner



## CONTENTS

Editorial Preface	vii
<b>Dov M. Gabbay</b>	
Relevance Logic	1
<b>Mike Dunn and Greg Restall</b>	
Quantum Logics	129
<b>Maria-Luisa Dalla Chiara and Roberto Giuntini</b>	
Combinators, Proofs and Implicational Logics	229
<b>Martin Bunder</b>	
Paraconsistent Logic	287
<b>Graham Priest</b>	
Index	395





## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London

Logic	IT			
	Natural language processing	Program control specification, verification, concurrency	Artificial intelligence	Logic programming
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs

<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical fragments</b>	Basic back-ground language	Program synthesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely useful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calculus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syntax	Modules. Combining languages	Logics of space and time	Combining features
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game semantics gaining ground		
<b>Object level/metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action-revision models</b>			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model





## RELEVANCE LOGIC

### 1 INTRODUCTION

#### 1.1 *Delimiting the topic*

The title of this piece is not ‘A Survey of Relevance Logic’. Such a project was impossible in the mid 1980s when the first version of this article was published, due to the development of the field and even the space limitations of the *Handbook*. The situation is if anything, more difficult now. For example Anderson and Belnap and Dunn’s two volume [1975; 1992] work *Entailment: The Logic of Relevance and Necessity*, runs to over 1200 pages, and is their summary of just some of the work done by them and their co-workers up to about the late 1980s. Further, the comprehensive bibliography (prepared by R. G. Wolf) contains over 3000 entries in work on relevance logic and related fields.

So, we need some way of delimiting our topic. To be honest the fact that *we* are writing this is already a kind of delimitation. It is natural that you shall find emphasised here the work that we happen to know best. But still rationality demands a less subjective rationale, and so we will proceed as follows.

Anderson [1963] set forth some open problems for his and Belnap’s system **E** that have given shape to much of the subsequent research in relevance logic (even much of the earlier work can be seen as related to these open problems, e.g. by giving rise to them). Anderson picks three of these problems as major: (1) the admissibility of Ackermann’s rule  $\gamma$  (the reader should not worry that he is expected to already know what this means), (2) the decision problems, (3) the providing of a semantics. Anderson also lists additional problems which he calls ‘minor’ because they have no ‘philosophical bite’. We will organise our remarks on relevance logic around three major problems of Anderson. The reader should be told in advance that each of these problems are closed (but of course ‘closed’ does not mean ‘finished’—closing one problem invariably opens another related problem). This gives then three of our sections. It is obvious that to these we must add an introduction setting forth at least some of the motivations of relevance logic and some syntactical specifications. To the end we will add a section which situates work in relevance logic in the wider context of study of other logical systems, since in the recent years it has become clear that relevance logics fit well among a wider class of ‘resource-conscious’ or ‘substructural’ logics [Schroeder-Heister and Došen, 1993; Restall, 2000] [and cite the S–H article in this volume]. We thus have the following table of contents:

1. Introduction
2. The Admissibility of  $\gamma$
3. Semantics
4. The Decision Problem
5. Looking About

We should add a word about the delimitation of our topic. There are by now a host of formal systems that can be said with some justification to be ‘relevance logics’. Some of these antedate the Anderson–Belnap approach, some are more recent. Some have been studied somewhat extensively, whereas others have been discussed for only a few pages in some journal. It would be impossible to describe all of these, let alone to assess in each and every case how they compare with the Anderson–Belnap approach. It is clear that the Anderson–Belnap-style logics have been the most intensively studied. So we will concentrate on the research program of Anderson, Belnap and their co-workers, and shall mention other approaches only insofar as they bear on this program. By way of minor recompense we mention that Anderson and Belnap [1975] have been good about discussing related approaches, especially the older ones.

Finally, we should say that our paradigm of a relevance logic throughout this essay will be the Anderson–Belnap system **R** or relevant implication (first devised by Belnap—see [Belnap, 1967a; Belnap, 1967b] for its history) and not so much the Anderson–Belnap favourite, their system **E** of entailment. There will be more about each of these systems below (they are explicitly formulated in Section 1.3), but let us simply say here that each of these is concerned to formalise a species of implication (or the conditional—see Section 1.2) in which the antecedent suffices *relevantly* for the consequent. The system **E** differs from the system **R** primarily by adding necessity to this relationship, and in this **E** is a modal logic as well as a relevance logic. This by itself gives good reason to consider **R** and not **E** as the paradigm of a relevance logic.<sup>1</sup>

## 1.2 *Implication and the Conditional*

Before turning to matters of logical substance, let us first introduce a framework for grammar and nomenclature that is helpful in understanding the ways that writers on relevance logic often express themselves. We draw

---

<sup>1</sup>It should be entered in the record that there are some workers in relevance logic who consider both **R** and **E** too strong for at least some purposes (see [Routley, 1977], [Routley *et al.*, 1982], and more recently, [Brady, 1996]).

heavily on the ‘Grammatical Propaedeutic’ appendix of [Anderson and Belnap, 1975] and to a lesser extent on [Meyer, 1966], both of which are very much recommended to the reader for their wise heresy from logical tradition.

Thus logical tradition (think of [Quine, 1953]) makes much of the grammatical distinction between ‘if, then’ (a connective), and ‘implies’ or its rough synonym ‘entails’ (transitive verbs). This tradition opposes

1. If today is Tuesday, then this is Belgium

to the pair of sentences

2. ‘Today is Tuesday’ implies ‘This is Belgium’,
3. That today is Tuesday implies that this is Belgium.

And the tradition insists that (1) be called a *conditional*, and that (2) and (3) be called *implications*.

Sometimes much philosophical weight is made to rest on this distinction. It is said that since ‘implies’ is a verb demanding nouns to flank it, that implication must then be a relation between the objects stood for by those nouns, whereas it is said that ‘if, then’ is instead a connective combining that implication (unlike ‘if, then’) is really a metalinguistic notion, either overtly as in (2) where the nouns are names of sentences, or else covertly as in (3) where the nouns are naming propositions (the ‘ghosts’ of linguistic entities). This last is then felt to be especially bad because it involves ontological commitment to propositions or some equally disreputable entities. The first is at least free of such questionable ontological commitments, but does raise real complications about ‘nested implications’, which would seem to take us into a meta-metalanguage, etc.

The response of relevance logicians to this distinction has been largely one of ‘What, me worry?’ Sometime sympathetic outsiders have tried to apologise for what might be quickly labelled a ‘use–mention confusion’ on the part of relevance logicians [Scott, 1971]. But ‘hard-core’ relevance logicians often seem to luxuriate in this ‘confusion’. As Anderson and Belnap [1975, p. 473] say of their ‘Grammatical Propaedeutic’: “the principle aim of this piece is to convince the reader that it is philosophically respectable to ‘confuse’ implication or entailment with the conditional, and indeed philosophically suspect to harp on the dangers of such a ‘confusion’. (The suspicion is that such harpists are plucking a metaphysical tune on merely grammatical strings.)”

The gist of the Anderson–Belnap position is that there is a generic conditional-implication notion, which can be carried into English by a variety of grammatical constructions. Implication itself can be viewed as a connective requiring prenominalisation: ‘that \_\_\_ implies that \_\_\_’, and as such it nests. It is an incidental feature of English that it favours sentences with main subjects and verbs, and ‘implies’ conforms to this reference by

the trick of disguising sentences as nouns by prenominalisation. But such grammatical prejudices need not be taken as enshrining ontological presuppositions.

Let us use the label ‘Correspondence Thesis’ for the claim that Anderson and Belnap come close to making (but do not actually make), namely, that *in general* there is nothing other than a purely grammatical distinction between sentences of the forms

4. If  $A$ , then  $B$ , and
5. That  $A$  implies that  $B$ .

Now undoubtedly the Correspondence Thesis overstates matters. Thus, to bring in just one consideration, [Castañeda, 1975, pp. 66 ff.] distinguishes ‘if  $A$  then  $B$ ’ from ‘ $A$  only if  $B$ ’ by virtue of an essentially pragmatic distinction (frozen into grammar) of ‘thematic’ emphases, which cuts across the logical distinction of antecedent and consequent. Putting things quickly, ‘if’ introduces a sufficient condition for something happening, something being done, etc. whereas ‘only if’ introduces a necessary condition. Thus ‘if’ (by itself or prefixed with ‘only’) always introduces the state of affairs thought of as a condition for something else, then something else being thus the focus of attention. Since ‘that  $A$  implies that  $B$ ’ is devoid of such thematic indicators, it is not equivalent at *every* level of analysis to either ‘if  $A$  then  $B$ ’ or ‘ $A$  only if  $B$ ’.

It is worth remarking that since the formal logician’s  $A \rightarrow B$  is equally devoid of thematic indicators, ‘that  $A$  implies that  $B$ ’ would seem to make a better reading of it than either ‘if  $A$  then  $B$ ’ or ‘ $A$  only if  $B$ ’. And yet it is almost universally rejected by writers of elementary logic texts as even an acceptable reading.

And, of course, another consideration against the Correspondence Thesis is produced by notorious examples like Austin’s

6. There are biscuits on the sideboard if you want some,

which sounds very odd indeed when phrased as an implication. Indeed, (6) poses perplexities of one kind or another for any theory of the conditional, and so should perhaps best be ignored as posing any special threat to the Anderson and Belnap account of conditionals. Perhaps it was Austin-type examples that led Anderson and Belnap [1975, pp. 491–492] to say “we think every use of ‘implies’ or ‘entails’ as a connective can be replaced by a suitable ‘if-then’; however, the converse may not be true”. They go on to say “But with reference to the uses in which we are primarily interested, we feel free to move back and forth between ‘if-then’ and ‘entails’ in a free-wheeling manner”.

Associated with the Correspondence Thesis is the idea that just as there can be contingent conditionals (e.g. (1)), so then the corresponding implications (e.g. (3)) must also be contingent. This goes against certain Quinean

tendencies to ‘regiment’ the English word ‘implies’ so that it stands only for *logical* implication. Although there is no objection to thus giving a technical usage to an ordinary English word (even requiring in this technical usage that ‘implication’ be a metalinguistic relation between sentences), the point is that relevance logicians by and large believe we are using ‘implies’ in the ordinary non-technical sense, in which a sentence like (3) might be true without there being any logical (or even necessary) implication from ‘Today is Tuesday’ to ‘This is Belgium’.

Relevance logicians are not themselves free of similar regimenting tendencies. Thus we tend to differentiate ‘entails’ from ‘implies’ on precisely the ground that ‘entails’, unlike ‘implies’, stands only for *necessary* implication [Meyer, 1966]. Some writings of Anderson and Belnap even suggest a more restricted usage for just *logical* implication, but we do not take this seriously. There does not seem to be any more linguistic evidence for thus restricting ‘entails’ than there would be for ‘implies’, though there may be at least more excuse given the apparently more technical history of ‘entails’ (in its logical sense—cf. The OED).

This has been an explanation of, if not an apology for, the ways in which relevance logicians often express themselves. but it should be stressed that the reader need not accept all, or any, of this background in order to make sense of the basic aims of the relevance logic enterprise. Thus, e.g. the reader may feel that, despite protestations to the contrary, Anderson, Belnap and Co. are hopelessly confused about the relationships among ‘entails’, ‘implies’, and ‘if-then’, but still think that their system **R** provides a good formalisation of the properties of ‘if-then’ (or at least ‘if-then relevantly’), and that they system **E** does the same for some strict variant produced by the modifier ‘necessarily’.

One of the reasons the recent logical tradition has been motivated to insist on the fierce distinction between implications and conditionals has to do with the awkwardness of reading the so-called ‘material conditional’  $A \rightarrow B$  as corresponding to any kind of implication (cf. [Quine, 1953]).

The material conditional  $A \rightarrow B$  can of course be defined as  $\neg A \vee B$ , and it certainly does seem odd, modifying an example that comes by oral tradition from Anderson, to say that:

7. Picking a guinea pig up by its tail implies that its eyes will fall out.

just on the grounds that its antecedent is false (since guinea pigs have no tails). But then it seems equally false to say that:

8. If one picks up a guinea pig by its tail, then its eyes will fall out.

And also both of the following appear to be equally false:

9. Scaring a pregnant guinea pig implies that all of her babies will be born tailless.

10. If one scares a pregnant guinea pig, then all of her babies will be born tailless.

It should be noted that there are other ways to react to the oddity of sentences like the ones above other than calling them simply false. Thus there is the reaction stemming from the work of Grice [1975] that says that at least the conditional sentences (8) and (10) above are true though nonetheless pragmatically odd in that they violate some rule based on conversational co-operation to the effect that one should normally say the strongest thing relevant, i.e. in the cases above, that guinea pigs have no tails (cf. [Fogelin, 1978, p. 136 ff.] for a textbook presentation of this strategy).

Also it should be noted that the theory of the ‘counterfactual’ conditional due to Stalnaker–Thomason, D. K. Lewis and others (cf. Chapter [??] of this *Handbook*), while it agrees with relevance logic in finding sentences like (8) (not (10) *false*, disagrees with relevance logic in the formal account it gives of the conditional.

It would help matters if there were an extended discussion of these competing theories (Anderson–Belnap, Grice, Stalnaker–Thomason–Lewis), which seem to pass like ships in the night (can three ships do this without strain to the image?) but there is not the space here. Such a discussion might include an attempt to construct a theory of a relevant counterfactual conditional (if  $A$  were to be the case, then *as a result*  $B$  would be the case). The rough idea would be to use say The Routley–Meyer semantics for relevance logic (cf. Section 3.7) in place of the Kripke semantics for modal logic, which plays a key role in the Stalnaker–Thomason–Lewis semantical account of the conditional (put the 3-placed alternativeness relation in the role of the usual 2-placed one). Work in this area is just starting. See the works of [Mares and Fuhrmann, 1995] and [Akama, 1997] which both attempt to give semantics for relevant counterfactuals.

Also any discussion relating to Grice’s work would surely make much of the fact that the theory of Grice makes much use of a basically unanalysed notion of relevance. One of Grice’s chief conversational rules is ‘be relevant’, but he does not say much about just what this means. One could look at relevance logic as trying to say something about this, at least in the case of the conditional.

Incidentally, as Meyer has been at great pains to emphasise, relevance logic gives, on its face anyway, no separate account of relevance. It is not as if there is a unary relevance operator (‘relevantly’).

One last point, and then we shall turn to more substantive issues. Orthodox relevance logic differs from classical logic not just in having an additional logical connective ( $\rightarrow$ ) for the conditional. If that was the only difference relevance logic would just be an ‘extension’ of classical logic, using the notion of Haack [1974], in much the same way as say modal logic is an extension of classical logic by the addition of a logical connective  $\square$

for necessity. The fact is (cf. Section 1.6) that although relevance logic contains all the same theorems as classical logic in the classical vocabulary say,  $\wedge, \vee, \neg$  (and the quantifiers), it nonetheless does not validate the same inferences. Thus, most notoriously, the disjunctive syllogism (cf. Section 2) is counted as invalid. Thus, as Wolf [1978] discusses, relevance logic does not fit neatly into the classification system of [Haack, 1974], and might best be called ‘quasi-extension’ of classical logic, and hence ‘quasi-deviant’. Incidentally, all of this applies only to ‘orthodox’ relevance logic, and not to the ‘classical relevance logics’ of Meyer and Routley (cf. Section 3.11).

### 1.3 Hilbert-style Formulations

We shall discuss first the pure implicational fragments, since it is primarily in the choice of these axioms that the relevance logics differ one from the other. We shall follow the conventions of Anderson and Belnap [Anderson and Belnap, 1975], denoting by ‘ $\mathbf{R}_{\rightarrow}$ ’ what might be called the ‘putative implicational fragment of  $\mathbf{R}$ ’. Thus  $\mathbf{R}_{\rightarrow}$  will have as axioms all the axioms of  $\mathbf{R}$  that only involve the implication connective. That  $\mathbf{R}_{\rightarrow}$  is in fact the implicational fragment of  $\mathbf{R}$  is much less than obvious since the possibility exists that the proof of a pure implicational formula could detour in an essential way through formulas involving connectives other than implication. In fact Meyer has shown that this does not happen (cf. his Section 28.3.2 of [Anderson and Belnap, 1975]), and indeed Meyer has settled in almost every interesting case that the putative fragments of the well-known relevance logics (at least  $\mathbf{R}$  and  $\mathbf{E}$ ) are the same as the real fragments. (Meyer also showed that this does not happen in one interesting case,  $\mathbf{RM}$ , which we shall discuss below.)

For  $\mathbf{R}_{\rightarrow}$  we take the rule *modus ponens* ( $A, A \rightarrow B \vdash B$ ) and the following axiom schemes.

$$\begin{array}{ll} A \rightarrow A & \text{Self-Implication} \quad (1) \\ (A \rightarrow B) \rightarrow [(C \rightarrow A) \rightarrow (C \rightarrow B)] & \text{Prefixing} \quad (2) \\ [A \rightarrow (A \rightarrow B)] \rightarrow (A \rightarrow B) & \text{Contraction} \quad (3) \\ [A \rightarrow (B \rightarrow C)] \rightarrow [B \rightarrow (A \rightarrow C)] & \text{Permutation.} \quad (4) \end{array}$$

A few comments are in order. This formulation is due to Church [1951b] who called it ‘The weak implication calculus’. He remarks that the axioms are the same as those of Hilbert’s for the positive implicational calculus (the implicational fragment of the intuitionistic propositional calculus  $\mathbf{H}$ ) except that (1) is replaced with

$$A \rightarrow (B \rightarrow A) \quad \text{Positive Paradox.} \quad (1')$$

(Recent historical investigation by Došen [1992] has shown that Orlov constructed an axiomatisation of the implication and negation fragment of  $\mathbf{R}$

in the mid 1920s, predating other known work in the area. Church and Moh, however, provided a Deduction Theorem (see Section 1.4) which is absent from Orlov's treatment.)

The choice of the implicational axioms can be varied in a number of informative ways. Thus putting things quickly, (2) Prefixing may be replaced by

$$(A \rightarrow B) \rightarrow [(B \rightarrow C) \rightarrow (A \rightarrow C)] \quad \text{Suffixing.} \quad (2')$$

(3) Contraction may be replaced by

$$[A \rightarrow (B \rightarrow C)] \rightarrow [(A \rightarrow B) \rightarrow (A \rightarrow C)] \quad \text{Self-Distribution,} \quad (3')$$

and (4) Permutation may be replaced by

$$A \rightarrow [(A \rightarrow B) \rightarrow B] \quad \text{Assertion.} \quad (4')$$

These choices of implicational axioms are 'isolated' in the sense that one choice does not affect another. Thus

**THEOREM 1.**  $\mathbf{R}_{\rightarrow}$  may be axiomatised with modus ponens, (1) *Self-Implication* and any selection of one from each pair  $\{(2), (2')\}$ ,  $\{(3), (3')\}$ , and  $\{(4), (4')\}$ .

**Proof.** By consulting [Anderson and Belnap, 1975, pp. 79–80], and fiddling. ■

There is at least one additional variant of  $\mathbf{R}_{\rightarrow}$  that merits discussion. It turns out that it suffices to have Suffixing, Contraction, and the pair of axiom schemes

$$[(A \rightarrow A) \rightarrow B] \rightarrow B \quad \text{Specialised Assertion,} \quad (4a)$$

$$A \rightarrow [(A \rightarrow A) \rightarrow A] \quad \text{Demodaliser.} \quad (4b)$$

Thus (4b) is just an instance of Assertion, and (4a) follows from Assertion by substitution  $A \rightarrow A$  for  $A$  and using Self-Implication to detach. That (4a) and (4b) together with Suffixing and Contraction yield Assertion (and, less interestingly, Self-Implication) can be shown using the fact proven in [Anderson and Belnap, 1975, Section 8.3.3], that these yield (letting  $\vec{A}$  abbreviate  $A_1 \rightarrow A_2$ )

$$\vec{A} \rightarrow [(\vec{A} \rightarrow B) \rightarrow B] \quad \text{Restricted-Assertion.} \quad (4'')$$

The point is that (4a) and (4b) in conjunction say that  $A$  is equivalent to  $(A \rightarrow A) \rightarrow A$ , and so every formula  $A$  has an equivalent form  $\vec{A}$  and so 'Restricted Assertion' reduces to ordinary Assertion.<sup>2</sup>

<sup>2</sup>There are some subtleties here. Detailed analysis shows that both Suffixing and Prefixing are needed to replace  $\vec{A}$  with  $A$  (cf. Section 1.3). Prefixing can be derived from the above set of axioms (cf. [Anderson and Belnap, 1975, pp. 77–78 and p. 26]).



Incidentally, no claim is made that this last variant of  $\mathbf{R}_{\rightarrow}$  has the same isolation in its axioms as did the previous axiomatisations. Thus, e.g. that Suffixing (and not Prefixing) is an axiom is important (a matrix of J. R. Chidgey's (cf. [Anderson and Belnap, 1975, Section 8.6]) can be used to show this.

The system  $\mathbf{E}$  of entailment differs primarily from the system  $\mathbf{R}$  in that it is a system of relevant strict implication. Thus  $\mathbf{E}$  is both a relevance logic and a modal logic. Indeed, defining  $\Box A =_{\text{df}} (A \rightarrow A) \rightarrow A$  one finds  $\mathbf{E}$  has something like the modality structure of  $\mathbf{S4}$  (cf. [Anderson and Belnap, 1975, Sections 4.3 and 10]).

This suggests that  $\mathbf{E}_{\rightarrow}$  can be axiomatised by dropping Demodaliser from the axiomatisation of  $\mathbf{R}_{\rightarrow}$ , and indeed this is right (cf. [Anderson and Belnap, 1975, Section 8.3.3], for this and all other claims about axiomatisations of  $\mathbf{E}_{\rightarrow}$ ).<sup>3</sup>

The axiomatisation above is a 'fixed menu' in that Prefixing cannot be replaced with Suffixing. There are other 'à la carte' axiomatisations in the style of Theorem 1.

**THEOREM 2.**  $\mathbf{E}_{\rightarrow}$  may be axiomatised with modus ponens, *Self-Implication* and any selection from each of the pairs  $\{\text{Prefixing}, \text{Suffixing}\}$ ,  $\{\text{Contraction}, \text{Self-Distribution}\}$  and  $\{\text{Restricted-Permutation}, \text{Restricted-Assertion}\}$  (one from each pair).

Another implicational system of less central interest is that of 'ticket entailment'  $\mathbf{T}_{\rightarrow}$ . It is motivated by Anderson and Belnap [1975, Section 6] as deriving from some ideas of Ryle's about 'inference tickets'. It was motivated in [Anderson, 1960] as 'entailment shorn of modality'. The thought behind this last is that there are two ways to remove the modal sting from the characteristic axiom of alethic modal logic,  $\Box A \rightarrow A$ . One way is to add Demodaliser  $A \rightarrow \Box A$  so as to destroy all modal distinctions. The other is to drop the axiom  $\Box A \rightarrow A$ . Thus the essential way one gets  $\mathbf{T}_{\rightarrow}$  from  $\mathbf{E}_{\rightarrow}$  is to drop Specialised Assertion (or alternatively to drop Restricted Assertion or Restricted Permutation, depending on which axiomatisation of  $\mathbf{E}_{\rightarrow}$  one has). But before doing so one must also add whichever one of Prefixing and Suffixing was lacking, since it will no longer be a theorem otherwise (this is easiest to visualise if one thinks of dropping Restricted permutation, since this is the key to getting Prefixing from Suffixing and *vice versa*). Also (and this is a strange technicality) one must replace Self-Distribution with its permuted form:

$$(A \rightarrow B) \rightarrow [[A \rightarrow (B \rightarrow C)] \rightarrow (A \rightarrow C)] \quad \text{Permuted Self-Distribution.} \quad (3'')$$

This is summarised in

---

<sup>3</sup>The actual history is backwards to this, in that the system  $\mathbf{R}$  was first axiomatised by [Belnap, 1967a] by adding Demodaliser to  $\mathbf{E}$ .

**THEOREM 3** (Anderson and Belnap [Section 8.3.2, 1975]).  $\mathbf{T}_{\rightarrow}$  is axiomatised using *Self-Implication, Prefixing, Sufficing, and either of {Contraction, Permuted Self-Distribution}*, with modus ponens.

There is a subsystem of  $\mathbf{E}_{\rightarrow}$  called  $\mathbf{TW}_{\rightarrow}$  (and  $\mathbf{P-W}$ , and  $\mathbf{T-W}$  in earlier nomenclature) axiomatised by dropping Contraction (which corresponds to the combinator  $\mathbf{W}$ ) from  $\mathbf{T}_{\rightarrow}$ . This has obtained some interest because of an early conjecture of Belnap's (cf. [Anderson and Belnap, 1975, Section 8.11]) that  $A \rightarrow B$  and  $B \rightarrow A$  are both theorems of  $\mathbf{TW}_{\rightarrow}$  only when  $A$  is the same formula as  $B$ . That Belnap's Conjecture is now Belnap's Theorem is due to the highly ingenious (and complicated) work of E. P. Martin and R. K. Meyer [1982] (based on the earlier work of L. Powers and R. Dwyer). Martin and Meyer's work also highlights a system  $\mathbf{S}_{\rightarrow}$  (for Syllogism) in which Self-Implication is dropped from  $\mathbf{TW}_{\rightarrow}$ .

Moving on now to adding the positive extensional connectives  $\wedge$  and  $\vee$ , in order to obtain  $\mathbf{R}_{\rightarrow, \wedge, \vee}$  (denoted more simply as  $\mathbf{R}^+$ ) one adds to  $\mathbf{R}_{\rightarrow}$  the axiom schemes

$$A \wedge B \rightarrow A, \quad A \wedge B \rightarrow B \quad \text{Conjunction Elimination} \quad (5)$$

$$[(A \rightarrow B) \wedge (A \rightarrow C)] \rightarrow (A \rightarrow B \wedge C) \quad \text{Conjunction Introduction} \quad (6)$$

$$A \rightarrow A \vee B, \quad B \rightarrow A \vee B \quad \text{Disjunction Introduction} \quad (7)$$

$$[(A \rightarrow C) \wedge (B \rightarrow C)] \rightarrow (A \vee B \rightarrow C) \quad \text{Disjunction Elimination} \quad (8)$$

$$A \wedge (B \vee C) \rightarrow (A \wedge B) \vee C \quad \text{Distribution} \quad (9)$$

plus the rule of adjunction ( $A, B \vdash A \wedge B$ ). One can similarly get the positive intuitionistic logic by adding these all to  $\mathbf{H}_{\rightarrow}$ .

Axioms (5)–(8) can readily be seen to be encoding the usual elimination and introduction rules for conjunction and disjunction into axioms, giving  $\wedge$  and  $\vee$  what might be called 'the lattice properties' (cf. Section 3.3). It might be thought that  $A \rightarrow (B \rightarrow A \wedge B)$  might be a better encoding of conjunction introduction than (6), having the virtue that it allows for the dropping of adjunction. This is a familiar axiom for intuitionistic (and classical) logic, but as was seen by Church [1951b], it is only a hair's breadth away from Positive Paradox ( $A \rightarrow (B \rightarrow A)$ ), and indeed yields it given (5) and Prefixing. For some mysterious reason, this observation seemed to prevent Church from adding extensional conjunction/disjunction to what we now call  $\mathbf{R}_{\rightarrow}$  (and yet the need for adjunction in the Lewis formulations of modal logic where the axioms are all strict implications was well-known).

Perhaps more surprising than the need for adjunction is the need for axiom (9). It would follow from the other axioms if only we had Positive Paradox among them. The place of Distribution in  $\mathbf{R}$  is continually problematic. It causes inelegancies in the natural deduction systems (cf. Section 1.5) and is an obstacle to finding decision procedures (cf. Section 4.8). Incidentally, all of the usual distributive laws follow from the somewhat 'clipped' version

(9).

The rough idea of axiomatising  $\mathbf{E}^+$  and  $\mathbf{T}^+$  is to add axiom schemes (5)–(9) to  $\mathbf{E}_{\rightarrow}$  and  $\mathbf{T}_{\rightarrow}$ . This is in fact precisely right for  $\mathbf{T}^+$ , but for  $\mathbf{E}^+$  one needs also the axiom scheme (remember  $\Box A =_{\text{df}} (A \rightarrow A) \rightarrow A$ ):

$$\Box A \wedge \Box B \rightarrow \Box(A \wedge B) \quad (10)$$

This is frankly an inelegance (and one that strangely enough disappears in the natural deduction context of Section 1.5). It is needed for the inductive proof that necessitation ( $\vdash C \Rightarrow \vdash \Box C$ ) holds, handling the case where  $C$  just came by adjunction (cf. [Anderson and Belnap, 1975, Sections 21.2.2 and 23.4]). There are several ways of trying to conceal this inelegance, but they are all a little *ad hoc*. Thus, e.g. one could just postulate the rule of necessitation as primitive, or one could strengthen the axiom of Restricted Permutation (or Restricted Assertion) to allow that  $\vec{A}$  be a conjunction  $(A_1 \rightarrow A_1) \wedge (A_2 \rightarrow A_2)$ .

As Anderson and Belnap [1975, Section 21.2.2] remark, if propositional quantification is available,  $\Box A$  could be given the equivalent definition  $\forall p(p \rightarrow p) \rightarrow A$ , and then the offending (10) becomes just a special case of Conjunction Introduction and becomes redundant.

It is a good time to advertise that the usual zero-order and first-order relevance logics can be outfitted with a couple of optional convenience features that come with the higher-priced versions with propositional quantifiers. Thus, e.g. the propositional constant  $t$  can be added to  $\mathbf{E}^+$  to play the role of  $\forall p(p \rightarrow p)$ , governed by the axioms.

$$(t \rightarrow A) \rightarrow A \quad (11)$$

$$t \rightarrow (A \rightarrow A), \quad (12)$$

and again (10) becomes redundant (since one can easily show  $(t \rightarrow A) \leftrightarrow [(A \rightarrow A) \rightarrow A]$ ).

Further, this addition of  $t$  is conservative in the sense that it leads to no new  $t$ -free theorems (since in any given proof  $t$  can always be replaced by  $(p_1 \rightarrow p_1) \wedge \dots \wedge (p_n \rightarrow p_n)$  where  $p_1, \dots, p_n$  are all the propositional variables appearing in the proof — cf. [Anderson and Belnap, 1975]).

Axiom scheme (11) is too strong for  $\mathbf{T}^+$  and must be weakened to

$$t. \quad (11\mathbf{T})$$

In the context of  $\mathbf{R}^+$ , (11) and (11 $\mathbf{T}$ ) are interchangeable. and in  $\mathbf{R}^+$ , (12) may of course be permuted, letting us characterise  $t$  in a single axiom as ‘the conjunction of all truths’:

$$A \leftrightarrow (t \rightarrow A) \quad (13)$$

(in  $\mathbf{E}$ ,  $t$  may be thought of as ‘the conjunction of all necessary truths’).

‘Little  $t$ ’ is distinguished from ‘big  $T$ ’, which can be conservatively added with the axiom scheme

$$A \rightarrow T \quad (14)$$

(in intuitionistic or classical logic  $t$  and  $T$  are equivalent).

Additionally useful is a binary connective  $\circ$ , labelled variously ‘intensional conjunction’, ‘fusion’, ‘consistency’ and ‘cotenability’. these last two labels are appropriate only in the context of  $\mathbf{R}$ , where one can define  $A \circ B =_{\text{df}} \neg(A \rightarrow \neg B)$ . One can add  $\circ$  to  $\mathbf{R}^+$  with the axiom scheme:

$$[(A \circ B) \rightarrow C] \leftrightarrow [A \rightarrow (B \rightarrow C)] \quad \text{Residuation (axiom).} \quad (15)$$

This axiom scheme is too strong for other standard relevance logics, but Meyer and Routley [1972] discovered that one can always add conservatively the two way rule

$$(A \circ B) \rightarrow C \dashv \vdash A \rightarrow (B \rightarrow C) \quad \text{Residuation (rule)} \quad (16)$$

(in  $\mathbf{R}^+$  (16) yields (15)). Before adding negation, we mention the positive fragment  $\mathbf{B}^+$  of a kind of minimal (Basic) relevance logic due to Routley and Meyer (cf. Section 3.9).  $\mathbf{B}^+$  is just like  $\mathbf{TW}^+$  except for finding the axioms of Prefixing and Suffixing too strong and replacing them by rules:

$$A \rightarrow B \vdash (C \rightarrow A) \rightarrow (C \rightarrow B) \quad \text{Prefixing (rule)} \quad (17)$$

$$A \rightarrow B \vdash (B \rightarrow C) \rightarrow (A \rightarrow C) \quad \text{Suffixing (rule)} \quad (18)$$

As for negation, the full systems  $\mathbf{R}$ ,  $\mathbf{E}$ , etc. may be formed adding to the axiom schemes for  $\mathbf{R}^+$ ,  $\mathbf{E}^+$ , etc. the following<sup>4</sup>

$$(A \rightarrow \neg A) \rightarrow \neg A \quad \text{Reductio} \quad (19)$$

$$(A \rightarrow \neg B) \rightarrow (B \rightarrow \neg A) \quad \text{Contraposition} \quad (20)$$

$$\neg \neg A \rightarrow A \quad \text{Double Negation.} \quad (21)$$

Axiom schemes (19) and (20) are intuitionistically acceptable negation principles, but using (21) one can derive forms of reductio and contraposition that are intuitionistically rejectable. Note that (19)–(21) if added to  $\mathbf{H}^+$  would give the full intuitionistic propositional calculus  $\mathbf{H}$ .

In  $\mathbf{R}$ , negation can alternatively be defined in the style of Johansson, with  $\neg A =_{\text{df}} (A \rightarrow f)$ , where  $f$  is a false propositional constant, cf. [Meyer, 1966]. Informally,  $f$  is the disjunction of all false propositions (the ‘negation’ of  $t$ ). Defining negation thus, axiom schemes (19) and (20) become theorems

<sup>4</sup>Reversing what is customary in the literature, we use  $\neg$  for the standard negation of relevance logic, reserving  $\sim$  for the ‘Boolean negation’ discussed in Section 3.11. We do this so as to follow the notational policies of the *Handbook*.

(being instances of Contraction and Permutation, respectively). But scheme (21) must still be taken as an axiom.

Before going on to discuss quantification, we briefly mention a couple of other systems of interest in the literature.

Given that **E** has a theory of necessity riding piggyback on it in the definition  $\Box A =_{\text{df}} (A \rightarrow A) \rightarrow A$ , the idea occurred to Meyer of adding to **R** a primitive symbol for necessity  $\Box$  governed by the **S4** axioms.

$$\begin{aligned} \Box A &\rightarrow A && (\Box 1) \\ \Box(A \rightarrow B) &\rightarrow (\Box A \rightarrow \Box B) && (\Box 2) \\ \Box A \wedge \Box B &\rightarrow \Box(A \wedge B) && (\Box 3) \\ \Box A &\rightarrow \Box \Box A, && (\Box 4) \end{aligned}$$

and the rule of Necessitation ( $\vdash A \Rightarrow \vdash \Box A$ ).

His thought was that **E** could be exactly translated into this system **R**<sup>□</sup> with entailment defined as strict implication. That this is subtly not the case was shown by Maksimova [1973] and Meyer [1979b] has shown how to modify **R**<sup>□</sup> so as to allow for an exact translation.

Yet one more system of interest is **RM** (cf. Section 3.10) obtained by adding to **R** the axiom scheme

$$A \rightarrow (A \rightarrow A) \quad \text{Mingle.} \quad (22)$$

Meyer has shown somewhat surprisingly that the pure implicational system obtained by adding Mingle to **R** is not the implicational fragment of **RM**, and he and Parks have shown how to axiomatise this fragment using a quite unintelligible formula (cf. [Anderson and Belnap, 1975, Section 8.18]). Mingle may be replaced equivalently with the converse of Contraction:

$$(A \rightarrow B) \rightarrow (A \rightarrow (A \rightarrow B)) \quad \text{Expansion.} \quad (23)$$

Of course one can consider ‘mingled’ versions of **E**, and indeed it was in this context that McCall first introduced mingle, albeit in the strict form (remember  $\vec{A} = A_1 \rightarrow A_2$ ),

$$\vec{A} \rightarrow (\vec{A} \rightarrow \vec{A}) \quad \vec{\text{Mingle}} \quad (24)$$

(cf. [Dunn, 1976c]).

We finish our discussion of axiomatics with a brief discussion of first-order relevance logics, which we shall denote by **RQ**, **EQ**, etc. We shall presuppose a standard definition of first-order formula (with connectives  $\neg, \wedge, \vee, \rightarrow$  and quantifiers  $\forall, \exists$ ). For convenience we shall suppose that we have two denumerable stocks of variables: the bound variables  $x, y$ , etc.

and the free variables (sometimes called parameters)  $a, b$ , etc. The bound variables are never allowed to have unbound occurrences.

The quantifier laws were set down by Anderson and Belnap in accord with the analogy of the universal quantifier with a conjunction (or its instances), and the existential quantifier as a disjunction. In view of the validity of quantifier interchange principles, we shall for brevity take only the universal quantifier  $\forall$  as primitive, defining  $\exists xA =_{\text{df}} \neg\forall x\neg A$ . We thus need

$$\forall xA \rightarrow A(a/x) \quad \forall\text{-elimination} \quad (25)$$

$$\forall x(A \rightarrow B) \rightarrow (A \rightarrow \forall xB) \quad \forall\text{-introduction} \quad (26)$$

$$\forall x(A \vee B) \rightarrow A \vee \forall xB \quad \text{Confinement.} \quad (27)$$

If there are function letters or other term forming operators, then (25) should be generalised to  $\forall xA \rightarrow A(t/x)$ , where  $t$  is any term (subject to our conventions that the ‘bound variables’  $x, y$ , etc. do not occur (‘free’) in it). Note well that because of our convention that ‘bound variables’ do not occur free, the usual proviso that  $x$  does not occur free in  $A$  in (26) and (27) is automatically satisfied. (27) is the obvious ‘infinite’ analogy of Distribution, and as such it causes as many technical problems for **RQ** as does Distribution for **R** (cf. Section 4.8). Finally, as an additional rule corresponding to adjunction, we need:

$$\frac{A(a/x)}{\forall xA} \quad \text{Generalisation.} \quad (28)$$

There are various more or less standard ways of varying this formulation. Thus, e.g. (cf. Meyer, Dunn and Leblanc [1974]) one can take all universal generalisations of axioms, thus avoiding the need for the rule of Generalisation. Also (26) can be ‘split’ into two parts:

$$\forall x(A \rightarrow B) \rightarrow (\forall xA \rightarrow \forall xB) \quad (26a)$$

$$A \rightarrow \forall xA \quad \text{Vacuous Quantification} \quad (26b)$$

(again note that if we allowed  $x$  to occur free we would have to require that  $x$  not be free in  $A$ ).

The most economical formulation is due to Meyer [1970]. It uses only the axiom scheme of  $\forall$ -elimination and the rule.

$$\frac{A \rightarrow B \vee C(a/x)}{A \rightarrow B \vee \forall xC} \quad (a \text{ cannot occur in } A \text{ or } B) \quad (29)$$

which combines (26)–(28).

#### 1.4 Deduction Theorems in Relevance Logic

Let **X** be a formal system, with certain formulas of **X** picked out as *axioms* and certain (finitary) relations among the formulas of **X** picked out as *rules*.

(For the sake of concreteness,  $\mathbf{X}$  can be thought of as any of the Hilbert-style systems of the previous section.) Where  $\Gamma$  is a list of formulas of  $\mathbf{X}$  (thought of as *hypotheses*) it is customary to define a *deduction from*  $\Gamma$  to be a sequence  $B_1, \dots, B_n$ , where for each  $B_i$  ( $1 \leq i \leq n$ ), either (1)  $B_i$  is in  $\Gamma$ , or (2)  $B_i$  is an axiom of  $\mathbf{X}$ , or (3)  $B_i$  ‘follows from’ earlier members of the sequence, i.e.  $R(B_{j_1}, \dots, B_{j_k}, B_i)$  holds for some  $(k+1)$ —any rule  $R$  of  $\mathbf{X}$  and  $B_{j_1}, \dots, B_{j_k}$  all precede  $B_i$  in the sequence  $B_1, \dots, B_n$ . A formula  $A$  is then said to be *deducible* from  $\Gamma$  just in case there is some deduction from  $\Gamma$  terminating in  $A$ . We symbolise this as  $\Gamma \vdash_{\mathbf{X}} A$  (often suppressing the subscript).

A *proof* is of course a deduction from the empty set, and a theorem is just the last item in a proof. There is the well-known

DEDUCTION THEOREM (Herbrand). *If  $A_1, \dots, A_n, A \vdash_{\mathbf{H}_{\rightarrow}} B$ , then we have also  $A_1, \dots, A_n \vdash_{\mathbf{H}_{\rightarrow}} A \rightarrow B$ .*

This theorem is proven in standard textbooks for classical logic, but the standard inductive proof shows that in fact the Deduction Theorem holds for any formal system  $\mathbf{X}$  having *modus ponens* as its sole rule and  $\mathbf{H}_{\rightarrow} \subseteq \mathbf{X}$  (i.e. each instance of an axiom scheme of  $\mathbf{H}_{\rightarrow}$  is a theorem of  $\mathbf{X}$ ). Indeed  $\mathbf{H}_{\rightarrow}$  can be motivated as the minimal pure implicational calculus having *modus ponens* as its sole rule and satisfying the Deduction Theorem. This is because the axioms of  $\mathbf{H}_{\rightarrow}$  can all be derived as theorems in any formal system  $\mathbf{X}$  using merely *modus ponens* and the supposition that  $\mathbf{X}$  satisfies the Deduction Theorem. Thus consider as an example:

- |  |                         |
|--|-------------------------|
| (1) $A, B \vdash A$                          | Definition of $\vdash$  |
| (2) $A \vdash B \rightarrow A$               | (1), Deduction Theorem  |
| (3) $\vdash A \rightarrow (B \rightarrow A)$ | (2), Deduction Theorem. |

Thus the most problematic axiom of  $\mathbf{H}_{\rightarrow}$  has a simple ‘*a priori* deduction’, indeed one using only the Deduction Theorem, not even *modus ponens* (which is though needed for more sane axioms like Self-Distribution).

It might be thought that the above considerations provide a very powerful argument for motivating intuitionistic logic (or at least some logic having the same implicational fragment) as The One True Logic. For what else should an implication do but satisfy *modus ponens* and the Deduction Theorem?

But it turns out that there is another sensible notion of deduction. This is what is sometimes called a *relevant deduction*. (Anderson and Belnap [1975, Section 22.2.1] claim that this is the *only* sensible notion of deduction, but we need not follow them in that). If there is anything that sticks out in the *a priori* deduction of Positive Paradox above it is that in (1),  $B$  was not *used* in the deduction of  $A$ .

A number of researchers have been independently bothered by this point and have been motivated to study a relevant implication that goes hand in

hand with a notion of relevant deduction. This, in this manner Moh [1950] and Church [1951b] came up with what is in effect  $\mathbf{R}_{\rightarrow}$ . And Anderson and Belnap [1975, p. 261] say “In fact, the search for a suitable deduction theorem for Ackermann’s systems . . . provided the impetus leading us to the research reported in this book.” This research program begun in the late 1950s took its starting point in the system(s) of Ackermann [1956], and the bold stroke separating the Anderson–Belnap system  $\mathbf{E}$  from Ackermann’s system  $\Pi'$  was basically the dropping of Ackermann’s rule  $\gamma$  so as to have an appropriate deduction theorem (cf. Section 2.1).

Let us accordingly define a deduction of  $B$  from  $A_1, \dots, A_n$  to be *relevant with respect to a given hypothesis  $A_i$*  just in case  $A_i$  is actually *used* in the given deduction of  $B$  in the sense (paraphrasing [Church, 1951b]) that there is a chain of inferences connecting  $A_i$  with the final formula  $B$ . This last can be made formally precise in any number of ways, but perhaps the most convenient is to flag  $A_i$  with say a  $\sharp$  and to pass the flag along in the deduction each time *modus ponens* is applied to two items at least one of which is flagged. It is then simply required that the last step of the deduction ( $B$ ) be flagged. Such devices are familiar from various textbook presentations of classical predicate calculus when one wants to keep track whether some hypothesis  $A_i(x)$  was used in the deduction of some formula  $B(x)$  to which one wants to apply Universal Generalisation.

We shall define a deduction of  $B$  from  $A_1, \dots, A_n$  to be *relevant simpliciter* just in case it is relevant with respect to each hypothesis  $A_i$ . A practical way to test for this is to flag each  $A_i$  with a different flag (say the subscript  $i$ ) and then demand that all of the flags show up on the last step  $B$ .

We can now state a version of the

**RELEVANT DEDUCTION THEOREM** (Moh, Church). *If there is a deduction in  $\mathbf{R}_{\rightarrow}$  of  $B$  from  $A_1, \dots, A_n, A$  that is relevant with respect to  $A$ , then there is a deduction in  $\mathbf{R}_{\rightarrow}$  of  $A \rightarrow B$  from  $A_1, \dots, A_n$ . Furthermore the new deduction will be ‘as relevant’ as the old one, i.e. any  $A_i$  that was used in the given deduction will be used in the new deduction.*

**Proof.** Let the given deduction be  $B_1, \dots, B_k$ , and let it be given with a particular analysis as to how each step is justified. By induction we show for each  $B_i$  that if  $A$  was used in obtaining  $B_i$  ( $B_i$  is flagged), then there is a deduction of  $A \rightarrow B_i$  from  $A_1, \dots, A_n$ , and otherwise there is a deduction of  $B_i$  from those same hypotheses. The tedious business of checking that the new deduction is as relevant as the old one is left to the reader. We divide up cases depending on how the step  $B_i$  is justified.

*Case 1.*  $B_i$  was justified as a hypothesis. Then neither  $B_i$  is  $A$  or it is some  $A_j$ . But  $A \rightarrow A$  is an axiom of  $\mathbf{R}_{\rightarrow}$  (and hence deducible from  $A_1, \dots, A_n$ ), which takes care of the first alternative. And clearly on the second alternative  $B_i$  is deducible from  $A_1, \dots, A_n$  (being one of them).



*Case 2.*  $B_i$  was justified as an axiom. Then  $A$  was not used in obtaining  $B_i$ , and of course  $B_i$  is deducible (being an axiom).

*Case 3.*  $B_i$  was justified as coming from preceding steps  $B_j \rightarrow B_i$  and  $B_j$  by *modus ponens*. There are four subcases depending on whether  $A$  was used in obtaining the premises.

*Subcase 3.1.*  $A$  was used in obtaining both  $B_j \rightarrow B_i$  and  $B_j$ . Then by inductive hypothesis  $A_1, \dots, A_n \vdash_{\mathbf{R}\rightarrow} A \rightarrow (B_j \rightarrow B_i)$  and  $A_1, \dots, A_n \vdash_{\mathbf{R}\rightarrow} A \rightarrow B_j$ . So  $A \rightarrow B$  may be obtained using the axiom of Self-Distribution.

*Subcase 3.2.*  $A$  was used in obtaining  $B_j \rightarrow B_i$  but not  $B_j$ . Use the axiom of Permutation to obtain  $A \rightarrow B_i$  from  $A \rightarrow (B_j \rightarrow B_i)$  and  $B_j$ .

*Subcase 3.3.*  $A$  was not used in obtaining  $B_j \rightarrow B_i$  but was used for  $B_j$ . Use the axiom of Prefixing to obtain  $A \rightarrow B_i$  from  $B_j \rightarrow B_i$  and  $A \rightarrow B_j$ .

*Subcase 3.4.*  $A$  was not used in obtaining either  $B_j \rightarrow B_i$  nor  $B_j$ . Then  $B_i$  follows from these using just *modus ponens*.

Incidentally,  $\mathbf{R}\rightarrow$  can easily be verified to be the minimal pure implicational calculus having *modus ponens* as sole rule and satisfying the Relevant Deduction Theorem, since each of the axioms invoked in the proof of this theorem can be easily seen to be theorems in any such system (cf. the next section for an illustration of sorts).

There thus seem to be at least two natural competing pure implicational logics  $\mathbf{R}\rightarrow$  and  $\mathbf{H}\rightarrow$ , differing only in whether one wants one's deductions to be relevant or not.<sup>5</sup> ■

Where does the Anderson–Belnap's [1975] preferred system  $\mathbf{E}\rightarrow$  fit into all of this? The key is that the implication of  $\mathbf{E}\rightarrow$  is both a strict and a relevant implication (cf. Section 1.3 for some subtleties related to this claim). As such, and since Anderson and Belnap have seen fit to give it the modal structure of the Lewis system  $\mathbf{S4}$ , it is appropriate to recall the appropriate deduction theorem for  $\mathbf{S4}$ .

MODAL DEDUCTION THEOREM [Barcan Marcus, 1946] *If  $A_1 \rightarrow B_1, \dots, A_n \rightarrow B_n, A \vdash_{\mathbf{S4}} B$  ( $\rightarrow$  here denotes strict implication), then  $A_1 \rightarrow B_1, \dots, A_n \rightarrow B_n \vdash_{\mathbf{S4}} A \rightarrow B$ .*

The idea here is that in general in order to derive the strict (*necessary*) implication  $A \rightarrow B$  one must not only be able to deduce  $B$  from  $A$  and some other hypotheses but furthermore those other hypotheses must be supposed to be necessary. And in  $\mathbf{S4}$  since  $A_i \rightarrow B_j$  is equivalent to  $\Box(A_i \rightarrow B_j)$ , requiring those additional hypotheses to be strict implications at least suffices for this.

Thus we could only hope that  $\mathbf{E}\rightarrow$  would satisfy the

---

<sup>5</sup>This seems to differ from the good-humoured polemical stand of Anderson and Belnap [1975, Section 22.2.1], which says that the first kind of 'deduction', which they call (pejoratively) 'Official deduction', is no kind of deduction at all.

MODAL RELEVANT DEDUCTION THEOREM [Anderson and Belnap, 1975]  
*If there is a deduction in  $\mathbf{E}_{\rightarrow}$  of  $B$  from  $A_1 \rightarrow B_1, \dots, A_n \rightarrow B_n, A$  that is relevant with respect to  $A$ , then there is a deduction in  $\mathbf{E}_{\rightarrow}$  of  $A \rightarrow B$  from  $A_1 \rightarrow B_1, \dots, A_n \rightarrow B_n$  that is as relevant as the original.*

The proof of this theorem is somewhat more complicated than its unmodalised counterpart which we just proved (cf. [Anderson and Belnap, 1975, Section 4.21] for a proof).

We now examine a subtle distinction (stressed by Meyer—see, for example, [Anderson and Belnap, 1975, pp. 394–395]), postponed until now for pedagogical reasons. We must ask, how many hypotheses can dance on the head of a formula? The question is: given the list of hypotheses  $A, A$ , do we have one hypothesis or two? When the notion of a *deduction* was first introduced in this section and a ‘list’ of hypotheses  $\Gamma$  was mentioned, the reader would naturally think that this was just informal language for a set. And of course the set  $\{A, A\}$  is identical to the set  $\{A\}$ . Clearly  $A$  is relevantly deducible from  $A$ . The question is whether it is so deducible from  $A, A$ . We have then two different criteria of use, depending on whether we interpret hypotheses as grouped together into lists that distinguish multiplicity of occurrences (sequences)<sup>6</sup> or sets. This issue has been taken up elsewhere of late, with other accounts of deduction appealing to ‘resource consciousness’ [Girard, 1987; Troelstra, 1992; Schroeder-Heister and Došen, 1993] as motivating some non-classical logics. Substructural logics in general appeal to the notion that the number of times a premise is used, or even more radically, the *order* in which premises are used, matter.

At issue in  $\mathbf{R}$  and its neighbours is whether  $A \rightarrow (A \rightarrow A)$  is a correct relevant implication (coming by two applications of ‘The Deduction Theorem’ from  $A, A \vdash A$ ). This is in fact not a theorem of  $\mathbf{R}$ , but it is the characteristic axiom of  $\mathbf{RM}$  (cf. Section 1.3). So it is important that in the Relevant Deduction Theorem proved for  $\mathbf{R}_{\rightarrow}$  that the hypotheses  $A_1, \dots, A_n$  be understood as a sequence in which the same formula may occur more than once. One can prove a version of the Relevant Deduction Theorem with hypotheses understood as collected into a set for the system  $\mathbf{RMO}_{\rightarrow}$ , obtained by adding  $A \rightarrow (A \rightarrow A)$  to  $\mathbf{R}_{\rightarrow}$  (but the reader should be told that Meyer has shown that  $\mathbf{RMO}_{\rightarrow}$ , is *not* the implicational fragment of  $\mathbf{RM}$ , cf. [Anderson and Belnap, 1975, Section 8.15]).<sup>7</sup>

<sup>6</sup>Sequences are not quite the best mathematical structures to represent this grouping since it is clear that the order of hypotheses makes no difference (at least in the case of  $\mathbf{R}$ ). Meyer and McRobbie [1979] have investigated ‘firesets’ (finitely repeatable sets) as the most appropriate abstraction.

<sup>7</sup>Arnon Avron has defended this system,  $\mathbf{RMO}_{\rightarrow}$ , as a natural way to characterise relevant implication [Avron, 1986; Avron, 1990a; Avron, 1990b; Avron, 1990c; Avron, 1992]. In Avron’s system, conjunction and disjunction are *intensional* connectives, defined in terms of the implication and negation of  $\mathbf{RMO}_{\rightarrow}$ . As a result, they do

Another consideration pointing to the *naturalness* of  $\mathbf{R}_{\rightarrow}$  is its connection to the  $\lambda I$ -calculus. A formula is a theorem of  $\mathbf{R}_{\rightarrow}$  if and only if it is the type of a closed term of the  $\lambda I$ -calculus as defined by Church. A  $\lambda I$  term is a  $\lambda$  term in which every lambda abstraction binds at least one free variable. So,  $\lambda x.\lambda y.xy$  has type  $A \rightarrow ((A \rightarrow B) \rightarrow B)$ , and so, is a theorem of  $\mathbf{R}_{\rightarrow}$ , while  $\lambda x.\lambda y.x$ , has type  $A \rightarrow (B \rightarrow A)$ , which is an intuitionistic theorem, but not an  $\mathbf{R}_{\rightarrow}$  theorem. This is reflected in the  $\lambda$  term, in which the  $\lambda y$  does not bind a free variable.

We now briefly discuss what happens to deduction theorems when the pure implication systems  $\mathbf{R}_{\rightarrow}$  and  $\mathbf{E}_{\rightarrow}$  are extended to include other connectives, especially  $\wedge$ .  $\mathbf{R}$  will be the paradigm, its situation extending straight-forwardly to  $\mathbf{E}$ . The problem is that the full system  $\mathbf{R}$  seems not to be formulable with *modus ponens* as the sole rule; there is also need for adjunction ( $A, B \vdash A \wedge B$ ) (cf. Section 1.3).

Thus when we think of proving a version of the Relevant Deduction Theorem for the full system  $\mathbf{R}$ , it would seem that we are forced to think through once more the issue of when a hypothesis is used, this time with relation to adjunction. It might be thought that the thing to do would be to pass the flag  $\ddagger$  along over an application of adjunction so that  $A \wedge B$  ends up flagged if either of the premises  $A$  or  $B$  was flagged, in obvious analogy with the decision concerning *modus ponens*.

Unfortunately, that decision leads to disaster. For then the deduction  $A, B \vdash A \wedge B$  would be a relevant one (both  $A$  and  $B$  would be ‘used’), and two applications of ‘The Deduction Theorem’ would lead to the thesis  $A \rightarrow (B \rightarrow A \wedge B)$ , the undesirability of which has already been remarked.

A more appropriate decision is to count hypotheses as used in obtaining  $A \wedge B$  just when they were used to obtain both premises. This corresponds to the axiom of Conjunction Introduction  $(C \rightarrow A) \wedge (C \rightarrow B) \rightarrow (C \rightarrow A \wedge B)$ , which thus handles the case in the inductive proof of the deduction theorem when the adjunction is applied. This decision may seem *ad hoc* (perhaps ‘use’ simpliciter is not quite the right concept), but it is the only decision to be made unless one wants to say that the hypothesis  $A$  can (in the presence of the hypothesis  $B$ ) be ‘used’ to obtain  $A \wedge B$  and hence  $B$  (passing on the flag from  $A$  this way is something like laundering dirty money).

This is the decision that was made by Anderson and Belnap in the context of natural deduction systems (see next section), and it was applied by Kron [1973; 1976] in proving appropriate deduction theorems for  $\mathbf{R}$ ,  $\mathbf{E}$  (and  $\mathbf{T}$ ). It should be said that the appropriate Deduction Theorem requires simultaneous flagging of the hypothesis (distinct flags being applied to each formula occurrence, say using subscripts in the manner of the ‘practical suggestion’ after our definition of relevant deduction for  $\mathbf{R}_{\rightarrow}$ ), with the requirement that all of the subscripts are passed on to the conclusion. So the

---

not have all of the distributive lattice properties of traditional relevance logics.

Deduction Theorem applies only to *fully* relevant deductions, where every premise is used (note that no such restriction was placed on the Relevant Deduction Theorem for  $\mathbf{R}_{\rightarrow}$ ).

An alternative stated in Meyer and McRobbie [1979] would be to adjust the definition of deduction, modifying clause (2) so as to allow as a step in a deduction any theorem (not just axiom) of  $\mathbf{R}$ , and to restrict clause (3) so that the only rule allowed in moving to later steps is *modus ponens*.<sup>8</sup> This is in effect to restrict adjunction to theorems, and reminds one of similar restrictions in the context of deduction theorems of similarly restricting the rules of necessitation and universal generalisation. It has the virtue that the Relevant Deduction Theorem and its proof are the same as for  $\mathbf{R}_{\rightarrow}$ . (Incidentally, Meyer's and Kron's sense of deduction coincide when all of  $A_1, \dots, A_n$  are used in deducing  $B$ ; this is obvious in one direction, and less than obvious in the other.)

There are yet two other versions of the deduction theorem that merit discussion in the context of relevance logic (relevance logic, as Meyer often points out, allows for many distinctions).

First in Belnap [1960b] and Anderson and Belnap [1975], there is a theorem (stated for  $\mathbf{E}$ , but we will state it for our paradigm  $\mathbf{R}$ ) called The Entailment Theorem, which says that  $A_1, \dots, A_n$  'entails'  $B$  iff  $\vdash_{\mathbf{R}} (A_1 \wedge \dots \wedge A_n) \rightarrow B$ . A formula  $B$  is defined in effect to be *entailed* by hypothesis  $A_1, \dots, A_n$  just in case there is a deduction of  $B$  using their conjunction  $A_1 \wedge \dots \wedge A_n$ . Adjunction is allowed, but subject to the restriction that the conjunctive hypothesis was used in obtaining both premises. The Entailment Theorem is clearly implied by Kron's version of the Deduction Theorem.

The last deduction theorem for  $\mathbf{R}$  we wish to discuss is the

ENTHYMEMATIC DEDUCTION THEOREM (Meyer, Dunn and Leblanc [1974]).  
If  $A_1, \dots, A_n, A \vdash_{\mathbf{R}} B$ , then  $A_1, \dots, A_n \vdash_{\mathbf{R}} A \wedge t \rightarrow B$ .

Here ordinary deducibility is all that is at issue (no insistence on the hypotheses being used). It can either be proved by induction, or cranked out of one of the more relevant versions of the deduction theorem. Thus it falls out of the Entailment Theorem that

$$\vdash_{\mathbf{R}} X \wedge A \wedge T \rightarrow B,$$

where  $X$  is the conjunction of  $A_1, \dots, A_n$ , and  $T$  is the conjunction of all the axioms of  $\mathbf{R}$  used in the deduction of  $B$ . But since  $\vdash_{\mathbf{R}} t \rightarrow T$ , we have  $\vdash_{\mathbf{R}} X \wedge A \wedge t \rightarrow B$ .

---

<sup>8</sup>Of course this requires we give an independent characterisation of *proof* (and *theorem*), since we can no longer define a *proof* as a deduction from zero premisses. We thus define a proof as a sequence of formulas, each of which is either an axiom or follows from preceding items by either *modus ponens* or adjunction (!).

However, the following **R** theorem holds:

$$\vdash_{\mathbf{R}} (X \wedge A \wedge t \rightarrow B) \rightarrow (X \wedge t \rightarrow (A \wedge t \rightarrow B)).$$

So  $\vdash_{\mathbf{R}} X \wedge t \rightarrow (A \wedge t \rightarrow B)$ , which leads (using  $\vdash_{\mathbf{R}} t$ ) to  $X \vdash_{\mathbf{R}} A \wedge t \rightarrow B$ , which dissolving the conjunction gives the desired

$$A_1, \dots, A_n \vdash_{\mathbf{R}} A \wedge t \rightarrow B.$$

In view of the importance of the notion, let us symbolise  $A \wedge t \rightarrow B$  as  $A \rightarrow_t B$ . This functions as a kind of ‘enthymematic implication’ ( $A$  and some truth *really implies*  $B$ ) and there will be more about Anderson, Belnap and Meyer’s investigations of this concept in Section 1.7. Let us simply note now that in the context of deduction theorems, it functions like intuitionistic implication, and allows us in  $\mathbf{R}_{\rightarrow}$  to have two different kinds of implication, each well motivated in its relation to the two different kinds of deducibility (ordinary and relevant).<sup>9</sup> For a more extensive discussion of deduction theorems in relevance logics and related systems, more recent papers by Avron [1991] and Brady [1994] should be consulted.

### 1.5 Natural Deduction Formulations

We shall be very brief about these since natural deduction methods are amply discussed by Anderson and Belnap [1975], where such methods in fact are used as a major motivation for relevance logic. Here we shall concentrate on a natural deduction system  $N\mathbf{R}$  for **R**.

The main idea of natural deduction (cf. Chapters [[were I.1 and I.2]] of the *Handbook*) of course is to allow the making of temporary hypotheses, with some device usually being provided to facilitate the book-keeping concerning the use of hypotheses (and when their use is ‘discharged’). Several textbooks (for example, [Suppes, 1957] and [Lemmon, 1965])<sup>10</sup> have used the device of in effect subscripting each hypothesis made with a distinct numeral, and then passing this numeral along with each application of a rule, thus keeping track of which hypothesis are used. When a hypothesis is discharged, the subscript is dropped. A line obtained with no subscripts is a ‘theorem’ since it depends on no hypotheses.

Let us then let  $\alpha, \beta$ , etc. range over classes of numerals. The rules for  $\rightarrow$  are then naturally:

$$\frac{\frac{A \rightarrow B_{\alpha}}{A_{\beta}} \quad [\rightarrow E]}{B_{\alpha \cup \beta}} \quad \frac{\frac{A_{\{k\}}}{\vdots} \quad B_{\alpha}}{A \rightarrow B_{\alpha - \{k\}}} \quad [\rightarrow I]}{A \rightarrow B_{\alpha - \{k\}}} \quad (\text{provided } k \in \alpha)$$

<sup>9</sup>In **E** enthymematic implication is like **S4** strict implication. See [Meyer, 1970a].

<sup>10</sup>The idea actually originates with [Feys and Ladrière, 1955].

Two fussy, really incidental remarks must be made. First, in the rule  $\rightarrow E$  it is to be understood that the premises need not occur in the order listed, nor need they be adjacent to each other or to the conclusion. Otherwise we would need a rule of ‘Repetition’, which allows the repeating of a formula with its subscripts as a later line. (Repetition is trivially derivable given our ‘non-adjacent’ understanding of  $\rightarrow E$ —in order to repeat  $A_\alpha$ , just prove  $A \rightarrow A$  and apply  $\rightarrow E$ .) Second, it is understood that we have what one might call a rule of ‘Hypothesis Introduction’: anytime one likes one can write a formula as a line with a new subscript (perhaps most conveniently, the line number).

Now a non-fussy remark must be made, which is really the heart of the whole matter. In the rule for  $\rightarrow I$ , a *proviso* has been attached which has the effect of requiring that the hypothesis  $A$  was actually used in obtaining  $B$ . This is precisely what makes the implication relevant (one gets the intuitionistic implication system  $\mathbf{H}_\rightarrow$  if one drops this requirement). The reader should find it instructive to attempt a proof of Positive Paradox ( $A \rightarrow (B \rightarrow A)$ ) and see how it breaks down for  $\mathbf{NR}_\rightarrow$  (but succeeds in  $\mathbf{NH}_\rightarrow$ ). The reader should also construct proofs in  $\mathbf{NR}_\rightarrow$  of all the axioms in one of the Hilbert-style formulations of  $\mathbf{R}_\rightarrow$  from Section 1.3.

Then the equivalence of  $\mathbf{R}_\rightarrow$  in its Hilbert-style and natural deduction formulations is more or less self-evident given the Relevant Deduction Theorem (which shows that the rule  $\rightarrow I$  can be ‘simulated’ in the Hilbert-style system, the only point at issue).

Indeed it is interesting to note that Lemmon [1965], who *seems* to have the same *proviso* on  $\rightarrow I$  that we have for  $\mathbf{NR}_\rightarrow$  (his actual language is a bit informal), does not prove Positive Paradox until his *second* chapter adding conjunction (and disjunction) to the implication-negation system he developed in his first chapter. His proof of Positive Paradox depends finally upon an ‘irrelevant’  $\wedge I$  rule. The following is perhaps the most straightforward proof in his system (differing from the proof he actually gives):

(1)	$A_1$	Hyp
(2)	$B_2$	Hyp
(3)	$A \wedge B_{1,2}$	1, 2, $\wedge I$ ?
(4)	$A_{1,2}$	3, $\wedge E$
(5)	$B \rightarrow A_1$	2, 4, $\rightarrow I$
(6)	$A \rightarrow (B \rightarrow A)$	1, 5, $\rightarrow I$ .

We think that the manoeuvre used in getting  $B$ ’s 2 to show up attached to  $A$  in line (4) should be compared to laundering dirty money by running it through an apparently legitimate business. The correct ‘relevant’ versions

of the conjunction rules are instead

$$\frac{A_\alpha}{A \wedge B_\alpha} \quad [\wedge I] \quad \frac{A \wedge B_\alpha}{A_\alpha} \quad \frac{A \wedge B_\alpha}{B_\alpha} \quad [\wedge E]$$

What about disjunction? In **R** (also **E**, etc.) one has de Morgan's Laws and Double Negation, so one can simply define  $A \vee B = \neg(\neg A \wedge \neg B)$ . One might think that settling down in separate int-elim rules for  $\vee$  would then only be a matter of convenience. Indeed, one can find in [Anderson and Belnap, 1975] in effect the following rules:

$$\frac{A_\alpha}{A \vee B_\alpha} \quad \frac{B_\alpha}{A \vee B_\alpha} \quad [\vee I] \quad \begin{array}{c} A \vee B_\alpha \\ \vdots \\ A_k \\ \vdots \\ C_{\beta \cup \{k\}} \\ B_h \\ \vdots \\ C_{\beta \cup \{h\}} \\ \hline C_{\alpha \cup \beta} \end{array} \quad [\vee E]$$

But (as Anderson and Belnap point out) these rules are insufficient. From them one cannot derive the following

$$\frac{A \wedge (B \vee C)_\alpha}{(A \wedge B) \vee C_\alpha} \text{ Distribution.}$$

And so it must be taken as an additional rule (even if disjunction is defined from conjunction and negation).

This is clearly an unsatisfying, if not unsatisfactory, state of affairs. The customary motivation behind int-elim rules is that they show how a connective may be introduced into and eliminated from argumentative discourse (in which it has no essential occurrence), and thereby give the connective's role or meaning. In this context the Distribution rule looks very much to be regretted.

One remedy is to modify the natural deduction system by allowing hypotheses to be introduced in two different ways, 'relevantly' and 'irrelevantly'. The first way is already familiar to us and is what requires a subscript to keep track of the relevance of the hypothesis. It requires that the hypotheses introduced this way will *all* be used to get the conclusion. The second way involves only the weaker promise that at least *some* of the hypotheses so introduced will be used. This suggestion can be formalised by

allowing several hypotheses to be listed on a line, but with a single relevance numeral attached to them as a bunch. Thus, schematically, an argument of the form

$$\begin{array}{l} (1) \quad A, B_1 \\ (2) \quad C, D_2 \\ \quad \quad \vdots \\ (k) \quad E_{1,2} \end{array}$$

should be interpreted as establishing

$$A \wedge B \rightarrow (C \wedge D \rightarrow E).$$

Now the natural deduction rules must be stated in a more general form allowing for the fact that more than one formula can occur on a line. Key among these would be the new rule:

$$\frac{\begin{array}{l} \Gamma, A \vee B_\alpha \\ \quad \quad \vdots \\ \quad \quad \Gamma, A_k \\ \quad \quad \quad \vdots \\ \quad \quad \Delta_{\beta \cup \{k\}} \\ \quad \quad \Gamma, B_l \\ \quad \quad \quad \vdots \\ \quad \quad \Delta_{\beta \cup \{l\}} \end{array} \quad [VE']}{\Delta_{\alpha \cup \beta}}$$

It is fairly obvious that this rule has Distribution built into it. Of course, other rules must be suitably modified. It is easiest to interpret the formulas on a line as grouped into a set so as not to have to worry about ‘structural rules’ corresponding to the commutation and idempotence of conjunction.

The rules  $\rightarrow I$ ,  $\rightarrow E$ ,  $\vee I$ ,  $\vee E$ ,  $\wedge I$ , and  $\wedge E$  can all be left as they were (or except for  $\rightarrow I$  and  $\rightarrow E$ , trivially generalised so as to allow for the fact that the premises might be occurring on a line with several other ‘irrelevant’ premises), but we do need one new structural rule:

$$\frac{\Gamma_\alpha \quad \Delta_\alpha}{\Gamma, \Delta_\alpha} \quad [\text{Comma } I]$$

Once we have this it is natural to take the conjunction rules in ‘Ketonen form’:

$$\frac{\Gamma, A, B_\alpha}{\Gamma, A \wedge B_\alpha} \quad [\wedge I']$$



$$\frac{\Gamma, A \wedge B_\alpha}{\Gamma, A, B_\alpha} [\wedge E']$$

with the rule

$$\frac{\Gamma, \Delta_\alpha}{\Gamma_\alpha} [\text{Comma } E]$$

It is merely a tedious exercise for the reader to show that this new system  $N'\mathbf{R}$  is equivalent to  $N\mathbf{R}$ . Incidentally,  $N'\mathbf{R}$  was suggested by reflection upon the Gentzen System  $LR^+$  of Section 4.9.

Before leaving the question of natural deduction for  $\mathbf{R}$ , we would like to mention one or two technical aspects. First, the system of Prawitz [1965] differs from  $\mathbf{R}$  in that it lacks the rule of Distribution. This is perhaps compensated for by the fact that Prawitz can prove a normal form theorem for proofs in his system. A different system yet is that of [Pottinger, 1979], based on the idea that the correct  $\wedge I$  rule is

$$\frac{\begin{array}{c} A_\alpha \\ B_\beta \end{array}}{A \wedge B_{\alpha \cup \beta}}$$

He too gets a normal form theorem. We conjecture that some appropriate normal form theorem is provable for the system  $N'\mathbf{R}^+$  on the well-known analogy between cut-elimination and normalisation and the fact that cut-elimination has been proven for  $LR^+$  (cf. Section 4.9). Negation though would seem to bring extra problems, as it does when one is trying to add it to  $LR^+$ .

One last set of remarks, and we close the discussion of natural deduction. The system  $N\mathbf{R}$  above differs from the natural deduction system for  $\mathbf{R}$  of Anderson and Belnap [1975]. Their system is a so-called ‘Fitch-style’ formalism, and so named  $F\mathbf{R}$ . The reader is presumed to know that in this formalism when a hypothesis is introduced it is thought of as starting a subproof, and a line is drawn along the left of the subproof (or a box is drawn around the subproof, or some such thing) to demarcate the scope of the hypothesis. If one is doing a natural deduction system for classical or intuitionistic logic, subproofs or dependency numerals can either one be used to do essentially the same job of keeping track the use of hypotheses (though dependency numerals keep more careful track, and that is why they are so useful for relevant implication).

Mathematically, a Fitch-style proof is a nested structure, representing the fact that subproofs can contain further subproofs, etc. But once one has dependency numerals, this extra structure, at least for  $\mathbf{R}$ , seems otiose, and so we have dispensed with it. The story for  $\mathbf{E}$  is more complex, since

on the Anderson and Belnap approach **E** differs from **R** only in what is allowed to be ‘reiterable’ into subproof. Since implication in **E** is necessary as well as relevant, the story is that in deducing  $B$  from  $A$  in order to show  $A \rightarrow B$ , one should only be allowed to use items that have been assumed to be necessarily true, and that these can be taken to be formulas of the form  $C \rightarrow D$ . So only formulas of this form can be reiterated for use in the subproof from  $A$  to  $B$ . Working out how best to articulate this idea using only dependency numerals (no lines, boxes, etc.) is a little messy. This concern to keep track of how premises are used in a proof by way of labels has been taken up in a general way by recent work on *Labelled Deductive Systems* [D’Agostino and Gabbay, 1994; Gabbay, 1997].

We would be remiss not to mention other formulations of natural deduction systems for relevance logics and their cousins. A different generalisation of Hunter’s natural deduction systems (which follows more closely the Gentzen systems for positive logics — see Section 4.9) is in [Read, 1988; Slaney, 1990].<sup>11</sup>

### 1.6 Basic Formal Properties of Relevance Logic

This section contains a few relatively simple properties of relevance logics, proofs for which can be found in [Anderson and Belnap, 1975]. With one exception (the ‘Ackermann Properties’—see below), these properties all hold for both the system **R** and **E**, and indeed for most of the relevance logics defined in Section 1.3. For simplicity, we shall state these properties for sentential logics, but appropriate versions hold as well for their first-order counterparts.

First we examine the REPLACEMENT THEOREM *For both R and E*,

$$\vdash (A \leftrightarrow B) \wedge t \rightarrow (\chi(A) \leftrightarrow \chi(B)).$$

Here  $\chi(A)$  is any formula with perhaps some occurrences of  $A$  and  $\chi(B)$  is the result of perhaps replacing one or more of those occurrences by  $B$ . The proof is by a straightforward induction on the complexity of  $\chi(A)$ , and one clear role of the conjoined  $t$  is to imply  $\chi \rightarrow \chi$  when  $\chi(= \chi(A))$  contains no occurrences of  $A$ , or does but none of them is replaced by  $B$ . It might be thought that if these degenerate cases are ruled out by requiring that some actual occurrence of  $A$  be replaced by  $B$ , then the need for  $t$  would vanish. This is indeed true for the implication-negation (and of course the pure implication) fragments of **R** and **E**, but not for the whole systems in virtue of the non-theoremhood of what V. Routley has dubbed ‘Factor’:

<sup>11</sup>The reader should be informed that still other natural deduction formalisms for **R** of various virtues can be found in [Meyer, 1979b] and [Meyer and McRobbie, 1979].

1.  $(A \rightarrow B) \rightarrow (A \wedge \chi \rightarrow B \wedge \chi)$ .

Here the closest one can come is to

2.  $(A \rightarrow B) \wedge t \rightarrow (A \wedge \chi \rightarrow B \wedge \chi)$ ,

the conjoined  $g$  giving the force of having  $\chi \rightarrow \chi$  in the antecedent, and the theorem  $(A \rightarrow B) \wedge (\chi \rightarrow \chi) \rightarrow (A \wedge \chi \rightarrow B \wedge \chi)$  getting us home. (2) of course is just a special case of the Replacement Theorem. Of more ‘relevant’ interest is the

**VARIABLE SHARING PROPERTY.** If  $A \rightarrow B$  is a theorem of **R** (or **E**), then there exists some sentential variable  $p$  that occurs in both  $A$  and  $B$ . This is understood by Anderson and Belnap as requiring some commonality of meaning between antecedent and consequent of logically true relevant implications. The proof uses an ingenious logical matrix, having eight values, for which see [Anderson and Belnap, 1975, Section 22.1.3]. There are discussed both the original proof of Belnap and an independent proof of Dončenko, and strengthening by Maksimova. Of modal interest is the

**ACKERMANN PROPERTY.** No formula of the form  $A \rightarrow (B \rightarrow C)$  ( $A$  containing no  $\rightarrow$ ) is a theorem of **E**. The proof again uses an ingenious matrix (due to Ackermann) and has been strengthened by Maksimova (see [Anderson and Belnap, 1975, Section 22.1.1 and Section 22.1.2]) (contributed by J. A. Coffa) on ‘fallacies of modality’.

### 1.7 First-degree Entailments

A *zero degree formula* contains only the connectives  $\wedge, \vee$ , and  $\neg$ , and can be regarded as either a formula of relevance logic or of classical logic, as one pleases. A *first degree implication* is a formula of the form  $A \rightarrow B$ , where both  $A$  and  $B$  are zero-degree formulas: Thus first degree implications can be regarded as either a restricted fragment of some relevance logic (say **R** or **E**) or else as expressing some metalinguistic logical relation between two classical formulas  $A$  and  $B$ . This last is worth mention, since then even a classical logician of Quinean tendencies (who remains unconverted by the considerations of Section 1.2 in favour of nested implications) can still take first degree logical relevant implications to be legitimate.

A natural question is what is the relationship between the provable first-degree implications of **R** and those of **E**. It is well-known that the corresponding relationship between classical logic and some normal modal logic, say **S4** (with the  $\rightarrow$  being the material conditional and strict implication, respectively), is that they are identical in their first degree fragments. The same holds of **R** and **E** (cf. [Anderson and Belnap, 1975, Section 2.42]).

This fragment, which we shall call **R<sub>fd<sub>e</sub></sub>** (Anderson and Belnap [1975] call it **E<sub>fd<sub>e</sub></sub>**) is stable (cf. [Anderson and Belnap, 1975, Section 7.1]) in the sense

that it can be described from a variety of perspectives. For some semantical perspectives see Sections 3.3 and 3.4. We now consider some syntactical perspectives of more than mere ‘orthographic’ significance.

The perhaps least interesting of these perspectives is a ‘Hilbert-style’ presentation of  $\mathbf{R}_{\mathbf{fde}}$  (cf. [Anderson and Belnap, 1975, Section 15.2]). It has the following axioms:

- |  |                          |
|--|--------------------------|
| 3. $A \wedge B \rightarrow A, A \wedge B \rightarrow B$  | Conjunction Elimination  |
| 4. $A \rightarrow A \vee B, B \rightarrow A \vee B$      | Disjunction Introduction |
| 5. $A \wedge (B \vee C) \rightarrow (A \wedge B) \vee C$ | Distribution             |
| 6. $A \rightarrow \neg\neg A, \neg\neg A \rightarrow A$  | Double Negation          |

It also has gobs of rules:

- |   |                          |
|---|--------------------------|
| 7. $A \rightarrow B, B \rightarrow C \vdash A \rightarrow C$          | Transitivity             |
| 8. $A \rightarrow B, A \rightarrow C \vdash A \rightarrow B \wedge C$ | Conjunction Introduction |
| 9. $A \rightarrow C, B \rightarrow C \vdash A \vee B \rightarrow C$   | Disjunction Introduction |
| 10. $A \rightarrow B \vdash \neg B \rightarrow \neg A$                | Contraposition.          |

More interesting is the characterisation of Anderson and Belnap [1962b; 1975] of  $\mathbf{R}_{\mathbf{fde}}$  as ‘tautological entailments’. The root idea is to consider first the ‘primitive entailments’.

$$11. A_1 \wedge \dots \wedge A_m \rightarrow B_1 \vee \dots \vee B_n,$$

where each  $A_i$  and  $B_j$  is either a sentential variable or its negate (an ‘atom’) and make it a necessary and sufficient criterion for such a primitive entailment to hold that some  $A_i$  actually be identically the same formula as some  $B_j$  (that the entailment be ‘tautological’ in the sense that  $A_i$  is repeated). This rules out both

$$12. p \wedge \neg p \rightarrow q,$$

$$13. p \rightarrow q \vee \neg q,$$

where there is no variable sharing, but also such things as

$$14. p \wedge \neg p \wedge q \rightarrow \neg q,$$

where there is (of course all of (12)–(14) are valid classically, where a primitive entailment may hold because of atom sharing or because either the antecedent is contradictory or else the consequent is a logical truth).

Now the question remains as to which non-primitive entailments to count as valid. Both relevance logic and classical logic agree on the standard

count as valid. Both relevance logic and classical logic agree on the standard ‘normal form equivalences’: commutation, association, idempotence, distribution, double negation, and de Morgan’s laws. So the idea is, given a candidate entailment  $A \rightarrow B$ , by way of these equivalences,  $A$  can be put into disjunctive normal form and  $B$  may be put into conjunctive normal form, reducing the problem to the question of whether the following is a valid entailment:

$$15. A_1 \vee \cdots \vee A_k \rightarrow B_1 \wedge \cdots \wedge B_h.$$

But simple considerations (on which both classical and relevance logic agree) having to do with conjunction and disjunction introduction and elimination show that (15) holds if for each disjunct  $A_i$  and conjunct  $B_j$ , the primitive entailment  $A_i \rightarrow B_j$  is valid. For relevance logic this means that there must be atom sharing between the conjunction  $A_i$  and the disjunction  $B_j$ .

This criterion obviously counts the Disjunctive Syllogism

$$16. \neg p \wedge (p \vee q) \rightarrow q,$$

as an invalid entailment, for using distribution to put its antecedent into disjunctive normal form, (16) is reduced to

$$16' (\neg p \wedge p) \vee (\neg p \wedge q) \rightarrow q.$$

But by the criterion of tautological entailments,

$$17. \neg p \wedge p \rightarrow q,$$

which is required for the validity of (16'), is rejected.

Another pleasant characterisation of  $\mathbf{R}_{\mathbf{fde}}$  is contained in [Dunn, 1976a] using a simplification of Jeffrey’s ‘coupled trees’ method for testing classically valid entailments. The idea is that to test  $A \rightarrow B$  one works out a truth-tree for  $A$  and a truth tree for  $B$ . One then requires that every branch in the tree for  $A$  ‘covers’ some branch in the tree for  $B$  in the sense that every atom in the covered branch occurs in the covering branch. This has the intuitive sense that every way in which  $A$  might be true is also a way in which  $B$  would be true, whether these ways are logically possible or not, since ‘closed’ branches (those containing contradictions) are not exempt as they are in Jeffrey’s method for classical logic. This coupled-trees approach is ultimately related to the Anderson–Belnap tautological entailment method, as is also the method of [Dunn, 1980b] which explicates an earlier attempt of Levy to characterise entailment (cf. also [Clark, 1980]).

### 1.8 Relations to Familiar Logics

There is a sense in which relevance logic contains classical logic.

**ZDF THEOREM** (Anderson and Belnap [1959a]). *The zero-degree formulas (those containing only the connectives  $\wedge, \vee, \neg$ ) provable in **R** (or **E**) are precisely the theorems of classical logic.*

The proof went by considering a ‘cut-free’ formulation of classical logic whose axioms are essentially just excluded middles (which are theorems of **R** / **E**) and whose rules are all provable first-degree relevant entailments (cf. Section 2.7). This result extends to a first-order version [Anderson and Belnap Jr., 1959b]. (The admissibility of  $\gamma$  (cf. Section 2) provides another route to the proof to the ZDF Theorem.)

There is however another sense in which relevance logic does not contain classical logic:

**FACT** (Anderson and Belnap [1975, Section 25.1]). **R** (and **E**) lack as a derivable rule Disjunctive Syllogism:

$$\neg A, A \vee B \vdash B.$$

This is to say there is no deduction (in the standard sense of Section 1.4) of  $B$  from  $\neg A$  and  $A \vee B$  as premises. This is of course the most notorious feature of relevance logic, and the whole of Section 2 is devoted to its discussion.

Looking now in another direction, Anderson and Belnap [1961] began the investigation of how to translate intuitionistic and strict implication into **R** and **E**, respectively, as ‘enthymematic’ implication. Anderson and Belnap’s work presupposed the addition of propositional quantifiers to, let us say **R**, with the subsequent definition of ‘ $A$  intuitionistically implies  $B$ ’ (in symbols  $A \supset B$ ) as  $\exists p(p \wedge (A \wedge p \rightarrow B))$ . This has the sense that  $A$  together with some truth relevantly implies  $B$ , and does seem to be at least in the neighbourhood of capturing Heyting’s idea that  $A \supset B$  should hold if there exists some ‘construction’ (the  $p$ ) which adjoined to  $A$  ‘yields’ (relevant implication)  $B$ . Meyer in a series of papers [1970a; 1973] has extended and simplified these ideas, using the propositional constant  $t$  in place of propositional quantification, defining  $A \supset B$  as  $A \wedge t \rightarrow B$ . If a propositional constant  $F$  for the intuitionistic absurdity is introduced, then intuitionistic negation can be defined in the style of Johansson as  $\neg A =_{\text{df}} A \supset F$ . As Meyer has discovered one must be careful what axiom one chooses to govern  $F$ .  $F \rightarrow A$  or even  $F \supset A$  is too strong. In intuitionistic logic, the absurd proposition *intuitionistically* implies only the *intuitionistic* formulas, so the correct axiom is  $F \supset A^*$ , where  $A^*$  is a translation into **R** of an intuitionistic formula. Similar translations carry **S4** into **E** and classical logic into **R**.

2 THE ADMISSIBILITY OF  $\gamma$ 2.1 Ackermann's Rule  $\gamma$ 

The first mentioned problem for relevance logics in Anderson's [1963] seminal 'open problems' paper is the question of 'the admissibility of  $\gamma$ '. To demystify things a bit it should be said that  $\gamma$  is simply *modus ponens* for the material conditions ( $\neg A \vee B$ ):

$$1. \frac{A \quad \neg A \vee B}{B}.$$

It was the third listed rule of Ackermann's [1956] system of *strenge Implikation* ( $\alpha, \beta, \gamma$ ; 1st, 2nd, 3rd). This was the system Anderson and Belnap 'tinkered with' to produce **E** (Ackermann also had a rule  $\delta$  which they replaced with an axiom).

The major part of Anderson and Belnap's 'tinkering' was the extremely bold step of simply deleting  $\gamma$  as a primitive rule, on the well-motivated ground that the corresponding object language formula

$$2. A \wedge (\neg A \vee B) \rightarrow B$$

is not a theorem of **E**.

It is easy to see that (2) could not be a theorem of either **E** or **R**, since it is easy to prove in those systems

$$3. A \wedge \neg A \rightarrow A \wedge (\neg A \vee B)$$

(largely because  $\neg A \rightarrow \neg A \vee B$  is an instance of an axiom), and of course (3) and (2) yield by transitivity the 'irrelevancy'

$$4. A \wedge \neg A \rightarrow B.$$

The inference (1) is obviously related to the Stoic principle of the *disjunctive syllogism*:

$$5. \frac{\neg A \quad A \vee B}{B}.$$

Indeed, given the law of double negation (and replacement) they are equivalent, and double negation is never at issue in the orthodox logics. Thus **E** and **R** reject

$$6. \neg A \wedge (A \vee B) \rightarrow B$$

as well as (2).

This rejection is typically the hardest thing to swallow concerning relevance logics. One starts off with some pleasant motivations about relevant implication and using subscripts to keep track of whether a hypothesis has actually been used (as in Section 1.5), and then one comes to the point where one says ‘and of course we have to give up the disjunctive syllogism’ and one loses one’s audience. Please do not stop reading! We shall try to make this rejection of *disjunctive syllogism* as palatable as we can.

(See [Belnap and Dunn, 1981; Restall, 1999] for related discussions, and also discussion of [Anderson and Belnap, 1975, Section 16.1]); see Burgess [1981] for an opposing point of view.

## 2.2 The Lewis ‘Proof’

One reason that *disjunctive syllogism* has figured so prominently in the controversy surrounding relevance logic is because of the use it was put to by C. I. Lewis [Lewis and Langford, 1932] in his so-called ‘independent proof’: that a contradiction entails any sentence whatsoever (taken by Anderson and Belnap as a clear breakdown of relevance). Lewis’s proof (with our notations of justification) goes as follows:

- |     |                   |                                   |
|-----|-------------------|-----------------------------------|
| (1) | $p \wedge \neg p$ |                                   |
| (2) | $p$               | 2, $\wedge$ -Elimination          |
| (3) | $\neg p$          | 1, $\wedge$ -Elimination          |
| (4) | $p \vee q$        | 2, $\vee$ -Introduction           |
| (5) | $q$               | 3, 4 <i>disjunctive syllogism</i> |

Indeed one can usefully classify alternative approaches to relevant implication according to how they reject the Lewis proof. Thus, e.g. Nelson rejects  $\wedge$ -Elimination and  $\vee$ -Introduction, as does McCall’s connexive logic. Parry, on the other hand, rejects only  $\vee$ -Introduction. Geach, and more recently, Tennant [1994], accept each step, but says that ‘entailment’ (relevant implication) is not transitive. It is the genius of the Anderson–Belnap approach to see disjunctive syllogism as the culprit and the sole culprit.<sup>12</sup>

Lewis concludes his proof by saying, “If by (3),  $p$  is false; and, by (4), at least one of the two,  $p$  and  $q$  is true, then  $q$  must be true”. As is told in [Dunn, 1976a], Dunn was saying such a thing to an elementary logic class one time (with no propaganda about relevance logic) when a student yelled out, “But  $p$  was the true one—look again at your assumption”.

<sup>12</sup>Although this point is complicated, especially in some of their earlier writings (see, e.g. [Anderson and Belnap Jr., 1962a]) by the claim that there is a kind of fallacy of ambiguity in the Lewis proof. the idea is that if  $\vee$  is read in the ‘intensional’ way (as  $\neg A \rightarrow B$ ), then the move from (3) and (4) to (5) is OK (it’s just *modus ponens* for the relevant conditional), but the move from (2) to (4) is not (now being a paradox of implication rather than ordinary disjunction introduction).



That student had a point. Disjunctive syllogism is not obviously appropriate to a situation of inconsistent information—where  $p$  is assumed (given, believed, etc.) to be both true and false. This point has been argued strenuously in, e.g. [Routley and Routley, 1972; Dunn, 1976a] and Belnap [1977b; 1977a]. The first two of these develop a semantical analysis that lets both  $p$  and  $\neg p$  receive the value ‘true’ (as is appropriate to model the situation where  $p \wedge \neg p$  has been assumed true), and there will be more about these ideas in Section 3.4. The last is particularly interesting since it extends the ideas of Dunn [1976a] so as to provide a model of how a computer might be programmed as to make inferences from its (possibly inconsistent) database. One would not want trivially inconsistent information about the colour of your car that somehow got fed into the FBI’s computer (perhaps by pooled databases) to lead to the conclusion that you are Public Enemy Number One.

We would like to add yet one more criticism of *disjunctive syllogism*, which is sympathetic to many of the earlier criticisms.

We need as background to this criticism the natural deduction framework of [Gentzen, 1934] as interpreted by [Prawitz, 1965] and others. the idea (as in Section 1.5) is that each connective should come with rules that introduce it into discourse (as principal connective of a conclusion) and rules that eliminate it from discourse (as principal connective of a premise). further the ‘normalisation ideas of Prawitz, though of great technical interest and complication, boil down philosophically to the observation that an elimination rule should not be able to get out of a connective more than an introduction rule can put into the connective. This is just the old conservation Principle, ‘You can’t get something for nothing’, applied to logic.

The paradigm here is the introduction and elimination rules for conjunction. The introduction rule, from  $A, B$  to infer  $A \wedge B$  packs into  $A \wedge B$  precisely what the elimination rule, from  $A \wedge B$  to infer either  $A$  or  $B$  (separately), then unpacks.

Now the standard introduction rule for disjunction is this: from either  $A$  or  $B$  separately, infer  $A \vee B$ . We have no quarrel with an introduction rule. an introduction rule gives meaning to a connective and the only thing to watch out for is that the elimination rule does not take more meaning from a connective than the introduction rule gives to it (of course, one can also worry about the usefulness and/or naturalness of the introduction rules for a given connective, but that (*pace* [Parry, 1933]) seems not an issue in the case of disjunction.

In the Lewis ‘proof’ above, it is then clear that the *disjunctive syllogism* is the only conceivably problematic rule of inference. Some logicians (as indicated above) have queried the inferences from (1) to (2) and (4), and from (2) to (3), but from the point of view that we are now urging, this is simply wrongheaded. Like Humpty Dumpty, we use words to mean what *we* say. So there is nothing wrong with introducing connectives  $\wedge$  and  $\vee$

via the standard introduction rules. Other people may want connectives for which *they* provide different introduction (and matching elimination) rules, but that is *their* business. We want the standard (‘extensional’) senses of  $\wedge$  and  $\vee$ .

Now the d.s. is a very odd rule when viewed as an elimination rule for  $\vee$  parasitical upon the standard introduction rules (whereas the constructive dilemma, the usual  $\vee$ -Elimination rule is not at all odd). Remember that the introduction rules provide the actual inferences that are to be stored in the connective’s battery as potential inferences, perhaps later to be released again as actual inferences by elimination rules. The problem with the *disjunctive syllogism* is that it can release inferences from  $\vee$  that it just does not contain. (In another context, [Belnap, 1962] observed that Gentzen-style rules for a given connective should be ‘conservative’, i.e. they should not create new inferences not involving the given connective.)

Thus the problem with the *disjunctive syllogism* is just that  $p \vee q$  might have been introduced into discourse (as it is in the Lewis ‘proof’) by  $\vee$ -Introduction from  $p$ . So then to go on to infer  $q$  from  $p \vee q$  and  $\neg p$  by the *disjunctive syllogism* would be legitimate only if the inference from  $p$ ,  $\neg p$  to  $q$  were legitimate. But this is precisely the point at issue. At the very least the Lewis argument is circular (and not independent).<sup>13</sup>

### 2.3 The Admissibility of $\gamma$

Certain rules of inference are sometimes ‘admissible’ in formal logics in the sense that whenever the premises are *theorems*, so is the conclusion a *theorem*, although these rules are nonetheless invalid in the sense that the premises may be true while the conclusion is not. Familiar examples are the rule of substitution in propositional logic, generalisation in predicate logic, and necessitation in modal logic. Using this last as paradigm, although the inference from  $A$  to  $\Box A$  (necessarily  $A$ ) is clearly invalid and would indeed vitiate the entire point of modal logic, still for the (‘normal’) modal logics, whenever  $A$  is a theorem so is  $\Box A$  (and indeed their motivation would be somehow askew if this did not hold).

Anderson [1963] speculated that something similar was afoot with respect to the rule  $\gamma$  and relevance logic. Anderson hoped for a ‘sort of lucky accident’, but the admissibility of  $\gamma$  seems more crucial to the motivation of **E** and **R** than that. Kripke [1965] gives a list of four conditions that a propositional calculus must meet in order to have a normal characteristic matrix, one of which is the admissibility of  $\gamma$ .<sup>14</sup> ‘Normal’ is meant in the

<sup>13</sup>This is a new argument on the side of Anderson and Belnap [1962b, pp. 19, 21].

<sup>14</sup>The other conditions are that it be consistent, that it contain all classical tautologies, and that it be ‘complete in the sense of Halldén’. **R** and **E** can be rather easily seen to have the first two properties (see Section 1.8 for the bit about classical tautologies), but the last is rather more difficult (see Section 3.11).

sense of Church, and boils down to being able to divide up its elements into the ‘true’ and the ‘false’ with the operations of conjunction, disjunction, and negation treating truth and falsity in the style of the truth tables (a conjunction is true if both components are true, etc.). If one thinks of **E** (as Anderson surely did) as the logic of propositions with the logical operations, and surely this should divide itself up into the true and the false propositions.<sup>15</sup>

#### 2.4 Proof(s) of the Admissibility of $\gamma$

There are by now at least four variant proofs of the admissibility of  $\gamma$  for **E** and **R**. The first three proofs (in chronological order: [Meyer and Dunn, 1969], [Routley and Meyer, 1973] and [Meyer, 1976a]) are all basically due to Meyer (with some help from Dunn on the first, and some help from Routley on the second), and all depend on the same first lemma. The last proof was obtained by Kripke in 1978 and is unpublished (see [Dunn and Meyer, 1989]).

All of the Meyer proofs are what Smullyan [1968] would call ‘synthetic’ in style, and are inspired by Henkin-style methods. The Kripke proof is ‘analytic’ in style, and is inspired by Kanger–Beth–Hintikka tableau-style methods. In actual detail, Kripke’s argument is modelled on completeness proofs for tableau systems, wherein a partial valuation for some open branch is extended to a total valuation. As Kripke has stressed, this avoids the apparatus of inconsistent theories that has hitherto been distinctive of the various proofs of  $\gamma$ ’s admissibility.

We shall sketch the third of Meyer’s proofs, leaving a brief description of the first and second for Section 3.11. Since they depend on semantical notions introduced there.

The strategy of all the Meyer proofs can be divided into two segments: The Way Up and The Way Down. Of course we start with the hypotheses that  $\vdash A$  and  $\vdash \neg A \vee B$ , yet assume not  $\vdash B$  for the sake of *reduction*. We shall be more precise in a moment, but The Way Up involves constructing in a Henkin-like manner a maximal theory  $T$  (containing all the logical theorems) with  $B \notin T$ . The problem though is that  $T$  may be inconsistent in the sense of having both  $C, \neg C \in T$  for some formula  $C$ . (Of course this could not happen in classical logic, for by virtue of the paradox of implication  $C \wedge \neg C \rightarrow B, B$  would be a member of  $T$  contrary to construction.) The Way Down fixes this by finding in effect some subtheory  $T' \subseteq T$  that is both complete and consistent, and indeed is a ‘truth set’ in the sense of [Smullyan, 1968] (Meyer has labelled it the Converse Lindenbaum Lemma). Thus for all formulas  $X$  and  $Y$ ,  $\neg X \in T'$  iff  $X \notin T'$ , and  $X \vee Y \in T'$  iff at

---

<sup>15</sup>This would be less obvious to Routley and Meyer [1976], and Priest [1987; 1995] who raise the ‘consistency of the world’ as a real problem.

least one of  $X$  and  $Y$  is in  $T'$ . So since  $\neg A \vee B \in T'$ , at least one of  $\neg A$  and  $B$  is in  $T'$ . But since  $A \in T'$ , then  $\neg A$  is not in  $T'$ . So  $B$  must be in  $T'$ .<sup>16</sup> But  $T'$  is a subset of  $T$ , which was constructed to keep  $B$  out. So  $B$  cannot be in  $T'$ , and so by *reductio* we obtain  $B$  as desired.

Enough of strategy! We now collect together a few notions needed for a more precise statement of The Way Up Lemma. Incidentally, we shall from this point on in our discussion of  $\gamma$  consider only the case of  $\mathbf{R}$ . Results for  $\mathbf{E}$  (and a variety of neighbours) hold analogously.

By an ‘ $\mathbf{R}$ -theory’ we mean a set of formulas  $T$  of  $\mathbf{R}$  closed under adjunction and logical relevant implication, i.e. such that

1. if  $A, B \in T$ , then  $A \wedge B \in T$ ;
2. if  $\vdash_{\mathbf{R}} A \rightarrow B$  and  $A \in T$ , then  $B \in T$ .

Note that an arbitrary  $\mathbf{R}$ -theory may lack some or all of the theorems of  $\mathbf{R}$  (in classical logic and most familiar logics this would be impossible because of the paradox of strict implication which says that a logical theorem is implied by everything). We thus need a special name for those  $\mathbf{R}$ -theories that contain all of the  $\mathbf{R}$ -theorems—those are called *regular*.<sup>17</sup> In this section, since we have no use of irregular theories and shall be talking only of  $\mathbf{R}$ , by a *theory* we shall always mean a regular  $\mathbf{R}$  theory (irregular  $\mathbf{R}$ -theories however play a great role in the completeness theorems of Section 3 below and there we shall have to be more careful about our distinctions).

A theory  $T$  is called *prime* if whenever  $A \vee B \in T$ , then  $A \in T$  or  $B \in T$ . The converse of this holds for *any* theory  $T$  in virtue of the  $\mathbf{R}$ -axioms  $A \rightarrow A \vee B$  and  $B \rightarrow A \vee B$  and property (2). A theory  $T$  is called *complete* if for every formula  $A$ ,  $A \in T$  or  $\neg A \in T$ , and called *consistent* if for no formula  $A$  do we have both  $A, \neg A \in T$ . In virtue of the  $\mathbf{R}$ -theorem  $A \vee \neg A$ , we have that all prime theories are complete. A consistent prime theory is called *normal*, and it should by now be apparent that a normal theory is a truth set in the sense of Smullyan given above.

Where  $\Gamma$  is a set of formulas, we write  $\Gamma \vdash_{\mathbf{R}} A$  to mean that  $A$  is deducible from  $\Gamma$  in the ‘official sense’ of there being a finite sequence  $B_1, \dots, B_n$ ,

<sup>16</sup>The proof as given here would appear to use *disjunctive syllogism* in the metalanguage at just this point, but it can be restructured (indeed we so restructured the original proofs [Meyer and Dunn, 1969]) so as to avoid at least such an explicit use of *disjunctive syllogism*. The idea is to obtain by distribution ( $A \in T'$  and  $A \notin T'$ ) or ( $B \in T'$  and  $B \notin T'$ ) from the hypothesis  $B \notin T'$ . The whole question of a ‘relevant’ version of the admissibility of  $\gamma$  is a complicated one, and admits of various interpretations. See [Belnap and Dunn, 1981; Meyer, 1978].

<sup>17</sup>It is interesting to note for regular theories, condition (2) may be replaced with the condition

(2') if  $A \in T$  and  $(A \rightarrow B) \in T$ , then  $B \in T$ , in virtue of the  $\mathbf{R}$ -theorem  $A \wedge (A \rightarrow B) \rightarrow B$ .

with  $B_n = A$  and each  $B_i$  being either a member of  $\Gamma$ , or an axiom of  $\mathbf{R}$ , or a consequence of earlier terms by *modus ponens* or adjunction (in context we shall often omit the subscript  $\mathbf{R}$ ). We write  $\Gamma \vdash_{\Delta} A$  to mean that  $\Gamma \cup \Delta \vdash_{\mathbf{R}} A$ , and quite standardly we write things like  $\Gamma, A \vdash_{\mathbf{R}} B$  in place of the more formal  $\Gamma \cup \{A\} \vdash_{\mathbf{R}} B$ . Note that for any theory  $T$ , writing  $\vdash_T A$  in place of  $\phi \vdash_T A$  boils down to saying that  $A$  is a theorem of  $T$  ( $A \in T$ ). Where  $\Delta$  is a set of formulas not necessarily a theory,  $\vdash_{\Delta} A$  can be thought of as saying that  $A$  is deducible from the ‘axioms’  $\Delta$ . The set  $\{A : \vdash_{\Delta} A\}$  is pretty intuitively the smallest theory containing the axioms  $\Delta$ , and we shall label it as  $Th(\Delta)$ .

We can now state and sketch a proof of the

**WAY UP LEMMA.** *Suppose not  $\vdash_{\mathbf{R}} A$ . Then there exists a prime theory  $T$  such that not  $\vdash_T A$ .*

**Proof.** Enumerate the formulas of  $\mathbf{R}$  :  $X_1, X_2, \dots$ . Define a sequence of sets of formulas by induction as follows.

$T_0 =$  set of theorems of  $\mathbf{R}$ .

$T_{i+1} = Th(T_i \cup \{X_{i+1}\})$  if it is not the case that  $T_i, X_{i+1} \vdash A$ ;

$T_i$ , otherwise.

Let  $T$  be the union of all these  $T_n$ ’s. It is easy to see as is standard that  $T$  is a theory not containing  $A$ . Also we can show that  $T$  is prime.

Thus suppose  $\vdash_T X \vee Y$  and yet  $X, Y \notin T$ . Then it is easy to see that since neither  $X$  nor  $Y$  could be added to the construction when their turn came up without yielding  $A$ , we have both

1.  $X \vdash_T A$ ,
2.  $Y \vdash_T A$ .

But by reasonably standard moves ( $\mathbf{R}$  has distribution), we get

3.  $X \vee Y \vdash_T A$ ,

and so  $\vdash_T A$  contrary to the construction. ■

**THE WAY DOWN LEMMA.** *Let  $T'$  be a prime theory. Then there exists a normal theory  $T \subseteq T'$ .*

The concept we need is that of a ‘metavaluation’ (more precisely as we use it here a ‘quasi-metavaluation’, but we shall not bother the reader with such detail). The concept and its use *re*  $\gamma$  may be found in [Meyer, 1976a]. (See also Meyer [1971; 1976b] for other applications.) For simplicity we assume for a while that the only primitive connectives are  $\neg, \vee$  and  $\rightarrow$  ( $\wedge$  can be defined *via* de Morgan). A metavaluation  $v$  is a function from the set of formulas into the truth values  $\{0, 1\}$ , such that

1. for a propositional variable  $p$ ,  $v(p) = 1$  iff  $p \in T$ ;
2.  $v(\neg A) = 1$  iff both (a)  $v(A) = 0$  and (b)  $\neg A \in T$ ;
3.  $v(A \vee B) = 1$  iff either  $v(A) = 1$  or  $v(B) = 1$ .
4.  $v(A \rightarrow B) = 1$  iff both (a)  $v(A) = 0$  or  $v(B) = 1$ , and (b)  $A \rightarrow B \in T$ .

One surprising aspect of these conditions is the double condition in (2) that must be met for  $\neg A$  to be assigned the value 1. Not only must (a)  $A$  be assigned 0 (the usual ‘extensional condition’), but also (b)  $\neg A$  must be a theorem of  $T$  (the ‘intensional condition’). and of course there are similar remarks about (4). The condition in (1) also relies upon  $G$  (actually to a lesser extent than it might seem—when both  $p, \neg p \in T$ , it would not hurt to let  $v(p) = 0$ ).

We now set  $T' = \{A : v(A) = 1\}$ . The following lemma is useful, and has an easy proof by induction on complexity of formulas (the case when  $A$  is a negation evaluated as 0 uses the completeness of  $T$ ).

**COMPLETENESS LEMMA.** *If  $v(A) = 1$ , then  $A \in T$ . If  $v(A) = 0$ , then  $\neg A \in T$ .*

It is reasonably easy to see that  $T'$  is in fact a truth set. That it behaves OK with respect to disjunction can be read right off of clause (3) in the definition of  $v$ , so we need only look at negation where the issue is whether  $T'$  is both consistent and complete. It is clear from clause (2) that  $T'$  is consistent, but  $T'$  is also complete. Thus, suppose  $A \notin T'$ , then by the Completeness Lemma  $\neg A \in T$ . This is the intensional condition for  $v(\neg A) = 1$ , but our supposition that  $A \notin T'$  is just the extensional condition that  $v(A) = 0$ . Hence  $v(\neg A) = 1$ , i.e.  $\neg A \in T'$  as desired.

It is also reasonably easy to check that  $T'$  is an  $\mathbf{R}$ -theory. It is left to the reader to do the easy calculation that  $T'$  is closed under adjunction and  $\mathbf{R}$ -implication, i.e. that these preserve assignments by  $v$  of the value 1. Here we will illustrate the more interesting verification that the  $\mathbf{R}$ -axioms all get assigned the value 1. We shall not actually check all of them, but rather consider several typical ones.

First we check suffixing:  $(A \rightarrow B) \rightarrow [(B \rightarrow C) \rightarrow (A \rightarrow C)]$ . Suppose  $v$  assigns it 0. Since it is a theorem of  $\mathbf{R}$  and *a fortiori* of  $T$ , then it satisfies the intensional condition and so must fail to satisfy the extensional condition. So  $v(A \rightarrow B) = 1$  and  $v((B \rightarrow C) \rightarrow (A \rightarrow C)) = 0$ . By the Completeness Lemma, then  $(A \rightarrow B) \in T$ , and so by *modus ponens* from the very axiom in question (Suffixing) we have that  $(B \rightarrow C) \rightarrow (A \rightarrow C) \in T$ . So  $v((B \rightarrow C) \rightarrow (A \rightarrow C))$  satisfies the intensional condition, and so must fail to satisfy the extensional condition since it is 0. So  $v(B \rightarrow C) = 1$  and  $v(A \rightarrow C) = 0$ . By reasoning analogous to that above (one more *modus ponens*) we derive that  $v(A \rightarrow C)$  must finally fail to satisfy the

extensional condition, i.e.  $v(A) = 1$  and  $v(C) = 0$ . But clearly since all of  $v(A \rightarrow B) = 1$ ,  $v(B \rightarrow C) = 1$ ,  $v(A) = 1$ , then by the extensional condition,  $v(C) = 1$ , and we have a contradiction.

The reader might find it instructive in seeing how negation is handled to verify first the intuitionistically acceptable form of the Reductio axioms  $(A \rightarrow \neg A) \rightarrow \neg A$ , and then to verify its classical variant (used in some axiomatisations of **R**),  $(\neg A \rightarrow A) \rightarrow A$ . The first is easier. Also Classical Double Negation,  $\neg\neg A \rightarrow A$  is fun.

This completes the sketch of Meyer's latest proof of the admissibility of  $\gamma$  for **R**.

## 2.5 $\gamma$ for First-order Relevance Logics

The first proof of the admissibility of  $\gamma$  for first-order **R**, **E**, etc. (which we shall denote as **RQ**, etc.) was in Meyer, Dunn and Leblanc [1974], and uses algebraic methods analogous to those used for the propositional relevance logic in [Meyer and Dunn, 1969]. The proof we shall describe here though will again be Meyer's metavaluation-style proof.

The basic trick needed to handle first-order quantifiers is to produce this time a *first-order truth set*. Assuming that only the universal quantifier  $\forall$  is primitive (the existential can be defined:  $\exists x =_{\text{df}} \neg\forall x\neg$ ), this means we need

$$(\forall) \quad \forall x A \in T \text{ iff } A(a/x) \in T \text{ for all parameters (free variables) } a.$$

This is easily accommodated by adding a clause to the definition of the metavaluation  $v$  so that

$$5. \quad v(\forall x A) = 1 \text{ iff } v(A(a/x)) = 1 \text{ for all parameters } a.$$

This does not entirely fix things, for in proving the Completeness Lemma we have now in the induction to consider the case when  $A$  is of the form  $\forall x B$ . If  $v(\forall x B) = 1$ , then (by (5)),  $va(B(a/x)) = 1$  for all parameters  $a$ . By inductive hypothesis, for all  $a$ ,  $B(a/x) \in T$ . But, and here's the rub, this does not guarantee that  $\forall x B(a/x) \in T$ . We need to have constructed on The Way Up a theory  $T$  that is ' $\omega$ -complete' in just the sense that this guarantee is provided. ([Meyer *et al.*, 1974] call such a theory 'rich'.) Of course it is understood by 'theory' we now mean a 'regular **RQ**-theory', i.e. one containing all of the axioms of **RQ** and closed under its rules (see Section 1.3). Actually things can be arranged as in [Meyer *et al.*, 1974] so that generalisation is in effect built into the axioms so that the only rules can continue to be adjunction and *modus ponens*.

Thus we need the following

**WAY UP LEMMA FOR **RQ**.** *Suppose  $A$  is not a theorem of first-order **RQ**. Then there exists a prime rich theory  $T$  so that  $A \notin T$ .*

This lemma is Theorem 3 of [Meyer *et al.*, 1974], and its proof is of basically a Henkin style with one novelty. In usual Henkin proofs one can assure  $\omega$ -completeness by building into the construction of  $T$  that whenever  $\neg\forall xB$  is put in, then so is  $\neg B(a/x)$  for some *new* parameter  $a$ . This guarantees  $\omega$ -completeness since if  $B(a) \in T$  for all  $a$ , but  $\forall xB \notin T$ , then by completeness  $\neg\forall xB \in T$  and so by the usual construction  $\neg B(a) \in T$  for some  $a$ , and so *by consistency* (?)  $B(a) \notin T$  for some  $a$ , contradicting the hypothesis for  $\omega$ -completeness. But we of course have for relevance logics no guarantee that  $T$  is consistent, as has been remarked above.

The novelty then was to modify the construction so as to keep things out as well as put things in, though this last still was emphasised. Full symmetry with respect to ‘good guys’ and ‘bad guys’ was finally obtained by Belnap,<sup>18</sup> in what is called the Belnap Extension Lemma, which shall be stated after a bit of necessary terminology.

We shall call an ordered pair  $(\Delta, \Theta)$  of sets of formulas of **RQ** and ‘**RQ**-pair’. We shall say that one **RQ** pair  $(\Delta_1, \Theta_1)$  *extends* another  $(\Delta_0, \Theta_0)$  if  $\Delta_0 \subseteq \Delta_1$  and  $\Theta_0 \subseteq \Theta_1$ . An **RQ** pair is defined to be *exclusive* if for no  $A_1, \dots, A_m \in \Delta, B_1, \dots, B_n \in \Theta$  do we have  $\vdash A_1 \wedge \dots \wedge A_m \rightarrow B_1 \vee \dots \vee B_n$ . It is called *exhaustive* if for every formula  $A$ , either  $A \in \Delta$  or  $A \in \Theta$ .<sup>19</sup> It is now easiest to assume that  $\wedge$  and  $\exists$  are back as primitive. We call a set of formulas  $\Gamma$   *$\vee$ -prime* ( *$\wedge$ -prime*) if whenever  $A \vee B \in \Gamma$  ( $A \wedge B \in \Gamma$ ), at least one of  $A$  or  $B \in \Gamma$  (clearly  $\vee$ -primeness is the same as primeness). Analogously, we call  $\Gamma$   *$\exists$ -prime* ( *$\forall$ -prime*) if whenever  $\exists xA \in \Gamma$  ( $\forall xA \in \Gamma$ ), then  $A(a/x) \in \Gamma$  for some  $a$ . Given an **RQ** pair  $(\Delta, \Theta)$  we shall call  $\Delta$  ( $\Theta$ ) *completely prime* if  $\Delta$  is both  $\vee$ - and  $\exists$ -prime ( $\Theta$  is both  $\wedge$ - and  $\forall$ -prime). the pair  $(\Delta, \Theta)$  is called completely prime if both  $\Delta$  and  $\Theta$  are completely prime. We can now state the

**BELNAP EXTENSION LEMMA.** *Let  $(\Delta, \Theta)$  be an exclusive **RQ** pair. Then  $(\Delta, \Theta)$  can be extended to an exclusive, exhaustive, completely prime **RQ** pair  $(T, F)$  in a language just like the language of **RQ** except for having denumerably many new parameters.*

We shall not prove this lemma here, but simply remark that it is a surprisingly straightforward application of Henkin methods to construct a maximal **RQ**-pair and show it has the desired properties (indeed it simply symmetrises the usual Henkin construction of first-order classical logic).

<sup>18</sup>Belnap’s result is unpublished, although he communicated it to Dunn in 1973. Dunn circulated a write-up of it about 1975. It is cited in some detail in [Dunn, 1976d]. Gabbay [1976] contains an independent but precise analogue for the first-order intuitionistic logic with constant domain.

<sup>19</sup>We choose our terminology carefully, not calling  $(\Delta, \Theta)$  a ‘theory’, not using ‘consistency’ for exclusiveness, and not using ‘completeness’ for exhaustiveness. We do this so as to avoid conflict with our earlier (and more customary) usage of these terms and in this we differ on at least one term from usages on other occasions by Gabbay, Belnap, or Dunn.



In order to derive the **RQ** Way Up Lemma we simply set  $\Delta = \mathbf{RQ}$  and  $\Theta = \{A\}$  and extend it to the pair  $(T, F)$  using the Belnap Extension Lemma. It is easy to see that  $T$  is a (*regular*) **RQ**-theory, and clearly  $G$  is prime. but also  $T$  is  $\omega$ -complete. Thus suppose  $B(a/x) \in T$  for all  $a$ , but  $\forall x B \notin T$ . Then by exhaustiveness  $\forall x B \in FR$ . Then by  $\forall$ -primeness,  $B(a/x) \in FR$  for some  $a$ . But since  $\vdash_{\mathbf{RQ}} B(a/x) \rightarrow B(a/x)$ , this contradicts the exclusiveness of the pair  $(T, F)$ .

## 2.6 $\gamma$ for Higher-order Relevance Logics and Relevant Arithmetic

The whole point about  $\gamma$  being merely an *admissible* rule is that it might not hold for various extensions of **F** (cf. [Dunn, 1970] for actual counter examples). Thus, as we just saw, it was an achievement to show that  $\gamma$  continues to be admissible in **R** when it is extended to include first-order quantification. The question of the admissibility of  $\gamma$  naturally has great interest when **R** is further extended to include theories in the foundations of mathematics such as type theory (set theory) and arithmetic.

Meyer [1976a] contains investigations of the admissibility of  $\gamma$  for relevant type theory (**R <sup>$\omega$</sup>** ). We shall report nothing in the way of detail here except to observe that Meyer's result is invariant among various restrictions of the formulas  $A$  in the Comprehension Axiom scheme:

$$\exists X^{x+1} \forall y^n (X^{n+1}(y^n) \leftrightarrow A).$$

As for relevantly formulated arithmetic, most work has gone on in studying Meyer's systems **R <sup>$\sharp$</sup>** , **R <sup>$\sharp\sharp$</sup>**  and their relatives, based on Peano arithmetic, though Dunn has also considered a relevantly formulated version of Robinson Arithmetic [Anderson *et al.*, 1992]. Here we will recount the results for **R <sup>$\sharp$</sup>**  and **R <sup>$\sharp\sharp$</sup>**  for they are rather surprising. In a nutshell,  $\gamma$  is admissible in relevant arithmetics with the infinitary  $\omega$ -rule (from  $A(0)$ ,  $A(1)$ ,  $A(2)$ ,  $\dots$  to infer  $\forall x A(x)$ ), but not without it [Friedman and Meyer, 1992; Meyer, 1998].

The system **R <sup>$\sharp$</sup>**  is given by rewriting the traditional axioms of Peano arithmetic with relevant implication instead of material implication in the natural places. You get the following list of axioms

Identity	$y = z \rightarrow (x = y \rightarrow x = z)$
Successor	$x' = y' \rightarrow x = y$ $x = y \rightarrow x' = y'$ $0 \neq x'$
Addition	$x + 0 = x$ $x + y' = (x + y)'$
Multiplication	$x0 = 0$ $xy' = xy + x$
Induction	$A(0) \wedge \forall x (A(x) \rightarrow A(x')) \rightarrow \forall x A(x)$

which you add to those of  $\mathbf{RQ}$  in order to obtain an arithmetic theory. The question about the admissibility of  $\gamma$  was open for many years, until Friedman teamed up with Meyer to show that it is not [Friedman and Meyer, 1992]. The proof does not provide a direct counterexample to  $\gamma$ . Instead, it takes a more circuitous route. First, we need Meyer's classical containment result for  $\mathbf{R}^\sharp$ . When we map formulae in the extensional vocabulary of arithmetic to the language of  $\mathbf{R}^\sharp$  by setting  $\tau(x = y)$  to  $(x = y) \vee (0 \neq 0)$  and leaving the rest of the map to respect truth functions (so  $\tau(A \wedge B) = \tau(A) \wedge \tau(B)$ ,  $\tau(\neg A) = \neg\tau(A)$  and  $\tau(\forall xA) = \forall x\tau(A)$ ) then we have the following theorem:

$\tau(A)$  is a theorem of  $\mathbf{R}^\sharp$  iff  $A$  is a theorem of classical Peano arithmetic.

This is a subtle result. The proof goes through by showing, by induction, that  $\tau(A)$  is equivalent either to  $(A \wedge (0 = 0)) \vee (0 \neq 0)$  or to  $(A \vee (0 \neq 0)) \wedge (0 = 0)$ , and then that  $\gamma$  and the classical form of induction (with material implication in place of relevant implication) is valid for formulae of this form in  $\mathbf{R}^\sharp$ . Then, if we had the admissibility of  $\gamma$  for  $\mathbf{R}^\sharp$ , we could infer  $A$  from  $\tau(A)$ . (If  $\tau(A)$  is equivalent to  $(A \wedge (0 = 0)) \vee (0 \neq 0)$ , then we can use  $0 = 0$  and  $\gamma$  to derive  $A \wedge (0 = 0)$ , and hence  $A$ . Similarly for the other case).

The next significant result is that not all theorems of classical Peano arithmetic are theorems of  $\mathbf{R}^\sharp$ . Friedman provided a counterexample, which is simple enough to explain here. First, we need some simple preparatory results.

- $\mathbf{R}^\sharp$  is a conservative extension of the theory  $\mathbf{R}^{\sharp+}$  axiomatised by the negation free axioms of  $\mathbf{R}^\sharp$  [Meyer and Urbas, 1986].
- If classical Peano theorem is to be provable in  $\mathbf{R}^\sharp$  and if it contains no negations, then it must be provable in  $\mathbf{R}^{\sharp+}$ .
- Any theorem provable in  $\mathbf{R}^{\sharp+}$  must be provable in the classical positive system  $\mathbf{PA}^+$  which is based on classical logic, instead of  $\mathbf{R}$ .

The proofs of these results are relatively straightforward. The next result is due to Friedman, and it is much more surprising.

- The ring of complex numbers is a model of  $\mathbf{PA}^+$ .

The only difficult thing to show is that it satisfies the induction axiom. For any formula  $A(x)$  in the vocabulary of arithmetic, the set of complex numbers  $\alpha$  such that  $A(\alpha)$  is true is either finite or cofinite. If  $A(x)$  is atomic, then it is equivalent to a polynomial of the form  $f(x) = 0$ , and  $f$  must either have finitely many roots or be 0 everywhere. But the set of either finite or cofinite sets is closed under boolean operations, so no  $A(x)$

we can construct will have an extension which is neither finite or cofinite.) As a result, the induction axiom must be satisfied. For if  $A(0)$  holds and if  $A(x) \supset A(x')$  holds then there are infinitely many complex numbers  $\alpha$  such that  $A(\alpha)$ . So the extension of  $A$  is at least cofinite. But if there is a point  $\alpha$  such that  $A(\alpha)$  fails, then so would  $A(\alpha - 1)$ ,  $A(\alpha - 2)$  and so on by the induction step  $A(x) \supset A(x')$ , and this contradicts the cofinitude of the extension of  $A$ . As a result,  $A(\alpha)$  holds for *every*  $\alpha$ .

We can then use this surprising model of positive Peano arithmetic to construct a Peano theorem which is not a theorem of  $\mathbf{R}^\sharp$ . It is known that for any odd prime  $p$ , there is a positive integer  $y$  which is not a quadratic residue mod  $p$ . That is,  $\exists y \forall z \neg(y \equiv z^2 \pmod{p})$  is provable in Peano arithmetic. This formula can be rewritten in the language of arithmetic with a little work. However, the corresponding formula is false in the complex numbers, so it is not a theorem of  $\mathbf{PA}^+$ . Therefore it isn't a theorem of  $\mathbf{R}^\sharp$ , and by the conservative extension result, it is not a theorem of  $\mathbf{R}^\sharp$ . As a consequence,  $\mathbf{R}^\sharp$  is not closed under  $\gamma$ .

Where is the counterexample to  $\gamma$ ? Meyer's containment result provides a proof of  $\tau(B)$ , where  $B$  is the quadratic residue formula. The  $\gamma$  rule would allow us to derive  $B$  from  $\tau(B)$ , and it is here that  $\gamma$  must fail.

If we replace the induction axiom by the infinitary rule  $\omega$ , we can prove the admissibility of  $\gamma$  using a modification of the Belnap Extension Lemma for the Way Up and using the standard metavaluation technique for the Way Down. The modification of the Belnap Extension Lemma is due to Meyer [1998].

**BELNAP EXTENSION LEMMA, WITH WITNESS PROTECTION:**

Let  $(\Delta, \Theta)$  be an exclusive  $\mathbf{R}^\sharp$  pair in the language of arithmetic (that is, with 0 as the only constant). Then  $(\Delta, \Theta)$  can be extended to an exclusive, exhaustive, completely prime  $\mathbf{R}^\sharp$  pair  $(T, F)$  in the same language.

This lemma requires the  $\omega$ -rule for its proof. Consider the induction stage in which you wish to place  $\forall x A(x)$  in  $\Theta_i$ . The witness condition dictates that there be some term  $t$  such that  $A(t)$  also appear in  $\Theta_i$ . The  $\omega$ -rule ensures that we can do this without the need for a new term, for if no term  $0''\dots'$  could be consistently added to  $\Theta_i$ , then each  $A(0''\dots')$  is a consequence of  $\Delta_i$ , and by the  $\omega$ -rule, so is  $\forall x A(x)$ , contradicting the fact that we can add  $\forall x A(x)$  to  $\Theta_i$ . So, we know that some  $0''\dots'$  will do, and as a result, we need add no new constants to form the complete theory  $T$ . The rest of the way up lemma and the whole of the way down lemma can then be proved with little modification. (for details, see [Meyer, 1998]). Consequently,  $\gamma$  is admissible in  $\mathbf{R}^\sharp$ .

These have been surprising results, and important ones, for relevant arithmetic is an important 'test case' for accounts of relevance. It is a theory in which we can have some fairly clear idea of what it is for one formulae to

properly *follow from* another. In  $\mathbf{R}^\sharp$  and  $\mathbf{R}^{\sharp\sharp}$ , we have  $0 = 2 \rightarrow 0 = 4$  because there is an ‘arithmetically appropriate’ way to derive  $0 = 4$  from  $0 = 2$  — by multiplying both sides by 2. However, we cannot derive  $0 = 2 \rightarrow 0 = 3$ , and, correspondingly, there is no way to derive  $0 = 3$  from  $0 = 2$  using the resources of arithmetic. The only way to do it within the vocabulary is to appeal to the falsity of  $0 = 3$ , and this is not a relevantly acceptable move.  $0 \neq 3 \rightarrow (0 = 3 \rightarrow 0 = 2)$  does not have much to recommend as pattern of reasoning which respects the canons of relevance.

We are left with important questions. Are there axiomatisable extensions of  $\mathbf{R}^\sharp$  which are closed under  $\gamma$ ? Can theories like  $\mathbf{R}^\sharp$  and  $\mathbf{R}^{\sharp\sharp}$  be extended to deal with more interesting mathematical structures, while keeping account of some useful notion of relevance? Early work on this area, from a slightly different motivation (paraconsistency, not relevance) indicates that there are some interesting results at hand, but the area is not without its difficulties [Mortensen, 1995].

The admissibility of  $\gamma$  would also seem to be of interest for relevant type theory (even relevant second-order logic) with an axiom of infinity (see [Dunn, 1979b]).

One of the chief points of philosophical interest in showing the admissibility of  $\gamma$  for some relevantly formulated version of a classical theory relates to the question of the consistency of the classical theory (this was first pointed out in Meyer, Dunn and Leblanc [1974]). As we know from Gödel’s work, interesting classical theories cannot be relied upon to prove their own consistency. To exaggerate perhaps only a little, the consistency of systems like Peano (even Robinson) arithmetic must be taken in faith.

But using relevance logic in place of classical logic in formulating such theories gives us a new strategy of faith. It is conceivable that since relevance logic is weaker than classical logic, the consistency of the resultant theory might be easier to demonstrate. This has proved true at least in the sense of absolute consistency (some sentence is unprovable) as shown by [Meyer, 1976c] for Peano arithmetic using elementary methods. Classically of course there is no difference between absolute consistency and ordinary (negation) consistency (for no sentence are both  $A$  and  $\neg A$  provable), and if  $\gamma$  is admissible for the theory, then this holds for relevance logic, too. The interesting thing then would be to produce a proof of the admissibility of  $\gamma$ , which we know from Gödel would itself have to be non-elementary.

One could then imagine arguing with a classical mathematician in the following Pascal’s Wager sort of way [Dunn, 1980a].

Look. You have equally good reason to believe in the negation consistency of the classical system and the (relative) completeness of the relevant system. In both cases you have a non-elementary proof which secures your belief, but which might be mistaken. Consider the consequences in each case if it is mis-

taken. If you are using the classical system, disaster! Since even one contradiction classically implies everything, for each theorem you have proven, you might just as well have proven its negation. But if you are using the relevant system, things are not so bad. For at least large classes of sentences, it can be shown by elementary methods (Meyer's work) that not both the sentences and their negations are theorems.

### 2.7 Ackermann's $\gamma$ and Gentzen's Cut: Gentzen Systems as Relevance Logic

In [Meyer *et al.*, 1974] an analogy was noted between the role that the admissibility of  $\gamma$  plays in relevance logic and the role that cut elimination plays in Gentzen calculi (even those for classical systems). For the reader unfamiliar with Gentzen calculi, this subsection will make more sense after she has read Sections 4.6 and 4.7. The Gentzen system for the classical propositional calculus **LK** with the material conditional and negation as primitive (as is well-known, all of the other truth-functional connectives can be defined from these) may be obtained by adding to the rules of **LR** $\supset$  of Section 4.7. the rule of Thinning (see Section 4.6) on both the left and right. Gentzen also had as a primitive rule:

$$\frac{\alpha \vdash A, B \quad \gamma, A \vdash \delta}{\alpha, \gamma \vdash B, \delta}, \quad (\text{Cut})$$

which has as a special case

$$\frac{\vdash A \vdash B}{\vdash B}. \quad (1)$$

Since  $A \vdash B$  is derivable just when  $\vdash A \rightarrow B$  is derivable, and since in classical logic  $A \rightarrow B$  is equivalent to  $\neg A \vee B$ , (1) above is in effect

$$\frac{\vdash A \quad \vdash \neg A \vee B}{\vdash B}, \quad (1')$$

which is just  $\gamma$ .

All of the Gentzen rules except Cut have the Subformula Property: Every formula that occurs in the premises also occurs in the conclusion, though perhaps there as a subformula. Gentzen showed *via* his *Hauptsatz* that Cut was redundant—it could be eliminated without loss (hence this is often called the Elimination Theorem). Later writers have tended to think of Gentzen systems as lacking the Cut Rule, and so the Elimination Theorem is stated as showing that Cut is admissible in the sense that whenever the premises are derivable so is the conclusion. There is thus even a parallel

historical development with Ackermann's rule  $\gamma$  in relevance logic, since writers on relevance logic have tended to follow Anderson and Belnap's decision to drop  $\gamma$  as a primitive rule.

Note that the Subformula Property can be thought of as a kind of relation of relevance between premises and conclusion. Thus Cut as primitive destroys a certain kind of relevance property of Gentzen systems, just as  $\gamma$  as primitive destroys the relevance of premises to conclusion in relevance logic. The analogies become even clearer if we reformulate Gentzen's system according to the following ideas of [Schütte, 1956].

The basic objects of Gentzen's calculus  $LK$  were the *sequents*  $A_1, \dots, A_m \vdash B_1, \dots, B_n$ , where the  $A_i$ 's and  $B_j$ 's are formulas (any or all of which might be missing). Such a sequent may be interpreted as a statement to the effect that either one of the  $A_i$ 's is false or one of the  $B_j$ 's is true. To every such sequent there corresponds what we might as well call its 'right-handed counterpart':

$$\vdash \neg A_1, \dots, \neg A_m, B_1, \dots, B_n$$

It is possible to develop a calculus parallel to Gentzen's using only 'right-handed' sequents, i.e. those with empty left side. This is in effect what Schütte did, but with one further trick. Instead of working with a right-handed sequent  $\vdash A_1, \dots, A_m$ , which can be thought of as a sequence of formulas, he in effect replaced it with the single formula  $A_1 \vee \dots \vee A_m$ .<sup>20</sup>

With these explanations in mind, the reader should have no trouble in perceiving Schütte's calculus  $K_1$  as 'merely' a notational variant of Gentzen's original calculus  $LK$  (albeit, a highly ingenious one). Also Schütte's system had the existential quantifier which we have omitted here purely for simplicity. Dunn and Meyer [1989] treats it as well.

The axioms of  $K_1$  are all formulas of the form  $A \vee \neg A$ . The inference rules divide themselves into two types:

*Structural rules:*

$$\frac{\mathcal{M} \vee A \vee B \vee \mathcal{N}}{\mathcal{M} \vee B \vee A \vee \mathcal{N}} \text{ [Interchange]} \quad \frac{\mathcal{N} \vee A \vee A}{\mathcal{N} \vee A} \text{ [Contraction]}$$

*Operational rules:*

$$\frac{\mathcal{N}}{\mathcal{N} \vee B} \text{ [Thinning]} \quad \frac{\mathcal{N} \vee \neg A \quad \mathcal{N} \vee \neg B}{\mathcal{N} \vee \neg(A \vee B)} \text{ [de Morgan]} \quad \frac{\mathcal{N} \vee A}{\mathcal{N} \vee \neg \neg A} \text{ [Double Negation]}$$

It is understood in every case but that of Thinning that either both of  $\mathcal{M}$  and  $\mathcal{N}$  may be missing. Also there is an understanding in multiple disjunctions that parentheses are to be associated to the right.

<sup>20</sup>It ought be noted that similar "single sided" Gentzen systems find extensive use in the proof theory for Linear Logic [Girard, 1987; Troelstra, 1992].

In [Meyer *et al.*, 1974] it was said that the rule Cut is just  $\gamma$  ‘in peculiar notation’. In the context of Schütte’s formalism the notation is not even so different. Thus:

$$\frac{\mathcal{M} \vee A \quad \neg A \vee \mathcal{N}}{\mathcal{M} \vee \mathcal{N}} [\text{Cut}] \quad \frac{A \quad \neg A \vee B}{B} [\gamma].$$

Since either  $\mathcal{M}$  or  $\mathcal{N}$  may be missing, obviously  $\gamma$  is just a special case of Cut.

It is pretty easy to check that each of the rules above corresponds to a provable first-degree relevant implication. Indeed [Anderson and Belnap Jr., 1959a] with their ‘Simple Treatment’ formulation of classical logic (extended to quantifiers in [Anderson and Belnap Jr., 1959b]) independently arrived at a Cut-free system for classical logic much like Schütte’s (but with some improvements, i.e. they have more general axioms and avoid the need for structural rules). They used this to show that **E** contains all the classical tautologies as theorems, the point being that the Simple Treatment rules are all provable entailments in **E** (unlike the usual rule for axiomatic formulations of classical logic, *modus ponens* for the material conditional, i.e.  $\gamma$ ). Thus the later proven admissibility of  $\gamma$  was not needed for this purpose, although it surely can be so used. Schütte’s system can also clearly be adapted to the purpose of showing that classical logic is contained in relevance logic, and indeed [Belnap, 1960a] used **K**<sub>1</sub> (with its quantificational rules) to show that **EQ** contains all the theorems of classical first-order logic.

It turns out that one can give a proof of the admissibility of Cut for a classical Gentzen-style system, say Schütte’s **K**<sub>1</sub>, along the lines of a Meyer-style proof of the admissibility of  $\gamma$  (see [Dunn and Meyer, 1989], first reported in 1974).<sup>21</sup> We will not give many details here, but the key idea is to treat the rules of **K**<sub>1</sub> as rules of deducibility and not merely as theorem generating devices. Thus we define a *deduction* of  $A$  from a set of formulas  $\Gamma$  as a finite tree of formulas with  $A$  as its origin, members of  $\Gamma$  or axioms of **K**<sub>1</sub> at its tips, and such that each point that is not a tip follows from the points just above it by one of the rules of **K**<sub>1</sub> (this definition has to be slightly more complicated if quantifiers are present due to usual problems caused by generalisation). We can then inductively build a prime complete

---

<sup>21</sup>We hasten to acknowledge the nonconstructive character of this proof. In this our proof compares with that of Schütte [1956] (also proofs for related formalisms due to Anderson and Belnap, Beth, Hintikka, Kanger) in its uses of semantical (model-theoretic) notions, and differs from Gentzen’s. Like the proofs of Schütte *et al.* this proof really provides a completeness theorem. We may briefly label the difference between this proof and those of Schütte and the others by using (loosely) the jargon of Smullyan [1968]. Calling both Hilbert-style formalisms and their typical Henkin-style completeness proofs ‘synthetic’, and calling both Gentzen-style formalisms and their typical Schütte-style completeness proof ‘analytic’, it looks as if we can be said to have given an synthetic completeness proof for an analytic formalism.

theory (closed under deducibility) on The Way up, which will clearly be inconsistent since because of the ‘Subformula Property’ clearly, e.g.  $q$  is not deducible from  $p, \neg p$ . but this can be fixed on The Way Down by using metavaluation techniques so as to find a complete consistent subtheory.

In 1976 E. P. Martin, Meyer and Dunn extended and analogised the result of Meyer concerning the admissibility of  $\gamma$  for relevant type theory described in the last subsection, in much the same way as the  $\gamma$  argument for the first-order logic has been analogised here, so as to obtain a new proof of Takeuti’s Theorem (Cut-elimination for simple type theory). This unpublished proof dualises the proof of Takahashi and Prawitz (cf. [Prawitz, 1965]) in the same way that the proof here dualises the usual semantical proofs of Cut-elimination for classical first-order logic. This dualisation is vividly described by saying that in place of ‘Schütte’s Lemma’ that every semi- (partial-) valuation may be extended to a (total) valuation, there is instead the ‘Converse Schütte Lemma’ that every ‘ambi-valuation’ (sometimes assigns a sentence both the values 0, 1) may be restricted to a (consistent) valuation.

### 3 SEMANTICS

#### 3.1 Introduction

In Anderson’s [1963] ‘open problems’ paper, the last major question listed, almost as if an afterthought, was the question of the semantics of **E** and **EQ**. Despite this appearance Anderson said (p. 16) ‘the writer does *not* regard this question as “minor”; it is rather the principle large question remaining open’. Anderson cited approvingly some earlier work of Belnap’s (and his) on providing an algebraic semantics for first-degree entailments, and said (p. 16), ‘But the general problem of finding a semantics for the whole of **E**, with an appropriate completeness theorem, remains unsolved’.

It is interesting to note that Anderson’s paper appeared in the same *Acta Philosophica Fennica* volume as the now classic paper of Kripke [1963] which provided what is now simply called ‘Kripke-style’ semantics for a variety of modal logics (Kripke [1959a] of course provided a semantics for **S5**, but it lacked the accessibility relation  $R$  which is so versatile in providing variations).

When Anderson was writing his ‘open problems’ paper, the paradigm of a semantical analysis of a non-classical logic was probably still something like the work of McKinsey and Tarski [1948], which provided interpretations for modal logic and intuitionistic logic by way of certain algebraic structures analogous to the Boolean algebras that are the appropriate structures for classical logic. But since then the Kripke-style semantics (sometimes referred to as ‘possible-worlds semantics’, or ‘set-theoretical semantics’) seems



to have become the paradigm. Fortunately, **E** and **R** now have both an algebraic semantics and a Kripke-style semantics. We shall first distinguish in a kind of general way the differences between these two main approaches to semantics, before going on to explain the particular details of the semantics for relevant logics (again **R** will be our paradigm).

### 3.2 Algebraic vs. Set-theoretical Semantics

It is convenient to think of a logical system as having two distinct aspects syntax (well-formed strings of symbols, e.g. sentences) and semantics (what, e.g. these sentences mean, i.e. propositions). These double aspects compete with one another as can be seen in the competing usages ‘sentential calculus’ and ‘propositional calculus’, but we should keep firmly in mind both aspects.

Since sentences can be combined by way of connectives, say the conjunction sign  $\wedge$ , to form further sentences, typically there is for each logical system at least one natural algebra arising at the level of syntax, the algebra of sentences (if one has a natural logical equivalence relation there is yet another that one obtains by identifying logically equivalent sentences together into equivalence classes—the so-called ‘Lindenbaum algebra’). And since propositions can be combined by the corresponding logical operations, say conjunction, to form propositions, here is an analogous algebra of propositions.

Now undoubtedly some readers, who were taught to ‘Quine’ propositions from an early age, will have troubles with the above story. The same reader would most likely not find compelling any particular metaphysical account we might give of numbers. We ask that reader then to at least suspend *disbelief* in propositions so that we can get on with the mathematics.

There is an alternative approach to semantics which can be described by saying that rather than taking propositions as primitive, it ‘constructs’ them out of certain other semantical primitives. Thus there is as a paradigm of this approach the so-called ‘UCLA proposition’ as a set of ‘possible worlds’.<sup>22</sup> We here want to stress the general structural idea, not placing much emphasis upon the particular choice of ‘possible *worlds*’ as the semantical primitive. Various authors have chosen ‘reference points’, ‘cases’, ‘situations’, ‘set-ups’, etc.—as the name for the semantical primitive varying for sundry subtle reasons from author to author. We have both in relevance logic contexts have preferred ‘situations’, but in a show of solidarity we shall here join forces with the Routley’s [1972] in their use of ‘set-ups’.

Such ‘set-theoretical’ semantical accounts do not always explicitly verify such a construction of propositions. Indeed perhaps the more common approach is to provide an interpretation that says whether a formula  $A$  is

---

<sup>22</sup>Actually the germ of this idea was already in Boole (cf. [Dipert, 1978]), although apparently he thought of it as an analogy rather than as a reduction.

true and false at a given set-up  $S$  writing  $\phi(a, S) = T$  or  $S \models A$  or some such thing. Think of Kripke's [1963] presentation of his semantics for modal logic. But (unless one has severe ontological scruples about sets) one might just as well interpret  $A$  by assigning it a class of set-ups, writing  $\Phi(A)$  or  $|A|$  or some such thing. One can go from one framework to the other by way of equivalence

$$S \in |A| \text{ iff } S \models A.$$

### 3.3 Algebra of First-degree Relevant Implications

Given two propositions  $a$  and  $b$ , it is natural to consider the implication relation among them, which we write as  $a \leq b$  (' $a$  implies  $b$ '). It might be thought to be natural to write this the other way around as  $a \geq b$  on some intuition that  $a$  is the stronger or 'bigger' one if it implies  $b$ . Also it suggests  $a \supseteq b$  (' $b$  is contained in  $a$ '), which is a natural enough way to think of implication. There are good reasons though behind our by now almost universal choice (of course at one level it is just notation, and it doesn't matter what your convention is). Following the idea that a proposition might be identified with the set of cases in which it is true,  $a$  implies  $b$  corresponds to  $a \subseteq b$ , which has the same direction as  $a \leq b$ . Then conjunction  $\wedge$  corresponds to intersection  $\cap$ , and they have roughly the same symbol (and similarly for  $\vee$  and  $\cup$ ).

It is also natural to assume, as the notation suggests, that implication is a partial order, i.e.

- (p.o.1)  $a \leq a$  (Reflexivity),
- (p.o.2)  $a \leq b$  and  $b \leq a \Rightarrow a = b$  (Antisymmetry),
- (p.o.3)  $a \leq b$  and  $b \leq c \Rightarrow a \leq c$  (Transitivity).

It is natural also to assume that there are operations of conjunction  $\wedge$  and disjunction  $\vee$  that satisfy

- ( $\wedge$ lb)  $a \wedge b \leq a$ ,  $a \wedge b \leq b$ ,
- ( $\wedge$ glb)  $x \leq a$  and  $x \leq b \Rightarrow x \leq a \wedge b$ ,
- ( $\vee$ ub)  $a \leq a \vee b$ ,  $b \leq a \vee b$ ,
- ( $\vee$ lub)  $a \leq x$  and  $b \leq x \Rightarrow a \vee b \leq x$ .

Note that ( $\wedge$ lb) says that  $a \wedge b$  is a lower bound both of  $a$  and of  $b$ , and ( $\wedge$ glb) says it is the greatest such lower bound. Similarly  $a \vee b$  is the least upper bound of  $a$  and  $b$ .

A structure  $(L, \leq, \wedge, \vee)$  satisfying all the properties above is a well-known structure called a lattice. Almost any logic would be compatible with the assumption that propositions form a lattice (but there are exceptions, witness Parry's [1933] Analytic Implication which would reject ( $\vee$ ub)).

Lattices can be defined entirely operationally as structures  $(L, \wedge, \vee)$  with the relation  $a \leq b$  defined as  $a \wedge b = a$ . Postulates characterising the operations are:

$$\begin{aligned} \text{Idempotence:} & \quad a \wedge a = a, a \vee a = a \\ \text{Commutativity:} & \quad a \wedge b = b \wedge a, a \vee b = b \vee a \\ \text{Associativity:} & \quad a \wedge (b \wedge c) = (a \wedge b) \wedge c, a \vee (b \vee c) = (a \vee b) \vee c \\ \text{Absorption:} & \quad a \wedge (a \vee b) = a, a \vee (a \wedge b) = a. \end{aligned}$$

An (upper) *semi-lattice* is a structure  $(S, \vee)$ , with  $\vee$  satisfying Idempotence, Commutativity, and Associativity.

Given two lattices  $(L, \wedge, \vee)$  and  $(L', \wedge', \vee')$ , a function  $h$  from  $L$  into  $L'$  is called a (*lattice*) *homomorphism* if both  $h(a \wedge b) = h(a) \wedge' h(b)$  and  $h(a \vee b) = h(a) \vee' h(b)$ . If  $h$  is one-one,  $h$  is called an *isomorphism*.

Many logics (certainly orthodox relevance logic) would insist as well that propositions form a *distributive* lattice, i.e. that

$$a \wedge (b \vee c) \leq (a \wedge b) \vee c.$$

This implies the usual distributive laws  $a \wedge (b \vee c) = (a \wedge b) \wedge (a \wedge c)$  and  $a \vee (b \wedge c) = (a \vee b) \wedge (a \vee c)$ . (Again there are exceptions, important ones being quantum logic with its weaker orthomodular law, and linear logic with its rejection of even the orthomodular law.)

The paradigm example of a distributive lattice is a collection of sets closed under intersection and union (a so-called ‘ring’ of sets). Stone [1936] indeed showed that abstractly all distributive lattices can be represented in this way. Although we will not argue this here, it is natural to think that if propositions correspond to classes of cases, then conjunction should carry over to intersection and disjunction to union, and so productions should form a distributive lattice.

Certain subsets of lattices are especially important. A *filter* is a non-empty subset  $F$  such that

1.  $a, b \in F \Rightarrow a \wedge b \in F$ ,
2.  $a \in F$  and  $a \leq b \Rightarrow b \in F$ .

Filters are like theories. Note by easy moves that a filter satisfies

- 1'.  $a, b \in F \Leftrightarrow a \wedge b \in F$ ,
- 2'.  $a \in F$  or  $b \in F \Rightarrow a \vee b \in F$ .

When a filter also satisfies the converse of (2') it is called *prime*, and is like a prime theory. A filter that is not the whole lattice is called *proper*. Stone [1936] showed (using an equivalent of the Axiom of Choice) the

**PRIME FILTER SEPARATION PROPERTY.** In a distributive lattice, if  $a \not\leq b$ , then there exists a prime filter  $P$  with  $a \in P$  and  $b \notin P$ .

This is related to the Belnap Extension Lemma of Section 2.5.

So far we have omitted discussion of negation. This is because there is less agreement among logics as to what properties it should have.<sup>23</sup> There is, however, widespread agreement that it should at least have these:

1. (Contraposition)  $a \leq b \Rightarrow \neg b \leq \neg a$ ,
2. (Weak Double Negation)  $a \leq \neg\neg a$ .

These can both be neatly packaged in one law:

3. (Intuitionistic Contraposition)  $a \leq \neg b \Rightarrow b \leq \neg a$ .

We shall call any unary function  $\neg$  satisfying (3) (or equivalently (1) and (2)) a *minimal complement*. The intuitionists of course do not accept

4. (Classical Contraposition)  $\neg a \leq \neg b \Rightarrow b \leq a$ , or
5. (Classical Double Negation)  $\neg\neg a \leq a$ .

If one adds either of (4) or (5) to the requirements for a minimal complement one gets what is called a *de Morgan complement* (or quasi-complement), because, as can be easily verified, it satisfies all of de Morgan's laws

$$\begin{aligned} \text{(deM1)} \quad & \neg(a \wedge b) = \neg a \vee \neg b, \\ \text{(deM2)} \quad & \neg(a \vee b) = \neg a \wedge \neg b. \end{aligned}$$

Speaking in an algebraic tone of voice, de Morgan complement is just a (one-one) order-inverting mapping (a *dual automorphism*) of period two.

De Morgan complement captures many of the features of classical negation, but it misses

$$\begin{aligned} \text{(Irrelevance 1)} \quad & a \wedge \neg a \leq b, \\ \text{(Irrelevance 2)} \quad & a \leq b \vee \neg b. \end{aligned}$$

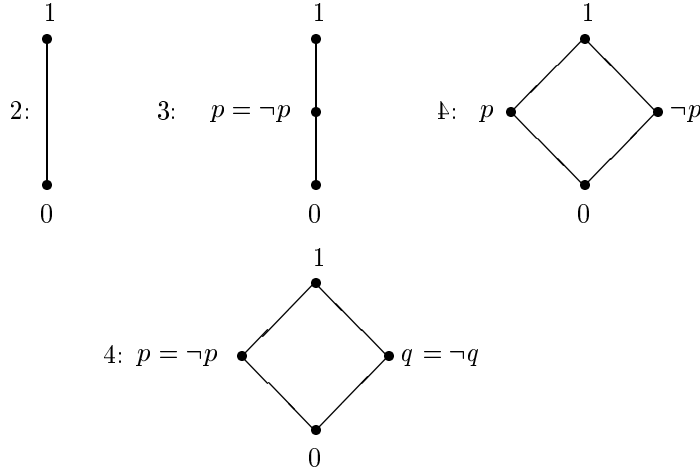
If (either of) these are added to a de Morgan complement it becomes a *Boolean complement*. If Irrelevance 1 is added to a minimal complement, it becomes a *Heyting complement* (or *pseudo-complement*).

A structure  $(L, \wedge, \vee, \neg)$ , where  $(L, \wedge, \vee)$  is a *distributive* lattice and  $\neg$  is a de Morgan (Boolean) complement is called a *de Morgan (Boolean) lattice*. Note that we did not try to extend this terminological framework to 'Heyting lattices', because in the literature a Heyting lattice requires an operation called 'relative pseudo-complementation' in addition to Heyting complementation (plain pseudo-complementation).

As an example of de Morgan lattices consider the following (here we use ordinary Hasse diagrams to display the order;  $a \leq b$  is displayed by putting  $a$  in a connected path below  $b$ ):

---

<sup>23</sup>Cf. Dunn [1994; 1996] wherein the various properties below are related to various ways of treating incompatibility between states of information.



The backwards numeral labelling the third lattice over is not a misprint. It signifies that not only has the de Morgan complement been obtained by inverting the order of the diagram, as in the order three (of course  $\neg I = \Theta$  and *vice versa*), but also by rotating it from right to left at the same time. 2 and 5 are Boolean lattices.

A *homomorphism (isomorphism)*  $h$  between de Morgan Lattice with de Morgan complements  $\neg$  and  $\neg'$  respectively is a lattice homomorphism (isomorphism) such that  $h(\neg a) = \neg' h(a)$ .

A valuation in a lattice outfitted with one or the other of these ‘complementations’ is a map  $v$  from the zero-degree formulas into its elements satisfying

$$\begin{aligned} v(\neg A) &= \neg v(A), \\ v(A \wedge B) &= v(A) \wedge v(B), \\ v(A \vee B) &= v(A) \vee v(B). \end{aligned}$$

Note that the occurrence of ‘ $\neg$ ’ on the left-hand side of the equation denotes the negation connective, whereas the occurrence on the right-hand side denotes the complementation operation in the lattice (similarly for  $\wedge$  and  $\vee$ ). Such ambiguities resolve themselves contextually.

A valuation  $v$  can be regarded as an interpretation of the formulas as propositions.

De Morgan lattices have become central to the study of relevance of logic, but they were antecedently studied, especially in the late 1950s by Moisil and Monteiro, by Białyński-Birula and Rasiowa (as ‘quasi-Boolean algebras’), and by Kalman (as ‘distributive  $i$ -lattices’) (see Anderson and Belnap [1975] or Rasiowa [1974] for references and information).

Belnap seems to have first recognised their significance for relevance logic, though his research favoured a special variety which he called an *intensionally complemented distributive lattice with truth filter* ('icdlw/TF'), shortened in Section 18 of [Anderson and Belnap, 1975] to just *intensional lattice*. An intensional lattice is a structure  $(L, \wedge, \vee, \neg, T)$ , where  $(L, \wedge, \vee, \neg)$  is a de Morgan lattice and  $T$  is a *truth-filter*, i.e.  $T$  is a filter which is complete in the sense  $T$  contains at least one of  $a$  and  $\neg a$  for each  $a \in L$ , and *consistent* in the sense that  $T$  contains no more than one of  $a$  and  $\neg a$ .

Belnap and Spencer [1966] showed that a necessary and sufficient condition for a de Morgan lattice to have a truth filter is that negation have no fixed point, i.e. for no element  $a$ ,  $a = \neg a$  (such a lattice was called an *icdl*). For Boolean algebras this is a non-degeneracy condition, assuring that the algebra has more than one element, the one element Boolean algebra being best ignored for many purposes. But the experience in relevance logic has been that de Morgan lattices where some elements are fixed points are extremely important (not all elements can be fixed points or else we do have the one element lattice).

The viewpoint of [Dunn, 1966] was to take general de Morgan lattices as basic to the study of relevance logics (though still results were analogised wherever possible to the more special icdl's). Dunn [1966] showed that upon defining a first-degree implication  $A \rightarrow B$  to be (*de Morgan*) *valid* iff for every valuation  $v$  in a de Morgan lattice,  $v(A) \leq v(B)$ ,  $A \rightarrow B$  is valid iff  $A \rightarrow B$  is a theorem of  $\mathbf{R}_{fde}$  (or  $\mathbf{E}_{fde}$ ). The analogous result for icdl's (in effect due to Belnap) holds as well.

Soundness ( $\vdash_{\mathbf{R}_{fde}} A \rightarrow B \Rightarrow A \rightarrow B$  is valid) is a more or less trivial induction on the length of proofs in  $\mathbf{R}_{fde}$  fragment—cf. [Anderson and Belnap, 1975, Section 18].

Completeness ( $A \rightarrow B$  valid  $\Rightarrow \vdash_{\mathbf{R}_{fde}} A \rightarrow B$ ) is established by proving the contrapositive. We suppose not  $\vdash_{\mathbf{R}_{fde}} A \rightarrow B$ . We then form the 'Lindenbaum algebra', which has as an element for each zero degree formula (zdf)  $X$ ,  $[X] =_{df} \{Y : Y \text{ is a zdf and } \vdash_{\mathbf{R}_{fde}} X \leftrightarrow Y\}$ . Operations are defined so that  $\neg[X] = [\neg X]$ ,  $[X] \wedge [Y] = [X \wedge Y]$ , and  $[X] \vee [Y] = [X \vee Y]$ , and we set  $[X] \leq [Y]$  whenever  $\vdash_{\mathbf{R}_{fde}} X \rightarrow Y$ . It is more or less transparent, given  $\mathbf{R}_{fde}$  formulated as it is, that the result is a de Morgan lattice. It is then easy to see that  $A \rightarrow B$  is invalidated by the *canonical* valuation  $v_c(X) = [X]$ , since clearly  $[A] \not\leq [B]$ .

The above kind of soundness and completeness result is really quite trivial (though not unimportant), once at least the logic has had its axioms chopped so that they look like the algebraic postulates merely written in a different notation. The next result is not so trivial.

**CHARACTERISATION THEOREM OF  $\mathbf{R}_{FDE}$  WITH RESPECT TO 4.**  $\vdash_{\mathbf{R}_{fde}} A \rightarrow B$  iff  $A \rightarrow B$  is valid in  $\mathcal{A}$ , i.e. for every valuation  $v$  in  $\mathcal{A}$ ,  $v(A) \leq v(B)$ .

**Proof.** Soundness follows from the trivial fact recorded above that  $\mathbf{R}_{fde}$  is

sound with respect to de Morgan lattices in general. For completeness we need the following:

*4-Valued Homomorphism Separation Property.* Let  $\mathcal{D}$  be a de Morgan lattice with  $a \not\leq b$ . Then there exists a homomorphism  $h$  of  $\mathcal{D}$  into 4 so that  $h(a) \not\leq h(b)$ .

Completeness will follow almost immediately from this result, for upon supposing that  $\text{not } \vdash_{\mathbf{R}_{\text{de}}} A \rightarrow B$ , we have  $v(A) = h[A] \not\leq h[B] = v(B)$  (the composition of a homomorphism with a valuation is transparently a valuation). So we go on to establish the Homomorphism Separation Property.

Assume that  $a \not\leq b$ . By the Prime Filter Separation Property, we know there is a prime filter  $P$  with  $a \in P$  and  $b \notin P$ . for a given element  $x$ , we define  $h(x)$  according to the following four possible ‘complementation patterns’ with respect to  $P$ .

1.  $x \in P, \neg x \notin P$ : set  $h(x) = 1$ ;
2.  $\neg x \in P, x \notin P$ : set  $h(x) = 0$ ;
3.  $x \in P, \neg x \in P$ : set  $h(x) = p$ ;
4.  $x \notin P, \neg x \notin P$ : set  $h(x) = q$ .

It is worth remarking that if  $\mathcal{D}$  is a Boolean lattice, (3) (inconsistency) and (4) (incompleteness) can never arise, which explains the well-known significance of 2 for Boolean homomorphism theory. Clearly these specifications assure that  $h(a) \in \{p, 1\}$  and  $h(b) \in \{q, 0\}$ , and so by inspection  $h(a) \not\leq h(b)$ . It is left to the reader to verify that  $h$  in fact is a homomorphism. (Hint to avoid more calculation: set  $[p] = \{p, 1\}$  and  $[q] = \{q, 0\}$  (the principal filters determined by  $p$  and  $q$ ). Observe that the definition of  $h$  above is equivalent to requiring of  $h$  that  $h(x) \in [p]$  iff  $x \in P$ , and  $h(x) \in [q]$  iff  $\neg a \notin P$ . Observe that if whenever  $i = p, q, y \in [i]$  iff  $z \in [i]$ , then  $y = z$ . Show for  $i = p, q$ ,  $h(a \wedge b) \in [i]$  iff  $h(a) \wedge h(b) \in [i]$ ,  $h(a \vee b) \in [i]$  iff  $h(a) \vee h(b) \in [i]$ , and  $h(\neg a) \in [i]$  iff  $\neg h(a) \in [i]$ . ■

### 3.4 Set-theoretical Semantics for First-degree Relevant Implication

Dunn [1966] (cf. also [Dunn, 1967]) considered a variety of (effectively equivalent) representations of de Morgan lattices as structures of sets. We shall here discuss the two of these that have been the most influential in the development of set-theoretical semantics for relevance logic.

The earliest one of these is due to Białynicki-Birula and Rasiowa [1957] and goes as follows. Let  $U$  be a non-empty set and let  $g : U \rightarrow U$  be such that it is of period two, i.e.

1.  $g(g(x)) = x$ , for all  $x \in U$ .

(We shall call the pair  $(U, g)$  and *involved set*— $g$  is the *involution*, and is clearly 1–1). Let  $Q(U)$  be a ‘ring’ of subsets of  $U$  (closed under  $\cap$  and  $\cup$ ) closed as well under the operation of ‘quasi-complement’

$$2. \neg X = U - g[X] (X \subseteq U).$$

$(Q(U), \cup, \cap, \neg)$  is called a *quasi-field of sets* and is a de Morgan lattice.

QUASI-FIELDS OF SETS THEOREM [Białynicki-Birula and Rasiowa, 1957].  
Every de Morgan lattice  $\mathcal{D}$  is isomorphic to a quasi-field of sets.

**Proof.** Let  $U$  be the set of all prime filters of  $\mathcal{D}$ , and let  $P$  range over  $U$ . Let  $\neg P \rightarrow \{\neg a : a \in P\}$ , and define  $g(P) = \mathcal{D} - \neg P$ . We leave to the reader to verify that  $U$  is closed under  $g$ . For each element  $a \in \mathcal{D}$ , set

$$f(a) = \{P : a \in P\}.$$

Clearly  $f$  is one–one because of the Prime Filter Separation Property, so we need only check that  $f$  preserves the operations.

*ad* $\wedge$ :  $P \in f(a \wedge b) \Leftrightarrow a \wedge b \in P \Leftrightarrow ((1')$  of Section 3.3)  $a \in P$  and  $b \in P \Leftrightarrow P \in f(a)$  and  $P \in f(b) \Leftrightarrow P \in f(a) \cap f(b)$ . So  $f(a \wedge b) = f(a) \cap f(b)$  as desired.

*ad* $\vee$ : The argument that  $f(a \vee b) = f(a) \cup f(b)$  is exactly parallel using (2') (or alternately this can be skipped using the fact that  $a \vee b = \neg(\neg a \wedge \neg b)$ ).

*ad* $\neg$ :  $P \in f(\neg a) \Leftrightarrow \neg a \in P \Leftrightarrow a \in \neg P \Leftrightarrow a \notin g(P) \Leftrightarrow g(P) \notin f(a) \Leftrightarrow P \notin g[f(a)] \Leftrightarrow P \in U - g[f(a)]$ .

We shall now discuss a second representation. Let  $U$  be a non-empty set and let  $R$  be a ring of subsets of  $U$  (closed under intersection and union, but not necessarily under complement, quasi-complement, etc.). by a *polarity* in  $R$  we mean an ordered pair  $X = (X_1, X_2)$  such that  $X_1, X_2 \in R$ . We define a relation and operations as follows, given polarities  $X = (X_1, X_2)$  and  $Y = (Y_1, Y_2)$ :

$$X \leq Y \Leftrightarrow X_1 \subseteq Y_1 \text{ and } Y_2 \subseteq X_2$$

$$X \wedge Y = (X_1 \cap Y_1, X_2 \cup Y_2)$$

$$X \vee Y = (X_1 \cup Y_1, X_2 \cap Y_2)$$

$$\neg X = (X_2, X_1).$$

By a field of polarities we mean a structure  $(P(R), \leq, \wedge, \vee, \neg)$  where  $P(R)$  is the set of all polarities in some ring of sets  $R$ , and the other components are defined as above. We leave to the reader the easy verification that every field of polarities is a de Morgan lattice. ■



We shall prove the following

**POLARITIES THEOREM** [Dunn, 1966]. *Every de Morgan lattice is isomorphic to a field of polarities.*

**Proof.** Given the previous representation, it clearly suffices to show that every quasi-field of sets is isomorphic to a field of polarities.

The idea is to set  $f(X) = (X, U - g[X])$ . Clearly  $f$  is one-one. We check that it preserves operations.

$$\begin{aligned} ad\wedge: f(X \cap Y) &= (X \cap Y, U - g[X \cap Y]) = (X \cap Y, (U - g[X]) \cap (U - g[Y])) = \\ &= (X, U - g[X]) \wedge (Y, U - g[Y]) = f(X) \wedge f(Y). \end{aligned}$$

$ad\vee$ : Similar.

$$\begin{aligned} ad\neg: f(\neg X) &= (\neg X, U - g(\neg X)) = (U - g[X], U - g(U - g[X])) = (U - \\ &= (U - g[X], X) = \neg f(X). \end{aligned}$$

■

We now discuss informal interpretations of the representation theorems that relate to semantical treatments of relevant first-degree implications familiar in the literature.

Routley and Routley [1972] presented a semantics for  $\mathbf{R}_{fde}$ , the main ingredients of which were a set  $K$  of ‘atomic set-ups’ (to be explained) on which was defined an involution  $*$ . An ‘atomic set-up’ is just a set of propositional variables, and it is used to determine inductively when complex formulas are also ‘in’ a given set-up. A set-up is explained informally as being like a possible world except that it is not required to be either consistent or complete. The Routley’s [1972] paper seems to conceive of set-ups very syntactically as literally being sets of formulas, but the Routley and Meyer [1973] paper conceives of them more abstractly. We shall think of them this latter way here so as to simplify exposition. The Routleys’ models can then be considered a structure  $(K, *, \vDash)$ , where  $K$  is a non-empty set,  $*$  is an involution on  $K$ , and  $\vDash$  is a relation from  $K$  to zero-degree formulas. We read ‘ $a \vDash A$ ’ as the formula  $A$  holds at the set-up  $a$ :

1.  $(\wedge \vDash) \quad a \vDash A \wedge B \Leftrightarrow a \vDash A \text{ and } a \vDash B$
2.  $(\vee \vDash) \quad a \vDash A \vee B \Leftrightarrow a \vDash A \text{ or } a \vDash B$
3.  $(\neg \vDash) \quad a \vDash \neg A \Leftrightarrow \text{not } a^* \vDash A.$

The connection of the Routleys’ semantics with quasi-fields of sets will become clear if we let  $(K, *)$  induce a quasi-field of sets  $Q$  with quasi-complement  $\neg$ , and let  $\vDash$  interpret sentences in  $Q$  subject to the following conditions:

$$1'. \quad |\wedge| \quad |A \wedge B| = |A| \cap |B|$$

$$2'. \quad |\vee| \quad |A \vee B| = |A| \cup |B|$$

$$3'. \quad |\neg| \quad |\neg A| = \neg|A|.$$

Clause  $(\wedge \vDash)$  results from clause  $|\wedge|$  by translating  $a \in |X|$  as  $a \vDash X$  (cf. Section 3.2). Thus clause  $|\wedge|$  says

$$a \in |A \wedge B| \Leftrightarrow a \in |A| \text{ and } a \in |B|,$$

i.e. it translates as clause  $(\wedge \vDash)$ . The case of disjunction is obviously the same. The case of negation is clearly of special interest, so we write it out.

Thus clause  $|\neg|$  says

$$\begin{aligned} a \in |\neg A| &\Leftrightarrow a \in \neg|A|, \\ &\Leftrightarrow a \in K - |A|^*, \\ &\Leftrightarrow a \notin |A|^*, \\ &\Leftrightarrow a^* \notin |A|. \end{aligned}$$

But the translation of this last is just clause  $(\neg \vDash)$ .

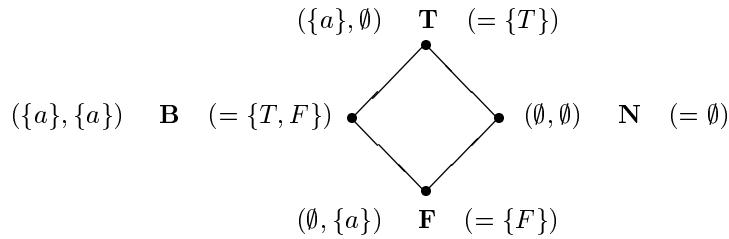
Of course the translation works both ways, so that the Routleys' semantics is just an interpretation in the quasi-fields of sets of Białynicki-Birula and Rasiowa written in different notation. Incidentally soundness and completeness of  $\mathbf{R}_{\text{fde}}$  relative to the Routleys' semantics follows immediately *via* the translation above from the corresponding theorem of the previous section *vis à vis* de Morgan lattices together with their representation as quasi-fields of sets. Of course the Routleys' conceived their results and derived them independently from the representation of Białynicki-Birula and Rasiowa.

We will not say very much here about what intuitive sense (if any) can be attached to the Routleys' use of the  $*$ -operator in their valational clause for negation. Indeed this question has had little extended discussion in the literature (though see [Meyer, 1979a; Copeland, 1979]). The Routleys' [1972] paper more or less just springs it on the reader, which led Dunn in [Dunn, 1976a] to describe the switching of  $a$  with  $a^*$  as 'a feat of prestidigitation'. Routley and Meyer [1973] contains a memorable story about how  $a^*$  'weakly asserts', i.e. fails to deny, precisely what  $a$  asserts, but one somehow feels that this makes the negation clause vaguely circular. Still, semantics often gives one this feeling and maybe it is just a question of degree. One way of thinking of  $a$  and  $a^*$  is to regard them as 'mirror images' of one another reversing 'in' and 'out'. Where one is inconsistent (containing both  $A$  and  $\neg A$ ), the other is incomplete (lacking both  $A$  and  $\neg A$ ), and *vice versa* (when  $a = a^*$ ,  $a$  is both consistent and complete and we have a situation appropriate to classical logic). Viewed this way the Routleys' negation clause

makes sense, but it does require some anterior intuitions about inconsistent and incomplete set-ups. More about the interpretation of this clause will be discussed in Section 5.1.

Let us now discuss the philosophical interpretation(s) to be placed on the representation of de Morgan lattices as fields of polarities. In Dunn [1966; 1971] the favoured interpretation of a polarity  $(X_1, X_2)$  was as a ‘proposition surrogate’,  $X_1$  consisting of the ‘topics’ the proposition gives definite positive information about and,  $X_2$  of the topics the proposition gives definite negative information about. A valuation of a zero degree formula in a de Morgan lattice can be viewed after a representation of the elements of the lattice as polarities as an assignment of positive and negative content to the formula. The ‘mistake’ in the ‘classical’ Carnap/Bar-Hillel approach to content is to take the content of  $\neg A$  to be the set-theoretical complement of the content of  $A$  (relative to a given universe of discourse). In general there is no easy relation between the content of  $A$  and that of  $\neg A$ . They may overlap, they may not be exhaustive. Hence the need for the double-entry bookkeeping done by proposition surrogates (polarities). If  $A$  is interpreted as  $(X_1, X_2)$ ,  $\neg A$  gets interpreted as the interchanged  $(X_2, X_1)$ .

Another semantical interpretation of the same mathematics is to be found in Dunn [1969; 1976a]. There given a polarity  $X = (X_1, X_2)$ ,  $X_1$  is thought of as the set of situations in which  $X$  is true and  $X_2$  as the set of situations in which  $X$  is false. These situations are conceived of as maybe inconsistent and/or incomplete, and so again  $X_1$  and  $X_2$  need not be set-theoretic complements. This leads in the case when the set of situations being assessed is a singleton  $\{a\}$  to a rather simple idea. The field of polarities looks like this



We have taken the liberty of labelling the points so as to make clear the informal meaning. (Thus the top is a polarity that is simply true in  $a$  and the bottom is one that is simply false, but the left-hand one is both true and false, and the right-hand one is neither.) Note that the de Morgan complement takes fixed points on both  $B$  and  $N$ . This is of course our old friend 4, which we know to be characteristic for  $\mathbf{R}_{fde}$ .

This leads to the idea of an ‘ambi-valuation’ as an assignment to sentences of one of the four values  $\mathbf{T}$ ,  $\mathbf{F}$ ,  $\mathbf{B}$ ,  $\mathbf{N}$ , conceived either as primitive or realised

as sets of the usual two truth values as suggested by the labelling. On this latter plan we have the valuation clauses (with double entry bookkeeping):

$$\begin{aligned}
 (\wedge) \quad & T \in v(A \wedge B) \Leftrightarrow T \in v(A) \text{ and } T \in v(B), \\
 & F \in v(A \wedge B) \Leftrightarrow F \in v(A) \text{ or } F \in v(B), \\
 (\vee) \quad & T \in v(A \vee B) \Leftrightarrow T \in v(A) \text{ or } T \in v(B), \\
 & F \in v(A \vee B) \Leftrightarrow F \in v(A) \text{ and } F \in v(B), \\
 (\neg) \quad & T \in v(\neg A) \Leftrightarrow F \in v(A), \\
 & F \in v(\neg A) \Leftrightarrow T \in v(A).
 \end{aligned}$$

We stress here (as in [Dunn, 1976a]) that all this talk of something's being both true and false or neither is to be understood epistemically and not ontologically. One can have inconsistent and or incomplete assumptions, information, beliefs, etc. and this is what we are trying to model to see what follows from them in an interesting (relevant!) way. Belnap [1977b; 1977a] calls the elements of the lattice 'told values' to make just this point, and goes on to develop (making connections with Scott's continuous lattices) a theory of 'a useful four-valued logic' for 'how a computer should think' without letting minor inconsistencies in its data lead to terrible consequences.

Before we leave the semantics of first-degree relevant implications, we should mention the interesting semantics of van Fraassen [1969] (see also Anderson and Belnap [1975, Section 20.3.1] and van Fraassen [1973]), which also has a double-entry bookkeeping device. We will not mention details here, but we do think it is an interesting problem to try to give a representation of de Morgan lattices using van Fraassen's facts so as to try to bring it under the same umbrella as the other semantics we have discussed here.

### 3.5 The Algebra of $\mathbf{R}$

This section is going to be brief. Dunn has already explicated on this topic in Section 28.2 of [Anderson and Belnap, 1975] and the interested reader should consult that and then Meyer and Routley [1972] for information about how to algebraise related weaker systems and how to give set-theoretical representations.

*De Morgan monoids* are a class of algebras that are appropriate to  $\mathbf{R}$  in the sense that (i) the Lindenbaum algebra of  $\mathbf{R}$  is one of them and (ii) all  $\mathbf{R}$  theorems are valid in them ((ii) gives soundness, and of course (i) delivers completeness by way of the canonical valuation). In thinking about de Morgan monoids it is essential that  $\mathbf{R}$  be equipped with the sentential constant  $\mathbf{t}$ . Also it is nice to think of fusion ( $\circ$ ) as a primitive connective, with even perhaps  $\rightarrow$  defined ( $A \rightarrow B =_{\text{df}} \neg(A \circ \neg B)$ ) but this is not essential since in  $\mathbf{R}$  (but not the weaker relevance logics) fusion can be defined as  $A \circ B =_{\text{df}} \neg(A \rightarrow \neg B)$ .

A de Morgan monoid is a structure  $\mathcal{D} = (D, \wedge, \vee, \neg, \circ, e)$  where

- (I)  $(D, \wedge, \vee, \neg)$  is a de Morgan lattice,
- (II)  $(D, \circ, e)$  is an Abelian monoid, i.e.  $\circ$  is a commutative, associative binary operation on  $D$  with  $e$  its identity, i.e.  $e \in D$  and  $e \circ a = a$  for all  $a \in D$ ,
- (III) the monoid is ordered by the lattice, i.e.  $a \circ (b \vee c) = (a \circ b) \vee (a \circ c)$ ,
- (IV)  $\circ$  is upper semi-idempotent ('square increasing'), i.e.  $a \leq a \circ a$ ,
- (V)  $a \circ b \leq c$  iff  $a \circ \neg c \leq \neg b$  (Antilogism).

De Morgan monoids were first studied in [Dunn, 1966] (although [Meyer, 1966] already had isolated some of the key structural features of fusion that they abstract). They also were described in [Meyer *et al.*, 1974] and used in showing  $\gamma$  admissible. Similar structures were investigated quite independently by Maksimova [1967; 1971].

The key trick in relating de Morgan monoids to  $\mathbf{R}$  is that they are residuated, i.e. there is a 'residual' operation  $\rightarrow$  so that

- (VI)  $a \circ b \leq c$  iff  $a \leq b \rightarrow c$ .

Indeed this operation turns out to be  $\neg(b \circ \neg c)$  (with the weaker systems or with positive  $\mathbf{R}$  it is important to postulate this law of the residual). Thus

- (1)  $a \circ b \leq c \Leftrightarrow b \circ a \leq c$       Commutativity
- (2)  $a \circ b \leq c \Leftrightarrow b \circ \neg a$       1, (V)
- (3)  $a \circ b \leq c \Leftrightarrow a \leq \neg(b \circ \neg c)$     2, de Morgan lattice.

As an illustration of the power of (VI) we show how the algebraic analogue of the Prefixing axiom follows from Associativity. First note that one can get from (III) the law of

- (Monotony)  $a \leq b \Rightarrow c \circ a \leq c \circ b$ .

Now getting down to Prefixing:

1.  $a \rightarrow b \leq a \rightarrow b$
2.  $(a \rightarrow b) \circ a \leq b$  1, (VI)
3.  $(c \rightarrow a) \circ c \leq a$  2, Substitution
4.  $(a \rightarrow b) \circ ((c \rightarrow a) \circ c) \leq b$  2,3, Monotony
5.  $((a \rightarrow b) \circ (c \rightarrow a)) \circ c \leq b$  4, Associatively
6.  $(a \rightarrow b) \circ (c \rightarrow a) \leq c \rightarrow b$  5, (VI)
7.  $a \rightarrow b \leq ((c \rightarrow a) \rightarrow (c \rightarrow b))$  6, (VI).

Incidentally, something better be said at this point about how validity in de Morgan monoids is defined. Unlike the case with  $\mathbf{R}_{fde}$ , there are theorems which are of the form  $A \rightarrow B$ , e.g.  $A \vee \neg A$ . We need some way of defining validity which is broader than insisting that always  $v(A) \leq v(B)$ . The identity  $e$  interprets the sentential constant  $t$ . By virtue of the  $\mathbf{R}$  axiom  $A \leftrightarrow (t \rightarrow A)$  characterising  $t$ , it makes sense to count all de Morgan monoid elements  $a$  such that  $e \leq a$  as ‘designated’, and to define  $A$  as valid iff  $v(A) \geq e$  for all valuations in all de Morgan monoids. We have the following law

$$a \leq b \Leftrightarrow e \leq a \rightarrow b,$$

which follows immediately from (VI) and the fact that  $e$  is the identity element. This means that (7) just above can be transformed into

$$e \leq (a \rightarrow b) \rightarrow ((c \rightarrow a) \rightarrow (c \rightarrow b))$$

validating prefixing as promised.

Other axioms of  $\mathbf{R}$  can be validated by similar moves. Commutativity validates Assertion, that  $e$  is the identity validates self-implication, square-increasingness validates Contraction, antilogism validates Contraposition, and the other axioms fall out of de Morgan lattice properties with lattice ordering and the residual law pitching in.

We shall not here investigate the ‘converse’ questions about how the fusion connective in  $\mathbf{R}$  is associative, etc. (that the Lindenbaum algebra of  $\mathbf{R}$  is indeed a de Morgan monoid (cf. Dunn’s Section 28.2.2 of [Anderson and Belnap, 1975])), but the proof is by ‘fiddling’ with contraposition being the key move.

Not as much is known about the algebraic properties of de Morgan monoids as one would like. Getting technical for a moment and using unexplained but standard terminology from universal algebra, it is known that de Morgan monoids are equationally definable (replace (V) with  $a \circ \neg(a \circ \neg b) \leq b$ , which can be replaced by the equation  $(a \circ \neg(a \circ \neg b)) \vee b = b$ ). So by a theorem of Birkhoff the class of de Morgan monoids is closed under sub-algebras, homomorphic images, and subdirect products. Further, given a de Morgan monoid  $\mathcal{D}$  with a prime filter  $P$  with  $e \in P$ , the relation  $a \approx b \Leftrightarrow (a \rightarrow b) \wedge (b \rightarrow a) \in P$  is a congruence, and the quotient algebra  $\mathcal{D}/\approx$  is subdirectly irreducible, and every de Morgan monoid is a subdirect product of such. It would be nice to have some independent interesting characterisation of the subdirectly irreducibles.

One significant recent result about the algebra of  $\mathbf{R}$  has been provided by John Slaney. He has shown that there are exactly 3088 elements in the free De Morgan monoid generated by the identity  $e$ . Or equivalently, in the language of  $\mathbf{R}$  including the constant  $t$ , there are exactly 3088 non-equivalent formulae free of propositional variables. The proof technique is

quite subtle, as generating a large algebra of 3088 elements is not feasible, even with computer assistance. Instead, Slaney attacked the problem using a “divide and conquer” technique [Slaney, 1985]. Since  $\mathbf{R}$  contains all formulae of the form  $A \vee \neg A$ , for any  $A$ , whenever  $L$  is a logic extending  $\mathbf{R}$ ,  $L = (L + A) \cap (L + \neg A)$ , where  $L + A$  is the result of adding  $A$  as an axiom to  $L$  and closing under modus ponens and adjunction. Given this simple result, we can proceed as follows.  $\mathbf{R}$  is  $(\mathbf{R} + f \rightarrow t) \cap (\mathbf{R} + \neg(f \rightarrow t))$ . Now it is not difficult to show that the algebra of  $\mathbf{R} + (f \rightarrow t)$  generated by  $t$  is the two element boolean algebra. Then you can restrict your attention to the algebra generated by  $t$  in the logic  $\mathbf{R} + \neg(f \rightarrow t)$ . If this has some characteristic algebra, then you can be sure that the elements freely generated by  $t$  in  $\mathbf{R}$  are bounded above by the number of elements in the direct product of the two algebras. To get the characteristic algebra of  $\mathbf{R} + \neg(f \rightarrow t)$ , Slaney goes on to divide and conquer again. He ends up considering six matrices, characterising six different extensions of  $\mathbf{R}$ . This would give him an upper bound on the number of constants (the matrices were size 2, 4, 6, 10, 10 and 14, so the bound was their product, 67200, well above 3088). Then you have to consider how many of these elements are generated by the identity in the direct product algebra. A reasonably direct argument shows that there are exactly 3088 elements generated in this way, so the result is proved.

### 3.6 The Operational Semantics (Urquhart)

This set-theoretical semantics is based upon an idea that occurred independently to Urquhart and Routley in the very late 1960s and early 70s. We shall discuss Routley’s contribution (as perfected by Meyer) in the next section and also just mention some related independent work of [Fine, 1974]. Here we concentrate upon the version of Urquhart [1972c] (cf. also Urquhart [1972b; 1972a; 1972d]).

Common to all the versions is the idea that one has some set  $K$  whose elements are ‘pieces of information’, and that there is a binary operation  $\circ$  on  $K$  that combines pieces of information. Also there is an ‘empty piece of information’  $0 \in K$ . We shall write  $x \vDash A$  to mean intuitively ‘ $A$  holds according to the piece of information  $x$ ’. The whole point of the semantics is disclosed in the valuational clause

$$(\rightarrow) \quad x \vDash A \rightarrow B \text{ iff } \forall y \in K \text{ (if } y \vDash A, \text{ then } x \circ y \vDash B).$$

The idea of the clause from left-to-right is immediately clear: if  $A \rightarrow B$  is given by the information  $x$ , then if  $A$  is given by  $y$ , then the combined piece of information  $x \circ y$  ought to give  $B$  (by *modus ponens*). The idea of the clause from right-to-left is to say that if this happens for all pieces of information  $y$ , this can only be because  $x$  gives us the information that  $A \rightarrow B$ .

Perhaps saying the whole point of the semantics is given in the clause  $(\rightarrow)$  along is an exaggeration. There are at least two quick surprises. The first is that we do not require (or want) a certain condition analogised from a condition required by Kripke's (relational) semantics for intuitionistic logic:

(The Hereditary Condition) If  $x \vDash A$ , then  $x \circ y \vDash A$ .

This would yield that if  $x \vDash A$ , then  $x \vDash B \rightarrow A$ , i.e. if  $y \vDash B$ , then  $x \circ y \vDash A$ . This would quickly involve us in irrelevance.

The other surprise is related to the failure of the Hereditary Condition: Validity cannot be defined as a formula's holding at *all* pieces of information in all models, since even  $A \rightarrow A$  would not then turn out to be valid. Thus  $x \vDash A \rightarrow A$  requires that if  $y \vDash A$  then  $x \circ y \vDash A$ . But this last is just a commuted form of the rejected Hereditary Condition, and there is no more reason to think it holds. We shall see in a moment that the appropriate definition of validity is to require that  $0 \vDash A$  for the empty piece of information in all models.

Enough talk of what properties  $\circ$  does not have! What property does it have? We have just been flirting with one of them. Clearly  $0 \vDash A \rightarrow A$  requires that if  $x \vDash A$  then  $0 \circ x \vDash A$ , and how more naturally would that be obtained than requiring that  $0$  be a (left) identity?

$$0 \circ x = x. \quad (\text{Identity})$$

This then seems the minimal algebraic condition on a model. Urquhart in fact requires others, all naturally motivated by the idea that  $\circ$  is the 'union' of pieces of information.

$$\begin{aligned} x \circ y &= y \circ x && (\text{Commutativity}) \\ x \circ (y \circ z) &= (x \circ y) \circ z && (\text{Associativity}) \\ x \circ x &= x. && (\text{Idempotence}) \end{aligned}$$

These conditions combined may be expressed by saying that  $(K, \circ, 0)$  is a '(join) semi-lattice with least element  $0$ ', and accordingly Urquhart's semantics is often referred to as the 'semi-lattice semantics'. It is well-known that every semi-lattice is isomorphic to a collection of sets with union as  $\circ$  and the empty set as  $0$  (map  $x$  to  $\{y : x \circ y = y\}$  so that henceforward  $\circ$  will be denoted by  $\cup$ ).

Each of the conditions above of course corresponds to an axiom of  $\mathbf{R}_{\rightarrow}$  when it is nicely axiomatised. Thus commutativity plays a natural role in verifying the validity of assertion. The following use of natural deduction



in the metalanguage makes this point nicely (we write ‘ $A, x$ ’ rather than  $x \vDash A$  for a notational analogy):

- |    |   |                         |
|----|---|-------------------------|
| 1. | $A, x$  | Hypothesis              |
| 2. | $A \rightarrow B, z$                                  | Hypothesis              |
| 3. | $B, x \cup z$   | 1, 2, ( $\rightarrow$ ) |
| 4. | $B, z \cup x$   | 3, Comm.                |
| 5. | $(A \rightarrow B) \rightarrow B, x$                  | 2, 4( $\rightarrow$ )   |
| 6. | $(A \rightarrow B) \rightarrow B, 0 \cup x$           | 5, Identity             |
| 7. | $A \rightarrow ((A \rightarrow B) \rightarrow B), 0.$ |                         |

The reader may find it amusing to write out an analogous pair of proofs for Prefixing, seeing how Associativity of  $\cup$  enters in, and for Contraction watching the Idempotence.<sup>24</sup>

The game has now been given away. There is some fiddling to be sure in proving a completeness theorem for  $\mathbf{R}_{\rightarrow}$  *re* the semi-lattice semantics, but basically the idea is that the semi-lattice semantics is just the system  $F\mathbf{R}_{\rightarrow}$  ‘written in the metalanguage’.

There is not a problem in extending the semi-lattice semantics so as to accommodate conjunction. The clause

$$x \vDash A \wedge B \text{ iff } x \vDash A \text{ and } x \vDash B \tag{\wedge}$$

does nicely. Somewhat strangely, the ‘dual’ clause

$$x \vDash A \vee B \text{ iff } x \vDash A \text{ of } x \vDash B \tag{\vee}$$

causes trouble. It is analogous to having the rule of  $\vee$ -Elimination  $N\mathbf{R}$  read:

$$\frac{\begin{array}{c} A \vee B, x \\ A, x \quad \text{Hyp.} \\ \vdots \\ c, x \cup y \\ B, x \quad \text{Hyp.} \\ \vdots \\ C, x \cup y \end{array}}{C, x \cup y.}$$

With this rule we can prove

$$\text{\#} \quad (A \rightarrow B \vee C) \wedge (B \rightarrow C) \rightarrow (A \rightarrow C),$$

---

<sup>24</sup>Though unfortunately verification of this last does not depend purely on Idempotence, but rather on  $(xy)y = xy$ , which of course is equivalent to Idempotence given Associativity and Identity. The verification of the formula  $A \wedge (A \rightarrow B) \rightarrow B$  ‘exactly’ uses Idempotence, but of course this is hardly a formula of the *implicational* fragment.

which is not a theorem of  $\mathbf{R}$  (see [Urquhart, 1972c]—the observation is Meyer’s and Dunn’s.)<sup>25</sup> And of course one can analogously verify that it is valid in the semi-lattice semantics.

Note that the condition (V) is not nearly as intuitive as the condition ( $\wedge$ ). The condition ( $\wedge$ ) is plausible for any piece of information  $x$ , at least if the relation  $x \vDash c$  does not require that  $C$  be *explicitly* contained in  $x$ . On the other hand the condition (V) is much less than natural. Does not it happen all the time that a piece of information  $x$  determines  $A \vee B$  to hold, without saying which? Is not this one of the whole points of disjunctions? Pieces of information  $x$  that satisfy (V) might be called ‘prime’ (in analogy with this epithet applied to theories of Section 2.4), and they have a kind of completeness or effeminateness that is rare in ordinary pieces of information. This by itself counts as no criticism of the semantics, since it is quite usual in semantical treatments to work with such idealised notions.

The condition (V) is not really as ‘dual’ to the condition ( $\wedge$ ) as one might think. Thus the formula

$$(\#d) \quad (B \wedge C \rightarrow A) \wedge (C \rightarrow B) \rightarrow (C \rightarrow A),$$

which is the dual of ( $\#$ ) is easily seen not to be valid in the semantics. This seems to be connected with another feature (problem?) of the semantics, to wit, no one has ever figured out how to add a natural semantical treatment of classical negation to the semantics (although it is straightforward to add a species of constructive negation—see [Urquhart, 1972c]).<sup>26</sup> The point of the connection is that ( $\#d$ ) would follow from ( $\#$ ) given classical contraposition principles, and yet the first is valid and the second one invalid in the positive semantics. So something about the positive semantics would have to be changed as well to accommodate negation.

The semi-lattice semantics has been extensively investigated in Charlewood [1978; 1981]. He fits it out with (two) natural deduction systems one with subscripts and one without. This last is in fact the (positive) system of Prawitz [1965], which Prawitz wrongly conjectured to be the same as Anderson and Belnap’s. Charlewood proves normalisation theorems (something that was anticipated by Prawitz for his system—incidentally the problem of normalisation for the Anderson–Belnap  $\mathbf{R}$  seems still open). Incidentally, one advantage of these natural deduction systems is that, unlike the Anderson–Belnap one for their system  $\mathbf{R}$  (cf. Section 1.5), they allow for a proof of distribution.

Charlewood also carries out in detail the engineering needed to implement K. Fine’s axiomatisation of the semi-lattice semantics. What is needed is

<sup>25</sup>It would be with  $C \rightarrow C$  as an additional conjunct in the antecedent.

<sup>26</sup>Charlewood and Daniels have investigated a combination of the semi-lattice semantics for the positive connectives and a four-valued treatment of negation in the style of [Dunn, 1976a]. they avoid the problem just described by in effect building into their definition of a model that it must satisfy classical contraposition. This does not seem to be natural.

to add to the Anderson–Belnap’s  $\mathbf{R}^+$  the following rule:

**R1:** From  $B_0 \wedge ((A_1 \wedge q_1, \dots, q_n \wedge A_n) \rightarrow X) \rightarrow ((B_1 \wedge q_1, \dots, B_n \wedge q_n) \rightarrow E)$  for  $X = B, C$ , and  $n \geq 0$  infer the same thing with  $B \vee C$  put in place of the displayed  $X$ , provided that the  $q_i$  are distinct and occur only where shown.

We forbear taking cheap shots at such an ungainly rule, the true elegance of which is hidden in the details of the completeness proof that we shall not be looking into. Obviously Anderson and Belnap’s  $\mathbf{R}$  is to be preferred when the issue is simplicity of Hilbert-style axiomatisations.<sup>27</sup>

### 3.7 The Relational Semantics (Routley and Meyer)

As was indicated in the last section, Routley too had the basic idea of the operational semantics at about the same time as Urquhart. Priority would be very hard to assess. At any rate Dunn first got details concerning both their work in early 1971, although J. Garson told him of Urquhart’s work in December of 1970 and he has seen references made to a typescript of Routley’s with a 1970 date on it (in [Charlewood, 1978]).

Meyer and Dunn were colleagues at the time, and Routley sent Meyer a somewhat incomplete draft of his ideas in early 1971. This was a courageous and open communication in response to our keen interest in the topic (instead he might have sat on it until it was perfected). The draft favoured the operational semantics, indeed the semi-lattice semantics, and was not clear that this was not the way to go to get Anderson and Belnap’s  $\mathbf{R}$ . But the draft started with a more general point of view suggesting the use of a 3-placed accessibility relation  $Rxyz$  (of course a 2-placed operation like  $\cup$  is a 3-placed relation, but not always conversely), with the following valuation clause for  $\rightarrow$ :

$$(\rightarrow) \quad x \vDash A \rightarrow B \text{ iff } \forall y, z \in K \text{ (if } Rxyz \text{ and } y \vDash A, \text{ then } z \vDash B).$$

Forgetting negation for the moment, the clauses for  $\wedge$  and  $\vee$  are ‘truth functional’, just as for the operational semantics.

Meyer, having observed with Dunn the lack of fit between the semi-lattice semantics and  $\mathbf{R}$ , was all primed to make important contributions to Routley’s suggestion. In particular he saw that the more general 3-placed relation approach could be made to work for all of  $\mathbf{R}$ . In interpreting  $Rxyz$  perhaps the best reading is to say that the combination of the pieces of information  $x$  and  $y$  (not necessarily the union) is a piece of information in  $z$  (in bastard symbols,  $x \circ y \leq z$ ). Routley himself called the  $x, y$ , etc.

<sup>27</sup>However, the semi-lattice semantics has been taken up and generalised in the field of substructural logics in the work of [Došen, 1988; Došen, 1989] and [Wansing, 1993].

‘set-ups’, and conceived of them as being something like possible worlds except that they were allowed to be inconsistent and incomplete (but always prime). On this reading  $Rxyz$  can be regarded as saying that  $x$  and  $y$  are compatible according to  $z$ , or some such thing.

Before going on we want to advertise some work that we are not going to discuss in any detail at all because of space limitations. The work of Fine [1974] independently covers some of the same ground as the Routley–Meyer papers, with great virtuosity making clear how to vary the central ideas for various purposes. The book of Gabbay [1976, see chapter 15] is also deserving of mention.

We now set out in more formal detail a version of the Routley–Meyer semantics for  $\mathbf{R}^+$  (negation will be reserved for the next section). The techniques are novel and the completeness proof quite complicated, so we shall be reasonably explicit about details. The presentation here is very much indebted to work (some unpublished) of Routley, Meyer and Belnap.

By an  $(\mathbf{R}^+)$  frame (or model structure) is meant a structure  $(K, R, 0)$ , where  $K$  is a non-empty set (the elements of which are called *set-ups*),  $R$  is a 3-placed relation on  $K, 0 \in K$ , all subject to some conditions we shall state after a few definitions. We define for  $a, b \in K, a \leq b$  (Routley and Meyer used  $>$ ) iff  $R0ab$ , and  $R^2abcd$  iff  $\exists x (Rabx \text{ and } Rxcd)$ . We also write this last as  $R^2(ab)cd$  and distinguish it from  $R^2a(bc)d =_{\text{df}} \exists x (Raxd \wedge Rbcx)$ . The variables  $a, b$ , etc. will be understood as ranging over the elements of some  $K$  fixed by the content of discussion.

Transcribing the conditions on the semi-lattice semantics as closely as we can into this framework we get the requirements

1. (Identity)  $R0aa$ ,
2. (Commutativity)  $Rabc \Rightarrow Rbac$ ,
3. (Associativity)  $R^2(ab)cd \Rightarrow R^2a(bc)d$ ,<sup>28</sup>
4. (Idempotence)  $Raaa$ .

It should be remarked that these conditions fail to pick up the whole strength of the corresponding semi-lattice conditions. Thus, e.g. Identity here only picks up  $0 \cdot a \leq a$  and not conversely, and similarly for Idempotence (also of course Commutativity and Associativity do not require any identity, but this is a slightly different point). We need for technical reasons one more condition:

5. (Monotony)  $Rabc$  and  $a' \leq a \Rightarrow Ra'bc$ .

---

<sup>28</sup>In the original equivalent conditions of Routley and Meyer [1973] this was instead ‘Pasch’s Law’:  $R^2abcd \Rightarrow R^2acbd$ . Also Monotony (condition (5) below) was misprinted there.

By a model we mean a structure  $M = (K, R, 0, \vDash)$ , where  $(K, R, 0)$  is a frame and  $\vDash$  is a relation from  $K$  to sentences of  $\mathbf{R}^+$  satisfying the following conditions:

- (1) (Atomic Hereditary Condition). For a propositional variable  $p$ , if  $a \vDash p$  and  $a \leq b$ , then  $b \vDash p$ .
- (2) (Valuational Clauses). For formulas  $A, B$ 
  - ( $\rightarrow$ )  $a \vDash A \rightarrow B$  iff  $\forall b, c \in K$  (if  $Rabc$  and  $b \vDash A$ , then  $c \vDash B$ );
  - ( $\wedge$ )  $a \vDash A \wedge B$  iff  $a \vDash A$  and  $a \vDash B$ ;
  - ( $\vee$ )  $a \vDash A \vee B$  iff  $a \vDash A$  or  $a \vDash B$ .

We shall say that  $A$  is *verified* on  $M$  if  $0 \vDash A$ , and that  $A$  *entails*  $B$  on  $M$  if  $\forall a \in K$  (if  $a \vDash A$ , then  $a \vDash B$ ). We say that  $A$  is *valid* if  $A$  is verified on all models.

It is easy to prove by an induction on  $A$ , the following (note how Monotony enters in):

**HEREDITARY CONDITION.** For an arbitrary formula  $A$ , if  $a \vDash A$  and  $a \leq b$ , then  $b \vDash A$ .

**VERIFICATION LEMMA.** *If in a given model  $(K, R, 0, \vDash)$   $A$  entails  $B$  in the sense that for every  $a \in K$ ,  $a \vDash A$  only if  $a \vDash B$ , then  $A \rightarrow B$  is verified in the model, i.e.  $0 \vDash A \rightarrow B$ .*

**Proof.** suppose that  $R0ab$  and  $a \vDash A$ . By the hypothesis of the Lemma,  $a \vDash B$ , and by the Hereditary Condition,  $b \vDash B$ , as is required for  $0 \vDash A \rightarrow B$ . ■

We are now in a position to prove the

**SOUNDNESS THEOREM.** *If  $\vdash_{\mathbf{R}} A$ , then  $A$  is valid.*

**Proof.** Most of this will be left to the reader. We first show that the axioms of  $\mathbf{R}^+$  are valid. Since they are all of the form  $A \rightarrow B$  we can simplify matters a little by using the Verification Lemma. As an illustration we verify Assertion (the reader may wish to compare this to the corresponding verification *vis à vis* the semi-lattice semantics of the last section).

To show  $A \rightarrow [(A \rightarrow B) \rightarrow B]$  is valid, it suffices by the Verification Lemma to assume  $a \vDash A$  and show  $a \vDash (A \rightarrow B) \rightarrow B$ . For this last we assume  $Rabc$  and  $b \vDash (A \rightarrow B)$ , and show  $c \vDash B$ . By Commutativity,  $Rabc$ . By ( $\rightarrow$ ) since we have  $b \vDash (A \rightarrow B)$  and  $a \vDash A$ , we get  $c \vDash B$  as desired.

The verification of the implicational axioms of Self-Implication and Prefixing are equally routine, falling right out of the Verification Lemma and

Associativity for the relation  $R$ . Unfortunately the verification of Contraction is a bit contrived (cf. note 24 above), so we give it here.

To verify Contraction, we assume that (1)  $a \vDash A \rightarrow .A \rightarrow B$  and show  $a \vDash A \rightarrow B$ . To show this last we assume that (2)  $Rabc$  and (3)  $b \vDash A$ , and show  $c \vDash B$ . From (2) we get, by Commutativity,  $Rbac$ . But  $Rbbb$  holds by Idempotence. so we have  $R^2(bb)ac$ . By Associativity we get  $R^2b(ba)c$ , i.e. for some  $x$ , both (4)  $Rbxc$  and (5)  $Rbax$ . by Commutativity, from (5) we get  $Rabx$ . Using  $(\rightarrow)$ , we obtain from this, (1), and (3) that (6)  $x \vDash A \rightarrow B$ . by Commutativity from (4) we get  $Rxbc$ , and from this, (6), and (3) we at last get the desired  $c \vDash B$ .

Verification of the conjunction and disjunction axioms is routine and is safely left to the reader.

It only remains to be shown then that the rules *modus ponens* and adjunction preserve validity. Actually something stronger holds. It is easy to see that for any  $a \in K$  (not just 0), if  $a \vDash A \rightarrow B$  and  $a \vDash A$ , then  $a \vDash B$  (by virtue of  $Raaa$ ), and of course it follows immediately from  $(\wedge)$  that if  $a \vDash A$  and  $a \vDash B$ , then  $a \vDash A \wedge B$ . ■

We next go about the business of establishing the

COMPLETENESS THEOREM. *If  $A$  is valid, then  $\vDash_{\mathbf{R}^+} A$ .*

The main idea of the proof is similar to that of the by now well-known Henkin-style completeness proofs for modal logic. We suppose that no  $\vDash_{\mathbf{R}^+} A$  and construct a so-called ‘canonical model’, the set-up of which are certain prime theories (playing the role of the maximal theories of modal logic). The base set-up 0 is constructed as a regular theory (for the terminology ‘regular’, ‘prime’, etc. consult Section 2.4; of course everything is relativised to  $\mathbf{R}^+$ ). From this point on for simplicity we shall assume that we are dealing with  $\mathbf{R}^+$  outfitted with the optional extra fusion connective  $\circ$  and the propositional constant  $t$  (recall these can be conservatively added — cf. Section 1.3). We then define  $Rabc$  to hold precisely when for all formulas  $A$  and  $B$ , whenever  $A \in a$  and  $B \in b$ , then  $A \circ B \in c$ .<sup>29</sup>

Let us look now at the details. Pick 0 as some prime regular theory  $T$  with  $A \notin T$ . We can derive that at least one such exists using the Belnap Extension Lemma (it was stated in Section 2.5 for  $\mathbf{RQ}$ , but it clearly holds for  $\mathbf{R}^+$  as well). thus set  $\Delta = \mathbf{R}^+$  and  $\theta = \{A\}$ .

Define  $K =$  set of prime theories,<sup>30</sup> and define the accessibility relation  $R$  canonically as above.

<sup>29</sup>The use of  $\circ$  and  $t$  is a luxury to make things prettier at least at the level of description. Thus, e.g. as we shall see, the associativity of  $\mathbf{R}$  follows from the associativity of  $\circ$ , and other mnemonically pleasant things happen. We could avoid its use by defining  $Rabc$  to hold whenever if  $A \in a$  and  $A \rightarrow B \in b$ , then  $B \in c$ . Incidentally, the valational clause for fusion is :  $x \vDash A \circ B$  iff for some  $a, b$  such that  $Rabx$ ,  $a \vDash A$  and  $b \vDash B$ . The valational clause for  $t$  is  $x \vDash t$  iff  $0 \leq x$ .

<sup>30</sup>One actually has a choice here. We have required of theories that they be closed under implications provable in 0, i.e. require of  $T$  that whenever  $A \in T$  and  $A \rightarrow B \in 0$ ,

**THEOREM 4.** *The canonically defined structure  $(K, 0, R)$  is an  $\mathbf{R}^+$  frame.*

**LEMMA 5.** *The relation  $R$  defined canonically above satisfies Identity, Commutativity, Idempotence, and Associativity.*

**PROOF.**

*ad Identity.* We need to show that  $R0aa$ , i.e. if  $X \in 0$  and  $A \in a$ , then  $X \circ A \in a$ . By virtue of the  $\mathbf{R}$ -theorem  $A \rightarrow t \circ A$ , we have  $t \circ A \in a$ . But using the  $\mathbf{R}$ -theorem  $X \rightarrow .t \rightarrow x$ , we have  $t \rightarrow X \in 0$ . By Monotony we have  $X \circ A \in a$  as desired.

*ad Commutativity.* Suppose  $Rabc$ . We need show  $Rbac$ , i.e. if  $B \in b$  and  $A \in a$ , then  $B \circ A \in c$ . From  $Rabc$ , it follows that  $A \circ B \in c$ . But by virtue of the  $\mathbf{R}$ -theorem  $A \circ B \rightarrow B \circ A$  (commutativity of  $\circ$ ) we have  $B \circ A \in C$ , as desired.

*ad Idempotence.* We need show  $Raaa$ , i.e. if  $A \in a$  and  $B \in a$ , then  $A \circ B \in a$ . This follows from the  $\mathbf{R}$ -theorem  $A \wedge B \rightarrow A \circ B$ , which follows ultimately from the square increasingness of  $\circ$ ,  $(X \rightarrow X \circ X)$ , as the proof sketch below makes clear.

1.  $A \wedge B \rightarrow A$  Axiom
2.  $A \wedge B \rightarrow B$  Axiom
3.  $(A \wedge B) \circ (A \wedge B) \rightarrow A \circ B$  1, 2, Monotony
4.  $A \wedge B \rightarrow A \circ B$  3, square increasingness

*ad Associativity.* This is by far the least trivial property. Let us then assume that  $R^2(ab)cd$ , i.e.  $\exists x(Rabx$  and  $Rxcd)$ . We need then show that there is a prime theory  $y$  such that  $Rayd$  and  $Rbcy$ , i.e.  $R^2a(bc)d$ .

Set  $y_0 = \{Y : \exists B \in b, C \in c : \vdash_{\mathbf{R}} B \circ C \rightarrow Y\}$ . (This is sometimes referred to as  $b \circ c$ ). Clearly the definition of  $y_0$  assures that  $Rbcy_0$ .

Observe that  $y_0$  is a theory.<sup>31</sup> Thus it is clear that  $y_0$  is closed under provable  $\mathbf{R}$ -implication, since this is just transitivity. We show it is also closed under adjunction. Thus suppose for some  $B, B' \in b, C, C' \in c, \models_{\mathbf{R}} B \circ C \rightarrow Y$  and  $\vdash_{\mathbf{R}} B' \circ C' \rightarrow Y'$ . Then  $\vdash_{\mathbf{R}} (B \circ C) \wedge (B' \circ C') \rightarrow Y \wedge Y'$  using easy properties of conjunction. But we have the  $\mathbf{R}$ -theorem  $(B \wedge B') \circ (C \wedge C') \rightarrow (B \circ C) \wedge (B' \circ C')$  (which follows basically from the one-way distribution of  $\circ$  over  $\wedge$ ,  $X \circ (Y \wedge Z) \rightarrow (X \circ Y) \wedge (X \circ Z)$ , which

then  $B \in T$ . The latter is a stronger requirement and leads to the 'smaller' *reduced* models of [Routley *et al.*, 1982], which are useful for various purposes.

<sup>31</sup>The presentation of Routley–Meyer [1973] is more elegant than ours, developing as they do properties of what they call the calculus of 'intensional  $\mathbf{R}$ -theories', showing that it is a partially ordered (under inclusion) commutative monoid ( $\circ$  as defined above) with identity 0. Further  $\circ$  is monotonous with respect to  $\leq$ , i.e. if  $a \leq b$  then  $c \circ a \leq c \circ b$ , and  $\circ$  is square increasing, i.e.  $a \leq a \circ a$ . Then defining  $Rabc$  to mean  $a \circ b \leq c$ , the requisite properties of  $\mathbf{R}$  fall right out.

follows basically from Monotony,  $Y_1 \rightarrow Y_2 \rightarrow X \circ Y_1 \rightarrow X \circ Y_2$ , which is easy). So by transitivity we get  $\vdash_{\mathbf{R}} (B \wedge B') \circ (C \wedge C') \rightarrow Y \wedge Y'$ , from which it follows that  $Y \wedge Y' \in y_0$  as promised ( $B \wedge B' \in b, C \wedge C' \in c$  of course, since  $b, c$  are closed under adjunction).

We next verify that  $Ra_0d$ . Suppose that  $A \in a$  and  $Y \in y_0$ . Then for some  $B \in b, C \in c, \vdash_{\mathbf{R}} B \circ C \rightarrow Y$ . Since  $Rabx, A \circ B \in x$ . And since  $Rxcd(A \circ B) \circ C \in d$ . By the associativity of  $\circ$  (since  $d$  is a theory), then  $A \circ (B \circ C) \in d$ . but by Monotony, since  $\vdash_{\mathbf{R}} B \circ C \rightarrow Y$ , we have  $\vdash_{\mathbf{R}} A \circ (B \circ C) \rightarrow A \circ Y$ . Hence  $A \circ Y \in d$ , as needed.

The reader is excused if he has lost the thread a bit and thinks that we are now finished verifying the associativity of  $R$ . We wanted some prime theory  $y$  which fills in the blanks

1.  $Ra\_d$  and
2.  $Rbc\_$ ,

and we have just finished verifying that  $y_0$  is a theory that does fill in the blanks. The kicker is that  $y_0$  need not be prime. So we work next at pumping up  $y_0$  to make it prime while continuing to fill in the blanks.

It clearly suffices to prove

**THE SQUEEZE LEMMA.** *Let  $a_0$  and  $y_0$  be theories that need not be prime, and let  $d$  be a prime theory. If  $Ra_0y_0d$ , then there exists a prime theory  $y$  such that (i)  $y_0 \subseteq y$  and (ii)  $Ra_0yd$ .*

This can be accomplished by a Lindenbaum-style construction like that of Section 2.3 (or alternatively Zorn's Lemma may be used as in Routley and Meyer [1973]). The idea is to define  $y$  as the union of a sequence of sets of formulas  $y_n$ , where (relative to some fixed enumeration of the formulas)  $y_{n+1}$  is defined inductively as  $y_n \cup \{A_{n+1}\}$  if  $Ra(y_n \cup \{A_{n+1}\})d$ , and otherwise  $y_{n+1}$  is just  $y_n$ .

But it is instructive to crank the existence of the given  $y$  out of the Belnap Extension Lemma for  $\mathbf{R}$ .

Thus set  $\Delta = y_0$  and  $\theta = \{A : \exists B(A \rightarrow B) \in a \text{ and } B \notin d\}$ . We need check that  $(\Delta, \theta)$  is exclusive.

We observe first that  $\theta$  is closed under disjunction. Thus suppose  $A_1, A_2 \in \theta$ . Then for some  $B_1, B_2, A_1 \rightarrow B_1, A_2 \rightarrow B_2 \in a$ , and yet  $B_1, B_2 \notin d$ . Then (since  $d$  is prime)  $B_1 \vee B_2 \notin d$ . but since  $a$  is a theory, then  $A_1 \vee A_2 \rightarrow B_1 \vee B_2$  by an appropriate theorem of  $\mathbf{R}$  in the proximity of the disjunction axioms. So  $A_1 \vee A_2 \in \theta$  as desired. Since  $\Delta$  is closed dually under adjunction (that was the point of observing above that  $y_0$  is a theory), this means that if the pair  $(\Delta, \theta)$  fails to be exclusive, then for some  $X \in \Delta, A \in \theta, \vdash_{\mathbf{R}} X \rightarrow A$ . So for some  $B, A \rightarrow B \in a$  and  $B \notin d$ . But since  $a$  is a theory, by transitivity we derive that  $X \rightarrow B \in a$ . But since  $Raxd$



and  $X \in x$ , we get  $(X \rightarrow B) \circ X \in d$ . But since  $\vdash_{\mathbf{R}} X \circ (X \rightarrow B) \rightarrow B$ , we have  $B \in d$ , contrary to the choice of  $B$ .

Now that we know  $(\Delta, \Theta)$  is an exclusive pair we apply the Belnap Extension Lemma to get a pair  $(y, y')$  with  $y_0 = \Delta \subseteq y$  and  $y$  a prime theory, completing the proof of the Squeeze Lemma, which actually does complete the proof that the relation  $R$  is Associativity.

*ad* Monotony. (Yes, we still have something left to do.) Let us suppose that  $R0a'a$  and  $Rabc$ , and show  $Ra'bc$ . Note that it follows from  $R0a'a$  that  $a' \leq a$ ,<sup>32</sup> from which it follows at once from  $Rabc$  and  $Ra'bc$ . Thus if  $X \in a'$  then since  $X \rightarrow X \in 0$ , then  $(X \rightarrow X) \circ X \in a$ . But since  $\vdash_{\mathbf{R}^+} (X \rightarrow X) \circ X \rightarrow X$ , then  $X \in a$ .

Having now finally verified that the canonical  $(K, 0, R)$  has all the properties of an  $\mathbf{R}^+$ -frame, we need now to define an appropriate relation  $\vDash$  on it. The natural definition is  $a \vDash A$  iff  $A \in a$ , but we need now to verify that this has the properties (1) and (2) required of  $\vDash$  above.

**THEOREM 2.** *The canonically defined  $(K, 0, R, \vDash)$  is indeed an  $\mathbf{R}$ -model.*

**Proof.** *ad* (1) (the Hereditary Condition). Suppose  $a \leq b$ , i.e.  $R0ab$ . We show that  $a \leq b$ , from which the Hereditary Condition immediately follows. Suppose then that  $A \in a$ . Since  $t \in 0, t \circ A \in b$ . But *via* the  $\mathbf{R}$ -theorem  $t \circ A \rightarrow A$ , we have  $A \in b$  as desired.

*ad* (2) (the valuation of clauses). The clauses  $(\wedge)$  and  $(\vee)$  are more or less immediate (primeness is of course needed for half of  $(\vee)$ ). The clause of interest is  $(\rightarrow)$ . Applying the canonical definition of  $\vDash$ , this amounts to

$$(\rightarrow_c) \quad A \rightarrow B \in a \text{ iff } \forall b, c (\text{if } Rabc \text{ and } A \in b, \text{ then } B \in c).$$

Left-to-right is argued as follows. Suppose  $A \rightarrow B \in a, Rabc, A \in b$ , and show  $B \in c$ .  $Rabc$  of course means canonically that whenever  $X \in a$  and  $Y \in b$ , then  $X \circ Y \in c$ . Setting  $X = A \rightarrow B$  and  $Y = A$ , we get  $A \circ (A \rightarrow B) \rightarrow C$ . Then using the  $\mathbf{R}^+$ -theorem

$$A \circ (A \rightarrow B) \rightarrow B, \text{ we obtain } B \in c.$$

Right-to-left is harder, and in fact involves the third (and last) application of the Belnap Extension Lemma in the proof of Completeness. Thus suppose contrapositively that  $A \rightarrow B \notin a$ . We need to construct prime theories  $b$  and  $c$ , with  $A \in b$  and  $B \notin c$ . We let  $\Delta_b = \text{Th}(\{A\})$  and set  $\Delta_c = a \circ \Delta_b$ , i.e.  $\{Z : \exists X \in a, \exists Y \in \Delta_b \vdash_{\mathbf{R}^+} X \circ Y \rightarrow Z\}$ . This is the same as  $\{Z : \exists X \in a \vdash_{\mathbf{R}^+} X \circ A \rightarrow Z\}$ . We set  $\theta_c = \{B\}$ . Clearly  $(\Delta_c, \theta_c)$  is an exclusive pair, for otherwise  $\vdash_{\mathbf{R}^+} X \circ A \rightarrow B$ , i.e.  $\vdash_{\mathbf{R}^+} X \rightarrow (A \rightarrow B)$  for some  $X \in a$ , and so  $A \rightarrow B \in a$  contrary to our supposition. We apply Belnap's Extension Lemma to get an exclusive pair  $(c, c')$  with  $\Delta_c \subseteq c$  and  $c$  prime theory. Note

<sup>32</sup>In the 'reduced models (cf. note 46) one can show that  $R0a'a$  iff  $a' \leq a$ .

that by definition of  $\Delta_b$  and  $\Delta_c$ ,  $Ra\Delta_b\Delta_c$ , and so  $Ra\Delta_b c$ . We are now in a position to apply the Squeeze Lemma getting a prime theory  $b \supseteq \Delta_b$  such that  $Rabc$ . Clearly  $A \in b$ , but also  $B \notin c$  since  $B \in \theta_c \subseteq c'$  ( $c$  and  $c'$  are exclusive).

This at last completes the proof of the Completeness Theorem for  $\mathbf{R}^+$ . ■

REMARK. It is fashionable these days to always prove *strong* completeness. This could have been done. Thus define  $A$  to be a *logical consequence* of a set of formulas  $\Gamma$  iff for every  $\mathbf{R}^+$ -model  $M$ , if  $0 \models B$  for every  $B \in \Gamma$ , then  $0 \models A$ . This is a kind of classical notion and should not be confused with some kind of relevant consequence. Thus, e.g. where  $B$  is a theorem of  $\mathbf{R}^+$ , since always  $0 \models B$ ,  $B$  will be a logical consequence of any set  $\Gamma$ . Define  $B$  to be deducible from  $\Gamma$  (again in a neo-classical sense) to mean  $B \in \text{Th}(\Gamma \cup \mathbf{R}^+)$ . Appropriate modifications of the work above will show that logical consequence is equivalent to deducibility.

### 3.8 Adding Negation to $\mathbf{R}^+$

We now discuss the Routley–Meyer semantics for the whole system  $\mathbf{R}$ . The idea is simply to add the Routley’s treatment of negation using the  $*$ -operator (discussed in Section 3.4). (This is not difficult and there is very little reason to segregate it off into this separate section, except that we thought that the treatment of  $\mathbf{R}^+$  was complicated enough.)

Thus an  $\mathbf{R}$ -frame is a structure,  $(K, R, 0, *)$  where  $(K, R, 0)$  is an  $\mathbf{R}^+$ -frame and  $K$  is closed under the unary operation  $*$  satisfying:

$$\begin{aligned} \text{(Period two)} \quad & A^{**} = a, \\ \text{(Inversion)} \quad & Rabc \Rightarrow Rac^*b^* \end{aligned}$$

For an  $\mathbf{R}$ -model the valuations clauses for the positive connectives are as for an  $\mathbf{R}$ -model, and we of course add

$$(\neg) \quad a \models \neg A \text{ iff } a^* \not\models A.$$

The soundness and completeness results are relatively easy modifications of those for  $\mathbf{R}^+$ . That  $*$  is of period two naturally is used in the verification of Double Negation and Inversion is central to the verification of Contraposition. For completeness,  $a^*$  is defined canonically as  $\{A : \neg A \notin a\}$  (cf. the definition of the analogue  $g[P]$  in the proof of Białyński–Birula and Rasiowa’s representation of de Morgan lattices in Section 3.4), and one of course has to show that  $a^*$  is a prime theory when  $a$  is. One also has to show that canonical  $*$  is of period two and satisfies (Inversion), and that canonical  $\models$  satisfies  $(\neg)$  above, i.e.  $A \in a \Leftrightarrow \neg A \notin a^*$ , i.e.  $\neg A \notin \{B : \neg B \notin a\}$ , i.e.  $\neg\neg A \in a$ , which of course just uses Double Negation.

It is worth remarking that since the canonical  $0$  is a prime regular theory, then since  $\vdash_{\mathbf{R}} A \vee \neg A$ , then  $0$  is complete (but not necessarily consistent—this is relevant to the development in Section 3.9). For your garden variety Routley–Meyer model (not necessarily canonical) notice also that  $0 \vDash A$  or  $0 \vDash \neg A$ . This follows ultimately from  $0^* \leq 0$ , i.e.  $R00^*0$ , proven below.

1.  $R0^*0^*0^*$
2.  $R0^*00$       1, (Inversion), (Period two)
3.  $R00^*0$       2, (Commutation).

Now  $0^* \leq 0$  means by the Hereditary Condition that if  $0 \vDash A$  then  $0^* \vDash A$ , i.e.  $0 \vDash \neg A$  as desired.

It should be said that although either the four-valued treatment or the  $*$ -operator treatment of negation work equally well for *first-degree* relevant implications (at least from a technical point of view), the  $*$ -operator treatment seems to win hands down in the context of all of  $\mathbf{R}$ . Meyer [1979a] has succeeded in giving a four-valued treatment of all of  $\mathbf{R}$ , but at the price of great technical complexity (e.g. the accessibility relation has to be made four-valued as well, and that is just for starters). Further, as Meyer points out, one's models still have to be closed under  $*$ , so it still can be said to sneak in the back door.

### 3.9 Routley–Meyer Semantics for $\mathbf{E}$ and other Neighbours of $\mathbf{R}$

Once one sets down a set of conditions on an accessibility relation, they can be played with in various ways so as to produce semantics for a wide variety of systems as the experience with modal logic has taught us. Also other features of the frames can be generalised.

We can here only give the flavour of a whole range of possible and actual results. In all the results below  $\vDash$  will satisfy the same conditions as for  $\mathbf{R}^+$  (or  $\mathbf{R}$ ) models (as appropriate). To begin with we follow Routley and Meyer [1973] with the description of a series of conditions on positive frames and corresponding axioms for propositional logic. They begin by requiring of a  $\mathbf{B}^+$ -frame  $(K, R, 0)$

- B1.  $a \leq a$
- B2.  $a \leq b$  and  $b \leq x \Rightarrow a \leq x$
- B3.  $a' \leq a$  and  $Rabc \Rightarrow Ra'bc$ .

B1–B2 of course say that  $\leq$  is a quasi-order, and B3 says something like that it is monotone.

‘**B**’ appears to be for ‘Basic’, for they regard the above postulates as a natural minimal set on their approach.<sup>33</sup> Gabbay [1976] investigates even weaker logics where no conditions at all are placed on the frame, but these have no theorems and are characterised only by rules of deducibility (unless Boolean negation and/or the Boolean material conditional is present, options which he does explore).

The sense in which the above postulates are minimal goes something like this. B3 is needed in proving the Hereditary Condition for implications, and the Hereditary Condition is needed in turn for verifying  $0 \vDash A \rightarrow A$  (indeed anything) so we have at least some minimal theorems. The Hereditary Condition is used in showing the equivalence of the verification of an implication in a model and entailment in that mode, i.e.  $0 \vDash A \rightarrow B$  iff  $\forall x \in K(x \vDash A \Rightarrow x \vDash B)$  (cf. Section 3.7 to see how these conditions were used to establish these facts about  $\mathbf{R}^+$ -models). What about B2? We think it is just a ‘freebie’. It seems to play no role in verifying axioms or rules, but the completeness proof can be made to yield canonical (‘reduced’) models (cf. note 3.7) that satisfy it, so why not have it? This seems to be what Routley *et. al.* [1982] say. It appears that B1 is even more a freebie.

It may be shown that  $A$  is a theorem of the system  $\mathbf{B}^+$  (formulated in Section 1.3) iff  $A$  is valid in all  $\mathbf{B}^+$  models.

Routley and Meyer establish the following correspondence between conditions on the accessibility relation  $R$  and axioms:

(1)	$Raaa$	$A \wedge (A \rightarrow B) \rightarrow B$
(2)	$Rabc \Rightarrow R^2a(ab)c$	$(A \rightarrow B) \wedge (B \rightarrow C) \rightarrow (A \rightarrow C)$
(3)	$R^2abcd \Rightarrow R^2a(bc)d$	$A \rightarrow B \rightarrow ([B \rightarrow C] \rightarrow [A \rightarrow C])$
(4)	$R^2abcd \Rightarrow R^2b(ac)d$	$A \rightarrow B \rightarrow ([C \rightarrow A] \rightarrow [C \rightarrow B])$
(5)	$Rabc \Rightarrow R^2abc$	$(A \rightarrow [A \rightarrow B]) \rightarrow (A \rightarrow B)$
(6)	$Ra0a$	$([A \rightarrow A] \rightarrow B) \rightarrow B$
(7)	$Rabc \Rightarrow Rbac$	$A \rightarrow ([A \rightarrow B] \rightarrow B)$
(8)	$0 \leq a$	$A \rightarrow (B \rightarrow B)$
(9)	$Rabc \Rightarrow b \leq c$	$A \rightarrow (B \rightarrow A)$ .

Routley and Meyer connect these conditions on accessibility relations to axioms extending the basic logic  $\mathbf{B}$ . The correspondence is more perspicuous when you consider the *structural rules* corresponding to each axiom or condition. We can express these as conditions on fusion:

---

<sup>33</sup>However, some notational confusion is possible, with Fine’s use of ‘**B**’ as another basic relevance logic differing slightly from Routley and Meyer’s usage [Fine, 1974]. For Fine,  $\mathbf{B}$  includes the law of the excluded middle, and for Routley and Meyer, it does not.

- |     |                                 |  |
|-----|---------------------------------|--|
| (1) | $Raaa$                          | $A \vdash A \circ A$                             |
| (2) | $Rabc \Rightarrow R^2a(ab)c$    | $A \circ B \vdash A \circ (A \circ B)$           |
| (3) | $R^2abcd \Rightarrow R^2a(bc)d$ | $(A \circ B) \circ C \vdash A \circ (B \circ C)$ |
| (4) | $R^2abcd \Rightarrow R^2b(ac)d$ | $(A \circ B) \circ C \vdash B \circ (A \circ C)$ |
| (5) | $Rabc \Rightarrow R^2abbc$      | $A \circ B \vdash (A \circ B) \circ B$           |
| (6) | $Ra0a$                          | $A \circ t \vdash A$                             |
| (7) | $Rabc \Rightarrow Rbac$         | $A \circ B \vdash B \circ A$                     |
| (8) | $0 \leq a$                      | $B \circ A \vdash B$ (or $A \vdash t$ )          |
| (9) | $Rabc \Rightarrow b \leq c$     | $A \circ B \vdash B$ .                           |

General recipes for translating between structural rules and conditions on accessibility relations are to be found in Restall [1998; 2000].

If one wants to add to  $\mathbf{B}^+$  any of the axioms on the right to get a sentential logic  $\mathbf{X}$ , one merely adds the corresponding conditions to those for a  $\mathbf{B}^+$  model to get the appropriate notion of an  $\mathbf{X}$ -model, with a resultant sound and complete semantics.

Some logics of particular interest arising in this way are (nomenclature as in [Anderson and Belnap, 1975]) (note well that  $\mathbf{T}$  has nothing to do with Feys'  $\mathbf{t}$  of modal logic fame):

$$\begin{aligned}
\mathbf{TW}^+ &: \mathbf{B}^+ + (3, 4) \\
\mathbf{T}^+ &: \mathbf{TW}^+ + (5) \\
\mathbf{E}^+ &: \mathbf{T}^+ + (6) \\
\mathbf{R}^+ &: \mathbf{E}^+ + (7) \\
\mathbf{H}^+ &: \mathbf{R}^+ + (8) \\
\mathbf{S4}^+ &: \mathbf{E}^+ + (8).
\end{aligned}$$

These are far from the most elegant formulations from a postulational point of view, being highly redundant (in particular the Prefixing and Suffixing *rules* of  $\mathbf{B}^+$  are supplanted already in  $\mathbf{TW}^+$  by the corresponding *axioms*. further the rule of Necessitation ( $A \vdash (A \rightarrow A) \rightarrow A$ ) is also redundant already in  $\mathbf{TW}^+$  (this is not so obvious—proof is by browsing through [Anderson and Belnap, 1975]).

What minimal conditions should be imposed on the  $*$ -operator when it is added to a  $\mathbf{B}^+$ -frame so as to give a  $\mathbf{B}$ -frame? Routley *et. al.* [1982] choose

B4.  $a^{**} = a$ , and

B5.  $a \leq b \Rightarrow b^* \leq a^*$ .

The minimality of B5 can be defended in terms of its being needed for showing that negations satisfy the Hereditary Condition. B4 would seem to have little place in a *minimal* system except for the fact that the dominant trend in relevance logic has been to keep classical double negation.<sup>34</sup>

<sup>34</sup>In fact, B5 is too strong for a purely *minimal* logic of negation. See Section 5.1 for more discussion on this.

One can get semantics for the full systems **TW**, **T**, etc. simply by adding the appropriate postulates to the conditions on a **B**-model.

We could go on, but will instead refer the reader to Routley *et al.* [1982], Fine [1974] and Gabbay [1976] for a variety of variations producing systems in the neighbourhood of **R**.

Some find the conditions on the “base point” 0 on frames rather puzzling or unintuitive. Why should the basic conditions on frames include conditions such as the fact that  $a \leq b$  defined as  $R0ab$  generate a partial order? Some recent work by Priest and Sylvan and extended by Restall has shown that these conditions can be done away with and the frames given an interpretation rather reminiscent of that of non-normal modal logics [Priest and Sylvan, 1992; Restall, 1993]. The idea is as follows. We have two sorts of set-ups in a frame — normal ones and non-normal ones. Then we split the treatment of implication along this division. Normal points are given an **S5**-like interpretation.

- $x \vDash A \rightarrow B$  iff for every  $y$  if  $y \vDash A$  then  $y \vDash B$

and non-normal points are given the condition which appeals to the ternary relation  $R$

- $x \vDash A \rightarrow B$  iff for every  $y$  and  $z$  where  $Rxyz$  if  $y \vDash A$  then  $z \vDash B$

The other connectives are treated in just the same way as in the original relational semantics. To prove soundness and completeness for this semantics, it is simplest to go through the original semantics — for it is not too difficult to show that this account is merely a notational variant, where we have set  $Rxyz$  iff  $y = z$  when  $x$  is a normal set-up. This satisfies all of the conditions in the original semantics, for we have set  $a \leq b$  to be simply  $a = b$ .

We turn now to one such system **RM** deserving of special treatment.

### 3.10 Algebraic and Set-theoretical Semantics for **RM**

**RM** has been described by Meyer as ‘the laboratory of relevance logic’. It plays a role somewhat like **S5** among modal logics, being a place where conjectures can be tested relatively easily (e.g. the admissibility of  $\gamma$  was first shown for **RM**). This then could be a very long section because **RM** is by far the best understood of the Anderson–Belnap style systems. We shall try to keep it short by being dogmatic. The interested reader can verify the results claimed by consulting Meyer’s Section 29.3 and Section 29.4 of [Anderson and Belnap, 1975] (see also [Dunn, 1970; Tokarz, 1980]).

In the first place the appropriate algebras for **RM** are the *idempotent* de Morgan monoids (strengthening  $a \leq a \circ a$  to  $a = a \circ a$ ). The subdirectly irreducible ones are all chains with de Morgan complement where  $a \circ b = a \wedge b$

if  $a \leq \neg b$ , and  $a \circ b = a \vee b$  otherwise. The designated elements are all elements  $a$  such that  $\neg a \leq a$ , and of course these must have a greatest lower bound to serve as the identity  $e$ . (This is just another description with  $\circ$  as primitive instead of  $\rightarrow$  of the ‘Sugihara matrices’ described in the publications cited above.) Meyer showed that if  $\vdash_{\mathbf{RM}} A$ , then  $A$  is valid in all the finite Sugihara matrices, establishing the finite model property for **RM**.

Dunn showed that every extension of **RM** closed under substitution and the rules of **R** has some finite Sugihara matrix as a characteristic matrix (**RM** is ‘pretabular’). A similar result was shown by Scroggs to hold for the modal logic **S5**, and researchers (particularly Maksimova) have obtained results characterising all such pretabular extensions of **S4** and of the intuitionistic logic. Curiously enough there are only finitely many, and it is an interesting open problem to find some similar results for **R**. **RM** corresponds to the super-system of the intuitionistic propositional calculus **LC** (indeed **LC** can be translated into **RM**; see [Dunn and Meyer, 1971]). Much study has been done of the ‘superintuitionistic’ calculi (with an emphasis on the decision problem), and it would be good to see some of the ideas of this carried over to the ‘super-relevant’ calculi. A small start was begun in [Dunn, 1979a].

Routley and Meyer [1973] add the postulate

$$0 \leq a \text{ or } 0 \leq a^*$$

to the requirement on an **R**-frame to get an **RM**-frame. Dunn [1979a] instead adds the requirement

$$Rac \Rightarrow a \leq c \text{ or } b \leq c,$$

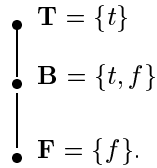
which neatly generalised to give a family of postulates yielding set-theoretical semantics for a denumerable family of weakenings of **RM** which are algebraised by adding various weakenings of idempotence ( $a^{n+1} = a^n$ ). It is an open problem whether **R** itself is the intersection of this family and whether they all have the finite model property (if so, **R** is decidable). Since **R** is undecidable, one of these must be false. However, it is unknown at the time of writing which one fails.

*Proof Sketch*

1.  $\neg X \rightarrow (\neg X \rightarrow \neg X)$  Mingle Axiom, Subst.
2.  $X \rightarrow (\neg X \rightarrow X)$  1, Permutation and Contraposition
3.  $(A \vee \neg A) \wedge (B \vee \neg B) \rightarrow \neg((A \vee \neg A) \wedge (B \vee \neg B)) \rightarrow ((A \vee \neg A) \wedge (B \vee \neg B))$  2, Subst.
4.  $\neg(A \vee \neg A) \vee \neg(B \vee \neg B) \rightarrow (A \vee \neg A) \wedge (B \vee \neg B)$  3, MP, de M
5.  $A \wedge \neg A \rightarrow B \vee \neg B$  4,  $\vee I, \wedge E$ , de M.

Kalman [1958] especially investigated de Morgan lattices with the property  $a \wedge \neg a \leq b \vee \neg b$ . We will call these *Kalman* lattices. He showed that every Kalman lattice is isomorphic to a subdirect product of the de Morgan lattice **3**. This implies a three-valued Homomorphism Separation Property for Kalman lattices (which also can be proven by modifying the proof of its four-valued analogue, noting that each ‘side’ of 4 is just a copy of 3). The representation in terms of polarities uses polarities  $X = (X_1, X_2)$  where  $X_1 \cup X_2 = U$ , i.e.  $X_1$  and  $X_2$  are exhaustive.

This means informally that  $X$  always receives at least one of the values true and false. This leads to a semantics using ambivaluations into the left-hand side of 4:



This idea leads to a simpler Kripke-style semantics for **RM** using an ordinary binary accessibility relation instead of the Routley–Meyer ternary one (actually this semantics antedates the Routley–Meyer one, the results having been presented in [Dunn, 1969]—cf. [Dunn, 1976b] for a full presentation. No details will be supplied here. This semantics has been generalised to first-order **RM** with a constant domain semantics [Dunn, 1976c]). The analogous question with Routley–Meyer semantics is now closed in the negative in the work of [Fine, 1989], which we consider in Section 3.12.

Meyer [1980] has used this ‘binary semantics’ to give a proof of an appropriate Interpolation Lemma for **RM**. (Unfortunately, interpolation fails for **E** and **R** [Urquhart, 1993].)

### 3.11 Spin Offs from the Routley–Meyer Semantics

The Routley–Meyer semantical techniques can be used to prove a variety of results concerning the system **R** and related logics which were either more



complicated using other methods (usually algebraic or Gentzen methods), or even impossible. Thus (cf. [Routley and Meyer, 1973]), it is possible to give a variety of conservative extension results (being careful in constructing the canonical model to use only connectives and sentential constant available in the fragment being extended). Also it is possible to give a proof of the admissibility of  $\gamma$  (see [Routley and Meyer, 1973]) that is easier than the original algebraic proof (though not as easy as Meyer's latest proof using metavaluations—cf. Section 2.4). Admissibility of  $\gamma$  amounts to showing that if  $A$  is refutable in a given  $\mathbf{R}$ -model  $(K, R, 0, \vDash)$  then  $A$  is refutable in a *normal*  $\mathbf{R}$ -model  $(K', R', 0', \vDash')$  (one where  $0'^* = 0'$ ) gotten by adding a new 'zero' and redefining  $R'$  and  $\vDash'$  in a certain way from  $R$  and  $\vDash$ .

Perhaps the most interesting new property to emerge this way is 'Halldén completeness', i.e. if  $\vdash_{\mathbf{R}} A \vee B$  and  $A$  and  $B$  share no propositional variables in common, then  $\vdash_{\mathbf{R}} A$  or  $\vdash_{\mathbf{R}} B$  ([Routley and Meyer, 1973, Section 2.3]).

Another direction that the Routley–Meyer semantics has taken quickly ends up in heresy: classical (Boolean) negation  $\sim$  can be added to  $\mathbf{R}$  with horrible theorems resulting like  $A \wedge \sim A \rightarrow B$ , and yet  $\mathbf{R}$  does not collapse to classical logic. Indeed no new theorems emerge in the original vocabulary of  $\mathbf{R}$ . The idea is to take a normal  $\mathbf{R}$ -model  $(K, R, 0, *, \vDash)$  and turn it in for a new  $\mathbf{R}$ -model  $(K', R', 0', *', \vDash')$ , whose  $0'$  is a new element  $K' = K \cup \{0'\}$ ,  $*'$  is like  $*$  but with  $0'*' = 0'$ , and  $R'$  is like  $R$  with the additional features:

1.  $R'0'ab$  iff  $R; a0'b$  iff  $a = b$ ,
2.  $R'ab0'$  iff  $a = b^*$ .

Also  $\vDash'$  is just like  $\vDash$  but with  $0' \vDash A$  if  $0 \vDash A$ .

The whole point of this exercise is to provide refuting  $\mathbf{R}$ -models for all non- $\mathbf{R}$ -theorems that have the property

$$a \leq b \text{ (i.e. } R0'ab) \Rightarrow a = b.$$

These are called 'classical  $\mathbf{R}$ -models' (first studied in Meyer and Routley [1973a; 1973b]) and upon them one can define

$$a \vDash \sim A \Leftrightarrow \text{not } a \vDash A.$$

One could not do this on ordinary  $\mathbf{R}$ -models without things coming apart at the seams, because in order to have the theorem  $\sim p \rightarrow \sim p$  valid, one would need the Hereditary Condition to hold for  $\sim p$ , i.e. if  $a \leq b$ , then if  $a \vDash \sim p$  then  $b \vDash \sim p$ , i.e. if  $a \vDash p$  then  $b \vDash p$ . But one has no reason to think that this is the case, since all one has is the converse coming from the fact that the Hereditary condition holds for  $p$ . The inductive proof the Hereditary condition breaks down in the presence of Boolean negation, but of course with classical  $\mathbf{R}$ -models the Hereditary Condition becomes vacuous and there is no need for a proof.

This leads to certain technical simplicities, e.g. it is possible to give Gödel–Lemmon style axiomatisations of relevance logics like the familiar ones for modal logics, where one takes among one’s axioms all classical tautologies (using  $\sim$ )—cf. [Meyer, 1974].

But it also leads to certain philosophical perplexities. For example, what was all the fuss Anderson and Belnap made against contradictions implying everything and disjunctive syllogism? Boolean negation trivially satisfies them, so what is the interest in de Morgan negation failing to satisfy them. Will the real negation please stand up?

A certain schism developed in relevance logic over just how Boolean negation should be regarded. See [Belnap and Dunn, 1981; Restall, 1999] for the ‘con’ side and [Meyer, 1978] for the ‘pro’ side.

Belnap and Dunn [1981] point out that although Meyer’s axiomatisations of  $\mathbf{R}$  with Boolean negation do not lead to any new *theorems* in the standard vocabulary of  $\mathbf{R}$ , they do lead to new derivable rules, e.g.  $A \wedge \neg A \vdash B$  and  $\neg A \wedge (A \vee B) \vdash B$  (note well that the negation here is de Morgan negation). This can be seen quite readily if one recognises that the semantic correlate of  $X \vdash Y$  is that  $0 \vDash X \Rightarrow 0 \vDash Y$  in all classical  $\mathbf{R}$ -models, and that since all such are normal,  $\neg$  behaves at 0 in these just like classical negation. We both think this point counts against enriching  $\mathbf{R}$  with Boolean negation, but Meyer [1978, note 21] thinks otherwise.

### 3.12 Semantics for $\mathbf{RQ}$

The question of how to extend these techniques to handle quantified relevance logics was open for a long time. The first significant results were by Routley, who showed that the obvious constant domain semantics were sufficient to capture  $\mathbf{BQ}$ , the natural first-order extension of  $\mathbf{B}$  [Routley, 1980b]. However, extending the result to deal with systems involving transitivity postulates in the semantics (such as  $Rabc \wedge Rcde \Rightarrow R^2abde$ ) proved difficult. To verify that the frame of prime theories on some constant domain actually satisfies this condition (given that the logic satisfies a corresponding condition, here the prefixing axiom) requires constructing a new prime theory  $x$  such that  $Rabx$  and  $Rxde$ . And there seems to be no general way to show that such a theory can be constructed using the domain shared by the other theories. This is not a problem for logics like  $\mathbf{BQ}$ , in which the frame conditions do not have conditions which, to be verified in the completeness proof, require the construction of new theories.

Fine showed that this is not merely a problem with our proof techniques. Logics like  $\mathbf{RQ}$ ,  $\mathbf{EQ}$ ,  $\mathbf{TQ}$  and even  $\mathbf{TWQ}$  are incomplete with respect to the constant domain semantics on the frames for the propositional logics [Fine, 1989]. He has given a technical argument for this, constructing a formula in the language of  $\mathbf{RQ}$  which is true in all constant domain models, but which is not provable. The argument is too detailed to give here. It consists of a

simple part, which shows that the formula

$$\begin{aligned} & ((p \rightarrow \exists x E x) \wedge \forall x ((p \rightarrow F x) \vee (G x \rightarrow H x))) \\ & \rightarrow (\forall x (E x \wedge F x \rightarrow q) \wedge \forall x ((E x \rightarrow q) \vee G x) \rightarrow \exists x H x \vee (p \rightarrow q)) \end{aligned}$$

is valid in the constant domain semantics. This is merely a tedious verification that there is no counterexample. The subtle part of his argument is the construction of a countermodel. Clearly the countermodel cannot be a constant domain frame. Instead, he constructs a frame with variable domains, in which each of the axioms of **RQ** is valid (and in which the rules preserve validity) but the offending formula fails. This is quite a tricky argument, for variable domain semantics tend not to verify **RQ**'s analogue to the Barcan formula

$$\forall x (p \rightarrow F x) \rightarrow (p \rightarrow \forall x F x)$$

But Fine constructs his example in such a way that this formula is valid, despite the variable domains.

Despite this problem, Fine has found a semantics with respect to which the logic **RQ** is sound and complete. This semantics rests on a different view of the quantifiers. For Fine's account, a statement of the form  $\forall x A(x)$  is true at a set-up not only when  $A(c)$  is true for each individual  $c$  in the domain of the set-up, but instead, when  $A(c)$  is true for an *arbitrary* individual  $c$ . In symbols,

$$a \models \forall x A(x) \text{ iff } (\exists a \uparrow)(\exists c \in D_{a \uparrow} - D_a)(a \uparrow \models A(c)).$$

That is, for every set-up  $a$  there are *expansions* of the form  $a \uparrow$  where we add new elements to the domain, but these are totally arbitrary. The frames Fine defines are rather complex, needing not only the  $\uparrow$  operator but also a corresponding  $\downarrow$  operator which cuts down the domain of a set-up, and an across operator  $\leftarrow$  which identifies points in setups ( $\rightarrow (a, \{c, d\})$  is the minimal extension of the set-up  $a$  in which the individuals  $c$  and  $d$  are identified. Instead of discussing the details of Fine's semantics, we refer the reader to his paper which introduced them [Fine, 1988]. Fine's work has received some attention, from Mares, who considers options for the semantics of identity [Mares, 1992]. However, it must be said that while the semantic structure pins down the behaviour of **RQ** and related systems exactly, it is not altogether clear whether the rich and complex structure of Fine's semantics is necessary to give a semantics for quantified relevance logics.

Whatever one's thoughts about the theoretical adequacy of Fine's semantics, they do raise some important issues for anyone who would give a semantic structure for quantified relevance logics. There are a number of issues to be faced and a number of options to be weighed up. One option is to give complete primacy to the frames for the propositional logics, and

to use the constant domain semantics on these frames. The task then is to axiomatise this extension. The task is also to give some interpretation of what the points in these semantic structures might be. For if they are theories (or prime theories) then the evaluation clauses for the quantifiers do not make a great deal of sense without further explanation. No-one thinks that a claim of the form  $\exists xA(x)$  can be a member of a theory only if there is an object in the language of the theory which satisfies  $A$  according to that theory. Nor are we so readily inclined to think that all theories need share the same domain of quantification.

If, on the other hand, we take the set-ups in frames to be quite like (some class of) theories, then we must face the issue of the relationships between these theories. No doubt, if  $\forall xA(x)$  is in some theory, then  $A(c)$  will be in that theory for any constant  $c$  in the language of the theory. However, the converse need not be the case.

Anyway, it is clear that there is a lot of work to be done in the semantics of relevance logics with quantifiers. One area which hasn't been explored at any depth, but which looks like it could bring some light is the semantics of positive quantified relevance logics. Without the distribution of the universal quantifier over disjunction, these systems are subsystems of intuitionistic logic.

## 4 THE DECISION PROBLEM

### 4.1 Background

When the original of this Handbook article was published back in 1985, without a doubt the outstanding open problem in relevance logics was the question as to whether there exists a decision procedure for determining whether formulas are theorems of the system **E** or **R**. Anderson [1963] listed it second among his now historic open problems (the first was the admissibility of Ackermann's rule  $\gamma$  discussed in Section 2). Through the work of Urquhart [1984], we now know that there is no such decision procedure.

Harrop [1965] lends interest to the decision problem with his remark that 'all "philosophically interesting" propositional calculi for which the decision problem has been solved have been found to be decidable ...'.<sup>35</sup> We now have a very good counterexample to Harrop's claim.

In this section we shall examine Urquhart's proof, but before we get there we shall also consider various fragments and subsystems of **R** for which there are decision procedures. **R** will be our paradigm throughout this discussion, though we will make clear how things apply to related systems.

---

<sup>35</sup>He continues somewhat more technically '... and none is known for which it has been proved that it does not possess the finite model property with recursive bound.'

## 4.2 Zero-degree Formulas

These are formulas containing only  $\wedge$ ,  $\vee$ , and  $\neg$ . As was explained in Section 1.7, the zero-degree theorems of  $\mathbf{R}$  (or  $\mathbf{E}$ ) are precisely the same as those of the classical propositional calculus, so of course the usual two valued truth tables yield a decision procedure.

## 4.3 First-degree Entailments

Two different (though related) ‘syntactical’ decision procedures were described for these in Section 1.7 (the method of ‘tautological entailments’ and the method of ‘coupled trees’). A ‘semantical’ decision procedure using a certain four element matrix  $\mathbf{4}$  is described in Section 3.3. The story thus told leaves out the historically (and otherwise) very important role of a certain eight element matrix  $M_0$  (cf. [Anderson and Belnap, 1975, Section 22.1.3]). This matrix is essential for the study of first-degree formulas and higher (see Section 4.4 below), in so much as it is impossible to define an implication operation on  $\mathbf{4}$  and pick out a proper subset of designated elements so as to satisfy the axioms of  $\mathbf{E}$  (*a fortiori*  $\mathbf{R}$ ). Indeed  $M_0$  was used in [Anderson and Belnap Jr., 1962b] and [Belnap, 1960b] to isolate the first-degree entailments of  $\mathbf{R}$ , and the formulation of Section 1.7 presupposes this use.

## 4.4 First-degree Formulas

These are ‘truth functions’ of first-degree entailments and/or formulas containing no  $\rightarrow$  at all (the ‘zero-degree formulas’). Belnap [1967a] gave a decision procedure using certain finite ‘products’ of  $M_0$ . No one such product is characteristic for  $\mathbf{D}_{\text{fdf}}$ , but every non-theorem of  $\mathbf{E}_{\text{fdf}}$  is refutable in some such products  $M_0^n$  (where  $n$  may in fact be computed as the largest number of first-degree entailments occurring in a disjunction once the candidate theorem has been put in conjunctive normal form). Hence  $\mathbf{E}_{\text{fdf}}$  has the finite model property which suffices of course for decidability (cf. [Harrop, 1965]). This is frankly one of the most difficult proofs to follow in the whole literature of relevance logics. A sketch may be found in [Anderson and Belnap, 1975, Section 19].

## 4.5 ‘Career Induction’

This is what Belnap has labelled the approach, exemplified in Sections 4.1–4.3 above of extending the positive solution to the decision problem ‘a degree at a time’. The last published word on the Belnap approach is to be found in his [1967b] where he examines entailments between conjunctions of first-degree entailments and first degree entailments.

Meyer [1979c], by an amazingly general and simple proof, shows that a positive answer to the decision problem for ‘second-degree formulas’ (no  $\rightarrow$  within the scope of an arrow within the scope of an  $\rightarrow$ ) is equivalent to finding a decision procedure for all of **R**.

#### 4.6 Implication Fragment

We now start another tack. Rather than looking at fragments of the whole system **R** delimited by complexity of formulas, we instead consider fragments delimited by the connectives which they contain. The earliest result of this kind is due to [Kripke, 1959b], who gave a Gentzen system for the implicational fragments of **E** and **R**, and showed them decidable. We shall here examine the implicational fragment of **R** (**R** $_{\rightarrow}$ ) in some detail as a kind of paradigm for this style of argument.<sup>36</sup>

The appropriate Gentzen calculus<sup>37</sup>  $LR_{\rightarrow}$  is the same as that given by Gentzen [1934] except for two trivial differences and one profound difference. The first trivial difference is the obvious one that we take only the operational rules for implication, and the second trivial difference consequent on this (with negation it would have to be otherwise) is that we can restrict our sequents to those with a single formula in the consequent. The profound difference is that we drop the structural rule variously called ‘thinning’ or ‘weakening’. This leaves:

*Axioms.*

$$A \vdash A.$$

*Structural Rules.*

$$\text{Permutation } \frac{\alpha, A, B, \beta, \vdash C}{\alpha, B, A, \beta, \vdash C} \quad \text{Contraction } \frac{\alpha, a, A \vdash B}{\alpha, A \vdash B}$$

*Operational Rules.*

$$(\vdash \rightarrow) \frac{\alpha, A \vdash B}{\alpha \vdash A \rightarrow B} \quad (\rightarrow \vdash) \frac{\alpha \vdash A \quad \beta, B \vdash C}{\alpha, \beta, A \rightarrow B \vdash C}.$$

It is easy to see why thinning would be a disaster for relevant implication.

---

<sup>36</sup>Actually this and various other results discussed below using Gentzen calculi presupposes ‘separation theorems’ due to Meyer, showing, e.g. as is relevant to this case, that all of the theorems containing only  $\rightarrow$  are provable from the axioms containing only  $\rightarrow$ .

<sup>37</sup>We do not follow Anderson and Belnap [1975] in calling Gentzen systems ‘consecution calculi’, much as their usage has to recommend it.

Thus:

$$\frac{\frac{\frac{A \vdash A}{A, B \vdash A} \text{Thinning}}{A \vdash B \rightarrow A} (\vdash \rightarrow)}{\vdash A \rightarrow (B \rightarrow A)} (\vdash \rightarrow)$$

It is desirable to prove ‘The Elimination Theorem’, which says that the following rule would be redundant (could be eliminated).

$$\text{(Cut)} \quad \frac{\alpha \vdash A \quad \beta, A \vdash B}{\alpha, \beta \vdash B}.$$

This is needed to show the equivalence of  $LR_{\rightarrow}$  to its usual Hilbert-style (axiomatic system ‘ $HR_{\rightarrow}$ ’  $\mathbf{R}_{\rightarrow}$  one of the formulations of Section 1.3). We will not pause on details here, but the principal question regarding the equivalence is whether *modus ponens* (The sole rule for  $HR_{\rightarrow}$ ) is admissible in the sense that whenever  $\vdash A$  and  $\vdash A \rightarrow B$  are both derivable in  $LR_{\rightarrow}$ , so is  $\vdash B$  (let  $\alpha$  and  $\beta$  be empty).

The strategy of the proof of the Elimination Theorem can essentially be that of Gentzen with one important but essentially minor modification. Thus, Gentzen actually proved something stronger than Cut elimination, namely,

$$\text{(Mix)} \quad \frac{\alpha \vdash A \quad \beta \vdash B}{\alpha, [\beta - A] \vdash B},$$

where  $[\beta - A]$  is the result of deleting *all* occurrences of  $A$  from  $\beta$ . This is useful in the induction, but sometimes it takes out too many occurrences of  $A$ . In Gentzen’s framework these could always be thinned back in, but of course this is not available with  $LR_{\rightarrow}$ . We thus instead generalise Cut to the rule

$$\text{(Fusion)} \quad \frac{\alpha \vdash A \quad \beta \vdash B}{\alpha, (\beta - A) \vdash B},$$

where  $\beta$  contains some occurrences of  $A$  and  $(\beta - A)$  is the result of deleting as many of those occurrences as one wishes (but at least one).

The main strategy of the decision procedure for  $LR_{\rightarrow}$  is to limit applications of the contraction rule so as to prevent a proof search from running on forever in the following manner: ‘Is  $p \vdash q$  derivable? Well it is if  $p, p \vdash q$  is derivable. Is  $p, p \vdash q$  derivable? Well it is if  $p, p, p \vdash q$  is, etc.’.

We need one simple notion before strategy can be achieved. We shall say that the sequent of  $\alpha' \vdash A$  is a *contraction* of sequent  $\alpha \vdash A$  just in

case  $\alpha' \vdash A$  can be derived from  $\alpha \vdash A$  by (repeated) applications of the rules Contraction and Permutation (with respect to this last it is helpful not even to distinguish two sequents that are mere permutations of one another). The idea that we now want to put in effect is to drop the rule Contraction, replacing it by building into the operational rules a limited amount of contraction (in the generalised sense just explained).

More precisely, the idea is to allow a contraction of the conclusion of an operational rule only in so far as the same result could not be obtained by first contracting the premises. A little thought shows that this means no change for the rule  $(\vdash \rightarrow)$ , and that the following will suffice for

$$(\rightarrow \vdash') \quad \frac{\alpha \vdash A \quad \beta, B \vdash C}{[\alpha, \beta, A \rightarrow B] \vdash C}$$

where  $[\alpha, \beta, A \rightarrow B]$  is any contraction of  $\alpha, \beta, A \rightarrow B$  such that :

1.  $A \rightarrow B$  occurs only 0, 1, or 2 times fewer than in  $\alpha, \beta, A \rightarrow B$ ;
2. Any formula other than  $A \rightarrow B$  occurs only 0 or 1 time fewer.

It is clear that after modifying  $LR_{\rightarrow}$  by building some limited contraction into  $(\rightarrow \vdash)$  in the manner just discussed, the following is provable by an induction on length of derivations:

**CURRY'S LEMMA.**<sup>38</sup> *If a sequent  $\Gamma'$  is a contraction of a sequent  $\Gamma$  and  $\Gamma$  has a derivation of length  $n$ , then  $\Gamma'$  has a derivation of length  $\leq n$ .*

Clearly this lemma shows that the modification of  $LR_{\rightarrow}$  leaves the same sequents derivable (since the lemma says the effect of contraction is retained). So henceforth we shall by  $LR_{\rightarrow}$  always mean the modified version.

Besides the use just adverted to, Curry's Lemma clearly shows that every derivable sequent has an *irredundant* derivation in the following sense: one containing no branch with a sequent  $\Gamma'$  below a sequent  $\Gamma$  of which it is a contraction.

We are finally ready to begin explicit talk about the decision procedure. Given a sequent  $\Gamma$ , one begins the test for derivability as follows (building

<sup>38</sup>This is named (following [Anderson and Belnap, 1975]) after an analogous lemma in [Curry, 1950] in relation to classical (and intuitionistic) Gentzen systems. There, with free thinning available, Curry proves his lemma with  $(\rightarrow \vdash)$  (in its singular version) stated as:

$$\frac{\Gamma, A \rightarrow B \vdash A \quad \Gamma, A \rightarrow B, B \vdash C}{\Gamma, A \rightarrow B \vdash C}.$$

This in effect *requires* the maximum contraction *permitted* in our statement of  $(\rightarrow \vdash)$  above, but this is OK since items contracted 'too much' can always be thinned back in. Incidentally, our statement of  $(\rightarrow \vdash)$  also differs somewhat from the statement of Anderson and Belnap [1975] or Belnap and Wallace [1961], in that we build in just the minimal amount of contraction needed to do the job.



a ‘complete proof search tree’): one places above  $\Gamma$  all possible premises or pairs of premises from which  $\Gamma$  follows by one of the rules. Note well that even with the little bit of contraction built into  $(\rightarrow\vdash)$  this will still be only a finite number of sequents. Incidentally, one draws lines from those premises to  $\Gamma$ . One continues in this way getting a tree. It is reasonably clear that if a derivation exists at all, then it will be formed as a subtree of this ‘complete proof search there’, by the paragraph just above, the complete proof search tree can be constructed to be irredundant. But the problem is that the complete proof search tree may be infinite, which would tend to louse up the decision procedure. There is a well-known lemma which begins to come to the rescue:

**KÖNIG’S LEMMA.** *A tree is finite iff both (1) there are only finitely many points connected directly by lines to a given point (‘finite fork property’) and (2) each branch is finite (‘finite branch property’).*

By the ‘note well’ in the paragraph above, we have (1). The question remaining then is (2), and this is where an extremely ingenious lemma of Kripke’s plays a role. To state it we first need a notion from Kleene. Two sequents  $\alpha \vdash A$  and  $\alpha' \vdash A$  are *cognate* just when exactly the same formulas (not counting multiplicity) occur in  $\alpha$  as in  $\alpha'$ . Thus, e.g. all of the following are cognate to each other:

- (1)  $X, Y \vdash A$
- (2)  $X, X, Y \vdash A$
- (3)  $X, Y, Y \vdash A$
- (4)  $X, X, Y, Y \vdash A$
- (5)  $X, X, X, Y, Y \vdash A$ .

We call the class of all sequents cognate to a given sequent a *cognition class*.

**KRIPKE’S LEMMA.** *Suppose a sequence of cognate sequents  $\Gamma_0, \Gamma_1, \dots$ , is irredundant in the sense that for no  $\Gamma_i, \Gamma_j$  with  $i < j$ , is  $\Gamma_i$  a contraction of  $\Gamma_j$ . Then the sequence is finite.*

We postpone elaboration of Kripke’s Lemma until we see what use it is to the decision procedure. First we remark an obvious property of  $LR_{\rightarrow}$  that is typical of Gentzen systems (that lack Cut as a primitive rule):

**SUBFORMULA PROPERTY.** *If  $\Gamma$  is a derivable sequent of  $LR_{\rightarrow}$ , then any formula occurring in any sequent in the derivation is a subformula of some formula occurring in  $\Gamma$ .*

This means that the number of cognition classes occurring in any derivation (and hence in each branch) is finite. But Kripke’s Lemma further shows

that only a finite number of members of each cognation class occur in a branch (this is because we have constructed the complete proof search tree to be irredundant). So every branch is finite, and so both conditions of König's lemma hold. Hence the complete proof search tree is finite and so there is a decision procedure.

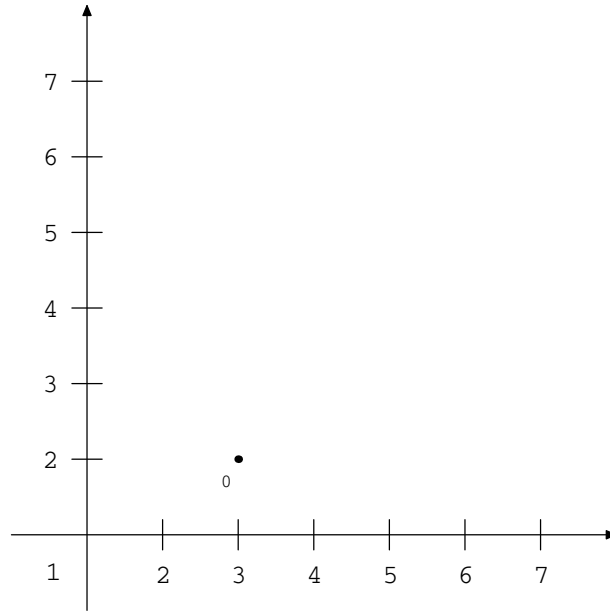


Figure 1. Sequents in the Plane

Returning now to Kripke's Lemma, we shall not present a proof (for which see [Belnap Jr. and Wallace, 1961] or [Anderson and Belnap, 1975]). Instead we describe how it can be geometrically visualised. For simplicity we consider sequents cognate to  $X, Y \vdash A$  ((1), (2), (3), etc. above). Each such sequent can be represented as a point in the upper right-hand quadrant of the co-ordinate plane (where origin is labelled with 1 rather than 0 since (1) is the minimal sequent in the cognation class). See Figure 1. Thus, e.g. (5) gets represented as '3  $X$  units' and '2  $Y$  units'.

Now given any sequent, say

$$(\Gamma_0) \quad X, X, X, Y, Y \vdash A$$

as a starting point one might try to build an irredundant sequence by first building up the number of  $Y$ 's tremendously (for purposes of keeping on the page we let this be to six rather than say a million). But in so doing one

has to reduce the number of  $X$ 's (say, to be strategic, by one). The graph now looks like 2 for the first two members of the sequence  $\Gamma_0, \Gamma_1$ .

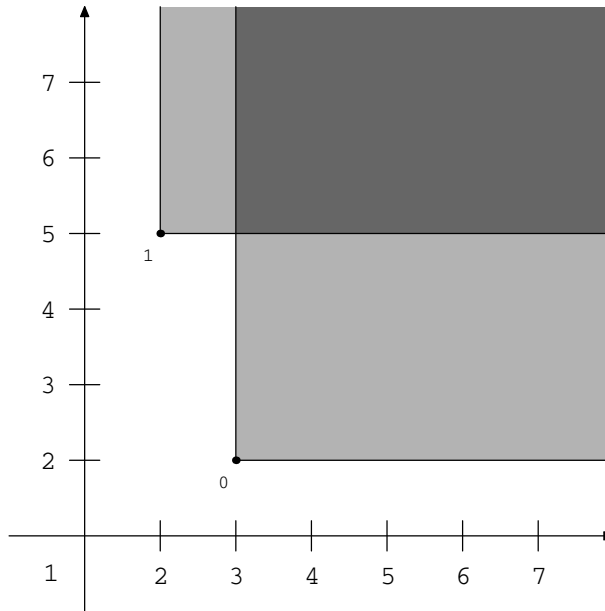


Figure 2. Descending Regions

The purpose of the intersecting lines at each point is to mark off areas (shaded in the diagram) into which no further points of the sequence may be placed. Thus if  $\Gamma_2$  were placed as indicated at the point (6, 5), it would reduce to  $\Gamma_0$ . What this means is that each new point must march either one unit closer to the  $X$  axis or one unit closer to the  $Y$  axis. Clearly after a finite number of points one or the other of the two axes must be 'bumped', and then after a short while the other must be bumped as well. When this happens there is no space left to play without the sequence becoming redundant.

The generalisation to the case of  $n$  formulas in the antecedent to Euclidean  $n$ -space is clear (this is with  $n$  finite—with  $n$  infinite no axis need ever be bumped).

Incidentally, Kripke's Lemma (as Meyer discovered) is equivalent to a theorem of Dickson about prime numbers: Let  $M$  be a set of natural numbers all of which are composed out of the first  $m$  primes. Then every  $n \in M$  is of the form  $P_1^{n_1} \cdot P_2^{n_2} \cdot \dots \cdot P_k^{n_k}$ , and hence (by unique decomposition) can be regarded as a sequence of the  $P_i$ 's in which each  $P_i$  is repeated  $n_i$  times. Divisibility corresponds then to contraction (at least neglecting the case

$n_i = 0$ ). Dickson's theorem says that if no member of  $M$  has a proper divisor in  $M$ , then  $M$  is finite.

Before going on to consider how the addition of connectives changes the complexity, let us call the reader's attention to a major open problem: It is still unknown whether the implication fragment of  $\mathbf{T}$  is decidable.

#### 4.7 Implication–Negation Fragment

The idea of  $\mathbf{LR}_{\rightarrow}$  is to accommodate the classical negation principles presenting  $\mathbf{R}$  in the same way that Gentzen [1934] accommodated them for classical logic: provide multiple right-hand sides for the sequents. This means that a sequent is of the form  $\alpha \vdash \beta$ , where  $\alpha$  and  $\beta$  are (possible empty) finite sequences of formulas. One adds structural rules for Permutation and Contraction on the right-hand side, reformulates  $(\vdash \rightarrow)$  and  $(\rightarrow \vdash)$  as follows

$$(\vdash \rightarrow) \frac{\alpha, A \vdash B, \beta}{\alpha \vdash A \rightarrow B, \beta} \quad (\rightarrow \vdash) \frac{\alpha \vdash A, \gamma \quad \beta, B \vdash \delta}{\alpha, \beta, A \rightarrow B \vdash \gamma, \delta},$$

and adds 'flip and flop' rules for negation:

$$(\vdash \neg) \frac{\alpha, A \vdash \beta}{\alpha \vdash \neg A, \beta} \quad (\neg \vdash) \frac{\alpha \vdash A, \beta}{\alpha, \neg A \vdash \beta}.$$

$\mathbf{LE}_{\rightarrow}$  is the same except that in the rule  $(\vdash \rightarrow)\beta$  must be empty and  $\alpha$  must consist only of formulas whose main connective is  $\rightarrow$ . The decision procedure for  $\mathbf{LE}_{\rightarrow}$  was worked out by Belnap and Wallace [1961] along basically the lines of the argument of Kripke just reported in the last section, and is clearly reported in [Anderson and Belnap, 1975, Section 13]. The modification to  $\mathbf{LR}_{\rightarrow}$  is straightforward (indeed  $\mathbf{LR}_{\rightarrow}$  is easier because one need not prove the theorem of p. 128 of [Anderson and Belnap, 1975], and so one can avoid all the apparatus there of 'squeezes'). McRobbie and Belnap [1979] have provided a nice reformulation of  $\mathbf{LR}_{\rightarrow}$  in an analytic tableau style, and Meyer has extended this to give analytic tableau for linear logic and other systems in the vicinity of  $\mathbf{R}$  [Meyer *et al.*, 1995].

#### 4.8 Implication–Conjunction Fragment, and $\mathbf{R}$ Without Distribution

This work is to be found in [Meyer, 1966]. The idea is to add to  $\mathbf{LR}_{\rightarrow}$  the Gentzen rules:

$$(\wedge \vdash) \frac{\alpha, A \vdash C}{\alpha, A \wedge B \vdash C} \quad \frac{\alpha, \beta \vdash C}{\alpha, A \wedge B \vdash C} \quad (\vdash \wedge) \frac{\alpha \vdash A \quad \alpha \vdash B}{\alpha \vdash A \wedge B}.$$

Again the argument for decidability is a simple modification of Kripke's.

Note that it is important that the rule  $(\wedge \vdash)$  is stated in two parts, and not as one ‘Ketonen form’ rule:

$$(K\wedge \vdash) \quad \frac{\alpha, A, B \vdash C}{\alpha, A \wedge B \vdash C}.$$

The reason is that without thinning it is impossible to derive the rule(s)  $(\wedge \vdash)$  from  $(K\wedge \vdash)$ .

Early on it was recognised that the distribution axiom

$$A \wedge (B \vee C) \rightarrow (A \vee B) \vee C$$

was difficult to derive from Gentzen-style rules for **E** and **R**. Thus Anderson [1963] saw this as the sticking point for developing Gentzen formulations, and Belnap [1960b, page 72]) says with respect to  $LE_{\rightarrow}$  that ‘the standard rules for conjunction and disjunction could be added ... the Elimination Theorem (suitably modified) remaining provable. However, [since distribution would not be derivable], the game does not seem worth the candle’. Meyer [1966] carried out such an addition to  $LR_{\rightarrow}$ , getting a system he called  $LR_{\rightarrow}$ –, whose Hilbert-style version is precisely **R** without the distribution axiom. He showed using a Kripke-style argument that this system is decidable. This system is now called **LR**, for “lattice **R**”.

Meyer [1966] also showed how **LR** can be translated into  $\mathbf{R}_{\rightarrow, \wedge}$  rather simply. Given a formula  $A$  in the language of  $\mathbf{LR}^+$ , let  $V$  be the set of variables in  $A$ , and let two atomic propositions  $p_t$  and  $p_f$  not in  $V$ . Set  $\neg A$  for the moment to be  $A \rightarrow p_f$ , to define a translation  $A'$  of  $A$  as follows.

$$\begin{aligned} p' &= p \\ t' &= p_t \\ (A \rightarrow B)' &= A' \rightarrow B' \\ (A \wedge B)' &= A' \wedge B' \\ (A \vee B)' &= \neg(\neg A' \wedge \neg B') \\ (A \circ B)' &= \neg(A' \rightarrow \neg B') \end{aligned}$$

then setting  $t(A) = \bigwedge \{p_t \rightarrow (p \rightarrow p) : p \in V \cup \{p_t, p_f\}\}$  and  $f(A) = \bigwedge \{\neg\neg p \rightarrow p : p \in V \cup \{p_t, p_f\}\}$ , we get the following theorem:

**TRANSLATION THEOREM** (Meyer). *If  $A$  is a formula in  $\mathbf{LR}^+$  then  $A$  is provable in  $\mathbf{LR}^+$  if and only if  $(t(A) \wedge f(A) \wedge p_t) \rightarrow A'$  is provable in  $\mathbf{R}_{\rightarrow, \wedge}$ .*

The proof is given in detail in [Urquhart, 1997], and we will not present it here.

Some recent work of Alasdair Urquhart has shown that although  $\mathbf{R}_{\rightarrow, \wedge}$  is decidable, it is *only just* decidable [Urquhart, 1990; Urquhart, 1997].

More formally, Urquhart has shown that given any particular formula in the language of  $\mathbf{R}_{\rightarrow, \wedge}$ , there is no primitive recursive bound on either the time or the space taken by a computation of whether or not that formula is a theorem. Presenting the proof here would take us too far away from the *logic* to be worthwhile, however we can give the reader the kernel of the idea behind Urquhart's result.

Urquhart follows work of [Lincoln *et al.*, 1992] by using a propositional logic to encode the behaviour of a *branching counter machines*. A counter machine has a finite number of *registers* (say,  $r_i$  for suitable  $i$ ) which each hold one non-negative integer, and some finite set of possible *states* (say,  $q_j$  for suitable  $j$ ). Machines are coded with a list of instructions, which enable you to *increment* or *decrement* registers, and test for registers' being zero. A *branching* counter machine dispenses with the test instructions and allows instead for machines to take multiple execution paths, by way of *forking* instructions. The instruction  $q_i + r_j q_k$  means "when in  $q_i$ , add 1 to register  $r_j$  and enter stage  $q_k$ ," and  $q_i - r_j q_k$  means "when in  $q_i$ , subtract 1 to register  $r_j$  (if it is non-empty) and enter stage  $q_k$ ," and  $q_i f q_j q_k$  is "when in  $q_i$ , fork into two paths, one taking state  $q_j$  and the other taking  $q_k$ ."

A machine configuration is a state, together with the values of each register. Urquhart uses the logic  $\mathbf{LR}$  to simulate the behaviour of a machine. For each register  $r_i$ , choose a distinct variable  $R_i$ , for each state  $q_j$  choose a distinct variable  $Q_j$ . The configuration  $\langle q_i; n_1, \dots, n_l \rangle$ , where  $n_i$  is the value of  $r_i$  is the formula

$$Q_i \circ R_1^{n_1} \circ \dots \circ R_l^{n_l}$$

and the instructions are modelled by sequents in the Gentzen system, as follows:

Instruction	Sequent
$q_i + r_j q_k$	$Q_i \vdash Q_k \circ R_j$
$q_i - r_j q_k$	$Q_i, R_j \vdash Q_k$
$q_i f q_j q_k$	$Q_i \vdash Q_j \vee Q_k$

Given a machine program (a set of instructions) we can consider what is provable from the sequents which code up those instructions. This set of sequents we can call the *theory* of the machine.  $Q_i \circ R_1^{n_1} \circ \dots \circ R_l^{n_l} \vdash Q_j \circ R_1^{m_1} \circ \dots \circ R_l^{m_l}$  is intended to mean that from state configuration  $\langle q_i; n_1, \dots, n_l \rangle$  all paths will go through configuration  $\langle q_j; m_1, \dots, m_l \rangle$  after some number of steps.

A branching counter machine *accepts* an initial configuration if when run on that configuration, all branches terminate at the final state  $q_f$ , with all registers taking the value zero. The corresponding condition in  $\mathbf{LR}$  will be the provability of

$$Q_i \circ R_1^{n_1} \circ \dots \circ R_l^{n_l} \vdash Q_m$$

This will *nearly* do to simulate branching counter machines, except for the fact that in **LR** we have  $A \vdash A \circ A$ . This means that each of our registers can be incremented as much as you like, provided that they are non-zero to start with. This means that each of our machines need to be equipped with every instruction of the form  $q_i > 0 + r_j q_i$ , meaning “if in state  $q_i$ , add 1 to  $r_j$ , provided that it is already nonzero, and remain in state  $q_i$ .”

Given these definitions, Urquhart is able to prove that a configuration is accepted in branching counter machine, if and only if the corresponding sequent is provable from the theory of that machine. But this is equivalent to a formula

$$\bigwedge \text{Theory}(M) \wedge t \rightarrow (Q_1 \rightarrow Q_m)$$

in the language of **LR**. It is then a short step to our complexity result, given the fact that there is no primitive recursive bound on determining acceptability for these machines. Once this is done, the translation of **LR** into  $\mathbf{R}_{\rightarrow \wedge}$  gives us our complexity result.

It is still unknown if  $\mathbf{R}_{\rightarrow}$  has similar complexity or whether it is a more tractable system.

Despite this complexity result, Kripke’s algorithm can be implemented with quite some success. The theorem prover **Kripke**, written by McRobbie, Thistlewaite and Meyer, implements Kripke’s decision procedure, together with some quite intelligent proof-search pruning, by means of finite models. If a branch is satisfiable in **RM3**, for example, there is no need to extend it to give a contradiction. This implementation *works* in many cases [Thistlewaite *et al.*, 1988]. Clearly, work must be done to see whether the horrific complexity of this problem in general can be transferred to results about *average case* complexity.

Finally, before moving to add distribution, we should mention that Linear Logic (see Section 5.5) also lacks distribution, and the techniques used in the theorem prover **Kripke** have application in that field also.

#### 4.9 Positive **R**

In this section we will examine extensions of the Gentzen technique to cover all of positive relevance logic. We know (see Section 4.12) that this will not provide decidability. However, they provide another angle on **R** and cousins. Dunn and Minc independently developed a Gentzen-style calculus (with some novel features) for **R** without negation ( $\mathbf{LR}^+$ ).<sup>39</sup> Belnap

---

<sup>39</sup>Dunn’s result was presented by title at a meeting of the Association for Symbolic Logic, December, 1969 (see [Dunn, 1974]), and the full account is to be found in [Anderson and Belnap, 1975, Section 28.5]. Minc [1972, earliest presentation said to there to be February 24] obtained essentially the same results (but for the system with a necessity operator). See also [Belnap Jr. *et al.*, 1980].

[1960b] had already suggested the idea of a Gentzen system in which antecedents were sequences of sequences of formulas, rather than just the usual sequences of formulas (in this section ‘sequence’ always means *finite* sequence). The problem was that the Elimination Theorem was not provable.  $LR^+$  goes a step ‘or two’ further, allowing an antecedent of a sequent instead to be a sequence of sequence of . . . sequences of formulas. More formally, we somehow distinguish two kinds of sequences, ‘intensional sequences’ and ‘extensional sequences’ (say prefix them with an ‘ $I$ ’ or an ‘ $E$ ’). an antecedent can then be an intensional sequence of formulas, an extensional sequence of the last mentioned, etc. or the same thing but with ‘intensional’ and ‘extensional’ interchanged. (We do not allow things to ‘pile up’, with, e.g. *intensional* sequences of *intensional* sequences—there must be alternation).<sup>40</sup> Extensional sequences are to be interpreted using ordinary ‘extensional’ conjunction  $\wedge$ , whereas intensional sequences are to be interpreted using ‘intensional conjunction’  $\circ$ , which may be defined in the full system  $\mathbf{R}$  as  $A \circ B = \neg(A \rightarrow \neg B)$ , but here it is taken as primitive—see below).

We state the rules, using commas for extensional sequences, semicolons for intensional sequences, and asterisks ambiguously for either; we also employ an obvious substitution notation.<sup>41</sup>

$$\text{Permutation} \quad \frac{\alpha[\beta * \gamma] \vdash A}{\alpha[\gamma * \beta] \vdash A} \quad \text{Contraction} \quad \frac{\alpha[\beta * \beta] \vdash A}{\alpha[\beta] \vdash A}$$

$$\text{Thinning} \quad \frac{\alpha[\beta] \vdash A}{\alpha[\beta, \gamma] \vdash A}, \quad \text{provided } \beta \text{ is non-empty.}$$

$$\frac{\alpha; A \vdash B}{\alpha \vdash A \rightarrow B} (\vdash \rightarrow) \quad \frac{\alpha \vdash A \quad \beta[B] \vdash C}{\beta[\alpha; A \rightarrow B] \vdash C} (\rightarrow \vdash)$$

$$\frac{\alpha \vdash A \quad \alpha \vdash B}{\alpha \vdash A \wedge B} (\vdash \wedge) \quad \frac{\alpha[A, B] \vdash C}{\alpha[A \wedge B] \vdash C} (\wedge \vdash)$$

<sup>40</sup>This differs from the presentation of [Anderson and Belnap, 1975] which allows such ‘pile ups’, and then adds additional structural rules to eliminate them. Belnap felt this was a clearer, more explicit way of handling things and he is undoubtedly right, but Dunn has not been able to read his own section since he rewrote it, and so return to the simpler, more sloppy form here.

<sup>41</sup>With the understanding that substitutions do not produce ‘pile ups’. Thus, e.g. a ‘substitution’ of an intensional sequence for an item in an intensional sequence does not produce an intensional sequence with an element that is an intensional sequence formed by juxtaposition. Again this differs from the presentation of [Anderson and Belnap, 1975, cf. note 28].



$$\frac{\alpha \vdash A}{\alpha \vdash A \vee B} (\vdash \vee) \quad \frac{\alpha \vdash B}{\alpha \vdash A \vee B} (\vdash \vee) \quad \frac{\alpha[a] \vdash C \quad \alpha[B] \vdash C}{\alpha[A \vee B] \vdash C} (\vee \vdash)$$

$$\frac{\alpha \vdash A \quad \beta \vdash B}{\alpha; \beta \vdash A \circ B} (\vdash \circ) \quad \frac{\alpha[a; B] \vdash C}{\alpha \circ B] \vdash C} (\circ \vdash)$$

For technical reasons (see below) we add the sentential constant  $t$  with the axiom  $\vdash t$  and the rule:

$$\frac{\alpha[B] \vdash A}{\alpha[\beta; t] \vdash A} (t \vdash)$$

The point of the two kind of sequences can now be made clear. Let us examine the classically (and intuitionistically) valid derivation:

- (1)  $\frac{A \vdash A}{A \vdash A}$  Axiom
- (2)  $\frac{A, B \vdash A}{A \vdash B \rightarrow A}$  Thinning
- (3)  $A \vdash B \rightarrow A$  ( $\vdash \rightarrow$ ).

It is indifferent whether (2) is interpreted as

- (2 $\wedge$ )  $(A \wedge B) \rightarrow A$ , or
- (2 $\rightarrow$ )  $A \rightarrow (B \rightarrow A)$ ,

because of the principles of exportation and importation. In  $LR^+$  however we may regard (2) as ambiguous between

- (2,)  $A, B \vdash A$  (extensional), and
- (2;)  $A; B \vdash A$  (intensional).

(2,) continues to be interpreted as (2 $\wedge$ ), but (2;) is interpreted as

- (2 $\circ$ )  $(A \circ B) \rightarrow A$ .

Now in  $\mathbf{R}$ , exportation holds for  $\circ$  but not for  $\wedge$  (importation holds for both). Thus the move from (2;) to (3) is valid, but not from (2,) to (3). On the other hand, in  $\mathbf{R}$ , the inference from  $A \rightarrow C$  to  $(A \wedge B) \rightarrow C$  is valid, whereas the inference to  $(A \circ B) \rightarrow C$  is not. Thus the move from (1) to (2,) is valid, but not the move from (1) to (2;). the whole point of  $LR^+$  is to allow some thinning, but only in extensional sequences.

This allows the usual classical derivation of the distribution axiom to go through, since clearly

$$A \wedge (B \vee C) \vdash (A \wedge B) \vee C$$

can be derived with no need of any but the usual extensional sequence. The following sketch of a derivation of distribution in the consequent is even

more illustrative of the powers of  $LR^+$  (permutations are left implicit; also the top half is left to the reader);

$$\frac{\frac{\frac{A, B \vdash (A \wedge B) \vee C \quad A, C \vdash (A \wedge B) \vee C}{X \vdash X} (\vee \vdash)}{X \vdash X \quad A, (X; X \rightarrow B \vee C) \vdash (A \wedge B) \vee C} (\rightarrow \vdash)}{(X; X \rightarrow A), (X; X \rightarrow A, X \rightarrow B \vee C) \vdash (A \wedge B) \vee C} (\rightarrow \vdash)}{\frac{X \rightarrow A, X \rightarrow B \vee C; X \vdash (A \wedge B) \vee C}{(X \rightarrow A) \wedge (X \rightarrow B \vee C); X \vdash (A \wedge B) \vee C}}{\vdash (X \rightarrow A) \wedge (X \rightarrow B \vee C) \rightarrow [X \rightarrow (A \wedge B) \vee C]}$$

$LR^+$  is equivalent to  $R^+$  in the sense that for any negation-free sentence  $A$  of  $R$ ,  $\vdash A$  is derivable in  $LR^+$  iff  $A$  is a theorem of  $R$ . The proofs of both halves of the equivalence are complicated by technical details. Right-to-left (the interpretation theorem) requires the addition of intensional conjunction as primitive, and then a lemma, due to R. K. Meyer, to the effect that this is harmless (a conservative extension). Left-to-right (the Elimination Theorem) is what requires the addition of the constant true sentence  $t$ . This is because the ‘Cut’ rule is stated as:

$$\frac{\alpha \vdash A \quad \beta(A) \vdash B}{\beta(\alpha) \vdash B},$$

where  $\beta(\alpha)$  is the result of replacing arbitrarily many occurrences of  $A$  in  $\beta(A)$  by  $\alpha$  if  $\alpha$  is non-empty, and otherwise by  $t$ .<sup>42</sup> Without this emendation of the Cut rule one could derive  $B \vdash A$  whenever  $\vdash A$  is derivable (for arbitrary  $B$ , relevant or not) as follows

$$\frac{\frac{A \vdash A}{\vdash A} \text{ Thinning}}{\frac{A, B \vdash A}{B \vdash A} \text{ Cut}}$$

Discussing decidability a bit, one problem seems to be that Kripke’s Lemma (appropriately modified) is just plain false. The following is a sequence of cognate sequents in just the two propositional variables  $X$  and  $Y$  which is irredundant in the sense that structural rules will not get you from a later member to an earlier member:

$$\frac{X; Y \vdash X \quad (X; Y), X \vdash X \quad ((X; Y), X); Y \vdash X \dots}{\dots}^{43}$$

<sup>42</sup>Considerations about the eliminability of occurrences of  $t$  are then needed to show the admissibility of *modus ponens*. This was at least the plan of [Dunn, 1974]. A different plan is to be found in [Anderson and Belnap, 1975, Section 28.5], where things are arranged so that sequents are never allowed to have empty left-hand sides (they have  $t$  there instead).

#### 4.10 Systems Without Contraction

Gentzen systems without the contraction rule tend to be more amenable to decision procedures than those with it. Clearly, all of the work in Kripke's Lemma is in keeping contraction under control. So it comes as no surprise that if we consider systems without contraction for intensional structure, decision procedures are forthcoming. If we remove the contraction rule from **LR** we get the system which has been known as **LRW** (**R** without **W** without distribution), and which is equivalent to the additive and multiplicative fragment of Girard's linear logic [Girard, 1987]. It is well known that this system is decidable. In the Gentzen system, define the *complexity* of a sequent to be the number of connectives and commas which appear in it. It is trivial to show that complexity never increases in a proof and that as a result, from any given sequent there are only a finite number of sequents which could appear in a proof of the original sequent (if there is one). This gives rise to a simple decision procedure for the logic. (Once the work has already been done in showing that Cut is eliminable.)

If we add the extensional structure which appears in the proof theories of traditional relevance logics then the situation becomes more difficult. However, work by Giambrone has shown that the Gentzen systems for positive relevance logics without contraction do in fact yield decision procedures [Giambrone, 1985]. In these systems we do have extensional contraction, so such a simple minded measure of complexity as we had before will not yield a result. In the rest of this section we will sketch Giambrone's ideas, and consider some more recent extensions of them to include negation. For details, the reader should consult his paper. The results are also in the second volume of *Entailment* [Anderson *et al.*, 1992].

Two sequents are *equivalent* just when you can get from one to the other by means of the invertible structural rules (intensional commutativity, extensional commutativity, and so on). A sequent is *super-reduced* if no equivalent sequent can be the premise of a rule of extensional contraction. A sequent is *reduced* if for any equivalent sequents which are the premise of a rule of extensional contraction, the conclusion of that rule is super-reduced. So, intuitively, a super-reduced sequent has no duplication in it, and a reduced sequent can have one part of it 'duplicated', but no more. Clearly any sequent is equivalent to a super-reduced sequent. The crucial lemma is that any super-reduced sequent has a proof in which every sequent appearing is

---

<sup>43</sup>Further, this is not just caused by a paucity of structural rules. Interpreting the sequents of formulas of  $\mathbf{R}^+$  ( $\wedge$  for comma,  $\circ$  for semicolon,  $\rightarrow$  for  $\vdash$ ) no later formula provably implies an earlier formula. Incidentally, one does need at least two variables (cf. R. K. Meyer [1970b]).

reduced. This is clear, for given any proof you can transform it into one in which every sequent is reduced without too much fuss.

As a result, we have gained as much control over extensional contraction as we need. Giambrone is able to show that only finitely many reduced sequents can appear in the proof of a given sequent, and as a result, the size of the proof-search tree is bounded, and we have decidability. This technique does not work for intensional contraction, as we do not have the result that every sequent is equivalent to an intensionally super-reduced sequent, in the absence of the mingle rule. While  $A \wedge A \vdash B$  is equivalent to  $A \vdash B$ , we do not have the equivalence of  $A \circ A \vdash B$  and  $A \vdash B$ , without mingle.

These methods can be extended to deal with negation. Brady [1991] constructs out of *signed* formulae  $TA$  and  $FA$  instead of formulae alone, and this is enough to include negation without spoiling the decidability property. Restall [1998] uses the techniques of Belnap's *Display Logic* (see Section 5.2) to provide an alternate way of modelling negation in sequent systems. These techniques show that the decidability of systems without intensional contraction are decidable, to a large extent independently of the other properties of the intensional structure.

#### 4.11 Various Methods Used to Attack the Decision Problem

Decision procedures can basically be subdivided into two types: syntactic (proof-theoretic) and semantic (model-theoretical). A paradigm of the first type would be the use of Gentzen systems, and a paradigm of the second would be the development of the finite model property. It seems fair to say, looking over the previous sections, that syntactic methods have dominated the scene when nested implications have been afoot, and that semantical methods have dominated when the issue has been first-degree implications and first-degree formulas.<sup>44</sup>

There are two well-known model-theoretic decision procedures used for such non-classical logics as the intuitionistic and modal logics. One is due to McKinsey and Tarski and is appropriate to algebraic models (matrices) (cf. [Lemmon, 1966, p. 56 ff.]), and the other (often called 'filtration') is due to Lemmon and Scott and is appropriate to Kripke-style models (cf. [Lemmon, 1966, p. 208 ff]). Actually these two methods are closely connected (equivalent?) in the familiar situation where algebraic model and Kripke models are duals. The problem is that neither seems to work with **E** and **R**. The difficulty is most clearly stated with **R** as paradigm. For the algebraic models the problem is given a de Morgan monoid  $(M, \wedge, \vee, \neg, \circ)$

---

<sup>44</sup>As something like 'the exception that proves the rule' it should be noted that Belnap's [1967a] work on first-degree formulas and slightly more complex formulas has actually been a subtle blend of model-theoretic (algebraic) and proof-theoretic methods.

and a finite de Morgan sublattice  $(D, \wedge', \vee, -')$ , how to define a new multiplicative operation  $\circ'$  on  $D$  so as to make it a de Morgan monoid and so for  $x, y \in D$ , if  $x \circ y \in D$  then  $x \circ y = x \circ' y$ . the chief difficulty is in satisfying the associative law. For the Kripke-style models (say the Routley–Meyer variety) the problem is more difficult to state (especially if the reader has skipped Section 3.7) but the basic difficulty is in the satisfying of certain requirements on the three-placed accessibility relation once set-ups have been identified into a finite number of equivalence classes by ‘filtration’. Thus, e.g. the requirement corresponding to the algebraic requirement of associativity is  $Raxy \ \& \ Rbcy \Rightarrow \exists y(Raby \ \& \ Rycx)$ <sup>45</sup> the problem in a nutshell is that after filtration one does not know that there exists the appropriate equivalence class  $\bar{y}$  needed to feed such an existentially hungry postulate.

The McKinsey-Tarski method has been used successfully by Maksimova [1967] with respect to a subsystem of **R**, which differs essentially only in that it replaces the nested form of the transitivity axiom

$$(A \rightarrow B) \rightarrow [(B \rightarrow C) \rightarrow (A \rightarrow C)]$$

by the ‘conjoined’ form

$$(A \rightarrow B) \wedge (B \rightarrow C) \rightarrow (A \rightarrow C).$$
<sup>46</sup>

Perhaps the most striking positive solution to the decision problem for a relevance logic is that provided for **RM** by Meyer (see [Anderson and Belnap, 1975, Section 29.3], although the result was first obtained by Meyer [1968]).<sup>47</sup> Meyer showed that a formula containing  $n$  propositional variables is a theorem of **RM** iff it is valid in the ‘Sugihara matrix’ defined on the non-zero integers from  $-n$  to  $+n$ . this result was extended by [Dunn, 1970] to show that every ‘normal’ extension of **RM** has some finite Sugihara matrix (with possibly 0 as an element) as a characteristic matrix. So clearly **RM** and its extensions have at least the finite model property. Cf. Section 3.10 for further information about **RM**.

Meyer [private communication] has thought that the fact that the decidability of **R** is equivalent to the solvability of the word problem for de Morgan monoids suggests that **R** might be shown to be undecidable by some suitable modification of the proof that the word problem for monoids is unsolvable. It turns out that this is technique is the one which pays off — although the proof is very complex. The complexity arises because there is an important disanalogy between monoids and de Morgan monoids in

<sup>45</sup>This is suggestively written (following Meyer) as  $Ra(bc)x \Rightarrow R(ab)cx$ .

<sup>47</sup>In fact neither McKinsey-Tarski methods nor filtration was used in this proof. We are no clearer now that they could not be used, and we think the place to start would be to try to apply filtration to the Kripke-style semantics for **RM** of [Dunn, 1976b], which uses a *binary* accessibility relation and seems to avoid the problems caused by ‘existentially hungry axioms’ for the ternary accessibility relation.

that in the latter the multiplicative operation is necessarily commutative (and the word problem for commutative monoids *is* solvable).<sup>48</sup> Still it has occurred to both Meyer and Dunn that it might be possible to define a new multiplication operation  $\times$  for both  $\circ$  and  $\wedge$  in such a way as to embed the free monoid into the free de Morgan monoid. This suspicion has turned out to be right, as we shall see in the next section.

#### 4.12 $\mathbf{R}$ , $\mathbf{E}$ and Related Systems

As is quite well known by now, the principal systems of relevance logic,  $\mathbf{R}$ ,  $\mathbf{E}$  and others, are undecidable. Alasdair Urquhart proved this in his ground breaking papers [Urquhart, 1983; Urquhart, 1984]. We have recounted earlier attempts to come to a conclusion on the decidability question. The insights that helped decide the issue came from an unexpected quarter — projective geometry. To see why projective geometry gave the necessary insights, we will first consider a simple case, the undecidability of the system  $\mathbf{KR}$ .  $\mathbf{KR}$  is given by adding  $A \wedge \neg A \rightarrow B$  to  $\mathbf{R}$ . A  $\mathbf{KR}$  frame is one satisfying the following conditions (given by adding the clause that  $a = a^*$  to the conditions for an  $\mathbf{R}$  frame).

$$\begin{array}{l} R0ab \text{ iff } a = b \quad Rabc \text{ iff } Rbac \text{ iff } Racb \text{ (total permutation)} \\ Raaa \text{ for each } a \quad R^2abcd \text{ only if } R^2acbd \end{array}$$

The clauses for the connectives are standard, with the proviso that  $a \vDash \neg A$  iff  $a \not\vDash A$ , since  $a = a^*$ .

Urquhart's first important insight was that  $\mathbf{KR}$  frames are quite like projective spaces. A *projective space*  $\mathcal{P}$  is a set  $P$  of points, and a collection  $L$  of subsets of  $P$  called *lines*, such that any two distinct points are on exactly one line, and any two distinct lines intersect in exactly one point. But we can define projective spaces instead through the ternary relation of collinearity. Given a projective space  $\mathcal{P}$ , its collinearity relation  $C$  is a ternary relation satisfying the condition:

$$Cabc \text{ iff } a = b = c, \text{ or } a, b \text{ and } c \text{ are distinct and they lie on a common line.}$$

If  $\mathcal{P}$  is a projective space, then its collinearity relation  $C$  satisfies the following conditions,

$$Caaa \text{ for each } a. \quad Cabc \text{ iff } Cbac \text{ iff } Cacb. \quad C^2abcd \text{ only if } C^2acbd.$$

---

<sup>48</sup>In this connection two things should be mentioned. First, Meyer [unpublished typescript, 1973] has shown that not all finitely generated de Morgan monoids are finitely presentable. Second, Meyer and Routley [1973c] have constructed a positive relevance logic  $\mathbf{Q}^+$  (the algebraic semantics for which dispenses with commutativity) and shown it undecidable.

provided that every line has at least four points (this last requirement is necessary to verify the last condition). Conversely, if we have a set with a ternary relation  $C$  satisfying these conditions, then the space defined with the original set as points and the sets  $l_{ab} = \{c : Cab\} \cup \{a, b\}$  where  $a \neq b$  as lines is a projective space.

Now the similarity with **KR** frames becomes obvious. If  $\mathcal{P}$  is a projective space, the frame  $\mathcal{F}(\mathcal{P})$  generated by  $\mathcal{P}$  is given by adjoining a new point 0, adding the conditions  $C0aa$ ,  $Ca0a$ , and  $Caa0$ , and by taking the extended relation  $C$  to be the accessibility relation of the frame.

Projective spaces have a naturally associated undecidable problem. The problem arises when considering the *linear subspaces* of projective spaces. A subspace of a projective space is a subset which is also a projective space under its inherited collinearity relation. Given any two linear subspaces  $\mathcal{X}$  and  $\mathcal{Y}$ , the subspace  $\mathcal{X} + \mathcal{Y}$  is the set of all points on lines through points in  $\mathcal{X}$  and points in  $\mathcal{Y}$ .

In **KR** frames there are propositions which play the role of linear subspaces in projective spaces. We need a convention to deal with the extra point 0, and we simply decree that 0 should be *in* every “subspace.” Then linear subspaces are equivalent to the *positive idempotents* in a frame. That is, they are the propositions  $X$  which are *positive* (so  $0 \in X$ ) and *idempotent* (so  $X = X \circ X$ ). Clearly for any sentence  $A$  and any **KR** model  $\mathcal{M}$ , the extension of  $A$ ,  $\|A\|$  in  $\mathcal{M}$  is a positive idempotent iff  $0 \models A \wedge (A \leftrightarrow A \circ A)$ . It is then not too difficult to show that if  $A$  and  $B$  are positive idempotents, so are  $A \circ B$  and  $A \wedge B$ , and that  $t$  and  $\top$  are positive idempotents.

Given a projective space  $\mathcal{P}$ , the lattice algebra  $\langle \mathcal{L}, \cap, + \rangle$  of all linear subspaces of the projective space, under intersection and  $+$  is a modular geometric lattice. That is, it is a complete lattice, satisfying these conditions:

**Modularity**  $a \geq c \Rightarrow (\forall b)(a \cap (b + c) \leq (a \cap b) + c)$

**Geometricity** Every lattice element is a join of atoms, and if  $a$  is an atom and  $X$  is a set where  $a \leq \Sigma X$  then there's some finite  $Y \subseteq X$ , where  $a \leq \Sigma Y$ .

The lattice of linear subspaces of a projective space satisfies these conditions, and that in fact, any modular geometric lattice is isomorphic to the lattice of linear subspaces of some projective space. Furthermore the lattice of positive idempotents of any **KR** frame is also a modular geometric lattice.

The undecidable problem which Urquhart uses to prove the undecidability of **KR** is now simple to state. Hutchinson [1973] and Lipshitz [1974] proved that

The word problem for any class of modular lattices which includes the subspace lattice of an infinite dimensional projective space is undecidable.

Given an infinite dimensional projective space in which every line includes at least four points  $\mathcal{P}$ , the logic of the frame  $(\mathcal{P})$  is said to be a *strong* logic. Our undecidability theorem then goes like this:

Any logic between **KR** and a strong logic is undecidable.

The proof is not too difficult. Consider a modular lattice problem

$$\text{If } v_1 = w_1 \dots v_n = w_n \text{ then } v = w$$

stated in a language with variables  $x_i$  ( $i = 1, 2, \dots$ ) constants 1 and 0, and the lattice connectives  $\cap$  and  $+$ . Fix a map into the language of **KR** by setting  $x_i^t = p_i$  for variables,  $0^t = t$ ,  $1^t = \top$ ,  $(v \cap w)^t = v^t \wedge w^t$  and  $(v + w)^t = v^t \circ w^t$ . The translation of our modular lattice problem is then the **KR** sentence

$$(B \wedge (v_1^t \leftrightarrow w_1^t) \wedge \dots \wedge (v_n^t \leftrightarrow w_n^t) \wedge t) \rightarrow (v^t \leftrightarrow w^t)$$

where the sentence  $B$  is the conjunction of all sentences  $p_i \wedge (p_i \leftrightarrow p_i \circ p_i)$  for each  $p_i$  appearing in the formulae  $v_j^t$  or  $w_j^t$ .

We will show that given a particular infinite dimensional projective space (with every line containing at least four points)  $\mathcal{P}$ , then the word problem is valid in the lattice of linear subspaces of  $\mathcal{P}$  if and only if its translation is provable in  $L$ , for any logic  $L$  intermediate between **KR** and the logic of the frame  $\mathcal{F}(\mathcal{P})$ .

If the translation of the word problem is valid in  $L$ , then it holds in the frame  $\mathcal{F}(\mathcal{P})$ . Consider the word problem. If it were invalid, then there would be linear subspaces  $x_1, x_2, \dots$  in the space  $\mathcal{P}$  such that each  $v_i = w_i$  would be true while  $v \neq w$ . Construct a model on the frame  $\mathcal{F}(\mathcal{P})$  as follows. Let the extension of  $p_i$  be the space  $x_i$  together with the point 0. It is then simple to show that  $0 \vDash B$ , as each  $p_i$  is a positive idempotent. In addition,  $0 \vDash t$ , and  $0 \vDash v_i^t \leftrightarrow w_i^t$ , for the extension of each  $v_i^t$  and  $w_i^t$  will be the spaces picked out by  $v_i$  and  $w_i$  (both with the obligatory 0 added). However, we would have  $0 \not\vDash v^t \leftrightarrow w^t$ , since the extensions of  $v^t$  and  $w^t$  were picked out to differ. This would amount to a counterexample to the translation of the word problem, which we said was valid. As a result, the word problem is valid in the space  $\mathcal{P}$ . The converse reasoning is similar.

Unfortunately, these techniques do not work for systems weaker than **KR**. The proof that positive idempotents are modular uses essentially the special properties of **KR**. Not every positive idempotent in **R** need be modular. But nonetheless, the techniques of the proof can be extended to apply to a much wider range of systems. Urquhart examined the structure of the modular lattice undecidability result, and he showed that you could make do with much less. You do not need to restrict your attention to modular lattices to construct an undecidable word problem. But to do that, you need to examine Lipshitz and Hutchinson's proof more carefully. In the rest of



this section, we will sketch the structure of Urquhart's undecidability proof. The techniques are quite involved, so we do not have the space to go into detail. For that, the reader is referred to Urquhart [1984].

Lipshitz and Hutchinson proved that the word problem for modular lattices was undecidable by embedding into that problem the already known undecidable word problem for semigroups. It is enough to show that a structure can define a "free associative binary operation", for then you will have the tools for representing arbitrary semigroup problems. (A semigroup is a set with an associative binary operation. An operation is a "free associative" operation if it satisfies those conditions satisfied by any associative operation but no more.) We will sketch how this can be done without resorting to a modular lattice.

The required structure is what is called a 0-structure, and a modular 4-frame defined within a 0-structure. An 0-structure is a set equipped with the following structure

- A semilattice with respect to  $\sqcap$ .
- With a binary operator  $+$  which is associative and commutative.
- And  $x \leq y \Rightarrow x + z \leq y + z$ .
- $0 + x = x$ .
- $y \geq 0 \Rightarrow x \sqcap (x + y) = x$ .

A 4-frame in a 0-structure is a set  $\{a_1, a_2, a_3, a_4\} \cup \{c_{ij} : i \neq j, i, j = 1, \dots, 4\}$  such that

- The  $a_i$ s are *independent*. If  $G, H \subseteq \{a_1, \dots, a_4\}$  then  $(\Sigma G) \sqcap (\Sigma H) = \Sigma(G \cap H)$  (where  $\Sigma \emptyset = 0$ )
- If  $G \subseteq \{a_1, \dots, a_4\}$  then  $\Sigma G$  is modular
- $a_i + a_i = a_i$
- $c_{ij} = c_{ji}$
- $a_i + a_j = a_i + c_{ik}; c_{ij} \sqcap a_j = 0$ , if  $i \neq j$
- $(a_i + a_k) \sqcap (c_{ij} + c_{jk}) = c_{ik}$  for distinct  $i, j, k$

Now, we are nearly at the point where we can define a semigroup structure. First, for each distinct  $i, j$ , we define the set  $L_{ij}$  to be  $\{x : x + a_j = a_i + a_j \text{ and } x \sqcap a_j = 0\}$ . Then if  $b \in L_{ij}$  and  $d \in L_{jk}$  where  $i, j, k$  are distinct, then we set  $b \otimes d = (b + d) \sqcap (a_i + a_k)$ , and it is not difficult to show that  $b \otimes d \in L_{ik}$ . Then, we can define a semigroup operation ' $\cdot$ ' on  $L_{12}$  by setting

$$x \cdot y = (x \otimes c_{23}) \otimes (c_{31} \otimes y)$$

Now it is quite an involved operation to show that this is in fact an associative operation, but it can be done. And in fact, in certain circumstances, the operation is a free associative operation. Given a countably infinite-dimensional vector space  $\mathcal{V}$ , its lattice of subspaces is a 0-structure, and it is possible to define a modular 4-frame in this lattice of subspaces, such that any countable semigroup is isomorphic to a subsemigroup of  $L_{12}$  under the defined associative operation. (Urquhart gives the complete proof of this result [Urquhart, 1984].)

The rest of the work of the undecidability proof involves showing that this construction can be modelled in a logic. Perhaps surprisingly, it can all be done in a weak logic like  $\mathbf{TW}[\wedge, \vee, \rightarrow, \top, \perp]$ . We can do without negation by picking out a distinguished propositional atom  $f$ , and by defining  $\neg A$  to be  $A \rightarrow f$ ,  $t$  to be  $\neg f$ , and  $A : B$  to be  $\neg(A \rightarrow \neg B)$ .  $A$  is a *regular* proposition iff  $\neg\neg A \leftrightarrow A$  is provable. The regular propositions form an 0-structure, under the assumption of the formula  $\Theta = \{R(t, f, \top, \perp), N(t, f, \top, \perp), \neg\top \leftrightarrow \perp\}$ , where  $R(A)$  is  $\neg\neg A \leftrightarrow A$ ,  $N(A)$  is  $(t \rightarrow A) \rightarrow A$ , and  $R(A, B, \dots)$  is  $R(A) \wedge R(B) \wedge \dots$  and similarly for  $N$ . In other words, we can show that the conditions for an 0-structure hold in the regular propositions, assuming  $\Theta$  as an extra premise. To interpret the 0-structure conditions we interpret  $\sqcap$  by  $\wedge$ ,  $+$  by  $:$  and 0 by  $t$ .

Now we need to model a 4-frame in the 0-structure. This can be done as we get just the modularity we need from another condition which is simple to state. Define  $K(A)$  to be  $R(A) \wedge (A \wedge \neg A \leftrightarrow \perp) \wedge (A \vee \neg A \leftrightarrow \top) \wedge (A : \neg A \leftrightarrow \neg A) \wedge (A \leftrightarrow A : A)$ . Then we can show the following

$$K(A), R(B, C), C \rightarrow A \vdash A \wedge (B : C) \leftrightarrow (A \wedge B) : C$$

In other words, if  $K(A)$ , then  $A$  is modular in the class of regular propositions. Then the conditions for a 4-frame are simple to state. We pick out our atomic propositions  $A_1, \dots, A_4$  and  $C_{12}, \dots, C_{34}$  which will do duty for  $a_1, \dots, a_4$  and  $c_{12}, \dots, c_{34}$ . Then, for example, one independence axiom is

$$(A_1 : A_2 : A_3) \wedge (A_2 : A_3 : A_4) \leftrightarrow (A_2 : A_3)$$

and one modularity condition is

$$K(A_1 : A_3 : A_4)$$

We will let  $\Pi$  be the conjunction of the statements that express that the propositions  $A_i$  and  $C_{ij}$  form a 4-frame in the 0-structure of regular propositions. So,  $\Theta \cup \Pi$  is a finite (but complex) set of propositions. In any algebra in which  $\Theta \cup \Pi$  is true, the lattice of regular propositions is a 0-structure, and the denotations of the propositions  $A_i$  and  $C_{ij}$  form a 4-frame. Finally, when coding up a semigroup problem with variables  $x_1, x_2, \dots, x_m$ , we will need formulae in the language which do duty for these variables. Thus we

need a condition which picks out the fact that  $p_i$  (standing for  $x_i$ ) is in  $L_{12}$ . We define  $L(p)$  to be  $(p : A_2 \leftrightarrow A_1 : A_2) \wedge (p \wedge A_2 \leftrightarrow t)$ . Then the semigroup operation on elements of  $L_{12}$  can be defined in terms of  $\wedge$  and  $:$  and the formulae  $A_i$  and  $C_{ij}$ . We assume that done, and we will simply take it that there is an operation  $\cdot$  on formulae which picks out the algebraic operation on  $L_{12}$ . This is enough for us to sketch the undecidability argument.

The deducibility problem for any logic between  $\mathbf{TW}[\wedge, \vee, \rightarrow, \top, \perp]$  and  $\mathbf{KR}$  is undecidable.

Take a semigroup problem which is known to be undecidable. It may be presented in the following way

$$\text{If } v_1 = w_1 \dots v_n = w_n \text{ then } v = w$$

where each term  $v_i, w_i$  is a term in the language of semigroups, constructed out of the variables  $x_1, x_2, \dots, x_m$  for some  $m$ . The translation of that problem into the language of  $\mathbf{TW}[\wedge, \vee, \rightarrow, \top, \perp]$  is the deducibility problem

$$\Theta, \Pi, L(p_1, \dots, p_m), v_1^t \leftrightarrow w_1^t, \dots, v_n^t \leftrightarrow w_n^t \vdash v^t \leftrightarrow w^t$$

where each the translation  $u^t$  of each term  $u$  is defined recursively by setting  $x_i^t$  to be  $p_i$ , and  $(u_1.u_2)^t$  to be  $u_1^t \cdot u_2^t$ .

Now the undecidability result will be immediate once we show that for any logic between  $\mathbf{TW}$  and  $\mathbf{KR}$  the word problem in semigroups is valid if and only if its translation is valid in that logic.

For left to right, if the word problem is valid in the theory of semigroups, its translation must be valid, for given the truth of  $\Theta$  and  $\Pi$  and  $L(p_1, \dots, p_m)$ , the operator  $\cdot$  is provably a semigroup operation on the propositions in  $L_{12}$  in the algebra of the logic, and the terms  $v_i$  and  $w_i$  satisfy the semigroup conditions. As a result, we must have  $v^t$  and  $w^t$  picking out the same propositions, hence we have a proof of  $v^t \leftrightarrow w^t$ .

Conversely, if the word problem is invalid, then it has an interpretation in the semigroup  $\mathcal{S}$  defined on  $L_{12}$  in the lattice of subspaces of an infinite dimensional vector space. The lattice of subspaces of this vector space is the 0-structure in our countermodel. However, we need a countermodel for our — the 0-structure is not a model of the whole of the logic, since it just models the regular propositions. How can we construct this? Consider the argument for  $\mathbf{KR}$ . There, the subspaces were the positive idempotents in the frame. The other propositions in the frame were *arbitrary* subsets of points. Something similar can work here. On the vector space, consider the subsets of points which are closed under multiplication (that is, if  $x \in \alpha$ , so is  $kx$ , where  $k$  is taken from the field of the vector space). This is a De Morgan algebra, defining conjunction and disjunction by means of intersection and union as is usual. Negation is modelled by set difference. The fusion  $\alpha \circ \beta$  of two sets of points is the set  $\{x + y : x \in \alpha \text{ and } y \in \beta\}$ . It is not too difficult

to show that this is commutative and associative, and square increasing, when the vector space is in a field of characteristic other than 2, since if  $x \in \alpha$  then  $x = \frac{1}{2}x + \frac{1}{2}x \in \alpha \circ \alpha$ . Then  $\alpha \rightarrow \beta$  is simply  $-(\alpha \circ -\beta)$ . It is not too difficult to show that this is an algebraic model for **KR**, and that the regular propositions in this model are exactly the subspaces of the vector space. It follows that our counterexample in the 0-structure is a counterexample in a model of **KR** to the translation of the word problem. As a result, the translation is not provable in **KR** or in any weaker logic.

This result applies to systems between **TW** and **KR**, and it shows that the deducibility problem is undecidable for any of these systems. In the presence of the *modus ponens* axiom  $A \wedge (A \rightarrow B) \wedge t \rightarrow B$ , this immediately yields the undecidability of the *theoremhood* problem, as the deducibility problem can be rewritten as a single formula.

$$(\Theta \wedge \Pi \wedge L(p_1, \dots, p_m) \wedge (v_1^t \leftrightarrow w_1^t) \wedge \dots \wedge (v_n^t \leftrightarrow w_n^t) \wedge t) \rightarrow (v^t \leftrightarrow w^t)$$

As a result, the theoremhood problem for logics between **T** and **KR** is undecidable. In particular, **R**, **E** and **T** are all undecidable.

The restriction to **TW** is necessary in the theorem. Without the prefixing and suffixing axioms, you cannot show that the lattice of regular propositions is closed under the ‘fusion-like’ connective ‘:’.

Before moving on to our next section, let us mention that these geometric methods have been useful not only in proving the undecidability of logics, but also in showing that interpolation fails in **R** and related logics [Urquhart, 1993].

## 5 LOOKING ABOUT

A lot of the work in relevance logics taking place in the late 1980’s and in the 1990’s has not focussed on Anderson’s core problems. Now that these have been more or less resolved, work has proceeded apace in other directions. In this section we will give an undeniably indiosyncratic and personal overview of what we think are some of the strategic directions of this recent research. The first two sections in this part deal with generalisations — first of semantics, and second of proof theory — which situate relevance logic into a wider setting. The next sections deal with neighbouring formal theories, and we end with one philosophical application of the machinery of relevance logics.

### 5.1 Gaggle Theory

The fusion connective  $\circ$  has played an important part in the study of relevance logics. This is because fusion and implication are tied together by

the *residuation condition*

$$a \leq b \rightarrow c \text{ iff } a \circ b \leq c$$

In addition, in the frame semantics, fusion and implication are tied to the same ternary relation  $R$ , implication with the universal condition and fusion with the existential condition.

This is an instance of a generalised Galois connection. Galois studied connections between functions on partially ordered sets. A Galois connection between two partial orders  $\leq$  on  $A$  and  $\leq'$  on  $B$  is a pair of functions  $f : A \rightarrow B$  and  $g : B \rightarrow A$  such that

$$b \leq' f(a) \text{ iff } a \leq g(b)$$

The condition tying together fusion and implication is akin to that tying together  $f$  and  $g$  for Galois. So, gaggle theory (for 'ggl': *generalised Galois logic*) studies these connections in their generality, and it turns out that relevance logics like  $\mathbf{R}$ ,  $\mathbf{E}$  and  $\mathbf{T}$  are a part of a general structure which not only includes other relevance logics, but also traditional modal logics, Jónsson and Tarski's Boolean algebras with operators [Jónsson and Tarski, 1951] and many other formal systems. Dunn has shown that if a logic has a family of  $n$ -ary connectives which are tied together with a generalised Galois connection, then the logic has a frame semantics in which those connectives are modelled using the one  $n+1$ -ary relation, in the way that fusion and implication are modelled by the same ternary relation in relevance logics [Dunn, 1991; Dunn, 1993a; Dunn, 1994].

In general, an  $n$ -ary connective  $f$  has a *trace*  $(\tau_1, \dots, \tau_n) \mapsto +$  if

- $f(c_1, \dots, \mathbf{1}, \dots, c_n) = \mathbf{1}$ , if  $\tau_i = +$  (where the  $\mathbf{1}$  is in position  $i$ ).
- $f(c_1, \dots, \mathbf{0}, \dots, c_n) = \mathbf{1}$ , if  $\tau_i = -$  (where the  $\mathbf{0}$  is in position  $i$ ).
- If  $a \leq b$ , and if  $\tau_i = +$  then  $f(c_1, \dots, a, \dots, c_n) \leq f(c_1, \dots, b, \dots, c_n)$ .
- If  $a \leq b$ , and if  $\tau_i = -$  then  $f(c_1, \dots, b, \dots, c_n) \leq f(c_1, \dots, a, \dots, c_n)$ .

We write this as  $T(f) = (\tau_1, \dots, \tau_n) \mapsto +$ . On the other hand, the connective  $f$  has *trace*  $(\tau_1, \dots, \tau_n) \mapsto -$  if

- $f(c_1, \dots, \mathbf{1}, \dots, c_n) = \mathbf{0}$ , if  $\tau_i = +$  (where the  $\mathbf{0}$  is in position  $i$ ).
- $f(c_1, \dots, \mathbf{0}, \dots, c_n) = \mathbf{0}$ , if  $\tau_i = -$  (where the  $\mathbf{1}$  is in position  $i$ ).
- If  $a \leq b$ , and if  $\tau_i = +$  then  $f(c_1, \dots, b, \dots, c_n) \leq f(c_1, \dots, a, \dots, c_n)$ .
- If  $a \leq b$ , and if  $\tau_i = -$  then  $f(c_1, \dots, a, \dots, c_n) \leq f(c_1, \dots, b, \dots, c_n)$ .

We write this as  $T(c) = (\tau_1, \dots, \tau_n) \mapsto -$ . Here are a few examples of traces of connectives. Conjunction-like connectives tend to be  $(-, -) \mapsto -$ , disjunction-like connectives tend to be  $(+, +) \mapsto +$ , necessity-like connectives tend to be  $+ \mapsto +$ , possibility-like connectives tend to be  $- \mapsto -$ , and negations can be either  $+ \mapsto -$  or  $- \mapsto +$  (and in many cases they are both).

Now we are nearly able to state the abstract law of residuation. First, we define  $S(f, a_1, \dots, a_n, b)$  as follows. If  $T(f) = (\dots) \mapsto +$ , then  $S(f, a_1, \dots, a_n, b)$  is the condition  $f(a_1, \dots, a_n) \leq b$ . If, on the other hand,  $T(f) = (\dots) \mapsto -$ , then  $S(f, a_1, \dots, a_n, b)$  is  $b \leq f(a_1, \dots, a_n)$ . Then, two connectives  $f$  and  $g$  are *contrapositives in place  $j$*  iff, if  $T(f) = (\tau_1, \dots, \tau_j, \dots, \tau_n) \mapsto \tau$ , then  $T(g) = (\tau_1, \dots, -\tau, \dots, \tau_n) \mapsto -\tau_j$ . (Where we define  $-+$  as  $-$  and  $--$  as  $+$ .) Two operators  $f$  and  $g$  satisfy the *abstract law of residuation* iff  $f$  and  $g$  are contrapositives in place  $j$ , and  $S(f, a_1, \dots, a_j, \dots, a_n, b)$  iff  $S(g, a_1, \dots, b, \dots, a_n, a_j)$ .

A collection of connectives in which there is some connective  $f$  such that every element of the collection satisfies the abstract law of residuation with  $f$ , is called a *founded family* of connectives. Dunn's major result is that if you have an algebra in which every connective is in a founded family, then the algebra is isomorphic to a subalgebra of the collection of propositions in a model in which each founded family of  $n$ -ary connectives shares an  $n + 1$ -ary relation. The soundness and completeness of the Routley–Meyer ternary relational semantics is for the implication-fusion fragment of relevance logics is an instance of this more general result.

The gaggle theoretic account of negation in relevance logics is interesting. We do not automatically get negation modelled by the Routley star — instead, being a unary connective, negation is modelled with a *binary* relation. One way negation can be modelled along gaggle theoretic lines is as follows. The De Morgan negation connective has trace  $- \mapsto +$ , so the gaggle theoretic result is that there is a binary relation  $C$  between set-ups such that

- $x \vDash \neg A$  iff for each  $y$  where  $xCy$ ,  $y \not\vDash A$

This is the general semantic structure which models negation connectives with trace  $- \mapsto +$ . Given a relation  $C$ , which we may read as ‘compatibility’, we can define another negation connective  $\sim$ , using  $C$ 's converse:

- $x \vDash \sim A$  iff for each  $y$  where  $yCx$ ,  $y \not\vDash A$

Then it follows that  $A \vdash \sim B$  iff  $B \vdash \neg A$ . For the De Morgan negation of relevance logics,  $\sim$  and  $\neg$  are the same, for the compatibility relation  $C$  is symmetric. But in more general settings, this need not hold.

The general perspective of gaggle theory not only opens up new formal systems to study — it also helps with interpreting the semantics. The

condition for  $\neg$  above can be read as follows:  $\neg A$  is true at  $x$  iff for each  $y$  compatible with  $x$ ,  $A$  is not true at  $y$ . This certainly sounds like a more palatable condition for negation than that using Routley star. We have an understanding of what it is for two set-ups (theories, worlds or situations) to be compatible, and the notion of compatibility is tied naturally to that of negation. Furthermore, the Routley star condition is an instance of this more general ‘compatibility’ condition. For any set-up  $a$ ,  $a^*$  can be seen as the set-up which ‘wraps up’ all set-ups compatible with  $a$ . We can argue whether there is such an all-encompassing set-up, but if there is, then the semantics for negation in terms of the compatibility relation is equivalent to that of the Routley star. And in addition, we have another means of explaining it.

Furthermore, once we have this generalised position from which to view negation, we can tinker with the binary accessibility relation in just the same way that modal logics are studied. Clearly if Boolean negation (written ‘ $\neg$ ’) is present, then  $\neg A$  is simply  $\Box\neg A$  for the positive modal operator  $\Box$  which uses  $C$  as its accessibility relation; and the study of these negation is dealt with using the techniques of modal logic. However, in relevance logics and other related systems, boolean negation is not present. And in this case the theory of negations arising from compatibility clauses like the one we have seen is a young and interesting subject in its own right. This perspective is pursued in Dunn [1994], and Restall [1999] develops a philosophical interpretation of the semantics.

## 5.2 Display Logic

Nuel Belnap has developed proof theoretical techniques which are quite similar to those from gaggle theory. Consider the general problem of providing a sequent calculus for logics like **R** and others. We have the choice of how to formulae sequents. If they are of the form  $X \vdash A$ , where  $X$  is a structured collection of formulae, and  $A$  is a formula, then we have the problem of how to state the introduction and elimination of negation rules in such a way as to make  $\neg\neg A$  equivalent to  $A$ . It is unclear how to do this while maintaining that the succedent of every sequent is a single formula. On the other hand, if we allow that sequents are of the form  $X \vdash Y$ , where now both  $X$  and  $Y$  are structured complexes of formulae, it is unclear how to state a cut rule which is both valid and admits of a cut-elimination proof in the style of Gentzen. If we are restricted to single formulae in the succedent position the rule is easy to state:

$$\frac{X \vdash A \quad Y(A) \vdash B}{Y(X) \vdash B}$$

but in the presence of multiple succedents it is unclear how to state the rule generally enough to be eliminable yet strictly enough to be valid under

interpretation. If there is only one sort of structuring in the consequent this might be possible, in the way used in the proof theories of classical or linear logic, for example:

$$\frac{X \vdash A, Y \quad X', A \vdash Y'}{X, X' \vdash Y, Y'}$$

But if we have  $X \vdash Y(A)$  and  $X'(A) \vdash Y'$  where the indicated instances of  $A$  are buried under multiple sorts of structure, then what is the appropriate conclusion of a cut rule?  $X'(X) \vdash Y(Y')$  will not do in general, for it is invalid in many instances. For example, in  $\mathbf{R}$  if  $Y(A)$  is  $A \wedge B$  and  $X'(A)$  is  $A \circ D$ , then we have  $A \wedge B \vdash A \wedge B$  and  $A \circ D \vdash A \circ D$ , but we don't have  $(A \wedge B) \circ D \vdash (A \circ D) \wedge B$  in general. (Consider the case where  $B = A$ .  $A \circ D$  needn't imply  $A$ .)

The alternative examined by Belnap is to make do with Cut where the cut formula is “displayed” in both premises of the rule.

$$\frac{X \vdash A \quad A \vdash Y}{X \vdash Y}$$

In order to get away with this, a system needs to be such that whenever you need to use a cut you can. The way Belnap does this is by requiring what he calls the “display condition”. The display condition is satisfied iff for every formula, every sequent including that formula is equivalent (using invertible rules) to one in which that formula is either the entire antecedent or the entire succedent of the sequent. For Belnap's original formulation, this is achieved by having a binary structuring connective  $\circ$  (not to be confused with the sentential connective  $\circ$ ) and a unary connective  $*$ . The display rules were as follows:

$$\begin{aligned} X \circ Y \vdash Z &\iff X \vdash *Y \circ Z \\ X \vdash Y \circ Z &\iff X \circ *Y \vdash Z \iff X \vdash Z \circ Y \\ X \vdash Y &\iff *Y \vdash *X \iff **X \vdash Y \end{aligned}$$

A structure is in antecedent position if it is in the left under an even number of stars, or in the right under an odd number of stars. If it is not in antecedent position, it is in succedent position. The star is read as negation, and the circle is read as conjunction in antecedent position, and disjunction in succedent position. The display postulates are a reworking of conditions like the residuation condition for fusion and implication. Here we have the conditions that  $a \circ b \leq c$  iff  $a \leq \sim b + c$  (where  $x + y$  is the *fission* of  $x$  and  $y$ ).

Belnap's system allows that different families of structural connectives can be used for different families of connectives in the language. For example, when  $\circ$  and  $*$  are read intensionally, we can have the following rules for



implication:

$$\frac{X \circ A \vdash B}{X \vdash A \rightarrow B} \quad \frac{X \vdash A \quad B \vdash Y}{A \rightarrow B \vdash *X \circ Y}$$

If the properties of  $\circ$  vary, so do the properties of the connective  $\rightarrow$ . We can give  $\circ$  properties of extensional conjunction in order to get a material conditional. Or conditions can be tightened, to give  $\rightarrow$  modal properties. It is clear that the family of structural connectives (here  $\circ$  and  $*$ ) act in analogously to accessibility relations on frames. However, the connections with gaggle theory run deeper, however. It can be shown a connective introduced in with rules without side conditions, and in a way which ‘mimics’ structural connectives (just as here  $A \rightarrow B$  mimics  $*X \circ Y$  in consequent position) must have a definable trace. Any implication satisfying those rules will have trace  $(-, +) \mapsto +$ , for example. For more details of this connection and a general argument, see Restall’s paper [Restall, 1995a].

Display logic gives these systems a natural cut-free proof theory, for Belnap has shown that under a broad set of conditions, any proof theory with this structure will satisfy cut-elimination. So again, just as with gaggle theory, we have an example of the way that the study of relevance logics like **R** and **E** have opened up into a more general theory of logics with similar structures.

### 5.3 Paraconsistency

Relevance logics are paraconsistent, in that argument forms such as  $A \wedge \neg A \vdash B$  are taken to be invalid. As a result, relevance logics have been seen to be important for the study of paraconsistent theories. [[See Priest’s article in this volume]]. Relevance logics are suited to applications for which a paraconsistent notion of consequence is needed however, not all logics are equal in this regard. For example, paraconsistentists have often considered the topic of naïve theories of sets and of truth (any predicate yields the set of things satisfying that predicate, the proposition  $p$  is true if and only if  $p$ ). With a relevance logic at hand, you can avoid the inference to triviality from contradictions such as that arising from the liar

This proposition is not true.

(from which you can deduce that it is true, and hence that it isn’t) and Russell’s paradox ( $\{x : x \notin x\}$  both is and is not a member of itself). However, the *Curried* forms of these paradoxes

If this proposition is true then there is a Santa Claus.

and  $\{x : (x \in x) \rightarrow P\}$  are more difficult to deal with. These yield arguments for the existence of Santa Claus and the truth of  $P$  (which was

arbitrary) in logics like  $\mathbf{R}$ , or any others with theorems related to the rule of contraction. The theoremhood of propositions such as  $(A \rightarrow (A \rightarrow B)) \rightarrow (A \rightarrow B)$  and  $A \wedge (A \rightarrow B) \rightarrow B$  rule out a logic for service in the cause of paraconsistent theories like these [Meyer *et al.*, 1979].

However, this has not deterred some hardier souls in considering weaker relevance logics which do not allow one to deduce triviality in these theories. Some work has been done to show that in some logics these theories are consistent, and in others, though inconsistent, not everything is a theorem [Brady, 1989].

Another direction of paraconsistency in which techniques of relevance logics have borne fruit is in the more computational area of reasoning with inconsistent information. The techniques of first degree entailment have found a home in the study of “bilattices” by Melvin Fitting and others, who seen in them a suitable framework for reasoning under the possibility of inconsistent information [Fitting, 1989].

#### 5.4 *Semantic Neighbours*

Another area in which research has grown in the recent years has been toward connections with other fields. It has turned out that seemingly completely unrelated fields have studied structures remarkably like those studied in relevance logics. These neighbours are helpful, not only for giving independent evidence for the fact that relevance logicians have been studying *something* worthwhile, but also because of the different insights they can bring to bear on theorising. In this section we will see just three of the neighbours which can shed light on work in relevance logics.

The first connection comes with Barwise and Perry’s situation semantics [1983]. For Barwise and Perry, utterances classify situations (parts of the world) which may be incomplete with regard to their semantic ‘content’. Consider the claim that Max saw Queensland win the Sheffield Shield”. How is this to be understood? For the Barwise and Perry of *Situations and Attitudes* [Barwise and Perry, 1983], this was to be parsed as expressing a relationship between Max and a *situation*, where a situation is simply a restricted part of the world. Situations are parts of the world and they support information. Max saw a situation and in this situation, Queensland won the Sheffield Shield. If, in this very situation, Queensland beat South Australia, then Max saw Queensland beat South Australia.

This shows why for this account situations have to be (in general) *restricted* bits of the world. The situation Max saw had better not be one in which Paul Keating lost the 1996 Federal Election, lest it follow from the fact that Max witnessed Queensland’s victory that he also witnessed Keating’s defeat, and surely *that* would be an untoward conclusion. Let’s denote this relationship between situations and the information they support as follows. We’ll abbreviate the claim that the situation  $s$  supports the

information that  $A$  by writing ' $s \models A$ ', and we'll write its negation, that  $s$  doesn't support the information that  $A$  by writing ' $s \not\models A$ '. This is standard in the situation theoretic literature. The information carried by these situations has, according to Barwise and Perry, a kind of logical coherence. For them, infons are closed under conjunction and disjunction, and  $s \models A \wedge B$  if and only if  $s \models A$  and  $s \models B$ , and  $s \models A \vee B$  if and only if  $s \models A$  or  $s \models B$ . However, negation is a different story — clearly situations don't support the traditional equivalence between  $s \models \neg A$  and  $s \not\models A$  (where  $\neg A$  is the negation of  $A$ ), for our situation witnessed by Max supports neither the infon "Keating won the 1996 election" nor its negation.

What to do? Well, Barwise and Perry suggest that negation interacts with conjunction and disjunction in the familiar ways —  $\neg(A \vee B)$  is (equivalent to)  $\neg A \wedge \neg B$ , and  $\neg(A \wedge B)$  is (equivalent to)  $\neg A \vee \neg B$ . And similarly,  $\neg\neg A$  is (equivalent to)  $A$ . This gives us a logic of sorts of negation — it is *first degree entailment*. Now for Barwise and Perry, there are no *actual* situations in which  $s \models A \wedge \neg A$  (the world is not self-contradictory). However, they agree that it is helpful to consider *abstract* situations which allow this sort of inconsistency. So, Barwise and Perry have an independent motivation for a semantic account of first-degree entailment. (More work has gone on to consider other connections between situation theory and relevance logics [Mares, 1997; Restall, 1994; Restall, 1995b].)

Another connection with a parallel field has come from completely different areas of research. The semantic structures of relevance logics have close cousins in the models for the Lambek Calculus and in Relation algebras. Let's consider relation algebras first.

A relation algebra is a Boolean algebra with some extra operations, a binary operation which denotes composition of relation, a unary operation  $\smile$ , for the converse of a relation, and a constant 1 for the identity relation. There is a widely accepted axiomatisation of the variety **RA** of relation algebras. A relation algebra is set  $R$  with operations  $\wedge, \vee, -, 1, \circ, \smile$  such that

- $\langle R, \wedge, \vee, - \rangle$  is a boolean algebra.
- $\smile$  is an automorphism on the algebra, satisfying  $a^{\smile\smile} = a$ ,  $(a \wedge b)^{\smile} = a^{\smile} \wedge b^{\smile}$ ,  $-(a^{\smile}) = (-a)^{\smile}$ .
- $\circ$  is associative, with a left and right identity 1, satisfying  $(a \vee b) \circ c = (a \circ c) \vee (b \circ c)$ ,  $a \circ (b \vee c) = (a \circ b) \vee (a \circ c)$ .
- $\smile$  and  $\circ$  are connected by setting  $(a \circ b)^{\smile} = b^{\smile} \circ a^{\smile}$ .

These conditions are satisfied by the class of relations on any base set (that is, by any *concrete* relation algebra). However, not every algebra satisfying these equations is isomorphic to a subalgebra of a concrete relation algebra.

These algebras are quite similar to de Morgan monoids. If we define  $\neg A$  to be  $\neg(a)^\smile$  or  $\neg(-a)^\smile$  then the conjunction, disjunction,  $\neg$ , 1 fragment is that of first degree entailment. We do not have  $a \leq b \vee \neg b$ , and nor do we have  $a \wedge \neg a \leq b$ . Consider the relation  $a$ :

$a$	$x$	$y$
$x$	1	1
$y$	0	1

Then  $\neg a$  is the following relation

$a$	$x$	$y$
$x$	0	1
$y$	0	0

So we don't have  $b \leq a \vee \neg a$  for every  $b$ , and nor do we have  $a \wedge \neg a \vee b$ . (However, we do have  $1 \leq a \vee \neg a$ .)

The class of relation algebras have a natural form of implication to go along with the fusionlike connective  $\circ$ . If we define  $a \rightarrow b$  to be  $\neg(\neg b \circ a)$ , then we have the residuation condition  $a \circ b \leq c$  iff  $a \leq b \rightarrow c$ . However, that is not the only implication-like connective we may define. If we set  $b \leftarrow a$  to be  $\neg(a \circ \neg b)$ , then  $a \circ b \leq c$  iff  $b \leq c \leftarrow a$ . Since  $\circ$  is not, in general, commutative, we have two residuals.

In logics like **R** this is not possible, for the left and the right residuals of fusion are the same connective. However, in systems in the vicinity of **E**, these implication operations come apart. This is mirrored by the behaviour on frames, since we can define  $B \leftarrow A$  by setting  $x \vDash B \leftarrow A$  iff for each  $y, z$  where  $Ryxz$  if  $y \vDash A$  then  $z \vDash B$ . This will be another residual for fusion, and it will not agree with  $\rightarrow$  in the absence of commutativity of  $R$  (if  $Rxyz$  then  $Ryxz$ ).<sup>49</sup>

It was hoped for some time that relation algebras would give an interesting model for logics like **R**. However, there does not seem to be a natural class of relations for which composition is commutative and square increasing. (The class of symmetric relations will not do. Even if  $a = a^\smile$  and  $b = b^\smile$ , it does not follow that  $a \circ b = b \circ a$ . You merely get that  $a \circ b = a^\smile \circ b^\smile = (b \circ a)^\smile$ .) Considered as a logic, **RA** is a sublogic of **R** (ignoring boolean negation for the moment). It is not a sublogic of **E**, since in **RA**,  $a = 1 \rightarrow a$ . Another difference between **RA** and typical relevance logics is the behaviour of contraposition. We do not have  $a \rightarrow b = \neg b \rightarrow \neg a$ . Instead,  $a \rightarrow b = \neg a \leftarrow \neg b$ .

A final connection between **RA** and relevance logics is in the issue of semantics. As we stated earlier, not all relation algebras are representable as subalgebras of concrete relation algebras. However, Dunn has shown

<sup>49</sup>We should flag here that in the relevance logic literature, [Meyer and Routley, 1972] seems to have been the first to consider both left- and right-residuals for fusion.

that all relation algebras *are* representable by algebras of propositions on a particular class of Routley–Meyer frames [Dunn, 1993b]. This is the first representation theorem for **RA**, and it shows that the semantical techniques of relevance logics have a wider scope than applications to **R**, **E** and their immediate neighbours.

In a similar vein, Dunn and Meyer [1997] have provided a Routley–Meyer style frame semantics for combinatory logic. The key idea here is that the ternary relation  $R$  satisfies *no* special conditions, but these properties are encoded by *combinators*, which are modelled by special propositions on frames.

Lambek’s *categorial grammar* is also similar to relevance logics, though this time it is introduced with frames, not algebras [Lambek, 1958; Lambek, 1961]. Here, the points in frames are pieces of syntax, and the ‘propositions’ are syntactic classifications of various kinds. For example, the classifications into noun phrases, verbs, and sentences. The interest comes with the way in which these classifications can be combined. For example  $A \circ B$  can be defined, where we say  $x \vDash A \circ B$  iff  $x$  is a concatenation of two strings  $y$  and  $z$ , where  $y \vDash A$  and  $z \vDash B$ . We can also define ‘slicing’ operations, setting  $x \vDash A \setminus B$  iff for each  $y$  where  $y \vDash A$ ,  $yx \vDash B$ ; and  $x \vDash B / A$  iff for each  $y$  where  $y \vDash A$ ,  $xy \vDash B$ . These are obviously analogues for  $\circ$  and  $\rightarrow$  in relevance logics, and again, we have a ‘left’ and ‘right’ residuals for fusion. In these frames  $Rxyz$  iff  $xy = z$ . So the Lambek calculus gives us an independently motivated interpretation of a class of Routley–Meyer frames. This connection has been explored by Kurtonina [1995], which is a helpful sourcebook of some recent work on ternary frames in connection with the Lambek calculus and related logics.

If you like, you can enrich the logic of strings with conjunction and disjunction, and if you do it in the obvious way (using the same clauses as in relevance logics) you get a formal logic quite like **RA** [Restall, 1994]. But more importantly, the conditions for conjunction and disjunction may be independently motivated. A string is of type  $A \vee B$  just when it is of type  $A$  or of type  $B$ . A string is of type  $A \wedge B$  just when it is of type  $A$  and of type  $B$ . The resulting logic is clearly interpretable, but it was a number of years before a proof theory was found for it. Here the techniques for the Gentzenisation for positive relevance logics are appropriate, and the proof theory can be found by utilising the proof theory for **R**<sup>+</sup>, and removing the commutativity and contraction of the intensional bunching operation. The resulting proof theory captures exactly the Lambek calculus enriched with conjunction and disjunction. In addition, the techniques of Giambrone show that the resulting logic is decidable [Restall, 1994].

### 5.5 *Linear Logic*

The burgeoning phenomenon of linear logic is one which has a number of formal similarities to relevance logics [Girard, 1987; Troelstra, 1992]. Linear logic is the study of systems in the vicinity of **LRW** (**R** without contraction, without distribution). This is proof-theoretically a very stable system. It is simple to show that it is decidable. Girard’s innovation, however, is to extend the proof theory with a modal operator  $!$  which allows intuitionistic logic to be modelled inside linear logic. This operation is given as follows, in single-succedent Gentzen systems.

$$\frac{X \vdash B}{X, !A \vdash B} \quad \frac{X; A \vdash B}{X; !A \vdash B} \quad \frac{!X \vdash B}{!X \vdash !B} \quad \frac{X, !A, !A \vdash B}{X, !A \vdash B}$$

Given this proof theory it is possible to show that  $A \Rightarrow B$  defined as  $!A \rightarrow B$  is an intuitionistic implication. This is similar to Meyer’s result that  $A \wedge t \rightarrow B$  is an intuitionistic implication in **R** (indeed,  $!A$  defined as  $A \wedge t$  satisfies each of the conditions for  $!$  above in **R**, but not in systems without contraction). However, nothing like it holds in relevance logics without contraction.

Linear logic also brings with it many new algebraic structures and models in category theory. None of these models have been mined to see if they can bring any ‘relevant’ insight. However, some transfer has gone on in the other direction — Allwein and Dunn [1993] have shown that the multiplicative and additive fragment of linear logic can be given a Routley–Meyer style semantics. This is not a simple job, as the absence of the distribution of (additive) conjunction over disjunction means that at least one of these connectives (in this case, disjunction) must take a non-standard interpretation.

### 5.6 *Relevant Predication*

There has been one major way in which relevance logics have been used in application to philosophical issues, and this application makes a good topic to end this article. The topic is Dunn’s work on relevant predication [Dunn, 1987].<sup>50</sup> The guiding idea is that a theory of relevant implication will give you some way of marking out the distinction between the way that Socrates’ wisdom is a property of Socrates, in the way that Socrates’ wisdom is not a property of Bill Clinton.

Classical first order logic is not good at marking out such a distinction, for if  $Wx$  stands for ‘ $x$  is wise’, and  $s$  stands for Socrates, and  $c$  stands for

<sup>50</sup>The reference [Dunn, 1987] is of course “Relevant Predication”: Of course all work has precursors, in this instance (largely unpublished) thoughts in the 1970’s by N. Belnap, J. Freeman, and most importantly R. K. Meyer and A. Urquhart (and Dunn). Cf. Sec. 9 of [Dunn, 1987] for some history.

Bill Clinton, then  $Wx$  is true of  $x$  iff it is wise, and  $(Ws \wedge x = x) \vee Ws$  is true of something iff Socrates is wise. Why is one a ‘real’ property and the other not? The guiding idea for relevant predication is the following distinction. It is true that if  $x$  is Socrates then  $x$  is wise. However, it is not true that if  $x$  is Bill Clinton then Socrates is wise. At least, it is plausible that this conditional fail, when read ‘relevantly’. This can be cashed out formally as follows.  $F$  is a *relevant property of  $a$*  (written  $(\rho xFx)a$ ) if and only if  $(\forall x)(x = a \rightarrow Fx)$ .

Given this definition, if  $F$  is a relevant property of  $a$  then  $Fa$  holds (quite clearly) and if  $F$  and  $G$  are relevant properties of  $a$  then so is their conjunction, and the disjunction of any relevant property with anything at all is still a relevant property.

Furthermore, one can define what it is for a relation to truly be a relation between objects. If  $Hx$  is ‘ $x$ ’s height is over 1 meter’, and  $Ly$  is ‘ $y$  is a logician’ then, it is true that Greg’s height is over 1 meter and Mike is a logician. However, it would be bizarre to hold that in this there is a real relation that holds between Greg and Mike because of this fact. We would have the following

$$\forall x \forall y (x = g \wedge y = m \rightarrow Hx \wedge Ly)$$

(assuming that  $(\rho xHx)g$  and  $(\rho yLy)m$ ) but it need not follow that

$$\forall x \forall y (x = g \rightarrow (y = m \rightarrow Hx \wedge Ly))$$

for there is no reason that  $Hx$  should follow from  $y = m$ , even given that  $x = g$  holds. There is no connection between ‘ $y$ ’s being  $m$ ’ and  $Hg$ . This latter proposition is a good candidate for expressing that there is a real relationship holding between  $g$  and  $m$ . In other words, we can define  $(\rho xyLxy)ab$  to be

$$\forall x \forall y (x = a \rightarrow (y = b \rightarrow Lxy))$$

to express the holding of a relevant relation. For more on relevant predication, consult Dunn’s series of papers [Dunn, 1987; Dunn, 1990a; Dunn, 1990b]

Relevance logics are very good at telling you what follows from what as a matter of logic — and in this case, the logical structure of relevant predication and relations. However, more work needs to be done to see in what it consists to say that a relevant implication is *true*. For that, we need a better grip on how to understand the models of relevance logics. It is our hope that this chapter will help people in this aim, and to bring the technique of relevance logics to a still wider audience.

## ACKNOWLEDGEMENTS

*Dunn's acknowledgements from the first edition:* I wish to express my thanks and deep indebtedness to a number of fellow toilers in the relevant vineyards, for information and discussion over the years. These include Richard Routley and Alasdair Urquhart and especially Nuel D. Belnap, Jr. and Robert K. Meyer, and of course Alan Ross Anderson, to whose memory I dedicate this essay. I also wish to thank Yong Auh for his patient and skilful help in preparing this manuscript, and to thank Nuel Belnap, Lloyd Humberstone, and Allen Hazen for corrections, although all errors and infelicities are to be charged to me.

*Our acknowledgements from the second edition:* Thanks to Bob Meyer, John Slaney, Graham Priest, Nuel Belnap, Richard Sylvan, Ed Mares, Rajeev Goré, Errol Martin, Chris Mortensen, Uwe Petersen and Pragati Jain for helpful conversations and correspondence on matters relevant to what is discussed here. Thanks too to Jane Spurr, who valiantly typed in the first edition of the article, to enable us to more easily create this version. This edition is dedicated to Richard Sylvan, who died while this essay was being written. Relevance (and *relevant*) logic has lost one of its most original and productive proponents.

J. Michael Dunn

*Indiana University*

Greg Restall

*Macquarie University*

## BIBLIOGRAPHY

- [Ackermann, 1956] Wilhelm Ackermann. Begründung einer strengen implikation. *Journal of Symbolic Logic*, 21:113–128, 1956.
- [Akama, 1997] Seiki Akama. Relevant counterfactuals and paraconsistency. In *Proceedings of the First World Conference on Paraconsistency, Gent, Belgium.*, 1997.
- [Allwein and Dunn, 1993] Gerard Allwein and J. Michael Dunn. A kripke semantics for linear logic. *Journal of Symbolic Logic*, 58(2):514–545, 1993.
- [Anderson and Belnap Jr., 1959a] A. R. Anderson and N. D. Belnap Jr. Modalities in Ackermann's 'rigorous implication'. *Journal of Symbolic Logic*, 24:107–111, 1959.
- [Anderson and Belnap Jr., 1959b] A. R. Anderson and N. D. Belnap Jr. A simple proof of Gödel's completeness theorem. *Journal of Symbolic Logic*, 24:320–321, 1959. (Abstract.).
- [Anderson and Belnap Jr., 1961] A. R. Anderson and N. D. Belnap Jr. Enthymemes. *Journal of Philosophy*, 58:713–723, 1961.
- [Anderson and Belnap Jr., 1962a] A. R. Anderson and N. D. Belnap Jr. The pure calculus of entailment. *Journal of Symbolic Logic*, 27:19–52, 1962.
- [Anderson and Belnap Jr., 1962b] A. R. Anderson and N. D. Belnap Jr. Tautological entailments. *Philosophical Studies*, 13:9–24, 1962.
- [Anderson and Belnap, 1975] Alan Ross Anderson and Nuel D. Belnap. *Entailment: The Logic of Relevance and Necessity*, volume 1. Princeton University Press, Princeton, 1975.



- [Anderson *et al.*, 1992] Alan Ross Anderson, Nuel D. Belnap, and J. Michael Dunn. *Entailment: The Logic of Relevance and Necessity*, volume 2. Princeton University Press, Princeton, 1992.
- [Anderson, 1960] A. R. Anderson. Entailment shorn of modality. *Journal of Symbolic Logic*, 25:388, 1960. (Abstract.).
- [Anderson, 1963] A. R. Anderson. Some open problems concerning the system  $E$  of entailment. *Acta Philosophica Fennica*, 16:7–18, 1963.
- [Avron, 1986] Arnon Avron. On purely relevant logics. *Notre Dame Journal of Formal Logic*, 27:180–194, 1986.
- [Avron, 1990a] Arnon Avron. Relevance and paraconsistency — a new approach. *Journal of Symbolic Logic*, 55:707–732, 1990.
- [Avron, 1990b] Arnon Avron. Relevance and paraconsistency — a new approach. part II: The formal systems. *Notre Dame Journal of Formal Logic*, 31:169–202, 1990.
- [Avron, 1990c] Arnon Avron. Relevance and paraconsistency — a new approach. part III: Cut-free gentzen-type systems. *Notre Dame Journal of Formal Logic*, 32:147–160, 1990.
- [Avron, 1991] Arnon Avron. Simple consequence relations. *Information and Computation*, 92:105–139, 1991.
- [Avron, 1992] Arnon Avron. Whither relevance logic? *Journal of Philosophical Logic*, 21:243–281, 1992.
- [Barcan Marcus, 1946] R. C. Barcan Marcus. The deduction theorem in a functional calculus of first-order based on strict implication. *Journal of Symbolic Logic*, 11:115–118, 1946.
- [Barwise and Perry, 1983] Jon Barwise and John Perry. *Situations and Attitudes*. MIT Press, Bradford Books, 1983.
- [Belnap and Dunn, 1981] Nuel D. Belnap and J. Michael Dunn. Entailment and the disjunctive syllogism. In F. Fløistad and G. H. von Wright, editors, *Philosophy of Language / Philosophical Logic*, pages 337–366. Martinus Nijhoff, The Hague, 1981. Reprinted as Section 80 in *Entailment* Volume 2, [Anderson *et al.*, 1992].
- [Belnap and Spencer, 1966] Nuel D. Belnap and J. H. Spencer. Intensionally complemented distributive lattices. *Portugaliae Mathematica*, 25:99–104, 1966.
- [Belnap Jr. and Wallace, 1961] N. D. Belnap Jr. and J. R. Wallace. A decision procedure for the system  $E_{\neg}$  of entailment with negation. Technical Report 11, Contract No. SAR/609 (16), Office of Naval Research, New Haven, 1961. Also published as [Belnap and Wallace, 1965].
- [Belnap Jr. *et al.*, 1980] N. D. Belnap Jr., A. Gupta, and J. Michael Dunn. A consecution calculus for positive relevant implication with necessity. *Journal of Philosophical Logic*, 9:343–362, 1980.
- [Belnap, 1960a] Nuel D. Belnap. EQ and the first-order functional calculus. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 6:217–218, 1960.
- [Belnap, 1960b] Nuel D. Belnap. A formal analysis of entailment. Technical Report Contract No. SAR/Nonr. 609(16), Office of Naval Research, New Haven, 1960.
- [Belnap, 1962] Nuel D. Belnap. Tonk, plonk and plink. *Analysis*, 22:130–134, 1962.
- [Belnap, 1967a] Nuel D. Belnap. Intensional models for first degree formulas. *Journal of Symbolic Logic*, 32:1–22, 1967.
- [Belnap, 1967b] Nuel D. Belnap. Special cases of the decision problem of relevant implication. *Journal of Symbolic Logic*, 32:431–432, 1967. (Abstract.).
- [Belnap, 1977a] Nuel D. Belnap. How a computer should think. In G. Ryle, editor, *Contemporary Aspects of Philosophy*. Oriol Press, 1977.
- [Belnap, 1977b] Nuel D. Belnap. A useful four-valued logic. In J. Michael Dunn and George Epstein, editors, *Modern Uses of Multiple-Valued Logics*, pages 8–37. Reidel, Dordrecht, 1977.
- [Białynicki-Birula and Rasiowa, 1957] A. Białynicki-Birula and H. Rasiowa. On the representation of quasi-boolean algebras. *Bulletin de L'académie Polonaise des Sciences*, 5:259–261, 1957.
- [Brady, 1989] Ross T. Brady. The non-triviality of dialectical set theory. In Graham Priest, Richard Routley, and Jean Norman, editors, *Paraconsistent Logic: Essays on the Inconsistent*, pages 437–470. Philosophia Verlag, 1989.

- [Brady, 1991] Ross T. Brady. Gentzenization and decidability of some contraction-less relevant logics. *Journal of Philosophical Logic*, 20:97–117, 1991.
- [Brady, 1994] Ross T. Brady. Rules in relevant logic — I: Semantic classification. *Journal of Philosophical Logic*, 23:111–137, 1994.
- [Brady, 1996] Ross T. Brady. Relevant implication and the case for a weaker logic. *Journal of Philosophical Logic*, 25:151–183, 1996.
- [Burgess, 1981] J. P. Burgess. Relevance: A fallacy? *Notre Dame Journal of Formal Logic*, 22:97–104, 1981.
- [Castañeda, 1975] H. N. Castañeda. *Thinking and Doing: The Philosophical Foundations of Institutions*. Reidel, Dordrecht, 1975.
- [Charlewood, 1978] G. W. Charlewood. *Representations of Semilattice Relevance Logic*. PhD thesis, University of Toronto, 1978.
- [Charlewood, 1981] G. W. Charlewood. An axiomatic version of positive semi-lattice relevance logic. *Journal of Symbolic Logic*, 46:233–239, 1981.
- [Church, 1951a] Alonzo Church. The weak positive implication calculus. *Journal of Symbolic Logic*, 16:238, 1951. Abstract of “The Weak Theory of Implication” [Church, 1951b].
- [Church, 1951b] Alonzo Church. The weak theory of implication. In A. Menne, A. Wilhelmly, and H. Angsil, editors, *Kontrolliertes Denken: Untersuchungen zum Logikkalkül und zur Logik der Einzelwissenschaften*, pages 22–37. Kommissions-Verlag Karl Alber, Munich, 1951. Abstracted in “The Weak Positive Implication Calculus” [Church, 1951a].
- [Clark, 1980] M. Clark. The equivalence of tautological and ‘strict’ entailment: proof of an amended conjecture of lewy’s. *Journal of Symbolic Logic*, 9:9–15, 1980.
- [Copeland, 1979] B. J. Copeland. On when a semantics is not a semantics: some reasons for disliking the Routley-Meyer semantics for relevance logic. *Journal of Philosophical Logic*, 8:399–413, 1979.
- [Curry, 1950] Haskell B. Curry. *A Theory of Formal Deducibility*, volume 6 of *Notre Dame Mathematical Lectures*. Notre Dame University Press, 1950.
- [D’Agostino and Gabbay, 1994] Marcello D’Agostino and Dov M. Gabbay. A generalization of analytic deduction via labelled deductive systems. part I: Basic substructural logics. *Journal of Automated Reasoning*, 13:243–281, 1994.
- [Dipert, 1978] R. R. Dipert. *Development and Crisis in Late Boolean Logic; The Deductive Logics of Peirce, Jevons and Schröder*. PhD thesis, Indiana University, 1978.
- [Došen, 1988] Kosta Došen. Sequent systems and groupoid models, part 1. *Studia Logica*, 47:353–386, 1988.
- [Došen, 1989] Kosta Došen. Sequent systems and groupoid models, part 2. *Studia Logica*, 48:41–65, 1989.
- [Došen, 1992] Kosta Došen. The first axiomatisation of relevant logic. *Journal of Philosophical Logic*, 21:339–356, 1992.
- [Dunn and Meyer, 1971] J. Michael Dunn and Robert K. Meyer. Algebraic completeness results for Dummett’s LC and its extensions. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 17:225–230, 1971.
- [Dunn and Meyer, 1989] J. Michael Dunn and Robert K. Meyer. Gentzen’s cut and Ackermann’s  $\gamma$ . In Richard Sylvan and Jean Norman, editors, *Directions in Relevant Logic*. Kluwer, Dordrecht, 1989.
- [Dunn and Meyer, 1997] J. Michael Dunn and Robert K. Meyer. Combinators and structurally free logic. *Logic Journal of the IGPL*, 5:505–357, 1997.
- [Dunn, 1966] J. Michael Dunn. *The Algebra of Intensional Logics*. PhD thesis, University of Pittsburgh, 1966.
- [Dunn, 1967] J. Michael Dunn. The effective equivalence of certain propositions about De Morgan lattices. *Journal of Symbolic Logic*, 32:433–434, 1967. (Abstract).
- [Dunn, 1969] J. Michael Dunn. Natural language versus formal language. In *Joint APA-APL Symposium, New York, December 17, 1969*.
- [Dunn, 1970] J. Michael Dunn. Algebraic completeness for  $R$ -mingle and its extensions. *Journal of Symbolic Logic*, 35:1–13, 1970.
- [Dunn, 1971] J. Michael Dunn. An intuitive semantics for first degree relevant implications (abstract). *Journal of Symbolic Logic*, 36:363–363, 1971.

- [Dunn, 1974] J. Michael Dunn. A ‘Gentzen’ system for positive relevant implication. *Journal of Symbolic Logic*, 38:356–357, 1974. (Abstract).
- [Dunn, 1976a] J. Michael Dunn. Intuitive semantics for first-degree entailments and “coupled trees”. *Philosophical Studies*, 29:149–168, 1976.
- [Dunn, 1976b] J. Michael Dunn. A Kripke-style semantics for  $R$ -mingle using a binary accessibility relation. *Studia Logica*, 35:163–172, 1976.
- [Dunn, 1976c] J. Michael Dunn. Quantification and RM. *Studia Logica*, 35:315–322, 1976.
- [Dunn, 1976d] J. Michael Dunn. A variation on the binary semantics for RM. *Relevance Logic Newsletter*, 1:56–67, 1976.
- [Dunn, 1979a] J. Michael Dunn.  $R$ -mingle and beneath: Extensions of the Routley-Meyer semantics for  $R$ . *Notre Dame Journal of Formal Logic*, 20:369–376, 1979.
- [Dunn, 1979b] J. Michael Dunn. A theorem in 3-valued model theory with connections to number theory, type theory, and relevant logic. *Studia Logica*, 38:149–169, 1979.
- [Dunn, 1980a] J. Michael Dunn. Relevant Robinson’s arithmetic. *Studia Logica*, 38:407–418, 1980.
- [Dunn, 1980b] J. Michael Dunn. A sieve for entailments. *Journal of Philosophical Logic*, 9:41–57, 1980.
- [Dunn, 1987] J. Michael Dunn. Relevant predication 1: The formal theory. *Journal of Philosophical Logic*, 16:347–381, 1987.
- [Dunn, 1990a] J. Michael Dunn. Relevant predication 2: Intrinsic properties and internal relations. *Philosophical Studies*, 60:177–206, 1990.
- [Dunn, 1990b] J. Michael Dunn. Relevant predication 3: Essential properties. In J. Michael Dunn and Anil Gupta, editors, *Truth or Consequences*, pages 77–95. Kluwer, 1990.
- [Dunn, 1991] J. Michael Dunn. Gaggle theory: An abstraction of Galois connections and residuation with applications to negation and various logical operations. In *Logics in AI, Proceedings European Workshop JELIA 1990*, volume 478 of *Lecture Notes in Computer Science*. Springer-Verlag, 1991.
- [Dunn, 1993a] J. Michael Dunn. Partial-gaggles applied to logics with restricted structural rules. In Peter Schroeder-Heister and Kosta Došen, editors, *Substructural Logics*. Oxford University Press, 1993.
- [Dunn, 1993b] J. Michael Dunn. A representation of relation algebras using Routley-Meyer frames. Technical report, Indiana University Logic Group Preprint Series, IULG-93-28, 1993.
- [Dunn, 1994] J. Michael Dunn. Star and perp: Two treatments of negation. In James E. Tomberlin, editor, *Philosophical Perspectives*, volume 7, pages 331–357. Ridgeview Publishing Company, Atascadero, California, 1994.
- [Dunn, 1996] J. Michael Dunn. Generalised ortho negation. In Heinrich Wansing, editor, *Negation: A Notion in Focus*, pages 3–26. Walter de Gruyter, Berlin, 1996.
- [Feys and Ladrière, 1955] R. Feys and J. Ladrière. Notes to *recherches sur la déduction logique*, 1955. Translation of “Untersuchungen über das logische Schliessen” [Gentzen, 1934], Paris.
- [Fine, 1974] Kit Fine. Models for entailment. *Journal of Philosophical Logic*, 3:347–372, 1974.
- [Fine, 1988] Kit Fine. Semantics for quantified relevance logic. *Journal of Philosophical Logic*, 17:27–59, 1988.
- [Fine, 1989] Kit Fine. Incompleteness for quantified relevance logics. In Jean Norman and Richard Sylvan, editors, *Directions in Relevant Logic*, pages 205–225. Kluwer Academic Publishers, Dordrecht, 1989.
- [Fitting, 1989] Melvin C. Fitting. Bilattices and the semantics of logic programming. *Journal of Logic Programming*, 11(2):91–116, 1989.
- [Fogelin, 1978] R. J. Fogelin. *Understanding Arguments*. Harcourt, Brace, Jovanovich, New York, 1978.
- [Friedman and Meyer, 1992] Harvey Friedman and Robert K. Meyer. Whither relevant arithmetic? *Journal of Symbolic Logic*, 57:824–831, 1992.
- [Gabbay, 1976] D. M. Gabbay. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. Reidel, Dordrecht, 1976.

- [Gabbay, 1997] Dov M. Gabbay. *Labelled Deductive Systems*. Number 33 in Oxford Logic Guides. Oxford University Press, 1997.
- [Gentzen, 1934] Gerhard Gentzen. Untersuchungen über das logische schliessen. *Math. Zeitschrift*, 39:176–210 and 405–431, 1934. Translated in *The Collected Papers of Gerhard Gentzen* [Gentzen, 1969].
- [Gentzen, 1969] Gerhard Gentzen. *The Collected Papers of Gerhard Gentzen*. North Holland, 1969. Edited by M. E. Szabo.
- [Giambrone, 1985] Steve Giambrone.  $TW_+$  and  $RW_+$  are decidable. *Journal of Philosophical Logic*, 14:235–254, 1985.
- [Girard, 1987] Jean-Yves Girard. Linear logic. *Theoretical Computer Science*, 50:1–101, 1987.
- [Grice, 1975] H. P. Grice. Logic and conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics: Speech Acts*, volume 3, pages 41–58. Academic Press, New York, 1975. Reprinted in Jackson [Jackson, 1991].
- [Haack, 1974] S. Haack. *Deviant Logic: Some Philosophical Issues*. Cambridge University Press, 1974. Reissued as *Deviant Logic, Fuzzy Logic: Beyond the Formalism* [Haack, 1996].
- [Haack, 1996] Susan Haack. *Deviant Logic, Fuzzy Logic: Beyond the Formalism*. Cambridge University Press, 1996.
- [Harrop, 1965] R. Harrop. Some structure results for propositional calculi. *Journal of Symbolic Logic*, 30:271–292, 1965.
- [Hutchinson, 1973] G. Hutchinson. Recursively unsolvable word problems of modular lattices and diagram chasing. *Journal of Algebra*, pages 385–399, 1973.
- [Jackson, 1991] Frank Jackson, editor. *Conditionals*. Oxford Readings in Philosophy. Oxford, 1991.
- [Jónsson and Tarski, 1951] Bjarni Jónsson and Alfred Tarski. Boolean algebras with operators: Part I. *American Journal of Mathematics*, 73:891–939, 1951.
- [Kalman, 1958] J. A. Kalman. Lattices with involution. *Transactions of the American Mathematical Society*, 87:485–491, 1958.
- [Kripke, 1959a] Saul A. Kripke. A completeness theorem in modal logic. *Journal of Symbolic Logic*, 24:1–15, 1959.
- [Kripke, 1959b] Saul A. Kripke. The problem of entailment. *Journal of Symbolic Logic*, 24:324, 1959. Abstract.
- [Kripke, 1963] Saul A. Kripke. Semantical considerations on modal logic. *Acta Philosophica Fennica*, 16:83–94, 1963.
- [Kripke, 1965] Saul A. Kripke. Semantical analysis of modal logic II: Non-normal modal propositional calculi. In Addison et al., editor, *The Theory of Models*, pages 206–220. North Holland Publishing Co, 1965.
- [Kron, 1973] A. Kron. Deduction theorems for relevant logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 19:85–92, 1973.
- [Kron, 1976] A. Kron. Deduction theorems for  $T$ ,  $E$  and  $R$  reconsidered. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 22:261–264, 1976.
- [Kurtonina, 1995] Natasha Kurtonina. *Frames and Labels: A Modal Analysis of Categorical Inference*. PhD thesis, Institute for Logic, Language and Computation, University of Utrecht, 1995.
- [Lambek, 1958] Joachim Lambek. The mathematics of sentence structure. *American Mathematical Monthly*, 65:154–170, 1958.
- [Lambek, 1961] Joachim Lambek. On the calculus of syntactic types. In R. Jacobsen, editor, *Structure of Language and its Mathematical Aspects*, Proceedings of Symposia in Applied Mathematics, XII. American Mathematical Society, 1961.
- [Lemmon, 1965] E. J. Lemmon. *Beginning Logic*. Nelson, 1965.
- [Lemmon, 1966] E. J. Lemmon. Algebraic semantics for modal logics, I and II. *Journal of Symbolic Logic*, 31:46–55 and 191–218, 1966.
- [Lewis and Langford, 1932] C. I. Lewis and C. H. Langford. *Symbolic Logic*. The Century Co, New York and London, 1932.
- [Lincoln et al., 1992] P. Lincoln, J. Mitchell, A. Scenrov, and N. Shankar. Decision problems for propositional linear logic. *Annals of Pure and Applied Logic*, 56:239–311, 1992.

- [Lipshitz, 1974] L. Lipshitz. The undecidability of the word problems for projective geometries and modular lattices. *Transactions of the American Mathematical Society*, pages 171–180, 1974.
- [Maksimova, 1967] L. L. Maksimova. O modeláh isčisléníá E. *Algébra i Logika, Séminar*, 6:5–20, 1967. (English title: On Models of the System E).
- [Maksimova, 1971] L. L. Maksimova. An interpolation and separation theorem for the logical systems  $e$  and  $r$ . *Algebra and Logic*, 10:232–241, 1971.
- [Maksimova, 1973] L. L. Maksimova. A semantics for the calculus  $E$  of entailment. *Bulletin of the section of Logic*, 2:18–21, 1973. Published by Polish Academy of Sciences, Institute of Philosophy and Sociology.
- [Mares and Fuhrmann, 1995] Edwin D. Mares and André Fuhrmann. A relevant theory of conditionals. *Journal of Philosophical Logic*, 24:645–665, 1995.
- [Mares, 1992] Edwin Mares. Semantics for relevance logic with identity. *Studia Logica*, 51:1–20, 1992.
- [Mares, 1997] Edwin Mares. Relevant logic and the theory of information. *Synthese*, 109:345–360, 1997.
- [Martin and Meyer, 1982] E. P. Martin and R. K. Meyer. Solution to the P-W problem. *Journal of Symbolic Logic*, 47:869–886, 1982.
- [McKinsey and Tarski, 1948] J. C. C. McKinsey and A. Tarski. Some theorems about the sentential calculi of Lewis and Heyting. *Journal of Symbolic Logic*, 13:1–15, 1948.
- [McRobbie and Belnap Jr., 1979] M. A. McRobbie and N. D. Belnap Jr. Relevant analytic tableaux. *Studia Logica*, 38:187–200, 1979.
- [Meyer and Dunn, 1969] Robert K. Meyer and J. Michael Dunn.  $E$ ,  $R$  and  $\gamma$ . *Journal of Symbolic Logic*, 34:460–474, 1969.
- [Meyer and Leblanc, 1970] Robert K. Meyer and H. Leblanc. A semantical completeness result for relevant quantification theories. *Journal of Symbolic Logic*, 35:181, 1970. Abstract.
- [Meyer and McRobbie, 1979] Robert K. Meyer and Michael A. McRobbie. Firesets and relevant implication. Technical Report 3, Logic Group, RSSS, Australian National University, 1979.
- [Meyer and Routley, 1972] Robert K. Meyer and Richard Routley. Algebraic analysis of entailment. *Logique et Analyse*, 15:407–428, 1972.
- [Meyer and Routley, 1973a] Robert K. Meyer and Richard Routley. Classical relevant logics I. *Studia Logica*, 32:51–66, 1973.
- [Meyer and Routley, 1973b] Robert K. Meyer and Richard Routley. Classical relevant logics II. *Studia Logica*, 33:183–194, 1973.
- [Meyer and Routley, 1973c] Robert K. Meyer and Richard Routley. An undecidable relevant logic. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 19:289–397, 1973.
- [Meyer and Urbas, 1986] Robert K. Meyer and Igor Urbas. Conservative extension in relevant arithmetics. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 32:45–50, 1986.
- [Meyer et al., 1974] Robert K. Meyer, J. Michael Dunn, and H. Leblanc. Completeness of relevant quantification theories. *Notre Dame Journal of Formal Logic*, 15:97–121, 1974.
- [Meyer et al., 1979] Robert K. Meyer, Richard Routley, and J. Michael Dunn. Curry's paradox. *Analysis*, 39:124–128, 1979.
- [Meyer et al., 1995] Robert K. Meyer, Michael A. McRobbie, and Nuel D. Belnap. Linear analytic tableaux. In *Proceedings of the Fourth Workshop on Theorem Proving with Analytic Tableaux and Related Methods*, volume 918 of *Lecture Notes in Computer Science*, 1995.
- [Meyer, 1966] Robert K. Meyer. *Topics in Modal and Many-valued Logic*. PhD thesis, University of Pittsburgh, 1966.
- [Meyer, 1968] Robert K. Meyer. An undecidability result in the theory of relevant implication. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 14:255–262, 1968.
- [Meyer, 1970a] Robert K. Meyer.  $E$  and  $S4$ . *Notre Dame Journal of Formal Logic*, 11:181–199, 1970.

- [Meyer, 1970b] Robert K. Meyer.  $R_I$ —the bounds of finitude. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, 16:385–387, 1970.
- [Meyer, 1971] Robert K. Meyer. On coherence in modal logics. *Logique et Analyse*, 14:658–668, 1971.
- [Meyer, 1973] Robert K. Meyer. Intuitionism, entailment, negation. In H. Leblanc, editor, *Truth, Syntax and Modality*, pages 168–198. North Holland, 1973.
- [Meyer, 1974] Robert K. Meyer. New axiomatics for relevant logics — I. *Journal of Philosophical Logic*, 3:53–86, 1974.
- [Meyer, 1976a] R. K. Meyer. Ackermann, Takeuti and Schnitt;  $\gamma$  for higher-order relevant logics. *Bulletin of the Section of Logic*, 5:138–144, 1976.
- [Meyer, 1976b] Robert K. Meyer. Metacompleteness. *Notre Dame Journal of Formal Logic*, 17:501–517, 1976.
- [Meyer, 1976c] Robert K. Meyer. Relevant arithmetic. *Bulletin of the Section of Logic*, 5:133–137, 1976.
- [Meyer, 1978] Robert K. Meyer. Why I am not a relevantist. Technical Report 1, Logic Group, RISS, Australian National University, 1978.
- [Meyer, 1979a] R. K. Meyer. A Boolean-valued semantics for  $r$ . Technical Report 4, Logic Group, RISS, Australian National University, 1979.
- [Meyer, 1979b] Robert K. Meyer. Arithmetic formulated relevantly. Unpublished manuscript. Some of the results appear in *Entailment* Volume 2 [Anderson *et al.*, 1992], 1979.
- [Meyer, 1979c] Robert K. Meyer. Career induction stops here (and here = 2). *Journal of Philosophical Logic*, 8:361–371, 1979.
- [Meyer, 1980] Robert K. Meyer. Relevantly interpolating in RM. Technical Report 9, Logic Group, RISS, Australian National University, 1980.
- [Meyer, 1998] Robert K. Meyer.  $\supset E$  is admissible in ‘true’ relevant arithmetic. *Journal of Philosophical Logic*, 27:327–351, 1998.
- [Minc, 1972] G. Minc. Cut-elimination theorem in relevant logics. In J. V. Matijasevic and O. A. Silenko, editors, *Issledovaniá po konstruktivnoj matematiké i matematičeskoj logike V*, pages 90–97. Izdatel’stvo “Nauka”, 1972. (English translation in “Cut-Elimination Theorem in Relevant Logics” [Minc, 1976]).
- [Minc, 1976] G. Minc. Cut-elimination theorem in relevant logics. *The Journal of Soviet Mathematics*, 6:422–428, 1976. (English translation of the original article [Minc, 1972]).
- [Mortensen, 1995] Chris Mortensen. *Inconsistent Mathematics*. Kluwer Academic Publishers, 1995.
- [Parry, 1933] W. T. Parry. Ein axiomensystem für eine neue art von implikation (analytische implication). *Ergebnisse eines mathematischen Kolloquiums*, 4:5–6, 1933.
- [Pottinger, 1979] G. Pottinger. On analysing relevance constructively. *Studia Logica*, 38:171–185, 1979.
- [Prawitz, 1965] Dag Prawitz. *Natural Deduction: A Proof Theoretical Study*. Almqvist and Wiksell, Stockholm, 1965.
- [Priest and Sylvan, 1992] Graham Priest and Richard Sylvan. Simplified semantics for basic relevant logics. *Journal of Philosophical Logic*, 21:217–232, 1992.
- [Priest, 1987] Graham Priest. *In Contradiction: A Study of the Transconsistent*. Martinus Nijhoff, The Hague, 1987.
- [Priest, 1995] Graham Priest. *Beyond the Limits of Thought*. Cambridge University Press, Cambridge, 1995.
- [Quine, 1953] W. V. Quine. Three grades of modal involvement. In *Proceedings of the XIth International Congress of Philosophy*. North Holland, Amsterdam, 1953. Reprinted in *The Ways of Paradox*, Random House, New York, 1966.
- [Rasiowa, 1974] H. Rasiowa. *An Algebraic Approach to Non-classical Logics*. North Holland, 1974.
- [Read, 1988] Stephen Read. *Relevant Logic*. Basil Blackwell, Oxford, 1988.
- [Restall, 1993] Greg Restall. Simplified semantics for relevant logics (and some of their rivals). *Journal of Philosophical Logic*, 22:481–511, 1993.
- [Restall, 1994] Greg Restall. A useful substructural logic. *Bulletin of the Interest Group in Pure and Applied Logic*, 2(2):135–146, 1994.

- [Restall, 1995a] Greg Restall. Display logic and gaggle theory. *Reports in Mathematical Logic*, 29:133–146, 1995.
- [Restall, 1995b] Greg Restall. Information flow and relevant logics. In Jerry Seligman and Dag Westerstahl, editors, *Logic, Language and Computation: The 1994 Moraga Proceedings*, pages 463–477. CSLI Publications, 1995.
- [Restall, 1998] Greg Restall. Displaying and deciding substructural logics 1: Logics with contraposition. *Journal of Philosophical Logic*, 27:179–216, 1998.
- [Restall, 1999] Greg Restall. Negation in Relevant Logics: How I Stopped Worrying and Learned to Love the Routley Star. In Dov Gabbay and Heinrich Wansing, editors, *What is Negation?*, volume 13 of *Applied Logic Series*, pages 53–76. Kluwer Academic Publishers, 1999.
- [Restall, 2000] Greg Restall. *An Introduction to Substructural Logics*. Routledge, 2000.
- [Routley and Meyer, 1973] Richard Routley and Robert K. Meyer. Semantics of entailment. In Hugues Leblanc, editor, *Truth Syntax and Modality*, pages 194–243. North Holland, 1973. Proceedings of the Temple University Conference on Alternative Semantics.
- [Routley and Meyer, 1976] Richard Routley and Robert K. Meyer. Dialectal logic, classical logic and the consistency of the world. *Studies in Soviet Thought*, 16:1–25, 1976.
- [Routley and Routley, 1972] Richard Routley and Valerie Routley. Semantics of first-degree entailment. *Noûs*, 3:335–359, 1972.
- [Routley et al., 1982] Richard Routley, Val Plumwood, Robert K. Meyer, and Ross T. Brady. *Relevant Logics and their Rivals*. Ridgeview, 1982.
- [Routley, 1977] Richard Routley. Ultralogic as universal. *Relevance Logic Newsletter*, 2:51–89, 1977. Reprinted in *Exploring Meinong's Jungle* [Routley, 1980a].
- [Routley, 1980a] Richard Routley. *Exploring Meinong's Jungle and Beyond*. Philosophy Department, RSSS, Australian National University, 1980. Interim Edition, Departmental Monograph number 3.
- [Routley, 1980b] Richard Routley. Problems and solutions in the semantics of quantified relevant logics — I. In A. I. Arrduda, R. Chuaqui, and N. C. A. da Costa, editors, *Proceedings of the Fourth Latin American Symposium on Mathematical Logic*, pages 305–340. North Holland, 1980.
- [Schroeder-Heister and Došen, 1993] Peter Schroeder-Heister and Kosta Došen, editors. *Substructural Logics*. Oxford University Press, 1993.
- [Schütte, 1956] K. Schütte. Ein system des verknüpfenden schliessens. *Archiv für Mathematische Logik*, 2:55–67, 1956.
- [Scott, 1971] Dana Scott. On engendering an illusion of understanding. *Journal of Philosophy*, 68:787–807, 1971.
- [Shaw-Kwei, 1950] Moh Shaw-Kwei. The deduction theorems and two new logical systems. *Methodos*, 2:56–75, 1950.
- [Slaney, 1985] John K. Slaney. 3088 varieties: A solution to the Ackermann constant problem. *Journal of Symbolic Logic*, 50:487–501, 1985.
- [Slaney, 1990] John K. Slaney. A general logic. *Australasian Journal of Philosophy*, 68:74–88, 1990.
- [Smullyan, 1968] R. M. Smullyan. *First-Order Logic*. Springer-Verlag, Berlin, 1968. Reprinted by Dover Press, 1995.
- [Stone, 1936] M. H. Stone. The theory of representation for Boolean algebras. *Transactions of the American Mathematical Society*, 40:37–111, 1936.
- [Suppes, 1957] P. Suppes. *Introduction to Logic*. D. Van Nostrand Co, Princeton, 1957.
- [Tennant, 1994] Neil Tennant. The transmission of truth and the transitivity of deduction. In Dov Gabbay, editor, *What is a Logical System?*, volume 4 of *Studies in Logic and Computation*, pages 161–177. Oxford University Press, Oxford, 1994.
- [Thistlewaite et al., 1988] Paul Thistlewaite, Michael McRobbie, and Robert K. Meyer. *Automated Theorem Proving in Non-Classical Logics*. Wiley, New York, 1988.
- [Tokarz, 1980] M. Tokarz. *Essays in Matrix Semantics of Relevant Logics*. The Polish Academy of Science, Warsaw, 1980.
- [Troelstra, 1992] A. S. Troelstra. *Lectures on Linear Logic*. CSLI Publications, 1992.
- [Urquhart, 1972a] Alasdair Urquhart. The completeness of weak implication. *Theoria*, 37:274–282, 1972.

- [Urquhart, 1972b] Alasdair Urquhart. A general theory of implication. *Journal of Symbolic Logic*, 37:443, 1972.
- [Urquhart, 1972c] Alasdair Urquhart. Semantics for relevant logics. *Journal of Symbolic Logic*, 37:159–169, 1972.
- [Urquhart, 1972d] Alasdair Urquhart. *The Semantics of Entailment*. PhD thesis, University of Pittsburgh, 1972.
- [Urquhart, 1983] Alasdair Urquhart. Relevant implication and projective geometry. *Logique et Analyse*, 26:345–357, 1983.
- [Urquhart, 1984] Alasdair Urquhart. The undecidability of entailment and relevant implication. *Journal of Symbolic Logic*, 49:1059–1073, 1984.
- [Urquhart, 1990] Alasdair Urquhart. The complexity of decision procedures in relevance logic. In J. Michael Dunn and Anil Gupta, editors, *Truth or Consequences*, pages 77–95. Kluwer, 1990.
- [Urquhart, 1993] Alasdair Urquhart. Failure of interpolation in relevant logics. *Journal of Philosophical Logic*, 22:449–479, 1993.
- [Urquhart, 1997] Alasdair Urquhart. The complexity of decision procedures in relevance logic ii. Available from the author, University of Toronto, 1997.
- [van Fraassen, 1969] Bas van Fraassen. Facts and tautological entailments. *Journal of Philosophy*, 66:477–487, 1969. Reprinted in *Entailment* Volume 1 [Anderson and Belnap, 1975].
- [van Fraassen, 1973] Bas van Fraassen. Extension, intension and comprehension. In M. Munitz, editor, *Logic and Ontology*, pages 101–103. New York University Press, New York, 1973.
- [Wansing, 1993] Heinrich Wansing. *The Logic of Information Structures*. Number 681 in Lecture Notes in Artificial Intelligence. Springer-Verlag, 1993.
- [Wolf, 1978] R. G. Wolf. Are relevant logics deviant. *Philosophia*, 7:327–340, 1978.



MARIA LUISA DALLA CHIARA AND ROBERTO  
GIUNTINI

## QUANTUM LOGICS

### 1 INTRODUCTION

The official birth of quantum logic is represented by a famous article of Birkhoff and von Neumann “The logic of quantum mechanics” [Birkhoff and von Neumann, 1936]. At the very beginning of their paper, Birkhoff and von Neumann observe:

One of the aspects of quantum theory which has attracted the most general attention, is the novelty of the logical notions which it presupposes .... The object of the present paper is to discover what logical structures one may hope to find in physical theories which, like quantum mechanics, do not conform to classical logic.

In order to understand the basic reason why a non classical logic arises from the mathematical formalism of quantum theory (QT), a comparison with classical physics will be useful.

There is one concept which quantum theory shares alike with classical mechanics and classical electrodynamics. This is the concept of a mathematical “phase-space”. According to this concept, any physical system  $\mathcal{S}$  is at each instant hypothetically associated with a “point” in a fixed phase-space  $\Sigma$ ; this point is supposed to represent mathematically, the “state” of  $\mathcal{S}$ , and the “state” of  $\mathcal{S}$  is supposed to be ascertainable by “maximal” observations.

Maximal pieces of information about physical systems are called also *pure states*. For instance, in classical particle mechanics, a pure state of a single particle can be represented by a sequence of six real numbers  $\langle r_1, \dots, r_6 \rangle$  where the first three numbers correspond to the *position*-coordinates, whereas the last ones are the *momentum*-components.

As a consequence, the phase-space of a single particle system can be identified with the set  $\mathbb{R}^6$ , consisting of all sextuples of real numbers. Similarly for the case of compound systems, consisting of a finite number  $n$  of particles.

Let us now consider an *experimental proposition*  $\mathbf{P}$  about our system, asserting that a given physical quantity has a certain value (for instance: “the value of position in the  $x$ -direction lies in a certain interval”). Such

a proposition  $\mathbf{P}$  will be naturally associated with a subset  $X$  of our phase-space, consisting of all the pure states for which  $\mathbf{P}$  holds. In other words, the subsets of  $\Sigma$  seem to represent good mathematical representatives of experimental propositions. These subsets are called by Birkhoff and von Neumann *physical qualities* (we will say simply *events*). Needless to say, the correspondence between the set of all experimental propositions and the set of all events will be many-to-one. When a pure state  $p$  belongs to an event  $X$ , we will say that our system in state  $p$  *verifies* both  $X$  and the corresponding experimental proposition.

What about the structure of all events? As is well known, the power-set of any set is a *Boolean algebra*. And also the set  $\mathcal{F}(\Sigma)$  of all measurable subsets of  $\Sigma$  (which is more tractable than the full power-set of  $\Sigma$ ) turns out to have a Boolean structure. Hence, we may refer to the following Boolean algebra:

$$\mathcal{B} = \langle \mathcal{F}(\Sigma), \subseteq, \cap, \cup, -, \mathbf{1}, \mathbf{0} \rangle,$$

where:

- 1)  $\subseteq, \cap, \cup, -$  are, respectively, the set-theoretic inclusion relation and the operations intersection, union, relative complement;
- 2)  $\mathbf{1}$  is the total space  $\Sigma$ , while  $\mathbf{0}$  is the empty set.

According to a standard interpretation,  $\cap, \cup, -$  can be naturally regarded as a set-theoretic realization of the classical logical connectives *and*, *or*, *not*. As a consequence, we will obtain a classical semantic behaviour:

- a state  $p$  verifies a conjunction  $X \cap Y$  iff  $p \in X \cap Y$  iff  $p$  verifies both members;
- $p$  verifies a disjunction  $X \cup Y$  iff  $p \in X \cup Y$  iff  $p$  verifies at least one member;
- $p$  verifies a negation  $-X$  iff  $p \notin X$  iff  $p$  does not verify  $X$ .

To what extent can such a picture be adequately extended to QT? Birkhoff and von Neumann observe:

In quantum theory the points of  $\Sigma$  correspond to the so called “wave-functions” and hence  $\Sigma$  is ... a function-space, usually assumed to be Hilbert space.

As a consequence, we immediately obtain a basic difference between the quantum and the classical case. The *excluded middle principle* holds in

classical mechanics. In other words, pure states semantically decide any event: for any  $p$  and  $X$ ,

$$p \in X \text{ or } p \in -X.$$

QT is, instead, essentially probabilistic. Generally, pure states assign only probability-values to quantum events. Let  $\psi$  represent a pure state (a wave function) of a quantum system and let  $\mathbf{P}$  be an experimental proposition (for instance “the spin value in the  $x$ -direction is up”). The following cases are possible:

- (i)  $\psi$  assigns to  $\mathbf{P}$  probability-value 1 ( $\psi(\mathbf{P}) = 1$ );
- (ii)  $\psi$  assigns to  $\mathbf{P}$  probability-value 0 ( $\psi(\mathbf{P}) = 0$ );
- (iii)  $\psi$  assigns to  $\mathbf{P}$  a probability-value different from 1 and from 0 ( $\psi(\mathbf{P}) \neq 0, 1$ ).

In the first two cases, we will say that  $\mathbf{P}$  is *true (false)* for our system in state  $\psi$ . In the third case,  $\mathbf{P}$  will be *semantically indeterminate*.

Now the question arises: what will be an adequate mathematical representative for the notion of quantum experimental proposition? The most important novelty of Birkhoff and von Neumann’s proposal is based on the following answer: “The mathematical representative of any experimental proposition is a closed linear subspace of Hilbert space” (we will say simply a *closed subspace*).<sup>1</sup> Let  $\mathcal{H}$  be a (separable) Hilbert space, whose *unitary vectors* correspond to possible wave functions of a quantum system. The closed subspaces of  $\mathcal{H}$  are particular instances of subsets of  $\mathcal{H}$  that are closed under linear combinations and Cauchy sequences. Why are mere subsets of the phase-space not interesting in QT? The reason depends on the *superposition principle*, which represents one of the basic dividing line between the quantum and the classical case. Differently from classical mechanics, in quantum mechanics, finite and even infinite linear combinations of pure states give rise to new pure states (provided only some formal conditions are satisfied). Suppose three pure states  $\psi, \psi_1, \psi_2$  and let  $\psi$  be a linear combination of  $\psi_1, \psi_2$ :

$$\psi = c_1\psi_1 + c_2\psi_2.$$

---

<sup>1</sup>A *Hilbert space* is a vector space over a *division ring* whose elements are the real or the complex or the quaternionic numbers such that

- (i) An *inner product*  $(\cdot, \cdot)$  that transforms any pair of vectors into an element of the division ring is defined;
- (ii) the space is *metrically complete* with respect to the metrics induced by the inner product  $(\cdot, \cdot)$ .

A Hilbert space  $\mathcal{H}$  is called *separable* iff  $\mathcal{H}$  admits a countable basis.

According to the standard interpretation of the formalism, roughly this means that a quantum system in state  $\psi$  might verify with probability  $|c_1|^2$  those propositions that are certain for state  $\psi_1$  (and are not certain for  $\psi$ ) and might verify with probability  $|c_2|^2$  those propositions that are certain for state  $\psi_2$  (and are not certain for  $\psi$ ). Suppose now some pure states  $\psi_1, \psi_2, \dots$  each assigning probability 1 to a given experimental proposition  $\mathbf{P}$ , and suppose that the linear combination

$$\psi = \sum_i c_i \psi_i \quad (c_i \neq 0)$$

is a pure state. Then also  $\psi$  will assign probability 1 to our proposition  $\mathbf{P}$ . As a consequence, the mathematical representatives of experimental propositions should be closed under finite and infinite linear combinations. The closed subspaces of  $\mathcal{H}$  are just the mathematical objects that can realize such a role.

What about the algebraic structure that can be defined on the set  $C(\mathcal{H})$  of all mathematical representatives of experimental propositions (let us call them *quantum events*)? For instance, what does it mean *negation*, *conjunction* and *disjunction* in the realm of quantum events? As to negation, Birkhoff and von Neumann's answer is the following:

The mathematical representative of the *negative* of any experimental proposition is the *orthogonal complement* of the mathematical representative of the proposition itself.

The orthogonal complement  $X'$  of a subspace  $X$  is defined as the set of all vectors that are orthogonal to all elements of  $X$ . In other words,  $\psi \in X'$  iff  $\psi \perp X$  iff for any  $\phi \in X$ :  $(\psi, \phi) = 0$  (where  $(\psi, \phi)$  is the inner product of  $\psi$  and  $\phi$ ). From the point of view of the physical interpretation, the orthogonal complement (called also *orthocomplement*) is particularly interesting, since it satisfies the following property: for any event  $X$  and any pure state  $\psi$ ,

$$\psi(X) = 1 \text{ iff } \psi(X') = 0;$$

$$\psi(X) = 0 \text{ iff } \psi(X') = 1;$$

In other words,  $\psi$  assigns to an event  $X$  probability 1 (0, respectively) iff  $\psi$  assigns to the orthocomplement of  $X$  probability 0 (1, respectively). As a consequence, one is dealing with an operation that *inverts* the two extreme probability-values, which naturally correspond to the truth-values *truth* and *falsity* (similarly to the classical truth-table of negation).

As to conjunction, Birkhoff and von Neumann notice that this can be still represented by the set-theoretic intersection (like in the classical case). For,

the intersection  $X \cap Y$  of two closed subspaces is again a closed subspace. Hence, we will obtain the usual truth-table for the connective *and*:

$\psi$  verifies  $X \cap Y$  iff  $\psi$  verifies both members.

Disjunction, however, cannot be represented here as a set-theoretic union. For, generally, the union  $X \cup Y$  of two closed subspaces is not a closed subspace. In spite of this, we have at our disposal another good representative for the connective *or*: the *supremum*  $X \sqcup Y$  of two closed subspaces, that is the smallest closed subspace including both  $X$  and  $Y$ . Of course,  $X \sqcup Y$  will include  $X \cup Y$ .

As a consequence, we obtain the following structure

$$\mathcal{C}(\mathcal{H}) = \langle C(\mathcal{H}), \sqsubseteq, \sqcap, \sqcup, ', \mathbf{1}, \mathbf{0} \rangle,$$

where  $\sqsubseteq, \sqcap$  are the set-theoretic inclusion and intersection;  $\sqcup, '$  are defined as above; while  $\mathbf{1}$  and  $\mathbf{0}$  represent, respectively, the total space  $\mathcal{H}$  and the null subspace (the singleton of the null vector, representing the smallest possible subspace). An isomorphic structure can be obtained by using as a support, instead of  $C(\mathcal{H})$ , the set  $P(\mathcal{H})$  of all *projections*  $P$  of  $\mathcal{H}$ . As is well known projections (i.e. *idempotent* and *self-adjoint linear operators*) and closed subspaces are in one-to-one correspondence, by the projection theorem. Our structure  $\mathcal{C}(\mathcal{H})$  turns out to simulate a “quasi-Boolean behaviour”; however, it is not a Boolean algebra. Something very essential is missing. For instance, conjunction and disjunction are no more distributive. Generally,

$$X \sqcap (Y \sqcup Z) \neq (X \sqcap Y) \sqcup (X \sqcap Z).$$

It turns out that  $\mathcal{C}(\mathcal{H})$  belongs to the variety of all *orthocomplemented orthomodular lattices*, that are not necessarily distributive.

The failure of distributivity is connected with a characteristic property of disjunction in QT. Differently from classical (bivalent) semantics, a quantum disjunction  $X \sqcup Y$  may be true even if neither member is true. In fact, it may happen that a pure state  $\psi$  belongs to a subspace  $X \sqcup Y$ , even if  $\psi$  belongs neither to  $X$  nor to  $Y$  (see Figure 1).

Such a semantic behaviour, which may appear *prima facie* somewhat strange, seems to reflect pretty well a number of concrete quantum situations. In QT one is often dealing with alternatives that are semantically determined and true, while both members are, in principle, strongly indeterminate. For instance, suppose we are referring to some one-half spin particle (say an electron) whose spin may assume only two possible values: either *up* or *down*. Now, according to one of the *uncertainty principles*, the spin in the  $x$  direction ( $spin_x$ ) and the spin in the  $y$  direction ( $spin_y$ ) represent two strongly *incompatible* quantities that cannot be simultaneously

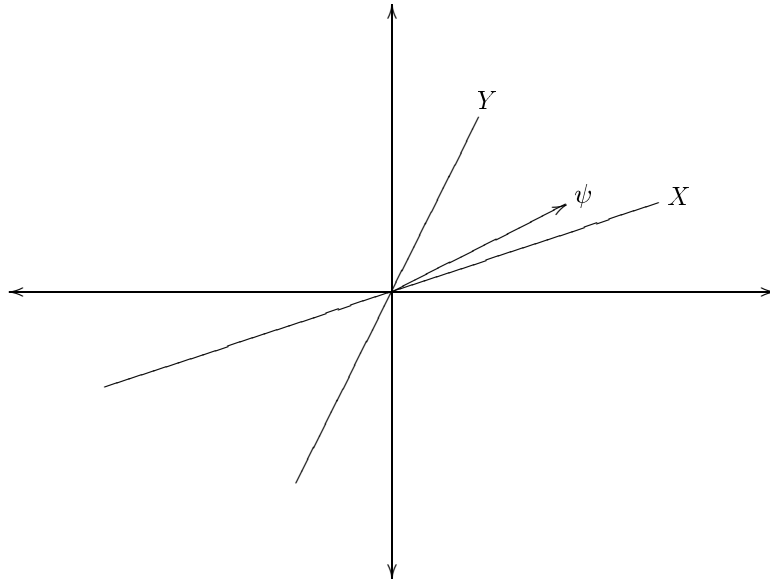


Figure 1. Failure of bivalence in QT

measured. Suppose an electron in state  $\psi$  verifies the proposition “ $spin_x$  is up”. As a consequence of the uncertainty principle both propositions “ $spin_y$  is up” and “ $spin_y$  is down” shall be strongly indeterminate. However the disjunction “either  $spin_y$  is up or  $spin_y$  is down” must be true.

Birkhoff and von Neumann’s proposal did not arouse any immediate interest, either in the logical or in the physical community. Probably, the quantum logical approach appeared too abstract for the foundational debate about QT, which in the Thirties was generally formulated in a more traditional philosophical language. As an example, let us only think of the famous discussion between Einstein and Bohr. At the same time, the work of logicians was still mainly devoted to classical logic.

Only twenty years later, after the appearance of George Mackey’s book *Mathematical Foundations of Quantum Theory* [Mackey, 1957], one has witnessed a “renaissance period” for the logico-algebraic approach to QT. This has been mainly stimulated by the contributions of Jauch, Piron, Varadarajan, Suppes, Finkelstein, Foulis, Randall, Greechie, Gudder, Beltrametti, Cassinelli, Mittelstaedt and many others. The new proposals are characterized by a more general approach, based on a kind of abstraction from the Hilbert space structures. The starting point of the new trends can be summarized as follows. Generally, any physical theory  $\mathbb{T}$  determines a class

of *event-state* systems  $\langle \mathcal{E}, S \rangle$ , where  $\mathcal{E}$  contains the events that may occur to our system, while  $S$  contains the states that a physical system described by the theory may assume. The question arises: what are the abstract conditions that one should postulate for any pair  $\langle \mathcal{E}, S \rangle$ ? In the case of QT, having in mind the Hilbert space model, one is naturally led to the following requirement:

- the set  $\mathcal{E}$  of events should be a good abstraction from the structure of all closed subspaces in a Hilbert space. As a consequence  $\mathcal{E}$  should be at least a  $\sigma$ -complete orthomodular lattice (generally non distributive).
- The set  $S$  of states should be a good abstraction from the *statistical operators* in a Hilbert space, that represent possible states of physical systems. As a consequence, any state shall behave as a *probability measure*, that assigns to any event in  $\mathcal{E}$  a value in the interval  $[0, 1]$ . Both in the concrete and in the abstract case, states may be either *pure* (maximal pieces of information that cannot be consistently extended to a richer knowledge) or *mixtures* (non maximal pieces of information).

In such a framework two basic problems arise:

- I) Is it possible to capture, by means of some abstract conditions that are required for any event-state pair  $\langle \mathcal{E}, S \rangle$ , the behaviour of the concrete Hilbert space pairs?
- II) To what extent should the Hilbert space model be absolutely binding?

The first problem gave rise to a number of attempts to prove a kind of *representation theorem*. More precisely, the main question was: what are the necessary and sufficient conditions for a generic event-state pair  $\langle \mathcal{E}, S \rangle$  that make  $\mathcal{E}$  isomorphic to the lattice of all closed subspaces in a Hilbert space?

Our second problem stimulated the investigation about more and more general quantum structures. Of course, looking for more general structures seems to imply a kind of discontent towards the standard quantum logical approach, based on Hilbert space lattices. The fundamental criticisms that have been moved concern the following items:

- 1) The standard structures seem to determine a kind of *extensional* collapse. In fact, the closed subspaces of a Hilbert space represent at the same time *physical properties* in an *intensional sense* and the *extensions* thereof (sets of states that certainly verify the properties in question). As happens in classical set theoretical semantics, there is no mathematical representative for physical properties in an intensional

sense. Foulis and Randall have called such an extensional collapse “the metaphysical disaster” of the standard quantum logical approach.

- 2) The lattice structure of the closed subspaces automatically renders the quantum proposition system closed under logical conjunction. This seems to imply some counterintuitive consequences from the physical point of view. Suppose two experimental propositions that concern two strongly incompatible quantities, like “the spin in the  $x$  direction is up”, “the spin in the  $y$  direction is down”. In such a situation, the intuition of the quantum physicist seems to suggest the following semantic requirement: the conjunction of our propositions has no definite meaning; for, they cannot be experimentally tested at the same time. As a consequence, the lattice proposition structure seems to be too strong.

An interesting weakening can be obtained by giving up the lattice condition: generally the *infimum* and the *supremum* are assumed to exist only for countable sets of propositions that are pairwise orthogonal. In the recent quantum logical literature an orthomodular partially ordered set that satisfies the above condition is simply called a *quantum logic*. At the same time, by *standard quantum logic* one usually means the complete orthomodular lattice based on the closed subspaces in a Hilbert space. Needless to observe, such a terminology that identifies a *logic* with a particular example of an algebraic structure turns out to be somewhat misleading from the strict logical point of view. As we will see in the next sections, different forms of quantum logic, which represent “genuine logics” according to the standard way of thinking of the logical tradition, can be characterized by convenient abstraction from the physical models.

## 2 ORTHOMODULAR QUANTUM LOGIC AND ORTHOLOGIC

We will first study two interesting examples of logic that represent a natural logical abstraction from the class of all Hilbert space lattices. These are represented respectively by *orthomodular quantum logic* (**OQL**) and by the weaker *orthologic* (**OL**), which for a long time has been also termed *minimal quantum logic*. In fact, the name “minimal quantum logic” appears today quite inappropriate, since a number of weaker forms of quantum logic have been recently investigated. In the following we will use **QL** as an abbreviation for both **OL** and **OQL**.

The language of **QL** consists of a denumerable set of sentential literals and of two primitive connectives:  $\neg$  (*not*),  $\wedge$  (*and*). The notion of *formula* of the language is defined in the expected way. We will use the following metavariables:  $p, q, r, \dots$  for sentential literals and  $\alpha, \beta, \gamma, \dots$  for formulas.



The connective disjunction ( $\vee$ ) is supposed defined via de Morgan's law:

$$\alpha \vee \beta := \neg (\neg \alpha \wedge \neg \beta).$$

The problem concerning the possibility of a well behaved conditional connective will be discussed in the next Section. We will indicate the basic metalogical constants as follows: *not*, *and*, *or*,  $\rightsquigarrow$  (*if...then*), *iff* (*if and only if*),  $\forall$  (for all),  $\exists$  (for at least one).

Because of its historical origin, the most natural characterization of **QL** can be carried out in the framework of an algebraic semantics. It will be expedient to recall first the definition of *ortholattice*:

**DEFINITION 1** (Ortholattice). An *ortholattice* is a structure  $\mathcal{B} = \langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$ , where

- (1.1)  $\langle B, \sqsubseteq, \mathbf{1}, \mathbf{0} \rangle$  is a bounded lattice, where  $\mathbf{1}$  is the *maximum* and  $\mathbf{0}$  is the *minimum*. In other words:
  - (i)  $\sqsubseteq$  is a partial order relation on  $B$  (reflexive, antisymmetric and transitive);
  - (ii) any pair of elements  $a, b$  has an *infimum*  $a \sqcap b$  and a *supremum*  $a \sqcup b$  such that:  
 $a \sqcap b \sqsubseteq a, b$  and  $\forall c: c \sqsubseteq a, b \rightsquigarrow c \sqsubseteq a \sqcap b$ ;  
 $a, b \sqsubseteq a \sqcup b$  and  $\forall c: a, b \sqsubseteq c \rightsquigarrow a \sqcup b \sqsubseteq c$ ;
  - (iii)  $\forall a: \mathbf{0} \sqsubseteq a; a \sqsubseteq \mathbf{1}$ .
- (1.2) the 1-ary operation  $'$  (called *orthocomplement*) satisfies the following conditions:
  - (i)  $a'' = a$  (double negation);
  - (ii)  $a \sqsubseteq b \rightsquigarrow b' \sqsubseteq a'$  (contraposition);
  - (iii)  $a \sqcap a' = \mathbf{0}$  (non contradiction).

Differently from Boolean algebras, ortholattices do not generally satisfy the distributive laws of  $\sqcap$  and  $\sqcup$ . There holds only

$$(a \sqcap b) \sqcup (a \sqcap c) \sqsubseteq a \sqcap (b \sqcup c)$$

and the dual form

$$a \sqcup (b \sqcap c) \sqsubseteq (a \sqcup b) \sqcap (a \sqcup c).$$

The lattice  $\langle C(\mathcal{H}), \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  of all closed subspaces in a Hilbert space  $\mathcal{H}$  is a characteristic example of a non distributive ortholattice.

**DEFINITION 2** (Algebraic realization for **OL**). An *algebraic realization* for **OL** is a pair  $\mathcal{A} = \langle \mathcal{B}, v \rangle$ , consisting of an ortholattice  $\mathcal{B} = \langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  and a *valuation*-function  $v$  that associates to any formula  $\alpha$  of the language an element (*truth-value*) in  $B$ , satisfying the following conditions:

- (i)  $v(\neg\beta) = v(\beta)'$ ;
- (ii)  $v(\beta \wedge \gamma) = v(\beta) \sqcap v(\gamma)$ .

DEFINITION 3 (Truth and logical truth). A formula  $\alpha$  is *true* in a realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  (abbreviated as  $\models_{\mathcal{A}} \alpha$ ) iff  $v(\alpha) = \mathbf{1}$ ;  $\alpha$  is a *logical truth* of **OL** ( $\models_{\mathbf{OL}} \alpha$ ) iff for any algebraic realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$ ,  $\models_{\mathcal{A}} \alpha$ .

When  $\models_{\mathcal{A}} \alpha$ , we will also say that  $\mathcal{A}$  is a *model* of  $\alpha$ ;  $\mathcal{A}$  will be called a *model* of a set of formulas  $T$  ( $\models_{\mathcal{A}} T$ ) iff  $\mathcal{A}$  is a model of any  $\beta \in T$ .

DEFINITION 4 (Consequence in a realization and logical consequence). Let  $T$  be a set of formulas and let  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  be a realization. A formula  $\alpha$  is a *consequence in  $\mathcal{A}$*  of  $T$  ( $T \models_{\mathcal{A}} \alpha$ ) iff for any element  $a$  of  $\mathcal{B}$ : if for any  $\beta \in T$ ,  $a \sqsubseteq v(\beta)$  then  $a \sqsubseteq v(\alpha)$ . A formula  $\alpha$  is a *logical consequence* of  $T$  ( $T \models_{\mathbf{OL}} \alpha$ ) iff for any algebraic realization  $\mathcal{A}$ :  $T \models_{\mathcal{A}} \alpha$ .

Instead of  $\{\alpha\} \models_{\mathbf{OL}} \beta$  we will write  $\alpha \models_{\mathbf{OL}} \beta$ . If  $T$  is finite and equal to  $\{\alpha_1, \dots, \alpha_n\}$ , we will obviously have:  $T \models_{\mathbf{OL}} \alpha$  iff  $v(\alpha_1) \sqcap \dots \sqcap v(\alpha_n) \sqsubseteq v(\alpha)$ . One can easily check that  $\models_{\mathbf{OL}} \alpha$  iff for any  $T$ ,  $T \models_{\mathbf{OL}} \alpha$ .

**OL** can be equivalently characterized also by means of a Kripke-style semantics, which has been first proposed by [Dishkant, 1972]. As is well known, the algebraic semantic approach can be described as founded on the following intuitive idea: interpreting a language essentially means associating to any sentence  $\alpha$  an abstract truth-value or, more generally, an abstract meaning (an element of an algebraic structure). In the Kripkean semantics, instead, one assumes that interpreting a language essentially means associating to any sentence  $\alpha$  the set of the *possible worlds* or *situations* where  $\alpha$  holds. This set, which represents the *extensional meaning* of  $\alpha$ , is called the *proposition* associated to  $\alpha$  (or simply *the proposition of  $\alpha$* ). Hence, generally, a Kripkean realization for a logic **L** will have the form:

$$\mathcal{K} = \langle I, \vec{R}_i, \vec{o}_j, \Pi, \rho \rangle,$$

where

- (i)  $I$  is a non-empty set of possible worlds possibly correlated by relations in the sequence  $\vec{R}_i$  and operations in the sequence  $\vec{o}_j$ . In most cases, we have only one binary relation  $R$ , called *accessibility* relation.
- (ii)  $\Pi$  is a set of sets of possible worlds, representing possible propositions of sentences. Any proposition and the total set of propositions  $\Pi$  must satisfy convenient closure conditions that depend on the particular logic.
- (iii)  $\rho$  transforms sentences into propositions preserving the logical form.

The Kripkean realizations that turn out to be adequate for **OL** have only one accessibility relation, which is reflexive and symmetric. As is well known, many logics, that are stronger than *positive logic*, are instead characterized by Kripkean realizations where the accessibility relation is at least reflexive and transitive. As an example, let us think of *intuitionistic logic*. From an intuitive point of view, one can easily understand the reason why semantic models with a reflexive and symmetric accessibility relation may be physically significant. In fact, physical theories are not generally concerned with *possible evolutions of states of knowledge* with respect to a constant world, but rather with *sets of physical situations* that may be *similar*, where *states of knowledge* must single out some *invariants*. And similarity relations are reflexive and symmetric, but generally not transitive.

Let us now introduce the basic concepts of a Kripkean semantics for **OL**.  
**DEFINITION 5 (Orthoframe).** An *orthoframe* is a relational structure  $\mathcal{F} = \langle I, R \rangle$ , where  $I$  is a non-empty set (called the set of *worlds*) and  $R$  (the *accessibility relation*) is a binary reflexive and symmetric relation on  $I$ .

Given an orthoframe, we will use  $i, j, k, \dots$  as variables ranging over the set of worlds. Instead of  $Rij$  (not  $Rij$ ) we will also write  $i \perp j$  ( $i \perp j$ ).

**DEFINITION 6 (Orthocomplement in an orthoframe).** Let  $\mathcal{F} = \langle I, R \rangle$  be an orthoframe. For any set of worlds  $X \subseteq I$ , the *orthocomplement*  $X'$  of  $X$  is defined as follows:

$$X' = \{i \mid \forall j(j \in X \rightsquigarrow j \perp i)\}.$$

In other words,  $X$  is the set of all worlds that are inaccessible to all elements of  $X'$ . Instead of  $i \in X'$ , we will also write  $i \perp X$  (and we will read it as “ $i$  is orthogonal to the set  $X$ ”). Instead of  $i \notin X'$ , we will also write  $i \not\perp X$ .

**DEFINITION 7 (Proposition).** Let  $\mathcal{F} = \langle I, R \rangle$  be an orthoframe. A set of worlds  $X$  is called a *proposition* of  $\mathcal{F}$  iff it satisfies the following condition:

$$\forall i [i \in X \text{ iff } \forall j(i \not\perp j \rightsquigarrow j \not\perp X)].$$

In other words, a proposition is a set of worlds  $X$  that contains all and only the worlds whose accessible worlds are not inaccessible to  $X$ . Notice that the conditional  $i \in X \rightsquigarrow \forall j(i \not\perp j \rightsquigarrow j \not\perp X)$  trivially holds for any set of worlds  $X$ .

Our definition of proposition represents a quite general notion of “possible meaning of a formula”, that can be significantly extended also to other logics. Suppose for instance, a Kripkean frame  $\mathcal{F} = \langle I, R \rangle$ , where the accessibility relation is at least reflexive and transitive (as happens in the Kripkean semantics for intuitionistic logic). Then a set of worlds  $X$  turns out to be a proposition (in the sense of Definition 7) iff it is *R-closed* (i.e.,

$\forall ij(i \in X \text{ and } Rij \leadsto j \in X)$ ). And  $R$ -closed sets of worlds represent precisely the possible meanings of formulas in the Kripkean characterization of intuitionistic logic.

LEMMA 8. *Let  $\mathcal{F}$  be an orthoframe and  $X$  a set of worlds of  $\mathcal{F}$ .*

$$(8.1) \quad X \text{ is a proposition of } \mathcal{F} \text{ iff } \forall i [i \notin X \leadsto \exists j (i \not\perp j \text{ and } j \perp X)]$$

$$(8.2) \quad X \text{ is a proposition of } \mathcal{F} \text{ iff } X = X''.$$

LEMMA 9. *Let  $\mathcal{F} = \langle I, R \rangle$  be an orthoframe.*

$$(9.1) \quad I \text{ and } \emptyset \text{ are propositions.}$$

$$(9.2) \quad \text{If } X \text{ is any set of worlds, then } X' \text{ is a proposition.}$$

$$(9.3) \quad \text{If } C \text{ is a family of propositions, then } \bigcap C \text{ is a proposition.}$$

DEFINITION 10 (Kripkean realization for **OL**). A *Kripkean realization* for **OL** is a system  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$ , where:

- (i)  $\langle I, R \rangle$  is an orthoframe and  $\Pi$  is a set of propositions of the frame that contains  $\emptyset, I$  and is closed under the orthocomplement  $'$  and the set-theoretic intersection  $\cap$ ;
- (ii)  $\rho$  is a function that associates to any formula  $\alpha$  a proposition in  $\Pi$ , satisfying the following conditions:

$$\rho(\neg\beta) = \rho(\beta)';$$

$$\rho(\beta \wedge \gamma) = \rho(\beta) \cap \rho(\gamma).$$

Instead of  $i \in \rho(\alpha)$ , we will also write  $i \models \alpha$  (or,  $i \models_{\mathcal{K}} \alpha$ , in case of possible confusions) and we will read: “ $\alpha$  is true in the world  $i$ ”. If  $T$  is a set of formulas,  $i \models T$  will mean  $i \models \beta$  for any  $\beta \in T$ .

THEOREM 11. *For any Kripkean realization  $\mathcal{K}$  and any formula  $\alpha$ :*

$$i \models \alpha \text{ iff } \forall j \not\perp i \exists k \not\perp j (k \models \alpha).$$

**Proof.** Since the accessibility relation is symmetric, the left to right implication is trivial. Let us prove  $i \not\models \alpha \leadsto \text{not} \forall j \not\perp i \exists k \not\perp j (k \models \alpha)$ , which is equivalent to  $i \notin \rho(\alpha) \leadsto \exists j \not\perp i \forall k \not\perp j (k \notin \rho(\alpha))$ . Suppose  $i \notin \rho(\alpha)$ . Since  $\rho(\alpha)$  is a proposition, by Lemma 8.1 there holds for a certain  $j$ :  $j \not\perp i$  and  $j \perp \rho(\alpha)$ . Let  $k \not\perp j$ , and suppose, by contradiction,  $k \in \rho(\alpha)$ . Since  $j \perp \rho(\alpha)$ , there follows  $j \perp k$ , against  $k \not\perp j$ . Consequently,  $\exists j \not\perp i \forall k \not\perp j (k \notin \rho(\alpha))$ . ■

LEMMA 12. *In any Kripkean realization  $\mathcal{K}$ :*

$$(12.1) \quad i \models \neg\beta \text{ iff } \forall j \not\perp i (j \not\models \beta);$$

$$(12.2) \quad i \models \beta \wedge \gamma \text{ iff } i \models \beta \text{ and } i \models \gamma.$$

DEFINITION 13 (Truth and logical truth). A formula  $\alpha$  is *true* in a realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  (abbreviated  $\models_{\mathcal{K}} \alpha$ ) iff  $\rho(\alpha) = I$ ;  $\alpha$  is a *logical truth* of **OL** ( $\models_{\text{OL}} \alpha$ ) iff for any realization  $\mathcal{K}$ ,  $\models_{\mathcal{K}} \alpha$ .

When  $\models_{\mathcal{K}} \alpha$ , we will also say that  $\mathcal{K}$  is a *model* of  $\alpha$ . Similarly in the case of a set of formulas  $T$ .

DEFINITION 14 (Consequence in a realization and logical consequence).

Let  $T$  be a set of formulas and let  $\mathcal{K}$  be a realization. A formula  $\alpha$  is a *consequence in  $\mathcal{K}$*  of  $T$  ( $T \models_{\mathcal{K}} \alpha$ ) iff for any world  $i$  of  $\mathcal{K}$ ,  $i \models T \curvearrowright i \models \alpha$ . A formula  $\alpha$  is a *logical consequence* of  $T$  ( $T \models_{\text{OL}} \alpha$ ) iff for any realization  $\mathcal{K}$ :  $T \models_{\mathcal{K}} \alpha$ . When no confusion is possible we will simply write  $T \models \alpha$ .

Now we will prove that the algebraic and the Kripkean semantics for **OL** characterize the same logic. Let us abbreviate the metalogical expressions “ $\alpha$  is a logical truth of **OL** according to the algebraic semantics”, “ $\alpha$  is a logical consequence in **OL** of  $T$  according to the algebraic semantics”, “ $\alpha$  is a logical truth of **OL** according to the Kripkean semantics”, “ $\alpha$  is a logical consequence in **OL** of  $T$  according to the Kripkean semantics”, by  $\stackrel{\text{A}}{\models}_{\text{OL}} \alpha$ ,  $T \stackrel{\text{A}}{\models}_{\text{OL}} \alpha$ ,  $\stackrel{\text{K}}{\models}_{\text{OL}} \alpha$ ,  $T \stackrel{\text{K}}{\models}_{\text{OL}} \alpha$ , respectively.

THEOREM 15.  $\stackrel{\text{A}}{\models}_{\text{OL}} \alpha$  iff  $\stackrel{\text{K}}{\models}_{\text{OL}} \alpha$ , for any  $\alpha$ .

The Theorem is an immediate corollary of the following Lemma:

LEMMA 16.

(16.1) For any algebraic realization  $\mathcal{A}$  there exists a Kripkean realization  $\mathcal{K}^{\mathcal{A}}$  such that for any  $\alpha$ ,  $\models_{\mathcal{A}} \alpha$  iff  $\models_{\mathcal{K}^{\mathcal{A}}} \alpha$ .

(16.2) For any Kripkean realization  $\mathcal{K}$  there exists an algebraic realization  $\mathcal{A}^{\mathcal{K}}$  such that for any  $\alpha$ ,  $\models_{\mathcal{K}} \alpha$  iff  $\models_{\mathcal{A}^{\mathcal{K}}} \alpha$ .

### Sketch of the proof

(16.1) The basic intuitive idea of the proof is the following: any algebraic realization can be canonically transformed into a Kripkean realization by identifying the set of worlds with the set of all non-null elements of the algebra, the accessibility-relation with the non-orthogonality relation in the algebra, and finally the set of propositions with the set of all *principal quasi-ideals* (i.e., the principal ideals, devoided of the zero-element). More precisely, given  $\mathcal{A} = \langle \mathcal{B}, v \rangle$ , the Kripkean realization  $\mathcal{K}^{\mathcal{A}} = \langle I, R, \Pi, \rho \rangle$  is defined as follows:

$$\begin{aligned} I &= \{b \in B \mid b \neq \mathbf{0}\}; \\ Rij &\text{ iff } i \not\perp j'; \\ \Pi &= \{\{b \in B \mid b \neq \mathbf{0} \text{ and } b \sqsubseteq a\} \mid a \in B\}; \end{aligned}$$

$$\rho(p) = \{b \in I \mid b \sqsubseteq v(p)\}.$$

One can easily check that  $\mathcal{K}^{\mathcal{A}}$  is a “good” Kripkean realization; further, there holds, for any  $\alpha$ :  $\rho(\alpha) = \{b \in B \mid b \neq \mathbf{0} \text{ and } b \sqsubseteq v(\alpha)\}$ . Consequently,  $\models_{\mathcal{A}} \alpha$  iff  $\models_{\mathcal{K}^{\mathcal{A}}} \alpha$ .

(16.2) Any Kripkean realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  can be canonically transformed into an algebraic realization  $\mathcal{A}^{\mathcal{K}} = \langle \mathcal{B}, v \rangle$  by putting:

$$\begin{aligned} B &= \Pi; \\ \text{for any } a, b \in B: & a \sqsubseteq b \text{ iff } a \subseteq b; \\ a' &= \{i \in I \mid i \perp a\}; \\ \mathbf{1} &= I; \quad \mathbf{0} = \emptyset; \\ v(p) &= \rho(p). \end{aligned}$$

It turns out that  $\mathcal{B}$  is an ortholattice. Further, for any  $\alpha$ ,  $v(\alpha) = \rho(\alpha)$ . Consequently:  $\models_{\mathcal{K}} \alpha$  iff  $\models_{\mathcal{A}^{\mathcal{K}}} \alpha$ . ■

**THEOREM 17.**  $T \stackrel{\mathbf{A}}{\underset{\mathbf{OL}}{\models}} \alpha$  iff  $T \stackrel{\mathbf{K}}{\underset{\mathbf{OL}}{\models}} \alpha$ .

**Proof.** In order to prove the left to right implication, suppose by contradiction:  $T \stackrel{\mathbf{A}}{\underset{\mathbf{OL}}{\models}} \alpha$  and  $T \not\stackrel{\mathbf{K}}{\underset{\mathbf{OL}}{\models}} \alpha$ . Hence there exists a Kripkean realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  and a world  $i$  of  $\mathcal{K}$  such that  $i \models T$  and  $i \not\models \alpha$ . One can easily see that  $\mathcal{K}$  can be transformed into  $\mathcal{K}^{\circ} = \langle I, R, \Pi^{\circ}, \rho \rangle$  where  $\Pi^{\circ}$  is the smallest subset of the power-set of  $I$ , that includes  $\Pi$  and is closed under infinitary intersection. Owing to Lemma 9.3,  $\mathcal{K}^{\circ}$  is a “good” Kripkean realization for  $\mathbf{OL}$  and for any  $\beta$ ,  $\rho(\beta)$  turns out to be the same proposition in  $\mathcal{K}$  and in  $\mathcal{K}^{\circ}$ . Consequently, also in  $\mathcal{K}^{\circ}$ , there holds:  $i \models T$  and  $i \not\models \alpha$ . Let us now consider  $\mathcal{A}^{\mathcal{K}^{\circ}}$ . The algebra  $\mathcal{B}$  of  $\mathcal{A}^{\mathcal{K}^{\circ}}$  is complete, because  $\Pi^{\circ}$  is closed under infinitary intersection. Hence,  $\bigcap \{\rho(\beta) \mid \beta \in T\}$  is an element of  $B$ . Since  $i \models \beta$  for any  $\beta \in T$ , we will have  $i \in \bigcap \{\rho(\beta) \mid \beta \in T\}$ . Thus there is an element of  $B$ , which is less or equal than  $v(\beta)(= \rho(\beta))$  for any  $\beta \in T$ , but is not less or equal than  $v(\alpha)(= \rho(\alpha))$ , because  $i \notin \rho(\alpha)$ . This contradicts the hypothesis  $T \stackrel{\mathbf{A}}{\underset{\mathbf{OL}}{\models}} \alpha$ .

The right to left implication is trivial. ■

Let us now turn to a semantic characterization of  $\mathbf{OQL}$ . We will first recall the definition of orthomodular lattice.

**DEFINITION 18** (Orthomodular lattice). An *orthomodular lattice* is an ortholattice  $\mathcal{B} = \langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  such that for any  $a, b \in B$ :

$$a \sqcap (a' \sqcup (a \sqcap b)) \sqsubseteq b.$$

Orthomodularity clearly represents a weak form of distributivity.

**LEMMA 19.** *Let  $\mathcal{B}$  be an ortholattice. The following conditions are equivalent:*

- (i)  $\mathcal{B}$  is orthomodular.
- (ii) For any  $a, b \in B$ :  $a \sqsubseteq b \iff b = a \sqcup (a' \sqcap b)$ .
- (iii) For any  $a, b \in B$ :  $a \sqsubseteq b$  iff  $a \sqcap (a \sqcap b)' = \mathbf{0}$ .
- (iv) For any  $a, b \in B$ :  $a \sqsubseteq b$  and  $a' \sqcap b = \mathbf{0} \iff a = b$ .

The property considered in (19(iii)) represents a significant weakening of the Boolean condition:

$$a \sqsubseteq b \text{ iff } a \sqcap b' = \mathbf{0}.$$

**DEFINITION 20** (Algebraic realization for **OQL**). An *algebraic realization* for **OQL** is an algebraic realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  for **OL**, where  $\mathcal{B}$  is an orthomodular lattice.

The definitions of truth, logical truth and logical consequence in **OQL** are analogous to the corresponding definitions of **OL**.

Like **OL**, also **OQL** can be characterized by means of a Kripkean semantics.

**DEFINITION 21** (Kripkean realization for **OQL**). A *Kripkean realization* for **OQL** is a Kripkean realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  for **OL**, where the set of propositions  $\Pi$  satisfies the *orthomodular property*:

$$X \not\subseteq Y \iff X \cap (X \cap Y)' \neq \emptyset.$$

The definitions of *truth*, *logical truth* and *logical consequence* in **OQL** are analogous to the corresponding definitions of **OL**. Also in the case of **OQL** one can show:

**THEOREM 22.**  $\stackrel{\mathbf{A}}{\underset{\mathbf{OQL}}{\vdash}} \alpha \text{ iff } \stackrel{\mathbf{K}}{\underset{\mathbf{OQL}}{\vdash}} \alpha.$

The Theorem is an immediate corollary of Lemma 16 and of the following lemma:

**LEMMA 23.**

- (23.1) If  $\mathcal{A}$  is orthomodular then  $\mathcal{K}^{\mathcal{A}}$  is orthomodular;
- (23.2) If  $\mathcal{K}$  is orthomodular then  $\mathcal{A}^{\mathcal{K}}$  is orthomodular.

**Proof.** (23.1) We have to prove  $X \not\subseteq Y \iff X \cap (X \cap Y)' \neq \emptyset$  for any propositions  $X, Y$  of  $\mathcal{K}^{\mathcal{A}}$ . Suppose  $X \not\subseteq Y$ . By definition of proposition in  $\mathcal{K}^{\mathcal{A}}$ :

$$\begin{aligned} X &= \{b \mid b \neq \mathbf{0} \text{ and } b \sqsubseteq x\} \text{ for a given } x; \\ Y &= \{b \mid b \neq \mathbf{0} \text{ and } b \sqsubseteq y\} \text{ for a given } y; \end{aligned}$$

Consequently,  $x \not\subseteq y$ , and by Lemma 19:  $x \sqcap (x \sqcap y)' \neq \mathbf{0}$ , because  $\mathcal{A}$  is orthomodular. Hence,  $x \sqcap (x \sqcap y)'$  is a world in  $\mathcal{K}^{\mathcal{A}}$ . In order to prove

$X \cap (X \cap Y)' \neq \emptyset$ , it is sufficient to prove  $x \sqcap (x \sqcap y)' \in X \cap (X \cap Y)'$ . There holds trivially  $x \sqcap (x \sqcap y)' \in X$ . Further,  $x \sqcap (x \sqcap y)' \in (X \cap Y)'$ , because  $(x \sqcap y)'$  is the generator of the quasi-ideal  $(X \cap Y)'$ . Consequently,  $x \sqcap (x \sqcap y)' \in X \cap (X \cap Y)'$ .

(23.2) Let  $\mathcal{K}$  be orthomodular. Then for any  $X, Y \in \Pi$ :

$$X \not\subseteq Y \rightsquigarrow X \cap (X \cap Y)' \neq \emptyset.$$

One can trivially prove:

$$X \cap (X \cap Y)' \neq \emptyset \rightsquigarrow X \not\subseteq Y.$$

Hence, by Lemma 19, the algebra  $\mathcal{B}$  of  $\mathcal{A}^{\mathcal{K}}$  is orthomodular.  $\blacksquare$

As to the concept of logical consequence, the proof we have given for **OL** (Theorem 17) cannot be automatically extended to the case of **OQL**. The critical point is represented by the transformation of  $\mathcal{K}$  into  $\mathcal{K}^\circ$  whose set of propositions is closed under infinitary intersection:  $\mathcal{K}^\circ$  is trivially a “good” **OL**-realization; at the same time, it is not granted that  $\mathcal{K}^\circ$  preserves the orthomodular property. One can easily prove:

THEOREM 24.  $T \stackrel{\mathcal{K}}{\underset{\text{OQL}}{=}} \alpha \rightsquigarrow T \stackrel{\mathcal{A}}{\underset{\text{OQL}}{=}} \alpha$ .

The inverse relation has been proved by [Minari, 1987]:

THEOREM 25.  $T \stackrel{\mathcal{A}}{\underset{\text{OQL}}{=}} \alpha \rightsquigarrow T \stackrel{\mathcal{K}}{\underset{\text{OQL}}{=}} \alpha$ .

Are there any significant structural relations between  $\mathcal{A}$  and  $\mathcal{K}^{\mathcal{A}^{\mathcal{K}}}$  and between  $\mathcal{K}$  and  $\mathcal{A}^{\mathcal{K}^{\mathcal{A}}}$ ? The question admits a very strong answer in the case of  $\mathcal{A}$  and  $\mathcal{K}^{\mathcal{A}^{\mathcal{K}}}$ .

THEOREM 26.  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  and  $\mathcal{A}^{\mathcal{K}^{\mathcal{A}}} = \langle \mathcal{B}^*, v^* \rangle$  are isomorphic realizations.

**Sketch of the proof** Let us define the function  $\psi : B \rightarrow B^*$  in the following way:

$$\psi(a) = \{b \mid b \neq \mathbf{0} \text{ and } b \sqsubseteq a\} \text{ for any } a \in B.$$

One can easily check that: (1)  $\psi$  is an isomorphism (from  $\mathcal{B}$  onto  $\mathcal{B}^*$ ); (2)  $v^*(p) = \psi(v(p))$  for any atomic formula  $p$ .  $\blacksquare$

At the same time, in the case of  $\mathcal{K}$  and  $\mathcal{K}^{\mathcal{A}^{\mathcal{K}}}$ , there is no natural correspondence between  $I$  and  $\Pi$ . As a consequence, one can prove only the weaker relation:

THEOREM 27. Given  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  and  $\mathcal{K}^{\mathcal{A}^{\mathcal{K}}} = \langle I^*, R^*, \Pi^*, \rho^* \rangle$ , there holds:

$$\rho^*(\alpha) = \{X \in \Pi \mid X \subseteq \rho(\alpha)\}, \text{ for any } \alpha.$$



In the class of all Kripkean realizations for **QL**, the realizations  $\mathcal{K}^A$  (which have been obtained by canonical transformation of an algebraic realization  $A$ ) present some interesting properties, which are summarized by the following theorem.

**THEOREM 28.** *In any  $\mathcal{K}^A = \langle I, R, \Pi, \rho \rangle$  there is a one-to-one correspondence  $\phi$  between the set of worlds  $I$  and the set of propositions  $\Pi - \{\emptyset\}$  such that:*

- (28.1)  $i \in \phi(i)$ ;
- (28.2)  $i \not\perp j$  iff  $\phi(i) \not\subseteq \phi(j)'$ ;
- (28.3)  $\forall X \in \Pi: i \in X$  iff  $\forall k \in \phi(i) (k \in X)$ .

**Sketch of the proof** Let us take as  $\phi(i)$  the quasi-ideal generated by  $i$ . ■

Theorem 28 suggests to isolate, in the class of all  $\mathcal{K}$ , an interesting subclass of Kripkean realizations, that we will call *algebraically adequate*.

**DEFINITION 29.** A Kripkean realization  $\mathcal{K}$  is *algebraically adequate* iff it satisfies the conditions of Theorem 28.

When restricting to the class of all algebraically adequate Kripkean realizations one can prove:

**THEOREM 30.**  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  and  $\mathcal{K}^{A^K} = \langle I^*, R^*, \Pi^*, \rho^* \rangle$  are isomorphic realizations; i.e., there exists a bijective function  $\psi$  from  $I$  onto  $I^*$  such that:

- (30.1)  $Rij$  iff  $R^*\psi(i)\psi(j)$ , for any  $i, j \in I$ ;
- (30.2)  $\Pi^* = \{\psi(X) \mid X \in \Pi\}$ , where  $\psi(X) := \{\psi(i) \mid i \in X\}$ ;
- (30.3)  $\rho^*(p) = \psi(\rho(p))$ , for any atomic formula  $p$ .

One can easily show that the class of all algebraically adequate Kripkean realizations determines the same concept of logical consequence that is determined by the larger class of all possible realizations.

The Kripkean characterization of **QL** turns out to have a quite natural physical interpretation. As we have seen in the Introduction, the mathematical formalism of quantum theory (QT) associates to any *physical system*  $\mathcal{S}$  a Hilbert space  $\mathcal{H}$ , while the *pure states* of  $\mathcal{S}$  are mathematically represented by unitary vectors  $\psi$  of  $\mathcal{H}$ . Let us now consider an elementary sublanguage  $\mathcal{L}^Q$  of QT, whose atomic formulas represent possible measurement reports (i.e., statements of the form “the value for the observable  $Q$  lies in the Borel set  $\Delta$ ”) and suppose  $\mathcal{L}^Q$  closed under the quantum logical connectives. Given a physical system  $\mathcal{S}$  (whose associated Hilbert space is  $\mathcal{H}$ ), one can define a natural Kripkean realization for the language  $\mathcal{L}^Q$  as follows:

$$\mathcal{K}^{\mathcal{S}} = \langle I, R, \Pi, \rho \rangle,$$

where:

- $I$  is the set of all pure states  $\psi$  of  $\mathcal{S}$ .
- $R$  is the non-orthogonality relation between vectors (in other words, two pure states are accessible iff their inner product is different from zero).
- $\Pi$  is the set of all propositions that is univocally determined by the set of all closed subspaces of  $\mathcal{H}$  (one can easily check that the set of all unitary vectors of any subspace is a proposition).
- For any atomic formula  $p$ ,  $\rho(p)$  is the proposition containing all the pure states that assign to  $p$  probability-value 1.

Interestingly enough, the accessibility relation turns out to have the following physical meaning:  $Rij$  iff  $j$  is a pure state into which  $i$  can be transformed after the performance of a physical measurement that concern an observable of the system.

### 3 THE IMPLICATION PROBLEM

Differently from most weak logics, **QL** gives rise to a critical “implication-problem”. All conditional connectives one can reasonably introduce in **QL** are, to a certain extent, anomalous; for, they do not share most of the characteristic properties that are satisfied by the *positive conditionals* (which are governed by a logic that is at least as strong as *positive logic*). Just the failure of a well-behaved conditional led some authors to the conclusion that **QL** cannot be a “real” logic. In spite of these difficulties, these days one cannot help recognizing that **QL** admits a set of different implicational connectives, even if none of them has a *positive* behaviour. Let us first propose a general semantic condition for a logical connective to be classified as an implication-connective.

DEFINITION 31. In any semantics, a binary connective  $\overset{*}{\rightarrow}$  is called an *implication-connective* iff it satisfies at least the two following conditions:

$$(31.1) \quad \alpha \overset{*}{\rightarrow} \alpha \text{ is always true (identity);}$$

$$(31.2) \quad \text{if } \alpha \text{ is true and } \alpha \overset{*}{\rightarrow} \beta \text{ is true then } \beta \text{ is true (modus ponens).}$$

In the particular case of **QL**, one can easily obtain:

LEMMA 32. *A sufficient condition for a connective  $\overset{*}{\rightarrow}$  to be an implication-connective is:*

- (i) *in the algebraic semantics: for any realization  $\mathcal{A} = \langle \mathcal{A}, v \rangle$ ,  $\models_{\mathcal{A}} \alpha \overset{*}{\rightarrow} \beta$  iff  $v(\alpha) \sqsubseteq v(\beta)$ ;*

- (ii) *in the Kripkean semantics: for any realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$ ,*  
 $\models_{\mathcal{K}} \alpha \overset{*}{\rightarrow} \beta$  *iff*  $\rho(\alpha) \subseteq \rho(\beta)$ .

In **QL** it seems reasonable to assume the sufficient condition of Lemma 32 as a minimal condition for a connective to be an implication-connective.

Suppose we have independently defined two different implication-connectives in the algebraic and in the Kripkean semantics. When shall we admit that they represent the “same logical connective”? A reasonable answer to this question is represented by the following convention:

**DEFINITION 33.** Let  $\overset{A}{*}$  be a binary connective defined in the algebraic semantics and  $\overset{K}{*}$  a binary connective defined in the Kripkean semantics:  $\overset{A}{*}$  and  $\overset{K}{*}$  represent the *same logical connective* iff the following conditions are satisfied:

- (33.1) given any  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  and given the corresponding  $\mathcal{K}^{\mathcal{A}} = \langle I, R, \Pi, \rho \rangle$ ,  $\rho(\alpha \overset{K}{*} \beta)$  is the quasi-ideal generated by  $v(\alpha \overset{A}{*} \beta)$ ;
- (33.2) given any  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  and given the corresponding  $\mathcal{A}^{\mathcal{K}} = \langle \mathcal{B}, v \rangle$ , there holds:  $v(\alpha \overset{A}{*} \beta) = \rho(\alpha \overset{K}{*} \beta)$ .

We will now consider different possible semantic characterizations of an implication-connective in **QL**. Differently from classical logic, in **QL** a material conditional defined by *Philo-law* ( $\alpha \rightarrow \beta := \neg\alpha \vee \beta$ ), does not give rise to an implication-connective. For, there are algebraic realizations  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  such that  $v(\neg\alpha \vee \beta) = \mathbf{1}$ , while  $v(\alpha) \not\sqsubseteq v(\beta)$ . Further, ortholattices and orthomodular lattices are not, generally, *pseudocomplemented* lattices: in other words, given  $a, b \in B$ , the maximum  $c$  such that  $a \sqcap c \sqsubseteq b$  does not necessarily exist in  $B$ . In fact, one can prove [Birkhoff, 1995] that any pseudocomplemented lattice is distributive.

We will first consider the case of *polynomial conditionals*, that can be defined in terms of the connectives  $\wedge, \vee, \neg$ . In the algebraic semantics, the minimal requirement of Lemma 32 restricts the choice only to five possible candidates [Kalmbach, 1983]. This result follows from the fact that in the orthomodular lattice freely generated by two elements there are only five polynomial binary operations  $\circ$  satisfying the condition  $a \sqsubseteq b$  iff  $a \circ b = \mathbf{1}$ . These are our five candidates:

- (i)  $v(\alpha \rightarrow_1 \beta) = v(\alpha)' \sqcup (v(\alpha) \sqcap v(\beta))$ .
- (ii)  $v(\alpha \rightarrow_2 \beta) = v(\beta) \sqcup (v(\alpha)' \sqcap v(\beta)')$ .
- (iii)  $v(\alpha \rightarrow_3 \beta) = (v(\alpha)' \sqcap v(\beta)) \sqcup (v(\alpha) \sqcap v(\beta)) \sqcup (v(\alpha)' \sqcap v(\beta)')$ .
- (iv)  $v(\alpha \rightarrow_4 \beta) = (v(\alpha)' \sqcap v(\beta)) \sqcup (v(\alpha) \sqcap v(\beta)) \sqcup ((v(\alpha)' \sqcup v(\beta)) \sqcap v(\beta)')$ .

$$(v) \quad v(\alpha \rightarrow_5 \beta) = (v(\alpha)' \sqcap v(\beta)) \sqcup (v(\alpha)' \sqcap v(\beta)') \sqcup (v(\alpha) \sqcap (v(\alpha)' \sqcup v(\beta))).$$

The corresponding five implication-connectives in the Kripkean semantics can be easily obtained. It is not hard to see that for any  $i$  ( $1 \leq i \leq 5$ ),  $\rightarrow_i$  represents the same logical connective in both semantics (in the sense of Definition 33).

**THEOREM 34.** *The polynomial conditionals  $\rightarrow_i$  ( $1 \leq i \leq 5$ ) are implication-connectives in **OQL**; at the same time they are not implication-connectives in **OL**.*

**Proof.** Since  $\rightarrow_i$  represent the same connective in both semantics, it will be sufficient to refer to the algebraic semantics. As an example, let us prove the theorem for  $i = 1$  (the other cases are similar). First we have to prove  $v(\alpha) \sqsubseteq v(\beta)$  iff  $\mathbf{1} = v(\alpha \rightarrow_1 \beta) = v(\alpha)' \sqcup (v(\alpha) \sqcap v(\beta))$ , which is equivalent to  $v(\alpha) \sqsubseteq v(\beta)$  iff  $v(\alpha) \sqcap (v(\alpha) \sqcap v(\beta))' = \mathbf{0}$ . From Lemma 19, we know that the latter condition holds for any pair of elements of  $B$  iff  $\mathcal{B}$  is orthomodular. This proves at the same time that  $\rightarrow_1$  is an implication-connective in **OQL**, but cannot be an implication-connective in **OL**. ■

Interestingly enough, each polynomial conditional  $\rightarrow_i$  represents a good weakening of the classical material conditional. In order to show this result, let us first introduce an important relation that describes a “Boolean mutual behaviour” between elements of an orthomodular lattice.

**DEFINITION 35** (Compatibility).

Two elements  $a, b$  of an orthomodular lattice  $\mathcal{B}$  are *compatible* iff

$$a = (a \sqcap b') \sqcup (a \sqcap b).$$

One can prove that  $a, b$  are compatible iff the subalgebra of  $\mathcal{B}$  generated by  $\{a, b\}$  is Boolean.

**THEOREM 36.** *For any algebraic realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  and for any  $\alpha, \beta$ :*

$$v(\alpha \rightarrow_i \beta) = v(\alpha)' \sqcup v(\beta) \text{ iff } v(\alpha) \text{ and } v(\beta) \text{ are compatible.}$$

As previously mentioned, Boolean algebras are pseudocomplemented lattices. Therefore they satisfy the following condition for any  $a, b, c$ :

$$c \sqcap a \sqsubseteq b \text{ iff } c \sqsubseteq a \rightsquigarrow b,$$

where:  $a \rightsquigarrow b := a' \sqcup b$ .

An orthomodular lattice  $\mathcal{B}$  turns out to be a Boolean algebra iff for any algebraic realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$ , any  $i$  ( $1 \leq i \leq 5$ ) and any  $\alpha, \beta$  the following *import-export* condition is satisfied:

$$v(\gamma) \sqcap v(\alpha) \sqsubseteq v(\beta) \text{ iff } v(\gamma) \sqsubseteq v(\alpha \rightarrow_i \beta).$$

In order to single out a unique polynomial conditional, various weakenings of the import-export condition have been proposed. For instance the following condition (which we will call *weak import-export*):

$$v(\gamma) \sqcap v(\alpha) \sqsubseteq v(\beta) \text{ iff } v(\gamma) \sqsubseteq v(\alpha) \rightarrow_i v(\beta), \\ \text{whenever } v(\alpha) \text{ and } v(\beta) \text{ are compatible.}$$

One can prove [Hardegree, 1975; Mittelstaedt, 1972] that a polynomial conditional  $\rightarrow_i$  satisfies the weak import-export condition iff  $i = 1$ . As a consequence, we can conclude that  $\rightarrow_1$  represents, in a sense, the best possible approximation for a material conditional in quantum logic. This connective (often called *Sasaki-hook*) was originally proposed by [Mittelstaedt, 1972] and [Finch, 1970], and was further investigated by [Hardegree, 1976] and other authors. In the following, we will usually write  $\rightarrow$  instead of  $\rightarrow_1$  and we will neglect the other four polynomial conditionals.

Some important positive laws that are violated by our quantum logical conditional are the following:

$$\begin{aligned} &\alpha \rightarrow (\beta \rightarrow \alpha); \\ &(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow ((\alpha \rightarrow \beta) \rightarrow (\alpha \rightarrow \gamma)); \\ &(\alpha \rightarrow \beta) \rightarrow ((\beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \gamma)); \\ &(\alpha \wedge \beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow (\beta \rightarrow \gamma)); \\ &(\alpha \rightarrow (\beta \rightarrow \gamma)) \rightarrow (\beta \rightarrow (\alpha \rightarrow \gamma)). \end{aligned}$$

This somewhat “anomalous” behaviour has suggested that one is dealing with a kind of *counterfactual conditional*. Such a conjecture seems to be confirmed by some important physical examples. Let us consider again the class of the Kripkean realizations of the sublanguage  $\mathcal{L}^{\mathbf{Q}}$  of QT (whose atomic sentences express measurement reports). And let  $K^{\mathcal{S}} = \langle I, R, \Pi, \rho \rangle$  represent a Kripkean realization of our language, which is associated to a physical system  $\mathcal{S}$ . As [Hardegree, 1975] has shown, in such a case the conditional  $\rightarrow$  turns out to receive a quite natural counterfactual interpretation (in the sense of Stalnaker). More precisely, one can define, for any formula  $\alpha$ , a partial *Stalnaker-function*  $f_\alpha$  in the following way:

$$f_\alpha : \text{Dom}(f_\alpha) \rightarrow I,$$

where:

$$\text{Dom}(f_\alpha) = \{i \in I \mid i \not\perp \rho(\alpha)\}.$$

In other words,  $f_\alpha$  is defined for all and only the states that are not orthogonal to the proposition of  $\alpha$ .

If  $i \in \text{Dom}(f_\alpha)$ , then:

$$f_\alpha(i) = j^{P_{\rho(\alpha)}i},$$

where  $P_{\rho(\alpha)}$  is the projection that is uniquely associated with the closed subspace determined by  $\rho(\alpha)$ , and  $j^{P_{\rho(\alpha)}i}$  is the normalized vector determined by  $P_{\rho(\alpha)}i$ . There holds:

$$i \models \alpha \rightarrow \beta \text{ iff either } \forall j \not\models i(j \not\models \alpha) \text{ or } f_\alpha(i) \models \beta.$$

From an intuitive point of view, one can say that  $f_\alpha(i)$  represents the “pure state nearest” to  $i$ , that verifies  $\alpha$ , where “nearest” is here defined in terms of the metrics of the Hilbert space  $\mathcal{H}$ . By definition and in virtue of one of the basic postulates of QT (von Neumann’s *collapse of the wave function*),  $f_\alpha(i)$  turns out to have the following physical meaning: it represents the transformation of state  $i$  after the performance of a measurement concerning the physical property expressed by  $\alpha$ , provided the result was positive. As a consequence, one obtains:  $\alpha \rightarrow \beta$  is true in a state  $i$  iff either  $\alpha$  is impossible for  $i$  or the state into which  $i$  has been transformed after a positive  $\alpha$ -test, verifies  $\alpha$ .

Another interesting characteristic of our connective  $\rightarrow$ , is a *weak non monotonic* behaviour. In fact, in the algebraic semantics the inequality

$$v(\alpha \rightarrow \gamma) \sqsubseteq v(\alpha \wedge \beta \rightarrow \gamma)$$

can be violated (a counterexample can be easily obtained in the orthomodular lattice based on  $\mathbb{R}^3$ ). As a consequence:

$$\alpha \rightarrow \gamma \not\models \alpha \wedge \beta \rightarrow \gamma.$$

Polynomial conditionals are not the only significant examples of implication-connectives in **QL**. In the framework of a Kripkean semantic approach, it seems quite natural to introduce a conditional connective  $\multimap$ , that represents a kind of *strict implication*. Given a Kripkean realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  one would like to require:

$$i \models \alpha \multimap \beta \text{ iff } \forall j \not\models i(j \models \alpha \rightsquigarrow j \models \beta).$$

However such a condition does not automatically represent a correct semantic definition, because it is not granted that  $\rho(\alpha \multimap \beta)$  is an element of  $\Pi$ . In order to overcome this difficulty, let us first define a new operation in the power-set of an orthoframe  $\langle I, R \rangle$ .

**DEFINITION 37** (Strict-implication operation ( $\boxed{\multimap}$ )). Given an orthoframe  $\langle I, R \rangle$  and  $X, Y \subseteq I$ :

$$X \boxed{\multimap} Y := \{i \mid \forall j (i \not\models j \text{ and } j \in X \rightsquigarrow j \in Y)\}.$$

If  $X$  and  $Y$  are sets of worlds in the orthoframe, then  $X \boxed{\dashv} Y$  turns out to be a proposition of the frame.

When the set  $\Pi$  of  $\mathcal{K}$  is closed under  $\boxed{\dashv}$ , we will say that  $\mathcal{K}$  is a realization for a *strict-implication language*.

**DEFINITION 38** (Strict implication ( $\dashv$ )). If  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  is a realization for a strict-implication language, then

$$\rho(\alpha \dashv \beta) := \rho(\alpha) \boxed{\dashv} \rho(\beta).$$

One can easily check that  $\dashv$  is a “good” conditional. There follows immediately:

$$i \models \alpha \dashv \beta \text{ iff } \forall j \not\models i (j \models \alpha \curvearrowright j \models \beta).$$

Another interesting implication that can be defined in **QL** is represented by an entailment-connective.

**DEFINITION 39** (Entailment ( $\Rightarrow$ )). Given  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$ ,

$$\rho(\alpha \Rightarrow \beta) := \begin{cases} I, & \text{if } \rho(\alpha) \subseteq \rho(\beta); \\ \emptyset, & \text{otherwise.} \end{cases}$$

Since  $I, \emptyset \in \Pi$ , the definition is correct. One can trivially check that  $\Rightarrow$  is a “good” conditional. Interestingly enough, our strict implication and our entailment represent “good” implications also for **OL**.

The general relations between  $\rightarrow$ ,  $\dashv$  and  $\Rightarrow$  are described by the following theorem:

**THEOREM 40.** *For any realization  $\mathcal{K}$  for a strict-implication language of **OL**:*

$$\models_{\mathcal{K}} (\alpha \Rightarrow \beta) \Rightarrow (\alpha \dashv \beta).$$

*For any realization  $\mathcal{K}$  for a strict-implication language of **OQL**:*

$$\models_{\mathcal{K}} (\alpha \Rightarrow \beta) \Rightarrow (\alpha \rightarrow \beta); \quad \models_{\mathcal{K}} (\alpha \dashv \beta) \Rightarrow (\alpha \rightarrow \beta).$$

But the inverse relations do not generally hold!

Are the connectives  $\dashv$  and  $\Rightarrow$  definable also in the algebraic semantics? The possibility of defining  $\Rightarrow$  is straightforward.

**DEFINITION 41** (Entailment in the algebraic semantics). Given  $\mathcal{A} = \langle \mathcal{B}, v \rangle$ ,

$$v(\alpha \Rightarrow \beta) := \begin{cases} \mathbf{1}, & \text{if } v(\alpha) \sqsubseteq v(\beta); \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$

One can easily check that  $\rightarrow$  represents the same connective in the two semantics. As to  $\rightarrow$ , given  $\mathcal{A} = \langle \mathcal{B}, v \rangle$ , one would like to require:

$$v(\alpha \rightarrow \beta) = \bigsqcup \{b \in B \mid b \neq \mathbf{0} \text{ and } \forall c(c \neq \mathbf{0} \text{ and } b \not\sqsubseteq c' \text{ and } c \sqsubseteq v(\alpha) \curvearrowright c \sqsubseteq v(\beta))\}.$$

However such a definition supposes the algebraic completeness of  $\mathcal{B}$ . Further we can prove that  $\rightarrow$  represents the same connective in the two semantics only if we restrict our consideration to the class of all algebraically adequate Kripkean realizations.

#### 4 METALOGICAL PROPERTIES AND ANOMALIES

Some metalogical distinctions that are not interesting in the case of a number of familiar logics weaker than classical logic turn out to be significant for **QL** (and for non distributive logics in general).

We have already defined (both in the algebraic and in the Kripkean semantics) the concepts of *model* and of *logical consequence*. Now we will introduce, in both semantics, the notions of *quasi-model*, *weak consequence* and *quasi-consequence*. Let  $T$  be any set of formulas.

DEFINITION 42 (Quasi-model).

*Algebraic semantics*

A realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$   
is a quasi-model of  $T$  iff  
 $\exists a[a \in B \text{ and } a \neq \mathbf{0} \text{ and}$   
 $\forall \beta \in T(a \sqsubseteq v(\beta))]$ .

*Kripkean semantics*

A realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$   
is a *quasi-model* of  $T$  iff  
 $\exists i(i \in I \text{ and } i \models T)$ .

The following definitions can be expressed in both semantics.

DEFINITION 43 (Realizability and verifiability).  $T$  is *realizable* ( $\text{Real } T$ ) iff it has a quasi-model;  $T$  is *verifiable* ( $\text{Verif } T$ ) iff it has a model.

DEFINITION 44 (Weak consequence). A formula  $\alpha$  is a *weak consequence* of  $T$  ( $T \models \alpha$ ) iff any model of  $T$  is a model of  $\alpha$ .

DEFINITION 45 (Quasi-consequence). A formula  $\alpha$  is a *quasi-consequence* of  $T$  ( $T \approx \alpha$ ) iff any quasi-model of  $T$  is a quasi-model of  $\alpha$ .

One can easily check that the algebraic notions of verifiability, realizability, weak consequence and quasi-consequence turn out to coincide with the corresponding Kripkean notions. In other words,  $T$  is Kripke-realizable iff  $T$  is algebraically realizable. Similarly for the other concepts.

In both semantics one can trivially prove the following lemmas.

LEMMA 46.  $\text{Verif } T \curvearrowright \text{Real } T$ .

LEMMA 47.  $\text{Real } T$  iff for any contradiction  $\beta \wedge \neg\beta$ ,  $T \not\models \beta \wedge \neg\beta$ .



LEMMA 48.  $T \models \alpha \curvearrowright T \equiv \alpha$ ;  $T \models \alpha \curvearrowright T \approx \alpha$ .

LEMMA 49.  $\alpha \equiv \beta$  iff  $\neg\beta \approx \neg\alpha$ .

Most familiar logics, that are stronger than positive logic, turn out to satisfy the following metalogical properties, which we will call *Herbrand–Tarski*, *verifiability* and *Lindenbaum*, respectively.

- *Herbrand–Tarski*

$$T \models \alpha \text{ iff } T \equiv \alpha \text{ iff } T \approx \alpha$$

- *Verifiability*

$$\text{Ver } T \text{ iff Real } T$$

- *Lindenbaum*

$$\text{Real } T \curvearrowright \exists T^* [T \subseteq T^* \text{ and Compl } T^*], \text{ where}$$

$$\text{Compl } T \text{ iff } \forall \alpha [\alpha \in T \text{ or } \neg\alpha \in T].$$

The Herbrand–Tarski property represents a semantic version of the deduction theorem. The Lindenbaum property asserts that any semantically non-contradictory set of formulas admits a semantically non-contradictory complete extension. In the algebraic semantics, canonical proofs of these properties essentially use some versions of Stone-theorem, according to which any *proper filter*  $F$  in an algebra  $\mathcal{B}$  can be extended to a *proper complete filter*  $F^*$  (such that  $\forall a(a \in F^* \text{ or } a' \in F^*)$ ). However, Stone-theorem does not generally hold for non distributive orthomodular lattices! In the case of ortholattices, one can still prove that every proper filter can be extended to an *ultrafilter* (i.e., a *maximal filter* that does not admit any extension that is a proper filter). However, differently from Boolean algebras, ultrafilters need not be complete.

A counterexample to the Herbrand–Tarski property in **OL** can be obtained using the “non-valid” part of the distributive law. We know that (owing to the failure of distributivity in ortholattices):

$$\alpha \wedge (\beta \vee \gamma) \not\equiv (\alpha \wedge \beta) \vee (\alpha \wedge \gamma).$$

At the same time

$$\alpha \wedge (\beta \vee \gamma) \equiv (\alpha \wedge \beta) \vee (\alpha \wedge \gamma),$$

since one can easily calculate that for any realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  the hypothesis  $v(\alpha \wedge (\beta \vee \gamma)) = \mathbf{1}$ ,  $v((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \neq \mathbf{1}$  leads to a contradiction <sup>2</sup>.

<sup>2</sup>In **OQL** a counterexample in two variables can be obtained by using the failure of the contraposition law for  $\rightarrow$ . One has:  $\alpha \rightarrow \beta \not\equiv \neg\beta \rightarrow \neg\alpha$ . At the same time  $\alpha \rightarrow \beta \equiv \neg\beta \rightarrow \neg\alpha$ ; since for any realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  the hypothesis  $v(\alpha \rightarrow \beta) = \mathbf{1}$ , implies  $v(\alpha) \sqsubseteq v(\beta)$  and therefore  $v(\neg\beta \rightarrow \neg\alpha) = v(\beta) \sqcup (v(\alpha)' \sqcap v(\beta)') = v(\beta) \sqcup v(\beta)' = \mathbf{1}$ .

A counterexample to the verifiability-property is represented by the negation of the *a fortiori* principle for the quantum logical conditional  $\rightarrow$ :

$$\gamma := \neg(\alpha \rightarrow (\beta \rightarrow \alpha)) = \neg(\neg\alpha \vee (\alpha \wedge (\neg\beta \vee (\alpha \wedge \beta))))).$$

This  $\gamma$  has an algebraic quasi-model. For instance the realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$ , where  $\mathcal{B}$  is the orthomodular lattice determined by all subspaces of the plane (as shown in Figure 2). There holds:  $v(\gamma) = v(\alpha) \neq \mathbf{0}$ . But one can easily check that  $\gamma$  cannot have any model, since the hypothesis that  $v(\gamma) = \mathbf{1}$  leads to a contradiction in any algebraic realization of **QL**.

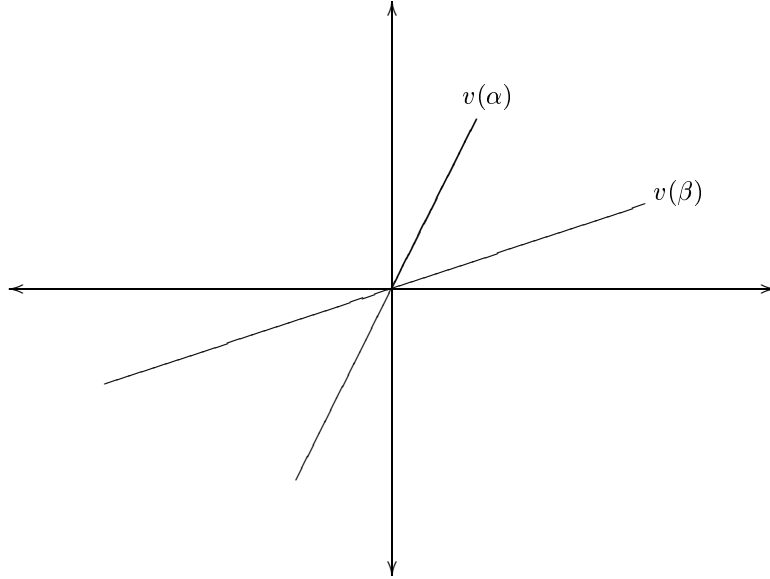


Figure 2. Quasi-model for  $\gamma$

The same  $\gamma$  also represents a counterexample to the Lindenbaum-property. Let us first prove the following lemma.

**LEMMA 50.** *If  $T$  is realizable and  $T \subseteq T^*$ , where  $T^*$  is realizable and complete, then  $T$  is verifiable.*

**Sketch of the proof** Let us define a realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  such that

(i)  $B = \{1, 0\}$ ;

(ii)

$$v(\alpha) = \begin{cases} 1, & \text{if } T^* \models \alpha; \\ 0, & \text{otherwise.} \end{cases}$$

Since  $T^*$  is realizable and complete,  $\mathcal{A}$  is a good realization and is trivially a model of  $T$ . ■

Now, one can easily show that  $\gamma$  violates Lindenbaum. Suppose, by contradiction, that  $\gamma$  has a realizable and complete extension. Then, by Lemma 50,  $\gamma$  must have a model, and we already know that this is impossible.

The failure of the metalogical properties we have considered represents, in a sense, a relevant “anomaly” of quantum logics. Just these anomalies suggest the following conjecture: the distinction between *epistemic logics* (characterized by Kripkean models where the accessibility relation is at least reflexive and transitive) and *similarity logics* (characterized by Kripkean models where the accessibility relation is at least reflexive and symmetric) seems to represent a highly significant dividing line in the class of all logics that are weaker than classical logic.

## 5 A MODAL INTERPRETATION OF **OL** AND **OQL**

**QL** admits a modal interpretation [Goldblatt, 1974; Dalla Chiara, 1981] which is formally very similar to the modal interpretation of intuitionistic logic. Any modal interpretation of a given non-classical logic turns out to be quite interesting from the intuitive point of view, since it permits us to associate a classical meaning to a given system of non-classical logical constants. As is well known, intuitionistic logic can be translated into the modal system **S4**. The modal basis that turns out to be adequate for **OL** is instead the logic **B**. Such a result is of course not surprising, since both the **B**-realizations and the **OL**-realizations are characterized by frames where the accessibility relation is reflexive and symmetric.

Suppose a modal language  $L^M$  whose alphabet contains the same sentential literals as **QL** and the following primitive logical constants: the classical connectives  $\sim$  (*not*),  $\wedge$  (*and*) and the modal operator  $\Box$  (*necessarily*). At the same time, the connectives  $\vee$  (*or*),  $\supset$  (*if ... then*),  $\equiv$  (*if and only if*), and the modal operator  $\Diamond$  (*possibly*) are supposed defined in the standard way.

The modal logic **B** is semantically characterized by a class of Kripkean realizations that we will call **B**-realizations.

**DEFINITION 51.** A **B**-realization is a system  $\mathcal{M} = \langle I, R, \Pi, \rho \rangle$  where:

- (i)  $\langle I, R \rangle$  is an orthoframe;
- (ii)  $\Pi$  is a subset of the power-set of  $I$  satisfying the following conditions:
  - $I, \emptyset \in \Pi$ ;
  - $\Pi$  is closed under the set-theoretic relative complement  $-$ , the set-theoretic intersection  $\cap$  and the modal operation  $\Box$ , which is defined as follows:

for any  $X \subseteq I$ ,  $\Box X := \{i \mid \forall j (Rij \leadsto j \in X)\}$ ;

- (iii)  $\rho$  associates to any formula  $\alpha$  of  $L^{\mathbf{M}}$  a proposition in  $\Pi$  satisfying the conditions:  $\rho(\sim \beta) = -\rho(\beta)$ ;  $\rho(\beta \wedge \gamma) = \rho(\beta) \cap \rho(\gamma)$ ;  $\rho(\Box \beta) = \Box \rho(\beta)$ .

Instead of  $i \in \rho(\alpha)$ , we will write  $i \models \alpha$ . The definitions of truth, logical truth and logical consequence for  $\mathbf{B}$  are analogous to the corresponding definitions in the Kripkean semantics for  $\mathbf{QL}$ .

Let us now define a translation  $\tau$  of the language of  $\mathbf{QL}$  into the language  $L^{\mathbf{B}}$ .

DEFINITION 52 (Modal translation of  $\mathbf{OL}$ ).

- $\tau(p) = \Box \Diamond p$ ;
- $\tau(\neg \beta) = \Box \sim \tau(\beta)$ ;
- $\tau(\beta \wedge \gamma) = \tau(\beta) \wedge \tau(\gamma)$ .

In other words,  $\tau$  translates any atomic formula as the necessity of the possibility of the same formula; further, the quantum logical negation is interpreted as the necessity of the classical negation, while the quantum logical conjunction is interpreted as the classical conjunction. We will indicate the set  $\{\tau(\beta) \mid \beta \in T\}$  by  $\tau(T)$ .

THEOREM 53. *For any  $\alpha$  and  $T$  of  $\mathbf{OL}$ :  $T \models_{\mathbf{OL}} \alpha$  iff  $\tau(T) \models_{\mathbf{B}} \tau(\alpha)$*

Theorem 53 is an immediate corollary of the following Lemmas 54 and 55.

LEMMA 54. *Any  $\mathbf{OL}$ -realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  can be transformed into a  $\mathbf{B}$ -realization  $\mathcal{M}^{\mathcal{K}} = \langle I^*, R^*, \Pi^*, \rho^* \rangle$  such that:  $I^* = I$ ;  $R^* = R$ ;  
 $\forall i (i \models_{\mathcal{K}} \alpha$  iff  $i \models_{\mathcal{M}^{\mathcal{K}}} \tau(\alpha)$ ).*

**Sketch of the proof** Take  $\Pi^*$  as the smallest subset of the power-set of  $I$  that contains  $\rho(p)$  for any atomic formula  $p$  and that is closed under  $I, \emptyset, -, \cap, \Box$ . Further, take  $\rho^*(p)$  equal to  $\rho(p)$ . ■

LEMMA 55. *Any  $\mathbf{B}$ -realization  $\mathcal{M} = \langle I, R, \Pi, \rho \rangle$  can be transformed into a  $\mathbf{OL}$ -realization  $\mathcal{K}^{\mathcal{M}} = \langle I^*, R^*, \Pi^*, \rho^* \rangle$  such that:  $I^* = I$ ;  $R^* = R$ ;  
 $\forall i (i \models_{\mathcal{K}^{\mathcal{M}}} \alpha$  iff  $i \models_{\mathcal{M}} \tau(\alpha)$ ).*

**Sketch of the proof** Take  $\Pi^*$  as the smallest subset of the power-set of  $I$  that contains  $\rho(\Box \Diamond p)$  for any atomic formula  $p$  and that is closed under  $I, \emptyset, ', \cap$  (where for any set  $X$  of worlds,  $X' := \{j \mid \text{not } Rij\}$ ). Further take  $\rho^*(p)$  equal to  $\rho(\Box \Diamond p)$ . The set  $\rho^*(p)$  turns out to be a proposition in the orthoframe  $\langle I^*, R^* \rangle$ , owing to the  $\mathbf{B}$ -logical truth:  $\Box \Diamond \alpha \equiv \Box \Diamond \Box \Diamond \alpha$ . ■

The translation of **OL** into **B** is technically very useful, since it permits us to transfer to **OL** some nice metalogical properties such as *decidability* and the *finite-model property*.

Does also **OQL** admit a modal interpretation? The question has a somewhat trivial answer. It is sufficient to apply the technique used for **OL** by referring to a convenient modal system **B<sup>o</sup>** (stronger than **B**) which is founded on a modal version of the orthomodular principle. Semantically **B<sup>o</sup>** can be characterized by a particular class of realizations. In order to determine this class, let us first define the concept of *quantum proposition* in a **B**-realization.

**DEFINITION 56.** Given a **B**-realization  $\mathcal{M} = \langle I, R, \Pi, \rho \rangle$  the set  $\Pi_Q$  of all *quantum propositions* of  $\mathcal{M}$  is the smallest subset of the power-set of  $I$  which contains  $\rho(\Box \Diamond p)$  for any atomic  $p$  and is closed under  $'$  and  $\cap$ .

**LEMMA 57.** *In any **B**-realization  $\mathcal{M} = \langle I, R, \Pi, \rho \rangle$ , there holds  $\Pi_Q \subseteq \Pi$ .*

**Sketch of the proof** The only non-trivial point of the proof is represented by the closure of  $\Pi$  under  $'$ . This holds since one can prove:  $\forall X \in \Pi (X' = \Box - X)$ . ■

**LEMMA 58.** *Given  $\mathcal{M} = \langle I, R, \Pi, \rho \rangle$  and  $\mathcal{K}^{\mathcal{M}} = \langle I, R, \Pi^*, \rho^* \rangle$ , there holds  $\Pi_Q = \Pi^*$ .*

**LEMMA 59.** *Given  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  and  $\mathcal{M}^{\mathcal{K}} = \langle I, R, \Pi^*, \rho^* \rangle$ , there holds  $\Pi \supseteq \Pi_Q^*$ .*

**DEFINITION 60.** A **B<sup>o</sup>**-realization is a **B**-realization  $\langle I, R, \Pi, \rho \rangle$  that satisfies the orthomodular property:

$$\forall X, Y \in \Pi_Q : X \not\subseteq Y \quad \curvearrowright \quad X \cap (X \cap Y)' \neq \emptyset.$$

We will also call the **B<sup>o</sup>**-realizations *orthomodular realizations*.

**THEOREM 61.** *For any  $T$  and  $\alpha$  of **OQL**:  $T \models_{\mathbf{OL}} \alpha$  iff  $\tau(T) \models_{\mathbf{B}^o} \tau(\alpha)$ .*

The Theorem is an immediate corollary of Lemmas 54, 55 and of the following Lemma:

**LEMMA 62.**

(62.1) *If  $\mathcal{K}$  is orthomodular then  $\mathcal{M}^{\mathcal{K}}$  is orthomodular.*

(62.2) *If  $\mathcal{M}$  is orthomodular then  $\mathcal{K}^{\mathcal{M}}$  is orthomodular.*

Unfortunately, our modal interpretation of **OQL** is not particularly interesting from a logical point of view. Differently from the **OL**-case, **B<sup>o</sup>** does not correspond to a familiar modal system with well-behaved metalogical properties. A characteristic logical truth of this logic will be a modal version of orthomodularity:

$$\alpha \wedge \sim \beta \supset \Diamond [\alpha \wedge \Box \sim (\alpha \wedge \beta)],$$

where  $\alpha, \beta$  are modal translations of formulas of **OQL** into the language  $L^M$ .

## 6 AN AXIOMATIZATION OF OL AND OQL

**QL** is an axiomatizable logic. Many axiomatizations are known: both in the Hilbert-Bernays style and in the Gentzen-style (*natural deduction* and *sequent-calculi*).<sup>3</sup> We will present here a **QL**-calculus (in the natural deduction style) which is a slight modification of a calculus proposed by [Goldblatt, 1974]. The advantage of this axiomatization is represented by the fact that it is formally very close to the algebraic definition of ortholattice; further it is independent of any idea of quantum logical implication.

Our calculus (which has no axioms) is determined as a set of *rules*. Let  $T_1, \dots, T_n$  be finite or infinite (possibly empty) sets of formulas. Any rule has the form

$$\frac{T_1 \vdash \alpha_1, \dots, T_n \vdash \alpha_n}{T \vdash \alpha}$$

(if  $\alpha_1$  has been inferred from  $T_1, \dots, \alpha_n$  has been inferred from  $T_n$ , then  $\alpha$  can be inferred from  $T$ ). We will call any  $T \vdash \alpha$  a *configuration*. The configurations  $T_1 \vdash \alpha_1, \dots, T_n \vdash \alpha_n$  represent the *premisses* of the rule, while  $T \vdash \alpha$  is the *conclusion*. As a limit case, we may have a rule, where the set of premisses is empty; in such a case we will speak of an *improper rule*. Instead of  $\frac{\emptyset}{T \vdash \alpha}$  we will write  $T \vdash \alpha$ ; instead of  $\emptyset \vdash \alpha$ , we will write  $\vdash \alpha$ .

### Rules of OL

(OL1)	$T \cup \{\alpha\} \vdash \alpha$	(identity)
(OL2)	$\frac{T \vdash \alpha, T^* \cup \{\alpha\} \vdash \beta}{T \cup T^* \vdash \beta}$	(transitivity)
(OL3)	$T \cup \{\alpha \wedge \beta\} \vdash \alpha$	( $\wedge$ -elimination)
(OL4)	$T \cup \{\alpha \wedge \beta\} \vdash \beta$	( $\wedge$ -elimination)
(OL5)	$\frac{T \vdash \alpha, T \vdash \beta}{T \vdash \alpha \wedge \beta}$	( $\wedge$ -introduction)
(OL6)	$\frac{T \cup \{\alpha, \beta\} \vdash \gamma}{T \cup \{\alpha \wedge \beta\} \vdash \gamma}$	( $\wedge$ -introduction)

---

<sup>3</sup>Sequent calculi for different forms of quantum logic will be described in Section 17.

- (OL7)  $\frac{\{\alpha\} \vdash \beta, \{\alpha\} \vdash \neg\beta}{\neg\alpha}$  (absurdity)
- (OL8)  $T \cup \{\alpha\} \vdash \neg\neg\alpha$  (weak double negation)
- (OL9)  $T \cup \{\neg\neg\alpha\} \vdash \alpha$  (strong double negation)
- (OL10)  $T \cup \{\alpha \wedge \neg\alpha\} \vdash \beta$  (Duns Scotus)
- (OL11)  $\frac{\{\alpha\} \vdash \beta}{\{\neg\beta\} \vdash \neg\alpha}$  (contraposition)

DEFINITION 63 (Derivation). A *derivation* of **OL** is a finite sequence of configurations  $T \vdash \alpha$ , where any element of the sequence is either the conclusion of an improper rule or the conclusion of a proper rule whose premisses are previous elements of the sequence.

DEFINITION 64 (Derivability). A formula  $\alpha$  is *derivable* from  $T$  ( $T \vdash_{\text{OL}} \alpha$ ) iff there is a derivation such that the configuration  $T \vdash \alpha$  is the last element of the derivation.

Instead of  $\{\alpha\} \vdash_{\text{OL}} \beta$  we will write  $\alpha \vdash_{\text{OL}} \beta$ . When no confusion is possible, we will write  $T \vdash \alpha$  instead of  $T \vdash_{\text{OL}} \alpha$ .

DEFINITION 65 (Logical theorem). A formula  $\alpha$  is a *logical theorem* of **OL** ( $\vdash_{\text{OL}} \alpha$ ) iff  $\emptyset \vdash_{\text{OL}} \alpha$ .

One can easily prove the following syntactical lemmas.

LEMMA 66.  $\alpha_1, \dots, \alpha_n \vdash \alpha$  iff  $\alpha_1 \wedge \dots \wedge \alpha_n \vdash \alpha$ .

LEMMA 67. *Syntactical compactness.*

$T \vdash \alpha$  iff  $\exists T^* \subseteq T$  ( $T^*$  is finite and  $T^* \vdash \alpha$ ).

LEMMA 68.  $T \vdash \alpha$  iff  $\exists \alpha_1, \dots, \alpha_n$  ( $\alpha_1 \in T$  and ... and  $\alpha_n \in T$  and  $\alpha_1 \wedge \dots \wedge \alpha_n \vdash \alpha$ ).

DEFINITION 69 (Consistency).  $T$  is an *inconsistent* set of formulas if  $\exists \alpha (T \vdash \alpha \wedge \neg\alpha)$ ;  $T$  is *consistent*, otherwise.

DEFINITION 70 (Deductive closure). The *deductive closure*  $\overline{T}$  of a set of formulas  $T$  is the smallest set which includes the set  $\{\alpha \mid T \vdash \alpha\}$ .  $T$  is called *deductively closed* iff  $T = \overline{T}$ .

DEFINITION 71 (Syntactical compatibility). Two sets of formulas  $T_1$  and  $T_2$  are called *syntactically compatible* iff

$$\forall \alpha (T_1 \vdash \alpha \curvearrowright T_2 \not\vdash \neg\alpha).$$

The following theorem represents a kind of “weak Lindenbaum theorem”.

**THEOREM 72.** *Weak Lindenbaum theorem.*

*If  $T \not\vdash \neg\alpha$ , then there exists a set of formulas  $T^*$  such that  $T^*$  is compatible with  $T$  and  $T^* \vdash \alpha$ .*

**Proof.** Suppose  $T \not\vdash \neg\alpha$ . Take  $T^* = \{\alpha\}$ . There holds trivially:  $T^* \vdash \alpha$ . Let us prove the compatibility between  $T$  and  $T^*$ . Suppose, by contradiction,  $T$  and  $T^*$  incompatible. Then, for a certain  $\beta$ ,  $T^* \vdash \beta$  and  $T \vdash \neg\beta$ . Hence (by definition of  $T^*$ ),  $\alpha \vdash \beta$  and by contraposition,  $\neg\beta \vdash \neg\alpha$ . Consequently, because  $T \vdash \neg\beta$ , one obtains by transitivity:  $T \vdash \neg\alpha$ , against our hypothesis. ■

We will now prove a soundness and a completeness theorem with respect to the Kripkean semantics.

**THEOREM 73.** *Soundness theorem.*

$$T \vdash \alpha \rightsquigarrow T \models \alpha.$$

**Proof.** Straightforward. ■

**THEOREM 74.** *Completeness theorem.*

$$T \models \alpha \rightsquigarrow T \vdash \alpha.$$

**Proof.** It is sufficient to construct a *canonical model*  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  such that:

$$T \vdash \alpha \text{ iff } T \models_{\mathcal{K}} \alpha.$$

As a consequence we will immediately obtain:

$$T \not\vdash \alpha \rightsquigarrow T \not\models_{\mathcal{K}} \alpha \rightsquigarrow T \not\models \alpha.$$

*Definition of the canonical model*

- (i)  $I$  is the set of all consistent and deductively closed sets of formulas;
- (ii)  $R$  is the compatibility relation between sets of formulas;
- (iii)  $\Pi$  is the set of all propositions in the frame  $\langle I, R \rangle$ ;
- (iv)  $\rho(p) = \{i \in I \mid p \in i\}$ .

In order to recognize that  $\mathcal{K}$  is a “good” **OL**-realization, it is sufficient to prove that: (a)  $R$  is reflexive and symmetric; (b)  $\rho(p)$  is a proposition in the frame  $\langle I, R \rangle$ .

The proof of (a) is immediate (reflexivity depends on the consistency of any  $i$ , and symmetry can be shown using the weak double negation rule).



In order to prove (b), it is sufficient to show (by Lemma 8.1):  $i \notin \rho(p) \curvearrowright \exists j \not\vdash i (j \perp \rho(p))$ . Let  $i \notin \rho(p)$ . Then (by definition of  $\rho(p)$ ):  $p \notin i$ ; and, since  $i$  is deductively closed,  $i \not\vdash p$ . Consequently, by the weak Lindenbaum theorem (and by the strong double negation rule), for a certain  $j$ :  $j \not\vdash i$  and  $\neg p \in j$ . Hence,  $j \perp \rho(p)$ .

LEMMA 75. *Lemma of the canonical model.*

*For any  $\alpha$  and any  $i \in I$ ,  $i \models \alpha$  iff  $\alpha \in i$ .*

**Sketch of the proof** By induction on the length of  $\alpha$ . The case  $\alpha = p$  holds by definition of  $\rho(p)$ . The case  $\alpha = \neg\beta$  can be proved by using Lemma 12.1 and the weak Lindenbaum theorem. The case  $\alpha = \beta \wedge \gamma$  can be proved using the  $\wedge$ -introduction and the  $\wedge$ -elimination rules. ■

Finally we can show that  $T \vdash \alpha$  iff  $T \models_{\mathcal{K}} \alpha$ . Since the left to right implication is a consequence of the soundness-theorem, it is sufficient to prove:  $T \not\vdash \alpha \curvearrowright T \not\models_{\mathcal{K}} \alpha$ . Let  $T \not\vdash \alpha$ ; then, by Duns Scotus,  $T$  is consistent. Take  $i := \overline{T}$ . There holds:  $i \in I$  and  $T \subseteq i$ . As a consequence, by the Lemma of the canonical model,  $i \models T$ . At the same time  $i \not\models \alpha$ . For, should  $i \models \alpha$  be the case, we would obtain  $\alpha \in i$  and by definition of  $i$ ,  $T \vdash \alpha$ , against our hypothesis. ■

An axiomatization of **OQL** can be obtained by adding to the **OL**-calculus the following rule:

$$(OQL) \quad \alpha \wedge \neg(\alpha \wedge \neg(\alpha \wedge \beta)) \vdash \beta. \quad (\text{orthomodularity})$$

All the syntactical definitions we have considered for **OL** can be extended to **OQL**. Also Lemmas 66, 67, 68 and the weak Lindenbaum theorem can be proved exactly in the same way. Since **OQL** admits a material conditional, we will be able to prove here a *deduction theorem*:

THEOREM 76.  $\alpha \vdash_{\text{OQL}} \beta$  iff  $\vdash_{\text{OQL}} \alpha \rightarrow \beta$ .

This version of the deduction-theorem is obviously not in contrast with the failure in **QL** of the semantical property we have called Herbrand-Tarski. For, differently from other logics, here the syntactical relation  $\vdash$  does not correspond to the weak consequence relation!

The soundness theorem can be easily proved, since in any orthomodular realization  $\mathcal{K}$  there holds:

$$\alpha \wedge \neg(\alpha \wedge \neg(\alpha \wedge \beta)) \models_{\mathcal{K}} \beta.$$

As to the completeness theorem, we need a slight modification of the proof we have given for ‘**OL**. In fact, should we try and construct the canonical model  $\mathcal{K}$ , by taking  $\Pi$  as the set of all possible propositions of the

frame, we would not be able to prove the orthomodularity of  $\mathcal{K}$ . In order to obtain an orthomodular canonical model  $\mathcal{K} = \{I, R, \Pi, \rho\}$ , it is sufficient to define  $\Pi$  as the set of all propositions  $X$  of  $\mathcal{K}$  such that  $X = \rho(\alpha)$  for a certain  $\alpha$ . One immediately recognizes that  $\rho(p) \in \Pi$  and that  $\Pi$  is closed under  $'$  and  $\cap$ . Hence  $\mathcal{K}$  is a “good” **OL**-realization. Also for this  $\mathcal{K}$  one can easily show that  $i \models \alpha$  iff  $\alpha \in i$ . In order to prove the orthomodularity of  $\mathcal{K}$ , one has to prove for any propositions  $X, Y \in \Pi$ ,  $X \not\subseteq Y \leadsto X \cap (X \cap Y)' \neq \emptyset$ ; which is equivalent (by Lemma 19) to  $X \cap (X \cap (X \cap Y)')' \subseteq Y$ . By construction of  $\Pi$ ,  $X = \rho(\alpha)$  and  $Y = \rho(\beta)$  for certain  $\alpha, \beta$ . By the orthomodular rule there holds  $\alpha \wedge \neg(\alpha \wedge \neg(\alpha \wedge \beta)) \vdash \beta$ . Consequently, for any  $i \in I$ ,  $i \models \alpha \wedge \neg(\alpha \wedge \neg(\alpha \wedge \beta)) \leadsto i \models \beta$ . Hence,  $\rho(\alpha) \cap (\rho(\alpha) \cap (\rho(\alpha) \cap \rho(\beta))')' \subseteq \rho(\beta)$ .

Of course, also the canonical model of **OL** could be constructed by taking  $\Pi$  as the set of all propositions that are “meanings” of formulas. Nevertheless, in this case, we would lose the following important information: the canonical model of **OL** gives rise to an algebraically complete realization (closed under infinitary intersection).

## 7 THE INTRACTABILITY OF ORTHOMODULARITY

As we have seen, the proposition-ortholattice in a Kripkean realization  $\mathcal{K} = \langle I, R, \Pi, \rho \rangle$  does not generally coincide with the (algebraically) *complete* ortholattice of *all* propositions of the orthoframe  $\langle I, R \rangle$ .<sup>4</sup> When  $\Pi$  is the set of all propositions,  $\mathcal{K}$  will be called *standard*. Thus, a *standard orthomodular Kripkean realization* is a standard realization, where  $\Pi$  is orthomodular. In the case of **OL**, every non standard Kripkean realization can be naturally extended to a standard one (see the proof of Theorem 17). In particular,  $\Pi$  can be always embedded into the complete ortholattice of all propositions of the orthoframe at issue. Moreover, as we have learnt from the completeness proof, the canonical model of **OL** is standard. In the case of **OQL**, instead, there are various reasons that make significant the distinction between standard and non standard realizations:

- (i) Orthomodularity is not elementary [Goldblatt, 1984]. In other words, there is no way to express the orthomodular property of the ortholattice  $\Pi$  in an orthoframe  $\langle I, R \rangle$  as an elementary (first-order) property.
- (ii) It is not known whether every orthomodular lattice is embeddable into a complete orthomodular lattice.
- (iii) It is an open question whether **OQL** is characterized by the class of all standard orthomodular Kripkean realizations.

<sup>4</sup>For the sake of simplicity, we indicate briefly by  $\Pi$  the ortholattice  $\langle \Pi, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$ . Similarly, in the case of other structures dealt with in this section.

- (iv) It is not known whether the *canonical model* of **OQL** is standard. Try and construct a canonical realization for **OQL** by taking  $\Pi$  as the set of all possible propositions (similarly to the **OL**-case). Let us call such a realization a *pseudo canonical realization*. Do we obtain in this way an **OQL**-realization, satisfying the orthomodular property? In other words, is the pseudo canonical realization a model of **OQL**?

In order to prove that **OQL** is characterized by the class of all standard Kripkean realizations it would be sufficient to show that the canonical model belongs to such a class. Should orthomodularity be elementary, then, by a general result proved by Fine, this problem would amount to showing the following statement: there is an elementary condition (or a set thereof) implying the orthomodularity of the standard pseudo canonical realization. Result (i), however, makes this way definitively unpracticable.

Notice that a positive solution to problem (iv) would automatically provide a proof of the full equivalence between the algebraic and the Kripkean consequence relation ( $T \stackrel{A}{\models}_{\mathbf{OQL}} \alpha$  iff  $T \stackrel{K}{\models}_{\mathbf{OQL}} \alpha$ ). If **OQL** is characterized by a standard canonical model, then we can apply the same argument used in the case of **OL**, the ortholattice  $\Pi$  of the canonical model being orthomodular. By similar reasons, also a positive solution to problem (ii) would provide a direct proof of the same result. For, the orthomodular lattice  $\Pi$  of the (not necessarily standard) canonical model of **OQL** would be embeddable into a complete orthomodular lattice.

We will now present Goldblatt's result proving that orthomodularity is not elementary. Further, we will show how orthomodularity leaves defeated one of the most powerful embedding technique: the MacNeille completion method.

**Orthomodularity is not elementary**

Let us consider a first-order language  $L^2$  with a single predicate denoting a binary relation  $R$ . Any frame  $\langle I, R \rangle$  (where  $I$  is a non-empty set and  $R$  any binary relation) will represent a classical realization of  $L^2$ .

DEFINITION 77 (Elementary class).

- (i) Let  $\Gamma$  be a class of frames. A possible property  $P$  of the elements of  $\Gamma$  is called *first-order* (or *elementary*) iff there exists a sentence  $\eta$  of  $L^2$  such that for any  $\langle I, R \rangle \in \Gamma$ :

$$\langle I, R \rangle \models \eta \text{ iff } \langle I, R \rangle \text{ has the property } P.$$

- (ii)  $\Gamma$  is said to be an *elementary class* iff the property of being in  $\Gamma$  is an elementary property of  $\Gamma$ .

Thus,  $\Gamma$  is an elementary class iff there is a sentence  $\eta$  of  $L^2$  such that

$$\Gamma = \{\langle I, R \rangle \mid \langle I, R \rangle \models \eta\}.$$

DEFINITION 78 (Elementary substructure). Let  $\langle I_1, R_1 \rangle, \langle I_2, R_2 \rangle$  be two frames.

(a)  $\langle I_1, R_1 \rangle$  is a *substructure* of  $\langle I_2, R_2 \rangle$  iff the following conditions are satisfied:

- (i)  $I_1 \subseteq I_2$ ;
- (ii)  $R_1 = R_2 \cap (I_1 \times I_1)$ ;

(b)  $\langle I_1, R_1 \rangle$  is an *elementary substructure* of  $\langle I_2, R_2 \rangle$  iff the following conditions hold:

- (i)  $\langle I_1, R_1 \rangle$  is a substructure of  $\langle I_2, R_2 \rangle$ ;
- (ii) For any formula  $\alpha(x_1, \dots, x_n)$  of  $L^2$  and any  $i_1, \dots, i_n$  of  $I_1$ :

$$\langle I_1, R_1 \rangle \models \alpha[i_1, \dots, i_n] \text{ iff } \langle I_2, R_2 \rangle \models \alpha[i_1, \dots, i_n].$$

In other words, the elements of the “smaller” structure satisfy exactly the same  $L^2$ -formulas in both structures. The following Theorem [Bell and Slomson, 1969] provides a useful criterion to check whether a substructure is an elementary substructure.

THEOREM 79. *Let  $\langle I_1, R_1 \rangle$  be a substructure of  $\langle I_2, R_2 \rangle$ . Then,  $\langle I_1, R_1 \rangle$  is an elementary substructure of  $\langle I_2, R_2 \rangle$  iff whenever  $\alpha(x_1, \dots, x_n, y)$  is a formula of  $L^2$  (in the free variables  $x_1, \dots, x_n, y$ ) and  $i_1, \dots, i_n$  are elements of  $I_1$  such that for some  $j \in I_2$ ,  $\langle I_2, R_2 \rangle \models \alpha[i_1, \dots, i_n, j]$ , then there is some  $i \in I_1$  such that  $\langle I_2, R_2 \rangle \models \alpha[i_1, \dots, i_n, i]$ .*

Let us now consider a *pre-Hilbert space*<sup>5</sup>  $\mathcal{H}$  and let  $\mathcal{H}^+ := \{\psi \in \mathcal{H} \mid \psi \neq \underline{0}\}$ , where  $\underline{0}$  is the null vector. The pair

$$\langle \mathcal{H}^+, \perp \rangle$$

is an orthoframe, where  $\forall \psi, \phi \in \mathcal{H}^+$ :  $\psi \not\perp \phi$  iff the inner product of  $\psi$  and  $\phi$  is different from the null vector  $\underline{0}$  (i.e.,  $(\psi, \phi) \neq \underline{0}$ ). Let  $\Pi(\mathcal{H})$  be the ortholattice of all propositions of  $\langle \mathcal{H}^+, \perp \rangle$ , which turns out to be isomorphic to the ortholattice  $\mathcal{C}(\mathcal{H})$  of all (not necessarily closed) subspaces of  $\mathcal{H}$  (a proposition is simply a subspace devoided of the null vector). The following deep Theorem, due to Amemiya and Halperin [Varadarajan, 1985] permits

<sup>5</sup>A *pre-Hilbert space* is a vector space over a division ring whose elements are the real or the complex or the quaternionic numbers such that an inner product (which transforms any pair of vectors into an element of the ring) is defined. Differently from Hilbert spaces, pre-Hilbert spaces need not be metrically complete.

us to characterize the class of all Hilbert spaces in the larger class of all pre-Hilbert spaces, by means of the orthomodular property.

**THEOREM 80** (Amemiya–Halperin Theorem).  *$\mathcal{C}(\mathcal{H})$  is orthomodular iff  $\mathcal{H}$  is a Hilbert space.*

In other words,  $\mathcal{C}(\mathcal{H})$  is orthomodular iff  $\mathcal{H}$  is metrically complete.

As is well known [Bell and Slomson, 1969], the property of “being metrically complete” is not elementary. On this basis, it will be highly expected that also the orthomodular property is not elementary. The key-lemma in Goldblatt’s proof is the following:

**LEMMA 81.** *Let  $Y$  be an infinite-dimensional (not necessarily closed) subspace of a separable Hilbert space  $\mathcal{H}$ . If  $\alpha$  is any formula of  $L^2$  and  $\psi_1, \dots, \psi_n$  are vectors of  $Y$  such that for some  $\phi \in \mathcal{H}$ ,  $\langle \mathcal{H}^+, \perp \rangle \models \alpha[\psi_1, \dots, \psi_n, \phi]$ , then there exists a vector  $\psi \in Y$  such that  $\langle \mathcal{H}^+, \perp \rangle \models \alpha[\psi_1, \dots, \psi_n, \psi]$ .*

As a consequence one obtains:

**THEOREM 82.** *The orthomodular property is not elementary.*

**Proof.** Let  $\mathcal{H}$  be any *metrically incomplete* pre-Hilbert space. Let  $\overline{\mathcal{H}}$  be its metric completion. Thus  $\mathcal{H}$  is an infinite-dimensional subspace of the Hilbert space  $\overline{\mathcal{H}}$ . By Lemma 81 and by Theorem 79,  $\langle \mathcal{H}^+, \perp \rangle$  is an elementary substructure of  $\langle \overline{\mathcal{H}}^+, \perp \rangle$ . At the same time, by Amemiya–Halperin’s Theorem,  $\mathcal{C}(\mathcal{H})$  cannot be orthomodular, because  $\mathcal{H}$  is metrically incomplete. However,  $\mathcal{C}(\overline{\mathcal{H}})$  is orthomodular. As a consequence, orthomodularity cannot be expressed as an elementary property. ■

### The embeddability problem

As we have seen in Section 2, the class of all propositions of an orthoframe is a complete ortholattice. Conversely, the representation theorem for ortholattices states that every ortholattice  $\mathcal{B} = \langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  is embeddable into the complete ortholattice of all propositions of the orthoframe  $\langle B^+, \perp \rangle$ , where:  $B^+ := B - \{\mathbf{0}\}$  and  $\forall a, b \in B$ :  $a \perp b$  iff  $a \not\sqsubseteq b'$ . The embedding is given by the map

$$h : a \mapsto \langle a \rangle,$$

where  $\langle a \rangle$  is the quasi-ideal generated by  $a$ . In other words:  $\langle a \rangle = \{b \neq \mathbf{0} \mid b \sqsubseteq a\}$ .

One can prove the following Theorem:

**THEOREM 83.** *Let  $\mathcal{B} = \langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  be an ortholattice.  $\forall X \sqsubseteq B$ ,  $X$  is a proposition of  $\langle B^+, \perp \rangle$  iff  $X = l(u(X))$ , where:*

$$u(Y) := \{b \in B^+ \mid \forall a \in Y : a \sqsubseteq b\} \text{ and } l(Y) := \{b \in B^+ \mid \forall a \in Y : b \sqsubseteq a\}.$$

Accordingly, the complete ortholattice of all propositions of the orthoframe  $\langle B^+, \perp \rangle$  is isomorphic to the *MacNeille completion* (or *completion by cuts*) of  $\mathcal{B}$  [Kalmbach, 1983].<sup>6</sup> At the same time, orthomodularity (similarly to distributivity and modularity) is not preserved by the MacNeille completion, as the following example shows [Kalmbach, 1983].

Let  $\mathcal{C}_{(2)}^0(\mathbb{R})$  be the class of all continuous complex-valued functions  $f$  on  $\mathbb{R}$  such that

$$\int_{-\infty}^{+\infty} |f(x)|^2 dx < \infty$$

Let us define the following bilinear form  $(\cdot, \cdot) : \mathcal{C}_{(2)}^0(\mathbb{R}) \times \mathcal{C}_{(2)}^0(\mathbb{R}) \rightarrow \mathbb{C}$  (representing an inner product):

$$(f, g) = \int_{-\infty}^{+\infty} f^*(x)g(x)dx,$$

where  $f^*(x)$  is the complex conjugate of  $f(x)$ . It turns out that  $\mathcal{C}_{(2)}^0(\mathbb{R})$ , equipped with the inner product  $(\cdot, \cdot)$ , gives rise to a metrically incomplete infinite-dimensional pre-Hilbert space. Thus, by Amemiya–Halperin’s Theorem (Theorem 80), the algebraically complete ortholattice  $\mathcal{C}(\mathcal{C}_{(2)}^0(\mathbb{R}))$  of all subspaces of  $\mathcal{C}_{(2)}^0(\mathbb{R})$  cannot be orthomodular. Now consider the sublattice  $\mathcal{FI}$  of  $\mathcal{C}(\mathcal{C}_{(2)}^0(\mathbb{R}))$ , consisting of all finite or cofinite dimensional subspaces. It is not hard to see that  $\mathcal{FI}$  is orthomodular. One can prove that  $\mathcal{C}(\mathcal{C}_{(2)}^0(\mathbb{R}))$  is *sup-dense* in  $\mathcal{FI}$ ; in other words, any  $X \in \mathcal{C}(\mathcal{C}_{(2)}^0(\mathbb{R}))$  is the *sup* of a set of elements of  $\mathcal{FI}$ . Thus, by a theorem proved by McLaren [Kalmbach, 1983], the MacNeille completion of  $\mathcal{C}(\mathcal{C}_{(2)}^0(\mathbb{R}))$  is isomorphic to the MacNeille completion of  $\mathcal{FI}$ . Since  $\mathcal{C}(\mathcal{C}_{(2)}^0(\mathbb{R}))$  is algebraically complete, the MacNeille completion of  $\mathcal{C}(\mathcal{C}_{(2)}^0(\mathbb{R}))$  is isomorphic to  $\mathcal{C}(\mathcal{C}_{(2)}^0(\mathbb{R}))$  itself. As a consequence,  $\mathcal{FI}$  is orthomodular, while its MacNeille completion is not.

## 8 HILBERT QUANTUM LOGIC AND THE ORTHOMODULAR LAW

As we have seen, the prototypical models of **OQL** that are interesting from the physical point of view are based on the class  $\mathbb{H}$  of all Hilbert lattices, whose support is the set  $\mathcal{C}(\mathcal{H})$  of all closed subspaces of a Hilbert space  $\mathcal{H}$ . Let us call *Hilbert quantum logic* (**HQL**) the logic that is semantically characterized by  $\mathbb{H}$ . A question naturally arises: do **OQL** and **HQL** represent one and the same logic? As proved by [Greechie, 1981],<sup>7</sup> this question

<sup>6</sup>The *MacNeille completion* of an ortholattice  $\mathcal{B} = \langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  is the lattice whose support consists of all  $X \subseteq B$  such that  $X = l(u(X))$ , where:  $u(Y) := \{b \in B \mid \forall a \in Y : a \sqsubseteq b\}$  and  $l(Y) := \{b \in B \mid \forall a \in Y : b \sqsubseteq a\}$ . Clearly the only difference between the proposition-lattice of the frame  $\langle B^+, \perp \rangle$  and the Mac Neille completion of  $\mathcal{B}$  is due to the fact that propositions do not contain  $\mathbf{0}$ .

<sup>7</sup>See also [Kalmbach, 1983].

has a negative answer: there is a lattice-theoretical equation (the so-called *orthoarguesian law*) that holds in  $\mathbb{H}$ , but fails in a particular orthomodular lattice. As a consequence, **OQL** does not represent a faithful logical abstraction from its quantum theoretical origin.

**DEFINITION 84.** Let  $\Gamma$  be a class of orthomodular lattices. We say that **OQL** is *characterized* by  $\Gamma$  iff for any  $T$  and any  $\alpha$  the following condition is satisfied:

$$T \models_{\mathbf{OQL}} \alpha \text{ iff for any } \mathcal{B} \in \Gamma \text{ and any } \mathcal{A} = \langle \mathcal{B}, v \rangle : T \models_{\mathcal{A}} \alpha.$$

In order to formulate the orthoarguesian law in an equational way, let us first introduce the notion of *Sasaki projection*.

**DEFINITION 85** (The Sasaki projection).

Let  $\mathcal{B}$  be an orthomodular lattice and let  $a, b$  be any two elements of  $\mathcal{B}$ . The *Sasaki projection* of  $a$  onto  $b$ , denoted by  $a \pitchfork b$ , is defined as follows:

$$a \pitchfork b := (a \sqcup b') \sqcap b.$$

It is easy to see that two elements  $a, b$  of an orthomodular lattice are compatible ( $a = (a \sqcap b') \sqcup (a \sqcap b)$ ) iff  $a \pitchfork b = a \sqcap b$ . Consequently, in any Boolean lattice,  $\pitchfork$  coincides with  $\sqcap$ .

**DEFINITION 86** (The orthoarguesian law).

$$a \sqsubseteq b \sqcup \{(a \pitchfork b') \sqcap [(a \pitchfork c') \sqcup ((b \sqcup c) \sqcap ((a \pitchfork b') \sqcup (a \pitchfork c')))]\} \quad (\text{OAL})$$

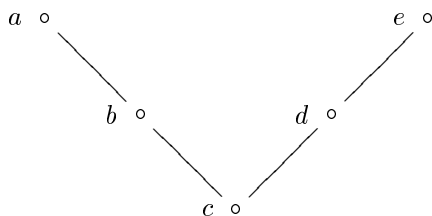
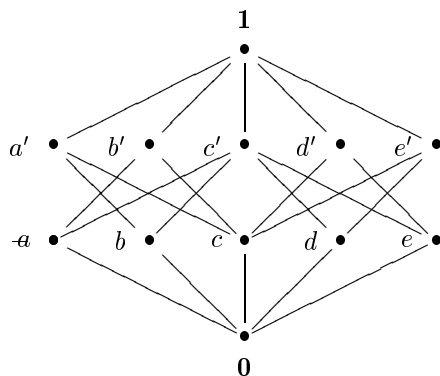
Greechie has proved that (OAL) holds in  $\mathbb{H}$  but fails in a particular finite orthomodular lattice. In order to understand Greechie's counterexample, it will be expedient to illustrate the notion of *Greechie diagram*.

Let us first recall the definition of *atom*.

**DEFINITION 87** (Atom). Let  $\mathcal{B} = \langle B, \sqsubseteq, \mathbf{1}, \mathbf{0} \rangle$  any bounded lattice. An *atom* is an element  $a \in B - \{\mathbf{0}\}$  such that:

$$\forall b \in B : \mathbf{0} \sqsubseteq b \sqsubseteq a \curvearrowright b = \mathbf{0} \text{ or } a = b.$$

Greechie diagrams are *hypergraphs* that permits us to represent particular orthomodular lattices. The representation is essentially based on the fact that a finite Boolean algebra is completely determined by its atoms. A Greechie diagram of an orthomodular lattice  $\mathcal{B}$  consists of points and lines. Points are in one-to-one correspondence with the atoms of  $\mathcal{B}$ ; lines are in one-to-one correspondence with the maximal Boolean subalgebras<sup>8</sup> of  $\mathcal{B}$ . Two lines are crossing in a common atom. For example, the Greechie diagram pictured in Figure 3. represents the orthomodular lattice  $\mathcal{G}_{12}$  (Figure 4).

Figure 3. The Greechie diagram of  $\mathcal{G}_{12}$ Figure 4. The orthomodular lattice  $\mathcal{G}_{12}$



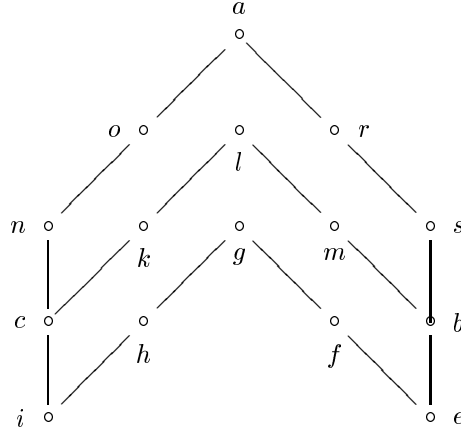


Figure 5. The Greechie diagram of  $\mathcal{B}_{30}$

Let us now consider a particular finite orthomodular lattice, called  $\mathcal{B}_{30}$ , whose Greechie diagram is pictured in Figure 3.

**THEOREM 88.** (OAL) fails in  $\mathcal{B}_{30}$ .

**Proof.** There holds:  $a \frown b' = (a \sqcup b) \sqcap b' = s' \sqcap b' = e$ ,  $a \frown c' = (a \sqcup c) \sqcap c' = n' \sqcap c' = i$  and  $b \sqcup c = l'$ . Thus,

$$\begin{aligned}
 & b \sqcup \{ (a \frown b') \sqcap [ (a \frown c') \sqcup ( (b \sqcup c) \sqcap ( (a \frown b') \sqcup (a \frown c') ) ) ] \} \\
 & \quad = b \sqcup \{ e \sqcap [ i \sqcup ( l' \sqcap ( e \sqcup i ) ) ] \} \\
 & \quad = b \sqcup \{ e \sqcap [ i \sqcup ( l' \sqcap g' ) ] \} \\
 & \quad = b \sqcup ( e \sqcap ( i \sqcup \mathbf{0} ) ) \\
 & \quad = b \sqcup ( e \sqcap i ) \\
 & \quad = b \\
 & \quad \not\sqsubseteq a.
 \end{aligned}$$

■

Hence, there are two formulas  $\alpha$  and  $\beta$  (whose valuations in a convenient realization represent the left- and right- hand side of (OAL), respectively) such that  $\alpha \not\vdash_{\mathbb{H}} \beta$ . At the same time, for any  $\mathcal{C}(\mathcal{H}) \in \mathbb{H}$  and for any realization  $\mathcal{A} = \langle \mathcal{C}(\mathcal{H}), v \rangle$ , there holds:  $\alpha \models_{\mathcal{A}} \beta$ .

As a consequence, **OQL** is not characterized by  $\mathbb{H}$ . Accordingly, **HQL** is definitely stronger than **OQL**. We are faced with the problem of finding

<sup>8</sup>A *maximal Boolean subalgebra* of an ortholattice  $\mathcal{B}$  is a Boolean subalgebra of  $\mathcal{B}$ , that is not a proper subalgebra of any Boolean subalgebra of  $\mathcal{B}$ .

out a calculus, if any, that turns out to be sound and complete with respect to  $\mathbb{H}$ . The main question is whether the class of all formulas valid in  $\mathbb{H}$  is recursively enumerable. In order to solve this problem, it would be sufficient (but not necessary) to show that the canonical model of **HQL** is isomorphic to the subdirect product of a class of Hilbert lattices. So far, very little is known about this question.

### Lattice characterization of Hilbert lattices

As mentioned in the Introduction, the algebraic structure of the set  $\mathcal{E}$  of all events in an event-state system  $\langle \mathcal{E}, S \rangle$  is usually assumed to be a  $\sigma$ -complete orthomodular lattice. Hilbert lattices, however, satisfy further important structural properties. It will be expedient to recall first some standard lattice theoretical definitions. Let  $\mathcal{B} = \langle B, \sqsubseteq, \mathbf{1}, \mathbf{0} \rangle$  be any bounded lattice.

**DEFINITION 89 (Atomicity).** A bounded lattice  $\mathcal{B}$  is *atomic* iff  $\forall a \in B - \{\mathbf{0}\}$  there exists an atom  $b$  such that  $b \sqsubseteq a$ .

**DEFINITION 90 (Covering property).** Let  $a, b$  be two elements of a lattice  $\mathcal{B}$ . We say that  $b$  *covers*  $a$  iff  $a \sqsubseteq b, a \neq b$ , and  $\forall c \in B : a \sqsubseteq c \sqsubseteq b \iff a = c$  or  $b = c$ .

A lattice  $\mathcal{B}$  satisfies the *covering property* iff  $\forall a, b \in B : a$  covers  $a \sqcap b \iff a \sqcup b$  covers  $b$ .

**DEFINITION 91 (Irreducibility).** Let  $\mathcal{B}$  be an orthomodular lattice.  $\mathcal{B}$  is said to be *irreducible* iff

$$\{a \in B \mid \forall b \in B : a \text{ is compatible with } b\} = \{\mathbf{0}, \mathbf{1}\}.$$

One can prove the following theorem:

**THEOREM 92.** *Any Hilbert lattice is a complete, irreducible, atomic orthomodular lattice, which satisfies the covering property.*

Are these conditions sufficient for a lattice  $\mathcal{B}$  to be isomorphic to (or embeddable into) a Hilbert lattice? In other words, is it possible to capture lattice-theoretically the structure of Hilbert lattices? An important result along these lines is represented by the so-called *Piron–McLaren’s coordinatization theorem* [Varadarajan, 1985].

**THEOREM 93 (Piron–McLaren coordinatization theorem).** *Any orthomodular lattice  $\mathcal{B}$  (of length<sup>9</sup> at least 4) that is complete, irreducible, atomic with the covering property, is isomorphic to the orthomodular lattice of all  $(\cdot, \cdot)$ -closed subspaces of a Hilbertian space  $\langle \mathcal{V}, \theta, (\cdot, \cdot), \mathbf{D} \rangle$ .*<sup>10</sup>

<sup>9</sup>The *length* of a lattice  $\mathcal{B}$  is the supremum over the numbers of elements of all the chains of  $\mathcal{B}$ , minus 1.

<sup>10</sup>A *Hilbertian space* is a 4-tuple  $\langle \mathcal{V}, \theta, (\cdot, \cdot), \mathbf{D} \rangle$ , where  $\mathcal{V}$  is a vector space over a division ring  $\mathbf{D}$ ,  $\theta$  is an involutive antiautomorphism on  $\mathbf{D}$ , and  $(\cdot, \cdot)$  (to be interpreted

Do the properties of the *coordinatized lattice*  $\mathcal{B}$  restrict the choice to one of the real, the complex or the quaternionic numbers ( $\mathbb{Q}$ ) and therefore to a classical Hilbert space? Quite unexpectedly, [Keller, 1980] proved a negative result: there are lattices that satisfy all the conditions of Piron–McLaren’s Theorem; at the same time, they are coordinatized by Hilbertian spaces over non-archimedean division rings. Keller’s counterexamples have been interpreted by some authors as showing the definitive impossibility for the quantum logical approach to capture the Hilbert space mathematics. This impossibility was supposed to demonstrate the failure of the quantum logic approach in reaching its main goal: the “bottom-top” reconstruction of Hilbert lattices. Interestingly enough, such a negative conclusion has been recently contradicted by an important result proved by Solèr [1995]: Hilbert lattices can be characterized in a lattice-theoretical way. Solèr’s result is essentially based on the following Theorem:

**THEOREM 94.** *Let  $\langle \mathcal{V}, \theta, (\cdot, \cdot), \mathbf{D} \rangle$  be an infinite-dimensional Hilbertian space over a division ring  $\mathbf{D}$ . Suppose our space includes a  $k$ -orthogonal set  $\{\psi_i\}_{i \in \mathbb{N}}$ , i.e., a family of vectors of  $\mathcal{V}$  such that  $\forall i : (\psi_i, \psi_i) = k$  and  $\forall i, j (i \neq j) : (\psi_i, \psi_j) = 0$ . Then  $\langle \mathcal{V}, \theta, (\cdot, \cdot), \mathbf{D} \rangle$  is a classical Hilbert space.*

As a consequence, the existence of  $k$ -orthogonal sets characterizes Hilbert spaces in the class of all Hilbertian spaces. The point is that the existence of such sets admits of a purely lattice-theoretic characterization, by means of the so-called *angle bisecting condition* [Morash, 1973]. Accordingly, every lattice which satisfies the angle bisecting condition (in addition to the usual conditions of Piron–McLaren’s Theorem) is isomorphic to a classical Hilbert lattice.

## 9 FIRST-ORDER QUANTUM LOGIC

The most significant logical and metalogical peculiarities of **QL** arise at the sentential level. At the same time the extension of sentential **QL** to a first-order logic seems to be quite natural. Similarly to the case of sentential **QL**, we will characterize first-order **QL** both by means of an algebraic and a Kripkean semantics.

Suppose a standard first-order language with predicates  $P_m^n$  and individual constants  $a_m$ .<sup>11</sup> The primitive logical constants are the connectives  $\neg, \wedge$  and the universal quantifier  $\forall$ . The concepts of *term*, *formula* and *sentence*

---

as an inner product) is a definite symmetric  $\theta$ -bilinear form on  $\mathcal{V}$ . Let  $X$  be any subset of  $\mathcal{V}$  and let  $X' := \{\psi \in \mathcal{V} \mid \forall \phi \in X, (\psi, \phi) = 0\}$ ;  $X$  is called *( $\cdot, \cdot$ )-closed* iff  $X = X''$ . The following condition is required to hold: for any  $(\cdot, \cdot)$ -closed set  $X$  of  $\mathcal{V}$ ,  $\mathcal{V} = X + X' := \{\psi + \phi : \psi \in X, \phi \in X'\}$ .

If  $\mathbf{D}$  is either  $\mathbb{R}$  or  $\mathbb{C}$  or  $\mathbb{Q}$  and the antiautomorphism  $\theta$  is continuous, then  $\langle \mathcal{V}, \theta, (\cdot, \cdot), \mathbf{D} \rangle$  turns out to be a classical Hilbert space.

<sup>11</sup>For the sake of simplicity, we do not assume functional symbols.

are defined in the usual way. We will use  $x, y, z, x_1, \dots, x_n, \dots$  as metavariables ranging over the individual variables, and  $t, t_1, t_2, \dots$  as metavariables ranging over terms. The existential quantifier  $\exists$  is supposed defined by a generalized de Morgan law:

$$\exists x\alpha := \neg\forall x\neg\alpha.$$

**DEFINITION 95** (Algebraic realization for first-order **OL**). An *algebraic realization* for (first-order) **OL** is a system  $\mathcal{A} = \langle B^C, D, v \rangle$  where:

- (i)  $B^C = \langle B^C, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  is an ortholattice closed under infinitary *infimum* ( $\prod$ ) and *supremum* ( $\sqcup$ ) for any  $F \subseteq B^C$  such that  $F \in C$  ( $C$  being a particular family of subsets of  $B^C$ ).
- (ii)  $D$  is a non-empty set (disjoint from  $B$ ) called the *domain* of  $\mathcal{A}$ .
- (iii)  $v$  is the *valuation*-function satisfying the following conditions:
  - for any constant  $a_m$ :  $v(a_m) \in D$ ; for any predicate  $P_m^n$ ,  $v(P_m^n)$  is an  $n$ -ary attribute in  $\mathcal{A}$ , i.e., a function that associates to any  $n$ -tuple  $\langle \mathbf{d}_1, \dots, \mathbf{d}_n \rangle$  of elements of  $D$  an element (*truth-value*) of  $B^c$ ;
  - for any *interpretation*  $\sigma$  of the variables in the domain  $D$  (i.e., for any function from the set of all variables into  $D$ ) the pair  $\langle v, \sigma \rangle$  (abbreviated by  $v^\sigma$  and called *generalized valuation*) associates to any term an element in  $D$  and to any formula a truth-value in  $B^c$ , according to the conditions:

$$\begin{aligned} v^\sigma(a_m) &= v(a_m) \\ v^\sigma(x) &= \sigma(x) \\ v^\sigma(P_m^n t_1, \dots, t_n) &= v(P_m^n)(v^\sigma(t_1), \dots, v^\sigma(t_n)) \\ v^\sigma(\neg\beta) &= v^\sigma(\beta)' \\ v^\sigma(\beta \wedge \gamma) &= v^\sigma(\beta) \sqcap v^\sigma(\gamma) \\ v^\sigma(\forall x\beta) &= \prod \{ v^{\sigma[x/\mathbf{d}]}(\beta) \mid \mathbf{d} \in D \}, \text{ where} \\ &\{ v^{\sigma[x/\mathbf{d}]}(\beta) \mid \mathbf{d} \in D \} \in C \end{aligned}$$

( $\sigma[x/\mathbf{d}]$  is the interpretation that associates to  $x$  the individual  $\mathbf{d}$  and differs from  $\sigma$  at most in the value attributed to  $x$ ).

**DEFINITION 96** (Truth and logical truth). A formula  $\alpha$  is *true* in  $\mathcal{A} = \langle B^C, D, v \rangle$  (abbreviated as  $\models_{\mathcal{A}} \alpha$ ) iff for any interpretation of the variables  $\sigma$ ,  $v^\sigma(\alpha) = \mathbf{1}$ ;  $\alpha$  is a *logical truth* of **OL** ( $\models_{\mathbf{OL}} \alpha$ ) iff for any  $\mathcal{A}$ ,  $\models_{\mathcal{A}} \alpha$ .

**DEFINITION 97** (Consequence in a realization and logical consequence). Let  $\mathcal{A} = \langle B^C, D, v \rangle$  be a realization. A formula  $\alpha$  is a *consequence* of  $T$  in  $\mathcal{A}$  (abbreviated  $T \models_{\mathcal{A}} \alpha$ ) iff for any element  $a$  of  $B^c$  and any interpretation

$\sigma$ : if for any  $\beta \in T$ ,  $a \sqsubseteq v^\sigma(\beta)$ , then  $a \sqsubseteq v^\sigma(\alpha)$ ;  $\alpha$  is a logical consequence of  $T$  ( $T \models_{\text{OL}} \alpha$ ) iff for any realization  $\mathcal{A}$ :  $T \models_{\mathcal{A}} \alpha$ .

**DEFINITION 98** (Kripkean realization for (first-order) **OL**). A Kripkean realization for (first-order) **OL** is a system  $\mathcal{K} = \langle I, R, \Pi^C, U, \rho \rangle$  where:

- (i)  $\langle I, R, \Pi^C \rangle$  satisfies the same conditions as in the sentential case; further  $\Pi^C$  is closed under infinitary intersection for any  $F \subseteq \Pi^c$  such that  $F \in C$  (where  $C$  is a particular family of subsets of  $\Pi^C$ );
- (ii)  $U$ , called the *domain* of  $\mathcal{K}$ , is a non-empty set, disjoint from the set of worlds  $I$ . The elements of  $U$  are *individual concepts*  $\mathbf{u}$  such that for any world  $i$ :  $\mathbf{u}(i)$  is an *individual* (called the *reference* of  $\mathbf{u}$  in the world  $i$ ). An individual concept  $\mathbf{u}$  is called *rigid* iff for any pairs of worlds  $i, j$ :  $\mathbf{u}(i) = \mathbf{u}(j)$ . The set  $U_i = \{\mathbf{u}(i) \mid \mathbf{u} \in U\}$  represents the *domain of individuals in the world  $i$* . Whenever  $U_i = U_j$  for all  $i, j$  we will say that the realization  $\mathcal{K}$  has a *constant domain*.
- (iii)  $\rho$  associates a meaning to any individual constant  $a_m$  and to any predicate  $P_m^n$  according to the following conditions:

$\rho(a_m)$  is an individual concept in  $U$ .

$\rho(P_m^n)$  is a *predicate-concept*, i.e. a function that associates to any  $n$ -tuple of individual concepts  $\langle \mathbf{u}_1, \dots, \mathbf{u}_n \rangle$  a proposition in  $\Pi^C$ ;

- (iv) for any interpretation of the variables  $\sigma$  in the domain  $U$ , the pair  $\langle \rho, \sigma \rangle$  (abbreviated as  $\rho^\sigma$  and called *valuation*) associates to any term  $t$  an individual concept in  $U$  and to any formula a proposition in  $\Pi^C$  according to the conditions:

$$\begin{aligned} \rho^\sigma(x) &= \sigma(x) \\ \rho^\sigma(a_m) &= \rho(a_m) \\ \rho^\sigma(P_m^n t_1, \dots, t_n) &= \rho(P_m^n)(\rho^\sigma(t_1), \dots, \rho^\sigma(t_n)) \\ \rho^\sigma(\neg\beta) &= \rho^\sigma(\beta)' \\ \rho^\sigma(\beta \wedge \gamma) &= \rho^\sigma(\beta) \cap \rho^\sigma(\gamma) \\ \rho^\sigma(\forall x\beta) &= \bigcap \{ \rho^{\sigma[x/\mathbf{u}]}(\beta) \mid \mathbf{u} \in U \}, \text{ where} \\ &\{ \rho^{\sigma[x/\mathbf{u}]}(\beta) \mid \mathbf{u} \in U \} \in C. \end{aligned}$$

For any world  $i$  and any interpretation  $\sigma$  of the variables, the triplet  $\langle \rho, i, \sigma \rangle$  (abbreviated as  $\rho_i^\sigma$ ) will be called a *world-valuation*.

**DEFINITION 99** (Satisfaction).  $\rho_i^\sigma \models \alpha$  ( $\rho_i^\sigma$  satisfies  $\alpha$ ) iff  $i \in \rho^\sigma(\alpha)$ .

**DEFINITION 100** (Verification).  $i \models \alpha$  ( $i$  verifies  $\alpha$ ) iff for any  $\sigma$ :  $\rho_i^\sigma \models \alpha$ .

DEFINITION 101 (Truth and logical truth).  $\models_{\mathcal{K}} \alpha$  ( $\alpha$  is *true* in  $\mathcal{K}$ ) iff for any  $i$ :  $i \models \alpha$ ;

$\models_{\mathbf{OL}} \alpha$  ( $\alpha$  is a *logical truth* of **OL**) iff for any  $\mathcal{K}$ :  $\models_{\mathcal{K}} \alpha$ .

DEFINITION 102 (Consequence in a realization and logical consequence).

$T \models_{\mathcal{K}} \alpha$  iff for any  $i$  of  $\mathcal{K}$  and any  $\sigma$ :  $\rho_i^\sigma \models T \curvearrowright \rho_i^\sigma \models \alpha$ ;

$T \models_{\mathbf{OL}} \alpha$  iff for any realization  $\mathcal{K}$ :  $T \models_{\mathcal{K}} \alpha$ .

The algebraic and the Kripkean characterization for first-order **OQL** can be obtained, in the obvious way, by requiring that any realization be orthomodular.

In both semantics for first-order **QL** one can prove a coincidence lemma:

LEMMA 103. *Given  $\mathcal{A} = \langle \mathcal{B}^C, D, v \rangle$  and  $\mathcal{K} = \langle I, R, \Pi^C, U, \rho \rangle$ :*

(103.1) *If  $\sigma$  and  $\sigma^*$  coincide in the values attributed to the variables occurring in a term  $t$ , then  $v^\sigma(t) = v^{\sigma^*}(t)$ ;  $\rho^\sigma(t) = \rho^{\sigma^*}(t)$ .*

(103.2) *If  $\sigma$  and  $\sigma^*$  coincide in the values attributed to the free variables occurring in a formula  $\alpha$ , then  $v^\sigma(\alpha) = v^{\sigma^*}(\alpha)$ ;  $\rho^\sigma(\alpha) = \rho^{\sigma^*}(\alpha)$ .*

One can easily prove, like in the sentential case, the following lemma:

LEMMA 104.

(104.1) *For any algebraic realization  $\mathcal{A}$  there exists a Kripkean realization  $\mathcal{K}^{\mathcal{A}}$  such that for any  $\alpha$ :  $\models_{\mathcal{A}} \alpha$  iff  $\models_{\mathcal{K}^{\mathcal{A}}} \alpha$ . Further, if  $\mathcal{A}$  is orthomodular then  $\mathcal{K}^{\mathcal{A}}$  is orthomodular.*

(104.2) *For any Kripkean realization  $\mathcal{K}$ , there exists an algebraic realization  $\mathcal{A}^{\mathcal{K}}$  such that for any  $\alpha$ :  $\models_{\mathcal{K}} \alpha$  iff  $\models_{\mathcal{A}^{\mathcal{K}}} \alpha$ . Further, if  $\mathcal{K}$  is orthomodular then  $\mathcal{A}^{\mathcal{K}}$  is orthomodular.*

An axiomatization of first-order **OL** (**OQL**) can be obtained by adding to the rules of our **OL** (**OQL**)-sentential calculus the following new rules:

(PR1)  $T \cup \{\forall x\alpha\} \vdash \alpha(x/t)$ , where  $\alpha(x/t)$  indicates a legitimate substitution).

(PR2)  $\frac{T \vdash \alpha}{T \vdash \forall x\alpha}$  (provided  $x$  is not free in  $T$ ).

All the basic syntactical notions are defined in the expected way. One can prove that any consistent set of sentences  $T$  admits of a consistent *inductive* extension  $T^*$ , such that  $T^* \vdash \forall x\alpha(t)$  whenever for any closed term  $t$ ,  $T^* \vdash \alpha(t)$ . The “weak Lindenbaum theorem” can be strengthened as follows: for any sentence  $\alpha$ , if  $T \not\vdash \neg\alpha$  then there exists a consistent and

inductive  $T^*$  such that:

$$T \text{ is syntactically compatible with } T^* \text{ and } T^* \vdash \alpha.^{12}$$

One can prove a soundness and a completeness theorem of our calculus with respect to the Kripkean semantics.

**THEOREM 105.** *Soundness.*

$$T \vdash \alpha \quad \curvearrowright \quad T \models \alpha.$$

**Proof.** Straightforward. ■

**THEOREM 106.** *Completeness.*

$$T \models \alpha \quad \curvearrowright \quad T \vdash \alpha.$$

**Sketch of the proof** Like in the sentential case, it is sufficient to construct a canonical model  $\mathcal{K} = \langle I, R, \Pi^C, U, \rho \rangle$  such that  $T \vdash \alpha$  iff  $T \models_{\mathcal{K}} \alpha$ .

*Definition of the canonical model*

- (i)  $I$  is the set of all consistent, deductively closed and inductive sets of sentences expressed in a common language  $\mathcal{L}^K$ , which is an extension of the original language;
- (ii)  $R$  is determined like in the sentential case;
- (iii)  $U$  is a set of rigid individual concepts that is naturally determined by the set of all individual constants of the extended language  $\mathcal{L}^K$ . For any constant  $c$  of  $\mathcal{L}^K$ , let  $\mathbf{u}^c$  be the corresponding individual concept in  $U$ . We require: for any world  $i$ ,  $\mathbf{u}^c(i) = c$ . In other words, the reference of the individual concept  $\mathbf{u}^c$  is in any world the constant  $c$ . We will indicate by  $c^{\mathbf{u}}$  the constant corresponding to  $\mathbf{u}$ .

- (iv)  $\rho(a_m) = \mathbf{u}^{a_m}$ ;  
 $\rho(P_m^n)(\mathbf{u}_1^{c_1}, \dots, \mathbf{u}_n^{c_n}) = \{i \mid P_m^n c_1, \dots, c_n \in i\}$ .

Our  $\rho$  is well defined since one can prove for any sentence  $\alpha$  of  $\mathcal{L}^K$ :

$$i \not\vdash \alpha \quad \curvearrowright \quad \exists j \not\vdash i : j \vdash \neg \alpha.$$

As a consequence,  $\rho^\sigma(P_m^n t_1, \dots, t_n)$  is a possible proposition.

- (v)  $\Pi^C$  is the set of all “meanings” of formulas (i.e.,  $X \in \Pi^C$  iff  $\exists \alpha \exists \sigma (X = \rho^\sigma(\alpha))$ );  $C$  is the set of all sets  $\{\rho^{\sigma[x/\mathbf{u}]}(\beta) \mid \mathbf{u} \in U\}$  for any formula  $\beta$ .

---

<sup>12</sup>By Definition 71,  $T$  is syntactically compatible with  $T^*$  iff there is no formula  $\alpha$  such that  $T \vdash \alpha$  and  $T^* \vdash \neg \alpha$ .

One can easily check that  $\mathcal{K}$  is a “good” realization with a constant domain.

LEMMA 107. *Lemma of the canonical model.*

For any  $\alpha$ , any  $i \in I$  and any  $\sigma$ :

$$\rho_i^\sigma \models \alpha \text{ iff } \alpha^\sigma \in i,$$

where  $\alpha^\sigma$  is the sentence obtained by substituting in  $\alpha$  any free variable  $x$  with the constant  $c^{\sigma(x)}$  corresponding to the individual concept  $\sigma(x)$ .

**Sketch of the proof.** By induction on the length of  $\alpha$ . The cases  $\alpha = P_m^n t_1, \dots, t_n$ ,  $\alpha = \neg\beta$ ,  $\alpha = \beta \wedge \gamma$  are proved by an obvious transformation of the sentential argument. Let us consider the case  $\alpha = \forall x\beta$  and suppose  $x$  occurring in  $\beta$  (otherwise the proof is trivial). In order to prove the left to right implication, suppose  $\rho_i^\sigma \models \forall x\beta$ . Then, for any  $\mathbf{u}$  in  $U$ ,  $\rho_i^{\sigma[x/\mathbf{u}]} \models \beta(x)$ . Hence, by inductive hypothesis,  $\forall \mathbf{u} \in U$ ,  $[\beta(x)]^{\sigma[x/\mathbf{u}]} \in i$ . In other words, for any constant  $c^{\mathbf{u}}$  of  $i$ :  $[\beta(x)]^\sigma(x/c^{\mathbf{u}}) \in i$ . And, since  $i$  is inductive and deductively closed:  $\forall x\beta(x)^\sigma \in i$ . In order to prove the right to left implication, suppose  $[\forall x\beta(x)]^\sigma \in i$ . Then, [by (PR1)], for any constant  $c$  of  $i$ :  $[\beta(x/c)]^\sigma \in i$ . Hence by inductive hypothesis: for any  $\mathbf{u}^c \in U$ ,  $\rho_i^{\sigma[x/\mathbf{u}^c]} \models \beta(x)$ , i.e.,  $\rho_i^\sigma \models \forall x\beta(x)$ . On this ground, similarly to the sentential case, one can prove  $T \vdash \alpha$  iff  $T \models_{\mathcal{K}} \alpha$ . ■

First-order **QL** can be easily extended (in a standard way) to a first-order logic with identity. However, a critical problem is represented by the possibility of developing, within this logic, a satisfactory *theory of descriptions*. The main difficulty can be sketched as follows. A natural condition to be required in any characterization of a  $\iota$ -operator is obviously the following:

$$\begin{aligned} & \exists x \{ \beta(x) \wedge \forall y [(\beta(y) \wedge x = y) \vee (\neg\beta(y) \wedge \neg x = y)] \wedge \alpha(x) \} \\ & \text{is true} \quad \rightsquigarrow \quad \alpha(\iota x \beta(x)) \text{ is true.} \end{aligned}$$

However, in **QL**, the truth of the antecedent of our implication does not generally guarantee the existence of a particular individual such that  $\iota x \beta$  can be regarded as a name for such an individual. As a counterexample, let us consider the following case (in the algebraic semantics): let  $\mathcal{A}$  be  $\langle \mathcal{B}, D, v \rangle$  where  $\mathcal{B}$  is the complete orthomodular lattice based on the set of all closed subspaces of the plane  $\mathbb{R}^2$ , and  $D$  contains exactly two individuals  $\mathbf{d}_1, \mathbf{d}_2$ . Let  $P$  be a monadic predicate and  $X, Y$  two orthogonal unidimensional subspaces of  $B$  such that  $v(P)(\mathbf{d}_1) = X$ ,  $v(P)(\mathbf{d}_2) = Y$ . If the equality predicate  $=$  is interpreted as the standard identity relation (i.e.,  $v^\sigma(t_1 = t_2) = \mathbf{1}$ , if  $v^\sigma(t_1) = v^\sigma(t_2)$ ;  $\mathbf{0}$ , otherwise), one can easily calculate:

$$v(\exists x [Px \wedge \forall y ((Py \wedge x = y) \vee (\neg Py \wedge \neg x = y))]) = \mathbf{1}.$$



However, for both individuals  $\mathbf{d}_1, \mathbf{d}_2$  of the domain, we have:

$$v^{\sigma[x/\mathbf{d}_1]}(Px) \neq 1, v^{\sigma[x/\mathbf{d}_2]}(Px) \neq 1.$$

In other words, there is no precise individual in the domain that satisfies the property expressed by our predicate  $P$ !

## 10 QUANTUM SET THEORIES AND THEORIES OF QUASISSETS

An important application of **QL** to set theory has been developed by [Takeuti, 1981]. We will sketch here only the fundamental idea of this application. Let  $\mathcal{L}$  be a standard set-theoretical language. One can construct *ortho-valued models* for  $\mathcal{L}$ , which are formally very similar to the usual *Boolean-valued models* for standard set-theory, with the following difference: the set of truth-values is supposed to have the algebraic structure of a complete orthomodular lattice, instead of a complete Boolean algebra. Let  $\mathcal{B}$  be a complete orthomodular lattice, and let  $\nu, \lambda, \dots$  represent ordinal numbers. An *ortho-valued (set-theoretical) universe*  $V$  is constructed as follows:

$$V^{\mathcal{B}} = \bigcup_{\nu \in \mathcal{O}_n} V(\nu), \text{ where:}$$

$$V(0) = \emptyset.$$

$$V(\nu+1) = \{g \mid g \text{ is a function and } Dom(g) \subseteq V(\nu) \text{ and } Rang(g) \subseteq B\}.$$

$$V(\lambda) = \bigcup_{\nu < \lambda} V(\nu), \text{ for any limit-ordinal } \lambda.$$

(  $Dom(g)$  and  $Rang(g)$  are the *domain* and the *range* of function  $g$ , respectively).

Given an ortho-valued universe  $V^{\mathcal{B}}$  one can define for any formula of  $\mathcal{L}$  the truth-value  $\llbracket \alpha \rrbracket^{\sigma}$  in  $\mathcal{B}$  induced by any interpretation  $\sigma$  of the variables in the universe  $V^{\mathcal{B}}$ .

$$\begin{aligned} \llbracket x \in y \rrbracket^{\sigma} &= \bigsqcup_{g \in Dom(\sigma(y))} \{ \sigma(y)(g) \sqcap \llbracket x = z \rrbracket^{\sigma[z/g]} \} \\ \llbracket x = y \rrbracket^{\sigma} &= \prod_{g \in Dom(\sigma(x))} \{ \sigma(x)(g) \rightsquigarrow \llbracket z \in y \rrbracket^{\sigma[z/g]} \} \sqcap \\ &\quad \prod_{g \in Dom(\sigma(y))} \{ \sigma(y)(g) \rightsquigarrow \llbracket z \in x \rrbracket^{\sigma[z/g]} \}. \end{aligned}$$

where  $\rightsquigarrow$  is the quantum logical conditional operation ( $a \rightsquigarrow b := a' \sqcup (a \sqcap b)$ , for any  $a, b \in B$ ).

A formula  $\alpha$  is called *true* in the universe  $V^{\mathcal{B}}$  ( $\models_{V^{\mathcal{B}}} \alpha$ ) iff  $\llbracket \alpha \rrbracket^{\sigma} = \mathbf{1}$ , for any  $\sigma$ .

Interestingly enough, the segment  $V(\omega)$  of  $V^{\mathcal{B}}$  turns out to contain some important mathematical objects, that we can call *quantum-logical natural numbers*.

The standard axioms of set-theory hold in  $\mathcal{B}$  only in a restricted form. An extremely interesting property of  $V^{\mathcal{B}}$  is connected with the notion of identity. Differently from the case of Boolean-valued models, the identity relation in  $V^{\mathcal{B}}$  turns out to be non-Leibnizian. For, one can choose an orthomodular lattice  $\mathcal{B}$  such that:

$$\not\equiv_{V^{\mathcal{B}}} x = y \rightarrow \forall z(x \in z \leftrightarrow y \in z).$$

According to our semantic definitions, the relation  $=$  represents a kind of “extensional equality”. As a consequence, one may conclude that two quantum-sets that are extensionally equal do not necessarily share all the same properties. Such a failure of the Leibniz-substitutivity principle in quantum set theory might perhaps find interesting applications in the field of intensional logics.

A completely different approach is followed in the framework of the theories of *quasisets* (or *quasets*). The basic aim of these theories is to provide a mathematical description for collections of microobjects, which seem to violate some characteristic properties of the classical identity relation.

In some of his general writings, Schrödinger discussed the inconsistency between the classical concept of physical object (conceived as an individual entity) and the behaviour of particles in quantum mechanics. Quantum particles – he noticed – lack individuality and the concept of identity cannot be applied to them, similarly to the case of classical objects.

One of the aims of the *theories of quasisets* (proposed by [da Costa *et al.*, 1992]) is to describe formally the following idea defended by Schrödinger: identity is generally not defined for microobjects. As a consequence, one cannot even assert that an “electron is identical with itself”. In the realm of microobjects only an *indistinguishability relation* (an equivalence relation that may violate the substitutivity principle) makes sense.

On this basis, different formal systems have been proposed. Generally, these systems represent convenient generalizations of a Zermelo–Fraenkel like set theory with *urelements*. Differently from the classical case, an urelement may be either a *macro* or a *microobject*. Collections are represented by *quasisets* and classical sets turn out to be limit cases of quasisets.

A somewhat different approach has been followed in the *theory of quasets* (proposed in [Dalla Chiara and Toraldo di Francia, 1993]).

The starting point is based on the following observation: physical kinds and compound systems in QM seem to share some features that are characteristic of intensional entities. Further, the relation between intensions and extensions turns out to behave quite differently from the classical semantic situations. Generally, one cannot say that a quantum intensional notion uniquely determines a corresponding extension. For instance, take the notion of *electron*, whose intension is well defined by the following physical property: mass =  $9.1 \times 10^{-28}$ g, electron charge =  $4.8 \times 10^{-10}$ e.s.u., spin

$= 1/2$ . Does this property determine a corresponding *set*, whose elements should be all and only the physical objects that satisfy our property at a certain time interval? The answer is negative. In fact, physicists have the possibility of recognizing, by theoretical or experimental means, whether a given physical system is an electron system or not. If yes, they can also enumerate all the quantum states available within it. But they can do so in a number of different ways. For example, take the spin. One can choose the  $x$ -axis and state how many electrons have spin up and how many have spin down. However, we could instead refer to the  $z$ -axis or any other direction, obtaining *different collections* of quantum states, all having the same cardinality. This seems to suggest that microobject systems present an irreducibly intensional behaviour: generally they do not determine precise extensions and are not determined thereby. Accordingly, a basic feature of the theory is a strong violation of the extensionality principle.

Quasets are convenient generalizations of classical sets, where both the extensionality axiom and Leibniz' principle of indiscernibles are violated. Generally a quaset has only a cardinal but not an ordinal number, since it cannot be well ordered.

## 11 THE UNSHARP APPROACHES

The unsharp approaches to QT (first proposed by [Ludwig, 1983] and further developed by Kraus, Davies, Mittelstaedt, Busch, Lahti, Bugajski, Beltrametti, Cattaneo and many others) have been suggested by some deep criticism of the standard logico-algebraic approach. Orthodox quantum logic (based on Birkhoff and von Neumann's proposal) turns out to be at the same time a *total* and a *sharp* logic. It is total because the *meaningful propositions* are represented as closed under the basic logical operations: the conjunction (disjunction) of two meaningful propositions is a meaningful proposition. Further, it is also sharp, because propositions, in the standard interpretation, correspond to *exact* possible properties of the physical system under investigation. These properties express the fact that "the value of a given observable lies in a certain *exact* Borel set".

As we have seen, the set of the physical properties, that may hold for a quantum system, is mathematically represented by the set of all closed subspaces of the Hilbert space associated to our system. Instead of closed subspaces, one can equivalently refer to the set of all *projections*, that is in one-to-one correspondence with the set of all closed subspaces. Such a correspondence leads to a collapse of different semantic notions, which Foulis and Randall described as the "metaphysical disaster" of orthodox QT. The collapse involves the notions of "experimental proposition", "physical property", "physical event" (which represent *empirical* and *intensional* concepts), and the notion of *proposition* as a *set* of states (which corresponds

to a typical *extensional* notion according to the tradition of standard semantics).

Both the total and the sharp character of **QL** have been put in question in different contexts. One of the basic ideas of the unsharp approaches is a “liberalization” of the mathematical counterpart for the intuitive notion of “experimental proposition”. Let  $P$  be a projection operator in the Hilbert space  $\mathcal{H}$ , associated to the physical system under investigation. Suppose  $P$  describes an experimental proposition and let  $W$  be a statistical operator representing a possible state of our system. Then, according to one of the axioms of the theory (the *Born rule*), the number  $\text{Tr}(WP)$  (the *trace* of the operator  $WP$ ) will represent the probability-value that our system in state  $W$  verifies  $P$ . This value is also called *Born probability*. However, projections are not the only operators for which a Born probability can be defined. Let us consider the class  $E(\mathcal{H})$  of all linear bounded operators  $E$  such that for any statistical operator  $W$ ,

$$\text{Tr}(WE) \in [0, 1].$$

It turns out that  $E(\mathcal{H})$  properly includes the set  $P(\mathcal{H})$  of all projections on  $\mathcal{H}$ . The elements of  $E(\mathcal{H})$  represent, in a sense, a “maximal” mathematical representative for the notion of experimental proposition, in agreement with the probabilistic rules of quantum theory. In the framework of the unsharp approach,  $E(\mathcal{H})$  has been called the set of all *effects*.<sup>13</sup> An important difference between projections and proper effects is the following: projections can be associated to *sharp* propositions having the form “the value for the observable  $A$  lies in the *exact* Borel set  $\Delta$ ”, while effects may represent also *fuzzy* propositions like “the value of the observable  $A$  lies in the *fuzzy* Borel set  $\Gamma$ ”. As a consequence, there are effects  $E$ , different from the null projection  $\mathbb{0}$ , such that no state  $W$  can verify  $E$  with probability 1. A limit case is represented by the *semitransparent effect*  $\frac{1}{2}\mathbb{I}$  (where  $\mathbb{I}$  is the identity operator), to which any state  $W$  assigns probability-value  $\frac{1}{2}$ .

From the intuitive point of view, one could say that moving to an unsharp approach represents an important step towards a kind of “second degree of fuzziness”. In the framework of the sharp approach, any physical event  $E$  can be regarded as a kind of “clear” property. Whenever a state  $W$  assigns to  $E$  a probability value different from 1 and 0, one can think that the semantic uncertainty involved in such a situation totally depends on the ambiguity of the state (first degree of fuzziness). In other words, even a pure state in QT does not represent a *logically complete information*, that is able to decide any possible physical event. In the unsharp approaches, instead, one take into account also “genuine ambiguous properties”. This second degree of fuzziness may be regarded as depending on the accuracy

<sup>13</sup>It is easy to see that an effect  $E$  is a projection iff  $E^2 := EE = E$ . In other words, projections are idempotent effects.

of the measurement (which tests the property), and also on the accuracy involved in the operational definition for the physical quantities which our property refers to.

## 12 EFFECT STRUCTURES

Different algebraic structures can be induced on the class  $E(\mathcal{H})$  of all effects. Let us first recall some definitions.

**DEFINITION 108** (Involutive bounded poset (lattice)). An involutive bounded poset (lattice) is a structure  $\mathcal{B} = \langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$ , where  $\langle B, \sqsubseteq, \mathbf{1}, \mathbf{0} \rangle$  is a partially ordered set (lattice) with maximum ( $\mathbf{1}$ ) and minimum ( $\mathbf{0}$ );  $'$  is a 1-ary operation on  $B$  such that the following conditions are satisfied: (i)  $a'' = a$ ; (ii)  $a \sqsubseteq b \iff b' \sqsubseteq a'$ .

**DEFINITION 109** (Orthoposet). An *orthoposet* is an involutive bounded poset that satisfies the non contradiction principle:

$$a \sqcap a' = \mathbf{0}.$$

**DEFINITION 110** (Orthomodular poset). An *orthomodular poset* is an orthoposet that is closed under the orthogonal sup ( $a \sqsubseteq b' \iff a \sqcup b$  exists) and satisfies the orthomodular property:

$$a \sqsubseteq b \iff \exists c \text{ such that } a \sqsubseteq c' \text{ and } b = a \sqcup c.$$

**DEFINITION 111** (Regularity). An involutive bounded poset (lattice)  $\mathcal{B}$  is *regular* iff  $a \sqsubseteq a'$  and  $b \sqsubseteq b' \iff a \sqsubseteq b'$ .

Whenever an involutive bounded poset  $\mathcal{B}$  is a lattice, then  $\mathcal{B}$  is regular iff it satisfies the *Kleene condition*:

$$a \sqcap a' \sqsubseteq b \sqcup b'.$$

The set  $E(\mathcal{H})$  of all effects can be naturally structured as an involutive bounded poset:

$$\mathcal{E}(\mathcal{H}) = \langle E(\mathcal{H}), \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle,$$

where

- (i)  $E \sqsubseteq F$  iff for any state (statistical operator)  $W$ ,  $\text{Tr}(WE) \leq \text{Tr}(WF)$  (in other words, any state assigns to  $E$  a probability-value that is less or equal than the probability-value assigned to  $F$ );
- (ii)  $\mathbf{1}, \mathbf{0}$  are the identity ( $\mathbb{I}$ ) and the null ( $\mathbb{0}$ ) projection, respectively;
- (iii)  $E' = \mathbf{1} - E$ .

One can easily check that  $\sqsubseteq$  is a partial order,  $'$  is an order-reversing involution, while  $\mathbf{1}$  and  $\mathbf{0}$  are respectively the maximum and the minimum with respect to  $\sqsubseteq$ . At the same time this poset fails to be a lattice. Differently from projections, some pairs of effects have no infimum and no supremum as the following example shows [Greechie and Gudder, n.d.]:

EXAMPLE 112. Let us consider the following effects (in the matrix-representation) on the Hilbert space  $\mathbb{R}^2$ :

$$E = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{2} \end{pmatrix} \quad F = \begin{pmatrix} \frac{3}{4} & 0 \\ 0 & \frac{1}{4} \end{pmatrix} \quad G = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{4} \end{pmatrix}$$

It is not hard to see that  $G \sqsubseteq E, F$ . Suppose, by contradiction, that  $L = E \sqcap F$  exists in  $E(\mathbb{R}^2)$ . An easy computation shows that  $L$  must be equal to  $G$ . Let

$$M = \begin{pmatrix} \frac{7}{16} & \frac{1}{8} \\ \frac{1}{8} & \frac{3}{16} \end{pmatrix}$$

Then  $M$  is an effect such that  $M \sqsubseteq E, F$ ; however,  $M \not\sqsubseteq L$ , which is a contradiction.

In order to obtain a lattice structure, one has to embed  $\mathcal{E}(\mathcal{H})$  into its *MacNeille completion*  $\overline{\mathcal{E}(\mathcal{H})}$ .

### The MacNeille completion of an involutive bounded poset

Let  $\langle B, \sqsubseteq, \mathbf{1}, \mathbf{0} \rangle$  be an involutive bounded poset. For any non-empty subset  $X$  of  $B$ , let  $l(X)$  and  $u(X)$  represent respectively the set of all lower bounds and the set of all upper bounds of  $X$ . Let  $MC(B) := \{X \subseteq B \mid X = u(l(X))\}$ . It turns out that  $X \in MC(B)$  iff  $X = X''$ , where  $X' := \{a \in B \mid \forall b \in X : a \sqsubseteq b'\}$ . Moreover, the structure

$$\overline{\mathcal{B}} = \langle MC(B), \subseteq, ', \{\mathbf{0}\}, B \rangle$$

is a complete involutive bounded lattice (which is regular if  $\mathcal{B}$  is regular), where  $X \sqcap Y = X \cap Y$  and  $X \sqcup Y = (X \cup Y)''$ .

It turns out that  $\mathcal{B}$  is embeddable into  $\overline{\mathcal{B}}$ , via the map  $h : a \rightarrow \langle a \rangle$ , where  $\langle a \rangle$  is the principal ideal generated by  $a$ . Such an embedding preserves the *infimum* and the *supremum*, when existing in  $\mathcal{B}$ .

The Mac Neille completion of an involutive bounded poset does not generally satisfies the non contradiction principle ( $a \sqcap a' = \mathbf{0}$ ) and the excluded middle principle ( $a \sqcup a' = \mathbf{1}$ ). As a consequence, differently from the projection case, the Mac Neille completion of  $\mathcal{E}(\mathcal{H})$  is not an ortholattice. Apparently, our operation  $'$  turns out to behave as a *fuzzy negation*, both in the case of  $\mathcal{E}(\mathcal{H})$  and of its Mac Neille completion. This is one of the reasons why proper effects (that are not projections) may be regarded as

representing *unsharp physical properties*, possibly violating the non contradiction principle.

The effect poset  $\mathcal{E}(\mathcal{H})$  can be naturally extended to a richer structure, equipped with a new complement  $\sim$ , that has an intuitionistic-like behaviour:

$E^\sim$  is the projection operator  $P_{Ker(E)}$  whose range is the kernel  $Ker(E)$  of  $E$ , consisting of all vectors that are transformed by the operator  $E$  into the null vector.

By definition, the intuitionistic complement of an effect is always a projection. In the particular case, where  $E$  is a projection, it turns out that:  $E' = E^\sim$ . In other words, the fuzzy and the intuitionistic complement collapse into one and the same operation.

The structure  $\langle \mathcal{E}(\mathcal{H}), \sqsubseteq, ', \sim, \mathbf{1}, \mathbf{0} \rangle$  turns out to be a particular example of a *Brouwer Zadeh poset* [Cattaneo and Nisticò, 1986].

**DEFINITION 113.** A *Brouwer–Zadeh poset* (simply a *BZ-poset*) is a structure  $\langle B, \sqsubseteq, ', \sim, \mathbf{1}, \mathbf{0} \rangle$ , where

(113.1)  $\langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  is a regular involutive bounded poset;

(113.2)  $\sim$  is a 1-ary operation on  $B$ , which behaves like an intuitionistic complement:

- (i)  $a \sqcap a^\sim = \mathbf{0}$ .
- (ii)  $a \sqsubseteq a^{\sim\sim}$ .
- (iii)  $a \sqsubseteq b \iff b^\sim \sqsubseteq a^\sim$ .

(113.3) The following relation connects the fuzzy and the intuitionistic complement:

$$a^{\sim'} = a^{\sim\sim}$$

**DEFINITION 114.** A *Brouwer Zadeh lattice* is a BZ-poset that is also a lattice.

**The Mac Neille completion of a BZ-poset**

Let  $\mathcal{B} = \langle B, \sqsubseteq, ', \sim, \mathbf{1}, \mathbf{0} \rangle$  be a BZ-poset and let  $\overline{\mathcal{B}}$  the Mac Neille completion of the regular involutive bounded poset  $\langle B, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$ . For any non-empty subset  $X$  of  $B$ , let

$$X^\sim := \{a \in B \mid \forall b \in X : a \sqsubseteq b^\sim\}.$$

It turns out that  $\overline{\mathcal{B}} = \langle MC(B), \subseteq, ', \sim, \{\mathbf{0}\}, B \rangle$  is a complete BZ-lattice [Giuntini, 1991], which  $\mathcal{B}$  can be embedded into, via the map  $h$  defined above.

Another interesting way of structuring the set of all effects can be obtained by using a particular kind of partial structure, that has been called *effect algebra* [Foulis and Bennett, 1994] or *unsharp orthoalgebra* [Dalla Chiara and Giuntini, 1994]. Abstract effect algebras are defined as follows:

DEFINITION 115. An *effect algebra* is a partial structure  $\mathcal{A} = \langle A, \boxplus, \mathbf{1}, \mathbf{0} \rangle$  where  $\boxplus$  is a partial binary operation on  $A$ . When  $\boxplus$  is defined for a pair  $a, b \in A$ , we will write  $\exists(a \boxplus b)$ . The following conditions hold:

- (i) *Weak commutativity*  
 $\exists(a \boxplus b) \curvearrowright \exists(b \boxplus a)$  and  $a \boxplus b = b \boxplus a$ .
- (ii) *Weak associativity*  
 $[\exists(b \boxplus c) \text{ and } \exists(a \boxplus (b \boxplus c))] \curvearrowright [\exists(a \boxplus b) \text{ and } \exists((a \boxplus b) \boxplus c)]$   
and  $a \boxplus (b \boxplus c) = (a \boxplus b) \boxplus c$ .
- (iii) *Strong excluded middle*  
For any  $a$ , there exists a unique  $x$  such that  $a \boxplus x = \mathbf{1}$ .
- (iv) *Weak consistency*  
 $\exists(a \boxplus \mathbf{1}) \curvearrowright a = \mathbf{0}$ .

From an intuitive point of view, our operation  $\boxplus$  can be regarded as an *exclusive disjunction* (*aut*), which is defined only for pairs of logically incompatible events.

An orthogonality relation  $\perp$ , a partial order relation  $\sqsubseteq$  and a generalized complement  $'$  can be defined in any effect algebra.

DEFINITION 116. Let  $\mathcal{A} = \langle A, \boxplus, \mathbf{1}, \mathbf{0} \rangle$  be an effect algebra and let  $a, b \in A$ .

- (i)  $a \perp b$  iff  $a \boxplus b$  is defined in  $A$ .
- (ii)  $a \sqsubseteq b$  iff  $\exists c \in A$  such that  $a \perp c$  and  $b = a \boxplus c$ .
- (iii) The *generalized complement* of  $a$  is the unique element  $a'$  such that  $a \boxplus a' = \mathbf{1}$  (the definition is justified by the strong excluded middle condition).

The category of all effect algebras turns out to be (categorically) equivalent to the category of all *difference posets*, which have been first studied in [Kôpka and Chovanec, 1994] and further investigated in [Dvurečenskij and Pulmannová, 1994].

Effect algebras that satisfy the non contradiction principle are called *orthoalgebras*:

DEFINITION 117. An *orthoalgebra* is an effect algebra  $\mathcal{B} = \langle B, \boxplus, \mathbf{1}, \mathbf{0} \rangle$  such that the following condition is satisfied:



*Strong consistency*

$$\exists (a \boxplus a) \curvearrowright a = \mathbf{0}.$$

In other words:  $\mathbf{0}$  is the only element that is orthogonal to itself.

In order to induce the structure of an effect algebra on  $E(\mathcal{H})$ , it is sufficient to define a partial sum  $\boxplus$  as follows:

$$\exists (E \boxplus F) \text{ iff } E + F \in E(\mathcal{H}),$$

where  $+$  is the usual sum-operator. Further:

$$\exists (E \boxplus F) \curvearrowright E \boxplus F = E + F.$$

It turns out that the structure  $\langle E(\mathcal{H}), \boxplus, \mathbb{I}, \mathbf{0} \rangle$  is an effect algebra, where the generalized complement of any effect  $E$  is just  $\mathbb{I} - E$ . At the same time, this structure fails to be an orthoalgebra.

Any abstract effect algebra

$$\mathcal{A} = \langle A, \boxplus, \mathbf{1}, \mathbf{0} \rangle$$

can be naturally extended to a kind of total structure, that has been termed *quantum MV-algebra* (abbreviated as QMV-algebra) [Giuntini, 1996].

Before introducing QMV-algebras, it will be expedient to recall the definition of MV-algebra. As is well known, *MV-algebras* (*multi-valued algebras*) have been introduced by Chang [1957] in order to provide an algebraic proof of the completeness theorem for Łukasiewicz' infinite-many-valued logic  $L_{\aleph}$ . A "privileged" model of this logic is based on the real interval  $[0, 1]$ , which gives rise to a particular example of a totally ordered (or linear) MV-algebra.

Both MV-algebras and QMV-algebras are total structures having the following form:

$$\mathcal{M} = (M, \oplus, *, \mathbf{1}, \mathbf{0})$$

where:

- (i)  $\mathbf{1}, \mathbf{0}$  represent the certain and the impossible propositions (or alternatively the two extreme truth values);
- (ii)  $*$  is the negation-operation;
- (iii)  $\oplus$  represents a disjunction (*or*) which is generally non idempotent ( $a \oplus a \neq a$ ).

A (generally non idempotent) conjunction (*and*) is then defined via de Morgan law:

$$a \odot b := (a^* \oplus b^*)^*.$$

On this basis, a pair consisting of an idempotent conjunction *et* ( $\mathfrak{m}$ ) and of an idempotent disjunction *vel* ( $\mathfrak{u}$ ) is then defined:

$$a \mathfrak{m} b := (a \oplus b^*) \odot b$$

$$a \mathfrak{u} b := (a \odot b^*) \oplus b.$$

In the concrete MV-algebra based on  $[0, 1]$ , the operations are defined as follows:

(i)  $\mathbf{1} = 1$ ;  $\mathbf{0} = 0$ ;

(ii)  $a^* = 1 - a$ ;

(iii)  $\oplus$  is the *truncated sum*:

$$a \oplus b = \begin{cases} a + b, & \text{if } a + b \leq 1; \\ 1, & \text{otherwise.} \end{cases}$$

In this particular case, it turns out that:

$$a \mathfrak{m} b = \text{Min}\{a, b\}$$

(*a et b* is the minimum between *a* and *b*).

$$a \mathfrak{u} b = \text{Max}\{a, b\}$$

(*a vel b* is the maximum between *a* and *b*).

A standard abstract definition of MV-algebras is the following [Mangani, 1973]:

DEFINITION 118. An *MV-algebra* is a structure  $\mathcal{M} = (M, \oplus, *, \mathbf{1}, \mathbf{0})$ , where  $\oplus$  is a binary operation,  $*$  is a unary operation and  $\mathbf{0}$  and  $\mathbf{1}$  are special elements of  $M$ , satisfying the following axioms:

(MV1)  $(a \oplus b) \oplus c = a \oplus (b \oplus c)$

(MV2)  $a \oplus \mathbf{0} = a$

(MV3)  $a \oplus b = b \oplus a$

(MV4)  $a \oplus \mathbf{1} = \mathbf{1}$

(MV5)  $(a^*)^* = a$

(MV6)  $\mathbf{0}^* = \mathbf{1}$

(MV7)  $a \oplus a^* = \mathbf{1}$

(MV8)  $(a^* \oplus b)^* \oplus b = (a \oplus b^*)^* \oplus a$

In other words, an MV-algebra represents a particular weakening of a Boolean algebra, where  $\oplus$  and  $\odot$  are generally non idempotent.

A partial order relation can be defined in any MV-algebra in the following way:

$$a \preceq b \text{ iff } a \mathbin{\&}\! \mathbin{\&}\! b = a.$$

Some important properties of MV-algebras are the following:

- (i) the structure  $\langle M, \preceq, *, \mathbf{1}, \mathbf{0} \rangle$  is a bounded involutive distributive lattice, where  $a \mathbin{\&}\! \mathbin{\&}\! b$  ( $a \mathbin{\cup}\! \mathbin{\cup}\! b$ ) is the *inf* (*sup*) of  $a, b$ ;
- (ii) the non-contradiction principle and the excluded middle principles for  $*, \mathbin{\&}\! \mathbin{\&}\!, \mathbin{\cup}\! \mathbin{\cup}\!$  are generally violated:  $a \mathbin{\cup}\! \mathbin{\cup}\! a^* \neq \mathbf{1}$  and  $a \mathbin{\&}\! \mathbin{\&}\! a^* \neq \mathbf{0}$  are possible. As a consequence, MV algebras permit us to describe *fuzzy* and *paraconsistent* situations;
- (iii)  $a^* \oplus b = \mathbf{1}$  iff  $a \preceq b$ . In other words: similarly to the Boolean case, “not- $a$  or  $b$ ” represents a *good material implication*;
- (iv) every MV-algebra is a subdirect product of totally ordered MV-algebras [Chang, 1958];
- (v) an equation holds in the class of all MV-algebras iff it holds in the concrete MV-algebra based on  $[0, 1]$  [Chang, 1958].

Let us now go back to our effect-structure  $\langle E(\mathcal{H}), \boxplus, \mathbf{1}, \mathbf{0} \rangle$ . The partial operation  $\boxplus$  can be extended to a total operation  $\oplus$  that behaves like a truncated sum. For any  $E, F \in E(\mathcal{H})$ :

$$E \oplus F = \begin{cases} E + F, & \text{if } \exists(E \boxplus F); \\ \mathbf{1}, & \text{otherwise.} \end{cases}$$

Further, let us put:

$$E^* = \mathbb{I} - E.$$

The structure  $\mathcal{E}(\mathcal{H}) = \langle E(\mathcal{H}), \oplus, *, \mathbf{1}, \mathbf{0} \rangle$  turns out to be “very close” to an MV-algebra. However, something is missing:  $\mathcal{E}(\mathcal{H})$  satisfies the first seven axioms of our definition (MV1-MV7); at the same time one can easily check that the axiom (MV8) (usually called “Łukasiewicz axiom”) is violated. For instance, let us consider two non trivial projections  $P, Q$  such that  $P$  is not orthogonal to  $Q^*$  and  $Q$  is not orthogonal to  $P^*$ . Then, by definition of  $\oplus$ , we have that  $P \oplus Q^* = \mathbb{I}$  and  $Q \oplus P^* = \mathbb{I}$ . Hence:  $(P^* \oplus Q)^* \oplus Q = Q \neq P = (P \oplus Q^*)^* \oplus P$ .

As a consequence, Łukasiewicz axiom must be conveniently weakened to obtain a representation for our concrete effect structure. This can be done by means of the notion of QMV-algebra.

DEFINITION 119. A *quantum MV-algebra* (QMV-algebra) is a structure  $\mathcal{M} = (M, \oplus, *, \mathbf{1}, \mathbf{0})$  where  $\oplus$  is a binary operation,  $*$  is a 1-ary operation, and  $\mathbf{0}, \mathbf{1}$  are special elements of  $M$ . For any  $a, b \in M$ :  $a \odot b := (a^* \oplus b^*)^*$ ,  $a \pitchfork b := (a \oplus b^*) \odot a$ ,  $a \uplus b := (a \odot b^*) \oplus b$ . The following axioms are required:

- (QMV1)  $a \oplus (b \oplus c) = (b \oplus a) \oplus c$ ,
- (QMV2)  $a \oplus a^* = \mathbf{1}$ ,
- (QMV3)  $a \oplus \mathbf{0} = a$ ,
- (QMV4)  $a \oplus \mathbf{1} = \mathbf{1}$ ,
- (QMV5)  $a^{**} = a$ ,
- (QMV6)  $\mathbf{0}^* = \mathbf{1}$ ,
- (QMV7)  $a \oplus [(a^* \pitchfork b) \pitchfork (c \pitchfork a^*)] = (a \oplus b) \pitchfork (a \oplus c)$ .

The operations  $\pitchfork$  and  $\uplus$  of a QMV-algebra  $\mathcal{M}$  are generally non commutative. As a consequence, they do not represent lattice-operations. It is not difficult to prove that a QMV-algebra  $\mathcal{M}$  is an MV-algebra iff for all  $a, b \in M$ :  $a \pitchfork b = b \pitchfork a$ .

At the same time, any QMV-algebra  $\mathcal{M} = (M, \oplus, *, \mathbf{1}, \mathbf{0})$  gives rise to an involutive bounded poset  $\langle M, \preceq, *, \mathbf{1}, \mathbf{0} \rangle$ , where the partial order relation is defined like in the MV case.

One can easily show that QMV-algebras represent a “good abstraction” from the effect-structures:

THEOREM 120. *The structure  $\mathcal{E}(\mathcal{H}) = \langle E(\mathcal{H}), \oplus, *, \mathbf{1}, \mathbf{0} \rangle$  (where  $\oplus, *, \mathbf{1}, \mathbf{0}$  are the operations and the special elements previously defined) is a QMV-algebra.*

The QMV-algebra  $\mathcal{E}(\mathcal{H})$  cannot be linear. For, one can easily check that any linear QMV-algebra collapses into an MV-algebra.

In spite of this, our algebra of effects turns out to satisfy some weak forms of linearity.

DEFINITION 121. A QMV-algebra  $\mathcal{M}$  is called *weakly linear* iff  $\forall a, b \in M$ :  $a \pitchfork b = b$  or  $b \pitchfork a = a$ .

DEFINITION 122. A QMV-algebra  $\mathcal{M}$  is called *quasi-linear* iff  $\forall a, b \in M$ :  $a \pitchfork b = a$  or  $a \pitchfork b = b$ .

It is easy to see that every quasi-linear QMV-algebra is weakly linear, but not the other way around (because  $\pitchfork$  is not commutative).

A very strong relation connects the class of all effect algebras with the class of all quasi-linear QMV-algebras: every effect algebra can be uniquely transformed into a quasi-linear QMV-algebra and viceversa.

Let  $\mathcal{B} = \langle B, \boxplus, \mathbf{1}, \mathbf{0} \rangle$  be an effect algebra. The operation  $\boxplus$  can be extended to a total operation

$$\overline{\boxplus} : B \times B \rightarrow B$$

in the following way:

$$a \overline{\boxplus} b := \begin{cases} a \boxplus b, & \text{if } \exists(a \boxplus b); \\ \mathbf{1}, & \text{otherwise.} \end{cases}$$

The resulting structure  $\langle B, \overline{\boxplus}, ', \mathbf{1}, \mathbf{0} \rangle$  will be denoted by  $\mathcal{B}^{qmv}$ .

Viceversa, let  $\mathcal{M} = \langle M, \oplus, *, \mathbf{1}, \mathbf{0} \rangle$  be a QMV-algebra. Then, one can define a partial operation  $\overline{\oplus}$  on  $M$  such that

$$\begin{aligned} \text{Dom}(\overline{\oplus}) &:= \{ \langle a, b \rangle \in M \times M \mid a \preceq b^* \}. \\ \exists(a \overline{\oplus} b) &\curvearrowright a \overline{\oplus} b = a \oplus b. \end{aligned}$$

The resulting structure  $\langle M, \overline{\oplus}, \mathbf{1}, \mathbf{0} \rangle$  will be denoted by  $\mathcal{M}^{ea}$ .

**THEOREM 123.** [Gudder, 1995; Giuntini, 1995] *Let  $\mathcal{B} = \langle B, \boxplus, \mathbf{1}, \mathbf{0} \rangle$  be an effect algebra and let  $\mathcal{M} = \langle M, \oplus, *, \mathbf{1}, \mathbf{0} \rangle$  be a QMV-algebra.*

- (i)  $\mathcal{B}^{qmv}$  is a quasi-linear QMV-algebra;
- (ii)  $\mathcal{M}^{ea}$  is an effect algebra;
- (iii)  $(\mathcal{B}^{qmv})^{ea} = \mathcal{B}$ ;
- (iv)  $\mathcal{M}$  is quasi-linear iff  $(\mathcal{M}^{ea})^{qmv} = \mathcal{M}$ ;
- (v)  $\mathcal{B}^{qmv}$  is the unique quasi-linear QMV-algebra such that  $\overline{\boxplus}$  extends  $\boxplus$  and  $a \preceq b$  in  $\mathcal{B}^{qmv}$  implies  $a \sqsubseteq b$  in  $\mathcal{B}$ .

As a consequence, the effect algebra  $\mathcal{E}(\mathcal{H})$  of all effects on a Hilbert space  $\mathcal{H}$  determines a quasi-linear QMV-algebra  $\mathcal{E}(\mathcal{H})^{qmv} = \langle \mathcal{E}(\mathcal{H}), \oplus, *, \mathbf{1}, \mathbf{0} \rangle$ , where

$$E \oplus F = \begin{cases} E + F, & \text{if } \exists(E \boxplus F); \\ \mathbf{1}, & \text{otherwise,} \end{cases}$$

and

$$E^* = \mathbf{1} - E = E'.$$

These different ways of inducing a structure on the set of all unsharp physical properties have suggested different logical abstractions. In the following sections, we will investigate some interesting examples of unsharp quantum logics.

### 13 PARACONSISTENT QUANTUM LOGIC

Paraconsistent quantum logic (**PQL**) represents the most obvious unsharp weakening of orthologic. In the algebraic semantics, this logic is characterized by the class of all realizations based on an involutive bounded lattice, where the non contradiction principle ( $a \sqcap a' = \mathbf{0}$ ) is possibly violated.

In the Kripkean semantics, instead, **PQL** is characterized by the class of all realizations  $\langle I, R, \Pi, \rho \rangle$ , where the accessibility relation  $R$  is symmetric (but not necessarily reflexive), while  $\Pi$  behaves like in the **OL** - case. Any pair  $\langle I, R \rangle$ , where  $R$  is a symmetric relation on  $I$ , will be called *symmetric frame*. Differently from **OL** and **OQL**, a world  $i$  of a **PQL** realization may verify a contradiction. Since  $R$  is generally not reflexive, it may happen that  $i \in \rho(\alpha)$  and  $i \perp \rho(\alpha)$ . Hence:  $i \models \alpha \wedge \neg\alpha$ .

All the other semantic definitions are given like in the case of **OL**, *mutatis mutandis*. On this basis, one can show that our algebraic and Kripkean semantics characterize the same logic.

An axiomatization of **PQL** can be obtained by dropping the *absurdity rule* and the *Duns Scotus rule* in the **OL** calculus. Similarly to **OL**, our logic **PQL** satisfies the finite model property and is consequently decidable.

Hilbert-space realizations for **PQL** can be constructed, in a natural way, both in the algebraic and in the Kripkean semantics. In the algebraic semantics, take the realizations based on the Mac Neille completion of an involutive bounded poset having the form

$$\langle E(\mathcal{H}), \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle,$$

where  $\mathcal{H}$  is any Hilbert space. In the Kripkean semantics, consider the realizations based on the following frames

$$\langle E(\mathcal{H}) - \{\mathbf{0}\}, \not\perp \rangle,$$

where  $\not\perp$  represents the non orthogonality relation between effects ( $E \not\perp F$  iff  $E \not\sqsubseteq F'$ ). Differently from the projection case, here the accessibility relation is symmetric but generally non-reflexive. For instance, the semi-transparent effect  $\frac{1}{2}\mathbb{I}$  (representing the prototypical ambiguous property) is a fixed point of the generalized complement  $'$ ; hence  $\frac{1}{2}\mathbb{I} \perp \frac{1}{2}\mathbb{I}$  and  $(\frac{1}{2}\mathbb{I})' \perp (\frac{1}{2}\mathbb{I})'$ . From the physical point of view, possible worlds are here identified with possible pieces of information about the physical system under investigation. Any information may be either maximal (a pure state) or non maximal (a mixed state); either sharp (a projection) or unsharp (a proper effect). Violations of the non contradiction principle are determined by unsharp (ambiguous) pieces of knowledge. Interestingly enough, proper mixed states (which cannot be represented as projections) turn out to coincide with particular effects. In other words, within the unsharp approach, it is possible to represent both states and events by a unique kind of mathematical object, an effect.

**PQL** represents a somewhat rough logical abstraction from the class of all effect-realizations. An important condition that holds in all effect realizations is represented by the *regularity property* (which may fail in a generic **PQL**-realization).

DEFINITION 124. An algebraic **PQL** realization  $\langle B, v \rangle$  is called *regular* iff the involutive bounded lattice  $\mathcal{B}$  is regular ( $a \sqcap a' \sqsubseteq b \sqcup b'$ ).

The regularity property can be naturally formulated also in the framework of the Kripkean semantics:

DEFINITION 125. A **PQL** Kripkean realization  $\langle I, R, \Pi, \rho \rangle$  is *regular* iff its frame  $\langle I, R \rangle$  is *regular*. In other words,  $\forall i, j \in I: i \perp i$  and  $j \perp j \curvearrowright i \perp j$ .

One can prove that a symmetric frame  $\langle I, R \rangle$  is regular iff the involutive bounded lattice of all propositions of  $\langle I, R \rangle$  is regular. As a consequence, an algebraic realization is regular iff its Kripkean transformation is regular and viceversa (where the Kripkean [algebraic] transformation of an algebraic [Kripkean] realization is defined like in **OL**).

On this basis one can introduce a proper extension of **PQL**: *regular paraconsistent quantum logic* (**RPQL**). Semantically **RPQL** is characterized by the class of all regular realizations (both in the algebraic and in the Kripkean semantics). The calculus for **RPQL** is obtained by adding to the **PQL**-calculus the following rule:

$$\alpha \wedge \neg\alpha \vdash \beta \vee \neg\beta \quad (\text{Kleene rule})$$

A completeness theorem for both **PQL** and **RPQL** can be proved, similarly to the case of **OL**. Both logics **PQL** and **RPQL** admit a natural modal translation (similarly to **OL**). The suitable modal system which **PQL** can be transformed into is the system **KB**, semantically characterized by the class of all symmetric frames. A convenient strengthening of **KB** gives rise to a regular modal system, that is suitable for **RPQL**.

An interesting question concerns the relation between **PQL** and the orthomodular property.

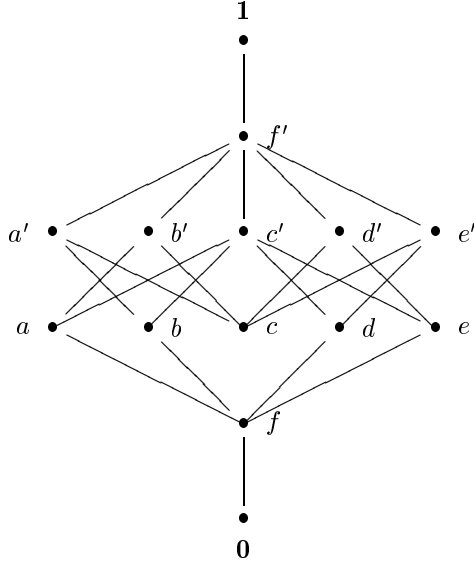
Let  $\mathcal{B} = \langle A, \sqsubseteq, ', \mathbf{1}, \mathbf{0} \rangle$  be an ortholattice. By Lemma 19 the following three conditions (expressing possible definitions of the orthomodular property) turn out to be equivalent:

- (i)  $\forall a, b \in B: a \sqsubseteq b \curvearrowright b = a \sqcup (a' \sqcap b)$ ;
- (ii)  $\forall a, b \in B: a \sqsubseteq b$  and  $a' \sqcap b = \mathbf{0} \curvearrowright a = b$ ;
- (iii)  $\forall a, b \in B: a \sqcap (a' \sqcup (a \sqcap b)) \sqsubseteq b$ .

However, this equivalence breaks down in the case of involutive bounded lattices. One can prove only:

LEMMA 126. *Let  $\mathcal{B}$  be an involutive bounded lattice. If  $\mathcal{B}$  satisfies condition (i), then  $\mathcal{B}$  satisfies conditions (ii) and (iii).*

**Proof.** (i) implies (ii): trivial. Suppose (i); we want to show that (iii) holds. Now,  $a' \sqsubseteq a' \sqcup b' = (a \sqcap b)'$ . Therefore, by (i),  $(a \sqcap b)' = a' \sqcup (a \sqcap (a \sqcap b))'$ . By de Morgan law:  $a \sqcap b = (a \sqcap (a' \sqcup (a \sqcap b))) \sqsubseteq b$ . ■

Figure 6.  $\mathcal{G}_{14}$ 

LEMMA 127. *Any involutive bounded lattice  $\mathcal{B}$  that satisfies condition (iii) is an ortholattice.*

**Proof.** Suppose (iii). It is sufficient to prove that  $\forall a, b \in B: a \sqcap a' \sqsubseteq b$ . Now,  $a \sqcap a' \sqsubseteq a, a'$ . Moreover,  $a' \sqsubseteq a' \sqcup (a \sqcap b)$ . Therefore, by (iii),  $a \sqcap a' \sqsubseteq a \sqcap (a' \sqcup (a \sqcap b)) \sqsubseteq b$ . Thus,  $\forall a \in B: a \sqcap a' = \mathbf{0}$ . ■

As a consequence, we can conclude that there exists no proper orthomodular paraconsistent quantum logic when orthomodularity is understood in the sense (i) or (iii). A residual possibility for a proper paraconsistent quantum logic to be orthomodular is orthomodularity in the sense (ii). In fact, the lattice  $\mathcal{G}_{14}$  (see Figure 6) is an involutive bounded lattice which turns out to be orthomodular (ii) but not orthomodular (i).

Since  $f \sqcap f' = f \neq \mathbf{0}$ ,  $\mathcal{G}_{14}$  cannot be an ortholattice. Hence,  $\mathcal{G}_{14}$  is neither orthomodular (i) nor orthomodular (iii). However,  $\mathcal{G}_{14}$  is trivially orthomodular (ii) since the premiss of condition (ii) is satisfied only in the trivial case where both  $a, b$  are either  $\mathbf{0}$  or  $\mathbf{1}$ .

Hilbert space realizations for orthomodular paraconsistent quantum logic can be constructed in the algebraic semantics by taking as support the following proper subset of the set of all effects:

$$I(\mathcal{H}) := \{a\mathbb{I} \mid a \in [0, 1]\} \cup P(\mathcal{H}).$$



In other words, a possible meaning of the formula is either a sharp property (projection) or an unsharp property that can be represented as a multiple of the universal property ( $\mathbb{I}$ ).

The set  $I(\mathcal{H})$  determines an orthomodular involutive regular bounded lattice, where the partial order is the partial order of  $\mathcal{E}(\mathcal{H})$  restricted to  $I(\mathcal{H})$ , while the fuzzy complement is defined like in the class of all effects ( $E' := \mathbb{I} - E$ ).

An interesting feature of **PQL** is represented by the fact that this logic turns out to be a common sublogic in a wide class of important logics. In particular, **PQL** is a sublogic of Girard's linear logic [Girard, 1987], of Łukasiewicz' infinite many-valued logic and of some relevant logics.

As we will see in Section 17, **PQL** represents the most natural quantum logical extension of a quite weak and general logic, that has been called *basic logic*.

## 14 THE BROUWER–ZADEH LOGICS

The *Brouwer–Zadeh logics* (called also *fuzzy intuitionistic logics*) represent natural abstractions from the class of all **BZ**-lattices (defined in Section 12). As a consequence, a characteristic property of these logics is a splitting of the connective “not” into two forms of negation: a *fuzzy-like* negation, that gives rise to a *paraconsistent* behaviour and an *intuitionistic-like* negation. The fuzzy “not” represents a weak negation, that inverts the two extreme truth-values (truth and falsity), satisfies the double negation principle but generally violates the non-contradiction principle. The second “not” is a stronger negation, a kind of necessitation of the fuzzy “not”.

We will consider two forms of Brouwer–Zadeh logic: **BZL** (*weak Brouwer–Zadeh logic*) and **BZL**<sup>3</sup> (*strong Brouwer–Zadeh logic*). The language of both **BZL** and **BZL**<sup>3</sup> is an extension of the language of **QL**. The primitive connectives are: the *conjunction* ( $\wedge$ ), the *fuzzy negation* ( $\neg$ ), the *intuitionistic negation* ( $\sim$ ).

*Disjunction* is metatheoretically defined in terms of conjunction and of the fuzzy negation:

$$\alpha \vee \beta := \neg(\neg\alpha \wedge \neg\beta).$$

A *necessity* operator is defined in terms of the intuitionistic and of the fuzzy negation:

$$L\alpha := \sim \neg\alpha.$$

A *possibility* operator is defined in terms of the necessity operator and of the fuzzy negation:

$$M\alpha := \neg L\neg\alpha.$$

Let us first consider our weaker logic **BZL**. Similarly to **OL** and **PQL**, also **BZL** can be characterized by an algebraic and a Kripkean semantics.

DEFINITION 128 (Algebraic realization for **BZL**). An *algebraic realization* of **BZL** is a pair  $\langle \mathcal{B}, v \rangle$ , consisting of a BZ-lattice  $\langle B, \sqsubseteq, ', \sim, \mathbf{1}, \mathbf{0} \rangle$  and a valuation-function  $v$  that associates to any formula  $\alpha$  an element in  $B$ , satisfying the following conditions:

- (i)  $v(\neg\beta) = v(\beta)'$
- (ii)  $v(\sim\beta) = v(\beta)\sim$
- (iii)  $v(\beta \wedge \gamma) = v(\beta) \sqcap v(\gamma)$ .

The definitions of truth, consequence in an algebraic realization for **BZL**, logical truth and logical consequence are given similarly to the case of **OL**.

A Kripkean semantics for **BZL** has been first proposed in [Giuntini, 1991]. A characteristic of this semantics is the use of frames with two accessibility relations.

DEFINITION 129. A *Kripkean realization* for **BZL** is a system

$\mathcal{K} = \langle I, \not\sim, \not\sim, \Pi, \rho \rangle$  where:

- (i)  $\langle I, \not\sim, \not\sim \rangle$  is a frame with a non empty set  $I$  of possible worlds and two accessibility relations:  $\not\sim$  (the *fuzzy accessibility* relation) and  $\not\sim$  (the *intuitionistic accessibility* relation).

Two worlds  $i, j$  are called *fuzzy-accessible* iff  $i \not\sim j$ . They are called *intuitionistically-accessible* iff  $i \not\sim j$ . Instead of *not*( $i \not\sim j$ ) and *not*( $i \not\sim j$ ), we will write  $i \perp j$  and  $i \not\sim j$ , respectively.

The following conditions are required for the two accessibility relations:

- (ia)  $\langle I, \not\sim \rangle$  is a regular symmetric frame;
- (ib) any world is fuzzy-accessible to at least one world:

$$\forall i \exists j : i \not\sim j.$$

- (ic)  $\langle I, \not\sim \rangle$  is an orthoframe;
- (id) Fuzzy accessibility implies intuitionistic accessibility:

$$i \not\sim j \rightsquigarrow i \not\sim j.$$

- (ie) Any world  $i$  has a kind of “twin-world”  $j$  such that for any world  $k$ :

- (a)  $i \not\sim k$  iff  $j \not\sim k$
- (b)  $i \not\sim k \rightsquigarrow j \not\sim k$ .

For any set  $X$  of worlds, the *fuzzy-orthogonal* set  $X'$  is defined like in **OL**:

$$X' = \{i \in I \mid \forall j \in X : i \perp j\}.$$

Similarly, the *intuitionistic orthogonal* set  $X^\sim$  is defined as follows:

$$X^\sim = \{i \in I \mid \forall j \in X : i \not\leq j\}.$$

The notion of *proposition* is defined like in **OL**. It turns out that a set of worlds  $X$  is a proposition iff  $X = X''$ .

One can prove that for any set of worlds  $X$ , both  $X'$  and  $X^\sim$  are propositions. Further, like in **OL**,  $X \sqcap Y$  (the greatest proposition included in the propositions  $X$  and  $Y$ ) is  $X \cap Y$ , while  $X \sqcup Y$  (the smallest proposition including  $X$  and  $Y$ ) is  $(X \cup Y)''$ .

- (ii)  $\Pi$  is a set of propositions that contains  $I$ , and is closed under  $'$ ,  $\sim$ ,  $\sqcap$ .
- (iii)  $\rho$  associates to any formula a proposition in  $\Pi$  according to the following conditions:

$$\begin{aligned} \rho(\neg\beta) &= \rho(\beta)'; \\ \rho(\sim\beta) &= \rho(\beta)^\sim; \\ \rho(\beta \wedge \gamma) &= \rho(\beta) \sqcap \rho(\gamma). \end{aligned}$$

**THEOREM 130.** *Let  $\langle I, \leq, \not\leq \rangle$  be a BZ-frame (i.e. a frame satisfying the conditions of Definition 129(i)) and let  $\Pi^0$  be the set of all propositions of the frame. Then, the structure  $\langle \Pi^0, \subseteq, ', \sim, \emptyset, I \rangle$  is a complete BZ-lattice such that for any set  $\Gamma \subseteq \Pi^0$ :*

$$\text{inf}(\Gamma) := \sqcap \Gamma = \bigcap \Gamma \quad \text{and} \quad \text{sup}(\Gamma) := \sqcup \Gamma = \left( \bigcup \Gamma \right)''.$$

As a consequence, the proposition-structure  $\langle \Pi, \subseteq, ', \sim, \emptyset, I \rangle$  of a **BZL** realization, turns out to be a BZ-lattice.

The definitions of truth, consequence in a Kripkean realization, logical truth and logical consequence, are given similarly to the case of **OL**.

One can prove, with standard techniques, that the algebraic and the Kripkean semantics for **BZL** characterize the same logic.

We will now introduce a calculus that represents an adequate axiomatization for the logic **BZL**. The most intuitive way to formulate our calculus is to present it as a modal extension of the axiomatic version of regular paraconsistent quantum logic **RPQL**. (Recall that the modal operators of **BZL** are defined as follows:  $L\alpha := \sim \neg \alpha$ ;  $M\alpha := \neg L\neg \alpha$ ).

### Rules of **BZL**.

The BZL-calculus includes, besides the rules of **RPQL** the following modal rules:

$$(BZ1) \quad L\alpha \vdash \alpha$$

$$(BZ2) \quad L\alpha \vdash LL\alpha$$

$$(BZ3) \quad ML\alpha \vdash L\alpha$$

$$(BZ4) \quad \frac{\alpha \vdash \beta}{L\alpha \vdash L\beta}$$

$$(BZ5) \quad L\alpha \wedge L\beta \vdash L(\alpha \wedge \beta)$$

$$(BZ6) \quad \emptyset \vdash \neg(L\alpha \wedge \neg L\alpha)$$

The rules (BZ1)-(BZ5) give rise to a  $\mathbf{S}_5$ -like modal behaviour. The rule (BZ6) (the non-contradiction principle for necessitated formulas) is, of course, trivial in any classical modal system.

One can prove a soundness and completeness Theorem with respect to the Kripkean semantics (by an appropriate modification of the corresponding proofs for  $\mathbf{QL}$ ).

Characteristic logical properties of  $\mathbf{BZL}$  are the following:

- (a) like in  $\mathbf{PQL}$ , the distributive principles, Duns Scotus, the non-contradiction and the excluded middle principles (for the fuzzy negation) break down;
- (b) like in intuitionistic logic, we have:

$$\begin{aligned} \models_{\mathbf{BZL}} \sim(\alpha \wedge \sim \alpha); \quad \not\models_{\mathbf{BZL}} \alpha \vee \sim \alpha; \quad \alpha \models_{\mathbf{BZL}} \sim \sim \alpha; \quad \sim \sim \alpha \not\models_{\mathbf{BZL}} \alpha; \\ \sim \sim \sim \alpha \models_{\mathbf{BZL}} \sim \alpha; \quad \alpha \models_{\mathbf{BZL}} \beta \wedge \sim \beta \models_{\mathbf{BZL}} \sim \alpha; \end{aligned}$$

- (c) moreover, we have:

$$\sim \alpha \models_{\mathbf{BZL}} \neg \alpha; \quad \neg \alpha \not\models_{\mathbf{BZL}} \sim \alpha; \quad \neg \sim \alpha \models_{\mathbf{BZL}} \sim \sim \alpha;$$

One can prove that  $\mathbf{BZL}$  has the finite model property; as a consequence it is decidable [Giuntini, 1992].

### The ortho-pair semantics

Our stronger logic  $\mathbf{BZL}^3$  has been suggested by a form of fuzzy-intuitionistic semantics, that has been first studied in [Cattaneo and Nisticò, 1986]. The intuitive idea, underlying this semantics (which has some features in common with Klaua's *partielle Mengen* and with Dunn's *polarities*) can be sketched as follows: one supposes that interpreting a language means associating to any sentence two *domains of certainty*: the domain of the situations

where our sentence certainly holds, and the domain of the situations where our sentence certainly does not hold. Similarly to Kripkean semantics, the situations we are referring to can be thought of as a kind of possible worlds. However, differently from the standard Kripkean behaviour, the positive domain of a given sentence does not generally determine the negative domain of the same sentence. As a consequence, propositions are here identified with particular *pairs* of sets of worlds, rather than with particular sets of worlds.

Let us again assume the **BZL** language. We will define the notion of *realization with positive and negative certainty domains* (shortly *ortho-pair realization*) for a **BZL** language.

**DEFINITION 131.** An *ortho-pair realization* is a system  $\mathcal{O} = \langle I, R, \Omega, v \rangle$ , where:

- (i)  $\langle I, R \rangle$  is an orthoframe.
- (ii) Let  $\Pi^0$  be the set of all propositions of the orthoframe  $\langle I, R \rangle$ . As we already know, this set gives rise to an ortholattice with respect to the operations  $\sqcap, \sqcup$  and  $'$  (where  $\sqcap$  is the set-theoretic intersection).

An *orthopairproposition* of  $\langle I, R \rangle$  is any pair  $\langle A_1, A_0 \rangle$ , where  $A_1, A_0$  are propositions in  $\Pi^0$  such that  $A_1 \subseteq A'_0$ . An orthopairproposition  $\langle A_1, A_0 \rangle$  is called *exact* iff  $A_0 = A'_1$  (in other words,  $A_0$  is maximal). The following operations and relations can be defined on the set of all orthopairpropositions:

- (iia) The fuzzy complement:

$$\langle A_1, A_0 \rangle^{\odot} := \langle A_0, A_1 \rangle .$$

- (iib) The intuitionistic complement:

$$\langle A_1, A_0 \rangle^{\ominus} := \langle A_0, A'_0 \rangle .$$

- (iic) The orthopairpropositional conjunction:

$$\langle A_1, A_0 \rangle \overline{\sqcap} \langle B_1, B_0 \rangle := \langle A_1 \sqcap B_1, A_0 \sqcup B_0 \rangle .$$

- (iid) The orthopairpropositional disjunction:

$$\langle A_1, A_0 \rangle \underline{\sqcup} \langle B_1, B_0 \rangle := \langle A_1 \sqcup B_1, A_0 \sqcap B_0 \rangle .$$

- (iie) The infinitary conjunction:

$$\overline{\sqcap}_n \{ \langle A_1^n, A_0^n \rangle \} := \left\langle \bigcap_n \{ A_1^n \}, \underline{\sqcup}_n \{ A_0^n \} \right\rangle .$$

(iif) The infinitary disjunction:

$$\bigsqcup_n \{\langle A_1^n, A_0^n \rangle\} := \left\langle \bigsqcup_n \{A_1^n\}, \bigcap_n \{A_0^n\} \right\rangle.$$

(iig) The necessity operator:

$$\Box(\langle A_1, A_0 \rangle) := \langle A_1, A_1' \rangle.$$

(iih) The possibility operator:

$$\Diamond(\langle A_1, A_0 \rangle) := (\Box(\langle A_1, A_0 \rangle^{\odot}))^{\odot}.$$

(iik) The order-relation:

$$\langle A_1, A_0 \rangle \sqsubseteq \langle B_1, B_0 \rangle \text{ iff } A_1 \subseteq B_1 \text{ and } B_0 \subseteq A_0.$$

(iii)  $\Omega$  is a set of orthopairpropositions, that is closed under  $\odot, \ominus, \overline{\phantom{x}}, \underline{\phantom{x}}$  and  $\mathbf{0} := \langle \emptyset, I \rangle$ .

(iv)  $v$  is a valuation-function that maps formulas into orthopairpropositions according to the following conditions:

$$\begin{aligned} v(\neg\beta) &= v(\beta)^{\odot}; \\ v(\sim\beta) &= v(\beta)^{\ominus}; \\ v(\beta \wedge \gamma) &= v(\beta)\overline{\phantom{x}}v(\gamma). \end{aligned}$$

The other basic semantic definitions are given like in the algebraic semantics. One can prove the following Theorem:

**THEOREM 132.** *Let  $\langle I, R \rangle$  be an orthoframe and let  $\Omega^0$  be the set of all orthopairpropositions of  $\langle I, R \rangle$ . Then, the structure  $\langle \Omega^0, \subseteq, \odot, \ominus, \langle \emptyset, I \rangle, \langle I, \emptyset \rangle \rangle$  is a complete BZ-lattice with respect to the infinitary conjunction and disjunction defined above. Further, the following conditions are satisfied: for any  $\langle A_1, A_0 \rangle, \langle B_1, B_0 \rangle \in \Omega^0$ :*

- (i)  $\Box \langle A_1, A_0 \rangle = \langle A_1, A_0 \rangle^{\odot\infty}$ .
- (ii)  $\langle A_1, A_0 \rangle^{\ominus} = \Box(\langle A_1, A_0 \rangle^{\odot})$ .
- (iii)  $\Diamond \langle A_1, A_0 \rangle = \langle A_1, A_0 \rangle^{\ominus\odot}$ .
- (iv)  $(\langle A_1, A_0 \rangle \overline{\phantom{x}} \langle B_1, B_0 \rangle)^{\ominus} = \langle A_1, A_0 \rangle^{\ominus} \overline{\phantom{x}} \langle B_1, B_0 \rangle^{\ominus}$ .  
(Strong de Morgan law)
- (v)  $(\langle A_1, A_0 \rangle \overline{\phantom{x}} \langle B_1, B_0 \rangle^{\ominus\ominus}) \subseteq (\langle A_1, A_0 \rangle^{\odot\odot} \overline{\phantom{x}} \langle B_1, B_0 \rangle)$ .

Accordingly, in any ortho-pair realization the set of all orthopairpropositions  $\Omega^0$  gives rise to a BZ-lattice. As a consequence, one can immediately

prove a soundness theorem with respect to the ortho-pair semantics. Does perhaps the ortho-pair semantics characterizes the logic **BZL**? The answer to this question is negative. As a counterexample, let us consider an instance of the fuzzy excluded middle and an instance of the intuitionistic excluded middle applied to the same formula  $\alpha$ :

$$\alpha \vee \neg\alpha \quad \text{and} \quad \alpha \vee \sim \alpha.$$

One can easily check that they are logically equivalent in the ortho-pair semantics. For, given any ortho-pair realization  $\mathcal{O}$ , there holds:

$$\alpha \vee \neg\alpha \models_{\mathcal{O}} \alpha \vee \sim \alpha \quad \text{and} \quad \alpha \vee \sim \alpha \models_{\mathcal{O}} \alpha \vee \neg\alpha.$$

However, generally

$$\alpha \vee \neg\alpha \not\models_{\mathbf{BZL}} \alpha \vee \sim \alpha.$$

For instance, let us consider the following algebraic **BZL**-realization  $\mathcal{A} = \langle \mathcal{B}, v \rangle$ , where the support  $\mathcal{B}$  of is the real interval  $[0, 1]$  and the algebraic structure on  $\mathcal{B}$  is defined as follows:

$$\begin{aligned} a \sqsubseteq b & \text{ iff } a \leq b; \\ a' & = 1 - a; \\ a^\sim & = \begin{cases} 1, & \text{if } a = 0; \\ 0, & \text{otherwise.} \end{cases} \\ \mathbf{1} & = 1; \quad \mathbf{0} = 0. \end{aligned}$$

Suppose for a given sentential literal  $p$ :  $0 < v(p) < 1/2$ . We will have  $v(p \vee \sim p) = \text{Max}(v(p), 0) = v(p) < 1/2$ . But  $v(p \vee \neg p) = \text{Max}(v(p), 1 - v(p)) = 1 - v(p) \geq 1/2$ . Hence:  $v(p \vee \sim p) < v(p \vee \neg p)$ .

As a consequence, the orthopair-semantics characterizes a logic stronger than **BZL**. We will call this logic **BZL**<sup>3</sup>. The name is due to the characteristic three-valued features of the ortho-pair semantics.

Our logic **BZL**<sup>3</sup> is axiomatizable. A suitable calculus can be obtained by adding to the **BZL**-calculus the following rules.

*Rules of **BZL**<sup>3</sup>.*

$$(BZ^31) \quad L(\alpha \vee \beta) \vdash L\alpha \vee L\beta$$

$$(BZ^32) \quad \frac{L\alpha \vdash \beta, \alpha \vdash M\beta}{\alpha \vdash \beta}$$

The following rules turn out to be derivable:

$$(DR1) \quad \frac{L\alpha \vdash \beta, M\alpha \vdash M\beta}{\alpha \vdash \beta}$$

$$(DR2) \quad M\alpha \wedge M\beta \vdash M(\alpha \wedge \beta)$$

$$(DR3) \quad \sim(\alpha \wedge \beta) \vdash \sim\alpha \vee \sim\beta$$

The validity of a strong de Morgan's principle for the connective  $\sim$  (DR3) shows that this connective represents, in this logic, a kind of strong “super-intuitionistic” negation (differently from **BZL**, where the strong de Morgan law fails, like in intuitionistic logic).

One can prove a soundness and a completeness theorem of our calculus with respect to the ortho-pair semantics.

THEOREM 133 (Soundness theorem).

$$T \vdash_{\mathbf{BZL}^3} \alpha \quad \curvearrowright \quad T \models_{\mathbf{BZL}^3} \alpha.$$

**Proof.** By routine techniques. ■

THEOREM 134. *Completeness theorem.*

$$T \models_{\mathbf{BZL}^3} \alpha \quad \curvearrowright \quad T \vdash_{\mathbf{BZL}^3} \alpha.$$

### Sketch of the proof

Instead of  $T \models_{\mathbf{BZL}^3} \alpha$  and  $T \vdash_{\mathbf{BZL}^3} \alpha$ , we will shortly write  $T \models \alpha$  and  $T \vdash \alpha$ . It is sufficient to construct a canonical model  $\mathcal{O} = \langle I, R, \Omega, v \rangle$  such that:

$$T \models_{\mathcal{O}} \alpha \quad \curvearrowright \quad T \vdash \alpha.$$

(The other way around follows from the soundness theorem).

#### *Definition of the canonical model*

(i)  $I$  is the set of all possible sets  $i$  of formulas satisfying the following conditions:

- (ia)  $i$  is *non contradictory* with respect to the fuzzy negation  $\neg$ : for any  $\alpha$ , if  $\alpha \in i$ , then  $\neg\alpha \notin i$ ;
- (ib)  $i$  is *L-closed*: for any  $\alpha$ , if  $\alpha \in i$ , then  $L\alpha \in i$ ;
- (ic)  $i$  is *deductively closed*: for any  $\alpha$ , if  $i \vdash \alpha$ , then  $\alpha \in i$ .

(ii) The accessibility relation  $R$  is defined as follows:

$$Rij \text{ iff for any formula } \alpha: \alpha \in i \quad \curvearrowright \quad \neg\alpha \notin j.$$

(In other words,  $i$  and  $j$  are not contradictory with respect to the fuzzy negation).

Instead of *not Rij*, we will write  $i \perp j$ .

(iii)  $\Omega$  is the set of all orthopairpropositions of  $\langle I, R \rangle$ .



(iv) For any atomic formula  $p$ :

$$v(p) = \langle v_1(p), v_0(p) \rangle ,$$

where:

$$v_1(p) = \{i \mid i \vdash p\} \quad \text{and} \quad v_0(p) = \{i \mid i \vdash \neg p\} .$$

$\mathcal{O}$  is well defined since one can prove the following Lemmas:

LEMMA 135. *R is reflexive and symmetric.*

LEMMA 136. *For any  $\alpha, \{i \mid i \vdash \alpha\}$  is a proposition of the orthoframe  $\langle I, R \rangle$ .*

LEMMA 137. *For any  $\alpha, \{i \mid i \vdash \alpha\} \subseteq \{i \mid i \vdash \neg \alpha\}'$ .*

Further, one can prove

LEMMA 138. *For any  $\alpha, v(\alpha) = \langle v_1(\alpha), v_0(\alpha) \rangle$ , where:*

$$\begin{aligned} v_1(\alpha) &= \{i \mid i \vdash \alpha\} \\ v_0(\alpha) &= \{i \mid i \vdash \neg \alpha\} \end{aligned}$$

LEMMA 139. *For any formula  $\alpha$ :*

$$\mathbf{0} := \langle \emptyset, I \rangle = \langle \{i \mid i \vdash L\alpha \wedge \neg L\alpha\}, \{i \mid i \vdash \neg(L\alpha \wedge \neg L\alpha)\} \rangle .$$

LEMMA 140. *Let  $T = \{\alpha_1, \dots, \alpha_n, \dots\}$  be a set of formulas and let  $\alpha$  be any formula.*

$$\bigcap \{v_1(\alpha_n) \mid \alpha_n \in T\} \subseteq v_1(\alpha) \curvearrowright L\alpha_1, \dots, L\alpha_n, \dots \vdash \alpha .$$

As a consequence, one can prove:

LEMMA 141. *Lemma of the canonical model*

$$T \models_{\mathcal{O}} \alpha \curvearrowright T \vdash \alpha .$$

Suppose  $T \models_{\mathcal{O}} \alpha$ . Hence (by definition of consequence in a given realization): for any orthopairproposition  $\langle A_1, A_0 \rangle \in \Omega$ , if for all  $\alpha_n \in T$ ,  $\langle A_1, A_0 \rangle \sqsubseteq v(\alpha_n)$ , then  $\langle A_1, A_0 \rangle \sqsubseteq v(\alpha)$ .

The propositional lattice, consisting of all orthopairpropositions of  $\mathcal{O}$  is complete (see Theorem 132). Hence:  $\prod_n \{v(\alpha_n) \mid \alpha_n \in T\} \sqsubseteq v(\alpha)$ . In other words, by definition of  $\sqsubseteq$ :

- (i)  $\bigcap \{v_1(\alpha_n) \mid \alpha_n \in T\} \subseteq v_1(\alpha)$ ;
- (ii)  $v_0(\alpha) \subseteq \bigsqcup \{v_0(\alpha_n) \mid \alpha_n \in T\}$ .

Thus, by (i) and by Lemma 140:  $L\alpha_1, \dots, L\alpha_n, \dots \vdash \alpha$ . Consequently, there exists a finite subset  $\{\alpha_{n_1}, \dots, \alpha_{n_k}\}$  of  $T$  such that  $L\alpha_{n_1} \wedge \dots \wedge L\alpha_{n_k} \vdash \alpha$ . Hence, by the rules for  $\wedge$  and  $L$ :  $L(\alpha_{n_1} \wedge \dots \wedge \alpha_{n_k}) \vdash \alpha$ .

At the same time, we obtain from (ii) and by Lemma 138:  $v_1(\neg\alpha) \sqsubseteq \bigsqcup\{v_1(\neg\alpha_n) \mid \alpha_n \in T\}$ .

Whence, by de Morgan,

$$v_1(\neg\alpha) \subseteq \left[ \bigcap \{(v_1(\neg\alpha_n))' \mid \alpha_n \in T\} \right]' .$$

Now, one can easily check that in any realization:  $v_1(\neg\alpha)' = v_1(M\alpha)$ . As a consequence:  $v_1(\neg\alpha) \subseteq [\bigcap\{v_1(M\alpha_n) \mid \alpha_n \in T\}]'$ . Hence, by contraposition:

$$\bigcap\{v_1(M\alpha_n) \mid \alpha_n \in T\} \subseteq (v_1(\neg\alpha))'$$

and

$$\bigcap\{v_1(M\alpha_n) \mid \alpha_n \in T\} \subseteq v_1(M\alpha).$$

Consequently, by Lemma 140 and by the  $S_5$ -rules:

$$LM\alpha_1, \dots, LM\alpha_n, \dots \vdash M\alpha, \quad M\alpha_1, \dots, M\alpha_n, \dots \vdash M\alpha.$$

By syntactical compactness, there exists a finite subset  $\{\alpha_{m_1}, \dots, \alpha_{m_h}\}$  of  $T$  such that  $M\alpha_{m_1}, \dots, M\alpha_{m_h} \vdash M\alpha$ . Whence, by the rules for  $\wedge$  and  $M$ :  $M(\alpha_{m_1} \wedge \dots \wedge \alpha_{m_h}) \vdash M\alpha$ . Let us put  $\gamma_1 = \alpha_{n_1} \wedge \dots \wedge \alpha_{n_k}$  and  $\gamma_2 = \alpha_{m_1} \wedge \dots \wedge \alpha_{m_h}$ . We have obtained:  $L\gamma_1 \vdash \alpha$  and  $M\gamma_2 \vdash M\alpha$ . Whence,  $L\gamma_1 \wedge L\gamma_2 \vdash \alpha$ ,  $L(\gamma_1 \wedge \gamma_2) \vdash \alpha$ ,  $M\gamma_1 \wedge M\gamma_2 \vdash M\alpha$ ,  $M(\gamma_1 \wedge \gamma_2) \vdash M\alpha$ . From  $L(\gamma_1 \wedge \gamma_2) \vdash \alpha$ , and  $M(\gamma_1 \wedge \gamma_2) \vdash M\alpha$  we obtain, by the derivable rule (DR1):  $\gamma_1 \wedge \gamma_2 \vdash \alpha$ . Consequently:  $T \vdash \alpha$ .  $\blacksquare$

Similarly to other forms of quantum logic, also  $\mathbf{BZL}^3$  admits an algebraic semantic characterization [Giuntini, 1993] based on the notion of  $\mathbf{BZ}^3$ -lattice.

**DEFINITION 142.** A  $\mathbf{BZ}^3$ -lattice is a  $\mathbf{BZ}$ -lattice  $\mathcal{B} = \langle B, \sqsubseteq, ', \sim, \mathbf{1}, \mathbf{0} \rangle$ , which satisfies the following conditions:

- (i)  $(a \sqcap b) \sim = a \sim \sqcup b \sim$ ;
- (ii)  $a \sqcap b \sim \sim \sqsubseteq a' \sim \sqcup b$ .

By Theorem 132, the set of all orthopairpropositions of an orthoframe determines a complete  $\mathbf{BZ}^3$ -lattice. One can prove the following representation theorem:

**THEOREM 143.** *Every  $\mathbf{BZ}^3$ -lattice is embeddable into the (complete)  $\mathbf{BZ}^3$ -lattice of all orthopairpropositions of an orthoframe.*

A slight modification of the proof of Theorem 17 permits us to show that ortho-pair semantics and the algebraic semantics strongly characterize the same logic.

One can prove that  $\mathbf{BZL}^3$  can be also characterized by means of a non standard version of Kripkean semantics [Giuntini, 1993].

Some problems concerning the Brouwer-Zadeh logics remain still open:

- 1) Is there any Kripkean characterization of the logic that is algebraically characterized by the class of all de Morgan **BZ**-lattices (i.e. **BZ**-lattices satisfying condition (i) of Definition 142)? In this framework, the problem can be reformulated in this way: is the (strong) de Morgan law elementary?
- 2) Is it possible to axiomatize a logic based on an infinite many-valued generalization of the ortho-pair semantics?
- 3) Find possible conditional connectives in **BZL**<sup>3</sup>.
- 4) Find an appropriate orthomodular extension of **BZL**<sup>3</sup>.

**Unsharp quantum models for BZL<sup>3</sup>**

The ortho-pair semantics has been suggested by the effect-structures in Hilbert-space QT. In this framework, natural quantum ortho-pair realizations for **BZL**<sup>3</sup> can be constructed. Let us refer again to the language  $\mathcal{L}^Q$  (whose atoms express possible measurement reports) and let  $\mathcal{S}$  be a quantum system whose associated Hilbert space is  $\mathcal{H}$ . As usual,  $E(\mathcal{H})$  will represent the set of all effects of  $\mathcal{H}$ . Now, an ortho-pair realization  $\mathcal{M}^{\mathcal{S}} = \langle I, R, \Omega, v \rangle$  (for the system  $\mathcal{S}$ ) can be defined as follows:

- (i)  $I$  is the set of all pure states of  $\mathcal{S}$  in  $\mathcal{H}$ .
- (ii)  $Rij$  iff for any effect  $E \in E(\mathcal{H})$  the following condition holds: whenever  $i$  assigns to  $E$  probability 1, then  $j$  assigns to  $E$  a probability different from 0.  
In other words,  $i$  and  $j$  are accessible iff they cannot be strongly distinguished by any physical property represented by an effect.
- (iii) The propositions of the orthoframe  $\langle I, R \rangle$  are determined by the set of all closed subspaces of  $\mathcal{H}$  (sharp properties), like in **OQL**.
- (iv)  $\Omega$  is the set of all orthopairpropositions of  $\langle I, R \rangle$ . Any effect  $E$  can be transformed into an orthopairproposition  $f(E) := \langle X_1^E, X_0^E \rangle$  of  $\Omega$ , where:

$$X_1^E := \{i \mid i \text{ assigns to } E \text{ probability } 1\};$$

$$X_0^E := \{i \mid i \text{ assigns to } E \text{ probability } 0\}.$$

In other words,  $X_1^E, X_0^E$  represent the positive and the negative domain of  $E$ , respectively. The map  $f$  turns out to preserve the order relation and the two complements:

$$E \sqsubseteq F \text{ iff } f(E) \sqsubseteq f(F).$$

$$f(E') = f(E)^{\odot} = \langle X_1^E, X_0^E \rangle^{\odot} = \langle X_0^E, X_1^E \rangle.$$

$$f(E^{\sim}) = f(E)^{\ominus} = \langle X_1^E, X_0^E \rangle^{\ominus} = \langle X_0^E, (X_0^E)' \rangle.$$

- (v) The valuation-function  $v$  follows the intuitive physical meaning of the atomic sentences. Let  $p$  express the assertion “the value for the observable  $A$  lies in the sharp (or fuzzy) Borel set  $\Delta$ ” and let  $E^p$  be the effect that is associated to  $p$  in  $\mathcal{H}$ . We define  $v$  as follows:

$$v(p) = f(E^p) = \langle X_1^{E^p}, X_0^{E^p} \rangle.$$

It is worthwhile to notice that our map  $f$  is not injective: different effects will be transformed into one and the same orthopairproposition. As a consequence, moving from effects to orthopairpropositions clearly determines a loss of information. In fact, orthopairpropositions are only concerned with the two extreme probability value  $(0,1)$ , a situation that corresponds to a three-valued semantics.

## 15 PARTIAL QUANTUM LOGICS

In Section 12, we have considered examples of partial algebraic structures, where the basic operations are not always defined. How to give a semantic characterization for different forms of quantum logic, corresponding respectively to the class of all effect algebras, of all orthoalgebras and of all orthomodular posets? We will call these logics: *unsharp partial quantum logic* (**UPaQL**), *weak partial quantum logic* (**WPaQL**) and *strong partial quantum logic* (**SPaQL**).

Let us first consider the case of **UPaQL**, that represents the “logic of effect algebras” [Dalla Chiara and Giuntini, 1995].

The language of **UPaQL** consists of a denumerably infinite list of atomic sentences and of two primitive connectives: the *negation*  $\neg$  and the *exclusive disjunction*  $\Psi$  (aut).

The set of sentences is defined in the usual way. A *conjunction* is metalinguistically defined, via de Morgan law:

$$\alpha \wedge \beta := \neg(\neg\alpha \Psi \neg\beta).$$

The intuitive idea underlying our semantics for **UPaQL** is the following: disjunctions and conjunctions are always considered “legitimate” from a mere linguistic point of view. However, semantically, a disjunction  $\alpha \Psi \beta$  will have the intended meaning only in the “well behaved cases” (where the values of  $\alpha$  and  $\beta$  are orthogonal in the corresponding effect orthoalgebra). Otherwise,  $\alpha \Psi \beta$  will have any meaning whatsoever (generally not connected with the meanings of  $\alpha$  and  $\beta$ ). As is well known, a similar semantic “trick” is used in some classical treatments of the description operator  $\iota$  (“the unique individual satisfying a given property”; for instance, “the present king of Italy”).



$$(UPa4) \quad \neg\neg\alpha \vdash \alpha \quad (\text{strong double negation})$$

$$(UPa5) \quad \frac{\alpha \vdash \beta}{\neg\beta \vdash \neg\alpha} \quad (\text{contraposition})$$

$$(UPa6) \quad \beta \vdash \alpha \vee \neg\alpha \quad (\text{excluded middle})$$

$$(UPa7) \quad \frac{\alpha \vdash \neg\beta \quad \alpha \vee \neg\alpha \vdash \alpha \vee \beta}{\neg\alpha \vdash \beta} \quad (\text{unicity of negation})$$

$$(UPa8) \quad \frac{\alpha \vdash \neg\beta \quad \alpha \vdash \alpha_1 \quad \alpha_1 \vdash \alpha \quad \beta \vdash \beta_1 \quad \beta_1 \vdash \beta}{\alpha \vee \beta \vdash \alpha_1 \vee \beta_1} \quad (\text{weak substitutivity})$$

$$(UPa9) \quad \frac{\alpha \vdash \neg\beta}{\alpha \vee \beta \vdash \beta \vee \alpha} \quad (\text{weak commutativity})$$

$$(UPa10) \quad \frac{\beta \vdash \neg\gamma \quad \alpha \vdash \neg(\beta \vee \gamma)}{\alpha \vdash \neg\beta} \quad (\text{weak associativity})$$

$$(UPa11) \quad \frac{\beta \vdash \neg\gamma \quad \alpha \vdash \neg(\beta \vee \gamma)}{\alpha \vee \beta \vdash \neg\gamma} \quad (\text{weak associativity})$$

$$(UPa12) \quad \frac{\beta \vdash \neg\gamma \quad \alpha \vdash \neg(\beta \vee \gamma)}{\alpha \vee (\beta \vee \gamma) \vdash (\alpha \vee \beta) \vee \gamma} \quad (\text{weak associativity})$$

$$(UPa13) \quad \frac{\beta \vdash \neg\gamma \quad \alpha \vdash \neg(\beta \vee \gamma)}{(\alpha \vee \beta) \vee \gamma \vdash \alpha \vee (\beta \vee \gamma)} \quad (\text{weak associativity})$$

The concepts of *derivation* and of *derivability* are defined in the expected way. In order to axiomatize weak partial quantum logic (**WPaQL**) it is sufficient to add a rule, which corresponds to a *Duns Scotus-principle*:

$$(WPaQL) \quad \frac{\alpha \vdash \neg\alpha}{\alpha \vdash \beta} \quad (\text{Duns Scotus})$$

Clearly, the Duns Scotus-rule corresponds to the strong consistency condition in our definition of orthoalgebra (see Definition 117). In other words, differently from **UPaQL**, the logic **WPaQL** forbids paraconsistent situations.

Finally, an axiomatization of strong partial quantum logic (**SPaQL**) can be obtained, by adding the following rule to (UPa1)-(UPa13), (WPa):

$$(SPaQL) \quad \frac{\alpha \vdash \neg\beta \quad \alpha \vdash \gamma \quad \beta \vdash \gamma}{\alpha \vee \beta \vdash \gamma}$$

In other words, (SPaQL) requires that the disjunction  $\vee$  behaves like a supremum, whenever it has the “right meaning”.

Let **PaQL** represent any instance of our three calculi. We will use the following abbreviations. Instead of  $\alpha \vdash_{\text{PaQL}} \beta$  we will write  $\alpha \vdash \beta$ . When  $\alpha$  and  $\beta$  are logically equivalent ( $\alpha \vdash \beta$  and  $\beta \vdash \alpha$ ) we will write  $\alpha \equiv \beta$ .

Let  $p$  represent a particular sentential literal of the language: **T** will be an abbreviation for  $p \vee \neg p$ ; while **F** will be an abbreviation for  $\neg(p \vee \neg p)$ .

Some important derivable rules of all calculi are the following:

$$(D1) \quad \mathbf{F} \vdash \beta, \beta \vdash \mathbf{T} \quad (\text{Weak Duns Scotus})$$

$$(D2) \quad \frac{\alpha \vdash \neg\beta}{\alpha \vdash \alpha \vee \beta} \quad (\text{weak sup rule})$$

$$(D3) \quad \frac{\alpha \vdash \beta}{\beta \equiv \alpha \vee \neg(\alpha \vee \neg\beta)} \quad (\text{orthomodular-like rule})$$

$$(D4) \quad \frac{\alpha \vdash \neg\gamma \quad \beta \vdash \neg\gamma \quad \alpha \vee \gamma \equiv \beta \vee \gamma}{\alpha \equiv \beta} \quad (\text{cancellation})$$

As a consequence, the following syntactical lemma holds:

LEMMA 145. *For any  $\alpha, \beta$ :  $\alpha \vdash \beta$  iff there exists a formula  $\gamma$  such that*

- (i)  $\alpha \vdash \neg\gamma$ ;
- (ii)  $\beta \equiv \alpha \vee \gamma$ .

In other words, the logical implication behaves similarly to the partial order relation in the effect algebras.

The following derivable rule holds for **WPaQL** and for **SPaQL**:

$$(D5) \quad \frac{\alpha \vdash \neg\beta \quad \alpha \vdash \gamma \quad \beta \vdash \gamma \quad \gamma \vdash \alpha \vee \beta}{\alpha \vee \beta \vdash \gamma}$$

Our calculi turn out to be adequate with respect to the corresponding semantic characterizations. Soundness proofs are straightforward. Let us sketch the proof of the completeness theorem for our weakest calculus (**UP-aQL**).

**THEOREM 146.** *Completeness.*

$$\alpha \models \beta \quad \curvearrowright \quad \alpha \vdash \beta.$$

**Proof.** Following the usual procedure, it is sufficient to construct a canonical model  $\mathcal{A} = \langle \mathcal{B}, v \rangle$  such that for any formulas  $\alpha, \beta$ :

$$\alpha \vdash \beta \quad \curvearrowright \quad \alpha \models_{\mathcal{A}} \beta.$$

*Definition of the canonical model.*

- (i) The algebra  $\mathcal{B} = \langle B, \boxplus, \mathbf{1}, \mathbf{0} \rangle$  is determined as follows:
  - (ia)  $B$  is the class of all equivalence classes of logically equivalent formulas:  $B := \{[\alpha]_{\equiv} \mid \alpha \text{ is a formula}\}$ . (In the following, we will write  $[\alpha]$  instead of  $[\alpha]_{\equiv}$ ).
  - (ib)  $[\alpha] \boxplus [\beta]$  is defined iff  $\alpha \vdash \neg\beta$ . If defined,  $[\alpha] \boxplus [\beta] := [\alpha \vee \beta]$ .
  - (ic)  $\mathbf{1} := [\mathbf{T}]$ ;  $\mathbf{0} := [\mathbf{F}]$ .
- (ii) The valuation function  $v$  is defined as follows:  $v(\alpha) = [\alpha]$ .

One can easily check that  $\mathcal{A}$  is a “good” model for our logic. The operation  $\boxplus$  is well defined (by the transitivity, contraposition and weak substitutivity rules). Further,  $\mathcal{B}$  is an effect algebra:  $\boxplus$  is weakly commutative and weakly associative, because of the corresponding rules of our calculus. The strong excluded middle axiom holds by definition of  $\boxplus$  and in virtue of the following rules: excluded middle, unicity of negation, double negation. Finally, the weak consistency axiom holds by weak Duns Scotus (D1) and by definition of  $\boxplus$ .

**LEMMA 147.** *Lemma of the canonical model*

$$[\alpha] \sqsubseteq [\beta] \quad \text{iff} \quad \alpha \vdash \beta.$$

**Sketch of the proof.** By definition of  $\sqsubseteq$  (in any effect algebra) one has to prove:

$$\alpha \vdash \beta \quad \text{iff} \quad \text{for a given } \gamma \text{ such that } [\alpha] \perp [\gamma] : [\alpha] \boxplus [\gamma] = [\beta].$$



This equivalence holds by Lemma 145 and by definition of  $\boxplus$ .

Finally, let us check that  $v$  is a “good” valuation function. In other words:

- (i)  $v(\neg\beta) = v(\beta)'$
- (ii)  $v(\beta \vee \gamma) = v(\beta) \boxplus v(\gamma)$ , if  $v(\beta) \boxplus v(\gamma)$  is defined.

(i) By definition of  $v$ , we have to show that  $[\neg\beta]$  is the unique  $[\gamma]$  such that  $[\beta] \boxplus [\gamma] = \mathbf{1} := [\mathbf{T}]$ . In other words,

- (ia)  $[\mathbf{T}] \sqsubseteq [\beta] \boxplus [\neg\beta]$ .
- (ib)  $[\mathbf{T}] \sqsubseteq [\beta] \boxplus [\gamma] \iff \neg\beta \equiv \gamma$ .

This holds by definition of the canonical model, by definition of  $\boxplus$  and by the following rules: double negation, excluded middle, unicity of negation.

(ii) Suppose  $v(\beta) \boxplus v(\gamma)$  is defined. Then  $\beta \vdash \neg\gamma$ . Hence, by definition of  $\boxplus$  and of  $v$ :  $v(\beta) \boxplus v(\gamma) = [\beta] \boxplus [\gamma] = [\beta \vee \gamma] = v(\beta \vee \gamma)$ .

As a consequence, we obtain:

$$\alpha \vdash \beta \iff [\alpha] \sqsubseteq [\beta] \iff v(\alpha) \sqsubseteq v(\beta) \iff \alpha \models_{\mathcal{A}} \beta$$

■

The completeness argument can be easily transformed, *mutatis mutandis* for the case of weak and strong partial quantum logic.

## 16 LUKASIEWICZ QUANTUM LOGIC

As we have seen in Section 12, the class  $E(\mathcal{H})$  of all effects on a Hilbert space  $\mathcal{H}$  determines a quasi-linear QMV-algebra. The theory of QMV-algebras suggests, in a natural way, the semantic characterization of a new form of quantum logic (called *Lukasiewicz quantum logic* (**LQL**)), which generalizes both **OQL** and **L<sub>N</sub>**.

The language of **LQL** contains the same primitive connectives as **WPaQL** ( $\vee, \neg$ ). The conjunction ( $\wedge$ ) is defined via de Morgan law (like in **WPaQL**). Further, a new pair of conjunction ( $\boxtimes$ ) and disjunction ( $\boxvee$ ) connectives are defined as follows:

$$\alpha \boxtimes \beta := (\alpha \vee \neg\beta) \wedge \beta$$

$$\alpha \boxvee \beta := \neg(\neg\alpha \boxtimes \neg\beta)$$

**DEFINITION 148.** A *realization* for **LQL** is a pair  $\mathcal{A} = \langle \mathcal{M}, v \rangle$ , where

- (i)  $\mathcal{M} = \langle M, \oplus, *, \mathbf{1}, \mathbf{0} \rangle$  is a QMV-algebra.
- (ii)  $v$  (the valuation-function) associates to any formula  $\alpha$  an element of  $M$ , satisfying the following conditions:

$$v(\neg\beta) = v(\beta)^*.$$

$$v(\beta \forall \gamma) = v(\beta) \oplus v(\gamma).$$

The other semantic definitions (truth, consequence in a given realization, logical truth, logical consequence) are given like in the **QL**-case.

**LQL** can be easily axiomatized by means of a calculus that simply mimics the axioms of QMV-algebras.

The quasi-linearity property, which is satisfied by the QMV-algebras of effects, is highly non equational. Thus, the following question naturally arises: is **LQL** characterized by the class of all quasi-linear QMV-algebras (QLQMV)? In the case of  $\mathbf{L}_\mathbb{R}$ , Chang has proved that  $\mathbf{L}_\mathbb{R}$  is characterized by the MV-algebra determined by the real interval  $[0, 1]$ . This MV-algebra is clearly quasi-linear, being totally ordered.

The relation between **LQL** and QMV algebras turns out to be much more complicated. In fact one can show that **LQL** cannot be characterized even by the class of all *weakly linear* QMV-algebras (WLQMV). Since WLQMV is strictly contained in QLQMV, there follows that **LQL** is not characterized by QLQMV. To obtain these results, something stronger is proved. In particular, we can show that:

- the variety of all QMV-algebras ( $\mathcal{QMV}$ ) strictly includes the variety generated by the class of all weakly linear QMV-algebras ( $\mathbb{HSP}(\text{WLQMV})$ ).
- $\mathbb{HSP}(\text{WLQMV})$  strictly includes the variety generated by the class of all quasi-linear QMV-algebras ( $\mathbb{HSP}(\text{QLQMV})$ ).

So far, little is known about the axiomatizability of the logic based on  $\mathbb{HSP}(\text{QLQMV})$ . In the case of  $\mathbb{HSP}(\text{WLQMV})$ , instead, one can prove that this variety is generated by the QMV-axioms together with the following axiom:

$$a = (a \oplus c \odot b^*) \text{ \textcircled{ \& } } (a \oplus c^* \odot b).$$

The problem of the axiomatizability of the logic based on  $\mathbb{HSP}(\text{QLQMV})$  is complicated by the fact that not every (quasi-linear) QMV-algebra  $\mathcal{M} = \langle M, \oplus, *, \mathbf{1}, \mathbf{0} \rangle$  admits of a “good polynomial conditional”, i.e., a polynomial binary operation  $\circ$  such that

$$a \circ b = 1 \text{ iff } a \preceq b.$$

Thus, it might happen that the notion of logical truth of the logic based on  $\mathbb{HSP}(\text{QLQMV})$  is (finitely) axiomatizable, while the notion of “logical entailment” ( $\alpha \models \beta$ ) is not.

We will now show that the QMV-algebra  $\mathcal{M}_4$  (see Figure 7 below) does not admit any good polynomial conditional. The operations of  $\mathcal{M}_4$  are

defined as follows:

		$\oplus$
<b>0</b>	<b>0</b>	<b>0</b>
<b>0</b>	<i>a</i>	<i>a</i>
<b>0</b>	<i>b</i>	<i>b</i>
<b>0</b>	<b>1</b>	<b>1</b>
<i>a</i>	<b>0</b>	<i>a</i>
<i>a</i>	<i>a</i>	<b>1</b>
<i>a</i>	<i>b</i>	<b>1</b>
<i>a</i>	<b>1</b>	<b>1</b>
<i>b</i>	<b>0</b>	<i>b</i>
<i>b</i>	<i>b</i>	<b>1</b>
<i>b</i>	<i>a</i>	<b>1</b>
<i>b</i>	<b>1</b>	<b>1</b>
<b>1</b>	<b>0</b>	<b>1</b>
<b>1</b>	<i>a</i>	<b>1</b>
<b>1</b>	<i>b</i>	<b>1</b>
<b>1</b>	<b>1</b>	<b>1</b>

	*
<b>0</b>	<b>1</b>
<i>a</i>	<i>a</i>
<i>b</i>	<i>b</i>
<b>1</b>	<b>0</b>

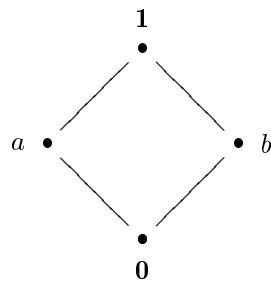


Figure 7.  $\mathcal{M}_4$

Let us consider the three-valued MV-algebra  $\mathcal{M}_3$ , whose operations are

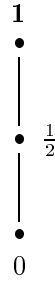


Figure 8.  $\mathcal{M}_3$

defined as follows:

		$\oplus$
0	0	0
0	$\frac{1}{2}$	$\frac{1}{2}$
0	1	1
$\frac{1}{2}$	0	$\frac{1}{2}$
$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{2}$	1	1
1	0	1
1	$\frac{1}{2}$	1
1	1	1

	*
0	1
$\frac{1}{2}$	$\frac{1}{2}$
1	0

It is easy to see that the map  $h : \mathcal{M}_4 \rightarrow \mathcal{M}_3$  such that  $\forall x \in \mathcal{M}_4$

$$h(x) := \begin{cases} 0, & \text{if } x = 0; \\ \frac{1}{2}, & \text{if } x = a \text{ or } x = b; \\ 1, & \text{otherwise} \end{cases}$$

is a homomorphism of  $\mathcal{M}_4$  into  $\mathcal{M}_3$ .

Suppose, by contradiction, that  $\mathcal{M}_4$  admits of a good polynomial conditional  $\rightarrow_{\mathcal{M}_4}$ . Since  $a \not\leq b$ , we have  $h(a \rightarrow_{\mathcal{M}_4} b) \neq 1$ . Thus,

$$1 \neq h(a \rightarrow_{\mathcal{M}_4} b) = h(a) \rightarrow_{\mathcal{M}_3} h(b) = \frac{1}{2} \rightarrow_{\mathcal{M}_3} \frac{1}{2} = 1,$$

contradiction.

17 QUANTUM LOGIC AND THE CUBE OF LOGICS  
(BY GIULIA BATTILOTTI AND CLAUDIA FAGGIAN)

Different forms of quantum logic can be axiomatized as sequent calculi [Dummett, 1976; Nishimura, 1980; Cutland and Gibbins, 1982; Tamura, 1988; Nishimura, 1994]. This permits us to investigate such logics more and more deeply from the proof-theoretical point of view. A sequent calculus for orthologic can be obtained from a calculus for classical logic, by requiring a special restriction on *contexts* in the rules that would permit to derive the distributive laws. The critical rules are the following: the introduction of disjunction on the left side, the introduction of conjunction on the right side, the rules concerning implication and negation. Such a restriction, however, determines some serious proof-theoretical difficulties, because quantum logic has a sufficiently strong negation that satisfies de Morgan's laws. The shortcoming becomes apparent when we try to prove the cornerstone result, represented by a cut-elimination theorem (which, essentially depends on the formulation of the rules that appear in our proofs).

A simple and compact sequent calculus for orthologic [Faggian and Sambin, 1997; Battilotti and Sambin, 1999], which admits cut-elimination by means of a neat procedure, can be obtained by a convenient strengthening of *basic logic*. This is a new logic that has been introduced in order to investigate a general structure for the space of logics [Sambin *et al.*, 1998].

In the framework of basic logic, constraints on contexts are not considered a limitation; on the contrary, they are regarded as a positive feature, which is called *visibility*. At the same time, negation is treated by exploiting the symmetry of the calculus: the main idea is to use Girard's linear negation, which can be interpreted as an orthocomplement in a quite natural way. This approach shows that orthologic (and non-distributive logics, in general) admits a proof-theory, which turns out to be simpler than the proof-theory for classical logic. Describing quantum logic in the framework of a uniform and general setting gives many advantages, since it permits us to study various logics and their mutual relations. In particular, we obtain a whole system of quantum logics (including *linear orthologic*); and for each of these logics we have a proof of the cut-elimination theorem. All this gives rise to a new formulation of classical logic [Faggian, 1997], with respect to which orthologic and the other quantum-like logics (created by this method) turn out to be characterizable as *substructural logics*. On this basis it is easy to compare different logics, and to prove embedding results [Battilotti, 1998].

**Basic logic and the cube of logics.**

As we already know, quantum logic represents a weakening of classical logic, obtained by dropping the distributive laws. There are at least two other important logics that are weaker than classical logic: intuitionistic logic and linear logic [Girard, 1987]. The situation can be sketched as pictured in Figure 9.

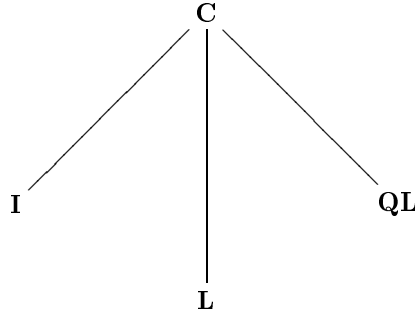


Figure 9. The most important weakenings of classical logic

It is natural to ask whether there exists a logic that represents a kind of common denominator for **Q**, **I** and **L**, in the same way as classical logic theoretically includes all the other logics. A solution to this problem has been found in terms of a suitable sequent calculus **B**, that represents a *basic logic*.

Differently from the calculi we have considered in the previous sections, a *sequent calculus* for a given logic **L** is based on *axioms* and *rules* that govern the behaviour of *sequents*. Any sequent has the form

$$M \vdash N$$

where  $M, N$  are (possibly empty) finite multisets of formulas.<sup>14</sup> Axioms are particular sequents. Any rule has the form

$$\frac{M_1 \vdash N_1 \quad \dots \quad M_n \vdash N_n}{M \vdash N}$$

where  $M_1 \vdash N_1, \dots, M_n \vdash N_n$  are the *premisses*, while  $M \vdash N$  is the *conclusion* of the rule. Rules can be *structural* or *operational*. Operational rules introduce a new connective, while structural rules deal only with the structure of the sequents (orders, repetitions, etc.).

A *derivation* is a sequence of sequents where any element is either an axiom or the conclusion of a rule whose premisses are previous elements of the sequence.

Basic logic has been introduced in [Battilotti and Sambin, 1999], and substantially reformulated in [Sambin *et al.*, 1998]. According to the sec-

<sup>14</sup>A *multiset* is a set of pairs such that the first element of every pair denotes any object, while the second element denotes the multiplicity of the occurrences of our object. Two multisets are equal if and only if all their pairs are equal, that is all their objects together with their multiplicities are equal.

ond formulation (we will follow here),<sup>15</sup> our logic is characterized by three strictly linked principles: *reflection*, *symmetry*, *visibility*. The reflection principle states the fact that logical constants are the result of importing into a formal system metalinguistic links between assertions, considered pre-existing. There is a method that leads to the rules of the calculus, starting from metalinguistic links between assertions. Such a method analyses the following equivalences, which assert a correspondence between language and metalanguage:

$$M \vdash \alpha \cdot \beta \quad \text{if and only if} \quad M \vdash \alpha \circ_R \beta$$

$$\alpha \cdot \beta \vdash N \quad \text{if and only if} \quad \alpha \circ_L \beta \vdash N$$

Here the generic sign “.”, corresponding to a metalinguistic link between assertions, is translated respectively into the connective  $\circ_R$ , when it appears on the right of the sign  $\vdash$ , and into the connective  $\circ_L$ , when it appears on the left. In **B**, the operational rules are completely determined by such equivalences. As a consequence, the meaning of a connective turns out to be semantically determined by the correspondence with a metalinguistic link, quite independently of any link with a context. Since every metalinguistic link is translated into a connective according to two specular ways, the system of rules, obtained by this method, turns out to be strongly symmetric. In fact **B** contains, for every axiom and for every (unary or binary) rule  $R$

$$\frac{M_i \vdash N_i}{M \vdash N} R$$

its symmetric rule  $R^s$ , given by

$$\frac{N_i^s \vdash M_i^s}{N^s \vdash M^s} R^s$$

where the map  $(-)^s$  is defined by induction (on the length of formulas), by putting  $\circ_R^s \equiv \circ_L$  and  $\circ_L^s \equiv \circ_R$ , given a suitable correspondence between propositional variables.

The third principle that **B** satisfies is the visibility property. A rule for a given connective is called *visible* when the principal formula and the corresponding *secondary formulas* appear in the rule without any context.<sup>16</sup>

<sup>15</sup>The formulation of the rules of **B** presented in [Sambin *et al.*, 1998] is based on finite lists rather than finite multisets of formulas; hence it contains in addition the structural rule of exchange. Here we prefer to use multisets, in order to obtain an easy comparison with sequent calculi for quantum logics.

<sup>16</sup>In any operational rule, the formula in the conclusion that contains the connective introduced by the rule itself is called the *principal formula*; the formulas in the premisses that are the components of the formula introduced by the rule are called the *secondary formulas*.

We shall see below the syntactical consequences of visibility; we stress here that semantically it corresponds to the fact that basic connectives have a primitive meaning, in accordance with the reflection principle.

As an example let us refer to a rule that plays an important role in the case of quantum logic. As is well known, in classical logic, disjunction is introduced on the left according to the following rule:

$$\frac{M, \alpha \vdash N \quad M, \beta \vdash N}{M, \alpha \vee \beta \vdash N}$$

In the case of **B**, instead, disjunction is introduced according to the following visible form:

$$\frac{\alpha \vdash N \quad \beta \vdash N}{\alpha \vee \beta \vdash N}$$

where the context  $M$  has disappeared.

From the intuitive point of view, one can read the difference between the two cases as follows: the rule typical of classical logic attaches a meaning to the connective  $\vee$  in presence of the link “,” with  $M$  (such a link is to be interpreted as a conjunction), whereas the visible rule is intended to explain the meaning of the connective  $\vee$  by referring only to the connective itself. In particular, the visible rule does not permit us to prove the equation that links conjunction and disjunction (the distributive law of  $\wedge$  with respect to  $\vee$ ). As a consequence, any sequent calculus for a quantum logic shall adopt the visible form for the rule that concerns the introduction of disjunction on the left. As to the other rules, visibility is not strictly necessary in order to obtain an adequate sequent calculus for quantum logic. However, a more convenient strategy permits us to axiomatize quantum logic, by adding only structural rules to basic logic, without any change in the rules for the connectives. In this way, we can preserve the characteristic properties of symmetry and visibility of **B**, that turn out to be highly convenient from the proof-theoretical point of view (as we will see below).

Basic logic **B** has no structural rules. As a consequence, **B** can be regarded as “the logic of connectives” from which various stronger logics can be obtained by adding suitable structural rules.

Let us now present the sequent calculus for **B**. Similarly to linear logic, the language of **B** contains two pairs of conjunctions and disjunctions: the *additive conjunction*  $\wedge$  and the *multiplicative conjunction*  $\otimes$ ; the *additive disjunction*  $\vee$  and the *multiplicative disjunction*  $\wp$ . Further there are two conditionals ( $\rightarrow$ ,  $\leftarrow$ ), and two pairs of propositional literals  $1$  and  $\top$ ,  $0$  and  $\perp$ .



**The basic sequent calculus B**
*Axioms*

$$\alpha \vdash \alpha$$

*Operational rules*

(Formation)	$\frac{\beta, \alpha \vdash N}{\beta \otimes \alpha \vdash N} \otimes L$	$\frac{M \vdash \alpha, \beta}{M \vdash \alpha \wp \beta} \wp R$
(Reflection)	$\frac{\alpha \vdash N_1 \quad \beta \vdash N_2}{\alpha \wp \beta \vdash N_1, N_2} \wp L$	$\frac{M_2 \vdash \beta \quad M_1 \vdash \alpha}{M_2, M_1 \vdash \beta \otimes \alpha} \otimes R$
(Formation)	$\frac{\vdash N}{1 \vdash N} 1L$	$\frac{M \vdash}{M \vdash \perp} \perp R$
(Reflection)	$\perp \vdash \perp L$	$\vdash 1 1R$
(Formation)	$\frac{\beta \vdash N \quad \alpha \vdash N}{\beta \vee \alpha \vdash N} \vee L$	$\frac{M \vdash \alpha \quad M \vdash \beta}{M \vdash \alpha \wedge \beta} \wedge R$
(Reflection)	$\frac{\alpha \vdash N}{\alpha \wedge \beta \vdash N} \wedge L$	$\frac{M \vdash \beta}{M \vdash \beta \vee \alpha} \vee R$
(Formation)	$0 \vdash N 0L$	$M \vdash \top \top R$
(Formation)	$\frac{\beta \vdash \alpha}{\beta \leftarrow \alpha \vdash} \leftarrow L$	$\frac{\alpha \vdash \beta}{\vdash \alpha \rightarrow \beta} \rightarrow R$
(Reflection)	$\frac{\vdash \alpha \quad \beta \vdash N}{\alpha \rightarrow \beta \vdash N} \rightarrow L$	$\frac{M \vdash \beta \quad \alpha \vdash}{M \vdash \beta \leftarrow \alpha} \leftarrow R$
(Order)	$\frac{\alpha \vdash \beta \quad \gamma \vdash \delta}{\beta \rightarrow \gamma \vdash \alpha \rightarrow \delta} \rightarrow U$	$\frac{\gamma \vdash \delta \quad \alpha \vdash \beta}{\gamma \leftarrow \beta \vdash \delta \leftarrow \alpha} \leftarrow U$

We will distinguish three main kinds of structural rules, labelled by the letters **L**, **R** and **S**. The extensions of **B** obtained by the addition of any combination of such rules can be organized in a cube, which is conceived as an architecture whose basis is **B** (see Figure 9).

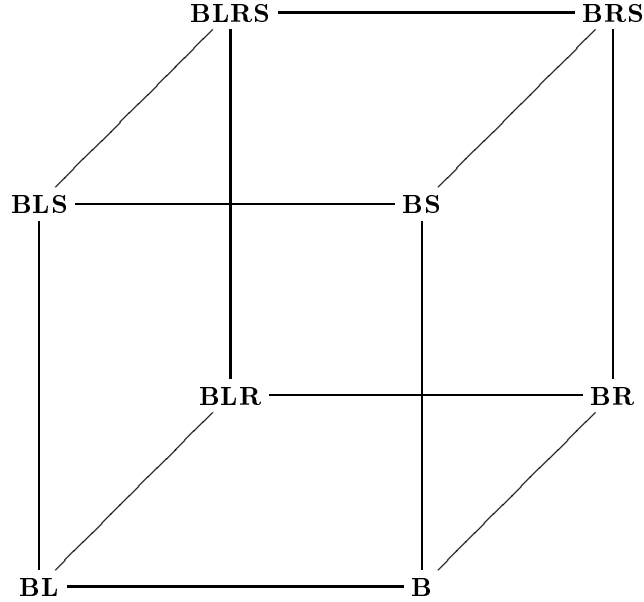


Figure 10. The cube of logics

In the cube, every logic with **S** satisfies the structural rules of weakening and contraction:<sup>17</sup>

$$\frac{M \vdash N}{M, O \vdash P, N} \textit{weakening} \quad \frac{M, O, O \vdash N, N, P}{M, O \vdash N, P} \textit{contraction}$$

Every logic with “**L**” allows left contexts in any inference rule; every logic with **R** allows right contexts in any inference rule. In particular, the cube solves our initial problem, sketched in Figure 11. In fact, vertex **BLRS**, opposed to **B** represents classical logic, vertex **BLR** and vertex **BLS** represent respectively Girard’s linear logic and intuitionistic logic; finally, vertex **BS** corresponds to paraconsistent quantum logic (see below). Moreover, since logics with **R** are simply the symmetric copy of logics with **L**, logics containing both **L** and **R** (**BLRS**, **BLR**) or logics containing neither **L** nor

<sup>17</sup>As we have seen, in **B** (as well as in linear logic) the connectives conjunction and disjunction are splitted into a *multiplicative* and an *additive* connective. Such a distinction depends on the fact that there are two ways of formulating contexts in any operational rule: this leads to a multiplicative and to an additive form for each rule. The multiplicative and additive formulation turn out to be equivalent, whenever the structural rules of weakening and contraction hold. Hence, the distinction plays an essential role in linear logic and a fortiori in basic logic (where weakening and contraction fail); at the same time, it vanishes in classical logic and in orthologic.

**R** (**BS**, **B**), are symmetric. The study of quantum logics finds place in the diagonal of symmetric logics, where a finer distinction of structural rules can be obtained.

**Sequent calculus for Orthologic.**

The logic **BS** is non-distributive. Let us consider the fragment of **BS** restricted to the connectives  $\wedge$  and  $\vee$ . If we want to obtain a quantum logic from it, what is still missing is an involutive negation, satisfying de Morgan.

This aim can be reached by extending the language and by adopting Girard's negation. The key point is to assume as primitive symbols of the language both the propositional variables and their duals. In other words, the propositional literals are assumed to be given in pairs, consisting of a positive element (written  $p$ ) and of a negative one (written  $p^\perp$ ). On this basis, the negation of a formula is defined as follows:

$$p^{\perp\perp} := p \quad (\alpha \wedge \beta)^\perp := \alpha^\perp \vee \beta^\perp \quad (\alpha \vee \beta)^\perp := \alpha^\perp \wedge \beta^\perp$$

By this choice, we obtain a calculus called *basic orthologic* and denoted by  ${}^\perp\mathbf{BS}$  (where the symbol  ${}^\perp$  reminds us that our calculus is applied to a dual language). Basic orthologic turns out to be equivalent to paraconsistent quantum logic (**PQL**). As we already know, **PQL** represents a weakening of orthologic, that is obtained by dropping the *non contradiction* and the *excluded middle* principles. Hence, in order to have a calculus for orthologic, it will be sufficient to add such principles to our  ${}^\perp\mathbf{BS}$ . This can be done by means of two new structural rules called *transfer*. The result is a calculus for orthologic, which will be denoted by  ${}^\perp\mathbf{O}$ .

The rules of  ${}^\perp\mathbf{O}$  are the following (where (i) -(v) are the rules of  ${}^\perp\mathbf{BS}$ <sup>18</sup> while (vi) express the *transfer* rules).

$$(i) \quad \alpha \vdash \alpha$$

$$(ii) \quad \frac{\alpha \vdash N \quad \beta \vdash N}{\alpha \vee \beta \vdash N} \vee L \qquad \frac{M \vdash \alpha \quad M \vdash \beta}{M \vdash \alpha \wedge \beta} \wedge R$$

$$(iii) \quad \frac{\alpha \vdash N}{\alpha \wedge \beta \vdash N} \quad \frac{\beta \vdash N}{\alpha \wedge \beta \vdash N} \wedge L \qquad \frac{M \vdash \alpha}{M \vdash \alpha \vee \beta} \quad \frac{M \vdash \beta}{M \vdash \alpha \vee \beta} \vee R$$

$$(iv) \quad \frac{M \vdash N}{M, O \vdash P, N} \textit{weakening}$$

<sup>18</sup>Note that, in  ${}^\perp\mathbf{BS}$ , weakening and contraction are redundant. In fact, one can show that such a calculus admits elimination of contraction. At the same time, weakening on the right and on the left can be simulated by  $\wedge L$  and  $\vee R$ , respectively. On this basis, **PQL** turns out to admit a very simple formulation, given by (i), (ii), (iii).

$$(v) \quad \frac{M, O, O \vdash N, N, P}{M, O \vdash N, P} \textit{contraction}$$

$$(vi) \quad \frac{M \vdash N}{M, N^\perp \vdash} \textit{tr1} \qquad \frac{M \vdash N}{\vdash M^\perp, N} \textit{tr2}$$

It is not hard to prove:

THEOREM 149.  $\perp\mathbf{BS}$  is a calculus for paraconsistent quantum logic.

THEOREM 150.  $\perp\mathbf{O}$  is a calculus for orthologic.

As we have seen, our calculus  $\perp\mathbf{O}$  contains both  $p, q, r, \dots$  and  $p^\perp, q^\perp, r^\perp, \dots$ . Moreover, for any rule of the calculus, the calculus shall contain also the symmetric one. As a consequence, whenever the calculus produces a derivation  $\Pi$ , it will also produce the dual derivation  $\Pi^\perp$ , obtained substituting every axiom  $\alpha \vdash \alpha$  with the axiom  $\alpha^\perp \vdash \alpha^\perp$  and every occurrence of a rule with an occurrence of its corresponding symmetric rule (e.g.  $\wedge R$  with  $\vee L$ ). On this basis there holds:

LEMMA 151. The following rule is derivable for  $\perp\mathbf{O}$ :

$$\frac{M \vdash N}{N^\perp \vdash M^\perp}$$

**Sketch of the proof** One can easily see that  $M \vdash N$  is derivable by a derivation  $\Pi$  if and only if  $N^\perp \vdash M^\perp$  is derivable by the symmetric derivation  $\Pi^\perp$ . ■

The structure of the calculus  $\perp\mathbf{O}$  permits us to prove the following cut-elimination result.

THEOREM 152.  $\perp\mathbf{O}$  admits the elimination of the cuts.

$$\frac{O \vdash \mu \quad M, \mu \vdash N}{M, O \vdash N} \textit{cutL} \qquad \frac{O \vdash \mu, P \quad \mu \vdash N}{O \vdash N, P} \textit{cutR}$$

**Sketch of the proof** Like in Gentzen, the cut-elimination procedure is obtained by induction on two parameters: the *degree* and the *rank* of the cut-formula<sup>19</sup>.

<sup>19</sup>Suppose a derivation and a sequent where a formula  $\alpha$  occurs. Let us consider the *paths* (i.e. the sequences of consecutive sequents) connecting that point with the point where the formula  $\alpha$  has been introduced, (by an axiom, or by weakening, or by an operational rule whose principal formula was  $\alpha$ ). The *rank* of this particular occurrence of  $\alpha$  is the maximum among the lengths of all these paths. In other words, the rank represents the 'maximum length' between the formula-occurrence we are examining and the point where that occurrence has been introduced.

The *degree* (or *length*) of a formula  $\alpha$  is the number of the connectives occurring in  $\alpha$ .

The calculus  ${}^{\perp}\mathbf{O}$  permits us to overcome in a simple way two questions that usually make cut elimination for orthologic so complicated: (i) constraints on contexts and (ii) negation. We give a sketch of the proof, considering the two points. The first problem is solved by visibility, while the second one is solved by symmetry.

- (i) As we have seen, in any calculus for quantum logic the rule that introduces  $\vee$  on the left (here indicated with  $\vee L$ ) must have an empty context on the left. Now consider, for a generic calculus, the derivation

$$\frac{\frac{\alpha \vdash \gamma \wedge \delta \quad \beta \vdash \gamma \wedge \delta}{\alpha \vee \beta \vdash \gamma \wedge \delta} \vee L \quad \frac{M, \gamma \vdash \Delta}{M, \gamma \wedge \delta \vdash \Delta} cutL}{M, \alpha \vee \beta \vdash \Delta} cutL$$

In this derivation, the cut-formula is principal on the right premiss; hence the right rank is 1. In such a situation, Gentzen's procedure to lower the rank must operate on the left; this would necessarily produce the two derivations

$$\frac{\alpha \vdash \gamma \wedge \delta \quad M, \gamma \wedge \delta \vdash \Delta}{M, \alpha \vdash \Delta} cutL \quad \frac{\beta \vdash \gamma \wedge \delta \quad M, \gamma \wedge \delta \vdash \Delta}{M, \beta \vdash \Delta} cutL$$

Now, one would like to conclude by applying  $\vee L$ , in order to obtain  $M, \alpha \vee \beta \vdash \Delta$ . However, this step is here not allowed, unless  $M$  is empty. Such a problem does not arise for the calculus  ${}^{\perp}\mathbf{O}$ , because, by visibility, every principal formula has an empty context.

- (ii) In  ${}^{\perp}\mathbf{O}$  the only rules about negation are the structural rules of transfer. Let us consider a derivation of the form:

$$\frac{O \vdash \mu^{\perp} \quad \frac{\frac{\vdots \Pi}{M \vdash \mu} tr1}{M, \mu^{\perp} \vdash} cutL}{M, O \vdash} cutL$$

We can reduce the rank in a quick way, by exploiting symmetry. In fact, Girard's negation has the nice property that every formula  $\alpha$  and its dual  $\alpha^{\perp}$  have exactly the same degree. The same idea can be extended to derivations, and hence to the rank of a cut. As we have seen in Lemma 151, whenever we have a derivation  $\Pi$  for the sequent  $M \vdash N$ , we also have the dual derivation  $\Pi^{\perp}$ , which derives  $N^{\perp} \vdash M^{\perp}$ . The two derivations  $\Pi$  and  $\Pi^{\perp}$  have exactly the same (symmetrical) structure. Hence in particular, if  $\mu$  is principal,  $\mu^{\perp}$  is principal. If  $\mu$  has rank  $r$ , then also  $\mu^{\perp}$  will have the same rank  $r$ . In

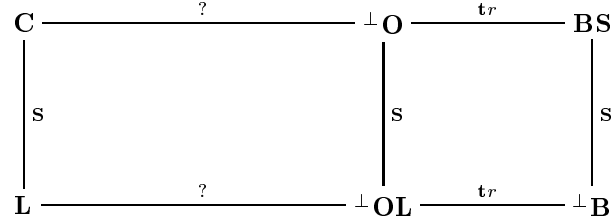


Figure 11. The diagonal of the cube

such a situation, in order to raise the cut rule, we can substitute  $\Pi^\perp$  by  $\Pi$  (*flipping derivation*). As a consequence, the initial derivation will be simply reduced to:

$$\frac{\frac{O \vdash \mu^\perp \quad \mu^\perp \vdash M^\perp}{O \vdash M^\perp} \text{tr1} \quad \frac{\vdots \Pi^\perp}{\mu^\perp \vdash M^\perp} \text{cutL}}{O, M \vdash} \text{tr1}$$

■

### Quantum logics and classical logic

We will now consider the *symmetric diagonal* of the cube in the diagram in Figure 11. In our diagram, the calculus  $\perp\mathbf{O}$  appears as an intermediate point between basic orthologic and classical logic. Similarly, we have another intermediate point between basic logic and linear logic: this is given by  $\perp\mathbf{B} + \mathbf{tr}$ , which represents the common denominator for orthologic and linear logic (we will call it “ortholinear logic”  $\perp\mathbf{OL}$ ). In the same way,  $\perp\mathbf{B}$  turns out to be the common denominator of basic orthologic and linear logic. On this basis, we obtain a whole system of quantum logics, which are all cut-free. The last of our logics,  $\perp\mathbf{B} + \mathbf{tr}$ , seems to be a good candidate in order to represent a *linear quantum logic* in the sense of Pratt [1993].

So far we have only dealt with a fragment of basic logic, which has no implication connective. By means of this linguistic restriction, we have easily proved the equivalence between our calculi and the usual formulations of paraconsistent quantum logic and of orthologic. However, the same methods can be naturally applied to the complete versions of our calculi, preserving cut-elimination and flipping of derivations. In this way, we will have a primitive implication connective  $\rightarrow$  (together with its dual  $\leftarrow$ ) in all the logics we have considered. An interesting question to be investigated concerns the possibility of physical interpretations of such new connectives.

In the diagram above, we have still a question mark concerning the path from orthologic to classical logic. Our question can be solved as follows:

**THEOREM 153.** *A calculus for classical logic is obtained from a calculus for orthologic by adding a pair of structural rules, named separation:*

$$(vii) \quad \frac{M, O \vdash}{M \vdash O^\perp} \text{ sep1} \qquad \frac{\vdash N, P}{N^\perp \vdash P} \text{ sep2}$$

It is easy to see that, in the framework of  $^\perp\mathbf{O}$ , separation rules allow us to derive the following full cut

$$\frac{M_1 \vdash \mu, N_1 \quad M_2, \mu \vdash N_2}{M_1, M_2 \vdash N_1, N_2} \text{ cut}$$

The converse is also true: full cut allows us to derive separation (by cutting with  $\vdash A, A^\perp$ ). In this sense, separation and cut rule are equivalent: adding either of them to orthologic gives rise to one and the same logic. Theorem 153 then expresses, with a more effective<sup>20</sup> content, the well known fact that adding a full cut rule to orthologic yields classical logic (cf. [Dummett, 1976], [Cutland and Gibbins, 1982]).

It is natural to ask what is the meaning of *sep*. In the same way as the *tr* rules are equivalent to *tertium non datur* and *non contradiction*, the *sep* rules turn out to be equivalent to *reductio ad absurdum*<sup>21</sup>

$$\frac{M, \alpha^\perp \vdash}{M \vdash \alpha} \text{ RAA}$$

Let us consider again our Figure 11, where the question marks have been substituted by *sep*. Given the logic  $\mathbf{B}$  as a basic calculus, which contains the fundamental rules for the connectives, several structural rules can be added: each rule permits us to reach a “superior” logic. The strongest element is represented by classical logic, which can be characterized as  $^\perp\mathbf{B} + \mathbf{S} + \mathbf{tr} + \mathbf{sep}$ . With respect to our formulation of classical logic (denoted by  $^\perp\mathbf{C}$ ) all the other logics in the diagram can be described as ‘substructural logics: for, they can be obtained by dropping some structural rules. This situation holds in particular for quantum logics, which turn out to be simpler and more basic than classical logic, from the proof-theoretical point of view.

As we have seen, the examples of quantum logic (we have considered so far) are, at the same time, substructural with respect to classical logic and substructural one with respect to the other. On this basis, on can prove

<sup>20</sup>For, in a sequent calculus cut should represent a *metarule*, that is should be eliminable.

<sup>21</sup>In [Gibbins, 1985], pag.361, Gibbins shows that dropping the rule RAA has a direct justification in terms of *quantum mechanics*, and this is the only case of direct justification, among all the rules which must be restricted in *quantum logic*.

some embedding theorems, by convenient restriction of our structural rules to suitable kinds of formulas, by means of special modalities. In the case of linear logic, *exponentials* have been introduced in order to express weakening and contraction. In the case of quantum logics, instead, we should obtain rules of separation and of transfer in a suitable way. How to express the separation rules in orthologic, in order to obtain an embedding of classical logic into orthologic? Given  $\perp\mathbf{O}$ , let us first assume in the language, besides the literals  $p$  and  $p^\perp$ , two new kinds of literals,  $\downarrow p$  and  $\downarrow p^\perp$ . This permits us to obtain a new kind of formulas, that will be named “separable formulas”, formulated as follows:

$$\downarrow(p) := \downarrow p \quad \downarrow(p^\perp) := \downarrow p^\perp \quad \downarrow(\downarrow p) := \downarrow p \quad \downarrow(\downarrow p^\perp) := \downarrow p^\perp$$

for literals;

$$\downarrow(\alpha \circ \beta) := \downarrow\alpha \circ \downarrow\beta$$

for every binary connective  $\circ$ .

Separable formulas are precisely those formulas that satisfy the separation rules, which are then defined as follows:

$$(vii') \quad \frac{M, \downarrow O \vdash}{M \vdash \downarrow O^\perp} \downarrow sep1 \qquad \frac{\vdash \downarrow N, P}{\downarrow N^\perp \vdash P} \downarrow sep2$$

where formulas in  $M$ ,  $N$  are any kind of formulas, while formulas in  $\downarrow M$ ,  $\downarrow N$  are separable formulas. We can now introduce the system  $\downarrow\perp\mathbf{O}$ , which is defined by the rules of  $\perp\mathbf{O}$  and by the rules  $\downarrow sep$ . In this system, the sign  $\downarrow$  plays the role of a *modality* (which behaves as an unary *monotonic* connective: if  $M \vdash N$ , is a derivable sequent in  $\downarrow\perp\mathbf{O}$ , then also  $\downarrow M \vdash \downarrow N$  is a derivable sequent).

Consider now the system  $\perp\mathbf{C}$  for classical logic, and let us describe  $\downarrow$  as a map from formulas of the language of  $\perp\mathbf{C}$  into formulas of the language of  $\downarrow\perp\mathbf{O}$ . It is easy to show, by induction on the depth of the derivation, that:

**THEOREM 154.** *For every  $M$ ,  $N$ ,  $M \vdash N$  is derivable in  $\perp\mathbf{C}$  if and only if  $\downarrow M \vdash \downarrow N$  is derivable in  $\downarrow\perp\mathbf{O}$ .*

**Sketch of the proof** A proof can be obtained by a natural transformation of a similar proof, given in [Battilotti, 1998] for the case of classical logic and basic orthologic. ■

As a consequence we obtain an embedding of  $\perp\mathbf{C}$  in  $\downarrow\perp\mathbf{O}$ . Formulas of the kind  $\downarrow\alpha$  can be interpreted as “the classical part of  $\downarrow\perp\mathbf{O}$ ”. Similarly to  $\perp$ , the sign  $\downarrow$  does not represent here a connective; therefore, there is no need of introduction rules. One can prove that sequents like  $\downarrow\alpha \vdash \alpha$  or like  $\alpha \vdash \downarrow\alpha$  are not provable (differently from the exponentials in linear logic).



In this way, the system  $\downarrow^\perp \mathbf{O}$  is simply a way to represent the *coexistence* of classical and quantum logic: it does not assert that “classical” propositions are stronger or weaker than “quantum” propositions.

## 18 CONCLUSION

Some general questions that have been often discussed in connection with (or against) quantum logic are the following:

- (a) Why quantum logics?
- (b) Are quantum logics helpful to solve the difficulties of QT?
- (c) Are quantum logics “real logics”? And how is their use compatible with the mathematical formalism of QT, based on classical logic?
- (d) Does quantum logic confirm the thesis that “logic is empirical”?

Our answers to these questions are, in a sense, trivial, and close to a position that Gibbins (1991) has defined a “quietist view of quantum logic”. It seems to us that quantum logics are not to be regarded as a kind of “clue”, capable of solving the main physical and epistemological difficulties of QT. This was perhaps an illusion of some pioneering workers in quantum logic. Let us think of the attempts to recover a *realistic interpretation* of QT based on the properties of the quantum logical connectives<sup>22</sup>.

Why quantum logics? Simply because “quantum logics are there!” They seem to be deeply incorporated in the abstract structures generated by QT. Quantum logics are, without any doubt, *logics*. As we have seen, they satisfy all the canonical conditions that the present community of logicians require in order to call a given abstract object *a logic*. A question that has been often discussed concerns the compatibility between quantum logic and the mathematical formalism of quantum theory, based on classical logic. Is the quantum physicist bound to a kind of “logical schizophrenia”? At first sight, the compresence of different logics in one and the same theory may give a sense of uneasiness. However, the splitting of the basic logical operations (negation, conjunction, disjunction,...) into different connectives with different meanings and uses is now a well accepted logical phenomenon, that admits consistent descriptions. Classical and quantum logic turn out to apply to different sublanguages of quantum theory, that must be sharply distinguished.

Finally, does quantum logic confirm the thesis that “logic is empirical”? At the very beginning of the contemporary discussion about the *nature of logic*, the claim that the “right logic” to be used in a given theoretical situation may depend also on experimental data appeared to be a kind of

---

<sup>22</sup>See for instance [Putnam, 1969]

extremistic view, in contrast with a leading philosophical tradition according to which a characteristic feature of logic should be its absolute independence from any content.

These days, an empirical position in logic is generally no longer regarded as a “daring heresy”. At the same time, as we have seen, we are facing not only a variety of logics, but even a variety of *quantum logics*. As a consequence, the original question seems to have turned to the new one : to what extent is it reasonable to look for the “right logic” of QT?

#### ACKNOWLEDGEMENTS

The authors are grateful to Giulia Battilotti and Claudia Faggian who wrote a section of this chapter.

M. L. Dalla Chiara  
*Università Firenze, Italy.*

R. Giuntini  
*Università Cagliari, Italy.*

#### BIBLIOGRAPHY

- [Battilotti, 1998] G. Battilotti. Embedding classical logic into basic orthologic with a primitive modality, *Logic Journal of the IGPL*, **6**, 383–402, 1998.
- [Battilotti and Sambin, 1999] G. Battilotti and G. Sambin. Basic logic and the cube of its extensions, in A. Cantini, E. Casari, and P. Minari (eds), *Logic and Foundations of Mathematics*, pp. 165–186. Kluwer, Dordrecht, 1999.
- [Bell and Slomson, 1969] J. L. Bell and A. B. Slomson. *Models and Ultraproducts: An Introduction*, North-Holland, Amsterdam, 1969.
- [Beltrametti and Cassinelli, 1981] E. Beltrametti and G. Cassinelli. *The Logic of Quantum Mechanics*, Vol. 15 of *Encyclopedia of Mathematics and its Applications*, Addison-Wesley, Reading, 1981.
- [Birkhoff, 1995] G. Birkhoff. *Lattice Theory*, Vol. 25 of *Colloquium Publications*, 3<sup>rd</sup> edn, American Mathematical Society, Providence, 1995.
- [Birkhoff and von Neumann, 1936] G. Birkhoff and J. von Neumann. The logic of quantum mechanics, *Annals of Mathematics* **37**, 823–843, 1936.
- [Cattaneo and Laudisa, 1994] G. Cattaneo and F. Laudisa. Axiomatic unsharp quantum mechanics, *Foundations of Physics* **24**, 631–684, 1994.
- [Cattaneo and Nisticò, 1986] G. Cattaneo and G. Nisticò. Brouwer–Zadeh posets and three-valued Lukasiewicz posets, *Fuzzy Sets and Systems* **33**, 165–190, 1986.
- [Chang, 1957] C. C. Chang. Algebraic analysis of many valued logics, *Transactions of the American Mathematical Society* **88**, 74–80, 1957.
- [Chang, 1958] C. C. Chang. A new proof of the completeness of Lukasiewicz axioms, *Transactions of the American Mathematical Society* **93**, 467–490, 1958.
- [Dalla Chiara, 1981] M. L. Dalla Chiara. Some metalogical pathologies of quantum logic, in E. Beltrametti and B. V. Fraassen (eds), *Current Issues in Quantum Logic*, Vol. 8 of *Ettore Majorana International Science Series*, Plenum, New York, pp. 147–159, 1981.
- [Cutland and Gibbins, 1982] N. Cutland and P. Gibbins. A regular sequent calculus for quantum logic in which  $\wedge$  and  $\vee$  are dual, *Logique et Analyse - Nouvelle Serie* - **25**(45), 221–248, 1982.

- [da Costa *et al.*, 1992] N. C. A. da Costa, S. French and D. Krause. The Schrödinger problem, in M. Bibtol and O. Darrigol (eds), *Erwin Schrödinger: Philosophy and the Birth of Quantum Mechanics*, Editions Frontières, pp. 445–460, 1992.
- [Dalla Chiara and Giuntini, 1994] M. L. Dalla Chiara and R. Giuntini. Unsharp quantum logics, *Foundations of Physics* **24**, 1161–1177, 1994.
- [Dalla Chiara and Giuntini, 1995] M. L. Dalla Chiara and R. Giuntini. The logics of orthoalgebras, *Studia Logica* **55**, 3–22, 1995.
- [Dalla Chiara and Toraldo di Francia, 1993] M. L. Dalla Chiara and G. Toraldo di Francia. Individuals, kinds and names in physics, in G. Corsi, M. L. Dalla Chiara and G. Ghirardi (eds), *Bridging the Gap: Philosophy, Mathematics, and Physics*, Kluwer Academic Publisher, Dordrecht, pp. 261–283, 1993.
- [Davies, 1976] E. B. Davies. *Quantum Theory of Open Systems*, Academic, New York, 1976.
- [Dishkant, 1972] H. Dishkant. Semantics of the minimal logic of quantum mechanics, *Studia Logica* **30**, 17–29, 1972.
- [Dummett, 1976] M. Dummett. Introduction to quantum logic, unpublished, 1976.
- [Dvurečenskij and Pulmannová, 1994] A. Dvurečenskij and S. Pulmannová. D-test spaces and difference poset, *Reports on Mathematical Physics* **34**, 151–170, 1994.
- [Faggian, 1997] C. Faggian. Classical proofs via basic logic, in *Proceedings of CSL '97*, pp. 203–219. *Lectures Notes in Computer Science 1414*, Springer, Berlin, 1997.
- [Faggian and Sambin, 1997] C. Faggian and G. Sambin. From basic logic to quantum logics with cut elimination, *International Journal of Theoretical Physics* **12**, 1997.
- [Finch, 1970] P. D. Finch. Quantum logic as an implication algebra, *Bulletin of the Australian Mathematical Society* **2**, 101–106, 1970.
- [Foulis and Bennett, 1994] D. J. Foulis and M. K. Bennett. Effect algebras and unsharp quantum logics, *Foundations of Physics* **24**, 1325–1346, 1994.
- [Gibbins, 1985] P. Gibbins. A user-friendly quantum logic, *Logique-et-Analyse.-Nouvelle-Serie* **28**, 353–362, 1985.
- [Gibbins, 1987] P. Gibbins. *Particles and Paradoxes - The Limits of Quantum Logic*, Cambridge University Press, Cambridge, 1987.
- [Girard, 1987] J. Y. Girard. Linear logic, *Theoretical Computer Science* **50**, 1–102, 1987.
- [Giuntini, 1991] R. Giuntini. A semantical investigation on Brouwer-Zadeh logic, *Journal of Philosophical Logic* **20**, 411–433, 1991.
- [Giuntini, 1992] R. Giuntini. Brouwer-Zadeh logic, decidability and bimodal systems, *Studia Logica* **51**, 97–112, 1992.
- [Giuntini, 1993] R. Giuntini. Three-valued Brouwer-Zadeh logic, *International Journal of Theoretical Physics* **32**, 1875–1887, 1993.
- [Giuntini, 1995] R. Giuntini. Quasilinear QMV algebras, *International Journal of Theoretical Physics* **34**, 1397–1407, 1995.
- [Giuntini, 1996] R. Giuntini. Quantum MV algebras, *Studia Logica* **56**, 393–417, 1996.
- [Goldblatt, 1984] R. H. Goldblatt. Orthomodularity is not elementary, *Journal of Symbolic Logic* **49**, 401–404, 1984.
- [Goldblatt, 1974] R. Goldblatt. Semantics analysis of orthologic, *Journal of Philosophical Logic* **3**, 19–35, 1974.
- [Greechie, 1981] R. J. Greechie. A non-standard quantum logic with a strong set of states, in E. G. Beltrametti and B. C. van Fraassen (eds), *Current Issues in Quantum Logic*, Vol. 8 of *Ettore Majorana International Science Series*, Plenum, New York, pp. 375–380, 1981.
- [Greechie and Gudder, n.d.] R. J. Greechie and S. P. Gudder. Effect algebra counterexamples, preprint, n. d.
- [Gudder, 1995] S. P. Gudder. Total extensions of effect algebras, *Foundations of Physics Letters* **8**, 243–252, 1995.
- [Hardegree, 1975] G. H. Hardegree. Stalnaker conditionals and quantum logic, *Journal of Philosophical Logic* **4**, 399–421, 1975.
- [Hardegree, 1976] G. M. Hardegree. The conditional in quantum logic, in P. Suppes (ed.), *Logic and Probability in Quantum Mechanics*, Reidel, Dordrecht, pp. 55–72, 1976.

- [Kalmbach, 1983] G. Kalmbach. *Orthomodular Lattices*, Academic Press, New York, 1983.
- [Keller, 1980] H. A. Keller. Ein nichtklassischer Hilbertscher Raum, *Mathematische Zeitschrift* **172**, 41–49, 1980.
- [Kôpka and Chovanec, 1994] F. Kôpka and F. Chovanec. D-posets, *Mathematica Slovaca* **44**, 21–34, 1994.
- [Kraus, 1983] K. Kraus. *States, Effects and Operations*, Vol. 190 of *Lecture Notes in Physics*, Springer, Berlin, 1983.
- [Ludwig, 1983] G. Ludwig. *Foundations of Quantum Mechanics*, Vol. 1, Springer, Berlin, 1983.
- [Mackey, 1957] G. Mackey. *The Mathematical Foundations of Quantum Mechanics*, Benjamin, New York, 1957.
- [Mangani, 1973] P. Mangani. Su certe algebre connesse con logiche a più valori, *Bollettino Unione Matematica Italiana* **8**, 68–78, 1973.
- [Minari, 1987] P. Minari. On the algebraic and Kripkean logical consequence relation for orthomodular quantum logic, *Reports on Mathematical Logic* **21**, 47–54, 1987.
- [Mittelstaedt, 1972] P. Mittelstaedt. On the interpretation of the lattice of subspaces of Hilbert space as a propositional calculus, *Zeitschrift für Naturforschung* **27a**, 1358–1362, 1972.
- [Morash, 1973] R. P. Morash. Angle bisection and orthoautomorphisms in Hilbert lattices, *Canadian Journal of Mathematics* **25**, 261–272, 1973.
- [Nishimura, 1980] H. Nishimura. Sequential method in quantum logic, *Journal of Symbolic Logic* **45**, 339–352, 1980.
- [Nishimura, 1994] H. Nishimura. Proof theory for minimal quantum logic I and II, *International Journal of Theoretical Physics* **33**, 102–113, 1427–1443, 1994.
- [Pratt, 1993] V. Pratt. Linear logic for generalized quantum mechanics. In *Proceedings of the Workshop on Physics and Computation*, pp. 166–180, IEEE, 1993.
- [Pták and Pulmannová, 1991] P. Pták and S. Pulmannová. *Orthomodular Structures as Quantum Logics*, number 44 in *Fundamental Theories of Physics*, Kluwer, Dordrecht, 1991.
- [Putnam, 1969] H. Putnam. Is logic empirical?, Vol. 5 of *Boston Studies in the Philosophy of Science*, Reidel, Dordrecht, pp. 216–241, 1969.
- [Sambin, 1996] G. Sambin. A new elementary method to represent every complete Boolean algebra, in A. Ursini, and P. Aglianò (eds), *Logic and Algebra*, Marcel Dekker, New York, pp. 655–665, 1996.
- [Sambin *et al.*, 1998] G. Sambin, G. Battilotti and C. Faggian. Basic logic: reflection, symmetry, visibility, *The Journal of Symbolic Logic*, to appear.
- [Solèr, 1995] M. P. Solèr. Characterization of Hilbert space by orthomodular spaces, *Communications in Algebra*, **23**, 219–243, 1995.
- [Takeuti, 1981] G. Takeuti. Quantum set theory, in E. G. Beltrametti and B. C. van Fraassen (eds), *Current Issues in Quantum Logic*, Vol. 8 of *Ettore Majorana International Science Series*, Plenum, New York, pp. 303–322, 1981.
- [Tamura, 1988] S. Tamura. A Gentzen formulation without the cut rule for ortholattices, *Kobe Journal of Mathematics* **5**, 133–150, 1988.
- [Varadarajan, 1985] V. S. Varadarajan. *Geometry of Quantum Theory*, 2 edn, Springer, Berlin, 1985.

MARTIN BUNDER

## COMBINATORS, PROOFS AND IMPLICATIONAL LOGICS

### 1 INTRODUCTION

In this chapter we first look at operators called **combinators**. These are very simple but extremely powerful. They provide a means of doing logic and mathematics without using variables, are powerful enough to allow the definition of all recursive functions and have more recently been used as a basis for certain “functional” computer languages.

We are interested in another use here which involves the **functional character** or **type** possessed by many combinators. Each type can be interpreted as a theorem of the intuitionistic implicative logic  $H_{\rightarrow}$  and combinators possessing that type can be interpreted as Hilbert-style proofs of that theorem. Weaker sets of combinators can be used to represent proofs in sublogics of  $H_{\rightarrow}$ , these include the substructural logics, such as the relevance logics  $R_{\rightarrow}$  and  $T_{\rightarrow}$ . There is a further interpretation of combinators and types as programs and specifications which we will not discuss here.

Next we look at lambda calculus. This also allows the definition of all recursive functions and has also been used in foundations of mathematics and computer language development. Many lambda terms also have types and these again are the theorems of  $H_{\rightarrow}$ . The lambda terms represent natural deduction style proofs of these theorems.

In the third section of this chapter we look at translations from combinators to lambda terms and vice versa. For the combinators and lambda terms that represent proofs in  $H_{\rightarrow}$  these translations are well known, for those corresponding to proofs in weaker logics they are quite new.

In a fourth section we develop a new algorithm which, given an implicational formula, allows us to find lambda terms representing natural deduction style proofs of the formula or demonstrates that the formula has no proof.

Most implicational substructural logics are specified by substructural rules or by axioms and not by rules in the natural deduction form. Our translation procedure, together with the algorithm, provides us with a simple constructive means of finding Hilbert-style proofs in many of these logics. As the translation procedure tells us which lambda terms are translatable into which sets of combinators, the algorithm can be directed to look only for the lambda terms of the appropriate kind. The algorithm is inherently finite; for any given formula, and for many logics, bounds for the proof searches can be written down.

The  $H_{\rightarrow}$  algorithm has been implemented by Anthony Dekker as Brouwer 7.9.0 (see [Dekker, 1996]) and, for any implicational formula, produces a  $\lambda$ -term proof (or even 50 alternative proofs) or a guarantee that there is no proof, virtually instantly. The implementation has more recently been extended by Martijn Oostdijk as LambdaCal2, (see [Oostdijk, 1996]) to cover the other implicational systems in this chapter as well as certain systems with other connectives. This implementation supplies combinator and lambda calculus proofs.

## 2 COMBINATORY LOGIC

Combinators are operators which manipulate arbitrary expressions by cancellation, duplication, bracketing and permutation. Combinators were first defined by Schönfinkel in his 1924 paper and rediscovered by Curry in [1930].

To illustrate their use we consider the following examples:

let  $Axy$  (rather than  $A(x, y)$ ) represent  $x + y$ . The commutative law for addition can then be written as

$$Axy = Ayx.$$

Given a combinator  $\mathbf{C}$  with the property:

$$\mathbf{C}xyz = xzy$$

this becomes

$$Axy = \mathbf{C}Axy$$

which could be simply written, without variables, as

$$A = \mathbf{C}A.$$

Given an identity combinator  $\mathbf{I}$ , i.e. one such that

$$\mathbf{I}x = x$$

we can write

$$0 + x = x$$

as

$$A0x = x$$

or

$$A0x = \mathbf{I}x$$

and so, without variables, as

$$A0 = \mathbf{I}.$$

$$x + 0 = x$$

would be

$$\mathbf{CA0} = \mathbf{I}.$$

Schönfinkel found that only two combinators,  $\mathbf{K}$  and  $\mathbf{S}$ , were enough to define all others.

We will now introduce these, other combinators, and our method of writing functional expressions (such as  $Axy$  rather than  $A(x, y)$ ) more formally.

### 2.1 Combinators and Application

DEFINITION 1 (Combinator).

1.  $\mathbf{K}$  and  $\mathbf{S}$  are combinators.
2. If  $X$  and  $Y$  are combinators so is  $(XY)$ .

(The operation in (2) is called **application**.)

Though it is possible, it is often not convenient to work without variables; we therefore introduce terms which are made up of combinators and variables using application. Other constants could also be included in (1) below.

DEFINITION 2 (Term).

1.  $\mathbf{K}$ ,  $\mathbf{S}$ ,  $x$ ,  $y$ ,  $z, \dots, x_1, x_2, \dots$  are terms
2. If  $X$  and  $Y$  are terms so is  $(XY)$ .

**Notation** We use association to the left for terms. This means that our  $Axy$  is short for  $((Ax)y)$ . A binary function over the real numbers such as  $A$  is therefore interpreted as a unary function from real numbers into the set of unary functions from real numbers to real numbers.

The process of going from  $\mathbf{CXYZ}$  to  $XZY$  or from  $\mathbf{IX}$  to  $X$  is called **reduction**. This is defined as follows:

DEFINITION 3 (Reduction). The relation  $X \triangleright Y$  ( $X$  reduces weakly to  $Y$ ) is defined as follows:

$$(\mathbf{K}) \quad \mathbf{KXY} \triangleright X$$

$$(\mathbf{S}) \quad \mathbf{SXYZ} \triangleright XZ(YZ)$$

$$(\rho) \quad X \triangleright X$$

$$(\mu) \quad X \triangleright Y \Rightarrow UX \triangleright UY$$

$$(\nu) \quad X \triangleright Y \Rightarrow XU \triangleright YU$$

$$(\tau) \quad X \triangleright Y \text{ and } Y \triangleright Z \Rightarrow X \triangleright Z$$

(**K**) and (**S**) are called the **reduction axioms** for **K** and **S**.

DEFINITION 4 (Weak equality).  $X = Y$  if this can be derived from the axioms and rules of Definition 3 with “=” instead of “ $\triangleright$ ”, together with

$$(\sigma) \quad X = Y \quad \Rightarrow \quad Y = X.$$

The formal system consisting of at least Definition 1 and the postulates in Definitions 3 we call **combinatory logic**. Other axioms and rules may be added.

We now show how the combinators we met earlier, and others, can be defined. We use “ $\equiv$ ” for “equals by definition”.

DEFINITION 5.

$$\begin{aligned} \mathbf{I} &\equiv \mathbf{SKK} \\ \mathbf{B} &\equiv \mathbf{S(KS)K} \\ \mathbf{C} &\equiv \mathbf{S(BBS)(KK)} \\ \mathbf{B}' &\equiv \mathbf{CB} \\ \mathbf{W} &\equiv \mathbf{SS(KI)} \\ \mathbf{S}' &\equiv \mathbf{B(BW)(BBB')} \end{aligned}$$

Each of these defined combinators has a characteristic reduction theorem:

THEOREM 6.

1.  $\mathbf{IX} \triangleright X$
2.  $\mathbf{BXYZ} \triangleright X(YZ)$
3.  $\mathbf{CXYZ} \triangleright XZY$
4.  $\mathbf{B'XYZ} \triangleright Y(XZ)$
5.  $\mathbf{WXY} \triangleright XYY$
6.  $\mathbf{S'XYZ} \triangleright YZ(XZ)$

**Proof.**

1.  $\mathbf{SKKX} \triangleright \mathbf{KX(KX)}$  by (**S**)
- so  $\mathbf{SKKX} \triangleright X$  by (**K**) and ( $\tau$ )
2.  $\mathbf{S(KS)KXYZ} \triangleright \mathbf{KSX(KX)YZ}$  by ( $\tau$ ) and ( $\nu$ )
- $\triangleright \mathbf{S(KX)YZ}$  by (**K**) and ( $\nu$ )
- $\triangleright \mathbf{KXZ(YZ)}$  by (**S**)
- $\triangleright X(YZ)$  by (**K**) and ( $\nu$ ).
- so  $\mathbf{S(KS)KXYZ} \triangleright X(YZ)$  by ( $\tau$ ) ■

If  $M(X_1, \dots, X_n)$  is a term made up by application using zero or more occurrences of each of  $X_1, \dots, X_n$ , we can find a combinator  $Z$  such that  $ZX_1X_2 \dots X_n \triangleright M(X_1, \dots, X_n)$ .



This property is called the “**combinatory completeness**” of the combinatory logic based on **K** and **S**.

The combinator  $Z$  above is represented by

$$Z \equiv [x_1]([x_2](\dots([x_n]M(x_1, \dots, x_n)) \dots))$$

where each  $[x_i](\dots)$  is called a **bracket abstraction**.

Bracket abstractions can be defined in various ways, a simple definition, involving **S** and **K**, is as follows:

DEFINITION 7 (Bracket abstraction  $[x_i]$ ).

- (i)  $[x_i] x_i \equiv \mathbf{I}$
- (k)  $[x_i] Y \equiv \mathbf{K}Y$  if  $x_i \notin Y$
- (η)  $[x_i]Yx_i \equiv Y$  if  $x_i \notin Y$
- (s)  $[x_i]YZ \equiv \mathbf{S}([x_i]Y)([x_i]Z)$ ,

where  $x_i \notin Y$  stands for  $x_i$  does not appear in  $Y$ .

The above clauses must be used in the order given i.e.  $(ik\eta s)$ . In the order  $(iks\eta)$ , we would always obtain  $\mathbf{S}([x_i]Y)\mathbf{I}$  for  $[x_i]Yx_i$ , if  $x_i \notin Y$ , instead of the simpler  $Y$ .

Repeated bracket abstraction as in  $[x_1]([x_2](\dots([x_n]M)\dots))$  we will write as  $[x_1, x_2, \dots, x_n]M$ .

EXAMPLE 8.

$$\begin{aligned} [x_1, x_2, x_3] x_3(x_1 x_3) &\equiv [x_1, x_2]\mathbf{S}([x_3]x_3)([x_3](x_1 x_3)) && \text{by (s)} \\ &\equiv [x_1, x_2]\mathbf{S}\mathbf{I}x_1 && \text{by (i) and (\eta)} \\ &\equiv [x_1]\mathbf{K}(\mathbf{S}\mathbf{I}x_1) && \text{by (k)} \\ &\equiv \mathbf{S}(\mathbf{K}\mathbf{K})(\mathbf{S}\mathbf{I}) && \text{by (k) and (\eta)}. \end{aligned}$$

$$\begin{aligned} \mathbf{S}(\mathbf{K}\mathbf{K})(\mathbf{S}\mathbf{I})x_1x_2x_3 &\triangleright \mathbf{K}\mathbf{K}x_1(\mathbf{S}\mathbf{I}x_1)x_2x_3 \\ &\triangleright \mathbf{K}(\mathbf{S}\mathbf{I}x_1)x_2x_3 \\ &\triangleright \mathbf{S}\mathbf{I}x_1x_3 \\ &\triangleright \mathbf{I}x_3(x_1x_3) \\ &\triangleright x_3(x_1x_3). \end{aligned}$$

Bracket abstraction has the following property which we call  $(\beta)$  for lambda abstraction in Section 3.

THEOREM 9.  $[x]X)Y \triangleright [Y/x]X$  where  $[Y/x]X$  is the result of substituting  $Y$  for all occurrences of  $x$  in  $X$ .

**Proof.** By a simple induction on the length of  $X$ . ■

From this theorem we get:

$$([x]X) x \triangleright X$$

and if  $x_1, \dots, x_n \notin X_1 X_2 \dots X_n$ ,

$$([x_1, \dots, x_n]X) X_1 \dots X_n \triangleright [X_1/x_1, \dots, X_n/x_n]X ,$$

where  $[X_1/x_1, \dots, X_n/x_n]X$  is the result of substituting simultaneously  $X_1$  for all occurrences of  $x_1$  in  $X$ ,  $\dots$ ,  $X_n$  for all occurrences of  $x_n$  in  $X$ .

## 2.2 Combinators, Types, Proofs and Theorems

If a term  $X$  is an element of a set  $\alpha$  (which we write as  $X \in \alpha$  or  $X : \alpha$  below) we have as

$$\begin{aligned} \mathbf{K}XY &= X, \\ \mathbf{K}XY &\in \alpha. \end{aligned}$$

If  $Y \in \beta$  we have that  $\mathbf{K}X$  is a function from the set  $\beta$  into the set  $\alpha$ , i.e. in the usual mathematical notation:

$$\mathbf{K}X : \beta \rightarrow \alpha ,$$

where  $\beta \rightarrow \alpha$  represents the set of all functions from  $\beta$  into  $\alpha$ .

From this it follows that  $\mathbf{K}$  is a function from  $\alpha$  into  $\beta \rightarrow \alpha$ , so we can write:

$$\mathbf{K} : \alpha \rightarrow (\beta \rightarrow \alpha) .$$

Sets such as the  $\alpha$ ,  $\beta$  and  $\alpha \rightarrow (\beta \rightarrow \alpha)$  above will be denoted by expressions called **types**. Types are defined as follows:

DEFINITION 10 (Types).

1.  $a, b, c, \dots$  are (atomic) types.
2. If  $\alpha$  and  $\beta$  are types so is  $(\alpha \rightarrow \beta)$ .

For types we use **association to the right**,  $\alpha \rightarrow (\beta \rightarrow \alpha)$  can therefore be written as  $\alpha \rightarrow \beta \rightarrow \alpha$ .

The type variables or atomic types can be interpreted as arbitrary sets, the compound types then represent sets of functions.

Above we arrived at  $\mathbf{K} : \alpha \rightarrow \beta \rightarrow \alpha$ ; such a derivation we call a **type assignment**, we call  $\alpha \rightarrow \beta \rightarrow \alpha$  the **type of  $\mathbf{K}$**  and  $\mathbf{K}$  an **inhabitant** of  $\alpha \rightarrow \beta \rightarrow \alpha$ .

Type assignments can be more formally derived from:

DEFINITION 11 (Type Assignment).

1. Variables can be assigned arbitrary types

2. If  $X : \alpha \rightarrow \beta$  and  $Y : \alpha$  then  $(XY) : \beta$ .
3. If  $Xx : \alpha$ ,  $x \notin X$  and  $x : \beta$  then  $X : \beta \rightarrow \alpha$ .

We will illustrate this by assigning a type to **S**.

If we let  $x : \alpha \rightarrow \beta \rightarrow \gamma$ ,  $y : \alpha \rightarrow \beta$  and  $z : \alpha$  we have by (2)  $xz : \beta \rightarrow \gamma$  and  $yz : \beta$  and so  $xz(yz) : \gamma$  which is, by (**S**),  $\mathbf{S}xyz : \gamma$ .

Now by (3)  $\mathbf{S}xy : \alpha \rightarrow \gamma$ ,  
 $\mathbf{S}x : (\alpha \rightarrow \beta) \rightarrow \alpha \rightarrow \gamma$   
 and  $\mathbf{S} : (\alpha \rightarrow \beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \beta) \rightarrow \alpha \rightarrow \gamma$ .

In particular we have

$$\mathbf{S} : (a \rightarrow b \rightarrow c) \rightarrow (a \rightarrow b) \rightarrow a \rightarrow c ,$$

and it can be seen, from the work above, that every type of **S** has to be a substitution instance of this. A type with this property we call the **principal type scheme** (PTS).

The PTS of **K** is given by

$$\mathbf{K} : a \rightarrow b \rightarrow a .$$

Notice that the types of **K** and **S** are exactly the axioms of  $H_{\rightarrow}$ , intuitionistic implicational logic, when  $\rightarrow$  is read as implication and  $\alpha$ ,  $\beta, \dots$  are read as well formed formulas or propositions.

Definition 11.2 and 3 can be rewritten as:

$$\rightarrow_e \frac{X : \alpha \rightarrow \beta \quad Y : \alpha}{XY : \beta}$$

and

$$\rightarrow_i \frac{\begin{array}{c} [x : \alpha] \\ \vdots \\ Xx : \beta \end{array}}{X : \alpha \rightarrow \beta} \quad (x \notin X)$$

We have, considering only the right hand sides of the  $:$ s, the rules of inference of a natural deduction formulation of  $H_{\rightarrow}$ .

If we take the types of **K** and **S** as axiom schemes and use only the parts of  $\rightarrow_e$  to the right of the  $:$ s, we have a Hilbert-style formulation of  $H_{\rightarrow}$ . What appears on the left hand side of the final step in such a proof gives us a unique representation of a proof of the theorem expressed on the right of the  $:$ .

EXAMPLE 12. If we abbreviate  $(\beta \rightarrow \gamma \rightarrow \delta) \rightarrow (\beta \rightarrow \gamma) \rightarrow \beta \rightarrow \delta$  by  $\alpha$  we have:

$$\begin{aligned}
\mathbf{K} &: \alpha \rightarrow (\gamma \rightarrow \delta) \rightarrow \alpha \\
\mathbf{S} &: \alpha \\
\mathbf{KS} &: (\gamma \rightarrow \delta) \rightarrow \alpha \\
\mathbf{S} &: ((\gamma \rightarrow \delta) \rightarrow \alpha) \rightarrow ((\gamma \rightarrow \delta) \rightarrow \beta \rightarrow \gamma \rightarrow \delta) \\
&\quad \rightarrow (\gamma \rightarrow \delta) \rightarrow (\beta \rightarrow \gamma) \rightarrow \beta \rightarrow \delta \\
\mathbf{S(KS)} &: ((\gamma \rightarrow \delta) \rightarrow \beta \rightarrow \gamma \rightarrow \delta) \rightarrow (\gamma \rightarrow \delta) \rightarrow (\beta \rightarrow \gamma) \rightarrow \beta \rightarrow \delta \\
\mathbf{K} &: (\gamma \rightarrow \delta) \rightarrow \beta \rightarrow \gamma \rightarrow \delta \\
\mathbf{S(KS)K} &: (\gamma \rightarrow \delta) \rightarrow (\beta \rightarrow \gamma) \rightarrow \beta \rightarrow \delta
\end{aligned}$$

We note that each  $\mathbf{S}$  or  $\mathbf{K}$  in  $\mathbf{S(KS)K}$  represents the use of an axiom and each application a use of  $\rightarrow_e$ . We also note that the above represents a type for the combinator  $\mathbf{B}$  of Definition 5.

Some combinators do not have types, for example if we want to find a type for  $\mathbf{SSS}$  we would proceed as follows:

$$\begin{aligned}
\text{let } \mathbf{S} &: (\beta \rightarrow \gamma \rightarrow \delta) \rightarrow (\beta \rightarrow \gamma) \rightarrow \beta \rightarrow \delta \\
\text{and } \mathbf{S} &: (\tau \rightarrow \rho \rightarrow \sigma) \rightarrow (\tau \rightarrow \rho) \rightarrow \tau \rightarrow \sigma
\end{aligned}$$

then we have putting  $\tau \rightarrow \rho \rightarrow \sigma = \beta$ ,  $\tau \rightarrow \rho = \gamma$  and  $\tau \rightarrow \sigma = \delta$ :

$$\mathbf{SS} : ((\tau \rightarrow \rho \rightarrow \sigma) \rightarrow \tau \rightarrow \rho) \rightarrow (\tau \rightarrow \rho \rightarrow \sigma) \rightarrow \tau \rightarrow \sigma .$$

Now with  $\mathbf{S} : (\mu \rightarrow \nu \rightarrow \xi) \rightarrow (\mu \rightarrow \nu) \rightarrow \mu \rightarrow \xi$ , we would need, to assign a type to  $\mathbf{SSS}$ :

$$\begin{aligned}
\tau \rightarrow \rho \rightarrow \sigma &= \mu \rightarrow \nu \rightarrow \xi , \\
\tau &= \mu \rightarrow \nu , \\
\text{and } \rho &= \mu \rightarrow \xi .
\end{aligned}$$

However this requires both  $\tau = \mu$  and  $\tau = \mu \rightarrow \nu$  which is impossible.

Also there are types that have no inhabitant, for example  $a \rightarrow b$  and  $((a \rightarrow b) \rightarrow a) \rightarrow a$ . In fact:

THEOREM 13.

1. If a type  $\tau$  has a combinator inhabitant,  $\tau$  is a theorem of the intuitionistic implicative logic  $H_{\rightarrow}$ .
2. If  $\tau$  is a theorem of  $H_{\rightarrow}$ , then it is the type of a combinator.

**Proof.**

1. It can be proved by an easy induction on the length of  $X$  that

$$x_1 : \alpha_1, \dots, x_n : \alpha_n \vdash X : \tau$$

implies

$$\alpha_1, \dots, \alpha_n \vdash \tau$$

2. It can be proved by an induction on the length of the deduction leading to

$$\alpha_1, \dots, \alpha_n \vdash \tau$$

that there are variables  $x_{11}, \dots, x_{1m_1}, \dots, x_{nm_n}$  and a term  $X$  such that

$$FV(X) \subseteq \{x_{11}, \dots, x_{nm_n}\}$$

and

$$x_{11} : \alpha_1, \dots, x_{1m_1} : \alpha_1, x_{21} : \alpha_2, \dots, x_{nm_n} : \alpha_n \vdash X : \alpha .$$

For more details on this see Hindley [1997, Section 6B2 and Section 6B5].

■

The isomorphism between inhabitants and types and proofs and theorems of  $H_{\rightarrow}$ , which can be extended to fit programs and specifications, is called the **Curry-Howard or Formulas-as types isomorphism**.

Curry was the first to recognise the relation between types and theorems of  $H_{\rightarrow}$  (see [Curry and Feys, 1958]). The idea was taken up and extended to other connectives and quantifiers in Lauchli [1965; 1970], Howard [1980] (but written in 1969), de Bruin [1970; 1980] and Scott [1970]. Recently it was extended to include a large amount of mathematics in Crossley and Shepherdson [1993].

### 2.3 Types and Weaker Logics

If we consider an arbitrary set of combinators  $Q$ , we can define the set of  **$Q$ -combinators** and  **$Q$ -terms** as follows:

DEFINITION 14 ( $Q$ -terms).

1. Elements of  $Q$  and  $x, y, z, \dots, x_1, x_2, \dots$  are  $Q$ -terms
2. If  $X$  and  $Y$  are  $Q$ -terms so is  $(XY)$ .

DEFINITION 15 ( $Q$ -combinators). A  $Q$ -term containing no variables is a  $Q$ -combinator.

The formal system consisting of at least Definition 15 and the postulates of Definition 3, with the reduction axioms (**K**) and (**S**) replaced by ones appropriate to  $Q$ , is called  $Q$ -**combinatory logic**.  $Q$ -**logic** is the implicational logic whose axioms are the types of the combinators in  $Q$  and whose rule is  $\rightarrow_e$ .

Thus combinatory logic, as defined before, is  $\{\mathbf{S}, \mathbf{K}\}$ -(or simply **SK**-) combinatory logic, terms are **SK**-terms and combinators are **SK**-combinators.

DEFINITION 16 (Weaker sets of combinators). A set  $Q_1$  of combinators is said to be **weaker** than a set  $Q_2$ , if for every  $Q_1$ -combinator  $X$  there is a  $Q_2$ -combinator  $Y$  with the same reduction theorem. (In that case we say  $X$  is  $Q_2$ -**definable**.) Also there must be a  $Q_2$ -combinator which is not  $Q_1$ -definable.

DEFINITION 17 (Weaker combinatory logics).  $Q_1$ -combinatory logic is **weaker than**  $Q_2$ -combinatory logic if  $Q_1$  is weaker than  $Q_2$ .

**BCKW**-combinatory logic is just as strong as **SK**-combinatory logic as **B, C, K** and **W** are (by Definition 5) **SK**-definable and **S** is also **BCKW**-definable ( $\mathbf{S} \equiv \mathbf{B}(\mathbf{B}\mathbf{W})(\mathbf{B}\mathbf{C}(\mathbf{B}\mathbf{B}))$ ).

**BCW** and **BCIW**-combinatory logics are both weaker than **SK**-combinatory logic as **S** is not definable using **B, C, I** and **W**.

It is clear that **BCK**- and **BCIW**-combinatory logics are not combinatorially complete.

The set of types of the **BCK**-combinators can easily be seen to be the set of theorems generated by  $\rightarrow_e$  using the types of **B, C** and **K**. This we call **BCK**-(implicational) logic, which is a subsystem of  $H_{\rightarrow}$ .

In general, if all the combinators of  $Q_1 \cup Q_2$  have types and  $Q_1$  is weaker than  $Q_2$  then  $Q_1$ -logic is weaker than  $Q_2$ -logic.

As before, given a  $Q$ -combinator, a  $Q$ -theorem and its proof can be read off.

EXAMPLE 18. Determine the type of **BC(BK)** and the **BCK**-proof of this as a **BCK**-theorem.

$$\begin{aligned}
 \mathbf{B} &: (\alpha \rightarrow \beta) \rightarrow (\gamma \rightarrow \alpha) \rightarrow \gamma \rightarrow \beta \\
 \mathbf{C} &: (\delta \rightarrow \tau \rightarrow \sigma) \rightarrow \tau \rightarrow \delta \rightarrow \sigma \\
 \mathbf{BC} &: (\gamma \rightarrow \delta \rightarrow \tau \rightarrow \sigma) \rightarrow \gamma \rightarrow \tau \rightarrow \delta \rightarrow \sigma \\
 &\quad (\text{Here } \alpha = \delta \rightarrow \tau \rightarrow \sigma, \beta = \tau \rightarrow \delta \rightarrow \sigma .) \\
 \mathbf{K} &: \mu \rightarrow \nu \rightarrow \mu \\
 \mathbf{BK} &: (\xi \rightarrow \mu) \rightarrow \xi \rightarrow \nu \rightarrow \mu \\
 &\quad (\text{Here } \alpha = \mu, \beta = \nu \rightarrow \mu \text{ and } \gamma = \xi .) \\
 \mathbf{BC(BK)} &: (\xi \rightarrow \mu) \rightarrow \nu \rightarrow \xi \rightarrow \mu \\
 &\quad (\text{Here } \gamma = \xi \rightarrow \mu, \delta = \xi, \tau = \nu \text{ and } \sigma = \mu)
 \end{aligned}$$

## 2.4 Combinator Reduction and Proof Reduction

We illustrate here what happens to (part of) a proof represented by a combinator  $\mathbf{KXY}$  or  $\mathbf{SXYZ}$  when this combinator is reduced by  $(\mathbf{K})$  or  $(\mathbf{S})$ .

The original proof involving  $\mathbf{KXY}$  must look like:

$$\frac{\frac{\frac{\mathbf{K} : \alpha \rightarrow \beta \rightarrow \alpha \quad X : \alpha \quad D_1}{\mathbf{KX} : \beta \rightarrow \alpha} \quad D_2}{\mathbf{KXY} : \alpha} \quad Y : \beta}{D_3}$$

with  $D_1$ ,  $D_2$  and  $D_3$  representing other proof steps.

With the reduction of  $\mathbf{KXY}$  to  $X$  the proof reduces (or normalises) to:

$$\frac{D_1}{X : \alpha} \quad D_3$$

If the proof involving  $\mathbf{SXYZ}$  is:

$$\frac{\frac{D_2 \quad \frac{\mathbf{S} : (\alpha \rightarrow \beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \beta) \rightarrow \alpha \rightarrow \gamma \quad X : \alpha \rightarrow \beta \rightarrow \gamma \quad D_1}{\mathbf{SX} : (\alpha \rightarrow \beta) \rightarrow \alpha \rightarrow \gamma}}{\mathbf{SXY} : \alpha \rightarrow \gamma} \quad D_3}{\mathbf{SXYZ} : \gamma} \quad Z : \alpha}{D_4}$$

With the reduction of  $\mathbf{SXYZ}$  to  $XZ(YZ)$  the proof reduces (or normalises) to:

$$\frac{\frac{\frac{D_1}{X : \alpha \rightarrow \beta \rightarrow \gamma} \quad D_3}{XZ : \beta \rightarrow \gamma} \quad \frac{\frac{D_2}{Y : \alpha \rightarrow \beta} \quad D_3}{YZ : \beta}}{XZ(YZ) : \gamma} \quad D_4$$

It may be, by the way, if  $D_3$  is a particularly long part of the proof, that the “reduced proof” is actually longer than the original. In the same way, if  $Z$  is long,  $XZ(YZ)$  may be longer than  $\mathbf{SXYZ}$ .

**DEFINITION 19.** If a term has no subterms of the form  $\mathbf{KXY}$  or  $\mathbf{SXYZ}$ , the term is said to be in **normal form**.

Not every combinator has a normal form, for example  $\mathbf{WI(WI)}$  and  $\mathbf{WW(WW)}$  do not, but every combinator that has a type also has a normal form.

A combinator in normal form will represent a normalised proof, these however are not unique. For example:

$$\mathbf{K} : a \rightarrow a \rightarrow a$$

and also

$$\mathbf{KI} : a \rightarrow a \rightarrow a.$$

### 3 LAMBDA CALCULUS

Lambda calculus, like combinatory logic, provides a means of representing all recursive functions. It is, these days, much used as the basis for functional computer languages. Extensions of the typed lambda calculus we introduce in Section 3.2 below, also have applications in program verification.

Lambda calculus was first developed by Church in the early 30s (see [Church, 1932; Church, 1933]) as part of a foundation of logic and mathematics. This was also the aim of Curry's "illative combinatory logic", but both Church's and Curry's extended systems proved to be inconsistent.

The use of the lambda calculus notation is best seen through an example such as the following:

If the value of the sin function at  $x$  is  $\sin x$  and the value of the log function at  $x$  is  $\log x$ , what is the function whose value at  $x$  is  $x^2$ ? Usually this function is also called  $x^2$ . The lambda calculus allows us to eliminate this ambiguity by using  $\lambda x.x^2$  for the name of the function.

#### 3.1 *Lambda terms and lambda reductions*

We will now set up the system more formally:

DEFINITION 20 (Lambda terms or  $\lambda$ -terms).

1. Variables are  $\lambda$ -terms.
2. If  $X$  is a  $\lambda$ -term and  $x$  a variable then, the **abstraction of  $X$  with respect to  $x$** ,  $(\lambda x.X)$  is a  $\lambda$ -term.
3. If  $X$  and  $Y$  are  $\lambda$ -terms then  $(XY)$ , the **application** of  $X$  to  $Y$ , is a  $\lambda$ -term.

$(\lambda x.X)$  is interpreted as the **function** whose value at  $x$  is  $X$ .

Given this we would expect the following to hold:

EXAMPLE 21.

$$\begin{aligned} (\lambda x. \sin x) &= \sin \\ ((\lambda x. \sin x)x) &= \sin x \\ ((\lambda x. \sin x)\pi) &= \sin \pi \\ ((\lambda x.x^2)2) &= 4 \\ ((\lambda x.2)x) &= 2 \end{aligned}$$



$(\lambda x.2)$  is the constant function whose value is 2.

For terms formed by application we use association to the left as for terms of combinatory logic. Repeated  $\lambda$ -abstraction as in  $(\lambda x_1.(\lambda x_2 \dots (\lambda x_n.X) \dots))$  we abbreviate to  $\lambda x_1.\lambda x_2 \dots \lambda x_n.X$  or to  $\lambda x_1 x_2 \dots x_n.X$ .

Note that while  $\lambda x_1 x_2 \dots x_n.X$  represents a function of  $n$  variables, it is also a function of one variable whose value,  $\lambda x_2 x_3 \dots x_n.X$  at  $x_1$ , is also a function (if  $n \geq 2$ ).

The process of simplifying  $(\lambda x. \sin x)x$  to  $\sin x$  or  $(\lambda x.x^2)2$  to  $2^2$  is called  $\beta$ -reduction. To explain this we need to define free and bound variables and, using these, a substitution operator.

As in combinatory logic we use  $\equiv$  for equality by definition or identity.

DEFINITION 22 (Free and bound variables, closed terms).

1.  $x$  is a **free** variable in  $x$ .
2. If  $x$  is **free** in  $Y$  or  $Z$  then  $x$  is free in  $(YZ)$ .
3. If  $x$  is **free** in  $Y$  and  $y \neq x$ ,  $x$  is free in  $\lambda y.Y$ .
4. Every  $x$  that appears in  $\lambda x.Y$  is **bound** in  $\lambda x.Y$ .

We write  $FV(X)$  for the set of free variables of  $X$ .

A **closed term** is one without free variables.

EXAMPLE 23.

1. If  $X \equiv \lambda xy.xyx$ ,  $X$  is a closed term and  $x$  and  $y$  are bound in  $X$ .
2. If  $X \equiv \lambda xy.xyzx(\lambda u.zx)w$ ,  $x$ ,  $y$  and  $u$  are bound in  $X$  and  $FV(X) = \{z, w\}$
3. If  $X \equiv x(\lambda x.xy)x$ , the first and last occurrences of  $x$  are **free occurrences** of  $x$ . Both the occurrences of  $x$  in  $\lambda x.xy$  are **bound**.

DEFINITION 24. ( $[Y/x]X$  - the result of substituting  $Y$  for all free occurrences of  $x$  in  $X$ )

1.  $[Y/x]x \equiv Y$
2.  $[Y/x]y \equiv y$  if  $y$  is an atom,  $x \neq y$
3.  $[Y/x](WZ) \equiv ([Y/x]W)([Y/x]Z)$
4.  $[Y/x](\lambda x.Z) \equiv \lambda x.Z$
5.  $[Y/x](\lambda y.Z) \equiv \lambda y.[Y/x]Z$  if  $y \neq x$   
and,  $y \notin FV(Y)$  or  $x \notin FV(Z)$ .

6.  $[Y/x]\lambda y.Z \equiv \lambda z.[Y/x][z/y]Z$   
if  $y \neq x$ ,  $y \in FV(Y)$ ,  $x \in FV(Z)$  and  $z \notin FV(YZ)$ .

As we saw in Example 21, a  $\lambda$ -term of the form  $(\lambda x.X)Y$  can be simplified or “**reduced**”. This kind of reduction is specified by the following axiom:

$$(\beta) (\lambda x.X)Y \triangleright [Y/x]X .$$

A  $\lambda$ -term of the form  $(\lambda x.X)Y$  is called a  $\beta$ -**redex**.

The following axiom allows a change of bound variables and is called an  $\alpha$ -reduction.

$$(\alpha) \lambda x.X \triangleright \lambda y.[y/x]X \quad \text{if } y \notin FV(X).$$

Rules  $(\mu)$ ,  $(\nu)$  and  $(\tau)$  of Definition 3 together with:

$$(\xi) X \triangleright Y \Rightarrow \lambda x.X \triangleright \lambda x.Y ,$$

allow us to perform  $\alpha$ - and  $\beta$ -reductions within a term.

DEFINITION 25 ( $\beta$ - and  $\beta\eta$ -reduction).

The reduction  $\triangleright$  specified by  $(\alpha)$ ,  $(\beta)$ ,  $(\rho)$ ,  $(\mu)$ ,  $(\nu)$ ,  $(\tau)$  and  $(\xi)$  is called  $\beta$ -**reduction**.

This becomes  $\beta\eta$ - (or just  $\eta$ -) **reduction** if the following axiom is added:

$$(\eta) \lambda x.Xx \triangleright X \quad \text{if } x \notin FV(X).$$

We denote the two forms of reduction, when we wish to distinguish them, by  $\triangleright_\beta$  and  $\triangleright_{\beta\eta}$ . If in a  $\beta\eta$ -reduction  $(\beta)$  is not used we sometimes write  $\triangleright_\eta$  instead of  $\triangleright_{\beta\eta}$ .

We will call  $\alpha$ -,  $\beta$ - and  $\eta$ - reductions  $\lambda$ -reductions to distinguish them from combinator reductions.

EXAMPLE 26.

$$1. \lambda x_1.x_1((\lambda x_2.x_3x_2x_2)x_1) \triangleright_\beta \lambda x_1.x_1(x_3x_1x_1)$$

$$\begin{aligned} 2. & \lambda x_3.(\lambda x_1.x_1x_2(\lambda x_2.x_3x_1x_2))(\lambda x_1.x_1) \\ & \triangleright_\eta \lambda x_3.(\lambda x_1.x_1x_2(x_3x_1))(\lambda x_1.x_1) \\ & \triangleright_\beta \lambda x_3.(\lambda x_1.x_1)x_2(x_3(\lambda x_1.x_1)) \\ & \triangleright_\beta \lambda x_3.x_2(x_3(\lambda x_1.x_1)) \end{aligned}$$

$$\begin{aligned} \text{so } & \lambda x_3.(\lambda x_1.x_1x_2(\lambda x_2.x_3x_1x_2))(\lambda x_1.x_1) \\ & \triangleright_{\beta\eta} \lambda x_3.x_2(x_3(\lambda x_1.x_1)) \end{aligned}$$

A term of the form  $\lambda x.Xx$  where  $x \notin FV(X)$ , is called an  $\eta$ -**redex**.

A term with no  $\beta$ -redexes is said to be in  $\beta$ -**normal form**. One without  $\beta$ - and  $\eta$ - redexes is in  $\beta\eta$ -**normal form**. One without  $\eta$ -redexes is said to

be in  $\eta$ -normal form.  $\beta$ -,  $\eta$ - and  $\beta\eta$ -normal forms are unique (see [Hindley and Seldin, 1986]).

DEFINITION 27 ( $\alpha$ -,  $\beta$ - and  $\beta\eta$ -equality).

The relation  $=_\alpha$  is specified by  $(\alpha)$ ,  $(\rho)$ ,  $(\mu)$ ,  $(\nu)$ ,  $(\tau)$  and  $(\xi)$ , all with  $=_\alpha$  for  $\triangleright$  and

$$(\sigma) X =_\alpha Y \Rightarrow Y =_\alpha X.$$

The relation  $=_\beta$  is specified by all the above postulates, as well as  $(\beta)$ , with  $=_\beta$  for  $\triangleright$  and  $=_\alpha$ .

The relation  $=_{\beta\eta}$  is specified by all the  $=_\beta$  postulates, as well as  $(\eta)$ , with  $=_{\beta\eta}$  replacing  $=_\beta$  and  $\triangleright$ .

Note that the weak equality of combinatory logic obeys all the above postulates (with  $[.]$  for  $\lambda.$ ) except  $(\xi)$  and  $(\eta)$ . It is possible to extend weak equality by means of some extra equations involving combinators to make  $(\xi)$  and/or  $(\eta)$  admissible (see [Curry and Feys, 1958, Section 6C4]). We will call the corresponding equalities for combinatory logic  $=_\beta$  and  $=_{\beta\eta}$  respectively.

### 3.2 Lambda Terms, Types, Proofs and Theorems

If a term  $X$  is an element of a set  $\alpha$  and the variable  $y$  is in a set  $\beta$ ,  $\lambda y.X$  will represent a function from  $\beta$  into  $\alpha$  i.e.  $\lambda y.X : \beta \rightarrow \alpha$ . We will denote sets such as these by the types introduced in Definition 10. We assign types to  $\lambda$ -terms in a similar way to the assignment to combinators.

DEFINITION 28 (Type Assignment).

1. Variables can be assigned arbitrary types.
2. If  $X : \alpha \rightarrow \beta$  and  $Y : \alpha$  then  $(XY) : \beta$
3. If  $X : \alpha$  and  $y : \beta$  then  $\lambda y.X : \beta \rightarrow \alpha$ .

EXAMPLE 29.

1. If  $x_2 : a \rightarrow a \rightarrow b$ ,  $x_3 : a$  then  $x_2 x_3 x_3 : b$  and  $\lambda x_3. x_2 x_3 x_3 : a \rightarrow b$ .  
If  $x_1 : (a \rightarrow b) \rightarrow c$  then  $x_1 (\lambda x_3. x_2 x_3 x_3) : c$ , so  $\lambda x_1 x_2. x_1 (\lambda x_3. x_2 x_3 x_3) : ((a \rightarrow b) \rightarrow c) \rightarrow (a \rightarrow a \rightarrow b) \rightarrow c$ .
2. If  $x_1 : \alpha \rightarrow \beta$  then for  $x_1 x_1$  to have a type we must have  $\alpha = \alpha \rightarrow \beta$  which is impossible. Hence  $x_1 x_1$  and so  $\lambda x_1. x_1 x_1$  have no types.

DEFINITION 30. If  $X$  is a closed lambda term and  $\vdash X : \alpha$  is derived by the type assignment rules then  $X$  is said to be an **inhabitant** of  $\alpha$  and  $\alpha$  a **type** of  $X$ .



If  $X$  is an atom,  $X = x_i$  for some  $i$ ,  $1 \leq i \leq n$  and  $\beta = \alpha_i$ , so (b) holds.

If (a) is obtained from

$$x_1 : \alpha_1, \dots, x_n : \alpha_n, x_{n+1} : \alpha_{n+1} \vdash Y : \alpha,$$

where  $\beta = \alpha_{n+1} \rightarrow \alpha$  and  $X = \lambda x_{n+1}.Y$  then by the induction hypothesis

$$\alpha_1, \dots, \alpha_n, \alpha_{n+1} \vdash \alpha$$

is valid in  $H_{\rightarrow}$  and so also (b).

$$\begin{aligned} \text{If } X \equiv UV \text{ where} \\ x_{i_1} : \alpha_{i_1}, \dots, x_{i_k} : \alpha_{i_k} \vdash U : \alpha \rightarrow \beta, \\ x_{j_1} : \alpha_{j_1}, \dots, x_{j_m} : \alpha_{j_m} \vdash V : \alpha \\ \text{and } \{x_{i_1} : \alpha_{i_1}, \dots, x_{i_n} : \alpha_{i_n}\} \cup \{x_{j_1} : \alpha_{j_1}, \dots, x_{j_m} : \alpha_{j_m}\} \\ = \{x_1 : \alpha_1, \dots, x_n : \alpha_n\} \end{aligned}$$

we have by the induction hypothesis in  $H_{\rightarrow}$ :

$$\alpha_{i_1}, \dots, \alpha_{i_k} \vdash \alpha \rightarrow \beta$$

and

$$\alpha_{j_1}, \dots, \alpha_{j_m} \vdash \alpha$$

and so (b).

2. We show by induction on the length of a proof in  $H_{\rightarrow}$  of

$$\alpha_1, \dots, \alpha_n \vdash \beta \tag{c}$$

that there is a term  $X$  with  $FV(X) \subseteq \{x_1, \dots, x_n\}$  such that

$$x_1 : \alpha_1, \dots, x_n : \alpha_n \vdash X : \beta \tag{d}$$

If  $\beta$  is one of the  $\alpha_i$ s this is obvious with  $X = x_i$ .

If (c) is obtained by the  $\rightarrow_i$  rule from

$$\alpha_1, \dots, \alpha_n, \gamma \vdash \delta$$

where  $\beta = \gamma \rightarrow \delta$ , then by the induction hypothesis we have

$$x_1 : \alpha_1, \dots, x_n : \alpha_n, x_{n+1} : \gamma \vdash Y : \delta$$

where  $FV(Y) \subseteq \{x_1, \dots, x_n, x_{n+1}\}$ .

Then (d) follows with  $X = \lambda x_{n+1}.Y$ .

If (c) is obtained from

$$\begin{array}{l} \alpha_{i_1} \dots, \alpha_{i_k} \vdash \alpha \rightarrow \beta \\ \text{and} \quad \alpha_{j_1} \dots, \alpha_{j_m} \vdash \alpha \\ \text{where} \quad \{\alpha_{i_1}, \dots, \alpha_{i_k}\} \cup \{\alpha_{j_1}, \dots, \alpha_{j_m}\} = \{\alpha_1, \dots, \alpha_n\} \end{array}$$

We have, by the induction hypothesis:

$$\begin{array}{l} x_{i_1} : \alpha_{i_1}, \dots, x_{i_k} : \alpha_{i_k} \vdash Y : \alpha \rightarrow \beta \\ x_{j_1} : \alpha_{j_1}, \dots, x_{j_m} : \alpha_{j_m} \vdash Z : \alpha \end{array}$$

where  $FV(Y) \subseteq \{x_{i_1}, \dots, x_{i_k}\}$  and  $FV(Z) \subseteq \{x_{j_1}, \dots, x_{j_m}\}$ .

(d) then follows with  $X = YZ$ . ■

### 3.3 Long Normal Forms

In Section 5 we will develop an algorithm, which, when given a type, will produce an inhabitant of this type if it has one. The inhabitant that is produced is in **long normal form**. This is defined below.

**DEFINITION 33** (Long normal form). A typed  $\lambda$ -term  $\lambda x_1 \dots x_n. x_i X_1 \dots X_k$  ( $n \geq 0$ ,  $k \geq 0$ ) is said to be in **long normal form** (lnf) if  $X_1, \dots, X_k$  are in lnf and have types  $\tau_1, \dots, \tau_k$  and  $x_i$  has type  $\tau_1 \rightarrow \dots \rightarrow \tau_k \rightarrow a$  where  $a$  is an atom.

**THEOREM 34.** *If  $X$  is a  $\lambda$ -term such that  $\vdash X : \alpha$ , then there is a term  $Y$  in lnf such that  $\vdash Y : \alpha$  and  $Y \triangleright_\eta X$ .*

**Proof.** By induction on the number of parts  $x_i X_1 \dots X_k$  of  $X$  that are not in lnf and are not the initial part of a term  $x_i X_1 \dots X_m$ , with  $m > k$ .

Consider the shortest of these.  $X_1, \dots, X_k$  must then be in lnf and if each  $X_j$  has type  $\tau_j$ ,  $x_i$  must have type  $\tau_1 \rightarrow \dots \rightarrow \tau_m \rightarrow a$  where  $m > k$ .

Let  $x_{p+1}, x_{p+2}, \dots, x_{p+m-k}$  be variables not free in  $x_i X_1 \dots X_k$  with types  $\tau_{k+1}, \tau_{k+2}, \dots, \tau_m$  respectively.

Then  $x_i X_1 \dots X_k x_{p+1} \dots x_{p+m-k}$  has type  $a$  and  $\lambda x_{p+1} \dots x_{p+m-k}. x_i X_1 \dots X_k x_{p+1} \dots x_{p+m-k}$  has type  $\tau_{k+1} \dots \rightarrow \tau_m \rightarrow a$ , the same as  $x_i X_1 \dots X_k$ .

This new term with the same type as  $x_i X_1, \dots, X_k$  is in lnf, so when it replaces  $x_i X_1 \dots X_k$  in  $X$ , there is one fewer part not in lnf.

Hence  $X$  can be expanded to a term  $Y$  in lnf such that  $Y \triangleright_\eta X$ . The types of the parts of  $X$  being changed are not affected so the type of  $Y$  will be the same as the type of  $X$ . ■

### 3.4 Lambda Reductions and Proof Reductions

We illustrate here what happens to (a part of) a proof represented by a  $\lambda$ -term  $(\lambda x.X)Y$  or  $\lambda y.Xy$  with  $y \notin FV(X)$  when this is reduced by  $(\beta)$  or  $(\eta)$ .

$$\frac{\frac{\frac{(\lambda)}{x : \alpha} D_1}{X : \beta} \quad - (1)}{\lambda x.X : \alpha \rightarrow \beta} \quad \frac{D_2}{Y : \alpha}}{(\lambda x.X)Y : \beta} D_3$$

reduces to:

$$\frac{\frac{D_2}{Y : \alpha} \quad [Y/x]D_1}{[Y/x]X : \beta} \quad D_3}{[[Y/x]X / (\lambda x.X)Y] D_3}$$

and

$$\rightarrow_e \frac{\frac{\frac{D_1}{X : \alpha \rightarrow \beta} \quad (\lambda)}{Xx : \beta} \quad x : \alpha}{\rightarrow_i \frac{\lambda x.Xx : \alpha \rightarrow \beta}{} - (1)} \quad (x \notin FV(X))}{D_2}$$

reduces to

$$\frac{D_1}{X : \alpha \rightarrow \beta} \quad [X / \lambda x.Xx] D_2$$

An expansion of a typed  $\lambda$ -term to Inf reverses the second reduction.

The effect of these reductions on a type assignment shown by the above is stated in the theorem below.

**THEOREM 35 (Subject Reduction Theorem).** *If*

$$x_1 : \tau_1, \dots, x_n : \tau_n \vdash X : \alpha$$

and

$$X \triangleright_{\beta\eta} Y \quad ,$$

then

$$x_1 : \tau_1, \dots, x_n : \tau_n \vdash Y : \alpha$$

**Proof.** A full proof of this is lengthy. A good outline appears in [Hindley and Seldin, 1986, Chapter 15]. ■

The converse of this is not true, for example  $(\lambda uv.v)(\lambda x.xx)$  has no type but it  $\beta$ -reduces to  $\lambda v.v$  which has type  $a \rightarrow a$ .

$\lambda xyz.(\lambda u.y)(xz)$  has type  $(c \rightarrow d) \rightarrow b \rightarrow c \rightarrow b$ , but not type  $a \rightarrow b \rightarrow c \rightarrow b$ , it reduces to  $\lambda xyz.y$  which has type  $a \rightarrow b \rightarrow c \rightarrow b$ .

The  $\eta$ -expansion used in forming a lnf in the proof of Theorem 34, illustrates the following limited form of the Subject Expansion Theorem.

**THEOREM 36.** *If*

$$x_1 : \tau_1, \dots, x_n : \tau_n \vdash X : \alpha \rightarrow \beta$$

*then*

$$x_1 : \tau_1, \dots, x_n : \tau_n \vdash \lambda x.Xx : \alpha \rightarrow \beta$$

*if*  $x \notin FV(X)$ .

**Proof.** See Hindley and Seldin [1986, Chapter 15]. ■

**References** Much more detail on the work in this section can be found in [Hindley, 1997, Chapter 2]. The system  $TA_\lambda$  discussed there is effectively the system we have introduced. See also [Barendregt, 1984, Appendix A].

## 4 TRANSLATIONS

As  $\lambda$ -terms and combinators describe the same set of (recursive) functions it is not surprising that for each  $\lambda$ -term there is a combinator representing the same function and, in a simple case, with the same reduction theorems (as in **(K)**, **(S)** or Theorem 6) and the same types.

For every combinator there is also a similar  $\lambda$ -term given by:

**DEFINITION 37** ( $(X_\lambda)$ ).

$$\begin{aligned} \mathbf{S}_\lambda &\equiv \lambda xyz.xz(yz) \\ \mathbf{K}_\lambda &\equiv \lambda xy.x \\ (XY)_\lambda &\equiv X_\lambda Y_\lambda. \end{aligned}$$

Any  $\lambda$ -term  $X$  can be translated into a combinator by taking parts of  $X$  of the form  $\lambda x_i \dots x_j.Y$  where  $Y$  contains no  $\lambda x_k$ s and changing these to  $[x_i, \dots, x_j]Y$  as defined in Definition 7, then further parts of the form  $\lambda x_i \dots x_j.Y$  can be changed in the same way until there are no  $\lambda x_k$ s left. If we call this translation  $*$  we might hope to have

$$X_{\lambda*} \equiv X$$



and

$$Y_{*\lambda} \equiv Y$$

for any combinator  $X$  and  $\lambda$ -term  $Y$ .

The former identity holds, but the latter one does not, in general.

EXAMPLE 38.

1.  $(\mathbf{SKK})_{\lambda*} \equiv (\lambda xyz.xz(yz))_*(\lambda uv.u)_*(\lambda ts.t)_*$   
 $\equiv ([x, y, z].xz(yz))([u.v]u)([t, s]t)$   
 $\equiv ([x, y].\mathbf{S}xy)([u].\mathbf{K}u)([t].\mathbf{K}t) \equiv \mathbf{SKK}$
2.  $(z(\lambda xyz.xyyz))_{*\lambda} \equiv z([x, y, z].xyyz)_\lambda$   
 $\equiv z([x, y].xyy)_\lambda \equiv z([x].\mathbf{S}x\mathbf{I})_\lambda \equiv z(\mathbf{SS}(\mathbf{KI}))_\lambda$   
 $\equiv z(\lambda xyz.xz(yz))(\lambda uvw.uw(vw))((\lambda st.s)(\lambda r.r))$   
 $\triangleright z(\lambda yzw.zw(yzw))(\lambda tr.r)$   
 $\triangleright z(\lambda zw.zw((\lambda tr.r)zw))$   
 $\triangleright z(\lambda zw.zww)$

In the second example above we have only

$$Y_{*\lambda} =_{\beta\eta} Y .$$

This turns out to be about as much as we can hope to have.

Weaker sets of combinators can be translated into  $\lambda$ -calculus in the same way as above with the translations of  $\mathbf{S}$  and  $\mathbf{K}$  in Definition 37 replaced by appropriate translations such as:

$$\mathbf{B}_\lambda \equiv \lambda xyz.x(yz) ,$$

for each element of the given basis set  $Q$ .

The reverse process however is not so simple. It is not clear which  $\lambda$ -terms can be translated, say, into  $\mathbf{BCIW}$  combinators, nor how to perform this translation.

We will resolve this problem later for several basis sets  $Q$ .

The process is important as, as was mentioned in the introduction, it provides a decision procedure and a constructive proof finding algorithm for axiomatic logics.

#### 4.1 $Q$ -Translation Algorithms

Trigg, Bunder and Hindley in [Trigg *et al.*, 1994] have extended the notion of abstractibility from that of Definition 7 for  $\mathbf{SK}$ -combinators to abstractibility for  $Q$ -combinators for various basis sets  $Q$ . These definitions will form part of our translation procedures.

First we will give the definition, from [Bunder, 1996], of a translation from  $\lambda$ -terms into  $Q$ -combinators.

DEFINITION 39. A mapping  $*$  from  $\lambda$ -terms to  $Q$ -terms is defined by:

1.  $X_* \equiv X$  ( $X$  an atom)
2.  $(XY)_* \equiv X_*Y_*$
3.  $(\lambda x.X)_* \equiv \lambda^*x.X_*$ ,

$\lambda^*x$  varies with  $Q$  but is, in simple cases, like the bracket abstraction  $[x]$  of Definition 7, a sequence of some of the following clauses:

- (i)  $\lambda^*x.x \equiv \mathbf{I}$
- (k)  $\lambda^*x.X \equiv \mathbf{K}X$  if  $x \notin FV(X)$
- ( $\eta$ )  $\lambda^*x.Xx \equiv X$  if  $x \notin FV(X)$
- (s)  $\lambda^*x.XY \equiv \mathbf{S}(\lambda^*x.X)(\lambda^*x.Y)$
- (b)  $\lambda^*x.XY \equiv \mathbf{B}X(\lambda^*x.Y)$  if  $x \notin FV(X)$
- (c)  $\lambda^*x.XY \equiv \mathbf{C}(\lambda^*x.X)$  if  $x \notin FV(Y)$

Note that, as before, the clauses in an algorithm  $*$  are used strictly in the order in which they appear.

Note also that the mapping starts by assigning a  $*$  to each  $\lambda$  in the term starting from the outermost ones and working inwards. Then, to the innermost terms  $\lambda^*x.Z$  i.e. those for which  $Z$  contains no  $\lambda^*$ s, we apply the appropriate abstraction clause. Later steps now evaluate terms  $\lambda^*x.Z_1$  where  $Z_1$  is a  $Q$ -term. The  $Q$ -combinators in  $Z_1$  arise from the evaluation of previous  $\lambda^*x.Z$ s.

EXAMPLE 40.

$$\begin{aligned} \lambda^{(ik\eta s)}xy.xzy &\equiv \lambda^{(ik\eta s)}x.xz \equiv \mathbf{SI}(\mathbf{K}z) \\ \lambda^{(isk\eta)}xy.xzy &\equiv \lambda^{(isk\eta)}x.\mathbf{S}(\mathbf{S}(\mathbf{K}x)(\mathbf{K}z))\mathbf{I} \\ &\equiv \mathbf{S}(\mathbf{S}(\mathbf{K}\mathbf{S})\mathbf{S}(\mathbf{S}(\mathbf{K}\mathbf{S})\mathbf{S}(\mathbf{K}\mathbf{K})\mathbf{I}))(\mathbf{S}(\mathbf{K}\mathbf{K})(\mathbf{K}z)))(\mathbf{K}\mathbf{I}) \end{aligned}$$

DEFINITION 41. A mapping  $*$  from  $\lambda$ -terms to  $Q$ -terms is said to be a **Q-translation algorithm** if:

- (A) For every  $Q$ -combinator  $X$ ,  $X_{\lambda^*}$  is defined and

$$X_{\lambda^*} \equiv X .$$

- (B) If for a  $\lambda$ -term  $Y$ ,  $Y_*$  is defined and is a  $Q$ -term, then there is a  $\lambda$ -term  $Y_1$  such that:

$$Y_{*\lambda} \triangleright_Q Y_1 \triangleleft_\eta Y$$

where  $\triangleright_Q$  means that only full or partial reductions involving the  $\lambda$ -versions of  $Q$ -combinators are used and  $\triangleleft_\eta$  involves only  $\eta$  and  $\alpha$ -reductions.

EXAMPLE 38.2 (again) In this example we had,

$$\begin{aligned} \text{with } * &\equiv (ik\eta s), \\ Y &\equiv z(\lambda xyz. xy yz), \\ \text{and } Y_{*\lambda} &\equiv z(\lambda zw. zww), \end{aligned}$$

so (B) holds with  $Y_1 =_\alpha Y_{*\lambda}$ .

EXAMPLE 42.  $(\mathbf{BCC})_{\lambda*} \equiv (\lambda xyz. x(yz))_*(\lambda uvw. uuv)_*(\lambda rst. rts)_*$

If  $*$  is  $(i\eta bc)$  we have

$$(\mathbf{BCC})_{\lambda*} \equiv \mathbf{BCC}$$

as required by (A).

If however  $*$  is  $(ikc)$   $(\lambda xyz. x(yz))_*$  is not definable so  $(ikc)$  is not a **BCI**-translation algorithm.

If  $*$  is  $(ibc)$  we have

$$\begin{aligned} \mathbf{B}_{\lambda*} &\equiv (\lambda xyz. x(yz))_* \\ &\equiv \mathbf{C}(\mathbf{BB}(\mathbf{BBI}))(\mathbf{C}(\mathbf{BBI})\mathbf{I}) \neq \mathbf{B}, \end{aligned}$$

so  $(ibc)$  is also not a **BCI**-translation algorithm.

(note that we do have  $\mathbf{B}_{\lambda*} =_{\beta\eta} \mathbf{B}$ .)

EXAMPLE 43. When  $*$  is  $(i\eta bc)$

$$\begin{aligned} (\lambda yz. z(\lambda x. yx))_{*\lambda} &\equiv (\lambda^* yz. z(\lambda^* x. yx))_\lambda \\ &\equiv (\lambda^* yz. zy)_\lambda \equiv (\lambda^* z. \mathbf{CI}z)_\lambda \\ &\equiv (\mathbf{CI})_\lambda \\ &\equiv (\lambda uvw. uuv)(\lambda x. x) \\ &\triangleright_{\mathbf{BCI}} \lambda vw. (\lambda x. x)wv \\ &\triangleright_{\mathbf{BCI}} \lambda vw. wv. \end{aligned}$$

while also  $\lambda yz. z(\lambda x. yx) \triangleright_\eta \lambda vw. wv$

The following theorem lists some properties that are preserved under the operations  $\lambda$  and  $*$ .

THEOREM 44.

1. If  $X$  and  $Y$  are  $Q$ -terms then

$$X \triangleright_Q Y \Rightarrow X_\lambda \triangleright_Q Y_\lambda .$$

If  $*$  is a  $Q$ -translation algorithm then:

2. for every  $Q$ -term  $X$ ,  $X_{\lambda*}$  is defined and

$$X_{\lambda*} \equiv X$$

3. For every pair of  $\lambda$ -terms  $U$  and  $V$  for which  $U_*$  and  $V_*$  are defined and are  $Q$ -terms,

$$U =_{\beta\eta} V \Leftrightarrow U_* =_{\beta\eta} V_*$$

4. If  $Y_*$  is defined and is a  $Q$ -term then

$$(\lambda x.Y)_* x =_{\beta\eta} Y_*$$

**Proof.**

1. It suffices to prove the result for  $X \equiv PX_1 \dots X_n$  where  $P$  is any  $Q$ -combinator and  $Y \equiv f(X_1, \dots, X_n)$  which results by a single  $P$ -reduction step. Then

$$\begin{aligned} X_\lambda &\equiv P_\lambda X_{1\lambda} \dots X_{n\lambda} \\ &\equiv (\lambda x_1 \dots x_n. f(x_1, \dots, x_n)) X_{1\lambda} \dots X_{n\lambda} \triangleright_Q f(X_{1\lambda}, \dots, X_{n\lambda}) \\ &\equiv Y_\lambda. \end{aligned}$$

2. By an easy induction using (A) and Definition 39.2.  
 3. By Curry and Feys [1958, Section 6C4, Theorem 1]:

$$U_* =_{\beta\eta} V_* \iff U_{*\lambda} =_{\beta\eta} V_{*\lambda}$$

so the result follows by (B).

4. By (3)  $((\lambda x.Y)x)_* =_{\beta\eta} Y_*$  so the result holds by Definition 39.2. ■

## 4.2 $Q$ -definability

We now come to the important question as to which  $\lambda$ -terms can be translated into  $Q$ -combinators, for a given  $Q$ . We first define  $Q$ -definability.

**DEFINITION 45.** A  $\lambda$ -term  $Y$  is  **$Q$ -definable** if there is a  $Q$ -translation algorithm  $*$  for which  $Y_*$  is defined and is a  $Q$ -term.

The following theorem relates the  $( )_\lambda$  operation to  $Q$ -definability.

**THEOREM 46.**

1. If  $Z$  is  $Q$ -definable there is a  $Q$ -term  $X$  such that  $Z =_{\beta\eta} X_\lambda$ .
2. If  $X$  is a  $Q$ -term and there is a  $Q$ -translation algorithm then  $X_\lambda$  is  $Q$ -definable.

**Proof.**

1. If  $Z$  is  $Q$ -definable there is a  $Q$ -translation algorithm  $*$  so that, by (B)

$$Z_{*\lambda} =_{\beta\eta} Z.$$

thus  $Z_*$  is the required  $X$ .

2. If  $X$  is a  $Q$ -term and  $*$  is a  $Q$ -translation algorithm, by Theorem 44.2  $X_{\lambda*} \equiv X$ , so clearly  $X_\lambda$  is  $Q$ -definable. ■

The importance of  $Q$ -definability, for implicational logics is that a typed  $\lambda$ -term  $Y$  is  $Q$ -definable if and only if its type is a theorem of  $Q$ -logic. In all the cases we deal with below we find that, for a given  $Q$ , we can find a  $Q$ -translation algorithm which translates all  $Q$ -definable  $\lambda$ -terms.

We now give translations algorithms for a number of sets of combinators. First we need a lemma.

LEMMA 47. *If  $U$  is a  $\lambda$ -term for which  $U_*$  is defined then*

1. *if  $*$  is the  $(i\eta ks)$  algorithm*  $(\lambda^*x.U_*)x \triangleright_{\mathbf{KS}} U_*$ .
2. *if  $*$  is the  $(i\eta kbc)$  algorithm*  $(\lambda^*x.U_*)x \triangleright_{\mathbf{BCK}} U_*$ .
3. *if  $*$  is the  $(i\eta bc)$  algorithm*  $(\lambda^*x.U_*)x \triangleright_{\mathbf{BCI}} U_*$ .
4. *if  $*$  is the  $(i\eta bcs)$  algorithm*  $(\lambda^*x.U_*)x \triangleright_{\mathbf{BCIW}} U_*$ .

**Proof.**

1. By induction on the length of  $U_*$ . If  $U_* \equiv x$

$$(\lambda^*x.U_*)x \equiv \mathbf{I}x \equiv \mathbf{SKK}x \triangleright_{\mathbf{KS}} x \equiv U_*.$$

If  $U_* \equiv U_{1*}x$  where  $x \notin FV(U_{1*})$ ,

$$(\lambda^*x.U_*)x \equiv U_{1*}x \equiv U_*.$$

If  $x \notin FV(U_*)$ ,

$$(\lambda^*x.U_*)x \equiv \mathbf{K}U_*x \triangleright_{\mathbf{KS}} U_*$$

If  $U_* \equiv U_{1*}U_{2*}$ , where  $x \in FV(U_{1*}U_{2*})$  and either  $x \in FV(U_{1*})$  or  $U_{2*} \neq x$ ,  $(\lambda^*x.U_*)x = \mathbf{S}(\lambda^*x.U_{1*})(\lambda^*x.U_{2*})x \triangleright_{\mathbf{KS}} (\lambda^*x.U_{1*})x((\lambda^*x.U_{2*})x) \triangleright_{\mathbf{KS}} U_{1*}U_{2*} \equiv U_*$  by the induction hypothesis.

Cases 2. to 4. are similar. ■

**THEOREM 48.**

1.  $(i\eta ks)$  is an **SK**-translation algorithm.
2.  $(i\eta kbc)$  is a **BCK**-translation algorithm.
3.  $(i\eta bc)$  is a **BCI**-translation algorithm.
4.  $(i\eta bcs)$  is a **BCIW**-translation algorithm.

**Proof.**

1. It is easy to check that (A) holds for the  $(i\eta ks)$  abstraction algorithm.

We prove (B) by induction on the length of the  $\lambda$ -term  $Y$ .

If  $Y$  is an atom  $Y_{*\lambda} \equiv Y$ .

If  $Y \equiv UV$ ,  $Y_{*\lambda} \equiv U_{*\lambda}V_{*\lambda}$  and by the inductive hypothesis we have a  $U_1$  and  $V_1$  such that  $Y_{*\lambda} \equiv U_{*\lambda}V_{*\lambda} \triangleright_{\mathbf{KS}} U_1V_1 \triangleleft_{\eta} UV \equiv Y$ .

If  $Y \equiv \lambda x.Xx$ , where  $x \notin FV(X)$ ,

$$Y_{*\lambda} \equiv (\lambda x.Xx)_{*\lambda} \equiv X_{*\lambda}.$$

By the induction hypothesis there is an  $X_1$  such that

$$Y_{*\lambda} \equiv X_{*\lambda} \triangleright_{\mathbf{KS}} X_1 \triangleleft_{\eta} X \triangleleft_{\eta} Y.$$

If  $Y = \lambda x.UV$  where  $x \in FV(U)$  or  $x \neq V$ , but  $x \in FV(UV)$ .

$$\begin{aligned} Y_{*\lambda} &\equiv \mathbf{S}_{\lambda}(\lambda x.U)_{*\lambda}(\lambda x.V)_{*\lambda} \\ &\equiv (\lambda uvx.ux(vx)) \\ &\quad (\lambda x.U)_{*\lambda}(\lambda x.V)_{*\lambda} \\ \triangleright_{\mathbf{SK}} \lambda x.((\lambda x.U)_{*\lambda}x)((\lambda x.V)_{*\lambda}x) &\equiv \lambda x.((\lambda^*x.U^*)x)_{\lambda}((\lambda^*x.V^*)x)_{\lambda} \\ &\quad \triangleright_{\mathbf{SK}} \lambda x.U_{*\lambda}V_{*\lambda} \end{aligned}$$

by Lemma 47 and Theorem 44.1.

Now by the induction hypothesis there exist  $U_1$  and  $V_1$  such that

$$\lambda x.U_{*\lambda}V_{*\lambda} \triangleright_{\mathbf{SK}} \lambda x.U_1V_1 \triangleleft_{\eta} \lambda x.UV \equiv Y .$$

Note that if  $x$  had not been chosen as the third bound variable in  $\mathbf{S}_{\lambda}$ , an extra  $\alpha$ -reduction would have been required from  $\lambda x.U_1V_1$  to reach the term obtained by the **SK**-reduction. We will use similar simplifications below.

If  $Y \equiv \lambda x.x$ ,

$$Y_{*\lambda} \equiv \mathbf{I}_{\lambda} =_{\alpha} Y .$$

If  $Y \equiv \lambda x.X$ , where  $x \notin FV(X)$ ,

$$\begin{aligned} Y_{*\lambda} &\equiv \mathbf{K}_\lambda X_{*\lambda} \\ &\equiv (\lambda y x.y) X_{*\lambda} \triangleright_{\mathbf{SK}} \lambda x.X_{*\lambda} . \end{aligned}$$

By the induction hypothesis there exists an  $X_1$  such that

$$X_{*\lambda} \triangleright_{\mathbf{SK}} X_1 \triangleleft_\eta X$$

so

$$Y_{*\lambda} \triangleright_{\mathbf{SK}} \lambda x.X_1 \triangleleft_\eta Y$$

2., 3. and 4. are similar. ■

EXAMPLE 49.

1.  $\lambda^{i\eta kbc} x_1 x_2 . x_2 (\lambda^{i\eta kbc} x_3 . x_1 x_4) \equiv \lambda^{i\eta kbc} x_1 x_2 . x_2 (\mathbf{K}(x_1 x_4))$   
 $\equiv \lambda^{i\eta kbc} x_1 . \mathbf{CI}(\mathbf{K}(x_1 x_4))$   
 $\equiv \mathbf{B}(\mathbf{CI})(\mathbf{BK}(\mathbf{CI}x_4))$
2.  $\lambda^{i\eta bc} x_1 x_2 . x_2 (\lambda^{i\eta bc} x_3 . x_3 x_1)$   
 $\equiv \lambda^{i\eta bc} x_1 x_2 . x_2 (\mathbf{CI}x_1)$   
 $\equiv \lambda^{i\eta bc} x_1 . \mathbf{CI}(\mathbf{CI}x_1)$   
 $\equiv \mathbf{B}(\mathbf{CI})(\mathbf{CI})$ .
3.  $\lambda^{i\eta kbc s} x_1 x_2 . x_2 (\lambda^{i\eta kbc s} x_3 . x_2 (x_1 x_3))$   
 $\equiv \lambda^{i\eta kbc s} x_1 x_2 . x_2 (\mathbf{B}x_2 x_1)$   
 $\equiv \lambda^{i\eta kbc s} x_1 . \mathbf{SI}(\mathbf{CB}x_1)$   
 $\equiv \mathbf{B}(\mathbf{SI})(\mathbf{CB})$ .

### 4.3 SK, BCI, BCK and BKIW Definable Terms

For many sets  $Q$  we can delineate the  $Q$ -definable terms. First we need some notation.

DEFINITION 50.

1.  $\Lambda$  is the set of all  $\lambda$ -terms.
2. A  $\lambda$ -term is in  $\text{Once}(i_1, \dots, i_n)$  if each  $x_{i_1}, \dots, x_{i_n}$  appears free exactly once in the term and if in every subterm  $\lambda x_{j_1} \dots x_{j_k} . Y$  of the term,  $Y$  is in  $\text{Once}(j_1, \dots, j_k)$ .
3. A  $\lambda$ -term is in  $\text{Once}^-(i_1, \dots, i_n)$  if each  $x_{i_1}, \dots, x_{i_n}$  appears free at most once in the term and if in every subterm  $\lambda x_{j_1} \dots x_{j_k} . Y$  of the term,  $Y$  is in  $\text{Once}^-(j_1, \dots, j_k)$ .

4. A  $\lambda$ -term is in  $\text{Once}^+(i_1, \dots, i_n)$  if each of  $x_{i_1}, \dots, x_{i_n}$  appears at least once in the term and if in every subterm  $\lambda x_{j_1} \dots x_{j_k}. Y$  of the term,  $Y$  is in  $\text{Once}^+(j_1, \dots, j_k)$ .

**THEOREM 51.**

1. The set of **SK**-definable terms is  $\Lambda$ .
2. The set of **BCI**-definable terms is  $\text{Once}(\ )$ .
3. The set of **BCK**-definable terms is  $\text{Once}^-(\ )$ .
4. The set of **BCIW**-definable terms is  $\text{Once}^+(\ )$ .

**Proof.**

1. is trivial.
2. If  $Y \in \text{Once}(i_1, \dots, i_n)$  for some  $i_1, \dots, i_n$  we show by induction on  $Y$  that  $Y$  is **BCI**-definable using the  $(i\eta bc)$  algorithm.

**Case 1**  $Y$  is an atom  $Y_{(i\eta bc)} \equiv Y$ .

**Case 2**  $Y \equiv UV$ ,  $U \in \text{Once}(j_1, \dots, j_r)$  and  $V \in \text{Once}(m_1, \dots, m_s)$ , where  $(j_1, \dots, j_r)$  and  $(m_1, \dots, m_s)$  are disjoint subsequences of  $(i_1, \dots, i_n)$  and  $r + s = n$ .

Then by the induction hypothesis  $U$  and  $V$  are **BCI**-definable using the  $(i\eta bc)$  algorithm and

$$Y_{(i\eta bc)} = U_{(i\eta bc)} V_{(i\eta bc)}.$$

**Case 3**  $Y \equiv \lambda x_p. Zx_p$  where  $x_p \notin FV(Z)$ .

$$Zx_p \in \text{Once}(i_1, \dots, i_n, p)$$

and by the induction hypothesis  $Zx_p$  is **BCI**-definable as  $Z_{(i\eta bc)}x_p$  and

$$Y_{(i\eta bc)} = Z_{(i\eta bc)}.$$

**Case 4**  $Y \equiv \lambda x_p. UV$ ,  $UV \in \text{Once}(i_1, \dots, i_n, p)$  and so  $x_p \in FV(V) - FV(U)$  or  $x_p \in FV(U) - FV(V)$ .

Hence, for similar disjoint sequences to the above we have  $U \in \text{Once}(j_1, \dots, j_r)$  and  $V \in \text{Once}(m_1, \dots, m_s, p)$  or  $U \in \text{Once}(j_1, \dots, j_r, p)$  and  $V \in \text{Once}(m_1, \dots, m_s)$ .

In the former case  $U, V$  and  $\lambda x_p. V$  are **BCI** definable and

$$Y_{(i\eta bc)} \equiv \mathbf{B}U_{(i\eta bc)}(\lambda^{(i\eta bc)}x_p.V_{(i\eta bc)}).$$



In the latter case  $U, V$  and  $\lambda x_p.U$  are **BCI**-definable and

$$Y_{(i\eta bc)} \equiv \mathbf{C}(\lambda^{(i\eta bc)} x_p.U_{(i\eta bc)})V_{(i\eta bc)}.$$

3. and 4. are similar, except that in 4.  $(m_1, \dots, m_s)$  and  $(j_1, \dots, j_r)$  need not be disjoint. ■

Note that in each of the four cases above the set of definable  $\lambda$ -terms leads us to a natural deduction style system for the logic.

In the case of **BCIW**-logic (the relevance logic  $R_{\rightarrow}$ ), the only  $\lambda$ -terms allowable are those in  $\text{Once}^+(\cdot)$ . These can be generated by allowing  $\lambda x.X$  to be defined only when  $x \in FV(X)$ . This restriction when translated to typed terms becomes:

$$\frac{\begin{array}{c} (A) \\ x : \alpha \\ D_1 \\ X : \beta \end{array}}{\lambda x.X : \alpha \rightarrow \beta} \text{---}(1)$$

only if  $x : \alpha$  is used in the proof  $D_1$ . This therefore gives the appropriate restriction on an  $R_{\rightarrow} \rightarrow$  introduction rule.

#### 4.4 Bases Without **C**

We now look at some basis sets that do not include (as defined or primitive) the combinator **C**. Its lack causes a problem.

Previously the algorithms we used in evaluating  $\lambda^* x_2.X$  did not affect the definability of  $\lambda^* x_1.\lambda^* x_2.X$ . When we are dealing with algorithms that do not include (c) (or (s)), this definability may fail depending on a choice of  $*$ .

If, for example, we define  $\lambda^* x_3.x_4 x_2(x_1 x_3)$  to be  $\mathbf{B}(x_4 x_2)x_1$ , using clause (b), we cannot easily define  $\lambda^* x_2.\mathbf{B}(x_4 x_2)x_1$ . If however we defined  $\lambda^* x_3.x_4 x_2(x_1 x_3)$  as  $\mathbf{B}'x_1(x_4 x_2)$ , using a clause (b'), we can define  $\lambda^* x_2.\lambda^* x_3.x_4 x_2(x_1 x_3)$  as  $\mathbf{B}(\mathbf{B}\mathbf{B}')x_1 x_4$  using (b). Clearly if (b) and (b') (and so **B** and **B'**) are both available in our translation algorithm, the choice of which to use would depend on the variables to be abstracted later.

If a subterm  $\lambda x_{i_n+1}.Y$  of a  $\lambda$ -term  $X$  is to be translated by  $*$  and is in the scope of (from left to right in  $X$ ):  $\lambda x_{i_1}, \dots, \lambda x_{i_n}$ , we will write this translation as  $\lambda_{x_{i_n+1}}^{x_{i_1}, \dots, x_{i_n}}.Y_*$ , so that the variables with respect to which we need to abstract later are flagged as:  $x_{i_n}$  to be done next, then  $x_{i_{n-1}}$  etc. These abstractions are of course tied to some set  $Q$  and to algorithm clauses which we still denote by  $*$ .

To ensure that the flagged  $x_{i_j}$ s are distinct we will assume that any  $\lambda$ -term  $X$  being translated has, if necessary, first been altered so that no  $\lambda x_k$  appears more than once in  $X$ .

For  $Q$ -translation algorithms  $*$ , where  $\mathbf{C}$  is not in  $Q$ , such as  $\mathbf{BB'I}$  and  $\mathbf{BB'IW}$ , we replace Definition 39 by:

DEFINITION 52.  $(x_{i_1}, \dots, x_{i_n}; Y)^*$  and in particular  $(; Y)^* \equiv Y_*$  are given by:

$$\begin{aligned} (x_{i_1}, \dots, x_{i_n}; P)^* &\equiv P \quad \text{if } P \text{ is an atom} \\ (x_{i_1}, \dots, x_{i_n}; PQ)^* &\equiv (x_{i_1}, \dots, x_{i_n}; P)^*(x_{i_1}, \dots, x_{i_n}; Q)^* \\ (x_{i_1}, \dots, x_{i_n}; \lambda_{x_{i_{n+1}}}.R)^* &\equiv \lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.(x_{i_1}, \dots, x_{i_n}, x_{i_{n+1}}; R)^* \\ (; \lambda_{x_{i_1}}.R)^* &\equiv \lambda^* x_{i_1} (x_{i_1}; R)^*. \end{aligned}$$

Note that Theorem 44 still holds under this definition.

EXAMPLE 53.

$$\begin{aligned} (& ; \lambda x_4 x_1 . x_3 (\lambda x_5 x_6 . x_1 (\lambda x_7 . x_1 x_5)) (\lambda x_2 . x_2 (\lambda x_9 . x_2)))^* \equiv \\ & \lambda^* x_4 \lambda_{x_1}^{x_4} . x_3 (\lambda_{x_5}^{x_4 x_1} . (\lambda_{x_6}^{x_4 x_1 x_5} . x_1 (\lambda_{x_7}^{x_4 x_1 x_5 x_6} . x_1 x_5))) (\lambda_{x_2}^{x_4 x_1} . x_2 (\lambda_{x_9}^{x_4 x_1 x_2} . x_2)) \end{aligned}$$

Before we can write down the algorithm clauses used for logics without  $\mathbf{C}$ , we need the notion of the index of a term with respect to a set of variables.

DEFINITION 54.  $(idx(M, i_1, \dots, i_n))$

$$idx(M, i_1, \dots, i_n) = \max \left\{ p \mid 1 \leq p \leq n \wedge x_{i_p} \in FV(M) \right\}$$

DEFINITION 55  $((\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.P))$ . This is defined using some or all of the following clauses, depending on  $*$ .

- (i)  $\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.x_{i_{n+1}} \equiv \mathbf{I}$
- ( $\eta$ )  $\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.Ux_{i_{n+1}} \equiv U$  if  $x_{i_{n+1}} \notin FV(U)$
- (b)  $\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.PQ \equiv \mathbf{BP}(\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.Q)$   
if  $idx(P, i_1, \dots, i_n) \leq idx(Q, i_1, \dots, i_n)$  or  $x_{i_1} \dots x_{i_n}$  is replaced by  $*$ ;  
and  $x_{i_{n+1}} \notin FV(P)$ .
- (b')  $\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.PQ \equiv \mathbf{B}'(\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.Q)P$   
if  $idx(P, i_1, \dots, i_n) > idx(Q, i_1, \dots, i_n)$  and  $x_{i_{n+1}} \notin FV(P)$ .
- (s)  $\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.PQ \equiv \mathbf{S}(\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.P)(\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.Q)$   
if  $idx(P, i_1, \dots, i_n) \leq idx(Q, i_1, \dots, i_n)$  or if  $x_{i_1} \dots x_{i_n}$  is replaced by  $*$
- (s')  $\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.PQ \equiv \mathbf{S}'(\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.Q)(\lambda_{x_{i_{n+1}}}^{x_{i_1}, \dots, x_{i_n}}.P)$   
if  $idx(Q, i_1, \dots, i_n) < idx(P, i_1, \dots, i_n)$

EXAMPLE 56.

$$\begin{aligned}
1. & \quad (; \lambda x_2 x_4 . x_2 (\lambda x_3 . (\lambda x_5 x_6 . x_5 (x_4 x_6))) x_3)^{(i\eta bb')} \\
& \equiv \lambda^* x_2 \lambda_{x_4}^{x_2} . x_2 (\lambda_{x_3}^{x_2 x_4} . (\lambda_{x_5}^{x_2 x_4 x_3} \lambda_{x_6}^{x_2 x_4 x_3 x_5} . x_5 (x_4 x_6))) x_3 \\
& \equiv \lambda^* x_2 \lambda_{x_4}^{x_2} . x_2 (\lambda_{x_3}^{x_2 x_4} . (\lambda_{x_5}^{x_2 x_4 x_3} . \mathbf{B}' x_4 x_5) x_3) \\
& \equiv \lambda^* x_2 \lambda_{x_4}^{x_2} x_2 (\lambda_{x_3}^{x_2 x_4} . \mathbf{B}' x_4 x_3) \\
& \equiv \lambda^* x_2 \lambda_{x_4}^{x_2} . x_2 (\mathbf{B}' x_4) \\
& \equiv \mathbf{B}' \mathbf{B}' \quad \text{where } * \text{ is } (i\eta bb')
\end{aligned}$$

$$\begin{aligned}
2. & \quad (; \lambda x_2 x_4 . x_2 (\lambda x_3 . x_2 x_4 (\lambda x_1 . x_3 (x_4 x_1))))^{(\eta ikbb' ss')} \\
& \equiv \lambda^* x_2 . \lambda_{x_4}^{x_2} . x_2 (\lambda_{x_3}^{x_2 x_4} . x_2 x_4 (\lambda_{x_1}^{x_2 x_4 x_3} . x_3 (x_4 x_1))) \\
& \equiv \lambda^* x_2 . \lambda_{x_4}^{x_2} . x_2 (\lambda_{x_3}^{x_2 x_4} . x_2 x_4 (\mathbf{B}' x_4 x_3)) \\
& \equiv \lambda^* x_2 . \lambda_{x_4}^{x_2} . x_2 (\mathbf{B} (x_2 x_4) (\mathbf{B}' x_4)) \\
& \equiv \lambda^* x_2 . \mathbf{B} x_2 (\mathbf{S}' \mathbf{B}' (\mathbf{B} \mathbf{B} x_2)) \\
& \equiv \mathbf{S} \mathbf{B} (\mathbf{B} (\mathbf{S}' \mathbf{B}' (\mathbf{B} \mathbf{B}))) \quad \text{where } * \text{ is } (\eta ikbb' ss') .
\end{aligned}$$

#### 4.5 Translation Algorithms and Definable Terms for $\mathbf{BB}'\mathbf{I}$ and $\mathbf{BB}'\mathbf{IW}$

Before we give the algorithms we need a lemma

LEMMA 57. *If  $Q$  is (i)  $\mathbf{BB}'\mathbf{I}$  or (ii)  $\mathbf{BB}'\mathbf{IW}$ ,  $X$  is a  $Q$ -term and  $\lambda_{x_{i_n+1}}^{x_{i_1} \dots x_{i_n}} . X$  is defined then*

$$\left( \lambda_{x_{i_n+1}}^{x_{i_1} \dots x_{i_n}} . X \right) x_{i_n+1} \triangleright_Q X .$$

**Proof.** By induction on the length of  $X$ .

Cases 1 to 4 apply where  $Q$  is  $\mathbf{BBI}'$  or  $\mathbf{BB}'\mathbf{IW}$ , Cases 5 and 6 only for  $\mathbf{BB}'\mathbf{IW}$ .

**Case 1**  $X \equiv x_{i_n+1}$ .

$$\left( \lambda_{x_{i_n+1}}^{x_{i_1} \dots x_{i_n}} . X \right) x_{i_n+1} \equiv \mathbf{I} x_{i_n+1} \triangleright_Q x_{i_n+1} \equiv X$$

**Case 2**  $X \equiv U x_{i_n+1}$  where  $x_{i_n+1} \notin FV(U)$

$$\left( \lambda_{x_{i_n+1}}^{x_{i_1} \dots x_{i_n}} . X \right) x_{i_n+1} \equiv U x_{i_n+1} \equiv X .$$

**Case 3**  $X \equiv UV$  where  $x_{i_n+1} \notin FV(U)$ ,  $x_{i_n+1} \neq V$ , and  $idx(U, i_1, \dots, i_n) \leq idx(V, i_1, \dots, i_n)$ .

$$\begin{aligned} (\lambda_{x_{i_{n+1}}}^{x_{i_1} \dots x_{i_n}} . X) x_{i_{n+1}} &\equiv \mathbf{BU} (\lambda_{x_{i_{n+1}}}^{x_{i_1} \dots x_{i_n}} . V) x_{i_{n+1}} \\ &\triangleright_Q U \left( (\lambda_{x_{i_{n+1}}}^{x_{i_1} \dots x_{i_n}} . V) x_{i_{n+1}} \right) \\ &\triangleright_Q UV \equiv X, \end{aligned}$$

by the induction hypothesis.

**Case 4**  $X \equiv UV$  where  $x_{i_{n+1}} \notin FV(U)$ ,  $x_{i_{n+1}} \not\equiv V$  and  $idx(V, i_1, \dots, i_n) < idx(U, i_1, \dots, i_n)$ . Similar to Case 3.

**Case 5**  $X \equiv UV$  where  $x_{i_{n+1}} \in FV(U) \cap FV(V)$  and  $idx(U, i_1, \dots, i_n) \leq idx(V, i_1, \dots, i_n)$ . Similar to Case 3.

**Case 6**  $X \equiv UV$  where  $x_{i_{n+1}} \in FV(U) \cap FV(V)$  and  $idx(V, i_1, \dots, i_n) < idx(U, i_1, \dots, i_n)$ . Similar to Case 3.  $\blacksquare$

**THEOREM 58.** *The following are translation algorithms:*

1.  $(; )^{(i\eta bb')}$ , for **BB'I**
2.  $(; )^{(i\eta bb' ss')}$ , for **BB'IW**

**Proof.** In each case (A) is obvious.

We will prove, for each algorithm  $*$  and each basis  $Q$ , if  $((x_{i_1}, \dots, x_{i_n}; Y)^*$  is defined, that there is a  $Y_1$  such that:

$$\left( (x_{i_1}, \dots, x_{i_n}; Y)^* \right)_\lambda \triangleright_Q Y_1 \triangleleft_\eta Y.$$

This we do by induction on the number  $k$  of clauses of  $*$  that are needed to evaluate  $(x_{i_1}, \dots, x_{i_n}; Y)^*$ .

If  $k = 0$  there are no  $\lambda$ s in  $Y$  and so

$$((x_{i_1}, \dots, x_{i_n}; Y)^*)_\lambda \equiv Y \equiv Y_1$$

If  $k > 0$  it is sufficient to consider a subterm  $(x_{i_1}, \dots, x_{i_m}; \lambda x_{i_{m+1}} . Z)^*$  ( $\equiv \lambda_{x_{m+1}}^{x_{i_1} \dots x_{i_m}} . Z$ ) of  $(x_{i_1}, \dots, x_{i_n}; Y)^*$  where  $Z$  contains no  $\lambda$ s and to show that there is a  $Z_1$  such that:

$$(\lambda_{x_{m+1}}^{x_{i_1} \dots x_{i_m}} . Z)_\lambda \triangleright_Q Z_1 \triangleleft_\eta \lambda x_{i_{m+1}} . Z$$

We then have, by Theorem 44.1,

$$((x_{i_1}, \dots, x_{i_n}; Y)^*)_\lambda \triangleright_Q ((x_{i_1}, \dots, x_{i_n}; Y')^*)_\lambda \triangleleft_\eta Y' \triangleleft_\eta Y$$

where  $Y'$  is  $Y$  with  $Z_1$  for  $\lambda x_{i_{m+1}} . Z$ .

Now  $Y'$  needs fewer than  $k$  clauses of  $*$  for its evaluation, so by the induction hypothesis we have a  $Y_1$  such that

$$((x_{i_1}, \dots, x_{i_n}; Y')^*)_\lambda \triangleright_Q Y_1 \triangleleft_\eta Y',$$

which provides the result.

To prove the above result for  $\lambda x_{i_{m+1}}.Z$  we consider 6 cases for **BB'IW**. The first 4 also apply to **BB'I**.

Note that in each case if  $U$  is  $Z$  or a subterm of  $Z$ , as this contains no  $\lambda$ s, we have  $U_* \equiv U$ .

(a) If  $Z \equiv x_{i_{m+1}}$ , then

$$\left( \lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.Z \right)_\lambda \equiv \mathbf{I}_\lambda \equiv \lambda x_{i_{m+1}}.Z$$

(b) If  $Z \equiv Ux_{i_{m+1}}$  where  $x_{i_{m+1}} \notin FV(U)$  then

$$\left( \lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.Z \right)_\lambda \equiv U_{*\lambda}.$$

By the induction hypothesis there is a  $U_1$ , such that

$$(\lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.Z)_\lambda \triangleright_Q U_{*\lambda} \triangleright_Q U_1 \triangleleft_\eta U \triangleleft_\eta \lambda x_{i_{m+1}}.Z$$

(c) If  $Z \equiv UV$ , where  $x_{i_{m+1}} \notin FV(U)$ ,  $V \neq x_{i_{m+1}}$  and  $idx(U, i_1, \dots, i_m) \leq idx(V, i_1, \dots, i_m)$

$$\begin{aligned} \left( \lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.Z \right)_\lambda &\equiv \left( BU \left( \lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.V \right) \right)_\lambda \\ &\equiv B\lambda U_{*\lambda} \left( \lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.V \right)_\lambda \\ &\triangleright_Q \lambda x_{i_{m+1}}.U_{*\lambda} \left( \left( \lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.V \right) x_{i_{m+1}} \right)_\lambda \end{aligned}$$

so by Theorem 44.1 and Lemma 57,

$$\left( \lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.Z \right)_\lambda \triangleright_Q \lambda x_{i_{m+1}}.U_{*\lambda} V_\lambda \equiv \lambda x_{i_{m+1}}.U_{*\lambda} V_{*\lambda}$$

Now by the induction hypothesis we have a  $U_1$  and  $V_1$  such that:

$$\left( \lambda_{x_{i_{m+1}}}^{x_{i_1}, \dots, x_{i_m}}.Z \right)_\lambda \triangleright_Q \lambda x_{i_{m+1}}.U_1 V_1 \triangleleft_\eta \lambda x_{i_{m+1}}.UV \equiv \lambda x_{i_{m+1}}.Z$$

(d), (e), (f) The cases  $Z \equiv UV$  where  $x_{i_{m+1}} \notin FV(U)$  and  $idx(U, i_1, \dots, i_m) > idx(V, i_1, \dots, i_m)$  or where  $x_{i_{m+1}} \in FV(U) \cap FV(V)$  are similar.  $\blacksquare$

Our classification of the **BB'I** and **BB'IW**-definable  $\lambda$ -terms involves a class  $HRM(i_1, \dots, i_n)$  of **hereditary right maximal terms** with respect to  $x_{i_1} \dots x_{i_n}$ , which we now define.

DEFINITION 59 ( $(HRM(i_1, \dots, i_n))$ ).

1. Every variable and every basis combinator is in  $HRM(i_1, \dots, i_n)$ .

2. If  $M, N \in HRM(i_1, \dots, i_n)$   
and  $idx(M, i_1, \dots, i_n) \leq idx(N, i_1, \dots, i_n)$   
then  $MN \in HRM(i_1, \dots, i_n)$ .
3. If  $M \in HRM(i_1, \dots, i_{n+1})$   
then  $\lambda x_{i_{n+1}}.M \in HRM(i_1, \dots, i_n)$ .

Strictly we should write  $HRM_Q(i_1, \dots, i_n)$ , but in each case below the basis  $Q$  will be clear from the context.

$HRM_{\mathbf{BB}'\mathbf{I}}(1, \dots, n)$  is  $HRM(x_1, \dots, x_n)$  of Hirokawa 1996. Our  $HRM(1, \dots, n)$  is also  $HRM_n$  of Trigg *et al.* [1994] where the basis is also taken from the context.

Before obtaining the classifications we need a lemma.

**LEMMA 60.** *If  $X$  is a  $\mathbf{BB}'\mathbf{IW}$ -term or  $X \equiv Y_\lambda$  where  $Y$  is a  $\mathbf{BB}'\mathbf{IW}$ -term and  $X \triangleright_{\mathbf{BB}'\mathbf{IW}} Z$  or  $Z \triangleright_\eta X$  then, if  $X$  is in  $HRM(i_1, \dots, i_n)$ , so is  $Z$ .*

**Proof.** If  $X$  is a  $\mathbf{BB}'\mathbf{IW}$ -term this is easy to prove by induction on the length of the  $\mathbf{BB}'\mathbf{IW}$ -reduction or of the  $\eta$ -expansion.

If  $X \equiv Y_\lambda$ , where  $Y$  is a  $\mathbf{BB}'\mathbf{IW}$ -term more single  $\mathbf{BB}'\mathbf{IW}$ -reductions are possible. For example instead of

$$\mathbf{B}'UVW \triangleright_{\mathbf{B}'} V(UW)$$

we can have  $(\lambda uvw.v(uw)) U_\lambda V_\lambda W_\lambda \triangleright_{\mathbf{B}'} (\lambda vw.v(U_\lambda w)) V_\lambda W_\lambda$

$$\begin{aligned} &\triangleright_{\mathbf{B}'} (\lambda w.V_\lambda(U_\lambda w)) W_\lambda \\ &\triangleright_{\mathbf{B}'} V_\lambda(U_\lambda W_\lambda), \end{aligned}$$

however in each case the membership of  $HRM(i_1, \dots, i_n)$  is preserved. Similarly for  $\eta$ -expansions. ■

**THEOREM 61.**

1. *The set of  $\mathbf{BB}'\mathbf{I}$ -definable terms is  $HRM(\ ) \cap Once(\ )$ .*
2. *The set of  $\mathbf{BB}'\mathbf{IW}$ -definable terms is  $HRM(\ ) \cap Once^+(\ )$ .*

**Proof.** Theorem 58 gives translation algorithms.

1. If  $Y \in HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n)$ , it is easy to show by induction on the length of  $Y$  that  $Y$  is  $\mathbf{BB}'\mathbf{I}$  definable by  $(x_{i_1}, \dots, x_{i_n}; Y)^{i\eta bb'}$ . This holds in particular when  $n = 0$ .

If  $Y$  is  $\mathbf{BB}'\mathbf{I}$ -definable then there is a  $\mathbf{BB}'\mathbf{I}$ -translation algorithm  $*$  and a  $\lambda$ -term  $Y_1$  such that

$$Y_{*\lambda} \triangleright_{\mathbf{BB}'\mathbf{I}} Y_1 \triangleleft_\eta Y$$

Now  $Y_*$  is a  $\mathbf{BB}'\mathbf{I}$ -term and is therefore in  $HRM(\ ) \cap Once(\ )$ . It follows by Lemma 60 that  $Y_1$  and  $Y$  are also in  $HRM(\ ) \cap Once(\ )$ .

2. is similar. ■

The  $(i\eta bb')$ -algorithm for **BB'I** was first used (for abstraction only) in [Helman, 1977]. It was also used by Hirokawa in his proof of the  $\Rightarrow$  half of (B) in [Hirokawa, 1996]. Hirokawa proved the result of Theorem 61.1 there.

## 5 THE BASES **BB'IK**, **BB'**, **BB'W** AND **BB'K**

The **BB'IK** $(i_1, \dots, i_n)$  abstractable terms of Trigg *et al* 1994 were terms obtainable from terms of  $HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n)$  by deleting certain variables. The **BB'IK** $(i_1, \dots, i_n)$ -translation algorithm that we develop here has as its first stage, a “**full ordering algorithm**” which reverses the deletion process by building up elements of a subclass of  $Once^-(i_1, \dots, i_n)$  to elements of  $HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n)$ . Such elements can then be translated by  $(;)^{(bb'in)}$ . A partial ordering algorithm, which builds up to elements of  $HRM(i_1, \dots, i_n) \cap Once^-(i_1, \dots, i_n)$  could also be used and requires only simple alterations to 1. and 2. below.

### THE FULL ORDERING ALGORITHM

**Aim** To extend, if possible, a  $\lambda$ -**BB'IK**-term  $Y \in Once^-(i_1, \dots, i_n)$  to a  $\lambda$ -**BB'IK**-term  $Y^o \in HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n)$  so that  $Y^o \triangleright_{\mathbf{KI}} Y$ .

1. If  $Y \equiv a$ , an atom not in  $\{x_{i_1}, \dots, x_{i_n}\}$   
 $Y^o \equiv \mathbf{K}_\lambda a (x_{i_1} x_{i_2} \dots x_{i_n})$
2. If  $Y \equiv x_{i_m}$ , and  $1 \leq m < n$  then  
 $Y^o \equiv \mathbf{K}_\lambda x_{i_m} (x_{i_1} \dots x_{i_{m-1}} x_{i_{m+1}} \dots x_{i_n})$
3. If  $Y \equiv x_{i_n}$ ,  
 $Y^o \equiv \mathbf{K}_\lambda \mathbf{I}_\lambda (x_{i_1} \dots x_{i_{n-1}}) x_{i_n}$
4. If  $Y \equiv \lambda x_{i_{n+1}}.Z$ , find, if possible,  $Z^o$  such that

$$Z^o \in HRM(i_1, \dots, i_{n+1}) \cap Once(i_1, \dots, i_{n+1})$$

and  $Z^o \triangleright_{\mathbf{KI}} Z$   
then  $Y^o \equiv \lambda x_{i_{n+1}}.Z^o$ .

5. If  $Y \equiv Zx_{i_n}$ , find, if possible,  $Z^o$  such that

$$Z^o \in HRM(i_1, \dots, i_{n-1}) \cap Once(i_1, \dots, i_{n-1})$$

and  $Z^o \triangleright_{\mathbf{KI}} Z$   
then  $Y^o \equiv Z^o x_{i_n}$ .

6. If  $Y \equiv UV$ , where  $V \not\equiv x_{i_n}$  find, by going back to (1), a term  $U^\circ$  and a subsequence  $(x_{j_1}, \dots, x_{j_r})$  of  $(x_{i_1}, \dots, x_{i_n})$  such that:

- (a)  $FV(U) \cap \{x_{i_1}, \dots, x_{i_n}\} \subseteq \{x_{j_1}, \dots, x_{j_r}\}$  and  $FV(V) \cap \{x_{j_1}, \dots, x_{j_r}\} = \emptyset$ .
- (b)  $U^\circ \in HRM(j_1, \dots, j_r) \cap Once(j_1, \dots, j_r)$
- (c)  $U^\circ \triangleright_{\mathbf{KI}} U$
- (d)  $j_r \neq i_n$
- (e)  $\max\{p \mid x_{j_p} \in \{x_{j_1}, \dots, x_{j_r}\} - FV(U)\}$  is minimal.
- (f) Given (e), the number of variables in  $\{x_{j_1}, \dots, x_{j_r}\} - FV(U)$  is minimal.

Now let  $(x_{s_1}, \dots, x_{s_t})$  be the sequence obtained from  $(x_{i_1}, \dots, x_{i_n})$  by removing  $(x_{j_1}, \dots, x_{j_r})$ .

Now if possible (i.e. by going back to (1)) find  $V^\circ$  such that

- (g)  $V^\circ \in HRM(s_1, \dots, s_t) \cap Once(s_1, \dots, s_t)$   
and
- (h)  $V^\circ \triangleright_{\mathbf{KI}} V$

then  $Y^\circ \equiv U^\circ V^\circ$ .

Choosing the maximal  $p$  in  $x_{j_p} \in FV(U^\circ) \cap \{x_{i_1}, \dots, x_{i_n}\}$  to be minimal in (e) and then using as few as possible variables new to  $U$  in (f), gives us maximal flexibility for expanding  $V$  to  $V^\circ$  using the remaining, especially the higher subscripted, variables. These clauses also ensure that a unique  $Y^\circ$  is produced by the algorithm. Other  $Y^\circ$ s satisfying the above aim may exist as well.

The algorithm is applied in two examples below.

EXAMPLE 62.  $Y \equiv x_3x_2x_1$  cannot be ordered relative to  $(1, 2, 3)$  or even  $(1, 2, 3, 4)$ , but relative to  $(1, 2, 3, 4, 5)$

$$Y^\circ \equiv x_3(\mathbf{K}_\lambda x_2 x_4)(\mathbf{K}_\lambda x_1 x_5)$$

EXAMPLE 63.

$$Y \equiv x_7x_0(\lambda x_9.x_5(x_4(x_3x_2)))x_1x_9$$

Relative to  $(0, 1, 2, \dots, 7, 8, 10, 11)$  we have

$$\begin{aligned} Y^\circ &\equiv x_7(\mathbf{K}_\lambda x_0 x_8)(\lambda x_9.x_5(x_4(x_3(\mathbf{K}_\lambda x_2 x_6))(\mathbf{K}_\lambda x_1(x_{10}x_{11}))))x_9 \\ &\in HRM(0, 1, 2, \dots, 8, 10, 11) \cap Once(0, 1, 2, \dots, 8, 10, 11) \end{aligned}$$

The  $\lambda$ -terms that are **BB'IK**-translatable will be represented in terms of a generalisation of the class  $HRM(i_1, \dots, i_n)$ . If it is not possible to extend a  $\lambda$ -**BB'IK**-term  $Y$  to a  $Y^\circ \in HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n)$ , it is always



possible to choose variables  $x_{i_{n+1}}, \dots, x_{i_m}$  and a  $Y^o \in HRM(i_1, \dots, i_m) \cap Once(i_1, \dots, i_m)$  so that  $Y^o \triangleright_{\mathbf{KI}} Y$ .

If  $(x_{j_1}, \dots, x_{j_r})$  is  $(x_{i_1}, \dots, x_{i_n})$  with the free variables of  $Y$  deleted and if we named the atom occurrences in  $Y$  from the leftmost to the rightmost  $a_1, \dots, a_p$  then  $Y^o$  could be defined as:  $Y$  with  $a_1$  replaced by  $\mathbf{K}a_1(x_{j_1} \dots x_{j_r} x_{i_{n+1}})$  and  $a_i$  ( $1 < i \leq p$ ) replaced by  $\mathbf{K}a_i x_{i_{n+i}}$ . Repeatedly using the full reordering algorithm, with  $n$  increased by one each time, will produce a minimal set of extra variables that need to be added to form a  $Y^o$ .

DEFINITION 64 (Potentially Right Maximal  $(i_1, \dots, i_n)$ - $\lambda$ -terms).

( $PRM(i_1, \dots, i_n)$ - $\lambda$ -terms)

1. If  $X$  is an atom  $X \in PRM()$ .
2.  $x_e \in PRM(e)$ .
3. If  $X \in PRM(i_1, \dots, i_{n-1})$  and  $x_j \notin FV(X)$  then  $X \in PRM(i_1, \dots, i_k, i_j, i_{k+1}, \dots, i_n)$  for  $0 \leq k \leq n$ .
4. If  $X \in PRM(i_1, \dots, i_{n+1})$  then  $\lambda x_{i_{n+1}}.X \in PRM(i_1, \dots, i_n)$ .
5. If  $X \in PRM(j_1, \dots, j_p)$  and  $Y \in PRM(r_1, \dots, r_q)$  where  $p = q = n = 0$  or  $r_q = i_n$ ,

$$\begin{aligned} \{j_1, \dots, j_p\} \cap \{r_1, \dots, r_q\} &= \emptyset, \\ FV(X) \cap \{x_{r_1}, \dots, x_{r_q}\} &= \emptyset, \\ FV(Y) \cap \{x_{j_1}, \dots, x_{j_p}\} &= \emptyset, \end{aligned}$$

and  $(i_1, \dots, i_n)$  is a merge of  $(j_1, \dots, j_p)$  and  $(r_1, \dots, r_q)$  (i.e.  $n = p + q$  and  $(i_1, \dots, i_n)$  has the elements of the two sequences with the orders preserved) then  $XY \in PRM(i_1, \dots, i_n)$ .

EXAMPLE 65.

$$x_1 \in PRM(1)$$

so

$$x_1 \in PRM(1, 5)$$

Similarly

$$x_2 \in PRM(2, 4)$$

and

$$x_3 \in PRM(3)$$

so

$$x_3 x_2 \in PRM(2, 3, 4)$$

and

$$x_3x_2x_1 \in PRM(1, 2, 3, 4, 5)$$

Note:

$$x_3x_2x_1 \notin PRM(1, 2, 3) \cup PRM(1, 2, 3, 4)$$

Note that the variables in  $\{x_{i_1}, \dots, x_{i_n}\} - FV(X)$  are used in the ordering of a term  $Y$  relative to  $(i_1, \dots, i_n)$ , in the same way these extra variable subscripts are needed to show  $X \in PRM(i_1, \dots, i_n)$ . The connection is given by the following lemmas.

LEMMA 66. *If  $Y$ , ordered relative to  $(i_1, \dots, i_n)$  by the full ordering algorithm, becomes  $Y^\circ$ , then*

$$Y^\circ \triangleright_{\mathbf{KI}} Y$$

where each single  $\mathbf{K}$ -reduction eliminates one or more of  $x_{i_1}, \dots, x_{i_n}$ .

**Proof.** Obvious from the algorithm. ■

LEMMA 67.  $Y \in PRM(i_1, \dots, i_n) \cap Once^-(i_1, \dots, i_n) \iff$  there is a  $Y^\circ \in HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n)$  defined by the full ordering algorithm.

**Proof.**  $\Rightarrow$  By induction on  $Y$ .

The case where  $Y$  is an atom is obvious.

If  $Y \equiv \lambda x_{i_{n+1}}.Z$ , then as each bounded variable of  $Y$  appears at most once in  $Y$ , we have  $Z \in PRM(i_1, \dots, i_{n+1}) \cap Once^-(i_1, \dots, i_{n+1})$ .

By the induction hypothesis we have an appropriate

$$Z^\circ \in HRM(i_1, \dots, i_{n+1}) \cap Once(i_1, \dots, i_{n+1})$$

and so a  $Y^\circ \equiv \lambda x_{i_{n+1}}.Z^\circ \in HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n)$ .

If  $Y \equiv Zx_{i_n}$ , then as each bounded variable of  $Y$  appears at most once in  $Y$ , we have  $Z \in PRM(i_1, \dots, i_{n-1}) \cap Once^-(i_1, \dots, i_{n-1})$ .

By the induction hypothesis we have an appropriate

$$Z^\circ \in HRM(i_1, \dots, i_{n-1}) \cap Once(i_1, \dots, i_{n-1})$$

and so a  $Y^\circ \equiv Z^\circ x_{i_n} \in HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n)$ .

If  $Y \equiv UV$  ( $V \neq x_{i_n}$ ) then we have  $(j_1, \dots, j_p)$  and  $(r_1, \dots, r_q)$  such that

$$\begin{aligned} \{j_1, \dots, j_p\} \cap \{r_1, \dots, r_q\} &= \emptyset, \\ FV(U) \cap \{x_{r_1}, \dots, x_{r_q}\} &= \emptyset, r_q = i_n, \text{ or } n = p = q = 0, \\ FV(V) \cap \{x_{j_1}, \dots, x_{j_p}\} &= \emptyset \\ U &\in PRM(j_1, \dots, j_p) \cap Once^-(j_1, \dots, j_p) \end{aligned}$$

and

$$V \in PRM(r_1, \dots, r_q) \cap Once^-(r_1, \dots, r_q)$$

Also the order of  $(j_1, \dots, j_p)$  and  $(r_1, \dots, r_q)$  is preserved in  $(i_1, \dots, i_n)$ , where  $\{i_1, \dots, i_n\} = \{j_1, \dots, j_p\} \cup \{r_1, \dots, r_q\}$ .

Of the sequences  $(j_1, \dots, j_p), (r_1, \dots, r_q)$  that satisfy these properties (and so (a), (b), (c), (d), (e) and (h) of (4) of the full ordering algorithm), choose those that also satisfy (e) and (f).

Then, as by the induction hypothesis we have:

$$U^\circ \in HRM(j_1, \dots, j_p) \cap Once(j_1, \dots, j_p)$$

and

$$V,^\circ \in HRM(r_1, \dots, r_q) \cap Once(r_1, \dots, r_q),$$

we have

$$Y^\circ \equiv U^\circ V^\circ \in HRM(i_1, \dots, i_n) \cap Once(i_1, \dots, i_n).$$

$\Leftarrow$  To prove this we only need to show, by the previous lemma, that  $\mathbf{K}_\lambda$  or  $\mathbf{I}_\lambda$ -reductions in  $\lambda$ -terms eliminating some of  $x_{i_1}, \dots, x_{i_n}$ , preserve membership of  $PRM(i_1, \dots, i_n)$  in a reduction  $T \triangleright_{\mathbf{BB}'\mathbf{IK}} R$ .

We prove this by induction on  $T$ .

If  $T$  is an atom or contains an  $\mathbf{I}_\lambda$  redex this is obvious.

If

$$T \equiv \mathbf{K}_\lambda W x_{i_s} \in PRM(i_1, \dots, i_n)$$

then

$$\mathbf{K}_\lambda W \in PRM(j_1, \dots, j_p)$$

and

$$x_{i_s} \in PRM(r_1, \dots, r_q)$$

where  $i_n = r_q$  and the other conditions apply.

From there it follows that  $x_{r_1}, \dots, x_{r_p} \notin FV(W)$ .

Also  $W \in PRM(j_1, \dots, j_p)$

and as  $(j_1, \dots, j_p)$  is a subsequence of  $(i_1, \dots, i_n)$ , by 3,

$$W \in PRM(i_1, \dots, i_n).$$

If  $T \equiv \lambda x_{i_{n+1}}.W$ ,  $R \equiv \lambda x_{i_{n+1}}.S$  and  $W \triangleright S$ ,

$$W \in PRM(i_1, \dots, i_{n+1})$$

and by the induction hypothesis  $S \in PRM(i_1, \dots, i_{n+1})$  and so

$R \in PRM(i_1, \dots, i_n)$ .

If  $T \equiv UV$ , where  $R \equiv WS$ ,  $U \triangleright_{\mathbf{BB}'\mathbf{IK}} W$  and  $V \triangleright_{\mathbf{BB}'\mathbf{IK}} S$ , we have: for some  $(j_1, \dots, j_p)$  and  $(r_1, \dots, r_q)$

$$\begin{aligned} U &\in PRM(j_1, \dots, j_p) \\ V &\in PRM(r_1, \dots, r_q) \end{aligned}$$

with the usual conditions.

By the induction hypothesis

$$W \in PRM(j_1, \dots, j_p)$$

and

$$S \in PRM(r_1, \dots, r_q)$$

and so

$$R \equiv WS \in PRM(i_1, \dots, i_n)$$

■

LEMMA 68. *If  $A \triangleright_\eta B$  and  $A \triangleright_{\mathbf{KI}} C$  then there is a term  $D$  such that  $B \triangleright_{\mathbf{KI}} D$  and  $C \triangleright_\eta D$ .*

**Proof.** By induction on the number of reduction steps in  $A \triangleright_\eta B$  and a secondary induction on the number of steps in  $A \triangleright_{\mathbf{KI}} C$  (i.e. a standard Church-Rosser theorem proof.) ■

THEOREM 69.  $(; Y^o)^{(i\eta bb')}$  is a **BB'IK**-translation algorithm.

**Proof.** (A) holds as before. By Theorem 58.1 there is a term  $Y_1$  such that  $((; Y^o)^{(i\eta bb')} )_\lambda \triangleright_{\mathbf{BB'I}} Y_1 \triangleleft_\eta Y^o$ .

Now also by Lemma 66

$$Y^o \triangleright_{\mathbf{KI}} Y,$$

so by Lemma 68 there is a  $Y_2$  such that

$$Y_1 \triangleright_{\mathbf{KI}} Y_2 \triangleleft_\eta Y$$

So  $((; Y^o)^{(i\eta bb')} )_\lambda \triangleright_{\mathbf{BB'IK}} Y_2 \triangleleft_\eta Y$ , i.e. (B) holds. ■

Thus  $(; Y^o)^{(i\eta bb')}$  is a **BB'IK**-translation algorithm.

THEOREM 70. *The set of **BB'IK**-translatable terms is  $PRM() \cap Once^-()$ .*

**Proof.** We have a **BB'IK**-translation algorithm by Theorem 69.

If  $Y \in PRM() \cap Once^-()$ , then by Lemma 67  $Y^o \in HRM() \cap Once^-()$  and so  $(; Y^o)^{(i\eta bb')}$  is a **BB'IK**-term, so  $Y$  is **BB'IK**-definable.

If  $Y$  is **BB'IK**-definable, the proof of  $Y \in PRM() \cap Once^-()$  proceeds as for Theorem 61. ■

We now consider some bases without **I**.

We define  $\Lambda^-$  as the set of all  $\lambda$ -terms whose  $\beta$ -normal forms do not have subterms of the form  $\lambda x_{i_1} \dots x_{i_n} . x_{i_j} x_{i_{j+1}} \dots x_{i_n}$  or  $\lambda x_{i_1} \dots x_{i_n} . x_{i_1} x_{i_1} x_{i_2} x_{i_3} \dots x_{i_n}$ .

THEOREM 71.

1.  $HRM(\ ) \cap \Lambda^- \cap Once$  is the set of  $\mathbf{BB}'$ -definable terms.
2.  $HRM(\ ) \cap \Lambda^- \cap Once^+$  is the set of  $\mathbf{BB}'\mathbf{W}$ -definable terms.
3.  $PRM(\ ) \cap \Lambda^- \cap Once^-$  is the set of  $\mathbf{BB}'\mathbf{K}$ -definable terms.

**Proof.** 3. Any  $\mathbf{BB}'\mathbf{K}$ -translation algorithm  $*$  must contain  $(\eta)$ , otherwise, for example,  $\mathbf{B}_{\lambda*}$  would not be definable. If a  $Q$ -translation algorithm  $*$  contains  $(\eta)$  it is easy to show, by induction on the length of a  $\lambda Q$ -term  $Y$ , with  $\eta$ -normal form  $Z$  that  $Y_* \equiv Z_*$ .

The full ordering algorithm is such that the combinator  $\mathbf{I}$  is used only when it is essential (i.e. just  $\mathbf{K}$ ,  $\mathbf{B}'$  and  $\mathbf{B}$  won't do) and it is clear that the only terms in  $\eta$ -normal form in  $PRM(\ ) \cap \Lambda^- \cap Once^-$  that can have an  $\mathbf{I}$  in their translation are of the form:  $\lambda x_{i_1} \dots x_{i_n} . x_{i_j} x_{i_{j+1}} \dots x_{i_n}$ .

Hence our result follows from Theorems 69 and 70.

1. and 2. are similar but simpler. ■

Note that in a  $\mathbf{BB}'\mathbf{IW}$  abstraction of a term only terms of the form  $\lambda x_{i_1} \dots x_{i_n} . x_{i_1} x_{i_2} \dots x_{i_n}$  and  $\lambda x_{i_1} \dots x_{i_n} . x_{i_1} x_{i_1} x_{i_2} \dots x_{i_n}$  can contain an  $\mathbf{I}$  and in  $\mathbf{BB}'\mathbf{I}$  abstraction only the former.

## 6 THE $\lambda$ PROOF FINDING ALGORITHMS

We have shown that for each theorem of a wide range of implicational logics there is a  $\lambda$ -term of a particular kind, depending on the logic, that has that theorem as a type. We have also shown that we can expand any such  $\lambda$ -term into long normal form.

If a theorem, or type,  $\tau$  takes the form  $\tau_1 \rightarrow \dots \rightarrow \tau_n \rightarrow a$ , where  $a$  is an atom, we know that any lnf inhabitant must take the form  $\lambda x_1 \dots x_n . X$ , where  $X$  has type  $a$  and has  $FV(X) \subseteq \{x_1, \dots, x_n\}$ . This is the basis for the Ben-Yelles algorithm (see [Ben-Yelles, 1979; Hindley, 1997; Bunder, 1995]), which constructs potential inhabitants of a given type from the outside in. As there may be several potential  $X$ s with type  $a$  the process can branch several times. There is, at least for  $\mathbf{SK}$  and sets of combinators without  $\mathbf{S}$  and  $\mathbf{W}$  a simple bound to this inhabitant, or proof, finding procedure. The version of the algorithm given in [Bunder, 1995], which is very efficient, has been implemented in [Dekker, 1996].

In this section we will be looking at an alternative inhabitant building algorithm, generally even more efficient, which builds the inhabitants of a type from the inside outwards. The given type provides the building blocks we use for this. This algorithm has been implemented in [Ostdijk, 1996].

### 6.1 The Variables and Subterms of an Inhabitant of a Type

In order to describe the way in which a  $\lambda$ -term inhabitant of a type is built up we need some notation.

DEFINITION 72.

1.  $\alpha$  is a positive subtype of a type  $\alpha$ .
2. If  $\beta$  is a positive subtype of  $\alpha$  or a negative subtype of  $\gamma$  then  $\beta$  is a negative subtype of  $\alpha \rightarrow \gamma$ .
3. If  $\beta$  is a negative subtype of  $\alpha$  or a positive subtype of  $\gamma$  then  $\beta$  is a positive subtype of  $\alpha \rightarrow \gamma$ .

DEFINITION 73. An occurrence of a positive (negative) subtype  $\beta$  of a type  $\tau$  is said to be **long** if the occurrence of  $\beta$  is not the right hand part of a positive (negative) subtype  $\alpha \rightarrow \beta$  of  $\tau$ .

All types other than atomic types are said to be **composite**.

DEFINITION 74. The rightmost atomic subtype of a type is known as its **tail**.

EXAMPLE 75.  $\tau = (a \rightarrow b) \rightarrow (b \rightarrow c) \rightarrow ((a \rightarrow b) \rightarrow c) \rightarrow c$  has  $b \rightarrow c, (a \rightarrow b) \rightarrow c$ , the second occurrence of  $a$  and the first occurrence of  $a \rightarrow b$  as long negative subtypes and  $\tau$ , the first occurrence of  $a$  and the second occurrences of  $b$  and  $a \rightarrow b$  as long positive subtypes of  $\tau$ .

DEFINITION 76.

$dn(\tau)$  = the number of distinct long negative subtypes of  $\tau$ .

$do(\tau)$  = the number of occurrences of long negative subtypes of  $\tau$ .

$dcp(\tau)$  = the number of distinct long positive composite subtypes of  $\tau$ .

$dapn(\tau)$  = the number of distinct atoms that are tails of both long positive and long negative subtypes of  $\tau$ .

$F(\tau) = 2^{dn(\tau)}(dapn(\tau) + dcp(\tau)) + dn(\tau)$ .

$G(\tau) = 2^{do(\tau)}(dapn(\tau) + dcp(\tau)) + do(\tau)$ .

$|\tau|$  = the total number of subtypes of  $\tau$ .

DEFINITION 77.

1.  $X$  is of  $\lambda$ -depth 0 in  $X$ .
2. If an occurrence of a term  $Y$  is of  $\lambda$ -depth  $d$  in  $X$  then  $Y$  is of  $\lambda$ -depth  $d$  in  $UX$  and  $XU$ , provided these are in  $\beta$ -normal form.
3. If an occurrence of a term  $Y$  is of  $\lambda$ -depth  $d$  in  $\lambda x_i.U$ , it is of  $\lambda$ -depth  $d$  in  $\lambda x_j x_i.U$ .

4. If an occurrence of a term  $Y$  is of  $\lambda$ -depth  $d$  in  $V \not\equiv \lambda x_i.U$  for any  $x_i$  or  $U$ , then  $Y$  is of  $\lambda$ -depth  $d + 1$  in  $\lambda x_j.V$ .

The two lemmas that now follow will help to tell us with what a long normal form inhabitant of a given type  $\tau$  can be constructed and also how to restrict the search for components.

LEMMA 78. *If  $X$  is a normal form inhabitant of  $\tau$ ,  $U$  is a subterm of  $X$  of type  $\alpha$  and  $V$  is a term of type  $\alpha$  with  $FV(V) \subseteq FV(U)$ , then the result of replacing  $U$  by  $V$  in  $X$  is another inhabitant of  $\tau$ .*

**Proof.** By a simple induction on the length of  $X$ . ■

EXAMPLE 79.

$$\lambda x_1 x_2. x_2(\lambda x_3. x_2(\lambda x_4. x_1 x_3 x_4)) : (a \rightarrow a \rightarrow b) \rightarrow ((a \rightarrow b) \rightarrow b) \rightarrow b$$

Here  $x_1 : a \rightarrow a \rightarrow b$ ,  $x_2 : (a \rightarrow b) \rightarrow b$ ,  $x_3 : a$  and  $x_4 : a$ , so

$$\lambda x_3. x_2(\lambda x_4. x_1 x_3 x_4) : a \rightarrow b$$

and

$$\lambda x_3. x_1 x_3 x_3 : a \rightarrow b.$$

So we have by Lemma 77:

$$\lambda x_1 x_2. x_2(\lambda x_3. x_1 x_3 x_3) : (a \rightarrow a \rightarrow b) \rightarrow ((a \rightarrow b) \rightarrow b) \rightarrow b.$$

LEMMA 80. *If  $\vdash Z : \tau$ , then there is a term  $X$  in long normal form, in which no two distinct variables have the same type, such that:*

1.  $\vdash X : \tau$
2. For every long subterm  $Y$  of  $X$  with  $FV(Y) = \{x_{i_1}, \dots, x_{i_k}\}$  we have:

$$x_{i_1} : \tau_{i_1}, \dots, x_{i_k} : \tau_{i_k} \vdash Y : \beta$$

where  $\beta$  is either a long occurrence of a composite positive subtype of  $\tau$  or an atom which is the tail of both a long positive and a long negative subtype of  $\tau$ .  $\tau_{i_1}, \dots, \tau_{i_k}$  are distinct long negative subtypes of  $\tau$ .

**Proof.**

1. By Theorem 34 there is an  $X'$  in *lnf* such that  $\vdash X' : \tau$ .

If  $X'$  now contains two variables  $x_k$  and  $x_e$  with the same type  $\beta$  we can change any part  $\lambda x_e.B(x_k, x_e)$  of  $X'$  to  $\lambda x_k.B(x_k, x_k)$  and any part  $\lambda x_e.B(x_e)$  to  $\lambda x_k.B(x_k)$ .

Let  $X$  be the term obtained when all possible changes of this kind have been made to  $X'$ . In  $X$  no two distinct variables will have the same type.

2. We prove this by induction on the  $\lambda$ -depth  $d$  of  $Y$  in  $X$ .

If  $X$  is formed by application, as it is in long normal form, it takes the form:

$$X \equiv x_i X_1 \dots X_m$$

and the type of  $X$  must be dependent on that of  $x_i$ , which does not agree with 1.

Hence  $X$  is not formed by application and if  $\tau = \tau_1 \rightarrow \dots \rightarrow \tau_n \rightarrow a$  takes the form

$$X \equiv \lambda x_1 \dots x_n. x_i X_1 \dots X_m.$$

(Note: some of  $x_1, \dots, x_n$  could be identical.)

If  $d = 0$  and  $Y$  is long, we must have  $Y \equiv X$  and the result holds with  $k = 0$  and  $\beta = \tau$ . If  $d = 1$  and  $X = \lambda x_1 \dots x_n. x_i X_1 \dots X_m$ ,  $Y$  will appear in  $x_i X_1 \dots X_m$ , not in the scope of any  $\lambda x_j$ s. Thus it appears in a part  $x_t Z_1 \dots Z_r Y Z_{r+2} \dots Z_q$  of  $X$  for some  $t(1 \leq t \leq n)$ .

If  $\tau_t = \beta_1 \rightarrow \dots \rightarrow \beta_q \rightarrow b$  we have,  $x_1 : \tau_1, \dots, x_n : \tau_n \vdash Y : \beta_{r+1}$ .

$\beta_{r+1}$  is a long negative subtype of  $\tau_t$  and so a long positive subtype of  $\tau$ .  $\tau_1, \dots, \tau_n$  are occurrences of long negative subtypes of  $\tau$ .

If  $\beta_{r+1}$  is an atom, then  $\beta_{r+1}$  must also be the tail of a long negative subtype of  $\tau$  (namely one of  $\tau_1, \dots, \tau_n$ ).

Leaving out the variables not free in  $Y$  and variables identical to others gives the result.

If  $d > 1$ ,

$X \equiv \lambda x_1 \dots x_n. x_i X_1 \dots X_{j-1} (\lambda x_{n+1} \dots x_k. x_s Z_1 \dots Z_p) X_{j+1} \dots X_m$ , where  $Y$  is a long subterm of  $x_s Z_1 \dots Z_p$ .

As  $x_s Z_1 \dots Z_p$  is long and only in the scope of  $\lambda x_1 \dots x_n \dots x_k$ , we have, if  $\tau_i = \gamma_1 \rightarrow \dots \rightarrow \gamma_m \rightarrow c$  and  $\gamma_j = \tau_{n+1} \rightarrow \dots \rightarrow \tau_k \rightarrow b$ :

$$x_1 : \tau_1, \dots, x_n : \tau_n, \dots, x_k : \tau_k \vdash x_s Z_1 \dots Z_p : b.$$

Thus  $\vdash \lambda x_1 \dots x_k. x_s Z_1 \dots Z_p : \tau_1 \rightarrow \dots \rightarrow \tau_k \rightarrow b$  where  $Y$  is a long subterm of  $\lambda x_1 \dots x_k. x_s Z_1 \dots Z_p$  of  $\lambda$ -depth  $d - 1$ .

Thus 2. holds by the induction hypothesis with  $\beta$  a long positive subtype of  $\tau_1 \rightarrow \dots \rightarrow \tau_k \rightarrow b$  and so of  $\tau$  or an atom which is the tail of a long positive as well as of a long negative subtype of  $\tau_1 \rightarrow \dots \rightarrow \tau_k \rightarrow b$  and so of  $\tau$ .

■

**Note** The first inhabitant given in Example 79 is a counterexample, due to Ryo Kashima, to an earlier version of Lemma 80. This claimed property 2.



for any inhabitant  $X$  of the given type  $\tau$  in long normal form, rather than just some  $X$ .

It follows from Lemmas 78 and 80 that an  $X$  in long normal form such that

$$\vdash X : \tau,$$

can be built up from subterms of the form

$$\lambda x_r \dots x_s. x_i X_1 \dots X_n,$$

where  $x_i : \tau_i$ ,  $\tau_i$  is a long negative subtype of  $\tau$ , and  $\tau_i$  has tail  $a$ , where  $a$  is an atom which is also the tail of a positive subtype of  $\tau$ .

The compound types of these subterms in long normal form must be among the long positive subtypes of  $\tau$ .

With this in mind we arrive at the following algorithm for finding inhabitants of types, i.e. proofs in intuitionistic implicative logic.

## 6.2 SK-logic ( $H \rightarrow$ )

*The  $H \rightarrow$  Decision Procedure or the SK Long Inhabitant Search Algorithm*

### Aim

Given a type  $\tau$ , to find a closed  $X$  in long normal form (if any) such that

$$\vdash X : \tau$$

### Step 1

To each distinct long negative subtype  $\tau_i$  of  $\tau$  assign a variable  $x_i$  giving a finite list:

$$x_1 : \tau_1, \dots, x_m : \tau_m$$

### Step 2

For each atomic type  $b$  that is the tail of both a long negative and a long positive subtype of  $\tau$ , form by application, (if possible) a lnf inhabitant  $Y$  of  $b$  if we don't already have a  $Y' : b$  with  $FV(Y') \subseteq FV(Y)$  in Step 1 or this or earlier Steps 2.

**Step 3**

For each long positive composite subtype  $\beta$  of  $\tau$ , form, by abstraction, with respect to some of  $x_1, \dots, x_m$ , a term  $Y$  in  $\text{Inf}$  such that  $Y : \beta$  if we don't already have a  $Y' : \beta$  with  $FV(Y') \subseteq FV(Y)$  in Step 1 or earlier Steps 3.

If one of these terms  $Y$  is closed and has type  $\tau$  we stop.

If there is no such term we continue with Steps 2 and 3 until we obtain no more terms with a “new” set of free variables or a new (atomic tail or long positive) type. If there are no new terms there is no solution  $X$ .

**Note**

In the work below and in all examples we will select our variables in Step 1 in the following order:  $x_1$  is assigned to the leftmost shallowest long negative subtype of  $\tau$ ,  $x_2$  to the next to leftmost shallowest long negative subtype etc. until the shallowest long negative subtypes are used up. The next variable  $x_{n+1}$  is assigned to the leftmost next shallowest long negative subtype and eventually  $x_m$  to the rightmost deepest long negative subtype.

In Example 81 below,  $x_1$  to  $x_4$  are assigned to the shallowest subtypes and  $x_5$  to the next shallowest (the deepest) subtype.

Because of this ordering of the variables of  $X$  any subterm  $Y$  formed by the algorithm will be in the scope of  $\lambda x_1 x_2 \dots x_n x_{p_1} x_{p_1+1} \dots x_{p_2} x_{p_3} \dots x_{p_4} \dots x_{p_q}$  where  $n < p_1 < p_2 < p_3 \dots < p_q$  and where each  $\lambda x_{p_{2i+1}} \dots \lambda x_{p_{2i+2}}$  represents a single set of abstractions.

EXAMPLE 81.

$$\tau = [((a \rightarrow b) \rightarrow d) \rightarrow d] \rightarrow [(a \rightarrow b) \rightarrow d \rightarrow e] \rightarrow [a \rightarrow a \rightarrow b] \rightarrow a \rightarrow b$$

**Step 1**  $x_1 : ((a \rightarrow b) \rightarrow d) \rightarrow d$ ,  $x_2 : (a \rightarrow b) \rightarrow d \rightarrow e$ ,  
 $x_3 : a \rightarrow a \rightarrow b$ ,  $x_4 : a$ ,  $x_5 : a \rightarrow b$ ,

**Step 2**  $x_3 x_4 x_4 : b$ ,  $x_5 x_4 : b$ ,

**Step 3**  $\lambda x_4. x_3 x_4 x_4 : a \rightarrow b$ ,  $\lambda x_4. x_5 x_4 : a \rightarrow b$ ,  $\lambda x_1 x_2 x_3 x_4. x_3 x_4 x_4 : \tau$ .

EXAMPLE 82.

$$\tau = ((a \rightarrow b) \rightarrow a) \rightarrow a.$$

**Step 1**  $x_1 : (a \rightarrow b) \rightarrow a$ ,  $x_2 : a$ .

**Step 2** No new terms can be formed by application.

**Step 3**  $\lambda x_1. x_2 : \tau$ .  
 No new terms can be formed.  $\lambda x_1. x_2$  is not closed so  $\tau$  has no inhabitants and no proof in  $H \rightarrow$ .

EXAMPLE 83.

$$\tau = [((a \rightarrow a) \rightarrow a) \rightarrow a] \rightarrow [(a \rightarrow a) \rightarrow a \rightarrow a] \rightarrow [a \rightarrow a \rightarrow a] \rightarrow a \rightarrow a.$$

**Step 1**  $x_1 : ((a \rightarrow a) \rightarrow a) \rightarrow a$ ,  $x_2 : (a \rightarrow a) \rightarrow a \rightarrow a$ ,  
 $x_3 : a \rightarrow a \rightarrow a$ ,  $x_4 : a$ ,  $x_5 : a \rightarrow a$ ,

**Step 2**  $x_3 x_4 x_4 : a$ ,  $x_5 x_4 : a$ ,

**Step 3**  $\lambda x_4. x_4 : a \rightarrow a$ ,  $\lambda x_5. x_5 x_4 : (a \rightarrow a) \rightarrow a$ ,  $\lambda x_1 x_2 x_3 x_4. x_3 x_4 x_4 : \tau$ .

**THEOREM 84.** *Given a type  $\tau$ , the **SK** long inhabitant search algorithm will, in finite time, produce an inhabitant of  $\tau$  or will demonstrate that  $\tau$  has no inhabitants. The algorithm will produce at most  $F(\tau)$  terms before terminating.*

**Proof.** It follows from the Weak Normalisation Theorem (see [Turing, 1942; Hindley, 1997]) that if  $\tau$  has an inhabitant, this inhabitant has a normal form and this will also have type  $\tau$ .

By Lemma 80,  $\tau$  will have an inhabitant  $X$  of the form prescribed there. We show that our **SK**-algorithm provides such an inhabitant  $X$  of  $\tau$ .

Step 1 of the **SK**-algorithm provides us with the largest set of variables  $x_1, \dots, x_m$  that, by Lemma 80, need appear in a solution for  $X$ .

Step 2 of the algorithm considers terms  $U = x_i X_1 \dots X_n$  with an atomic type having a particular subset of  $x_1, \dots, x_m$  as free variables. By Lemma 78 other terms  $x_j Y_1 \dots Y_k$  with the same atomic type and a superset of these free variables can at most produce alternative inhabitants and so do not need to be considered.

The total number of variables we can have is  $dn(\tau)$ . These are the terms generated by Step 1. The number of subsets of these is at most  $2^{dn(\tau)}$ , the number of atomic types we can have is at most  $dapn(\tau)$  so the number of terms generated by Steps 2 is at most  $2^{dn(\tau)}.dapn(\tau)$ .

Step 3 forms terms in long normal form which have composite long positive subtypes of  $\tau$  as types. There are  $dcp(\tau)$  of these and we can form at most one of these terms for each set of variables. Hence the most terms we can form using Steps 3 is  $2^{dn(\tau)}.dcp(\tau)$ .

The maximal number of terms formed using the algorithm is therefore  $dn(\tau) + 2^{dn(\tau)}.dapn(\tau) + dcp(\tau) = F(\tau)$ . ■

### Note

As  $dapn(\tau) + dcp(\tau) \leq dp(\tau)$  where  $dp(\tau)$  is the number of occurrences of distinct positive subtypes in  $\tau$ ,  $F(\tau) < dn(\tau) + 2^{dn(\tau)}.dp(\tau) < 2^{dn(\tau)+dp(\tau)} \leq 2^{d(\tau)}$  where  $d(\tau)$  is the number of long subtypes of  $\tau$  so  $F(\tau) < 2^{|\tau|}$ .

In Example 1,  $F(\tau) = 197$  while  $2^{d(\tau)} = 2^{12}$  and  $2^{|\tau|} = 2^{25}$ . The actual number of terms formed by the algorithm was 10.

### 6.3 BCK Logic

We now adapt our long Inhabitant Search Algorithm to search for **BCK**- $\lambda$ -terms in long normal form. By Theorem 51(3), these need to be elements of  $Once^-(\cdot)$ .

*The BCK Logic Decision Procedure or the BCK-Long Inhabitant Search Algorithm*

#### Aim

Given a type  $\tau$  to find a closed **BCK**-definable  $\lambda$ -term in long normal form (if any) such that

$$\vdash X : \tau.$$

**Step 1** To each occurrence of a long negative subtype  $\tau_i$  of  $\tau$  assign a variable  $x_i$  giving a finite list:

$$x_1 : \tau_1, \dots, x_m : \tau_m.$$

**Step 2** For each atomic type  $b$  that is the tail of both a long positive and a long negative subtype of  $\tau$  form (if possible), by application, from the terms we have so far, an inhabitant  $Y$  of  $b$  such that no free variables appear more than once in  $Y$ , if we don't already have a  $Y' : b$  with  $FV(Y') \subseteq FV(Y)$ .

**Step 3** For each long positive subtype  $\beta$  of  $\tau$ , form, by abstraction, with respect to some of  $x_1, \dots, x_m$ , a term  $Y$  such that  $Y : \beta$ , if we don't already have a  $Y' : \beta$  where  $FV(Y') \subseteq FV(Y)$ .

EXAMPLE 85.

$$\tau = [((a \rightarrow b) \rightarrow d) \rightarrow d] \rightarrow [(a \rightarrow b) \rightarrow d \rightarrow e] \rightarrow [a \rightarrow a \rightarrow b] \rightarrow a \rightarrow b$$

**Step 1**  $x_1 : ((a \rightarrow b) \rightarrow d) \rightarrow d$ ,  $x_2 : (a \rightarrow b) \rightarrow d \rightarrow e$ ,  
 $x_3 : a \rightarrow a \rightarrow b$ ,  $x_4 : a$ ,  $x_5 : a \rightarrow b$ ,  $x_6 : a$ ,

**Step 2**  $x_5 x_4 : b$ ,  $x_5 x_6 : b$ ,  $x_3 x_4 x_6 : b$ ,

**Step 3**  $\lambda x_4. x_5 x_4 : a \rightarrow b$ ,  
 $\lambda x_4. x_3 x_4 x_6 : a \rightarrow b$ ,  $\lambda x_6. x_3 x_4 x_6 : a \rightarrow b$ ,  $\lambda x_1 x_2 x_3 x_4. x_5 x_4 : \tau$ ,  
 $\lambda x_1 x_2 x_3 x_6. x_3 x_4 x_6 : \tau$ ,  $\lambda x_1 x_2 x_3 x_4. x_3 x_4 x_6 : \tau$ .

No new terms are generated by further uses of steps 2 and 3 and as the terms with type  $\tau$  are not closed, there is no **BCK** inhabitant of  $\tau$ . Note that  $\tau$  does have an **SK** inhabitant  $\lambda x_1 x_2 x_3 x_4. x_3 x_4 x_4$ , but this is not **BCK** because  $x_4$  is used twice.

EXAMPLE 86.

$$\tau = ((a \rightarrow a) \rightarrow a \rightarrow a \rightarrow b) \rightarrow a \rightarrow a \rightarrow b$$

**Step 1**  $x_1 : (a \rightarrow a) \rightarrow a \rightarrow a \rightarrow b, x_2 : a, x_3 : a, x_4 : a,$

**Step 2** No new terms are formed.

**Step 3**  $\lambda x_2.x_2 : a \rightarrow a,$

**Step 2**  $x_1(\lambda x_2.x_2)x_2x_3 : b,$   
 $x_1(\lambda x_2.x_2)x_2x_4 : b, x_1(\lambda x_2.x_2)x_3x_4 : b,$

**Step 3**  $\lambda x_1x_2x_3.x_1(\lambda x_2.x_2)x_2x_3 : \tau.$

**Notes:** 1. The **SK** algorithm would have produced only  $\lambda x_1x_2x_3.x_1(\lambda x_2.x_2)x_2x_3 : \tau$  which is not a **BCK**- $\lambda$ -term.

2. It was essential here to have a variable for each **distinct** long negative occurrence of  $a$  in  $\tau$ .

**LEMMA 87.** *If  $X$  is a **BCK** term which is a normal form inhabitant of  $\tau$ ,  $U$  a subterm of  $X$  of type  $\alpha$  and  $V$  a **BCK** term of type  $\alpha$ , in which no free variable appears more than once, with  $FV(V) \subseteq FV(U)$ , then the result of replacing  $U$  by  $V$  in  $X$  is another **BCK** inhabitant of  $\tau$ .*

**Proof.** As for Lemma 78. ■

**LEMMA 88.** *If  $\vdash_{\mathbf{BCK}} Z : \tau$ , then there is a term  $X$  in long normal form such that:*

1.  $\vdash_{\mathbf{BCK}} X : \tau$
2. For every long subterm  $Y$  of  $X$  with  $FV(Y) = \{x_{i_1}, \dots, x_{i_n}\}$  we have

$$x_{i_1} : \tau_{i_1}, \dots, x_{i_n} : \tau_{i_n} \vdash_{\mathbf{BCK}} Y : \beta,$$

where  $\beta$  is either a long occurrence of a composite positive subtype of  $\tau$  or an atom which is the tail of both a long positive and a long negative subtype of  $\tau$ .  $\tau_{i_1}, \dots, \tau_{i_n}$  are distinct occurrences of long negative subtypes of  $\tau$ .

**Proof.**

1. The formation of an  $X$  in long normal form is as in the proof of Lemma 80(1) except that the extra variables  $x_{m+1}, \dots, x_n$  that are chosen must not be free in  $x_iX_1, \dots, X_m$ , (otherwise  $\lambda x_{m+1} \dots x_n.x_iX_1, \dots, X_mx_{m+1}, \dots, x_n$  would not be a **BCK**-term). Also for the same reason, we don't identify distinct variables with the same type. We let  $X$  be a term obtained by the expansion of  $Z$  to long normal form.

2. As for Lemma 80(2) except that we have to show that we have at most one variable for each distinct occurrence of a long negative subtype of  $\tau$ . We extend the induction proof to include this.

When  $d = 1$ , we clearly have one variable for every long negative subtype  $\tau_1 \dots, \tau_n$  of  $\tau = \tau_1 \rightarrow \dots \rightarrow \tau_n \rightarrow a$  and no others.

When  $d > 1$  the subterm  $Y$  appears as  $Z_j$  in a term

$$\begin{aligned} & x_t U_1 \dots U_s \\ \text{where } & U_e = \lambda x_u \dots x_v. x_r Z_1 \dots Z_p \\ & x_t : \gamma_1 \rightarrow \dots \rightarrow \gamma_s \rightarrow c \\ \text{and } & \gamma_w = \delta_u \rightarrow \dots \rightarrow \delta_v \rightarrow d \\ \text{and } & 1 \leq j \leq p, \quad 1 \leq e \leq s \text{ and } 1 \leq w \leq s. \end{aligned}$$

The extra typed variables added at this stage are  $x_u : \delta_u, \dots, x_v : \delta_v$ .

By the inductive hypothesis we have that  $\gamma_1 \rightarrow \dots \rightarrow \gamma_s \rightarrow c$  is an occurrence of a long negative subtype of  $\tau$  and it therefore follows that the same holds for  $\delta_u, \dots, \delta_v$ . Note that as in **BCK** (and **BCI**) logic there can be only one free occurrence of the variable  $x_t$  in a term before we abstract with respect to  $x_t$ , so there can be no other use of  $x_t$  that might generate another set of variables with types  $\delta_u, \dots, \delta_v$  i.e. one occurrence of  $\gamma_1 \rightarrow \dots \rightarrow \gamma_q \rightarrow c$  in  $\tau$  generates at most one occurrence of a variable for each occurrence of  $\delta_u, \dots, \delta_v$  which are long negative subtypes of depth one lower than  $\gamma_1 \rightarrow \dots \rightarrow \gamma_q \rightarrow c$  in  $\tau$ . ■

**THEOREM 89.** *Given a type  $\tau$ , the **BCK** long inhabitant search algorithm will in finite time produce an inhabitant or will demonstrate that  $\tau$  has no **BCK**-inhabitants. The algorithm will produce at most  $G(\tau)$  terms before terminating.*

**Proof.** As for Theorem 84, except that Lemmas 87 and 88 replace Lemmas 78 and 80. ■

It might be thought that the **BCK** algorithm might require fewer than the maximal  $F(\tau)$  terms required for **SK**, in fact it requires more because there may be several variables with the same type. Even when this is not the case, both algorithms require at most one term for each given type and each set of variables. For **SK** some variables may appear several times, for **BCK** they may not. For **BCI** in addition all abstracted variables will have to appear in the term being abstracted.

The bounds  $F(\tau)$  and  $G(\tau)$  are not directly related to standard complexity measures. The algorithm will in fact generate some other terms but not record them because they have the same type and set of free variables to another already recorded.

### 6.4 BCI Logic

Again we can adapt the Long Inhabitant Search Algorithm. This time the inhabitants found must, by Theorem 51(2) be in *Once*( ).

*The BCI-Logic Decision Procedure or the BCI Long Inhabitant Search Algorithm*

#### Aim

Given a type  $\tau$  to find a closed **BCI**- $\lambda$ -term  $X$  in long normal form (if any) such that:

$$\vdash X : \tau.$$

#### Method

As for the **BCK**-algorithm except that Step 2 and 3 end in “ $FV(Y') = FV(Y)$ ” and in Step 3 we may only form  $\lambda x_i \dots x_j. Z$  if  $x_i, \dots, x_j$  occur free in  $Z$  exactly once each.

EXAMPLE 90.

$$\tau = ((a \rightarrow b) \rightarrow c) \rightarrow b \rightarrow d \rightarrow (c \rightarrow e) \rightarrow e$$

**Step 1**  $x_1 : (a \rightarrow b) \rightarrow c, x_2 : b, x_3 : d, x_4 : c \rightarrow e, x_5 : a$

**Step 2** No new terms are formed.

**Step 3**  $\lambda x_5. x_2$  is not a **BCI**- $\lambda$ -term, so no new terms are generated and so there is no **BCI**-proof of  $\tau$ .

EXAMPLE 91.

$$((a \rightarrow b \rightarrow c) \rightarrow d) \rightarrow (b \rightarrow a \rightarrow c) \rightarrow d.$$

**Step 1**  $x_1 : (a \rightarrow b \rightarrow c) \rightarrow d, x_2 : b \rightarrow a \rightarrow c, x_3 : a, x_4 : b,$

**Step 2**  $x_2 x_4 x_3 : c,$

**Step 3**  $\lambda x_3 x_4. x_2 x_4 x_3 : a \rightarrow b \rightarrow c,$

**Step 2**  $x_1 (\lambda x_3 x_4. x_2 x_4 x_3) : d,$

**Step 3**  $\lambda x_1 x_2. x_1 (\lambda x_3 x_4. x_2 x_4 x_3) : \tau.$

Example 86 also produced a **BCI**-term.

**THEOREM 92.** *Given a type  $\tau$ , the BCI long inhabitant search algorithm will, in finite time, produce an inhabitant or will demonstrate that  $\tau$  has no inhabitants. The algorithm will produce at most  $G(\tau)$  terms before terminating.*

**Proof.** Lemma 87 holds provided that  $V$  is a **BCI** term such that  $FV(V) = FV(U)$  and each free variable of  $V$  appears exactly once in  $V$ . If  $Z$  in Lemma 87 is a **BCI** term so is  $X$ .

The proof of the theorem now proceeds as for Theorem 89 except that in any substitution  $[V/U]X$  the above restrictions apply. Hence  $FV(Y') = FV(Y)$  was needed in Step 2 of the **BCI**-algorithm. In Steps 2 and 3 we must, for each subset of  $\{x_1, \dots, x_m\}$ , consider a term with those free variables with a particular atomic negative or long positive subtype of  $T$ . ■

### 6.5 **BCIW** Logic ( $R \rightarrow$ )

The **BCI**-search algorithm can easily be extended to **BCIW**-logic. By Theorem 51(4), the  $\lambda$ -terms required are from  $Once^+$  ().

#### *The **BCIW** Long Inhabitant Search Algorithm*

##### **Aim**

Given a type  $\tau$  to find a closed **BCIW**- $\lambda$ -term  $X$  in long normal form (if any) such that

$$\vdash X : \tau$$

##### **Method**

As for the **BCI** algorithm except that in Steps 2 and 3 each variable must appear in  $Y$  and  $Z$  at least once.

If the algorithm that we have to this stage fails, additional variables with the same types as the ones first given in Step 1 are added and the previous algorithm is repeated.

Note that, as it is not clear as to how many times new variables might need to be added, this method, while, as shown in Theorem 96, it leads to finding an inhabitant if there is one, does not constitute a decision procedure. The need for extra variables is illustrated in Example 95 below. This logic does have a decision procedure (see [Urquhart, 1990]), but its maximum complexity is related to Ackermann's function.

EXAMPLE 93.

$$\tau = [((a \rightarrow b) \rightarrow d) \rightarrow d] \rightarrow [(a \rightarrow b) \rightarrow d \rightarrow e] \rightarrow [a \rightarrow a \rightarrow b] \rightarrow a \rightarrow b$$

**Step 1**  $x_1 : ((a \rightarrow b) \rightarrow d) \rightarrow d, x_2 : (a \rightarrow b) \rightarrow d \rightarrow e,$   
 $x_3 : a \rightarrow a \rightarrow b, x_4 : a, x_5 : a \rightarrow b, x_6 : a$

**Step 2**  $x_3x_4x_4 : b, x_3x_4x_6 : b, x_3x_6x_6 : b, x_5x_4 : b, x_5x_6 : b$

**Step 3**  $\lambda x_4. x_3x_4x_4 : a \rightarrow b, \lambda x_4. x_3x_4x_6 : a \rightarrow b, \lambda x_6. x_3x_4x_6 : a \rightarrow b,$   
 $\lambda x_4. x_5x_4 : a \rightarrow b$



We cannot form  $\lambda x_1 x_2 x_3 x_4 . x_3 x_4 x_4$  as this is not a **BCIW**- $\lambda$ -term.

We can form no more terms by Step 2 as we have no terms of type  $d$  or  $(a \rightarrow b) \rightarrow d$  and no new terms of type  $a$ . Hence  $\tau$  has no **BCIW** inhabitant.

EXAMPLE 94.

$$\tau = (c \rightarrow a \rightarrow a \rightarrow a) \rightarrow c \rightarrow a \rightarrow a$$

**Step 1**  $x_1 : c \rightarrow a \rightarrow a \rightarrow a, x_2 : c, x_3 : a.$

**Step 2**  $x_1 x_2 x_3 x_3 : a.$

**Step 3**  $\lambda x_1 x_2 x_3 . x_1 x_2 x_3 x_3 : \tau.$

EXAMPLE 95.

$$\tau = c \rightarrow c \rightarrow (a \rightarrow a \rightarrow b) \rightarrow (c \rightarrow (a \rightarrow b) \rightarrow b) \rightarrow b$$

**Step 1**  $x_1 : c, x_2 : c, x_3 : a \rightarrow a \rightarrow b, x_4 : c \rightarrow (a \rightarrow b) \rightarrow b, x_5 : a$

**Step 2**  $x_3 x_5 x_5 : b$

**Step 3**  $\lambda x_5 . x_3 x_5 x_5 : a \rightarrow b$

**Step 2**  $x_4 x_1 (\lambda x_5 . x_3 x_5 x_5) : b,$   
 $x_4 x_2 (\lambda x_5 . x_3 x_5 x_5) : b,$

**Step 3** No new terms can be formed.

(Add to) Step 1  $x_6 : a$

**Step 2**  $x_3 x_5 x_6 : b, x_3 x_6 x_6 : b$

**Step 3**  $\lambda x_6 . x_3 x_5 x_6 : a \rightarrow b, \lambda x_5 . x_3 x_5 x_6 : a \rightarrow b$

**Step 2**  $x_4 x_1 (\lambda x_6 . x_3 x_5 x_6) : b, x_4 x_2 (\lambda x_6 . x_3 x_5 x_6) : b,$   
 $x_4 x_1 (\lambda x_5 . x_3 x_5 x_6) : b, x_4 x_2 (\lambda x_5 . x_3 x_5 x_6) : b,$

**Step 3**  $\lambda x_5 . x_4 x_1 (\lambda x_6 . x_3 x_5 x_6) : a \rightarrow b, \lambda x_5 . x_4 x_2 (\lambda x_6 . x_3 x_5 x_6) : a \rightarrow b,$

**Step 2**  $x_4 x_1 (\lambda x_5 . x_4 x_1 (\lambda x_6 . x_3 x_5 x_6)) : b,$   
 $x_4 x_2 (\lambda x_5 . x_4 x_1 (\lambda x_6 . x_3 x_5 x_6)) : b,$   
 $x_4 x_2 (\lambda x_5 . x_4 x_2 (\lambda x_6 . x_3 x_5 x_6)) : b,$

**Step 3**  $\lambda x_1 x_2 x_3 x_4 . x_4 x_2 (\lambda x_5 . x_4 x_1 (\lambda x_6 . x_3 x_5 x_6)) : \tau.$

Note that in Example 95  $x_4$  is used twice and so the one occurrence of  $a$  in  $c \rightarrow (a \rightarrow b) \rightarrow b$ , requires two variables of type  $a$ . If these were identified the resultant  $\lambda$ -term would no longer be **BCIW**-definable.

**THEOREM 96.** *Given a type  $\tau$ , the **BCIW** long inhabitant search algorithm will, in finite time, produce an inhabitant.*

**Proof.** Lemma 87 holds for **BCIW**-terms provided that we have  $FV(U) = FV(V)$ .

If  $Z$  in Lemma 88 is a **BCIW**-term so is  $X$ . In the counterpart to Lemma 88 the word “distinct” must also be dropped for the reasons illustrated in Example 95 above.

The proof of the theorem now proceeds as that of Theorem 91 except that multiple copies of variables may appear in substitutions and in terms formed by the algorithm.  $\blacksquare$

### 6.6 **BB'IW** Logic ( $T \rightarrow$ )

**BB'IW** search algorithm finds  $\lambda$ -terms of a given type that are in  $HRM() \cap Once^+(\cdot)$ , as required by Theorem 61.

#### *The **BB'IW** Long Inhabitant Search Algorithm*

##### **Aim**

Given a type  $\tau$  to find a closed **BB'IW**- $\lambda$ -term  $X$  in long normal form (if any) such that

$$\vdash X : \tau$$

**Step 1** As for **BCIW**.

**Step 2** For each atom  $b$  and for each subsequence  $(j_1, \dots, j_r)$  of  $(1, \dots, m)$  find one **BB'IW**- $\lambda$ -term  $Y \equiv x_{j_i} X_1 \dots X_k$  ( $k \geq 0, 1 \leq i \leq r$ ), such that  $Y \in HRM(j_1, \dots, j_r)$ ,  $Y : b$  and  $FV(Y) = \{x_{j_1}, \dots, x_{j_r}\}$ , if there is not already such a  $Y$ .

**Step 3** For each subsequence  $(j_1, \dots, j_r)$  of  $(1, \dots, m)$  and for each long positive subtype  $\beta$  of  $\tau$ , form a **BB'IW**- $\lambda$ -term  $Y$  by abstraction so that  $Y : \beta$  and  $FV(Y) = \{x_{j_1}, \dots, x_{j_r}\}$ , if we don't already have such a  $Y$ .

Now repeat steps 2 and 3 and if needs be add extra variables as for **BCIW**.

As with the **BCIW**-algorithm this does not, in general, provide a decision procedure.

EXAMPLE 97.

$$\tau = (a \rightarrow b \rightarrow c \rightarrow d) \rightarrow a \rightarrow c \rightarrow b \rightarrow d$$

**Step 1**  $x_1 : a \rightarrow b \rightarrow c \rightarrow d, x_2 : a, x_3 : c, x_4 : b$

**Step 2**  $x_1 x_2 x_4 x_3 : d$  and  $x_1 x_2 x_4 x_3 \in HRM(1, 2, 4, 3)$

**Step 3** The only term with a positive subtype of  $\tau$  that can be formed is  $\lambda x_1 x_2 x_3 x_4. x_1 x_2 x_4 x_3 : \tau$ , but this is not a **BB'IW**- $\lambda$ -term. Adding extra variables with the same types only allows us to generate this same (modulo- $\alpha$  conversion) inhabitant of  $\tau$ .

EXAMPLE 98.

$$\tau = (c \rightarrow a \rightarrow a \rightarrow a) \rightarrow c \rightarrow a \rightarrow a$$

The only **BCIW** definable- $\lambda$ -term inhabitant of  $\tau$  was  $\lambda x_1 x_2 x_3. x_1 x_2 x_3 x_3$ , this is also a **BB'IW** definable  $\lambda$ -term.

**THEOREM 99.** *Given a type  $\tau$ , the **BB'IW** long inhabitant search algorithm will, in finite time, produce an inhabitant.*

**Proof.** As for Theorem 96, except that we can replace subterms only by subterms belonging to the same class  $HRM(j_1, \dots, j_r)$  ( $1 \leq j_1 < \dots < j_r \leq m$ ). ■

### 6.7 **BB'I** Logic (*T-W, P-W*)

The search algorithm for this logic finds elements of the appropriate type in  $HRM(\ ) \cap Once(\ )$ , as required by Theorem 61(1).

*The **BB'I** Logic or T-W(P-W) Decision Procedure or the **BB'I** Long Inhabitant Search Algorithm*

#### **Aim**

Given a type  $\tau$  to find a closed **BB'I**- $\lambda$ -term  $X$  in long normal form (if any) such that

$$\vdash X : \tau$$

#### **Method**

As for **BB'IW** logic except that in the terms formed in **Step 2** no free variable may appear twice and that no extra variables need be added.

EXAMPLE 100.

$$\tau = (c \rightarrow a \rightarrow a \rightarrow a) \rightarrow c \rightarrow a \rightarrow a$$

The only **BB'IW** inhabitant of  $\tau$  is  $\lambda x_1 x_2 x_3. x_1 x_2 x_3 x_3$  and this is not a **BB'I**-definable  $\lambda$ -term. Thus  $\tau$  has no **BB'I** inhabitants.

EXAMPLE 101.

$$\tau = [(a \rightarrow a) \rightarrow a] \rightarrow (a \rightarrow a) \rightarrow a$$

**Step 1**  $x_1 : (a \rightarrow a) \rightarrow a, x_2 : a \rightarrow a, x_3 : a$

**Step 2**  $x_2 x_3 : a$

**Step 3**  $\lambda x_3. x_3 : a \rightarrow a \quad \lambda x_3. x_2 x_3 : a \rightarrow a$

**Step 2**  $x_1 (\lambda x_3. x_2 x_3) : a, x_1 (\lambda x_3. x_3) : a$

**Step 3**  $\lambda x_1 x_2. x_1(\lambda x_3. x_2 x_3) : \tau$

**THEOREM 102.** *Given a type  $\tau$  the **BB'I** long inhabitant search algorithm will, in finite time, produce an inhabitant or will demonstrate that  $\tau$  has no inhabitants. The algorithm will produce at most  $G(\tau)$  terms before terminating.*

**Proof.** As for Theorem 99, except that each variable  $x_i$  must appear exactly once in  $Y$  in any  $\lambda x_i. Y$  as with **BCI** logic. Also as in Theorem 89 and 92 the procedure can be bounded. Note that the number of subsequences of a sequence is the same as the number of subsets of the corresponding set. ■

### 6.8 **BB'IK** Logic

The search algorithm for this logic finds  $\lambda$ -terms of the appropriate type in  $PRM(\cdot) \cap Once^-(\cdot)$ .

*The **BB'IK** Logic Decision Procedure or the **BB'IK** Long Inhabitant Search Algorithm*

#### **Aim**

Given a type  $\tau$  to find a closed **BB'IK**- $\lambda$  term  $X$  in long normal form (if any) such that

$$\vdash X : \tau.$$

#### **Method**

As for **BB'I** logic except that in Step 2 *HRM* is replaced by *PRM*.

**EXAMPLE 103.**

$$\tau = b \rightarrow (b \rightarrow c) \rightarrow a \rightarrow c$$

**Step 1**  $x_1 : b, x_2 : b \rightarrow c, x_3 : a$

**Step 2**  $x_2 x_1 : c \quad (x_2 x_1 \in PRM(1, 2, 3))$

**Step 3**  $\lambda x_1 x_2 x_3. x_2 x_1 \in \tau.$

**THEOREM 104.** *Given a type  $\tau$  the **BB'IK** long inhabitant search algorithm will, in finite time, produce an inhabitant or will demonstrate that  $\tau$  has no inhabitant. The algorithm will produce at most  $G(\tau)$  terms before terminating.*

**Proof.** As for Theorem 99. ■

### 6.9 Some Other Logics

Bunder [1996] also gives the  $\lambda$ -terms definable in terms of the combinators **BB'**, **BT**, **BB'K**, **BIT**, **BITK**, **BITW**, **BTK**, and **BTW**. Those for **BCW** and **BB'W** can easily be found.

Inhabitant finding algorithms for these logics can easily be obtained as above. Decision procedures can be obtained for those without **W**.

*University of Wollongong, Australia*

### BIBLIOGRAPHY

- [Barendregt, 1984] H. P. Barendregt. *The Lambda Calculus*, North Holland, Amsterdam, 1984.
- [Ben-Yelles, 1979] C.-B. Ben Yelles. *Type Assignments in the Lambda Calculus*, Ph.D. thesis, University College, Swansea, Wales, 1979.
- [Bunder, 1996] M. W. Bunder. Lambda terms definable as combinators. *Theoretical Computer Science*, **169**, 3–21, 1996.
- [Bunder, 1995] M. W. Bunder. Ben-Yelles type algorithms and the generation of proofs in implicational logics. University of Wollongong, Department of Mathematics, Preprint Series no 3/95, 1995.
- [Church, 1932] A. Church. A set of postulates for the foundation of logic. *Annals of Mathematics*, **33**, 346–366, 1932.
- [Church, 1933] A. Church. A set of postulates for the foundation of logic. (second paper). *Annals of Mathematics*, **34**, 839–864, 1933.
- [Crossley and Shepherdson, 1993] J. N. Crossley and J. C. Shepherdson. Extracting programs from proofs by an extension of the Curry-Howard process. In *Logical Methods*, J. N. Crossley, J. B. Remmel, R. A. Shore and M. E. Sweedler, eds. pp. 222–288. Birkhäuser, Boston, 1993.
- [Curry, 1930] H. B. Curry. Grundlagen der Kombinatorischen Logik. *American Journal of Mathematics*, **52**, 509–536, 789–834, 1930.
- [Curry and Feys, 1958] H. B. Curry and R. Feys. *Combinatory Logic*, Vol 1, North Holland, Amsterdam, 1958.
- [de Bruin, 1970] N. G. de Bruin. *The Mathematical Language AUTOMATH, Its Usage, and Some of Its Extensions*. Vol. 125 of *Lecture Notes in Mathematics*, Springer Verlag, Berlin, 1970.
- [de Bruin, 1980] N. G. de Bruin. A survey of the project AUTOMATH. In *To H. B. Curry, Essays on Combinatory Logic, Lambda Calculus and Formalism*, J. R. Hindley and J. P. Seldin eds. pp. 576–606. Academic Press, London, 1980.
- [Dekker, 1996] A. H. Dekker. Brouwer 0.7.9- a proof finding program for intuitionistic, **BCI**, **BCK** and classical logic, 1996.
- [Helman, 1977] G. H. Helman. *Restricted lambda abstraction and the interpretation of some nonclassical logics*, Ph.D. thesis, Philosophy Department, University of Pittsburgh, 1977
- [Hindley, 1997] J. R. Hindley. *Basic Simple Type Theory*. Cambridge University Press, Cambridge, 1997.
- [Hindley and Seldin, 1986] J. R. Hindley and J. P. Seldin. *Introduction to Combinators and  $\lambda$ -Calculus*. Cambridge University Press, Cambridge, 1986.
- [Hirokawa, 1996] S. Hirokawa. The proofs of  $\alpha \rightarrow \alpha$  in  $P - W$ . *Journal of Symbolic Logic*, **61**, 195–211, 1996.
- [Howard, 1980] W. A. Howard. The formulae-as-types notion of construction. In *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, J. R. Hindley and J. P. Seldin eds. pp 479–490. Academic Press, London, 1980.

- [Lauchli, 1965] H. Lauchli. Intuitionistic propositional calculus and definably non-empty terms (abstract). *Journal of Symbolic Logic*, **30**, 263, 1965.
- [Lauchli, 1970] H. Lauchli. An abstract notion of realizability for which intuitionistic predicate calculus is complete. In *Intuitionism and Proof Theory*, A. Kino *et al.*, eds. pp. 227–234. North Holland, Amsterdam, 1970.
- [Oostdijk, 1996] M. Oostdijk. LambdaCal. 2—proof finding algorithms for intuitionistic and various (sub)classical propositional logics.  
Available on: <http://www.win.tue.nl/~martijno/lambdaCal/html>, 1996.
- [Schönfinkel, 1924] M. Schönfinkel. Über die Bausteine der mathematische Logik. *Mathematische Annale*, **92**, 305–316, 1924. English translation in *From Frege to Gödel*, J. van Heijenoort, ed. pp. 355–366. Harvard University Press, 1967.
- [Scott, 1970] D. S. Scott. Constructive validity, In Vol 125 of *Lecture notes in Mathematics*, pp. 237–275. Springer Verlag, Berlin, 1970.
- [Trigg *et al.*, 1994] P. Trigg, J. R. Hindley and M. W. Bunder. Combinatory abstraction using  $\mathbf{B}, \mathbf{B}'$  and friends. *Theoretical Computer Science*, **135**, 405–422, 1994.
- [Turing, 1942] A. M. Turing. R. O. Gandy An early proof of normalization by A. M. Turing. In *To H. B. Curry: Essays...*, pp. pp 453–455, 1980.
- [Urquhart, 1990] A. Urquhart. The complexity of decision procedures in relevant logic. In *Truth or Consequences: Essays in honour of Noel Belnap*. J. M. Dunn and A. Gupta, eds. pp. 61–75, Reidel, Dordrecht, 1990.

GRAHAM PRIEST

## PARACONSISTENT LOGIC

*Indeed, even at this stage, I predict a time when there will be mathematical investigations of calculi containing contradictions, and people will actually be proud of having emancipated themselves from ‘consistency’.* Ludwig Wittgenstein, 1930.<sup>1</sup>

### 1 INTRODUCTION

Paraconsistent logics are those which permit inference from inconsistent information in a non-trivial fashion. Their articulation and investigation is a relatively recent phenomenon, even by the standards of modern logic. (For example, there was no article on them in the first edition of the *Handbook*.) The area has grown so rapidly, though, that a comprehensive survey is already impossible. The aim of this article is to spell out the basic ideas and some applications. Paraconsistent logic has interest for philosophers, mathematicians and computer scientists. As befits the *Handbook*, I will concentrate on those aspects of the subject that are likely to be of more interest to philosopher-logicians. The subject also raises many important philosophical issues. However, here I shall tread over these very lightly—except in the last section, where I shall tread over them lightly.

I will start in part 2 by explaining the nature of, and motivation for, the subject. Part 3 gives a brief history of it. The next three parts explain the standard systems of paraconsistent logic; part 4 explains the basic ideas, and how, in particular, negation is treated; parts 5 and 6 discuss how this basic apparatus is extended to handle conditionals and quantifiers, respectively. In part 7 we look at how a paraconsistent logic may handle various other sorts of machinery, including modal operators and probability. The next two parts discuss the applications of paraconsistent logic to some important theories; part 8 concerns set theory and semantics; part 9, arithmetic. The final part of the essay, 10, provides a brief discussion of some central philosophical aspects of paraconsistency.

In writing an essay of this nature, there is a decision to be made as to how much detail to include concerning proofs. It is certainly necessary to include many proofs, since an understanding of them is essential for anything other than a relatively modest grasp of the subject. On the other hand, to prove everything in full would not only make the essay extremely long, but distract from more important issues. I hope that I have struck a happy *via media*.

---

<sup>1</sup>Wittgenstein [1975], p. 332.

Where proofs are given, the basic definitions and constructions are spelled out, and the harder parts of the proof worked. Routine details are usually left to the reader to check, even where this leaves a considerable amount of work to be done. In many places, particularly where the material is a dead end for the purposes of this essay, and is easily available elsewhere, I have not given proofs at all, but simply references. Those for whom a modest grasp of the subject is sufficient may, I think, skip all proofs entirely.

Paraconsistent logic is strongly connected with many other branches of logic. I have tried, in this essay, not to duplicate material to be found in other chapters of this *Handbook*, and especially, the chapter on Relevant Logic. At several points I therefore defer to these. There is no section of this essay entitled ‘Further Reading’. I have preferred to indicate in the text where further reading appropriate to any particular topic may be found.<sup>2</sup>

## 2 DEFINITION AND MOTIVATION

### 2.1 Definition

The major motivation behind paraconsistent logic has always been the thought that in certain circumstances we may be in a situation where our information or theory is inconsistent, and yet where we are required to draw inferences in a sensible fashion. Let  $\vdash$  be any relationship of logical consequence. Call it *explosive* if it satisfies the condition that for all  $\alpha$  and  $\beta$ ,  $\{\alpha, \neg\alpha\} \vdash \beta$ , *ex contradictione quodlibet* (ECQ). (In future I will omit set braces in this context.) Both classical and intuitionist logics are explosive. Clearly, if  $\vdash$  is explosive it is not a sensible inference relation in an inconsistent context, for applying it gives rise to *triviality*: everything. Thus, a minimal condition for a suitable inference relation in this context is that it not be explosive. Such inference relationships (and the logics that have them) have come to be called *paraconsistent*.<sup>3</sup>

Paraconsistency, so defined, is something of a minimal condition for a logic to be used as envisaged; and there are logics that are paraconsistent but not really appropriate for the use. For example, Johansson’s minimal logic is paraconsistent, but satisfies  $\alpha, \neg\alpha \vdash \neg\beta$ . One might therefore attempt a stronger constraint on the definition of ‘paraconsistent’, such as: for no syntactically definable class of sentences (e.g., negated sentences),  $\Sigma$ , do

---

<sup>2</sup>The most useful general reference is Priest *et al.* [1989] (though this is already a little dated). That book also contains a bibliography of paraconsistency up to about the mid-1980s.

<sup>3</sup>The word was coined by Miró Quesada at the Third Latin American Symposium on Mathematical Logic, in 1976. Note that a paraconsistent logic need not itself have an inconsistent set of logical truths: most do not. But there are some that do, e.g., any logic produced by adding the connexivist principle  $\neg(\alpha \rightarrow \neg\alpha)$  to a relevant logic at least as strong as *B*. See Mortensen [1984].



we have  $\alpha, \neg\alpha \vdash \sigma$ , for all  $\sigma \in \Sigma$ . This seems too strong, however. In many logics,  $\alpha, \neg\alpha \vdash \beta$ , for every logical truth,  $\beta$ . If the logic is decidable, then there is a clear sense in which the set of logical truths is syntactically characterisable. Yet such logics would still be acceptable for many paraconsistent purposes. Hence, this definition would seem to be too strong.<sup>4</sup>

In his [1974], da Costa suggests another couple of natural constraints on a paraconsistent logic, of a rather different nature. One is to the effect that the logic should not contain  $\neg(\alpha \wedge \neg\alpha)$  as a logical truth. The rationale for this is not spelled out. However, I take it that the idea is that if one has information that contains  $\alpha$  and  $\neg\alpha$  one does not want to have a logical truth that contradicts this. Why not though? Since one is not ruling out inconsistency *a priori*, there would seem to be nothing *a priori* against this (though maybe for particular applications one would not want the situation to arise). As a general condition, then, it seems too strong. And certainly a number of the logics that we will consider have  $\neg(\alpha \wedge \neg\alpha)$  as a logical truth.

Another of the constraints that da Costa suggests is to the effect that the logic should contain as much of classical—or at least intuitionist—logic, as does not interfere with its paraconsistent nature. The condition is somewhat vague, though its intent is clear enough; and again, it is too strong. It assumes that a paraconsistent logician must have no objection to other aspects of classical or intuitionist logic, and this is clearly not true. For example, a relevant logician might well object to paradoxes of implication, such as  $\alpha \rightarrow (\beta \rightarrow \alpha)$ .<sup>5</sup>

As an aside, let me clarify the relationship between relevant logics and paraconsistent logics. The motivating concern of relevant logic is somewhat different from that of paraconsistency, namely to avoid paradoxes of the conditional. Thus, one may take a relevant (propositional) logic to be one such that if  $\alpha \rightarrow \beta$  is a logical truth then  $\alpha$  and  $\beta$  share a propositional parameter. The interests of relevant and paraconsistent logics clearly converge at many points. Relevant logics and paraconsistent logics are not coextensive, however. There are many paraconsistent logics that are not relevant, as we shall see. The relationship the other way is more complex, since there are different ways of using a relevant logic to define a consequence relation. A natural way is to say that  $\alpha \vdash \beta$  iff  $\alpha \rightarrow \beta$  is a logical truth. Such a consequence relation is clearly paraconsistent. Another is to define logical consequence as deducibility, defined in the standard way, using some set of axioms and rules for the relevant logic. Such a consequence relation may, but need not, be relevant. For example, Ackermann's original formulation of  $E$  contained the rule  $\gamma$ : if  $\vdash \alpha$  and  $\vdash \neg\alpha \vee \beta$  then  $\vdash \beta$ . This gives explo-

---

<sup>4</sup>Further attempts to tighten up the definition of paraconsistency along these lines can be found in Batens [1980] (in the definition of 'A-destructive', p. 201, clause (i) should read  $\not\vdash_L A$ ), and Urbas [1990].

<sup>5</sup>Indeed, it is just this principle that ruins minimal logic for serious paraconsistent purposes. For  $\alpha$  and  $\alpha \rightarrow \perp$  (i.e.,  $\neg\alpha$ ) give  $\perp$ , and the principle then gives  $\beta \rightarrow \perp$ .

sion by an argument often called the ‘Lewis Independent Argument’, that we will meet in a moment.

Anyway, and to return from the digression: the definition of paraconsistency given here is weaker than sufficient to *guarantee* sensible application in inconsistent contexts; but an elegant stronger definition is not at hand, and since the one in question has become standard, I will use it to define the contents of this essay.

## 2.2 *Inconsistency and Dialetheism*

Numerous examples of inconsistent information/theories from which one might want to draw inferences in a controlled way have been offered by paraconsistent logicians. For example:

1. information in a computer data base;
2. various scientific theories;
3. constitutions and other legal documents;
4. descriptions of fictional (and other non-existent) objects;
5. descriptions of counterfactual situations.

The first of these is fairly obvious. As an example of the second, consider, e.g., Bohr’s theory of the atom, which required bound electrons both to radiate energy (by Maxwell’s equations) and not to (since they do not spiral inwards towards the nucleus). As an example of the third, just consider a constitution that gives persons of kind *A* the right to do something, *x*, and forbids persons of kind *B* from doing *x*. Suppose, then, that a person in both categories turns up. (We may assume that it had never occurred to the legislators that there might be such a person.) In the fourth case, the information (in, say, a novel or a myth) characterises an object, and turns out—deliberately or otherwise—to be inconsistent. To illustrate the fifth, suppose, for example, that we need to compute the truth of the conditional: if you were to square the circle, I would give you all my money. Applying the Ramsey-test, we see what follows from the antecedent (which is logically impossible), together with appropriate background assumptions. (And I would *not* give you all my money!)<sup>6</sup>

There is no suggestion here that in every case one must remain content with the inconsistent information in question. One might well like to remove

---

<sup>6</sup>Many of these examples are discussed further in Priest *et al.* [1989], ch. 18. The Bohr case is discussed in Brown [1993]. Another kind of example that is sometimes cited is the information provided by witnesses at a trial. I find this less persuasive. It seems to me that the relevant information here is all of the form: witness *x* says so and so. (That a witness is lying, or making an honest mistake, is always a possibility to be taken into account.) And any collection of statements of this form is quite consistent.

some of the inconsistent information in the data base; reject or revise the scientific theory; change the law to eliminate the inconsistency. But this is not possible in all of the cases given, e.g., for counterfactual conditionals with impossible antecedents. And even where it is, this not only may take time; it is often not clear how to do so satisfactorily. (The matter is certainly not algorithmic.) While we figure out how to do it, we may still be in a situation where inference is necessary, perhaps for practical ends, e.g., so that we can act on the information in the data base; or manipulate some piece of scientific technology; or make decisions of law (on other than an obviously inconsistent case). Moreover, since there is no decision procedure for consistency, there is no guarantee that any revision will achieve consistency. We cannot, therefore, be sure that we have succeeded. (This is particularly important in the case of the data base, where the deductions go on “behind our back”, and the need to revise may never become apparent.)

In cases of this kind, then, even though we may not, ideally, be satisfied with the inconsistent information, it may be desirable—indeed, practically necessary—to use a paraconsistent logic. Moreover, we know that many scientific theories are false; they may still be important because they make correct predications in most, or even all, cases; they may be good *approximations* to the truth, and so on. These points remain in force, even if the theories in question contain contradictions, and so are (thought to be) false for logical reasons. Of course, this is not so if the theories are trivial; but that’s the whole point of using a paraconsistent logic.

One can thus subscribe to the use of paraconsistent logics for some purposes without believing that inconsistent information or theories may be *true*. The view that some *are* true has come to be called *dialetheism*, a dialetheia being a true contradiction.<sup>7</sup> If the truth about some subject is dialethic then, clearly, a paraconsistent logic needs to be employed in reasoning about that subject. (I take it to be uncontroversial that the set of truths is not trivial. Why this is so, especially once one has accepted dialetheism is, however, a substantial question.)

Examples of situations that may give rise to dialetheias, and that have been proposed, are of several kinds, including:

1. certain kinds of moral and legal dilemmas;
2. borderline cases of vague predicates;
3. states of change.

Thus, one may suppose, in the legal example mentioned before, that a person who is *A* and *B* both has and has not the right to do *x*; or that in

---

<sup>7</sup>The term was coined by Priest and Routley in 1981. See Priest *et al.* [1989], p. xx. Note that some writers prefer ‘dialethism’.

a case of light drizzle it both is and is not raining; or that at the instant a moving object comes to rest, it both is and is not in motion.<sup>8</sup>

The most frequent and, arguably, most persuasive examples of dialetheias that have been given are the paradoxes of self-reference, such as the Liar Paradox and Russell's Paradox. What we have in such cases, are apparently sound arguments resulting in contradictions. There are many suggestions as to what is wrong with such arguments, but none of them is entirely happy. Indeed, in the case of the semantic paradoxes there is not (even after 2,000 years) any consensus concerning the most plausible way to go. This gives the thought that the arguments are, after all, sound, its appeal.<sup>9</sup>

Naturally, all the examples cited in this section are contestable. I will return to the issue of possible objections in the last part of this essay.

### 3 A BRIEF HISTORY OF PARACONSISTENT LOGIC

#### 3.1 *The Law of Non-contradiction and Paraconsistency*

During the history of Western Philosophy, there have been a number of figures who deliberately endorsed inconsistent views.<sup>10</sup> The earliest were some Presocratics, including Heraklitus. In the middle ages, some Neo-Platonists, such as Nicholas of Cusa, endorsed contradictory views. In the modern period, the most notable advocate of inconsistent views was Hegel.<sup>11</sup> These figures are relatively isolated, however. It is something of an understatement to say that the dominant orthodoxy in Western Philosophy has been strongly hostile to inconsistency.<sup>12</sup> Consistency has been taken to be pivotal to a number of fundamental notions, such as truth and rational belief. This antipathy to contradiction is, historically, due in large part to Aristo-

---

<sup>8</sup>Many of these examples are discussed further in Priest *et al.* [1989], ch. 18, 2.2. A discussion of transition states and legal dialetheias can be found in chs. 11 and 13 of Priest [1987]. Moral dilemmas are also discussed in Routley and Routley [1989]. The dialethic nature of vagueness is advocated in Peña [1989]. It has also been suggested that some contradictions in the Hegel/Marx tradition are dialethic. For a discussion of this, see Priest [1989a].

<sup>9</sup>For further discussion, see Routley [1979] and Priest [1987], chs. 1-3.

<sup>10</sup>And nearly every great philosopher has unwittingly endorsed inconsistent views.

<sup>11</sup>In each case, there is, of course, some—though, I would argue, misguided—possibility for exegetical attempts to render the views consistent. Other modern philosophers whose thought also appears to endorse inconsistency are Meinong and the later Wittgenstein. In their cases there is more scope for exegetical evasion. For further discussion on all these matters, see Priest *et al.* [1989], chs. 1, 2.

<sup>12</sup>Eastern philosophy has been notably less so—though there is, again, room for exegetical debate. The most natural interpretation of Jaina philosophy has them endorsing inconsistent positions. And major Buddhist logicians of the stature of Nāgārjuna held that it was quite possible for statements to be both true and false. Significant elements of inconsistency can also be found in Chinese philosophy. For further discussion of all this, see Priest *et al.* [1989], ch. 1, sect. 2.

tle's defense of the Law of Non-contradiction in the *Metaphysics*.<sup>13</sup> Given this situation, it may therefore be surprising that the orthodoxy against paraconsistency is a relatively recent phenomenon.

### 3.2 Paraconsistency Before the Twentieth Century

The major account of validity until this century was, of course, Aristotelian Syllogistic. Now, consider any sentences of the Syllogistic *E* and *I* forms; for example, 'No women are white' and 'Some women are white'. These are contradictories. But the inference from them to, e.g., 'All cows are black', is not a valid syllogism. Syllogistic is not, therefore, explosive: it is paraconsistent.

It might be suggested that it is more appropriate to look for explosion in accounts of propositional inference. Here the story is more complex, but the conclusion is similar. Aristotle had no elaborated account of propositional inference. However, there are comments that bear on the matter scattered through the *Organon*, and they have a distinctly paraconsistent flavour. For example, in the *Prior Analytics* (57<sup>b</sup>3), Aristotle states that contradictories cannot both entail the same thing. It would seem to follow that Aristotle did not endorse at least one of (in modern notation)  $\alpha \wedge \neg\alpha \vdash \alpha$  and  $\alpha \wedge \neg\alpha \vdash \neg\alpha$ . For contraposing (a move that Aristotle endorses immediately before), we obtain  $\alpha \vdash \neg(\alpha \wedge \neg\alpha)$  and  $\neg\alpha \vdash \neg(\alpha \wedge \neg\alpha)$ . Hence, not everything can follow from a contradiction. In fact, there are reasons to suppose that Aristotle held a view of negation according to which the negation of any claim cancels that claim out. A contradiction has, therefore, no content, and entails nothing. This view of negation (which would now be called 'connexivist') was endorsed by a number of subsequent logicians (notably Abelard) well into the late middle ages.<sup>14</sup>

A theory of propositional inference was worked out much more thoroughly by Stoic logicians, and the explosive nature of their theories is more plausible for the following reason. There is a famous argument for ECQ, often called the Lewis (independent) argument, after C. I. Lewis. This goes (in natural deduction form) as follows:

$$\frac{\alpha \quad \frac{\neg\alpha}{\neg\alpha \vee \beta}}{\beta}$$

<sup>13</sup>Book  $\Gamma$ , ch 4. The historical success of this defence is, however, out of all proportion to its intellectual weight. See Priest [1998e].

<sup>14</sup>Much of this and the rest of the material in this subsection is documented in Sylvan [2000], ch. 4. The discussion there is carried out in terms of the conditional, though it is equally applicable to the consequence relation.

The argument uses just two principles (three if you include the transitivity of deducibility): Addition ( $\alpha \vdash \alpha \vee \beta$ ) and the Disjunctive Syllogism ( $\alpha, \neg\alpha \vee \beta \vdash \beta$ ). As we shall see in due course, the Disjunctive Syllogism (DS) has, unsurprisingly, been rejected by most paraconsistent logicians. Now Stoic logicians endorsed just this principle. The explosive nature of their logic would therefore seem a good bet. Despite this, it probably was not: there is reason to suppose that their disjunction was an intensional one that required some kind of connection between  $\alpha$  and  $\beta$  for the truth of  $\alpha \vee \beta$ . If this is the case, Addition fails in general, as does the Lewis argument.

It is not known who discovered the Lewis argument. Martin [1985] conjectures that it was William of Soissons in the 12th Century. (It was certainly known to, and endorsed by, some later logicians, such as Scotus and Buridan.) At any rate, William was a member of a group of logicians called the Parvipontanians, who were known not only for living by a small bridge, but for defending ECQ. This group may therefore herald the arrival of explosion on the philosophical stage. Whether or not this is so, after this time, some logicians endorsed explosion, some rejected it, different orthodoxies ruling at different times and different places (though, possibly, the explosive view was more common). One group of logicians who rejected it is notable, since they very much prefigure modern paraconsistent logicians. This is the Cologne School of the late 15th Century, who argued against the DS on the ground that if you start by *assuming* that  $\alpha$  and  $\neg\alpha$ , then you cannot appeal to  $\alpha$  to rule out  $\neg\alpha$  as the DS manifestly does.

Notoriously, logic made little progress between the end of the Middle Ages and the start of the third great period in logic, towards the end of the 19th Century. With the work of logicians such as Boole and Frege, we see the mathematical articulation of an explosive logical theory that has come to be known, entirely inappropriately, as ‘classical logic’. Though, in its early years, many objected to its explosive features, it has achieved a hegemony (though never a universality) in the logical community, in a (historically) very brief space of time. Whether this is because the truth was definitively and transparently revealed, or because at the time it was the only game in town, history will tell.

### 3.3 *The Twentieth Century*

A feature of paraconsistent logic this century is that the idea appears to have occurred independently to many different people, at different times and places, working in ignorance of each other, and often motivated by somewhat different considerations. Some, notably, for example, da Costa, have been motivated by the idea that inconsistent theories might be of intrinsic importance. Others, notably the early relevant logicians, were motivated simply by the idea that explosion, as a property of entailment, is

just too counter-intuitive.<sup>15</sup>

The earliest paraconsistent logics (that I am aware of) were given by two Russians. The first of these was Vasil'ev. Starting about 1910, Vasil'ev proposed a modified Aristotelian syllogistic, according to which there is a new form:  $S$  is both  $P$  and not  $P$ . How, exactly, this form was to be interpreted is contentious, though, a problem exacerbated by the fact that he was not in a position to employ the techniques of modern logic. This is not true of the second logician, Orlov, who, in 1929, gave the first axiomatisation of the relevant logic  $R$ . Sadly, the work of neither Vasil'ev nor Orlov made any impact at the time.<sup>16</sup>

An important figure who did have a good deal of influence was the Polish logician and philosopher Lukasiewicz. Partly influenced by Meinong's account of impossible objects, Lukasiewicz clearly envisaged the construction of paraconsistent logics in his seminal 1910 critique of Aristotle on the Law of Non-contradiction.<sup>17</sup> And it was his erstwhile student, Jaškowski, who, in 1948, produced the first non-adjunctive paraconsistent logic.<sup>18</sup>

Paraconsistent logics were again, independently, proposed in South America in doctoral dissertations by Asenjo (1954, Argentina) and da Costa (1963, Brazil). Asenjo proposed the first many-valued paraconsistent logic. Da Costa gave axiom systems for a certain family of paraconsistent logics (the  $C$  systems), and produced the first quantified paraconsistent logic. Many co-workers, such as Arruda and Loparić, joined da Costa in the next 20 years, to produce an active school of paraconsistent logicians at Campinas (and later São Paulo). They developed non-truth-functional semantics for the  $C$  systems, and articulated the subject in various other ways; this included "rediscovering" Vasil'ev, taking up the work of Jaškowski, and formulating various other paraconsistent systems.<sup>19</sup>

Guided by considerations of relevance, an entirely different approach to paraconsistency was proposed in England by Smiley in [1959], who articulated the first filter logic. Starting at about the same time, and drawing on the earlier work of Ackermann and Church, Anderson and Belnap in the USA proposed a number of relevant paraconsistent logics of a different kind. A research school quickly grew up around them in Pittsburgh, which included co-workers such as Meyer and Dunn.<sup>20</sup> The algebraic semantics

---

<sup>15</sup>The later Wittgenstein was also sympathetic to paraconsistency for various reasons, though he never articulated a paraconsistent logic. See, e.g., Marconi [1984].

<sup>16</sup>On Vasil'ev see Priest *et al.* [1989], ch. 3, 2.2 and Arruda [1977]. On Orlov, see Anderson *et al.* [1992], p. xvii.

<sup>17</sup>A synopsis of this is published in English in Lukasiewicz [1971].

<sup>18</sup>For a discussion of Lukasiewicz and Jaškowski, see Priest *et al.* [1989], ch. 3, 2.1, 2.3. Jaškowski's work is translated into English in his [1969].

<sup>19</sup>Discussion and bibliography can be found in Priest *et al.* [1989], 5.6. The most accessible introduction to Asenjo's work is his [1966], and to da Costa's is his [1974]. Da Costa and Marconi [1989] reports much of the work of da Costa and his co-workers.

<sup>20</sup>The work of this school is recorded in Anderson and Belnap [1975], and Anderson *et*

for relevant logics, in particular, was inaugurated largely by Dunn's 1966 doctoral thesis.<sup>21</sup>

Investigation of things paraconsistent in Australia took off in the early 1970s with the discovery of world (intensional) semantics for negation by R. Routley (now Sylvan) and V. Routley (now Plumwood). This was developed into an intensional semantics for the Anderson/Belnap logics—and many others—by Routley,<sup>22</sup> Meyer (now in Australia), and a school that developed around them in Canberra, which included workers such as Brady and Mortensen. These semantics made the paraconsistent aspects of relevant logics plain.<sup>23</sup> Later in the 1970s the cudgel for dialetheism was taken up by Priest (now Priest) and Routley.<sup>24</sup>

By the mid-1970s the paraconsistent movement was a fully international one, with workers in all countries cooperating (though not necessarily agreeing!), and with logicians working in numerous countries other than the ones already mentioned, including Belgium, Bulgaria, Canada and Italy. Some feel for the state of the subject at the end of the 70s can be obtained from Priest *et al.* [1989].<sup>25</sup> The rest, as they say, is not history.

#### 4 BASIC TECHNIQUES OF PARACONSISTENT LOGICS

An understanding of most paraconsistent logics can be obtained by looking at the strategies employed in virtue of which ECQ fails. There are many techniques for achieving this end. In this part, I will describe the most fundamental. In the process, we will meet dozens of different systems of paraconsistent logic, often constructed along very different lines. It is therefore necessary to have some common medium for comparison. I have chosen to make this semantics, and will specify systems in terms of these. (I would warn straight away though, that many of the systems we will meet appeared first in proof theoretic terms. Indeed, some of the authors of these systems—e.g. Tennant—would privilege proof theory over semantics.) When I give details of corresponding proof theories, I will use the sort of proof theory (natural deduction, sequent calculus, or axiomatic) that seems most natural for the logic.

Because paraconsistency concerns only negation essentially, we can see the essentials of paraconsistent logics in languages with very little logical

---

*al.* [1992].

<sup>21</sup>See Anderson and Belnap [1975], and also the article on Relevant Logic in this volume of the *Handbook*.

<sup>22</sup>Whenever the name 'Routley' is used without initial in this essay, it refers to Sylvan.

<sup>23</sup>The work of this group is most accessible in Routley *et al.* [1982].

<sup>24</sup>See, e.g., Routley [1979]. Priest's early work on the area is most accessible in Priest [1987].

<sup>25</sup>Despite the date, all the work in the collection was finished by 1980. A number of papers produced at the same time, that were not included in this, were published in a special issue of *Studia Logica* on paraconsistent logics (**43** (1984), nos. 1 & 2).



apparatus. In this part, we will be concerned with a propositional language whose only connectives are negation,  $\neg$ , conjunction,  $\wedge$ , and disjunction,  $\vee$ . I will use lower case Roman letters, starting with  $p$ , for propositional parameters, lower case Greeks, starting with  $\alpha$ , for arbitrary formulas, and upper case Greeks for sets of formulas. I will use  $\models_C$ ,  $\models_I$ , and  $\models_{S5}$  for the consequence relations of classical logic, intuitionist logic and  $S5$ , respectively, and  $\models$  for the semantic consequence relation of whichever paraconsistent logic happens to be the topic of discussion. If a proof theory is involved, I will use  $\vdash$  for the corresponding notion of deducibility.

#### 4.1 Filtration

One of the simplest ways to prevent explosion is to filter it, and any other undesirables, out. Consider, for simplicity, the one-premise case. (Finite sets of premises can always be reduced to this by conjoining.) Let  $F(\alpha, \beta)$  be any relationship between formulas. Define an inference from  $\alpha$  to  $\beta$  to be *prevalid* iff  $\alpha \models_C \beta$  and  $F(\alpha, \beta)$ . The thought here is that for an inference to be correct, something more than classical truth-preservation is required, e.g., some *connection* between premise and conclusion. This is expressed by  $F$ . Usually, prevalidity is too weak as a notion of validity, since, in general, it is not closed under uniform substitution, and this is normally taken to be a desideratum for any notion of validity. However, closure can be ensured if we define an inference to be valid iff it is a uniform substitution instance of a prevalid inference.

What inferences are valid depends, of course, entirely on the filter,  $F$ . One that naturally and obviously gives rise to a paraconsistent logic is:  $F(\alpha, \beta)$  iff  $\alpha$  and  $\beta$  share a propositional parameter. (This collapses the notions of validity and prevalidity, since if  $\alpha$  and  $\beta$  share a propositional parameter, so do uniform substitution instances thereof.) This logic is not a very interesting paraconsistent one, however, since, as is clear,  $p \wedge \neg p \models \alpha$  where  $\alpha$  is any formula containing the parameter  $p$ .<sup>26</sup>

A different filter, proposed by Smiley [1959] is:  $F(\alpha, \beta)$  iff  $\alpha$  is not a (classical) contradiction and  $\beta$  is not a (classical) tautology.<sup>27</sup> (Note that, according to this definition,  $\alpha \wedge \neg \alpha / \alpha$  is not prevalid, but it is valid, since it is an instance of  $p \wedge q / p$ .) It is easy to see that on this account  $p \wedge \neg p$  does not entail  $q$ . The major notable feature of filter logics is that, in general, transitivity of deducibility breaks down.<sup>28</sup> For example, using

---

<sup>26</sup>A stronger filter is one to the effect that *all* the variables of the premise occur in the conclusion. This gives rise to a logic in the family of analytic implications. On this family, see Anderson and Belnap [1975], sect. 29.6.

<sup>27</sup>Filters of a related kind were also suggested by Geach and von Wright. See Anderson and Belnap [1975], sect. 20.1.

<sup>28</sup>Though it need not. First Degree Entailment, where transitivity holds, can be seen as a filter logic. See Dunn [1980].

Smiley's filter, it is easy to see that  $p \wedge \neg p \models p \wedge (\neg p \vee q)$ ,  $p \wedge (\neg p \vee q) \models q$ , but  $p \wedge \neg p \not\models q$ .

One of the most interesting filter logics, given by Tennant [1984], is obtained by generalising Smiley's approach. Let  $\Pi$  and  $\Sigma$  be sets of sentences, and let ' $\Pi \models_C \Sigma$ ' be understood in the natural way (every classical evaluation that makes every member of  $\Pi$  true makes some member of  $\Sigma$  true). Define the inference from  $\Pi$  to  $\Sigma$  to be prevalid iff:  $\Pi \models_C \Sigma$  and for no proper subsets of  $\Pi$  and  $\Sigma$ ,  $\Pi'$  and  $\Sigma'$ , respectively, do we have  $\Pi' \models_C \Sigma'$ . Validity is then defined by closing under substitution as before. In this account, a valid inference is one which is classically valid, and minimally so: there is no "noise" amongst premise and conclusion set.<sup>29</sup>

Tennant's  $\models$  is obviously non-monotonic (that is, adding extra premises may invalidate an inference). It also has the following property: if  $\Pi \models_C \Sigma$ , then there are subsets of  $\Pi$  and  $\Sigma$ ,  $\Pi'$  and  $\Sigma'$ , respectively, such that  $\Pi' \models \Sigma'$ . For if  $\Pi \models_C \Sigma$ , we can simply throw out premises and/or conclusions until this is no longer true; the result is a prevalid, and so valid, inference. In particular, if  $\Pi \models_C \alpha$  then for some  $\Pi' \subseteq \Pi$ ,  $\Pi' \models \alpha$  or  $\Pi' \models \phi$ . In the first case,  $\alpha$  follows validly from *part* of  $\Pi$ ; in the second, part of  $\Pi$  can be shown to be inconsistent by valid reasoning.

Filtration can also be applied proof theoretically: we start with classical proofs and throw out those that do not satisfy some specific criteria. Tennant's logic can be characterised proof-theoretically in just this way. For finite premises and conclusions, the valid inferences are exactly those that are provable in the Gentzen sequent calculus for classical logic, but which do not use the structural rules of dilution (thinning) and cut. Specifically, consider the sequent calculus whose basic sequents are of the form  $\alpha : \alpha$ , and whose rules are as follows. ( $\Pi_1, \Pi_2$  means  $\Pi_1 \cup \Pi_2$ ; similarly,  $\Pi, \alpha$  means  $\Pi \cup \{\alpha\}$ , and if something of this form occurs as a premise of a rule, it is to be understood that  $\alpha \notin \Pi$ ).

$$\frac{\Pi, \alpha : \Delta}{\Pi : \Delta, \neg \alpha} \qquad \frac{\Pi : \Delta, \alpha}{\Pi, \neg \alpha : \Delta}$$

$$\frac{\Pi, \alpha : \Delta}{\Pi, \alpha \wedge \beta : \Delta} \qquad \frac{\Pi, \beta : \Delta}{\Pi, \alpha \wedge \beta : \Delta} \qquad \frac{\Pi_1 : \Delta_1, \alpha \quad \Pi_2 : \Delta_2, \beta}{\Pi_1, \Pi_2 : \Delta_1, \Delta_2, \alpha \wedge \beta}$$

$$\frac{\Pi : \Delta, \alpha}{\Pi : \Delta, \alpha \vee \beta} \qquad \frac{\Pi : \Delta, \beta}{\Pi : \Delta, \alpha \vee \beta} \qquad \frac{\Pi_1, \alpha : \Delta_1 \quad \Pi_2, \beta : \Delta_2}{\Pi_1, \Pi_2, \alpha \vee \beta : \Delta_1, \Delta_2}$$

<sup>29</sup>The restriction of Tennant's approach to the one-premise, one-conclusion, case obviously gives Smiley's account. Smiley himself, handles the multiple-premise case, simply by conjoining. As Tennant points out ([1984], p. 199), this generates a different account from his. It is not difficult to check that  $p \vee q, \neg(p \vee q) \not\models p \wedge q$  for Smiley. (The conjoined antecedent is a contradiction; and any inference of which the conjoined form is a substitution instance is not classically valid.) But it is valid for Tennant, since it is a substitution instance of  $p \vee q, r \vee s, \neg(t \vee q), \neg(r \vee u) \models p \wedge s$ .

Then we have  $\Pi \models \Sigma$  iff the sequent  $\Pi : \Sigma$  is provable. For the proof see Tennant [1984].<sup>30</sup>

Tennant's account of inference seems to capture very nicely what one might call the 'essential core' of classical inference. As an inference engine to be applied to inconsistent information/theories, it could not be applied in the obvious way, however. This is because information is often heavily redundant. For example, for Tennant's  $\models$ , we do not have  $p, \neg p \vee q, q \models q$ . Yet given the information in the premises, it would certainly seem that we are entitled to infer  $q$ . Presumably, then, we would take  $\alpha$  to follow from  $\Sigma$  iff for some  $\Sigma' \subseteq \Sigma$ ,  $\Sigma' \models \alpha$ .<sup>31</sup> If we do this then more than transitivity fails; so does Adjunction. For  $\neg p, p \vee q \models q$  and  $\neg q, p \vee q \models p$ , hence both  $p$  and  $q$  follow from  $\{\neg p, p \vee q, \neg q\} = \Sigma$ . But for no subset of  $\Sigma$ ,  $\Sigma'$ , do we have  $\Sigma' \models p \wedge q$ . ( $\Sigma \models \phi$ , and if  $\Sigma'$  is a proper subset of  $\Sigma$ ,  $\Sigma' \not\models_C p \wedge q$ .) In this respect, Tennant's approach is similar to the next one that we will look at.

#### 4.2 Non-adjunction

All the other approaches that we will consider, except the last (algebraic logics) accept validity as defined simply in terms of model-preservation. Thus, given some notion of interpretation, call it a *model* of a sentence if the sentence holds in the interpretation; an interpretation is a model of a set of sentences if it is a model of every member of the set; and an inference is valid iff every model of the premises is a model of the conclusion. In particular, then, if explosion is to be avoided, it must be possible to have models for contradictions, which are not models of everything. Where the differences in the following approaches lie is in what counts as an interpretation, and what counts as holding in it.

For the next approach, an interpretation,  $I$ , is a Kripke interpretation of some modal logic, say  $S5$ , employing the usual truth conditions. Each world in an interpretation may be thought of as the world according to some party in a debate or discussion. This gives the approach its common name, *discussive* (or *discursive*) logic.  $I$  is a (discursive) model of sentence  $\alpha$  iff  $\alpha$  holds at some world in  $I$ , i.e.,  $\diamond\alpha$  holds in the model. Thus,  $\Sigma \models \alpha$  iff  $\alpha$  holds, discursively, in every discursive model of  $\Sigma$ , i.e., iff  $\diamond\Sigma \models_{S5} \diamond\alpha$ , where  $\diamond\Sigma$  is  $\{\diamond\alpha; \alpha \in \Sigma\}$ . This approach is that of Jaškowski [1969].<sup>32</sup> It is clear that discussive logic is paraconsistent, since we may have  $\diamond\alpha$  and  $\diamond\neg\alpha$

<sup>30</sup>The proof theory can be given a filtered natural deduction form too. Essentially, classical deductions that have a certain "normal form" pass through the filter. See Tennant [1980].

<sup>31</sup>Though if we do this, symmetry suggests that we should take  $\Pi$  to follow from  $\Sigma$  iff for some  $\Sigma' \subseteq \Sigma$  and  $\Pi' \subseteq \Pi$ ,  $\Sigma' \vdash \Pi'$ ; in this case paraconsistency is lost since  $\alpha, \neg\alpha \vdash \phi$ .

<sup>32</sup>Popper also seems to have had a similar idea in 1948. See his [1963], p. 321.

in an  $S5$  interpretation, without having  $\diamond\beta$ . For similar reasons, Adjunction ( $\alpha, \beta \models \alpha \wedge \beta$ ) fails. It should be noted, however, that  $\alpha \wedge \neg\alpha \models \beta$ , so the logic is not paraconsistent for conjoined contradictions.

A closely related approach can be found in Rescher and Brandom [1979]. They define validity as truth preservation in all worlds, but they augment the worlds of standard modal logic by inconsistent and complete worlds, constructed using operators  $\dot{\cup}$  and  $\dot{\cap}$ . Specifically, worlds are constructed recursively from standard worlds as follows. If  $W$  is a set of worlds,  $\dot{\cup}W$  is a world such that  $\alpha$  is true in  $\dot{\cup}W$  iff for some  $w$  in  $W$ ,  $\alpha$  is true in  $w$ ; and  $\dot{\cap}W$  is a world such that  $\alpha$  is true in  $\dot{\cap}W$  iff for all  $w$  in  $W$ ,  $\alpha$  is true in  $w$ . As is intuitively clear, inconsistent worlds just provide another way of expressing what holds in a Jaśkowski interpretation. Incomplete worlds appear more novel, but, in fact, add nothing. For if truth fails to be preserved in one of these, it fails to be preserved in one of the ordinary worlds which go into making it up. These ideas can be recast to show that the semantics of Rescher and Brandom, and of Jaśkowski are inter-translatable, and deliver the same notion of validity.<sup>33</sup>

A notable feature of discussive logic is that  $\Sigma \models \alpha$  iff for some  $\beta \in \Sigma$ ,  $\beta \models_C \alpha$ . (The proof from right to left is obvious. From left to right, suppose that for every  $\beta \in \Sigma$ ,  $\beta \not\models_C \alpha$ . Let  $w_\beta$  be a classical world where  $\beta$  holds but  $\alpha$  does not. If we take the interpretation whose worlds are  $\{w_\beta; \beta \in \Sigma\}$  this is a counter-model for  $\Sigma \models \alpha$ .) Thus, single-premise discussive inference is classical, and there is no essentially multiple-premise inference. One way to avoid the second of these features is to add an appropriate conditional connective. We will look at this later. Another way is to allow a certain amount of conjoining of premises. The question is how to do this in a controlled way so that explosion does not arise.

One suggestion, made by Rescher and Manor, is, in effect, to allow conjoining up to maximal consistency.<sup>34</sup> Given a set of premises,  $\Sigma$ , a maximally consistent subset (mcs) is any consistent subset,  $\Sigma'$ , such that if

---

<sup>33</sup>Proof: Suppose that, discursively,  $\Sigma \not\models \alpha$ . Then there is an interpretation such that for each  $\sigma \in \Sigma$ , there is some world,  $w_\sigma$ , such that  $\sigma$  is true in  $w_\sigma$ , but  $\alpha$  is not true in  $w_\sigma$ . Let  $w = \dot{\cup}\{w_\sigma; \sigma \in \Sigma\}$ , then  $w$  is a Rescher/Brandom counter-model. Conversely, suppose that  $\Sigma \not\models \alpha$  for Rescher and Brandom. Then there is some world such that for every  $\sigma \in \Sigma$ ,  $\sigma$  is true at  $w$ , but  $\alpha$  is not. We show that there is a Jaśkowski counter-model. The result is proved by recursion on the construction of Rescher/Brandom worlds. If  $w$  is a standard world, the result is clear. So suppose that  $w = \dot{\cap}W$ , where the result holds for all members of  $W$ . By definition, for every  $z \in W$ , and every  $\sigma \in \Sigma$ ,  $\sigma$  is true in  $z$ , but for some  $z$ ,  $\alpha$  is not true in  $z$ . Consider that  $z$ . This is a Rescher/Brandom counter-model to the inference. Hence, the result holds by recursion hypothesis. Alternatively, suppose that  $w = \dot{\cup}W$ , where the result holds for all members of  $W$ . By definition, for every  $\sigma \in \Sigma$ , there is some  $w_\sigma \in W$ , such that  $\sigma$  is true in  $w_\sigma$ , but  $\alpha$  is not. By recursion, there must be a Jaśkowski countermodel for the inference  $\sigma/\alpha$ .  $\sigma$  is true at some world in this, but  $\alpha$  is not. If we form the collection of worlds for all such  $\sigma$ , this then gives us a Jaśkowski counter-model to the original inference.

<sup>34</sup>Rescher and Manor [1970-1]. This takes off from the earlier work of Rescher [1964].

$\alpha \in \Sigma - \Sigma'$ ,  $\Sigma' \cup \{\alpha\}$  is inconsistent. We can now say that  $\alpha$  follows from  $\Sigma$  iff for some mcs  $\Sigma'$ ,  $\Sigma' \models_C \alpha$ . In possible world terms, we can rephrase this as follows. Let us say that an interpretation,  $I$ , *respects*  $\Sigma$  iff for every mcs  $\Sigma'$ , there is a world,  $w$ , in  $I$  such that  $\Sigma'$  is true in  $w$ . Then it is not difficult to see that this policy is a variant of discussive logic:  $\Sigma \models \alpha$  iff  $\alpha$  holds discussively in every interpretation that respects  $\Sigma$ . (If  $\alpha$  follows classically from some  $\Sigma'$ , then it holds in every discussive interpretation that respects  $\Sigma$ . Conversely, suppose that it follows from no  $\Sigma'$ . Then for each  $\Sigma'$  choose a world  $w_{\Sigma'}$  where  $\Sigma'$  is true, but  $\alpha$  is not. The interpretation containing all such  $w_{\Sigma'}$  is a countermodel.)

This policy is certainly stronger than simple discussive consequence. For example, it gives:  $p, q \models p \wedge q$ . In fact, if  $\Sigma$  is (classically) consistent then every classical consequence of  $\alpha$  is a consequence. But it is still non-adjunctive:  $p, \neg p \not\models p \wedge \neg p$ .<sup>35</sup>

A slightly different way of proceeding is provided by Schotch and Jennings.<sup>36</sup> Given a finite set,  $\Sigma$ , a *partition* is any family of disjoint sets, each of which is classically consistent, and whose union is  $\Sigma$ . The *level* of  $\Sigma$ ,  $l(\Sigma)$ , is the least  $n$  such that  $\Sigma$  can be partitioned into  $n$  sets (or, conventionally,  $\infty$  if there is no such  $n$ ).  $\Sigma \models \alpha$  iff  $l(\Sigma) = \infty$  or,  $l(\Sigma) = n$  and for any partition of  $\Sigma$  of size  $n$ ,  $\{\Sigma_i; 1 \leq i \leq n\}$ , there is an  $i$  such that  $\Sigma_i \models_C \alpha$ . As with the previous approach, this definition can be converted into discussive terms, by taking our models to be those that respect the premise set. But this time, an interpretation respects  $\Sigma$  iff for some partition of the level of  $\Sigma$ ,  $\{\Sigma_i; 1 \leq i \leq n\}$ , and every  $i$ , there is a world in the interpretation where  $\Sigma_i$  is true.

Leaving aside the fact that Schotch and Jennings consider only finite premise sets, one difference between their approach and the previous one concerns the consequences of sets,  $\Sigma$ , with single inconsistent members. Such sets have no partitions, and so explode for Schotch and Jennings. They still have mcscs though (e.g.,  $\phi$ ), and so do not explode for Rescher and Manor. If  $\Sigma$  has no single inconsistent member then Schotch and Jennings' consequence relation is included in that of Rescher and Manor. For if  $\{\Sigma_i; i \in n\}$  is a partition of the premises,  $\Sigma$ , and for some  $i$ ,  $\Sigma_i \models_C \alpha$ , then  $\Sigma_i$  can be extended to an mcs of  $\Sigma$ , and this classically entails  $\alpha$ . The converse is not true, however. Let  $\Sigma = \{p, \neg p, q, r\}$ . This has two mcscs,  $\{p, q, r\}$  and  $\{\neg p, q, r\}$ . Hence, for Rescher and Manor,  $q \wedge r$  follows. But  $\Sigma$  has level 2, and one partition is  $\{\{p, q\}, \{\neg p, r\}\}$ . Neither of these classically entails  $q \wedge r$ , so this does not follow for Schotch and Jennings (which seems wrong, intuitively).<sup>37</sup>

<sup>35</sup> Rescher and Manor also formulate a weaker policy of inference.  $\alpha$  follows from  $\Sigma$  iff for *all* mcs  $\Sigma'$ ,  $\Sigma' \models_C \alpha$ . This logic is clearly adjunctive.

<sup>36</sup> See their [1980], where they also discuss appropriate proof theories and modal connections.

<sup>37</sup> The same example shows that Schotch and Jennings'  $\models$ , unlike Rescher and Manor's,

Despite the differences, Schotch and Jennings' approach shares with that of Rescher and Manor the following features: for consistent sets, consequence coincides with classical consequence; Adjunction fails. For Schotch and Jennings, like Jaśkowski,  $\alpha \wedge \neg\alpha$  explodes. For Rescher and Manor, it has no consequences (other than tautologies). The logics that we will look at in subsequent sections are more discriminating concerning conjoined contradictions.

### 4.3 Interlude: Henkin Constructions

Before we move on to look at the other basic approaches to paraconsistent logic, I want to isolate a construction that we will have many occasions to use. In a standard Henkin proof for the completeness of an explosive logic, we construct a maximally consistent set of sentences, and use this to define an evaluation. In the construction of the set, we keep something out of it by putting its negation in. As might be expected, these techniques do not work in paraconsistent logic; but they can be generalised to do so. What plays the role of a maximally consistent set in a paraconsistent logic is a *prime theory*, where a set of sentences,  $\Sigma$ , is a theory iff it is closed under deducibility; and it is prime iff  $\alpha \vee \beta \in \Sigma \Rightarrow (\alpha \in \Sigma \text{ or } \beta \in \Sigma)$ . To keep something out in the construction of a prime theory, we have to exclude it explicitly. I now show how.

Assume that the proof theory is to be given in natural deduction terms. For definiteness I adopt the notational conventions of Prawitz [1965].<sup>38</sup> Consider the following rules for conjunction and disjunction:

$$\begin{array}{l} \vee I \quad \frac{\alpha \quad \beta}{\alpha \wedge \beta} \\ \\ \wedge E \quad \frac{\alpha \wedge \beta}{\alpha} \quad \frac{\alpha \wedge \beta}{\beta} \\ \\ \vee I \quad \frac{\alpha}{\alpha \vee \beta} \quad \frac{\beta}{\alpha \vee \beta} \end{array}$$

---

is non-monotonic, since we have  $q, r \models q \wedge r$ .

<sup>38</sup>In particular, something of the form:

$$\begin{array}{c} \alpha \\ \vdots \\ \beta \end{array}$$

in a rule indicates a subproof with  $\alpha$  as one assumption—though there may be others—and conclusion  $\beta$ . If  $\alpha$  is overlined, this means that the application of the rule discharges it.

$$\vee E \quad \frac{\alpha \vee \beta \quad \begin{array}{c} \bar{\alpha} \\ \vdots \\ \gamma \end{array} \quad \begin{array}{c} \bar{\beta} \\ \vdots \\ \gamma \end{array}}{\gamma}$$

Let  $\vdash$  be any proof theory that includes these rules. Write  $\Sigma \vdash \Pi$  to mean that there are members of  $\Pi$ ,  $\pi_1, \dots, \pi_n$ , such that  $\Sigma \vdash \pi_1 \vee \dots \vee \pi_n$ . Then if  $\Sigma \not\vdash \Pi$ , there are sets  $\Delta \supseteq \Sigma$  and  $\Gamma \supseteq \Pi$ , such that  $\Delta \not\vdash \Gamma$ , and  $\Delta$  is a prime theory. To prove this, we enumerate the formulas of the language:  $\beta_0, \beta_1, \beta_2, \dots$ , and define a sequence of sets  $\Sigma_n, \Pi_n$  ( $n \in \omega$ ) as follows.  $\Sigma_0 = \Sigma$ ;  $\Pi_0 = \Pi$ . If  $\Sigma_n \cup \{\beta_n\} \not\vdash \Pi_n$ , then  $\Sigma_{n+1} = \Sigma_n \cup \{\beta_n\}$  and  $\Pi_{n+1} = \Pi_n$ . Otherwise  $\Sigma_{n+1} = \Sigma_n$  and  $\Pi_{n+1} = \Pi_n \cup \{\beta_n\}$ .  $\Delta = \bigcup_{n \in \omega} \Sigma_n$ ;  $\Gamma = \bigcup_{n \in \omega} \Pi_n$ .

It is not difficult to check by induction that for all  $n$ ,  $\Sigma_n \not\vdash \Pi_n$ . (Suppose this holds for  $n$ ; if  $\Sigma_n \cup \{\beta_n\} \not\vdash \Pi_n$ , the result for  $n + 1$  is immediate. So suppose that  $\Sigma_n \cup \{\beta_n\} \vdash \Pi_n$  and  $\Sigma_{n+1} \vdash \Pi_{n+1}$ . Then  $\Sigma_n \vdash \{\beta_n\} \cup \Pi_n$ . By a sequence of moves that amount to “cut”,  $\Sigma_n \vdash \Pi_n$ , contrary to induction hypothesis.) By compactness, it follows that  $\Delta \not\vdash \Gamma$ .

It is also easy to check that  $\Delta$  is a prime theory. Suppose that  $\Delta \vdash \alpha$ , but  $\alpha \notin \Delta$ . Then for some  $n$ ,  $\Sigma_n \cup \{\alpha\} \vdash \Pi_n$ . Hence,  $\Delta \vdash \Gamma$ . Next, suppose that  $\alpha \vee \beta \in \Delta$ , but  $\alpha \notin \Delta$  and  $\beta \notin \Delta$ . Then for some  $m$  and  $n$ ,  $\Sigma_n \cup \{\alpha\} \vdash \Pi_n$  and  $\Sigma_m \cup \{\beta\} \vdash \Pi_m$ . Hence  $\Delta \cup \{\alpha \vee \beta\} \vdash \Gamma$ , and so  $\Delta \vdash \Gamma$ .

#### 4.4 Non-truth-functionality

Let us now return to the other basic approaches to paraconsistent logics. On the first of these, explosion is invalidated by employing a non-truth-functional account of negation. Typically, this account of negation is imposed on top of an orthodox account of positive logic. Thus, let an interpretation be a map,  $\nu$ , from the set of formulas to  $\{1, 0\}$ , satisfying just the following conditions:

$$\begin{aligned} \nu(\alpha \wedge \beta) &= 1 \text{ iff } \nu(\alpha) = 1 \text{ and } \nu(\beta) = 1 \\ \nu(\alpha \vee \beta) &= 1 \text{ iff } \nu(\alpha) = 1 \text{ or } \nu(\beta) = 1 \end{aligned}$$

In particular, the truth value of  $\neg\alpha$  is independent of that of  $\alpha$ . Validity is defined as truth preservation over all interpretations. It is obvious that explosion fails, since we may choose an evaluation that assigns both  $p$  and  $\neg p$  (and their conjunction) the value 1, whilst assigning  $q$  the value 0.

These semantics can be characterised very simply in natural deduction terms by just the rules  $\vee I$ ,  $\vee E$ ,  $\wedge I$  and  $\wedge E$ . Soundness is easy to check. For completeness, suppose that  $\Sigma \not\vdash \alpha$ . Then put  $\Pi = \{\alpha\}$ , and extend  $\Sigma$

to a prime theory,  $\Delta$ , with the same property, as in 4.3. Define a map,  $\nu$ , as follows:

$$\begin{aligned}\nu(\alpha) &= 1 \text{ if } \alpha \in \Delta \\ \nu(\alpha) &= 0 \text{ if } \alpha \notin \Delta\end{aligned}$$

It is easy to check that  $\nu$  is an interpretation. Hence, we have the result.

This system contains no inferences that involve negation essentially. For this reason,  $\neg$  can hardly be thought of as a negation functor. Stronger paraconsistent systems, where this is more plausibly the case, can be obtained by adding conditions on the semantics. The following are some examples:<sup>39</sup>

- (i) if  $\nu(\alpha) = 0$ ,  $\nu(\neg\alpha) = 1$
- (ii) if  $\nu(\neg\neg\alpha) = 1$ ,  $\nu(\alpha) = 1$
- (iii) if  $\nu(\alpha) = 1$ ,  $\nu(\neg\neg\alpha) = 1$
- (iv)  $\nu(\neg(\alpha \wedge \beta)) = \nu(\neg\alpha \vee \neg\beta)$
- (v)  $\nu(\neg(\alpha \vee \beta)) = \nu(\neg\alpha \wedge \neg\beta)$

Sound and complete rule systems can be obtained by adding the corresponding rules, which are, respectively:

- (i)  $\frac{}{\overline{\alpha \vee \neg\alpha}}$
- (ii)  $\frac{\neg\neg\alpha}{\alpha}$
- (iii)  $\frac{\alpha}{\neg\neg\alpha}$
- (iv)  $\frac{\neg(\alpha \wedge \beta)}{\neg\alpha \vee \neg\beta}$
- (v)  $\frac{\neg(\alpha \vee \beta)}{\neg\alpha \wedge \neg\beta}$

(Double underlining indicates a two-way rule of inference, and a zero premise rule, as in (i), can be thought of as an assumption that discharges itself.) The corresponding soundness and completeness proofs are simple extensions of the basic arguments.

These additions give the  $\wedge, \vee, \neg$ -fragments of various systems in the literature. (i) gives that of Batens' *PI* [1980]; (i) and (ii) that of da Costa's  $C_\omega$ ; <sup>40</sup> (i)-(v) that of Batens'  $PI^S$ . In  $PI^S$  every sentence is logically equivalent to one in Conjunctive Normal Form. This can be used to show that  $PI^S$

<sup>39</sup>Some others can be found in Loparić and da Costa [1984], and Beziau [1990].

<sup>40</sup>Semantics of the present kind for the da Costa systems were first proposed in da Costa and Alves [1977].



is a maximal paraconsistent logic, in the sense that any logic that extends it is not paraconsistent. (For details, see Batens [1980].)

Observe, for future reference, that if we add to  $PI$  or an extension thereof the condition:

$$\text{if } \nu(\alpha) = 1, \nu(\neg\alpha) = 0$$

then all interpretations are classical, and so we have classical logic. As may easily be checked, adding this is sound and complete with respect to the rule of inference:

$$\frac{\alpha \wedge \neg\alpha}{\beta}$$

Another major da Costa system,  $C_1$ , extends  $C_\omega$  in accordance with the following idea. It should be possible to express in the language the idea that a sentence,  $\alpha$ , behaves consistently; and for consistent sentences classical logic should apply. Let us write  $\neg(\alpha \wedge \neg\alpha)$  as  $\alpha^0$ . Then it is natural enough to suppose that  $\alpha^0$  expresses the consistency of  $\alpha$ . It does not, in any of the above systems, since we may have  $\alpha \wedge \neg\alpha \wedge \alpha^0$  true in an interpretation. This is exactly what is ruled out by the condition:

$$\text{(vi) } \nu(\alpha) = \nu(\neg\alpha) \text{ then } \nu(\alpha^0) = 0^{41}$$

( $\nu(\alpha) = \nu(\neg\alpha)$  iff both are 1, by semantic condition (i). Note that (i) also guarantees the converse of (vi):  $\nu(\alpha) \neq \nu(\neg\alpha)$  then  $\nu(\alpha^0) = 1$ .)

$C_1$  is obtained by adding (vi) to  $C_\omega$ , together with the following condition, which requires consistency to be preserved under syntactic constructions:

$$\text{(vii) if } \nu(\alpha^0) = \nu(\beta^0) = 1 \text{ then } \nu((\neg\alpha)^0) = \nu((\alpha \wedge \beta)^0) = \nu((\alpha \vee \beta)^0) = 1$$

The deduction rules that correspond to (vi) and (vii) are, respectively:

$$\text{(vi) } \frac{\alpha \wedge \neg\alpha \wedge \alpha^0}{\beta}$$

$$\text{(vii) } \frac{\alpha^0}{(\neg\alpha)^0} \quad \frac{\alpha^0 \quad \beta^0}{(\alpha \wedge \beta)^0} \quad \frac{\alpha^0 \quad \beta^0}{(\alpha \vee \beta)^0}$$

Soundness and completeness of the extensions are easily checked.

Now suppose that we have a piece of valid classical reasoning concerning formulas composed of parameters  $p_1, \dots, p_n$ . If we *assume*  $p_1^0, \dots, p_n^0$  then for

<sup>41</sup>Da Costa's actual condition is: if  $\nu(\alpha^0) = \nu(\beta \supset (\alpha \wedge \neg\alpha)) = 1$  then  $\nu(\beta) = 0$ . This is equivalent, given his account of the conditional.

every such formula,  $\alpha$ ,  $\alpha^0$  follows by the appropriate applications of the rules of (vii). Hence, whenever we have established  $\alpha \wedge \neg\alpha$  we may apply rule (vi) to give  $\beta$ . But the addition of this inference is sufficient to give classical logic, as I have already observed. Hence any valid classical reasoning may be recaptured formally by adding the appropriate consistency assumptions.<sup>42</sup>

One final comment on treating negation non-truth-functionally. It is a consequence of this that the substitutivity of provable equivalents breaks down in general. For example, even though  $\alpha$  is logically equivalent to  $\alpha \wedge \alpha$  there is no guarantee that the negations of these formulas have the same truth value in an interpretation.<sup>43</sup>

#### 4.5 Many Values

The previous approach sticks with the traditional two truth values, and obtains a paraconsistent logic by making negation non-truth-functional. The next approach retains truth functionality, but drops the idea that there are exactly two truth values. That is, such logics are many-valued.<sup>44</sup> A many-valued logic will be paraconsistent if it is possible for a formula and its negation both to take designated values (whilst not everything does). A natural way of obtaining this is to have a designated value that is a fixed point for negation. The simplest such logic is a three valued one with values,  $t$ ,  $b$ , and  $f$ , where  $t$  and  $b$  are designated, and the matrices are:

$\neg$	
$t$	$f$
$b$	$b$
$f$	$t$

$\wedge$	$t$	$b$	$f$
$t$	$t$	$b$	$f$
$b$	$b$	$b$	$f$
$f$	$f$	$f$	$f$

$\vee$	$t$	$b$	$f$
$t$	$t$	$t$	$t$
$b$	$t$	$b$	$b$
$f$	$t$	$b$	$f$

It will be noted that these are just the matrices of Lukasiewicz and Kleene's 3-valued logics, where the middle value is normally thought of as *undecidable*, or *neither true nor false*, and so not designated. It was the thought that this value might be read as *both true and false*—a natural enough thought, given dialetheism—and so be designated, that marks the start of many-valued paraconsistent logic. This was the approach proposed by Asenjo (see his [1966]), and others, e.g. Priest [1979], where the logic is called *LP*, a nomenclature that I will stick with here.

<sup>42</sup>It might be suggested that one ought not to take  $\alpha^0$  as expressing consistency unless it, itself, behaves consistently. This thought motivates the weaker da Costa system  $C_2$ , which is the same as  $C_1$ , except that  $\alpha^0$  is replaced everywhere by  $\alpha^0 \wedge \alpha^{00}$ . Of course, there is no reason to suppose that this expresses the consistency of  $\alpha$  unless it, itself, behaves consistently. This thought motivates the da Costa system  $C_3$  where  $\alpha^0$  is replaced everywhere by  $\alpha^0 \wedge \alpha^{00} \wedge \alpha^{000}$ . And so on for all the da Costa Systems  $C_i$ , for finite non-zero  $i$ .

<sup>43</sup>For a discussion of this in the context of da Costa's logics, see Urbas [1989].

<sup>44</sup>For a general discussion of many-valued logics, see the articles on the topic in this *Handbook*. See also, Rescher [1969].

$LP$  may be generalised in various different ways. One is as follows. If we let  $t = +1$ ,  $b = 0$  and  $f = -1$  then the truth conditions of  $LP$  are:

$$\begin{aligned}\nu(\neg\alpha) &= -\nu(\alpha) \\ \nu(\alpha \wedge \beta) &= \min\{\nu(a), \nu(\beta)\} \\ \nu(\alpha \vee \beta) &= \max\{\nu(a), \nu(\beta)\}\end{aligned}$$

The same conditions can be used for any set of integers,  $X$ , containing 0 and closed under  $-$ . The designated values are the non-negative values. Let us call this a *Sugihara generalisation*, after the person who, in effect, first proposed a matrix of this kind, where  $X$  was the set of all integers.<sup>45</sup>

Any Sugihara generalisation, though semantically different from  $LP$ , is essentially equivalent to it. Any  $LP$  countermodel is a Sugihara countermodel. But conversely, if we have a Sugihara countermodel, we can obtain an  $LP$  countermodel by mapping all positive values to  $+1$ , 0 to 0, and all negative values to  $-1$ . A little thought is sufficient to establish that the mapping respects the matrices and preserves designated values, as required.

A different way of generalising  $LP$  is as follows. If we let  $t = 1$ ,  $b = 0.5$  and  $f = 0$  then the truth conditions of  $LP$  are:

$$\begin{aligned}\nu(\neg\alpha) &= 1 - \nu(\alpha) \\ \nu(\alpha \wedge \beta) &= \min\{\nu(a), \nu(\beta)\} \\ \nu(\alpha \vee \beta) &= \max\{\nu(a), \nu(\beta)\}\end{aligned}$$

The same conditions can be used for any set of reals  $\{0, 0.5, 1\} \subseteq X \subseteq [0, 1]$ , which is closed under subtraction (of a greater by a lesser). For suitable choices of  $X$ , these are the matrices of the odd-numbered finite Łukasiewicz many-valued logics, and for  $X = [0, 1]$  they are the matrices of Łukasiewicz' continuum-valued logic. In Łukasiewicz' logics proper, the only designated value is 1, which does not give a paraconsistent logic. But if one takes the designated values to be  $\{x; a < x \leq 1\}$  (or  $\{x; a \leq x \leq 1\}$ ) then the logic will be paraconsistent provided that  $0 < a < 0.5$  (or  $0 < a \leq 0.5$ ). Let us call such logics *Łukasiewicz generalisations*. In a Łukasiewicz generalisation where the set of truth values is  $[0, 1]$ , these may naturally be thought of as degrees of truth. Hence, such a logic is a natural candidate for a paraconsistent fuzzy logic (logic of vagueness).<sup>46</sup>

It is not difficult to see that any Łukasiewicz generalisation is, in fact, equivalent to  $LP$ . As with the Sugihara generalisations, any  $LP$  countermodel is a Łukasiewicz countermodel; and conversely, any Łukasiewicz

<sup>45</sup>See Anderson and Belnap [1975], sect. 26.9.

<sup>46</sup>A variation on this theme is given by Peña in a number of papers. (See, e.g., Peña [1984].) Peña takes truth values to be an ordered set of more complex entities defined in terms of the interval  $[0, 1]$ .

countermodel can be collapsed into an  $LP$  countermodel by the mapping,  $f$ , defined thus:

$$\begin{aligned} f(x) &= 1 && \text{if } 1 - a \leq x \leq 1 \\ &= 0.5 && \text{if } a < x < 1 - a \\ &= 0 && \text{if } 0 \leq x \leq a \end{aligned}$$

or for the case where  $a$  is a designated value:

$$\begin{aligned} f(x) &= 1 && \text{if } 1 - a < x \leq 1 \\ &= 0.5 && \text{if } a \leq x \leq 1 - a \\ &= 0 && \text{if } 0 \leq x < a \end{aligned}$$

The generalisations of  $LP$  that we have considered in this section all, therefore, generate the same logic. What its proof theory is, we will see in the next.

#### 4.6 Relational Valuations

Standardly, semantic evaluations are thought of as functions from formulas to truth values, say, 0 and 1. Another way of invalidating explosion is to take them to be, not functions, but relations. A formula may then relate to both 0 and 1, another way of expressing the thought that a sentence is both true and false. Assuming that negation behaves as usual, this means that both  $p$  and  $\neg p$  may relate to 1, whilst an arbitrary formula may not. A natural way of spelling out this idea is as follows.

If  $P$  is the set of propositional parameters, an evaluation,  $\rho$ , is a subset of  $P \times \{0, 1\}$ . The evaluation is extended to a relation for all formulas by the familiar looking recursive clauses:

$$\begin{aligned} \neg\alpha\rho 1 &\text{ iff } \alpha\rho 0 \\ \neg\alpha\rho 0 &\text{ iff } \alpha\rho 1 \end{aligned}$$

$$\begin{aligned} \alpha \wedge \beta\rho 1 &\text{ iff } \alpha\rho 1 \text{ and } \beta\rho 1 \\ \alpha \wedge \beta\rho 0 &\text{ iff } \alpha\rho 0 \text{ or } \beta\rho 0 \end{aligned}$$

$$\begin{aligned} \alpha \vee \beta\rho 1 &\text{ iff } \alpha\rho 1 \text{ or } \beta\rho 1 \\ \alpha \vee \beta\rho 0 &\text{ iff } \alpha\rho 0 \text{ and } \beta\rho 0 \end{aligned}$$

Let us say that a formula,  $\alpha$ , is true in an interpretation,  $\rho$ , iff  $\alpha\rho 1$ , and false iff  $\alpha\rho 0$ ; then validity may be defined as truth preservation in all interpretations. According to this account, classical logic is just the special case where multi-valued relations have been forgotten.

These semantics are the Dunn semantics for the logic of First Degree Entailment, *FDE*.<sup>47</sup> In natural deduction terms, *FDE* can be characterised by the rules  $\wedge I$ ,  $\wedge E$ ,  $\vee I$  and  $\vee E$ , together with the rules:

$$\frac{\neg(\alpha \wedge \beta)}{\neg\alpha \vee \neg\beta} \quad \frac{\neg\alpha \wedge \neg\beta}{\neg(\alpha \vee \beta)} \quad \frac{\alpha}{\neg\neg\alpha}$$

Soundness is easily checked. For completeness, suppose that  $\Sigma \not\vdash \alpha$ . Extend  $\Sigma$  to a prime theory,  $\Delta$ , with the same property, as in 4.3. Now define an interpretation,  $\rho$ , thus:

$$\begin{aligned} p\rho 1 &\text{ iff } p \in \Delta \\ p\rho 0 &\text{ iff } \neg p \in \Delta \end{aligned}$$

A straightforward (joint) induction shows that this characterisation extends to all formulas. Completeness follows.

There are two natural restrictions that one may place upon Dunn evaluations:

- #1 for every  $p$ , there is at most one  $x$  such that  $p\rho x$
- #2 for every  $p$  there is at least one  $x$  such that  $p\rho x$

Both conditions extend from propositional parameters to all formulas, by a simple induction. Thus, the first condition ensures that the relation is *functional*; the second that it is *total*. A relation that satisfies both conditions is just a classical evaluation.

These extra conditions are sound and complete with respect to the extra rules:

$$\frac{\alpha \wedge \neg\alpha}{\beta} \quad \frac{}{\alpha \vee \neg\alpha}$$

respectively, as simple extensions of the completeness proofs demonstrate.

We can express the relational semantics in functional terms by taking an evaluation to be a function from formulas to *subsets* of  $\{1, 0\}$ , since there is an obvious isomorphism between relations,  $\rho$ , and functions,  $\nu$ , given by the condition:

$$\alpha\rho x \text{ iff } x \in \nu(\alpha)$$

---

<sup>47</sup>Published in Dunn [1976], though he discovered them somewhat earlier than this. In the present context, it might be better to call the system ‘Zero Degree Entailment’ since the language does not contain a conditional connective.

In this way, *FDE* can be seen as a many- (in fact, four-) valued logic.<sup>48</sup>

Restriction #2, which ensures that no formula takes the value  $\phi$ , gives a three-valued logic that is identical with *LP*. It is easy enough to check that the values  $\{1\}$ ,  $\{1, 0\}$ , and  $\{0\}$  work the same way as *t*, *b*, and *f*, respectively. I will make this identification in the rest of this essay. Restriction #1, which ensures that no formula takes the value  $\{1, 0\}$ , obviously gives an explosive logic, which is, in fact, the strong Kleene three-valued logic. This is therefore a logic dual to *LP*.<sup>49</sup>

A feature of these semantics for *LP* and *FDE* is that they are monotonic in the following sense. Let  $\nu_1$  and  $\nu_2$  be functional evaluations. If for all propositional parameters,  $p$ ,  $\nu_1(p) \subseteq \nu_2(p)$  then for all  $\alpha$ ,  $\nu_1(\alpha) \subseteq \nu_2(\alpha)$ . The proof of this is by a simple induction. One consequence of this for *LP* is worth remarking on. *LP* is clearly a sub-logic of classical logic, since it has the classical matrices as sub-matrices. The consequence relation of *LP* is weaker than that of classical logic, since it is paraconsistent. But the set of logical truths of *LP* is identical with that of classical logic. For suppose that  $\alpha$  is not valid in *LP*. Let  $\nu$  be an evaluation such that  $1 \notin \nu(\alpha)$ . Let  $\nu'$  be the interpretation that is the same as  $\nu$ , except that for every parameter,  $p$ , if  $\nu(p) = \{0, 1\}$ ,  $\nu'(p) = \{0\}$ . This is a classical evaluation; and by monotonicity,  $1 \notin \nu'(\alpha)$ , as required.

Another feature of these semantics is the evaluation that assigns every propositional parameter the value  $\{1, 0\}$ ,  $v_{\{1,0\}}$ ; and, in the four-valued case, the evaluation that gives every parameter the value  $\phi$ ,  $v_\phi$ . A simple induction shows that these properties extend to all formulas. Thus,  $v_{\{1,0\}}$  makes all formulas true—and false—and  $v_\phi$  makes every formula neither. In particular, then, *FDE* has no logical truths.<sup>50</sup>

#### 4.7 Possible Worlds

Yet another, closely connected, way of invalidating explosion is to treat negation as an intensional operator. This way was proposed by the Routleys in [1972]. A *Routley interpretation* is a structure,  $\langle W, *, \nu \rangle$ , where  $W$  is a set (of worlds),  $*$  is a map from  $W$  to  $W$ , and  $\nu$  maps sets of pairs comprising a world and propositional parameter to  $\{1, 0\}$ . (I will write  $\nu(w, \alpha)$  as  $\nu_w(\alpha)$ .) The truth conditions for conjunction and disjunction are the standard:

$$\nu_w(\alpha \wedge \beta) = 1 \text{ iff } \nu_w(\alpha) = 1 \text{ and } \nu_w(\beta) = 1$$

<sup>48</sup>In fact, the straight truth tables with values 1, 2, 3 and 4 were enunciated by Smiley. See Anderson and Belnap [1975], p. 161.

<sup>49</sup>I will usually use the functional semantic representation for *FDE* and *LP* in the rest of this essay. A word of warning, though: in the context of a dialethic metatheory, the functional approach may have consequences that the relational approach, proper, does not have. See Priest and Smiley [1993], p. 49ff.

<sup>50</sup>Further interesting properties of *LP* and *FDE* are established in Pynko [1995a] and [1995b].

$$\nu_w(\alpha \vee \beta) = 1 \text{ iff } \nu_w(\alpha) = 1 \text{ or } \nu_w(\beta) = 1$$

The truth conditions for negation are:

$$\nu_w(\neg\alpha) = 1 \text{ iff } \nu_{w^*}(\alpha) = 0$$

Note that if  $w^* = w$ , these conditions just reduce to the classical ones. A natural understanding of the  $*$  operator is a moot point.<sup>51</sup> I will return to the issue in a moment. Validity is defined in terms of truth preservation at all worlds of all interpretations.

In natural deduction terms, this system can be characterised by modifying that for *FDE* by dropping the rule for double negation, and replacing it with:

$$\frac{\begin{array}{c} \bar{\alpha} \\ \vdots \\ \beta \quad \neg\beta \end{array}}{\neg\alpha}$$

where, in the subproof, there are no undischarged assumptions other than  $\alpha$ . Soundness is easily checked. For completeness, suppose that  $\Sigma \not\vdash \alpha$ . Extend  $\Sigma$  to a prime theory,  $\Delta$ , with the same property, as in 4.3. Now define an interpretation,  $\langle W, *, \nu \rangle$ , where  $W$  is the set of all prime theories,  $*$  is defined by the condition:

$$\alpha \in \Delta^* \text{ iff } \neg\alpha \notin \Delta$$

and  $\nu$  is defined by:

$$\nu_\Delta(p) = 1 \text{ iff } p \in \Delta \quad (\#)$$

It is not difficult to check that if  $\Delta$  is a prime theory, so is  $\Delta^*$  and hence that  $*$  is well defined. First, suppose that  $\alpha \notin \Delta^*$  and  $\beta \notin \Delta^*$ . Then  $\neg\alpha$  and  $\neg\beta$  are in  $\Delta$ . Since  $\Delta$  is a theory,  $\neg\alpha \wedge \neg\beta \in \Delta$ , and so  $\neg(\alpha \vee \beta) \in \Delta$ . Hence,  $\alpha \vee \beta \notin \Delta^*$ . Next, suppose that  $\Delta^* \vdash \alpha$ , but  $\alpha \notin \Delta^*$ . Then for some  $\beta_1, \dots, \beta_n \in \Delta^*$ ,  $\beta_1 \wedge \dots \wedge \beta_n \vdash \alpha$ . Hence, by contraposition and De Morgan  $\neg\alpha \vdash \neg\beta_1 \vee \dots \vee \neg\beta_n$ . But  $\neg\alpha \in \Delta$ ; hence  $\neg\beta_1 \vee \dots \vee \neg\beta_n \in \Delta$ . Since  $\Delta$  is prime, for some  $1 \leq i \leq n$ ,  $\neg\beta_i \in \Delta$ , i.e.,  $\beta_i \notin \Delta^*$ . Contradiction.

An easy recursion shows that  $(\#)$  extends to all formulas. The result follows.

---

<sup>51</sup>For some discussion and references, see the article on Relevance Logic and Entailment in this *Handbook*.

The logic can be made stronger without (necessarily) ruining its paraconsistency by adding further conditions on  $*$ . The most notable is:  $w^{**} = w$ . This is sound and complete with respect to the additional rule:

$$\frac{\alpha}{\neg\neg\alpha}$$

as a simple extension of the completeness argument demonstrates.

These semantics are, in fact, very closely related to the those for *FDE* of the previous section. Given an *FDE* interpretation,  $\nu$ , define a Routley evaluation on the worlds  $w$  and  $w^*$ , as follows:

$$\begin{aligned} \nu_w(p) &= 1 \text{ iff } 1 \in \nu(p) \\ \nu_{w^*}(p) &= 1 \text{ iff } 0 \notin \nu(p) \end{aligned}$$

A simple induction shows that these conditions follow for all formulas. Conversely, we can turn the conditions into reverse. Given any Routley evaluation on a pair of worlds,  $w, w^*$ , define a Dunn evaluation by the conditions:

$$\begin{aligned} 1 \in \nu(p) &\text{ iff } \nu_w(p) = 1 \\ 0 \in \nu(p) &\text{ iff } \nu_{w^*}(p) \neq 1 \end{aligned}$$

Essentially the same induction shows that these conditions hold for all formulas. Hence, the two semantics are inter-translatable, and validate the same proof theories.<sup>52</sup> The translation also suggests a natural interpretation of the  $*$  operator.  $w^*$  is that world characterised by the set of untruths of  $w$ . (This is, of course, in general, distinct from the set of truths in a four-valued context.)

Under the above translation, the condition:  $1 \in \nu(p)$  or  $0 \in \nu(p)$ , which gives an *LP* interpretation, is equivalent to:  $\nu_w(p) = 1$  or  $\nu_{w^*}(p) \neq 1$ ; imposing which condition on an intensional interpretation therefore gives an intensional semantics for *LP*.

#### 4.8 Algebraic Semantics

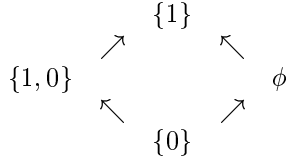
Let us now turn to the final approach to paraconsistent logics that we will consider, an algebraic one. In algebraic logic, an interpretation is a homomorphism,  $\nu$ , from sentences into some algebraic structure,  $\mathcal{A} = \langle A, \wedge, \vee, \neg \rangle$ ; i.e.,  $\nu(\neg\alpha) = \neg\nu(\alpha)$ ,  $\nu(\alpha \wedge \beta) = \nu(\alpha) \wedge \nu(\beta)$ , etc. (I will use the same signs for the connectives and the algebraic operations. Context, and the style of variable, will serve to disambiguate.) If the algebra is a lattice—as it usually

<sup>52</sup>Which shows that the contraposition rule is admissible in *FDE*, something that is not at all obvious.



is, and will be in all the cases we consider—the consequence relation of the logic is represented by the lattice order relation, defined in the usual way:  $a \leq b$  iff  $a \wedge b = a$ . Thus, a logic will be paraconsistent if it is possible in the algebra to have an  $a$  and  $b$  such that  $a \wedge \neg a \not\leq b$ .

Several of the logics that we have looked at can be algebraicised. Consider, for example, *FDE*. If we take the four-valued semantics for this, we can think of the values as a lattice whose Hasse diagram is as follows:



( $\wedge$  is lattice-meet;  $\vee$  is lattice join;  $\neg\{1,0\} = \{1,0\}$ , and  $\neg\phi = \phi$ .) This generalises to a *De Morgan algebra*. A De Morgan algebra is a structure  $\mathcal{A} = \langle A, \wedge, \vee \rangle$ , where  $\langle A, \wedge, \vee, \neg \rangle$  is a distributive lattice, and  $\neg$  is an involution of period 2, i.e.:

$$\begin{aligned} \neg\neg a &= a \\ a \leq b &\Rightarrow \neg b \leq \neg a \end{aligned}$$

The structures take their name from the fact that in every such algebra  $\neg(a \wedge b) = \neg a \vee \neg b$  holds, as do the other De Morgan laws.

Define an inference  $\alpha_1, \dots, \alpha_n / \beta$  to be algebraically valid iff for every homomorphism,  $\nu$ , into a De Morgan algebra,  $\mathcal{A}$ ,  $\nu(\alpha_1) \wedge \dots \wedge \nu(\alpha_n) \leq \nu(\beta)$ . Then the algebraically valid inferences are exactly those of *FDE*. It is easy to check that the rule system for *FDE* is sound with respect to these semantics. Completeness follows from completeness in the four-valued case. Alternatively, we can give a direct argument as follows.

Consider the relation  $\alpha \sim \beta$ , defined by:  $\alpha \vdash \beta$  and  $\beta \vdash \alpha$ . One can check that this is an equivalence relation, and a congruence on the logical operators (i.e., if  $\alpha_1 \sim \beta_1$  and  $\alpha_2 \sim \beta_2$  then  $\alpha_1 \wedge \alpha_2 \sim \beta_1 \wedge \beta_2$ , etc.).<sup>53</sup> If  $F$  is the set of formulas, define the quotient algebra,  $\mathcal{A} = \langle F / \sim, \wedge, \vee, \neg \rangle$ , where, if  $[\alpha]$  is the equivalence class of  $\alpha$ ,  $\neg[\alpha] = [\neg\alpha]$ ,  $[\alpha] \wedge [\beta] = [\alpha \wedge \beta]$ , etc. One can check that  $\mathcal{A}$  is a De Morgan lattice. Now, let  $\nu$  be the homomorphism that maps every  $\alpha$  to  $[\alpha]$ . If  $\nu(\alpha_1) \wedge \dots \wedge \nu(\alpha_n) \leq \nu(\beta)$ . Then  $[\alpha_1 \wedge \dots \wedge \alpha_n] \leq [\beta]$ , i.e.,  $[\alpha_1 \wedge \dots \wedge \alpha_n \wedge \beta] = [\alpha_1 \wedge \dots \wedge \alpha_n]$ . Hence,  $\alpha_1 \wedge \dots \wedge \alpha_n \vdash \alpha_1 \wedge \dots \wedge \alpha_n \wedge \beta$  and so  $\alpha_1 \wedge \dots \wedge \alpha_n \vdash \beta$ . Conversely, then, if  $\alpha_1 \wedge \dots \wedge \alpha_n \not\vdash \beta$  then  $\nu(\alpha_1) \wedge \dots \wedge \nu(\alpha_n) \not\leq \nu(\beta)$ , as required.

<sup>53</sup>The only tricky point concerns negation. For this, we need to appeal to the fact, which we have already noted, that if  $\alpha \vdash \beta$  then  $\neg\beta \vdash \neg\alpha$ . This can be established directly, by an induction on proofs.

It should be noted that not all the logics we have considered in previous sections algebraicise. In particular, the non-truth-functional logics resist this treatment in general. This is for the same reason that the substitutivity of provable equivalents breaks down: the semantic value of  $\neg\alpha$  is *entirely independent* of that of  $\alpha$ . It cannot, therefore, correspond to any well-defined algebraic operation.

The point can be made more precise in many cases. Suppose that  $\mathcal{A}$  is some algebraic structure for a logic, and consider any interpretation,  $\nu$ , with values in the algebra, such that for some  $p, q$  and  $r$ ,  $\nu(p) = \nu(q) \neq \nu(r)$ . Then the condition  $\nu(\alpha) = \nu(\beta)$  is a congruence relation on the set of formulas, and collapse by it gives a non-degenerate quotient algebra (i.e., an algebra that is neither a single-element algebra, nor the algebra of formulas). But many non-truth-functional logics can be shown to have no such thing. (See, e.g., Mortensen [1980].)

One final algebraic paraconsistent logic is worth noting. This is that of Goodman [1981]. A Heyting algebra can be thought of as a distributive lattice, with a bottom element,  $\perp$ , and an operator,  $\rightarrow$ , satisfying the condition:

$$a \wedge b \leq c \text{ iff } a \leq b \rightarrow c$$

(which makes  $\perp \rightarrow \perp$  the top element). We may define  $\neg a$  as  $a \rightarrow \perp$ .

Let  $\mathcal{T}$  be a topological space. Then a standard example of a Heyting algebra is the topological Heyting algebra  $\langle X, \wedge, \vee, \rightarrow, \perp \rangle$ , where  $X$  is the set of open sets in  $\mathcal{T}$ ,  $\wedge$  and  $\vee$  are intersection and union, respectively,  $\perp$  is  $\phi$ , and  $a \rightarrow b$  is  $(\overline{a} \vee b)^{\circ}$ —overlining denotes complementation and  $^{\circ}$  is the interior operator of the topology.  $\neg a$  is clearly  $\overline{a}^{\circ}$ .

It is well known that for finite sets of premises, Intuitionistic logic is sound and complete with respect to the class of Heyting algebras, in fact, with respect to the topological Heyting algebras. That is,  $\alpha_1, \dots, \alpha_n \models_I \beta$  iff for every homomorphism,  $\nu$ , into such an algebra,  $\nu(\alpha_1 \wedge \dots \wedge \alpha_n) \leq \nu(\beta)$ .<sup>54</sup>

The whole construction can be dualised in a natural way to give a paraconsistent logic. A dual Heyting algebra is a distributive lattice, with a top element,  $\top$ , and an operator,  $\leftarrow$ , satisfying the condition:

$$a \leq b \vee c \text{ iff } a \leftarrow b \leq c$$

(which makes  $\top \leftarrow \top$  the bottom element). We may define  $\neg a$  as  $\top \leftarrow a$ . As may be checked, if  $\mathcal{T}$  is a topological space, then the structure  $\langle X, \wedge, \vee, \leftarrow, \top \rangle$  is a dual Heyting algebra, where  $X$  is the set of *closed* sets of  $\mathcal{T}$ ,  $\wedge$  and  $\vee$  are intersection and union, respectively,  $\top$  is the whole space,

<sup>54</sup>See, e.g., Dummett [1977], 5.3.

and  $a \leftarrow b$  is  $(a \wedge \bar{b})^c$ —where  $c$  is the closure operator of the topology.  $\neg b$  is clearly  $\bar{b}^c$ .

The logic generated by dual Heyting algebras is dual to Intuitionistic logic. In particular, in Intuitionistic logic we have  $\alpha \wedge \neg\alpha \models \beta$ , but not  $\alpha \models \beta \vee \neg\beta$ ; and  $\alpha \models \neg\neg\alpha$ , but not  $\neg\neg\alpha \models \alpha$ . Thus in dual Intuitionist logic, we have  $\alpha \models \beta \vee \neg\beta$  but not  $\alpha \wedge \neg\alpha \models \beta$ ; and  $\neg\neg\alpha \models \alpha$  but not  $\alpha \models \neg\neg\alpha$ . For a topological counter-model to the first, consider the real line with its usual topology, and an interpretation,  $\nu$ , that maps  $p$  to  $[-1, +1]$ , and  $q$  to  $\phi$ . Then  $\nu(p \wedge \neg p) = \{-1, +1\} \not\subseteq \phi = \nu(q)$ . (This illustrates how the points in the set represented by  $p \wedge \neg p$  may be thought of as the points on the topological boundary between the set of points represented by  $p$  and the set of points represented by  $\neg p$ .) For a counter-model to the second, let  $\nu(p) = \{0\}$ . Then  $\nu(\neg\neg p) = \phi \not\subseteq \nu(p)$ .

If  $\models$  in dual Intuitionist logic, then  $\models_C \alpha$ , since the two-element Boolean algebra is a dual Heyting algebra. Conversely, if  $\alpha$  is any classical tautology, its dual,  $\alpha'$ , is a contradiction. Hence,  $\models_C \neg\alpha'$ . But then by a result of Glivenko,  $\models_I \neg\alpha'$ , and so  $\alpha' \models_I$ . Thus by duality, in dual Intuitionist logic  $\models \alpha$ . The logical truths of dual Intuitionist logic are therefore the same as those of classical logic.

It is worth noting that just as Intuitionist logic can be given an intensional semantics, namely Kripke semantics, so can dual Intuitionist logic; we simply dualise the Kripke construction. For further details of all the above, see Goodman [1981].<sup>55</sup>

## 5 CONDITIONAL CONNECTIVES

We have now looked at most of the basic techniques of paraconsistent logic, applied to languages containing only negation, conjunction and disjunction.<sup>56</sup> I will call this language the *basic* language. Next, we will look at some important extensions of these techniques (which do not ruin paraconsistency). In this part, we will start with the conditional, by which I mean some con-

<sup>55</sup>It is well known that in a certain well defined sense, Intuitionist logic can be seen as the “internal logic” of the category-theoretic structures called topoi. It is possible to dualise the construction involved there to show that dual Intuitionist logic has an equally good claim to that title. For details, see Mortensen [1995], who calls the  $\wedge, \vee, \neg$ -fragment of a dual Heyting algebra a ‘paraconsistent algebra’.

<sup>56</sup>There are others, such as the use of the techniques of combinatorial logic, but I will not go into these here. For details, one can consult, e.g., Bunder [1984]. There *ought* to be yet more. The discussion of connexivism in 3.2 suggests that there ought to be a distinctive connexivist approach to paraconsistency. To date, this has not emerged. The most articulated modern connexivist logic is due to McCall (see sect. 29.8 of Anderson and Belnap [1975], which can also be consulted for references to other discussions). Although this provides a connexivist treatment of the connective  $\rightarrow$ , the logic of the basic language is classical, and so explosive. Alternatively, one can formulate versions of relevant logic that contain connexivist principles. See Routley [1978] and Mortensen [1984].

nective,  $\rightarrow$  (if necessary, added to the basic language), satisfying, at least, *modus ponens*:  $\alpha, \alpha \rightarrow \beta \models \beta$ .

Although paraconsistency does not concern the conditional as such, many of the paraconsistent logics that we have looked at have distinctive approaches to the conditional. And this is no accident. If one identifies  $\alpha \rightarrow \beta$  with the material conditional,  $\alpha \supset \beta$ , defined in the usual way as  $\neg\alpha \vee \beta$ , then *modus ponens* reduces to the disjunctive syllogism. But in any logic where disjunction behaves normally and deducibility is transitive, the disjunctive syllogism must fail, or explosion would arise, due to the “Lewis independent argument”. Specifically, in all the logics we have looked at except filter logics and some of the non-adjunctive logics, the syllogism fails. In such logics, therefore, a distinct account of the conditional is required. For completeness’ sake, we will start by considering the others.

### 5.1 $\rightarrow$ as $\supset$

In filter logics, we may simply identify  $\rightarrow$  with  $\supset$ . Things then proceed as before. A one-premise inference in this language,  $\alpha/\beta$ , is prevalid iff it is classically valid and  $F(\alpha, \beta)$ . It is valid iff it is a substitution instance of a prevalid inference.<sup>57</sup>

In the natural extension of Tennant’s semantic approach, an inference from  $\Pi$  to  $\Sigma$  is prevalid iff  $\Pi \models_C \Sigma$  and for no proper subsets of  $\Pi$  and  $\Sigma$ ,  $\Pi'$  and  $\Sigma'$ , respectively,  $\Pi' \models_C \Sigma'$ . The natural extension of the proof theory is to add the conditional rules:

$$\frac{\Pi, \alpha : \beta, \Delta}{\Pi : \alpha \rightarrow \beta, \Delta} \quad \frac{\Pi_1, \alpha : \Delta_1 \quad \beta, \Pi_2 : \Delta_2}{\Pi_1, \Pi_2, \alpha \rightarrow \beta : \Delta_1, \Delta_2}$$

Unfortunately, the equivalence between these two approaches now fails. For, semantically,  $p \models \neg p \rightarrow q$  (though the system is still paraconsistent); but without dilution there is no proof of the sequent  $p : \neg p \rightarrow q$ . At this point, Tennant prefers to go with the proof theory rather than the semantics. He also prefers the intuitionist version, which allows at most one formula on the right-hand side of a sequent. For further details, including natural deduction versions of the proof theory, see Tennant [1987], ch. 23.

In [1992] Tennant suggests modifying the rule for the introduction of  $\rightarrow$  on the right.<sup>58</sup> The  $\alpha$  in the premise sequent is made optional, and the following rule is added.

<sup>57</sup>One can modify this approach, invoking the filter in the truth conditions of the conditional itself, to give logics of a more relevant variety. This is pursued in a number of the essays in *Philosophical Studies* 26 (1979), no. 2, a special issue on relatedness logics.

<sup>58</sup>In fact, he gives the natural deduction rules. The sequent rules described are the obvious equivalents.

$$\frac{\Pi, \alpha : \Delta}{\Pi : \alpha \rightarrow \beta, \Delta}$$

The exact relationship between these rules and the above semantics is as yet unresolved.

In the non-adjunctive logics of Rescher and Manor, and Schotch and Jennings:  $\rightarrow$  may again be identified with  $\supset$ , producing no novelties. The machinery of maximally consistent subsets and partitions carries straight over.

### 5.2 Discursive Implication

The situation is otherwise with discursive logic. Here a distinct approach is required, since, as we have already seen, the disjunctive syllogism fails discursively.

Given an  $S5$  interpretation, Jaśkowski adds a conditional,  $\rightarrow$  (often written as  $\supset_d$ , and called discursive implication), and defines  $\alpha \rightarrow \beta$  as  $\diamond\alpha \supset \beta$ .<sup>59</sup> It is easy to check that in discursive logic  $\alpha, \alpha \rightarrow \beta \models \beta$ , since  $\diamond\alpha, \diamond(\diamond\alpha \supset \beta) \models_{S5} \diamond\beta$  (and so there are essentially multi-premise inferences).

In fact, the logical truths of the pure  $\rightarrow$  fragment of discursive logic are the same as those of the pure  $\supset$  fragment of classical logic. For let  $\alpha_{\supset}$  be any sentence containing only  $\supset$ s, and let  $\alpha_{\rightarrow}$  be the corresponding sentence containing only  $\rightarrow$ s. In an  $S5$  interpretation with only one world,  $\alpha_{\supset}$  and  $\diamond\alpha_{\rightarrow}$  are equivalent. So if  $\alpha_{\supset}$  is not a classical logical truth,  $\diamond\alpha_{\rightarrow}$  is not a discursive one. Conversely, suppose that  $\alpha_{\supset}$  is a classical logical truth. We need to show that  $\diamond\alpha_{\rightarrow}$  is valid in every  $S5$  model. As may easily be checked, in  $S5$ ,  $\diamond(\diamond\alpha \supset \beta)$  is logically equivalent to  $\diamond\alpha \supset \diamond\beta$ . Hence, given  $\diamond\alpha_{\rightarrow}$ , we may “drive the  $\diamond$ s inwards” to obtain a logically equivalent sentence where the modal operator applies only to propositional parameters. But this is a substitution instance of  $\alpha_{\supset}$ , and hence valid in  $S5$ . This result does not carry over to the full language. For example,  $\not\models \alpha \rightarrow (\neg\alpha \rightarrow \beta)$ , since, as may be checked,  $\not\models_{S5} \diamond(\diamond\alpha \supset (\diamond\neg\alpha \supset \beta))$ .<sup>60</sup>

Full discursive logic can naturally be generalised in two obvious ways. The first is by using some modal logic other than  $S5$ . The second is by changing the definition of what it is for a sentence,  $\alpha$ , to hold discursively in an interpretation. We change this from  $\diamond\alpha$  holding to  $M\alpha$  holding, where  $M$  is some other modality (i.e., string of  $\diamond$ s and  $\Box$ s). For references and discussion, see Błaszczuk [1984] and Kotas and da Costa [1989].

<sup>59</sup> Given what amounts to Jaśkowski’s identification of truth with truth in some possible world, it might be more natural to define  $\alpha \rightarrow \beta$  as  $\diamond\alpha \rightarrow \diamond\beta$ . This would have just the same consequences.

<sup>60</sup> The natural definition of the biconditional,  $\alpha \leftrightarrow \beta$ , is  $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \alpha)$ . For reasons not explained, Jaśkowski defines it as  $(\alpha \rightarrow \beta) \wedge (\beta \rightarrow \diamond\alpha)$ . This asymmetric and counter-intuitive definition would seem to have no significant advantages.

### 5.3 Da Costa's $C$ -systems

The natural way of extending the non-truth functional semantics of 4.4 to include a conditional connective, in keeping with the idea that such logics are just the addition of a non-truth-functional negation to a standard positive logic, is to give  $\rightarrow$  the classical truth conditions:

$$\nu(\alpha \rightarrow \beta) = 1 \text{ iff } \nu(\alpha) = 0 \text{ or } \nu(\beta) = 1$$

(Note that  $\rightarrow$ , so defined, is distinct from  $\supset$ .) Adding this condition to the logics of 4.4 (except,  $C_\omega$ , which we will come to in a moment) gives the full (propositional) versions of the logics mentioned there; in particular it gives the da Costa logic  $C_1$  (and the other  $C_i$  for finite non-zero  $i$ ). In each case, a natural deduction system can be obtained by adding the rules:

$$\rightarrow E \quad \frac{\alpha \quad \alpha \rightarrow \beta}{\beta}$$

$$(a) \quad \frac{\beta}{\alpha \rightarrow \beta}$$

$$(b) \quad \frac{}{\alpha \vee (\alpha \rightarrow \beta)}$$

Soundness is proved as usual. The extension to the completeness proof amounts to checking that for a prime theory,  $\Sigma$ ,  $\alpha \rightarrow \beta \in \Sigma$  iff  $\alpha \notin \Sigma$  or  $\beta \in \Sigma$ . From left to right, the result follows by ( $\rightarrow E$ ). From right to left: if  $\beta \in \Sigma$  then the result follows from (a); if  $\alpha \notin \Sigma$  then  $(\alpha \rightarrow \beta) \in \Sigma$  by (b) and primeness.

If instead of (a) and (b), we add to any of these systems—except the ones with a consistency operator; I will come to these in a second—the rule:

$$\rightarrow I \quad \frac{\begin{array}{c} \bar{\alpha} \\ \vdots \\ \beta \end{array}}{\alpha \rightarrow \beta}$$

we obtain, not classical positive logic, but intuitionist positive logic. (These rules are well known to be complete with respect to this logic.) In particular, if we add  $\rightarrow I$  and  $\rightarrow E$  to the rule system for the basic language fragment of  $C_\omega$  we obtain da Costa's  $C_\omega$ .

The intuitionist conditional is not, of course, truth functional, but a valuational semantics for  $C_\omega$  can be obtained as follows. A *semi-valuation* is any function that satisfies the conditions for conjunction, disjunction and negation, plus:

if  $\nu(\alpha \rightarrow \beta) = 1$  then  $\nu(\alpha) = 0$  or  $\nu(\beta) = 1$   
 if  $\nu(\alpha \rightarrow \beta) = 0$  then  $\nu(\beta) = 0$

A valuation is any semi-valuation,  $\nu$ , satisfying the following condition. Let  $\alpha$  be of the form  $\alpha_1 \rightarrow (\alpha_2 \rightarrow (\alpha_3 \dots \rightarrow \alpha_n) \dots)$ , where  $\alpha_n$  is not itself of the form  $\beta \rightarrow \gamma$ . Then if  $\nu(\alpha) = 0$  there is a semi-valuation,  $\nu'$ , such that for all  $1 \leq i < n$ ,  $\nu'(\alpha_i) = 1$ , and  $\nu'(\alpha_n) = 0$ .  $C_\omega$  is sound and complete with respect to this notion of valuation. For details, see Loparić [1986].<sup>61</sup>

Changing the deduction rules for  $\rightarrow$  to the intuitionist ones, makes no difference for those logics that contain a consistency operator, and in particular, the da Costa logics  $C_i$  for finite  $i$ .<sup>62</sup> The reason, *in nuce*, is that the consistency operator allows us to define a negation with the properties of classical negation. As is well known, the addition of such a negation to positive intuitionist logic is not conservative, but produces classical logic. In more detail, the argument for  $C_1$  is as follows.<sup>63</sup>

Define  $\neg^* \alpha$  as  $\neg \alpha \wedge \alpha^\circ$ . Then it is easy to check that:

$$\nu(\neg^* \alpha) = 1 \text{ iff } \nu(\alpha) = 0$$

In particular, then,  $\neg^* \alpha$  satisfies the rules for classical negation:

$$\frac{}{\alpha \vee \neg^* \alpha} \qquad \frac{\alpha \wedge \neg^* \alpha}{\beta}$$

Given these, it is easy to show that  $\alpha \rightarrow \beta \dashv\vdash \neg^* \alpha \vee \beta$ . (Hint: from left to right, assume  $\alpha \vee \neg^* \alpha$  and argue by cases. From right to left, assume  $\alpha$  and  $\neg^* \alpha \vee \beta$ , and argue to  $\beta$  by cases.) Hence,  $\rightarrow$  has the classical truth conditions.

#### 5.4 Many-valued Conditionals

There are numerous ways to define a many-valued conditional operator. We will just look at two of the more systematic.<sup>64</sup>

Given a Sugihara generalisation of  $LP$ , one can define a conditional with the following truth conditions:

---

<sup>61</sup>A Kripke-style semantics for  $C_\omega$  can be found in Baaz [1986].

<sup>62</sup>This was first observed, in effect, by da Costa and Guillaume [1965].

<sup>63</sup>The argument for the other  $C_i$ s is similar.

<sup>64</sup>In the three-valued case, other definitions give the system of Asenjo and Tamburino [1975], and the  $J$  systems of D'Ottaviano and da Costa [1970]. A natural many-valued conditional, given the four-valued semantics of  $FDE$ , produces the system  $BN4$  of Brady [1982].

$$\begin{aligned} \nu(\alpha \rightarrow \beta) &= \nu(\neg\alpha \vee \beta) && \text{if } \nu(\alpha) \leq \nu(\beta) \\ &= \nu(\neg\alpha \wedge \beta) && \text{if } \nu(\alpha) > \nu(\beta) \end{aligned}$$

This definition gives rise to “semi-relevant” logics, i.e., logics that avoid the standard paradoxes of relevance, but are still not relevant.

In the case where the set of truth values is the set of all integers, this gives the Anderson/Belnap logic *RM*. Proof-theoretically, *RM* is obtained from the relevant logic *R*, which we will come to in the next section, by adding the “mingle” axiom:

$$\vdash \alpha \rightarrow (\alpha \rightarrow \alpha)$$

For details of proofs, see Anderson and Belnap [1975], sect. 29.3.

In the 3-valued case, where the set of truth values is  $\{-1, 0, +1\}$ , the conditions for  $\rightarrow$  give the matrix:

$\rightarrow$	+1	0	-1
+1	+1	-1	-1
0	+1	0	-1
-1	+1	+1	+1

and the stronger logic called *RM3*. This is sound and complete with respect to the axiomatic system obtained by augmenting the system *R* with the axioms:

$$\begin{aligned} &\vdash (\neg\alpha \wedge \beta) \rightarrow (\alpha \rightarrow \beta) \\ &\vdash \alpha \vee (\alpha \rightarrow \beta) \end{aligned}$$

For the proof, see Brady [1982].

Turning to the second systematic approach, consider any Łukasiewicz generalisation of *LP*. Łukasiewicz’ truth conditions for his conditional,  $\mapsto$ , are as follows:

$$\begin{aligned} \nu(\alpha \mapsto \beta) &= 1 && \text{if } \nu(\alpha) \leq \nu(\beta) \\ &= 1 - (\nu(\alpha) - \nu(\beta)) && \text{if } \nu(\alpha) > \nu(\beta) \end{aligned}$$

In the three-valued case, this gives the well known matrix:

$\mapsto$	1	0.5	0
1	1	0.5	0
0.5	1	1	0.5
0	1	1	1



Now the most notable feature of the Łukasiewicz definition, given that 0.5 is designated, is that *modus ponens* fails. For example, consider a valuation,  $\nu$ , where  $\nu(p) = 0.5$  and  $\nu(q) = 0$ . Then  $\nu(p \mapsto q) = 0.5$ . Hence  $p, p \mapsto q \not\models q$ . (*Modus ponens* is valid provided that the only designated value is 1, but then the logic is not paraconsistent.)

Kotas and da Costa [1978] get around this problem by adding to the language a new operator,  $\Delta$ , with the truth conditions:

$$\begin{aligned} \nu(\Delta\alpha) &= 1 && \text{if } \nu(\alpha) \text{ is designated} \\ &= 0 && \text{otherwise} \end{aligned}$$

and then define a conditional,  $\alpha \rightarrow \beta$ , as  $\Delta\alpha \mapsto \beta$ .<sup>65</sup> They point out the similarity of this definition to Jaśkowski's definition of discursive implication. (In fact, they use the symbol  $\diamond$  instead of  $\Delta$  because of this.)<sup>66</sup>

It is not difficult to check that *modus ponens* for  $\rightarrow$  holds. In fact, as Kotas and da Costa point out, the  $\wedge, \vee, \rightarrow$ -fragment of the logic is exactly positive classical logic. The easiest way to see this is just to collapse the designated values to 1, and the others to 0, to obtain classical truth tables.

### 5.5 Relevant $\rightarrow$ s

Given a Routley interpretation (say one for *FDE*, though the other cases will be similar), it is natural to treat  $\rightarrow$  intensionally. The simplest way is to give it the *S5* truth conditions:

$$\nu_w(\alpha \rightarrow \beta) = 1 \text{ iff for all } w' \in W (\nu_{w'}(\alpha) = 1 \Rightarrow \nu_{w'}(\beta) = 1)$$

Clearly, given an interpretation either  $\alpha \rightarrow \beta$  is true at all worlds, or at none. With the Routley  $*$  giving the semantics for negation, it follows that the same is true of negated conditionals. It also follows that  $\nu_w(\alpha \rightarrow \beta) = 1$  iff  $\nu_{w*}(\alpha \rightarrow \beta) = 1$  iff  $\nu_w\neg(\alpha \rightarrow \beta) \neq 1$ . Thus, the semantics validate the rules:

$$\text{LEM}_{\rightarrow} \quad \overline{(\alpha \rightarrow \beta) \vee \neg(\alpha \rightarrow \beta)}$$

$$\text{EFQ}_{\rightarrow} \quad \frac{\alpha \rightarrow \beta \quad \neg(\alpha \rightarrow \beta)}{\gamma}$$

and so are unsuitable for serious paraconsistent purposes. Moreover, even though there may be worlds where  $\alpha \wedge \neg\alpha$  is true, or where  $\alpha \vee \neg\alpha$  is false,

<sup>65</sup>In fact, their treatment is more general, since they consider the case in which the extension of  $\Delta$  may be other than the set of designated values.

<sup>66</sup>Peña [1984] defines an operator,  $F$ , on real numbers such that the value  $F\alpha$  is 0 if that of  $\alpha$  is greater than 0, and 1 otherwise; and then defines a conditional operator,  $\alpha C\beta$ , as  $F\alpha \vee \beta$ . The result is similar.

and so neither  $(\alpha \wedge \neg\alpha) \rightarrow \beta$  nor  $\alpha \rightarrow (\alpha \vee \neg\alpha)$  is valid, the system is not a relevant one since, e.g.,  $\models p \rightarrow (q \rightarrow q)$ .

These facts may both be changed by modifying the semantics, by adding a class of non-normal worlds. Thus, an interpretation is a structure  $\langle W, N, *, \nu \rangle$ . The worlds in  $N$  are called *normal*; the worlds in  $W - N$  ( $NN$ ) are called *non-normal*. Truth conditions are the same as before, except that at non-normal worlds, the truth value of a conditional is arbitrary. Technically,  $\nu$  assigns to every pair of world and propositional parameter a truth value, as before, but for every  $w \in NN$  and every conditional  $\alpha \rightarrow \beta$ , it now also assigns  $\alpha \rightarrow \beta$  a value at  $w$ . This provides the value of  $\alpha \rightarrow \beta$  at non-normal worlds (non-recursively). Validity is defined as truth preservation at all *normal* worlds of all interpretations.

If one thinks of the conditionals as entailments, then the non-normal worlds are those where the facts of logic may be different. Thus, one may think of non-normal worlds as logically impossible situations.<sup>67</sup>

The system described is called *H* in Routley and Loparić [1978].<sup>68</sup> It is sound and weakly complete (i.e., theorem-complete) with respect to the following axiom system.

$$\begin{aligned} &\vdash \alpha \rightarrow \alpha \\ &\vdash (\alpha \wedge \beta) \rightarrow \alpha \quad \vdash (\alpha \wedge \beta) \rightarrow \beta \\ &\vdash \beta \rightarrow (\alpha \vee \beta) \quad \vdash \beta \rightarrow (\alpha \vee \beta) \\ &\vdash \alpha \leftrightarrow \neg\neg\alpha \\ &\vdash (\neg\alpha \vee \neg\beta) \leftrightarrow \neg(\alpha \wedge \beta) \\ &\vdash (\neg\alpha \wedge \neg\beta) \leftrightarrow \neg(\alpha \vee \beta) \\ &\vdash (\alpha \wedge (\beta \vee \gamma)) \rightarrow ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \end{aligned}$$

$$\begin{aligned} &\text{If } \vdash \alpha \text{ and } \vdash \alpha \rightarrow \beta \text{ then } \vdash \beta \\ &\text{If } \vdash \alpha \text{ and } \vdash \beta \text{ then } \vdash \alpha \wedge \beta \\ &\text{If } \vdash \alpha \rightarrow \beta \text{ and } \vdash \beta \rightarrow \gamma \text{ then } \vdash \alpha \rightarrow \gamma \\ &\text{If } \vdash \alpha \rightarrow \beta \text{ then } \vdash \neg\beta \rightarrow \neg\alpha \\ &\text{If } \vdash \alpha \rightarrow \beta \text{ and } \vdash \alpha \rightarrow \gamma \text{ then } \vdash \alpha \rightarrow (\beta \wedge \gamma) \\ &\text{If } \vdash \alpha \rightarrow \gamma \text{ and } \vdash \beta \rightarrow \gamma \text{ then } \vdash (\alpha \vee \beta) \rightarrow \gamma \end{aligned}$$

Strong (i.e., deducibility-) completeness requires also the rules in disjunctive form.<sup>69</sup> The disjunctive form of the first is:  $\vdash \alpha \vee \gamma$  and  $\vdash (\alpha \rightarrow \beta) \vee \gamma$  then  $\vdash \beta \vee \gamma$ . The others are similar.<sup>70</sup>

<sup>67</sup> For a further discussion of non-normality, see Priest [1992].

<sup>68</sup> There are several other systems in the vicinity here. Some are obtained by varying the conditions on  $*$ . Others, sometimes called the Arruda - da Costa *P* systems, are obtained by retaining the positive logic and adding a non-truth-functional negation. For details, see Routley and Loparić [1978].

<sup>69</sup> Which are known to be admissible anyway.

<sup>70</sup> A sound and complete natural deduction system is an open question.

Soundness is proved as usual. The (strong) completeness proof is as follows. We first show by induction on proofs that if  $\alpha \vdash \beta$  then  $\alpha \vee \gamma \vdash \beta \vee \gamma$ . It quickly follows that if  $\alpha \vdash \gamma$  and  $\beta \vdash \gamma$  then  $\alpha \vee \beta \vdash \gamma$ . Now suppose that  $\Gamma \not\vdash \alpha$ . Extend  $\Gamma$  to a prime theory,  $\Theta$ , with the same property, as in 4.3. Call a set  $\Delta$  a  $\Theta$ -theory if it is prime, closed under adjunction, and  $\beta \rightarrow \gamma \in \Theta \Rightarrow (\beta \in \Delta \Rightarrow \gamma \in \Delta)$ . Note that  $\Theta$  is a  $\Theta$ -theory. Define the interpretation  $\langle W, N, *, \nu \rangle$ , where  $W$  is the set of  $\Theta$ -theories;  $N = \{\Theta\}$ ,  $\beta \in \Delta^*$  iff  $\neg\beta \notin \Delta$  (which is well-defined). If  $\Delta \in NN$  then  $\nu_\Delta(\beta \rightarrow \gamma) = 1$  iff  $\beta \rightarrow \gamma \in \Delta$ ; and for all  $\Delta$ :

$$\nu_\Delta(p) = 1 \text{ iff } p \in \Delta$$

Once it can be shown that this condition carries over to all formulas, the result follows as usual. This is proved by induction. The only difficult case concerns  $\rightarrow$  when  $\Delta = \Theta$ . From right to left, the result follows from the definition of  $W$ . From left to right, the result follows from the following lemma. If  $\beta \rightarrow \gamma \notin \Theta$  then there is a  $\Theta$ -theory,  $\Delta$ , such that  $\beta \in \Delta$  and  $\gamma \notin \Delta$ . To prove this, we proceed essentially as in 4.3, except that  $\Sigma \vdash \Pi$  is redefined. Let  $\Theta_{\rightarrow}$  be the set of conditionals in  $\Theta$ ; then  $\Sigma \vdash \Pi$  is now taken to mean that there are  $\sigma_1, \dots, \sigma_n \in \Sigma$  and  $\pi_1, \dots, \pi_m \in \Pi$  such that  $\Theta_{\rightarrow} \vdash (\sigma_1 \wedge \dots \wedge \sigma_n) \rightarrow (\pi_1 \vee \dots \vee \pi_m)$ . Now set  $\Sigma = \{\beta\}$ , and  $\Pi = \{\gamma\}$ , and proceed as in 4.3. The rest of the details are left as a (lengthy) exercise.<sup>71</sup>

If we add the Law of Excluded Middle to the axiom system:

$$\vdash \alpha \vee \neg\alpha$$

we obtain a logic that we will call  $HX$ . In virtue of the discussion in 4.7, one might suppose that this would be sound and complete if we add the condition: for all  $w$ , and parameters,  $p$ ,  $1 = \nu_w(p)$  or  $0 = \nu_w^*(p)$ . This condition indeed makes  $\alpha \vee \neg\alpha$  true in all worlds; but for just that reason, it also verifies the irrelevant  $\beta \rightarrow (\alpha \vee \neg\alpha)$ . To obtain  $HX$ , we place this constraint on just normal worlds. The semantics are then just right, as may be checked. For further details, see Routley and Loparić [1978]. Since normal worlds are now, in effect,  $LP$  interpretations,  $HX$  verifies all the logical truths of  $LP$  and so of classical logic.

A feature of this system is that substitutivity of equivalents breaks down. For example, as is easy to check,  $p \leftrightarrow q \not\vdash (r \rightarrow p) \leftrightarrow (r \rightarrow q)$ . This can be changed by taking the valuation function to work on propositions (i.e., set of worlds), rather than formulas.<sup>72</sup> The most significant feature of semantics of this kind is that there are no principles of inference that employ nested

<sup>71</sup>Details can be found in Priest and Sylvan [1992].

<sup>72</sup>For details see Priest [1992].

conditionals in an essential way. This is due entirely to the anarchic nature of non-normal worlds. In effect, *any* breakdown of logic is countenanced.

One way of putting a little order into the anarchy without destroying relevance, proposed by Routley and Meyer,<sup>73</sup> is by employing a ternary relation,  $R$ , to give the truth conditions of conditionals at non-normal worlds. An interpretation is now of the form  $\langle W, N, R, *, \nu \rangle$ . All is as before, except that  $\nu$  no longer gives the truth values of conditionals at non-normal worlds. Rather, for any  $w \in NN$ , the truth conditions are:

$$\nu_w(\alpha \rightarrow \beta) = 1 \text{ iff for all } x, y \in W, Rwx y \Rightarrow (\nu_x(\alpha) = 1 \Rightarrow \nu_y(\beta) = 1)$$

Note that this is just the standard condition for strict implication, except that the worlds of the antecedent ( $x$ ) and the consequent ( $y$ ) have become distinguished. What, exactly, the ternary relation,  $R$ , means, is still a matter for philosophical deliberation. Validity is again defined as truth preservation at all normal worlds.

These semantics give the basic system of affixing relevant logic,  $B$ . An axiom system therefor can be obtained by replacing the last two rules for  $H$  by the corresponding axioms:

$$\begin{aligned} &\vdash ((\alpha \rightarrow \beta) \wedge (\alpha \rightarrow \gamma)) \rightarrow (\alpha \rightarrow (\beta \wedge \gamma)) \\ &\vdash ((\alpha \rightarrow \gamma) \wedge (\beta \rightarrow \gamma)) \rightarrow ((\alpha \vee \beta) \rightarrow \gamma) \end{aligned}$$

and adding a rule that ensures replacement of equivalents:

$$\text{If } \vdash \alpha \rightarrow \beta \text{ and } \vdash \gamma \rightarrow \delta \text{ then } \vdash (\beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \delta)$$

The soundness and completeness proofs generalise those for  $H$ . Details can be found in Priest and Sylvan [1992].

We may form the system  $BX$  proof theoretically by adding the Law of Excluded Middle. Semantically, we proceed as with  $H$ , placing the appropriate condition on normal worlds.

As with modal logics, stronger logics can be obtained by placing conditions on the accessibility relation,  $R$ . In this way, most of the logics in the Anderson/Belnap family can be generated. Details can be found in Restall [1993]. The strongest of these is the logic  $R$ , an axiom system for which is as follows:

$$\begin{aligned} &\vdash \alpha \rightarrow \alpha \\ &\vdash (\alpha \rightarrow \beta) \rightarrow ((\beta \rightarrow \gamma) \rightarrow (\alpha \rightarrow \gamma)) \\ &\vdash \alpha \rightarrow ((\alpha \rightarrow \beta) \rightarrow \beta) \end{aligned}$$

<sup>73</sup>Initially, this was in Routley and Meyer [1973]. For further discussion of all the following, see the article on Relevant Logic in this volume of the *Handbook*.

$$\begin{aligned}
 &\vdash (\alpha \rightarrow (\alpha \rightarrow \beta)) \rightarrow (\alpha \rightarrow \beta) \\
 &\vdash (\alpha \wedge \beta) \rightarrow \beta, \quad \vdash (\alpha \wedge \beta) \rightarrow \alpha \\
 &\vdash ((\alpha \rightarrow \beta) \wedge (\alpha \rightarrow \gamma)) \rightarrow (\alpha \rightarrow (\beta \wedge \gamma)) \\
 &\vdash \beta \rightarrow (\alpha \vee \beta), \quad \vdash \alpha \rightarrow (\alpha \vee \beta) \\
 &\vdash ((\alpha \rightarrow \gamma) \wedge (\beta \rightarrow \gamma)) \rightarrow ((\alpha \vee \beta) \rightarrow \gamma) \\
 &\vdash (\alpha \wedge (\beta \vee \gamma)) \rightarrow ((\alpha \wedge \beta) \vee \gamma) \\
 &\vdash (\alpha \rightarrow \neg\beta) \rightarrow (\beta \rightarrow \neg\alpha) \\
 &\vdash \neg\neg\alpha \rightarrow \alpha
 \end{aligned}$$

with the rules of adjunction and *modus ponens*.

The equivalence between the Dunn 4-valued semantics and the Routley \* operation that we noted in 4.7 suggests another way of obtaining an intensional conditional connective. In the simplest case, an interpretation is a structure  $\langle W, \nu \rangle$  where  $W$  is a set of worlds and  $\nu$  is an evaluation of the parameters at worlds, but this time it is a Dunn 4-valued interpretation. The truth conditions for the basic language are as in 4.6, except that they are relativised to worlds. Thus, using the functional notation:

$$\begin{aligned}
 1 \in \nu_w(\neg\alpha) &\text{ iff } 0 \in \nu_w(\alpha) \\
 0 \in \nu_w(\neg\alpha) &\text{ iff } 1 \in \nu_w(\alpha)
 \end{aligned}$$

$$\begin{aligned}
 1 \in \nu_w(\alpha \wedge \beta) &\text{ iff } 1 \in \nu_w(\alpha) \text{ and } 1 \in \nu_w(\beta) \\
 0 \in \nu_w(\alpha \wedge \beta) &\text{ iff } 0 \in \nu_w(\alpha) \text{ or } 0 \in \nu_w(\beta)
 \end{aligned}$$

$$\begin{aligned}
 1 \in \nu_w(\alpha \vee \beta) &\text{ iff } 1 \in \nu_w(\alpha) \text{ or } 1 \in \nu_w(\beta) \\
 0 \in \nu_w(\alpha \vee \beta) &\text{ iff } 0 \in \nu_w(\alpha) \text{ and } 0 \in \nu_w(\beta)
 \end{aligned}$$

The natural truth and falsity conditions for  $\rightarrow$  are:

$$\begin{aligned}
 1 \in \nu_w(\alpha \rightarrow \beta) &\text{ iff for all } w' \in W, (1 \in \nu_{w'}(\alpha) \Rightarrow 1 \in \nu_{w'}(\beta)) \\
 0 \in \nu_w(\alpha \rightarrow \beta) &\text{ iff for some } w' \in W, 1 \in \nu_{w'}(\alpha) \text{ and } 0 \in \nu_{w'}(\beta)
 \end{aligned}$$

These semantics do not validate the undesirable:

$$\frac{\alpha \rightarrow \beta \quad \neg(\alpha \rightarrow \beta)}{\gamma}$$

as their \* counterparts do. But they are still not relevant. Relevant logics can be obtained by adding a class of non-normal worlds. The semantic values of conditionals at these may either be arbitrary, as with  $H$ , or, as with  $B$ , we may employ a ternary relation and give the conditions as follows:

$$\begin{aligned}
1 \in \nu_w(\alpha \rightarrow \beta) & \text{ iff for all } x, y \in W, Rwx y \Rightarrow (1 \in \nu_x(\alpha) \Rightarrow 1 \in \nu_y(\beta)) \\
0 \in \nu_w(\alpha \rightarrow \beta) & \text{ iff for some } x, y \in W, Rwx y, 1 \in \nu_x(\alpha) \text{ and } 0 \in \nu_y(\beta)
\end{aligned}$$

As usual, extra conditions may be imposed on  $R$ . This construction produces a family of relevant logics distinct from the usual ones, and one that has not been studied in great detail. One way in which it differs from the more usual ones is that contraposition of the conditional fails, though this can be rectified by modifying the truth conditions for  $\rightarrow$  by adding the clause: ‘and  $0 \in \nu_{w'}(\beta) \Rightarrow 0 \in \nu_{w'}(\alpha)$ ’ (or in the case of non-normal worlds employing a ternary relation: ‘and  $0 \in \nu_x(\beta) \Rightarrow 0 \in \nu_x(\alpha)$ ’). A more substantial difference concerns negated conditionals. Because of the falsity conditions of the conditional, all logics of this family validate  $\alpha \wedge \neg\beta \models \neg(\alpha \rightarrow \beta)$ . This is a natural enough principle, but absent from many of the logics obtained using the Routley  $*$ .

The more usual relevant logics can be obtained with the 4-valued semantics, but only by using some *ad hoc* device or other, such as an extra accessibility relation, or allowing only certain classes of worlds. For details, see Routley [1984] and Restall [1995].

### 5.6 $\rightarrow$ as $\leq$

There is a very natural way of employing any algebra which has an ordering relation to give a semantics for conditionals. One may think of the members of the algebra as propositions, or as Fregean senses. The relation  $\leq$  on the algebra can be thought of as an entailment relation, and it is then natural to take  $\alpha \rightarrow \beta$  to hold in some interpretation,  $\nu$ , iff  $\nu(\alpha) \leq \nu(\beta)$ . The problem, then, is to express the thought that  $\alpha \rightarrow \beta$  holds in algebraic terms. We obviously need an algebraic operator,  $\rightarrow$ , corresponding to the connective; but how is one to express the idea that  $a \rightarrow b$  holds when the algebra may have no maximal element?

A way to solve this problem for De Morgan algebras is to employ a designated member of the lattice,  $e$ , and take the things that hold in the algebra to be those whose values are  $\geq e$ .<sup>74</sup> While we are introducing new machinery, it is also useful algebraically to introduce another binary (groupoid) operator,  $\circ$ , often called ‘fusion’, whose significance we will come back to in a moment. We may also enrich the basic language to one containing a constant,  $e$ , and an operator,  $\circ$ , expressing the new algebraic features.

Thus, following Meyer and Routley [1972], let us call the structure  $\mathcal{A} = \langle \mathcal{D}, e, \rightarrow, \circ \rangle$  a *De Morgan groupoid* iff  $\mathcal{D}$  is a De Morgan algebra,  $\langle \mathcal{A}, \wedge, \vee, \neg \rangle$ , and for any  $a, b, c \in \mathcal{A}$ :

<sup>74</sup>A different way is to let  $T$  be a prime filter on the lattice, thought of as the set of all true propositions. We can then require that  $a \rightarrow b \in T$  iff  $a \leq b$ . For details, see Priest [1980].

$$\begin{aligned}
 e \circ a &= a \\
 a \circ b \leq c &\text{ iff } a \leq b \rightarrow c \\
 \text{if } a \leq b &\text{ then } a \circ c \leq b \circ c \text{ and } c \circ a \leq c \circ b \\
 a \circ (b \vee c) &= (a \circ b) \vee (a \circ c) \text{ and } (b \vee c) \circ a = (b \circ a) \vee (c \circ a)
 \end{aligned}$$

The first of these conditions ensures that  $e$  is a left identity on the groupoid. (Note that the groupoid may not be commutative.) And it, together with the second, ensure that  $a \leq b$  iff  $e \leq a \rightarrow b$ . The third and fourth ensure that  $\circ$  respects the lattice operations in a certain sense. The sense is question in that of a sort of conjunction, and this makes it possible to think of fusion as a kind of intensional conjunction.

An inference,  $\alpha_1, \dots, \alpha_n/\beta$ , is algebraically valid iff for every homomorphism,  $\nu$ , into a De Morgan groupoid,  $\nu(\alpha_1 \wedge \dots \wedge \alpha_n) \leq \nu(\beta)$ , i.e.,  $e \leq \nu((\alpha_1 \wedge \dots \wedge \alpha_n) \rightarrow \beta)$ .<sup>75</sup>

These semantics are sound and complete with respect to the relevant logic  $B$  of 5.5. Soundness is shown in the usual way, and completeness can be proved, as in 4.8, by constructing the Lindenbaum algebra, and showing that it is a De Morgan groupoid.

Stronger logics can be obtained, as usual, by adding further constraints. The condition:  $e \leq a \vee \neg a$  gives the law of excluded middle (and all classical tautologies). Additional constraints on  $\circ$  give the stronger logics in the usual relevant family, including  $R$ . Details of all the above can be found in Meyer and Routley [1972] (who also show how to translate between algebraic and world semantics).<sup>76</sup>

Before leaving the topic of conditionals in algebraic paraconsistent logics, a final comment on dual intuitionist logic. Goodman [1981] proves that in this logic there is no conditional operator (i.e., operator satisfying *modus ponens*) that can be defined in terms of  $\vee, \wedge$  and  $\neg$ ; and draws somewhat pessimistic conclusions from this concerning the usefulness of the logic. Such pessimism is not warranted, however. Exactly the same is true in relevant logic; this does not mean that a conditional operator cannot be added to the basic language. And as Mortensen notes,<sup>77</sup> given any algebraic structure with top ( $\top$ ) and bottom ( $\perp$ ) elements, the following conditions can always be used to define a conditional operator:

$$\begin{aligned}
 \nu(\alpha \rightarrow \beta) &= \top && \text{if } \nu(\alpha) \leq \nu(\beta) \\
 &= \perp && \text{otherwise}
 \end{aligned}$$

<sup>75</sup>A different notion of validity can be formulated using fusion thus:  $\nu(\alpha_1 \circ \dots \circ \alpha_n) \leq \nu(\beta)$ , i.e.,  $e \leq \nu((\alpha_1 \rightarrow (\dots \rightarrow (\alpha_n \rightarrow \beta)\dots))$ .

<sup>76</sup>See also Brink [1988]. A rather different algebraic approach which produces a relevant logic is given in Avro'n [1990]. This maintains an ordered structure, but dispenses with the lattice. The result is a logic closely related to the intensional fragment of  $RM$ .

<sup>77</sup>Mortensen [1995], p. 95.

Though this particular conditional is not suitable for robust paraconsistent purposes since it satisfies:  $\alpha \rightarrow \beta, \neg(\alpha \rightarrow \beta) \models \gamma$ .

### 5.7 *Decidability*

Before we leave the topic of propositional logics, let me review, briefly, the question of decidability for the logics that we have looked at. Unsurprisingly, most (though not all) are decidable, as the following decision procedures indicate. As will be clear, in many cases the procedures actually given could be greatly optimised.

Any filter logic is decidable if the filter is. Given any inference, we can effectively find the set of all inferences of which it is a uniform substitution instance. Provided that the filter is decidable, we can test each of these for prevalidity. If any of them is valid, the original inference is valid; otherwise not.

Smiley's filter is clearly decidable. So is Tennant's semantic filter. Given an inference with finite sets of premises and conclusions,  $\Sigma$  and  $\Pi$ , respectively, we can test the inference for classical validity. We may then test the inferences for all subsets of  $\Sigma$  and  $\Pi$ . (There is only a finite number of these.) If the original inference is valid, but its subinferences are not, it passes the test; otherwise not. Tennant's proof theory of 5.1 is also decidable. Anything provable has a Cut-free proof (since Cut is not a rule of proof). Decidability then follows as it does in the case of classical logic.

Turning to non-adjunctive logics: Jaśkowski's discursive logic is decidable; we may simply translate an inference into the corresponding one concerning *S5*, and use the *S5* decision procedure for this. The same obviously goes for any generalisation, provided only that the underlying modal logic is decidable.

Rescher and Manor's logic is decidable in the obvious way. Given any finite set of premises, we can compute all its subsets, the classical consistency of each of these, and hence determine which of the sets are maximally consistent. Once we have these, we can determine if any of them classically entails the conclusion. Similar comments apply to Schotch and Jennings' logic. Given any premise set, we can compute all its partitions, and so determine its level. For every partition of that size, we can test to see if one of its members classically entails the conclusion.

Non-truth-functional logics are also decidable by a simple procedure. Given an inference, we consider the set of all subformulas of the sentences involved (which is finite). We then consider all mappings from these to  $\{0, 1\}$ , the set of which is also finite. For each of these we go through and test whether it satisfies the appropriate constraints in the obvious way. Throwing away all those that do not, we see whether the conclusion holds



in all that remain.<sup>78</sup>

All finite many-valued logics are decidable by truth-tables. The infinite valued Lukasiewicz logics (and so their Kotas and da Costa augmentations) are not, in general, even axiomatisable, let alone decidable. (See Chang [1963].) This leaves *RM*. If there is a counter-model for an *RM* inference, there must be a number of maximum absolute value employed. Ignoring all the numbers in the model whose absolute size is greater than this gives a finite counter-model. Hence, *RM* has the finite model property. As is well known, any axiomatisable theory with this property is decidable. (Enumerate the theorems and the finite models simultaneously. Eventually we must find either a proof of a countermodel.)

Dual intuitionist logic is decidable since intuitionist logic is. We just compute the dual inference and test it with the intuitionist procedure.

This just leaves the logics of the relevant family. As we saw, the semantics of these can take either a world form or an algebraic form. The question of decidability here is the hardest and most sensitive. The weaker logics in the family are decidable, and can be shown to be so by semantic methods (such as filtration arguments) and/or proof theoretic ones (such as Gentzenisation plus Cut elimination).<sup>79</sup> The stronger ones, such as *R*, are not. Urquhart's [1984] proof of this fact contains one of the few applications of geometry to logic. A crucial principle in this context would seem to be contraction:  $(\alpha \rightarrow (\alpha \rightarrow \beta)) \rightarrow (\alpha \rightarrow \beta)$  (or various equivalent forms, such as  $(\alpha \wedge (\alpha \rightarrow \beta)) \rightarrow \beta$ ). Speaking very generally, systems without this principle are decidable; systems with it are not.

## 6 QUANTIFIERS

The novelty of paraconsistent logic lies, it is fair to say, almost entirely at the propositional level. However, if a logic is to be applied in any serious way, it must be quantificational. Most of the paraconsistent logics that we have considered extend in straightforward ways to quantified logics. In this section I will indicate how. Let us suppose that the propositional language is now augmented to a language, *L*, with predicates, constants, variables and the quantifiers  $\forall$  and  $\exists$  in the usual way. I will let the adicity of a predicate be shown by the context. Propositional parameters can be identified with predicates of adicity 0. I will write  $\alpha(x/t)$  to mean the result of substituting the term *t* for all free occurrences of *x*, any bound variables in  $\alpha$  having been relabelled, if necessary, to avoid clashes.

I will reserve the word 'sentence' for formulas without free variables. I will always define validity for inferences containing only sentences, though the accounts could always be extended to ones employing all formulas, in

<sup>78</sup>For the method applied to the da Costa systems, see da Costa and Alves [1977].

<sup>79</sup>See, respectively, Routley *et al.* [1982], sect. 5.9, and Brady [1991].

standard ways. Where quantifiers have an objectual interpretation, and the set of objects is  $D$ , I will assume—for the rest of this essay—that the language has been augmented by a set of constants in such a way that each member of the domain has a name. In particular, I will always assume that the names are the members of  $D$  themselves, and that each object names itself. This assumption is never essential, but it simplifies the notation.

### 6.1 *Filter and Non-adjunctive Logics*

In filter logics, we may simply take the filter to be a relation on the extended language. Smiley's filter works equally well, for example, when the notion of classical logical truth employed is that for first order, not propositional, logic. Similarly for Tennant's. In his case (without the conditional operator), the semantics are sound and complete with respect to the sequent calculus of 4.1 for the basic language, together with the usual rules for the quantifiers:

$$\frac{\Gamma : \alpha(x/c), \Delta}{\Gamma : \forall x\alpha, \Delta} \qquad \frac{\Gamma, \alpha : \Delta}{\Gamma, \forall x\alpha : \Delta}$$

$$\frac{\Gamma : \alpha, \Delta}{\Gamma : \exists x\alpha, \Delta} \qquad \frac{\Gamma, \alpha(x/c) : \Delta}{\Gamma, \exists x\alpha : \Delta}$$

where in the first and last of these,  $c$  does not occur in any formula in  $\Gamma$  or  $\Delta$ . For proofs, see Tennant [1984]. (With the conditional operator added, the situation is different, as we saw in 5.1.)

Non-adjunctive logic accommodates quantifiers in an obvious way. Consider discursive logic. An inference in the quantified language is discursively valid iff  $\diamond\Sigma \models_{CS5} \diamond\alpha$ , where  $CS5$  is constant-domain quantified  $S5$ . Clearly, any other quantified modal logic could be used to generalise this notion.<sup>80</sup>

Rescher and Manor's approach and Schotch and Jennings' also generalise in the obvious way, the classical notion of propositional consequence involved being replaced by the classical first-order notion. In the quantificational case, the usefulness of these logics is moot, since the computation of classically maximally consistent sets of premises, or partitions, is highly non-effective.

In all these logics, except Smiley's, the set of logical truths (in the appropriate vocabulary) coincides with that of classical quantifier logic; hence these logics are undecidable.<sup>81</sup>

<sup>80</sup> For details of quantified modal logic, see the article on that topic in this *Handbook*.

<sup>81</sup> I do not know whether Smiley's logic is decidable, though I assume that it is not.

### 6.2 Positive-plus Logics

Let us turn now to the logics that augment classical or intuitionist positive logic with a non-truth-functional negation. Since the semantics of these are not truth functional, the most natural quantifier semantics are not objectual, but substitutional. Let me illustrate this with the simplest non-truth-functional logic, with a classical conditional operator, but no semantic constraints on negation. Extensions of this to other cases are left as an exercise.

An interpretation is a pair  $\langle C, \nu \rangle$ .  $C$  is a set of constants, and  $L_C$  is the language  $L$  augmented by the constants  $C$ .  $\nu$  is a map from the sentences of  $L_C$  to  $\{1, 0\}$  satisfying the same conditions as in the propositional case, together with:

$$\begin{aligned} \nu(\forall x\alpha) &= 1 \text{ iff for every constant of } L_C, c, \nu(\alpha(x/c)) = 1 \\ \nu(\exists x\alpha) &= 1 \text{ iff for some constant of } L_C, c, \nu(\alpha(x/c)) = 1 \end{aligned}$$

An inference is valid iff it is truth-preserving in all interpretations.

The semantics are sound and complete with respect to the quantifier rules:

$$\forall I \quad \frac{\beta \vee \alpha(x/c)}{\beta \vee \forall x\alpha}$$

provided that  $c$  does not occur in  $\beta$ , or in any undischarged assumption on which the premise depends.

$$\forall E \quad \frac{\forall x\alpha}{\alpha(x/c)}$$

$$\exists I \quad \frac{\alpha(x/c)}{\exists x\alpha}$$

$$\exists E \quad \frac{\exists x\alpha \quad \overline{\alpha(x/c)} \quad \vdots \quad \beta}{\beta}$$

provided that  $c$  does not occur in  $\beta$  or in any undischarged assumption in the subproof.

Soundness is proved by a standard recursive argument. For completeness, call a theory,  $\Sigma$ , *saturated* in a set of constants,  $C$ , iff:

$$\exists x\alpha \in \Sigma \text{ iff for some } c \in C, \alpha(x/c) \in \Sigma$$

$$\forall x \alpha \in \Sigma \text{ iff for every } c \in C, \alpha(x/c) \in \Sigma$$

It is easy to check that if  $\Delta$  is a prime theory, saturated in  $C$ , then  $\langle C, \nu \rangle$  is an interpretation, where  $\nu$  is defined by:  $\nu(\alpha) = 1$  iff  $\alpha \in \Delta$ .

It remains to show that if  $\Sigma \not\vdash \alpha$  then  $\Sigma$  can be extended to a prime theory,  $\Delta$ , saturated in some set of constants,  $C$ , with the same property; and the result follows as in the propositional case, using  $\Delta$  to define the interpretation.

To show this, we augment the language with an infinite set of new constants,  $C$ , and then extend the proof of 4.3 as follows. Enumerate the formulas of  $L_C$ :  $\beta_0, \beta_1, \dots$ . If  $\forall x\beta$  or  $\exists x\beta$  occurs in the enumeration, and the constant  $c$  does not occur in any preceding formula, we will call  $\beta(x/c)$  a *witness*. Now, we run through the enumeration, as before, but this time, if we throw  $\exists x\beta$  into the  $\Sigma$  side, we also throw in a witness; and if we throw  $\forall x\beta$  into the  $\Pi$  side, we also throw in a witness. In proving that  $\Sigma_n \not\vdash \Pi_n$ , the only novelty is when a witness is present; and these can be ignored, by  $\exists E$  on the left, and  $\forall I$  on the right. The rest of the proof is as in 4.3. The saturation of  $\Delta$  in  $C$  follows from deductive closure and construction.

I observe that all the logics in this family contain positive classical quantifier logic, and so are undecidable.

### 6.3 Many-valued Logics

Most of the many-valued logics with numerical values that we considered in 4.5 and 5.4 had two particular properties. First, the truth value of a conjunction [disjunction] is the minimum [maximum] of the values of the conjuncts [disjuncts]. Second, the set of truth values is closed under greatest lower bounds (glbs) and least upper bounds (lubs), i.e., if  $Y \subseteq X$  then  $\text{glb}(Y) \in X$  and  $\text{lub}(Y) \in X$ . Any such logic can be extended to a quantified logic in a very natural way, merely by treating  $\forall$  and  $\exists$  as the “infinitary” generalisations of conjunction and disjunction, respectively.

Specifically, a quantifier interpretation adds to the propositional machinery, the pair  $\langle D, d \rangle$  where  $D$  is a non-empty domain of objects,  $d$  maps every constant into  $D$ , and if  $P$  is an  $n$ -place predicate,  $d$  maps  $P$  to a function from  $n$ -tuples of the domain into the set of truth-values. Every sentence,  $\alpha$ , can now be assigned a truth value,  $\nu(\alpha)$ , in the natural way. For atomic sentences,  $Pc_1 \dots c_n$ :

$$\nu(Pc_1 \dots c_n) = d(P) \langle d(c_1) \dots d(c_n) \rangle$$

The truth conditions for propositional connectives are as in the propositional logic. The truth conditions for the quantifiers are:

$$\begin{aligned}\nu(\forall x\alpha) &= \text{glb}\{\alpha(x/c); c \in D\} \\ \nu(\exists x\alpha) &= \text{lub}\{\alpha(x/c); c \in D\}\end{aligned}$$

Validity is defined in terms of preservation of designated values, as in the propositional case.

I will make just a few comments about what happens when these definitions are applied to the many-valued logics we have looked at. The quantified finite-valued logics of 4.5 all collapse into quantified *LP* (which we will come to in the next section), as extensions of the arguments given there, show. For a general theory of quantified finitely-many-valued logics, see Rosser and Turquette [1952]. Quantified *RM* we will come to in a later section. Infinite-valued Łukasiewicz logics are proof-theoretically problematic. For a start, standard quantifier rules may break down. In particular,  $\forall x\alpha$  may be undesignated, even though each substitution instance is designated. Thus,  $\forall I$  may fail. (Similarly for existential quantification.) Worse, as for their propositional counterparts, such logics are not even axiomatisable in general.<sup>82</sup>

#### 6.4 *LP and FDE*

The technique of extending a many-valued logic to a quantified one can be put in a slightly different, and possibly more illuminating, way for the logics with relational semantics, *LP* and *FDE*. An interpretation,  $\mathcal{I}$ , is a pair,  $\langle D, d \rangle$ , where  $D$  is the usual domain of quantification,  $d$  is a function that maps every constant into the domain, and every  $n$ -place predicate into a pair,  $\langle E_P, A_P \rangle$ , each member of which is a subset of the set of  $n$ -tuples of  $D, D^n$ .  $E_P$  is the *extension* of  $P$ ;  $A_P$  is the *anti-extension*. For *LP* interpretations, we require, in addition, that  $E_P \cup A_P = D^n$ . Truth values are now assigned to sentences in accord with the following conditions. For atomic sentences:

$$\begin{aligned}1 \in \nu(Pc_1 \dots c_n) &\text{ iff } \langle d(c_1), \dots, d(c_n) \rangle \in E_P \\ 0 \in \nu(Pc_1 \dots c_n) &\text{ iff } \langle d(c_1), \dots, d(c_n) \rangle \in A_P\end{aligned}$$

Truth/falsity conditions for connectives are as in the propositional case; and for the quantifiers:

$$\begin{aligned}1 \in \nu(\forall x\alpha) &\text{ iff for every } c \in D, 1 \in \nu(\alpha(x/c)) \\ 0 \in \nu(\forall x\alpha) &\text{ iff for some } c \in D, 0 \in \nu(\alpha(x/c))\end{aligned}$$

$$1 \in \nu(\exists x\alpha) \text{ iff for some } c \in D, 1 \in \nu(\alpha(x/c))$$

---

<sup>82</sup>See Chang [1963] for details.

$0 \in \nu(\exists x\alpha)$  iff for every  $c \in D$ ,  $0 \in \nu(\alpha(x/c))$

An inference is valid iff it is truth-preserving in all interpretations. It should be noted that if for every predicate,  $P$ ,  $E_P$  and  $A_P$  are exclusive and exhaustive, we have an interpretation of classical first order logic. All classical interpretations are therefore *FDE* (and *LP*) interpretations.

These semantics are sound and complete if we add to the rules for *LP* or *FDE*, the rules  $\forall I$ ,  $\forall E$ ,  $\exists I$  and  $\exists E$ , plus:

$$\frac{\forall x\neg\alpha}{\neg\exists x\alpha} \quad \frac{\exists x\neg\alpha}{\neg\forall x\alpha}$$

Soundness is established by the usual argument. For completeness, suppose that  $\Sigma \not\vdash \alpha$ . Extend  $\Sigma$  to a set  $\Delta$ , which is prime, deductively closed and saturated in a set of new constants, such that  $\Delta \not\vdash \alpha$ , as in 6.2. Then define an interpretation  $\langle D, d \rangle$  where  $D$  is the set of constants of the extended language,  $d$  maps any constant to itself, and for any predicate,  $P$ , its extension and anti-extension are defined as follows:

$$\begin{aligned} \langle c_1, \dots, c_n \rangle \in E_P &\text{ iff } Pc_1\dots c_n \in \Delta \\ \langle c_1, \dots, c_n \rangle \in A_P &\text{ iff } \neg Pc_1\dots c_n \in \Delta \end{aligned}$$

We now establish that for all formulas,  $\alpha$  :

$$\begin{aligned} 1 \in \nu(\alpha) &\text{ iff } \alpha \in \Delta \\ 0 \in \nu(\alpha) &\text{ iff } \neg\alpha \in \Delta \end{aligned}$$

The argument is a routine induction. Here are the cases for  $\forall$ .

$$\begin{aligned} \forall x\alpha \in \Delta &\Leftrightarrow \text{for all } c, \alpha(x/c) \in \Delta && \text{saturation} \\ &\Leftrightarrow \text{for all } c, 1 \in \nu(\alpha(x/c)) && \text{induction hypothesis} \\ &\Leftrightarrow 1 \in \nu(\forall x\alpha) && \text{truth conditions of } \forall \\ \neg\forall x\alpha \in \Delta &\Leftrightarrow \exists x\neg\alpha \in \Delta && \text{quantifier rules} \\ &\Leftrightarrow \text{for some } c, \neg\alpha(x/c) \in \Delta && \text{saturation} \\ &\Leftrightarrow \text{for some } c, 0 \in \nu(\alpha(x/c)) && \text{induction hypothesis} \\ &\Leftrightarrow 0 \in \nu(\forall x\alpha) && \text{truth conditions of } \forall \\ &\Leftrightarrow 1 \in \nu(\neg\forall x\alpha) && \text{truth conditions of } \neg \end{aligned}$$

The monotonicity property of the propositional logics *LP* and *FDE* carries over to the quantified case. If  $\mathcal{I}_1$  and  $\mathcal{I}_2$  are any interpretations, with truth value assignments  $\nu_1$  and  $\nu_2$ , define  $\mathcal{I}_1 \leq \mathcal{I}_2$  to mean that  $\mathcal{I}_1$  and  $\mathcal{I}_2$  have the same domain, and for every predicate,  $P$ , the extension (anti-extension) of  $P$  in  $\mathcal{I}_1$  is a subset of the extension (anti-extension) of  $P$  in

$\mathcal{I}_2$ . A simple induction shows that if  $\mathcal{I}_1 \leq \mathcal{I}_2$  then for all formulas,  $\alpha$  (in a language with a name for every member of the domain),  $\nu_1(\alpha) \subseteq \nu_2(\alpha)$ . As in 4.6, it follows that the set of logical truths of *LP* is exactly the same as that of classical first order logic. And *FDE* has no logical truths (just consider an interpretation that makes the extension and anti-extension of every predicate empty).

Since classical quantifier logic is not decidable, neither is quantified *LP*. If  $P$  is any  $n$ -place predicate, let  $P_{LEM}$  be the sentence:  $\forall x_1 \dots \forall x_n (Px_1 \dots x_n \vee \neg Px_1 \dots x_n)$ . If  $P_{LEM}$  is true in an interpretation, then the extension and anti-extension of  $P$ , exhaust the  $n$ -tuples of the domain. If  $\alpha$  is any formula, let  $\alpha_{LEM}$  be the conjunction of all formulas of the form  $P_{LEM}$ , where  $P$  occurs in  $\alpha$ . It follows that  $\alpha_{LEM} \models \alpha$  in *FDE* iff  $\models \alpha$  in *LP*. Hence, quantified *FDE* is undecidable too.

### 6.5 Relevant Logics

Turning to relevant logics, the issues are more complex. This is due to the fact that there are various approaches to these logics, the variety of the logics themselves, and the intrinsic complexities of the stronger logics.

Let us start with the world semantics. As we saw in 5.5, a world semantics for a relevant logic with the Routley operator is a structure  $\langle W, N, *, \nu, R \rangle$ , where  $W$  is a set of worlds,  $N$  is a subclass of normal worlds (the complement being  $NN$ ),  $*$  is the Routley operation (such that  $w = w^{**}$ ), and  $\nu$  assigns truth values to all propositional parameters at worlds. In the logic *H*, it also assigns values to conditionals at non-normal worlds. In stronger logics, the ternary relation  $R$  is present, and is used to specify the values of conditionals at non-normal worlds. When no constraints are placed on  $R$ , we have the logic *B*.

The simplest way of extending such semantics to those of a quantified language is by removing  $\nu$  from the structure and adding a domain of quantification,  $D$ , and a denotation function  $d$ .  $d$  specifies a denotation for each constant (same at each world) and an extension for each  $n$ -place predicate at each world,  $d_w(P) \subseteq D^n$ . Truth conditions are given in the standard way. In particular, for the quantifiers:

$$\begin{aligned} \nu(\forall x \alpha) &= 1 \text{ iff for every } c \in D, \nu(\alpha(x/c)) = 1 \\ \nu(\exists x \alpha) &= 1 \text{ iff for some } c \in D, \nu(\alpha(x/c)) = 1 \end{aligned}$$

An inference is valid iff it is truth-preserving in all *normal* worlds of all interpretations.<sup>83</sup>

---

<sup>83</sup>More complex semantics can be employed in the usual variety of ways employed in modal logic. (See the article on Quantified Modal Logic in this *Handbook*.) In particular, we might employ variable-domain semantics. This makes matters more complex.

Consider the following quantifier axioms and rules (where  $\vdash$  is now taken to indicate universal closure):

$$\begin{aligned} &\vdash \forall x\alpha \rightarrow \alpha(x/c) \\ &\vdash \alpha(x/c) \rightarrow \exists x\alpha \\ &\vdash \alpha \wedge \exists x\beta \rightarrow \exists x(\alpha \wedge \beta) \quad x \text{ not free in } \alpha \\ &\vdash \forall x(\alpha \vee \beta) \rightarrow (\alpha \vee \forall x\beta) \quad x \text{ not free in } \beta \end{aligned}$$

If  $\vdash \forall x(\alpha \rightarrow \beta)$  then  $\vdash \exists x\alpha \rightarrow \beta$   $x$  not free in  $\beta$

If  $\vdash \forall x(\alpha \rightarrow \beta)$  then  $\vdash \alpha \rightarrow \forall x\beta$   $x$  not free in  $\alpha$

It is easy to check that these axioms/rules are valid/truth-preserving for  $H$ . If they are added to the propositional axioms/rules for  $H$ , they are also complete. For the proof, see Routley and Loparić [1980].<sup>84</sup>

If we strengthen the two rules to conditionals (so that the first of these becomes  $\vdash \forall x(\alpha \rightarrow \beta) \rightarrow (\exists x\alpha \rightarrow \beta)$ , etc.) and add them to the rules for  $B$ , they are also sound and complete. The same is true for a number of the extensions of  $B$ , including  $BX$ . (For details, see Routley [1980a].) A notable exception to this fact is the system  $R$ . Though the system is sound, it is, perhaps surprisingly, not complete.<sup>85</sup> In fact, a proof-theoretic characterisation of constant domain quantified  $R$  is still an open problem. The axioms and rules are complete for the stronger semi-relevant system  $RM$  of 5.4.<sup>86</sup>

Since every relevant logic in the above family contains  $FDE$ , and this is undecidable, it follows that all the logics in this family are also undecidable.

## 6.6 Algebraic Logics

Given any algebraic logic, for which the appropriate algebraic structures are lattices, and in which conjunction and disjunction behave as lattice meet and join, there is, as with many-valued logics, a natural way to extend the machinery to quantifiers. An algebra is *complete* iff it is closed under least upper bounds ( $\bigvee$ ) and greatest lower bounds ( $\bigwedge$ ), i.e., if the domain of the algebra is  $A$  and  $B \subseteq A$  then  $\bigvee B \in A$  and  $\bigwedge B \in A$ . If  $\mathcal{A}$  is any algebraic structure of the required kind, with domain  $A$ , then an interpretation is a triple  $\langle \mathcal{A}, D, d \rangle$ , where  $D$  is the domain of quantification,  $d$  maps every constant into  $D$  and every  $n$ -place predicate into a function from  $D^n$  into

(The philosophical gain, however, is dubious: world relativised quantifiers can always be defined in constant-domain semantics, provided we have an Existence predicate.)

<sup>84</sup>If one works with a free-variable notion of deducibility, as Routley and Loparić do, one also has to add the rule of universal generalisation: if  $\vdash \alpha$  then  $\vdash \forall x\alpha$ .

<sup>85</sup>As Fine showed. Fine also produced a rather different semantics with respect to which it is complete. See Anderson *et al.* [1992], sects. 52 and 53.

<sup>86</sup>See Anderson *et al.* [1992], sect. 49.2.



A. Algebraic values are then assigned to all formulas in the usual way. In particular, for quantified sentences the conditions are:

$$\begin{aligned}\nu(\forall x\alpha) &= \bigwedge\{\nu(\alpha(x/c)); c \in D\} \\ \nu(\exists x\alpha) &= \bigvee\{\nu(\alpha(x/c)); c \in D\}\end{aligned}$$

I will comment on this construction for only two kinds of algebras. The first is when  $\mathcal{A}$  is a De Morgan groupoid, or strengthening thereof. In this case, the above semantics clearly give quantified relevant logics. Their relation to quantified relevant logics based on the intensional semantics has not, as far as I am aware, been investigated.

The second is where  $\mathcal{A}$  is a dual intuitionist algebra. In this case, the semantics give a quantified logic that is dual to quantified intuitionist logic. For details, see Goodman [1981].<sup>87</sup>

### 6.7 A Brief Look Back

Now that we have surveyed a large number of paraconsistent logics up to a quantified level—some very briefly—it would seem appropriate to look back for a moment and put the systems into some sort of perspective.

The logics we have looked at fall roughly and inexactly into four categories: non-transitive logics, non-adjunctive logics, non-truth-functional logics and relevant logics. (The most interesting many-valued systems are zero degree relevant logic, *FDE*, or closely related to it, like *LP*, and so may be classed in this family.) The non-transitive logics seem to be good for extracting the essential juice out of classical inferences, but do not really take inconsistent semantic structures seriously. Non-adjunctive logics may be just what one needs for certain applications (e.g., inferences in a data base, where one would not necessarily want to infer  $\alpha \wedge \neg\alpha$  from  $\alpha$  and  $\neg\alpha$ ); they also take inconsistent structure seriously, though conjoined contradictions are handled indiscriminately, which makes them unsuitable for many applications. Non-truth-functional logics contain the whole of classical (or at least intuitionist) positive logic, and so are useful when strong canons of positive reasoning are required. However, this very strength is a weakness when it comes to some important applications, as we shall see in connection with set theory. Undoubtedly the simplest and most robust paraconsistent logic is the logic *LP*. When conditional operators are required, the relevant logic *BX* is a good all-purpose paraconsistent logic. Its conditional operator is satisfactory for many purposes, but may be considered relatively weak. It may be strengthened to give stronger relevant logics; but this, too, may cause a problem for some applications, as we shall see.

---

<sup>87</sup>There is also a topos-theoretic account of quantification for dual intuitionistic logic. See Mortensen [1995], ch. 11.

## 7 OTHER EXTENSIONS OF THE BASIC APPARATUS

I now want to look at other extensions of the basic paraconsistent apparatus. One way or another, all the paraconsistent logics we have looked at can be extended appropriately. However, it is tedious to run through every case, especially when details are often obvious. Hence, I shall illustrate the extensions mainly with respect to just one logic. Since *LP* is simple and natural, it recommends itself for this purpose. I will comment on other logics occasionally, when there is a point to doing so.

### 7.1 Identity and Function Symbols

*LP*—and all the other logics with objectual semantics that we have looked at—can be extended to include function symbols and identity in the usual way. The denotation function,  $d$ , maps each  $n$ -place function symbol,  $f$ , to an  $n$ -place function on the domain. A denotation for every (closed) term,  $t$ , is then obtained by the usual recursive condition:

$$d(ft_1\dots t_n) = d(f)(d(t_1), \dots, d(t_n))$$

With functional terms present, the quantifier rules of proof are extended to arbitrary (closed) terms in the usual way.

If we require the extension of the identity predicate to be  $\{\langle x, x \rangle ; x \in D\}$  then this is sufficient to validate the usual laws of identity:

$$\frac{}{t = t} \quad \frac{t_1 = t_2 \quad \alpha(x/t_1)}{\alpha(x/t_2)}$$

This does not require identity statements to be consistent. In *LP* the anti-extension of identity is any set whose union with the extension exhausts  $D^2$ , and so a pair can be in both the extension and the anti-extension of the identity predicate. In other logics, negated identities can be taken care of by whatever mechanism is used for negation. The completeness proof for quantified *LP* can be extended to include function symbols and identity in the usual Henkin fashion.

I note that description operators can be added in the obvious ways, with the same panoply of options as in the classical case.<sup>88</sup>

### 7.2 Second-order Logic

Paraconsistent logics can also be extended to second order in the obvious ways. Consider *LP*. We add (monadic) second order variables,  $X, Y, \dots$  to

<sup>88</sup>See the article of Free Logics in this *Handbook*.

the first-order language. Then, given an interpretation,  $\langle D, d \rangle$ , we extend the language to one,  $L_D$ , such that every member of the domain has a name, and for every pair  $E, A$  such that  $E \cup A = D$  there is a predicate,  $P$ , with  $E$  and  $A$  as extension and anti-extension, respectively.<sup>89</sup> The truth/falsity conditions for the second order universal quantifier are then:

$$\begin{aligned} 1 \in \nu(\forall X\alpha) & \text{ iff for every } P \text{ in } L_D, 1 \in \nu(\alpha(X/P)) \\ 0 \in \nu(\forall X\alpha) & \text{ iff for some } P \text{ in } L_D, 0 \in \nu(\alpha(X/P)) \end{aligned}$$

The truth/falsity conditions for the existential quantifier are the dual ones.

Appropriate monotonicity carries over to second order  $LP$ . Recall from 6.4 that if  $\mathcal{I}_1$  and  $\mathcal{I}_2$  are any interpretations, with truth value assignments  $\nu_1$  and  $\nu_2$ ,  $\mathcal{I}_1 \leq \mathcal{I}_2$  means that  $\mathcal{I}_1$  and  $\mathcal{I}_2$  have the same domain, and for every predicate,  $P$ , the extension (anti-extension) of  $P$  in  $\mathcal{I}_1$  is a subset of the extension (anti-extension) of  $P$  in  $\mathcal{I}_2$ . The same sort of induction as in the first-order case shows that if  $\mathcal{I}_1 \leq \mathcal{I}_2$  then for all formulas,  $\alpha$ , in  $L_D$ ,  $\nu_1(\alpha) \subseteq \nu_2(\alpha)$ . (The predicates added in forming  $L_D$  have the same extension/anti-extension in both interpretations; and thus atomic sentences containing them satisfy the condition.)

In the second order case, and unlike the first order case, the logical truths of  $LP$  are distinct from their classical counterparts. For example, as is easy to check, in  $LP$ ,  $\models \exists X(Xa \wedge \neg Xa)$  (just consider the predicate which has  $D$  as both extension and anti-extension).<sup>90</sup> In fact, the logical truths of second order  $LP$  are inconsistent, since it is also a logical truth that  $\forall X(Xa \vee \neg Xa)$ , which is equivalent by quantifier rules and De Morgan to  $\neg \exists X(Xa \wedge \neg Xa)$ .

### 7.3 Modal Operators

All the logics may have modal operators added to them in one way or another. In the case of discursive logics, indeed, the semantics already provide for the possibility of alethic modal operators.

Adding modal operators to intensional logics where negation is handled by the Routley  $*$  operator is very natural, but suffers problems similar to those we witnessed at the start of 5.5 in connection with the conditional. Suppose we take an intensional interpretation and give the modal operators the natural  $S5$  conditions:

$$\nu_w(\Box\alpha) = 1 \text{ iff for every } w' \in W, \nu_{w'}(\alpha) = 1$$

<sup>89</sup>This is the natural policy, since properties are characterised semantically by an extension/anti-extension pair. As in the classical case, there are other policies, e.g., where only predicates corresponding to some restricted class of properties are added.

<sup>90</sup>Second order  $FDE$  is constructed in the obvious way. The same sentence is a logical truth of this, showing that, unlike the first order case, it has logical truths.

$$\nu_w(\diamond\alpha) = 1 \text{ iff for some } w' \in W, \nu_{w'}(\alpha) = 1$$

(or even  $N$  instead of  $W$ ). Then the truth values of modalised statements are the same at all worlds. Hence,  $\nu_w(\Box\alpha) = 1 \Leftrightarrow \nu_{w^*}(\Box\alpha) = 1 \Leftrightarrow \nu_w(\neg\Box\alpha) = 0$ . Hence  $\Box\alpha, \neg\Box\alpha \models \beta$ , and so the logic is not suitable for serious paraconsistent purposes. The problem does not arise if we attempt a modal logic weaker than  $S5$ , for then the truth conditions of modal operators are given employing a binary accessibility relation in the usual way, and the truth values of modal statements will vary across worlds. But, at least for some purposes, an  $S5$  modality is desirable.

These problems are avoided if we use the Dunn semantics for negation. The values of modalised formulas will still be the same at all worlds (in the  $S5$  case), but we may now have both  $\Box\alpha$  and  $\neg\Box\alpha$  true at a world. I will illustrate, again, with respect to  $LP$ . Let us start with the case where the binary accessibility relation is arbitrary, the three-valued analogue of the modal system  $K$ .

An interpretation is now a structure  $\langle W, R, \nu \rangle$ , where  $W$  is a set of worlds;  $R$  is a binary relation on  $W$ ; and for each parameter,  $p$ ,  $\nu_w(p) \in \{\{1\}, \{1, 0\}, \{0\}\}$ . Truth/falsity conditions for the propositional connectives are as in 5.5. The conditions for  $\Box$  are:

$$\begin{aligned} 1 \in \nu_w(\Box\alpha) & \text{ iff for every } w' \text{ such that } wRw', 1 \in \nu_{w'}(\alpha) \\ 0 \in \nu_w(\Box\alpha) & \text{ iff for some } w' \text{ such that } wRw', 0 \in \nu_{w'}(\alpha) \end{aligned}$$

and dually for  $\diamond$ .<sup>91</sup>

It is easy to check that at every world of an interpretation  $\Box\neg\alpha$  has the same truth value as  $\neg\diamond\alpha$ , and dually. In fact, we can simply define  $\diamond\alpha$  as  $\neg\Box\neg\alpha$ , and will do this in what follows.

To obtain a proof-theoretic characterisation for the logic, we add to the rules for  $LP$  the following (chosen to make the completeness proof simple):

$$\frac{\begin{array}{c} \overline{\gamma} \\ \vdots \\ \beta \end{array} \quad \Box\gamma}{\Box\beta} \quad \frac{\begin{array}{c} \overline{\beta} \\ \vdots \\ \delta \end{array} \quad \diamond\beta}{\diamond\delta}$$

where there are no other undischarged assumptions in the sub-proofs.

$$\frac{\frac{\Box(\beta \vee \delta)}{\Box\beta \vee \diamond\delta} \quad \frac{\Box\gamma \wedge \diamond\beta}{\diamond(\gamma \wedge \beta)}}{\Box\beta \vee \diamond\delta \quad \diamond(\gamma \wedge \beta)}$$

<sup>91</sup> If a conditional operator is required, we may add a class of non-normal worlds—and maybe a ternary accessibility relation—and proceed as in 5.5.

$$\frac{\Box\gamma_1 \wedge \dots \wedge \Box\gamma_n}{\Box(\gamma_1 \wedge \dots \wedge \gamma_n)} \quad \frac{\Diamond(\delta_1 \vee \dots \vee \delta_n)}{\Diamond\delta_1 \vee \dots \vee \Diamond\delta_n}$$

Soundness is easily checked. For completeness, suppose that  $\Sigma \not\vdash \alpha$ . Extend  $\Sigma$  to a prime, deductively closed theory,  $\Pi$ , with the same property, as in 4.3. Define an interpretation,  $\langle W, R, \nu \rangle$ , where  $W$  is the set of prime deductively closed theories;  $\Gamma R \Delta$  iff for all  $\beta$ :

$$\begin{aligned} \Box\beta \in \Gamma &\Rightarrow \beta \in \Delta \\ \beta \in \Delta &\Rightarrow \Diamond\beta \in \Gamma \end{aligned}$$

and  $\nu$  is defined by:

$$\begin{aligned} 1 \in \nu_\Gamma(p) &\text{ iff } p \in \Gamma \\ 0 \in \nu_\Gamma(p) &\text{ iff } \neg p \in \Gamma \end{aligned}$$

All that remains is to show that these conditions extend to all formulas. Completeness then follows as usual. This is established by induction. The only difficult case is that for  $\Box$ , which requires the following two-part lemma.

If  $\Box\beta \notin \Gamma$  then there is a  $\Delta \in W$  such that  $\Gamma R \Delta$  and  $\beta \notin \Delta$ . Proof: Let  $\Gamma_\Box = \{\gamma; \Box\gamma \in \Gamma\}$  and  $\Gamma_\Diamond = \{\delta; \Diamond\delta \notin \Gamma\}$ . Then  $\Gamma_\Box \not\vdash \beta, \Gamma_\Diamond$ , by the first and second pair of rules, and a bit of fiddling with the third. Extend  $\Gamma_\Box$  to a prime, deductively closed set,  $\Delta$ , with the same property, as in 4.3. The result follows.

If  $\Diamond\beta \in \Gamma$  then there is a  $\Delta \in W$  such that  $\Gamma R \Delta$  and  $\beta \in \Delta$ . Proof: Let  $\Gamma_\Box$  and  $\Gamma_\Diamond$  be as before. Then  $\Gamma_\Box, \beta \not\vdash \Gamma_\Diamond$ , by the first and second pair of rules, and a bit of fiddling with the third. Extend  $\Gamma_\Box, \beta$  to a prime, deductively closed set,  $\Delta$ , with the same property, as in 4.3. The result follows.

We can now prove the induction step for  $\Box$ :

$$\begin{aligned} \Box\beta \in \Gamma &\Leftrightarrow \forall \Delta \text{ s.t. } \Gamma R \Delta, \beta \in \Delta && \text{lemma in one direction} \\ &\Leftrightarrow \forall \Delta \text{ s.t. } \Gamma R \Delta, 1 \in \nu_\Delta(\beta) && \text{definition of } R \text{ in the other} \\ &\Leftrightarrow 1 \in \nu_\Gamma(\Box\beta) && \text{induction hypothesis} \\ \\ \neg\Box\beta \in \Gamma &\Leftrightarrow \Diamond\neg\beta \in \Gamma && \text{definition of } \Diamond \\ &\Leftrightarrow \exists \Delta(\Gamma R \Delta \text{ and } \neg\beta \in \Delta) && \text{lemma in one direction} \\ &\Leftrightarrow \exists \Delta(\Gamma R \Delta \text{ and } 0 \in \nu_\Delta(\beta)) && \text{definition of } R \text{ in the other} \\ &\Leftrightarrow 0 \in \nu_\Gamma(\Box\beta) && \text{induction hypothesis} \end{aligned}$$

Stronger modal logics can be obtained by placing conditions on  $R$ , and corresponding conditions on the proof theory. But even if we make  $R$  universal (so that for all  $x$  and  $y$ ,  $xRy$ ), and obtain the analogue of  $S5$ , we still do not get  $\Box\alpha, \neg\Box\alpha \models \beta$ . To see this, merely consider the interpretation with one world,  $w$ , which accesses itself; and where  $\nu_w(p) = \{1, 0\}$  and  $\nu_w(q) = \{0\}$ . It is easy to check that  $\nu_w(\Box p) = \nu_w(\neg\Box p) = \{1, 0\}$ . Hence,  $\Box p, \neg\Box p \not\models q$ .

The same treatment can be given to temporal operators. If we take these, as usual, to be  $F$  and  $G$  for the future, and  $P$  and  $H$  for the past, then (three-valued) tense logic gives  $F$  and  $G$  the same truth conditions as  $\Diamond$  and  $\Box$ , respectively; and  $P$  and  $H$  are the same, except that  $R$  is replaced by its converse,  $\check{R}$  (where  $x\check{R}y$  iff  $yRx$ ). Appropriate soundness and completeness proofs for the case where  $R$  is arbitrary are obtained by modifying the alethic modal argument,<sup>92</sup> and stronger tense logics are obtained by adding conditions on  $R$ , in the usual way.<sup>93</sup>

Let me also mention conditional operators,  $>$ , of the Lewis/Stalnaker variety. These are modal (binary) operators, and can be given  $LP$  (or  $FDE$ ) semantics in the same way that they are given a more usual semantics. For example, for the Stalnaker version, one extends interpretations with a selection function,  $f(w, \alpha)$ , thought of as selecting the nearest world to  $w$  where  $\alpha$  is true.  $\alpha > \beta$  is then true at  $w$  iff  $\beta$  is true at  $f(w, \alpha)$ . Details are left as a very non-trivial exercise.<sup>94</sup>

#### 7.4 The Paraconsistent Importance of Modal Operators

Let me digress from the technical details to say a little about why modal operators are important in the context of paraconsistency. The reason is simply that so many of the natural areas where one might want to apply a paraconsistent logic involve them.

Take alethic modalities first. Even though one might not think that there are any true contradictions, one might still take them to be possible, in the sense of holding in some situations, such as fictional or counterfactual ones. Thus, one might hold that for some  $p$ ,  $\Diamond(p \wedge \neg p)$ . This has a simple and obvious model in the above semantics. In this context, let me mention again the importance for counterfactual conditionals of worlds where the impossible holds; “impossible worlds” are just what one needs to evaluate such conditionals, according to the Lewis/Stalnaker semantics in whose direction I have just gestured.

Some have been tempted not just by the view that some contradictions

---

<sup>92</sup>See Priest [1982].

<sup>93</sup>See the article on Tense Logic in this *Handbook*.

<sup>94</sup>For a paraconsistent theory of conditionals of this kind, and of many other modal operators, that employs the Routley  $*$  to handle negation, see Routley [1989].

are possible, but by the view that *everything* is possible.<sup>95</sup> The valuation  $\nu_{\{1,0\}}$  assigns every formula the value  $\{1,0\}$ . (See 4.6). Hence, any interpretation that contains  $\nu_{\{1,0\}}$  as one world will verify  $\diamond\alpha$ , for all  $\alpha$ , at any world that accesses it.<sup>96</sup> If we interpret the modal operators  $\Box$  and  $\diamond$  as the deontic operators  $O$  (it is obligatory that) and  $P$  (it is permissible that), respectively, then the thesis that everything is possible becomes the nihilistic thesis that everything is permissible—what, according to Dostoevski, would be the case if there is no God.

Less exotically, standard deontic logic suffers badly from explosion.<sup>97</sup> Since in classical logic  $\alpha, \neg\alpha \models \beta$  it follows that  $O\alpha, O\neg\alpha \models O\beta$ : if you have inconsistent obligations then you are obliged to do everything. This is surely absurd. People incur inconsistent obligations; this may give rise to legal or moral dilemmas, but hardly to legal or moral anarchy.<sup>98</sup> And one does not have to believe in dialetheism to accept this. Unsurprisingly, deontic explosion fails, given the semantics of the previous section: just consider the interpretation where there is a single world,  $w$ ;  $R$  is universal;  $\nu_w(p) = \{1,0\}$  and  $\nu_w(q) = \{0\}$ . It is not difficult to check that  $\nu_w(Op) = \nu_w(O\neg p) = \{1,0\}$ , whilst  $\nu_w(Oq) = \{0\}$ .

What is often taken to be the basic possible-worlds deontic logic (called *KD* by Chellas [1980], p. 131) makes matters even worse, by requiring that in an interpretation the accessibility relation be serial: for all  $x$ , there is a  $y$  such that  $xRy$ . This validates the inference  $O\alpha/P\alpha$ . It also validates the inference  $O\neg\alpha/\neg O\alpha$ . Hence we have, classically,  $O\alpha, O\neg\alpha \models O\alpha \wedge \neg O\alpha \models \beta$ ; one who incurs inconsistent obligations renders the world trivial. Someone who believes that there are deontic dilemmas may just have to jettison the view that obligation entails permission, and so give up seriality. But on the above account one can retain seriality, and so both the above inferences; for  $O\alpha \wedge \neg O\alpha \not\models \beta$ , as the countermodel of the last paragraph shows.<sup>99</sup>

Another standard way of interpreting the modal operator  $\Box$  is as an epistemic operator,  $K$  (it is known that), or a doxastic operator  $B$  (it is believed that). In these cases, classically, one would almost certainly want to put extra constraints on the accessibility relation, though what these should be might be contentious: all can accept reflexivity ( $xRx$ ) for  $K$  (but not for  $B$ ) since this validates  $K\alpha \models \alpha$ . Whether one would want transitivity ( $(xRy \& yRz) \Rightarrow xRz$ ) is much more dubious for  $B$  and  $K$ , since this gives the

<sup>95</sup>E.g., Mortensen [1989].

<sup>96</sup>A similar, but slightly more complex, construction can be employed to the same effect if the logic has a conditional operator.

<sup>97</sup>See the article on Deontic Logic in this *Handbook* for details of Deontic Logic, including the possible-worlds approach.

<sup>98</sup>For further discussion, see Priest [1987], ch. 13.

<sup>99</sup>We have just been dealing with some of the “paradoxes of deontic logic”. There are many of these. Arguably, all of them—or at least all the serious ones—are avoided by using a paraconsistent logic with a relevant conditional. See Routley and Routley [1989].

highly suspect  $K\alpha \models KK\alpha$  and  $B\alpha \models BB\alpha$ . All this applies equally to the semantics of the previous section. Moreover, the paraconsistent semantics solve problems for doxastic logic of the same kind as for deontic logic. It is clear that people sometimes have inconsistent beliefs (if not knowledge). Standard semantics give  $B\alpha, B\neg\alpha \models B\beta$ . Yet patently someone may have inconsistent beliefs without believing everything.<sup>100</sup>

Observations such as this are particularly apt in the branch of *AI* known as knowledge representation, where it is common to use epistemic operators to model the information available to a computer. (See, e.g., a number of the essays in Halpern [1986].) Such information may well be inconsistent.

Finally, to tense operators. Whilst one does not have to be a dialetheist to hold that inconsistencies may be believed, obligatory, or true in some counterfactual situation, one does have to be, to believe that they were or will be true. Such views have certainly been held, however. Following Zeno, the whole dialectical tradition holds that contradictions arise in a state of change. To see one of the more plausible examples of this, just consider a state described by  $p$  which changes instantaneously at time  $t_0$  to a state described by  $\neg p$ . What is the state of affairs at  $t_0$ ? One answer is that at  $t_0$ ,  $p \wedge \neg p$  is true. Indeed, the contradictory state *is* the state of change.<sup>101</sup>

This can be modeled by the paraconsistent interpretation  $\langle W, R, \nu \rangle$ , where  $W$  is the set of real numbers (thought of as times);  $R$  is the standard ordering on the reals; and  $\nu$  is defined by the condition:

$$\begin{aligned} \nu_t(p) &= \{1\} && \text{if } t < t_0 \\ &= \{1, 0\} && \text{if } t = t_0 \\ &= \{0\} && \text{if } t > t_0 \end{aligned}$$

It is easy to check that this interpretation verifies the inference:  $p \wedge F\neg p / (p \wedge \neg p) \vee F(p \wedge \neg p)$ , which we might call ‘Zeno’s Principle’: change implies contradiction.

### 7.5 Probability

Probability is not a modal notion. But it, too, has paraconsistent significance. One of the most natural ways of constructing a paraconsistent probability theory is to extract one from a class of paraconsistent interpretations, in the manner of Carnap.<sup>102</sup>

<sup>100</sup>If you believe classical logic, then you might suppose that they are *rationally committed* to everything, but that is quite different. Even here, however, an explosive logic would seem to go astray. Dialetheism aside, situations such as the paradox of the preface, as well as more mundane things, would seem to show that one can be rationally committed to inconsistent propositions without being rationally committed to everything. See Priest [1987], sect. 7.4.

<sup>101</sup>See Priest [1982] and Priest [1987], ch. 11.

<sup>102</sup>See Carnap [1950].



A probabilistic interpretation is a pair,  $\langle I, \mu \rangle$ , where  $I$  is a class of interpretations for  $LP^{103}$  and  $\mu$  a finitely additive measure on  $I$ , that is, a function from subsets of  $I$  to non-negative real numbers such that:

$$\begin{aligned} \mu(\phi) &= 0 \\ \mu(X \cup Y) &= \mu(X) + \mu(Y) \quad \text{if } X \cap Y = \phi \end{aligned}$$

If  $\alpha$  is any sentence, let  $[\alpha] = \{\nu \in I; 1 \in \nu(\alpha)\}$ . For reasons that we will come to, we also require that for all  $\alpha$ ,  $\mu([\alpha]) \neq 0$ . There certainly are such interpretations and measures. For example, let  $I$  be any finite class that contains the trivial interpretation,  $\nu_{\{1,0\}}$ , where all sentences are true, and let  $\mu(X)$  be the cardinality of  $X$ . Then this condition is satisfied.

Given a probabilistic interpretation, we define a probability function,  $p$ , by:

$$p(\alpha) = \mu([\alpha]) / \mu(I)$$

It is easy to see that  $p$  satisfies all the standard conditions for a probability function, such as:

$$\begin{aligned} 0 &\leq p(\alpha) \leq 1 \\ \text{if } \alpha \models \beta &\text{ then } p(\alpha) \leq p(\beta) \\ \text{if } \models \alpha &\text{ then } p(\alpha) = 1 \\ p(\alpha \vee \beta) &= p(\alpha) + p(\beta) - p(\alpha \wedge \beta) \end{aligned}$$

except, of course:  $p(\neg\alpha) + p(\alpha) = 1$ . Since we have  $p(\alpha \wedge \neg\alpha) > 0$ , and  $p(\alpha \vee \neg\alpha) = 1$ , it follows that  $p(\alpha) + p(\neg\alpha) > 1$ .

By the construction, we have, in fact,  $p(\alpha) > 0$  for all  $\alpha$ . It might be suggested that a person whose personal probability function gives nothing the value zero would have to be very stupid—or at least credulous. But since  $p(\alpha)$  may be as small as one wishes, this hardly seems to follow. Moreover, giving nothing a zero probability signals an open-minded and undogmatic policy of belief. Arguably, this is the most rational policy.

Given a probability function, conditional probability can be defined in the usual way:

$$p(\alpha/\beta) = p(\alpha \wedge \beta) / p(\beta)$$

A singular advantage of this paraconsistent probability theory over standard accounts is that conditional probability is *always* defined, since the denominator is always non-zero.

<sup>103</sup> Again, many other paraconsistent logics could be used instead.

Perhaps the major application of probability theory is in framing an account of non-deductive inference. How, exactly, to do this is a moot question. But however one does it, a paraconsistent account of non-deductive inference can be framed in the same way, employing paraconsistent probability theory. For example, we may define the degree of (non-deductive) validity of the inference  $\alpha/\beta$  to be  $p(\beta/(\alpha \wedge \eta))$ , where  $\eta$  is our background evidence. As one would expect, deductively valid inferences come out as having maximal degree of inductive validity.

To compute the degree of validity of an inference, so defined, we would often need to employ Bayes' Theorem. Let us look at the paraconsistent two-hypothesis version of this. Suppose that we have two hypotheses,  $h_1$  and  $h_2$ , that are exclusive and exhaustive, in the sense that  $\models h_1 \vee h_2$  and  $\models \neg(h_1 \wedge h_2)$ , and that we wish to compute the probability of  $h_1$  on evidence,  $e$ , given the inverse probabilities of these hypotheses on the evidence (all relative to some background evidence,  $\eta$ , which we will ignore).

Note first that  $p(h_1/e) = p(h_1 \wedge e)/p(e) = p(e/h_1).p(h_1)/p(e)$ . It remains to compute  $p(e)$ . Since  $h_1 \vee h_2$  entails  $e \vee h_1 \vee h_2$  we have :

$$\begin{aligned} 1 &= p(e \vee h_1 \vee h_2) &= p(e) + p(h_1 \vee h_2) - p(e \wedge (h_1 \vee h_2)) \\ & &= p(e) + 1 - p(e \wedge (h_1 \vee h_2)) \end{aligned}$$

Hence:

$$\begin{aligned} p(e) &= p(e \wedge (h_1 \vee h_2)) \\ &= p((e \wedge h_1) \vee (e \wedge h_2)) \\ &= p(e \wedge h_1) + p(e \wedge h_2) - p(e \wedge h_1 \wedge h_2) \\ &= p(h_1).p(e/h_1) + p(h_2).p(e/h_2) - p(h_1 \wedge h_2).p(e/(h_1 \wedge h_2)) \end{aligned}$$

Thus:

$$p(h_1/e) = \frac{p(e/h_1).p(h_1)}{p(h_1).p(e/h_1) + p(h_2).p(e/h_2) - p(h_1 \wedge h_2).p(e/(h_1 \wedge h_2))}$$

This is the paraconsistent version of Bayes' Theorem. In the classical case, the last term of the denominator is zero, since  $\models \neg(h_1 \wedge h_2)$ ; but this is not so in the paraconsistent case. The theorem illustrates a general fact about paraconsistent probability theory: everything works as normal, except that we have to carry round some extra terms concerning the probabilities of certain contradictions which may be neglected in the classical case.

The extra complication may actually be a gain in some contexts. Let me mention one possible one; this concerns quantum mechanics. Quantum mechanics is known to suffer from various phenomena often called 'causal anomalies', a famous one of which is the two-slit experiment.<sup>104</sup> In this, a

<sup>104</sup>See, e.g., Haack [1974], ch. 8.

light is shone onto a screen through a mask with two slits. The intensity of light on any point on the screen is proportional to the probability that a photon hits it,  $\sigma$ , given that it goes through one slit,  $\alpha$ , or goes through the other,  $\beta$ . Let us write  $p(\alpha \vee \beta)$  as  $q$ . Then:

$$\begin{aligned} p(\sigma/(\alpha \vee \beta)) &= p(\sigma \wedge (\alpha \vee \beta))/p(\alpha \vee \beta) \\ &= p((\sigma \wedge \alpha) \vee (\sigma \wedge \beta))/q \\ &= p(\sigma \wedge \alpha)/q + p(\sigma \wedge \beta)/q - p(\sigma \wedge \alpha \wedge \beta)/q \end{aligned}$$

Classically, we know that  $\neg(\alpha \wedge \beta)$ , and so the last term may be ignored. For similar reasons,  $q = p(\alpha \vee \beta) = p(\alpha) + p(\beta)$ , and by symmetry we can arrange for  $p(\alpha)$  and  $p(\beta)$  to be equal. Hence:

$$\begin{aligned} p(\sigma/(\alpha \vee \beta)) &= p(\sigma \wedge \alpha)/2p(\alpha) + p(\sigma \wedge \beta)/2p(\beta) \\ &= \frac{1}{2}(p(\sigma/\alpha) + p(\sigma/\beta)) \end{aligned}$$

Thus, the intensity of light on the screen should be the average of the intensities of light going each slit independently (which can be determined by closing off the other). Exactly this is what is *not* found.

Standard quantum logic<sup>105</sup> avoids the result by rejecting the inference of distribution (i.e., the equivalence between  $\sigma \wedge (\alpha \vee \beta)$  and  $(\sigma \wedge \alpha) \vee (\sigma \wedge \beta)$ ), and so faulting the second line of the above proof. A paraconsistent solution is just to note that we cannot ignore the third term in the computation of  $p(\sigma/(\alpha \vee \beta))$ , even though we know that  $\neg(\alpha \wedge \beta)$ . In qualitative terms, what this means is that the photon has a non-zero probability of doing the impossible, and going through both slits simultaneously!

This application of paraconsistent probability theory to quantum mechanics is *highly* speculative. Whether it could be employed to resolve the other causal anomalies of quantum theory, let alone to predict the observations that are actually made, has not been investigated.<sup>106</sup>

## 7.6 The Classical Recapture

Most paraconsistent logicians have supposed that reasoning in accordance with classical logic is sometimes legitimate. Most, for example, have taken it that classical logic is perfectly acceptable in consistent situations. They have therefore proposed ways in which classical logic can be “recaptured” from a paraconsistent perspective.

The simplest such recapture occurs in non-adjunctive logics. As we noted in 4.2, single premise non-adjunctive reasoning is classical. Hence, classical

<sup>105</sup>See the article on this in the *Handbook*.

<sup>106</sup>For more on the above issues, including the effects of paraconsistent probability theory on confirmation theory, see Priest [1987], sect. 7.6, and Priest *et al.* [1989], pp. 376-9, 385-8.

reasoning can be regained simply by conjoining all premises. A different strategy is to employ a consistency operator, as is done in the da Costa logics  $C_i$ , for finite non-zero  $i$ . As we saw in 5.3, this can be employed to define a negation which behaves classically; hence classical reasoning can simply be interpreted within the system. This approach has problems for some applications, as we shall see when we come to look at set theory.

Yet another way to recapture classical reasoning, provided that a conditional operator is available, is to employ an absurdity constant,  $\perp$ , satisfying the condition  $\models \perp \rightarrow \alpha$ , for all  $\alpha$ . Such a constant makes perfectly good sense paraconsistently. Algebraically, it corresponds to the minimal value of an algebra (which can usually be added if it is not present already). In truth-preservation terms, there are two ways of handling its semantics. One is to require that  $\perp$  be untrue at every (world of every) evaluation. Its characteristic principle then holds vacuously. The other way (which may be preferable if one objects to vacuous reasoning) is simply to assign  $\perp$  (at a world) the value of the (infinitary) conjunction of all other formulas (at that world). A bit of juggling then usually verifies the characteristic principle. (The definition itself guarantees it only when  $\alpha$  does not contain  $\perp$ .)

Now let  $C$  be the set of all formulas of the form  $(\alpha \wedge \neg\alpha) \rightarrow \perp$ . Then an inference is classically valid iff it is enthymematically valid with  $C$  as the set of suppressed premises, in most paraconsistent logics. For if every member of  $C$  holds at (a world of) an interpretation, then the (world of the) interpretation is a classical one—or at least the trivial one—and hence if the premises of a classically valid inference are true at it, so is the conclusion. Thus, we have an enthymematic recapture.

Let us write  $\neg\alpha$  for  $\alpha \rightarrow \perp$ . In classical (and intuitionist) logic,  $\neg\alpha$  just is  $\neg\alpha$ . It might therefore be thought that provided a logic possesses  $\perp$ , we could simply interpret classical logic in it by identifying  $\neg\alpha$  with  $\neg\alpha$ . This thought would be incorrect, though. In many paraconsistent logics,  $\neg\alpha$  behaves quite differently from classical (and intuitionist) negation. What properties it has depends, of course, on the properties of  $\rightarrow$ . While it will always be the case that  $\alpha, \neg\alpha \models \beta$ , it will certainly not be true in general that  $\models \alpha \vee \neg\alpha$ , that  $\neg\neg\alpha \models \alpha$ , or even that  $\alpha \models \neg\neg\alpha$ . As an example of the last, consider an intensional interpretation for the logic  $H$ . (See 5.5.) Suppose that  $p$  is true at some normal world,  $w$ , but that at some non-normal world  $p \rightarrow \perp$  is true (and  $\perp$  is not). Then  $(p \rightarrow \perp) \rightarrow \perp$  fails at  $w$ .<sup>107</sup>

A final, and much less brute-force, way of recapturing classical logic starts from the idea that consistency is the norm. It is implicit in the paraconsistent enterprise that inconsistency can be contained. Instead of spreading everywhere, inconsistencies can exist isolated, as do singularities in a field

---

<sup>107</sup>It might be thought that the existence of the explosive connective ‘ $\perp$ ’ would cause problems for certain paraconsistent applications; notably, for example, for set theory. This is not the case, however, as we will see.

(of the kind found in physics, not agriculture). This metaphor suggests that even if inconsistencies are present they will be relatively rare. If it is *true* inconsistencies we are talking about, these will be even rarer—something that the classical logician can readily agree with!<sup>108</sup>

This suggests that consistency should be a default assumption, in the sense of non-monotonic logic. Many non-monotonic logics can be formulated by defining validity over some class of models, minimal with respect to violation of the default condition. In effect, we consider only those interpretations that are no more profligate in the relevant way than the information necessitates. In the case where it is consistency that is the default condition, we may define validity over models that are minimally inconsistent in some sense. I will illustrate, as usual, with respect to  $LP$ .<sup>109</sup>

Let  $\mathcal{I} = \langle D, d \rangle$  be an  $LP$  interpretation. Let  $\alpha \in \mathcal{I}!$  iff  $\alpha$  is  $Pd_1\dots d_n$ , where  $P$  is an  $n$ -place predicate and  $\langle d_1, \dots, d_n \rangle \in E_P \cap A_P$  in  $\mathcal{I}$ . (Recall that I am using members of the domain as names for themselves.)  $\mathcal{I}!$  is a measure of the inconsistency of  $\mathcal{I}$ . In particular,  $\mathcal{I}$  is a classical interpretation iff  $\mathcal{I}! = \phi$ . If  $\mathcal{I}_1$  and  $\mathcal{I}_2$  are  $LP$  interpretations, I will write  $\mathcal{I}_1 < \mathcal{I}_2$ , and say that  $\mathcal{I}_1$  is *more consistent than*  $\mathcal{I}_2$ , iff  $\mathcal{I}_1! \subset \mathcal{I}_2!$ . (The containment here is proper.)  $\mathcal{I}$  is a *minimally inconsistent (mi) model* of  $\Sigma$  iff  $\mathcal{I}$  is a model of  $\Sigma$  iff  $\mathcal{I}$  is a model of  $\Sigma$  and if  $\mathcal{J} < \mathcal{I}$ ,  $\mathcal{J}$  is not a model of  $\Sigma$ . Finally,  $\alpha$  is an *mi consequence* of  $\Sigma$  ( $\Sigma \models_m \alpha$ ) iff every mi model of  $\Sigma$  is a model of  $\alpha$ .

As is to be expected,  $\models_m$  is non-monotonic. For if  $p$  and  $q$  are atomic sentences, it is easy to check that  $\{p, \neg p \vee q\} \models_m q$ , but  $\{\neg p, p, \neg p \vee q\} \not\models_m q$ . Moreover, since all classical models (if there are any) are mi models, and all mi models are models, it follows that  $\Sigma \models \alpha \Rightarrow \Sigma \models_m \alpha \Rightarrow \Sigma \models_C \alpha$ . The implications are, in general, not reversible. For the first, note that  $\{p, \neg p \vee q\} \not\models q$ ; for the second, note that  $\{p, \neg p\} \not\models_m q$ . But if  $\Sigma$  is classically consistent, its mi models are exactly its classical models, and hence we have  $\Sigma \models_m \alpha \Leftrightarrow \Sigma \models_C \alpha$ : classical recapture.

$\models_m$  has various other interesting properties. For example, it can be shown that if the  $LP$  consequences of some set is non-trivial, so are its mi consequences *Reassurance*. For details, see Priest [1991a].<sup>110</sup>

<sup>108</sup>Though this is not so obvious once one accepts dialetheism. For a defence of the view given dialetheism, see Priest [1987], sect. 8.4.

<sup>109</sup>Though the first paraconsistent logician to employ this strategy was Batens [1989], who employs a non-truth-functional logic. Batens also considers the dynamical aspects of such default reasoning.

<sup>110</sup>In that paper, in the definition of  $<$ , a clause stating that the domains of  $\mathcal{I}_\infty$  and  $\mathcal{I}_\epsilon$  are the same is added. With this clause, the result concerning classical recapture is false (and that paper is mistaken). For example, if  $\alpha$  is  $\exists x Px \wedge \exists x \neg Px$ , then  $\langle D, d \rangle$  is an mi model, where  $D = \{a\}$ ,  $E_P = A_P = \{a\}$ , though this is not a classical model. (This was first noted by Diderik Batens, in correspondence.) As  $<$  is defined here,  $\{\forall x(Px \wedge \neg Px)\} \models_m \exists x \forall yx = y$ , which may be thought to be counter-intuitive. But if  $\forall x(Px \wedge \neg Px)$  is *all* the information we have, and inconsistencies are to be minimised, perhaps it is correct to infer that there is just one thing. Note that  $\{\forall x(Px \wedge \neg Px), \exists x Qx \wedge \exists x \neg Qx\} \not\models_m \exists x \forall yx = y$ . For  $\langle d, d \rangle$  is an mi model of the premises, where  $D = \{a, b\}$ ,  $E_P = A_P =$

## 8 SEMANTICS AND SET THEORY

The previous part gestured in the direction of various applications of paraconsistent logic. I want, in the next two parts, to look at some other applications in greater detail. These concern theories of particular mathematical significance. In this part I will deal with semantics and set theory.

Semantic and set-theoretic notions appear to be governed by simple and apparently obvious principles. In semantics, these concern truth,  $T$ , satisfaction,  $S$ , and denotation,  $D$ , and are:

$$\begin{aligned} T\text{-schema: } & T \langle \alpha \rangle \leftrightarrow \alpha \\ S\text{-schema: } & Sx \langle \beta \rangle \leftrightarrow \beta(y/x) \\ D\text{-schema: } & D \langle t \rangle x \leftrightarrow x = t \end{aligned}$$

where  $\alpha$  is any sentence,  $\beta$  is any formula with one free variable,  $y$ , and  $t$  is any closed term. Angle brackets indicate a name-forming device. In set theory the principle is the schema of set existence:

$$\text{Comprehension Schema: } \exists x \forall y (y \in x \leftrightarrow \beta)$$

where  $\beta$  is *any* formula not containing  $x$ . What the connective  $\leftrightarrow$  is in the above schemas, we will have to come back to.

Despite the fact that these schemas appear to be obvious, they all give rise to contradictions, as is well known: the paradoxes of self-reference, such as (respectively) the Liar Paradox, the Heterological Paradox, Berry's Paradox and Russell's Paradox. The usual approaches to set theory and semantics restrict the principles in some way. Such approaches are all unsatisfactory in one way or another, though I shall not discuss this here.<sup>111</sup>

A paraconsistent approach can simply leave the principles as they are, and allow the contradictions to arise. They need do no damage, because the logic is not explosive. Even so, not all paraconsistent logics are suitable as the underlying logics of these theories. For a start, if the above schemas are formulated with the material  $\equiv$  they give rise to a conjoined contradiction, so using a non-adjunctive logic (except Rescher and Manor's) explodes the theory.<sup>112</sup> And in the da Costa systems,  $C_i$ , for finite  $i$ , an operator behaving like classical negation,  $\neg^*$  can be defined (see 5.3). The usual arguments establish contradictions of the form  $\alpha \wedge \neg^* \alpha$ , and so again

---

$\{a, b\}, E_Q = \{a\}, A_Q = \{b\}$ . With the present definition, the proof of Reassurance for the first-order case, appropriately modified, still goes through.

<sup>111</sup>See, for example, Priest [1987], chs. 1, 2.

<sup>112</sup>Rescher and Brandom, [1980], p. 164, suggest splitting the biconditionals up into two non-conjoined conditionals.

the theories explode. Fortunately, there are other paraconsistent logics that will do the job.<sup>113</sup>

### 8.1 Truth Theory in LP

Let us start with the semantic case. I will deal with truth; similar remarks and constructions hold for the other semantic notions, but I will leave readers to ponder these for themselves. The first question we need to address is what connective it is that occurs in the biconditional of the  $T$ -schema. The first possibility is that it is a material biconditional,  $\equiv$ .<sup>114</sup>

Let us, then, suppose that we are dealing with the logic  $LP$ . We will need some machinery to handle self reference; a straightforward option is to let this be arithmetic. Hence, we suppose the language,  $L$ , to be that of first order arithmetic augmented by a one place predicate,  $T$ . To make things easy, we will assume that  $L$  has a function symbol for each primitive recursive function (and only those function symbols). Let  $T_0$  be the  $LP$  theory in this language which comprises the truths of first order arithmetic plus the  $T$ -schema.

The assumption that  $T_0$  contains all of arithmetic is obviously a very strong one, and means that the theory is not axiomatic. We could, instead, consider an axiomatic theory with some suitable fragment of arithmetic, but since a major part of our concern will be with what *cannot* be proved, it is useful to have the arithmetic part as strong as possible.

The first thing to note is that  $T_0$  is inconsistent. Given the resources of arithmetic, for any formula,  $\alpha$ , of one free variable,  $x$ , one can find, by the usual Gödel construction, a fixed-point formula,  $\beta$ , of the form  $\alpha(x/\langle\beta\rangle)$ .<sup>115</sup> Now, let  $\alpha$  be  $\neg Tx$  and let  $\beta$  be its fixed point. Then the  $T$ -schema gives us:  $T\langle\beta\rangle \equiv \beta$ , i.e.,  $T\langle\beta\rangle \equiv \neg T\langle\beta\rangle$ . Unpacking the definition of  $\equiv$ , in terms of  $\wedge$ ,  $\vee$ , and  $\neg$  and fiddling, gives exactly  $T\langle\beta\rangle \wedge \neg T\langle\beta\rangle$ .<sup>116</sup>

Despite being inconsistent,  $T_0$  is non-trivial. An easy way to see this is to observe, first of all, that if in any interpretation  $\nu(\alpha) = \{1, 0\}$  then  $\nu(\alpha \equiv \beta) = \{1, 0\}$ . Hence, an  $LP$  model for  $T_0$  can be obtained by letting the denotations of the arithmetic language be that of the standard interpretation of arithmetic—so that, in particular, the domain is  $N$ , the natural numbers; recall that classical interpretations are just special cases of  $LP$

<sup>113</sup>There are paraconsistent set theories based on da Costa's  $C$  systems. (See, e.g., Arruda [1980], da Costa [1986].) In these theories, the schemas have to be constrained, as they are classically. This takes away much of the appeal of a paraconsistent approach.

<sup>114</sup>It is natural to suppose that it ought to be a detachable conditional. Goodship [1996] argues that it is only a material conditional. Whether or not this is the case, it is certainly interesting to explore the two possibilities.

<sup>115</sup>See, e.g., Priest [1987], sect. 3.5.

<sup>116</sup>It is worth noting that for the  $S$ -schema, the fixed point machinery is unnecessary for the demonstration of inconsistency. For let  $\alpha$  be  $\neg Sxx$ . Then an instance of the  $S$ -schema is:  $S\langle\alpha\rangle\langle\alpha\rangle \equiv \neg S\langle\alpha\rangle\langle\alpha\rangle$ , and we can then proceed as before.

interpretations—and setting  $E_T$  and  $A_T$ , the extension and anti-extension of  $T$ , both to  $N$ . Call this interpretation  $\mathcal{I}_0$ . In  $\mathcal{I}_0$  every sentence of the form  $T \langle \alpha \rangle$  takes the value  $\{1, 0\}$ , and so by the observation concerning  $\equiv$ ,  $\mathcal{I}_0$  is a model for the  $T$ -schema, and so of all of  $T_0$ . The same interpretation shows that if  $\alpha$  is any arithmetic formula false in the standard model,  $T_0 \not\models \alpha$ .

$T_0$  is a relatively weak theory. In particular, it does not legitimate the two way rule of inference:

$$\frac{\alpha}{T \langle \alpha \rangle}$$

(just consider the south-north inference in  $\mathcal{I}_0$ , where  $\alpha$  is an arithmetic sentence false in the standard model).<sup>117</sup>

Let the theory obtained by replacing the material  $T$ -schema of  $T_0$  with this rule be called  $T_1$ .  $T_1$  is inconsistent. For choose an  $\alpha$  of the form  $\neg T \langle \alpha \rangle$ . The law of excluded middle gives  $T \langle \alpha \rangle \vee \neg T \langle \alpha \rangle$ , i.e.,  $T \langle \alpha \rangle \vee \alpha$ , which, applying the rule, gives  $T \langle \alpha \rangle$  and  $\alpha$ , i.e.,  $\neg T \langle \alpha \rangle$ .

We can construct a model for  $T_1$  as follows. If an interpretation assigns the standard denotations to all arithmetical language let us call it *arithmetical*. Any arithmetical interpretation is a model all of  $T_1$  except, perhaps, the  $T$ -schema. Let  $\mathcal{I}_1$  and  $\mathcal{I}_2$  be two arithmetical interpretations, with assignment functions  $\nu_1$  and  $\nu_2$ . Define  $\nu_1 \preceq \nu_2$  to mean that for all atomic sentences in the language,  $\alpha$ :

$$\begin{aligned} \nu_1(\alpha) = t &\Rightarrow \nu_2(\alpha) = t \\ \nu_1(\alpha) = f &\Rightarrow \nu_2(\alpha) = f \end{aligned}$$

If  $\nu_1 \preceq \nu_2$  then this condition extends to all formulas of  $L$ . For suppose that  $\nu_1 \preceq \nu_2$ . If  $n$  is in the extension of  $T$  in  $\mathcal{I}_2$  but not  $\mathcal{I}_1$ ; then  $\nu_2(Tn) = t$  or  $b$ , but  $\nu_1(Tn) = f$ , violating the condition. Similarly for anti-extensions. Hence,  $\mathcal{I}_2 \preceq \mathcal{I}_1$ . By monotonicity, for all  $\alpha$ ,  $\nu_2(\alpha) \subseteq \nu_1(\alpha)$ . The conclusion follows. For suppose that  $\nu_2(\alpha) \neq t$ . Then  $\alpha$  is false (i.e.,  $b$  or  $f$ ) in  $\mathcal{I}_2$ ; hence  $\alpha$  is false in  $\mathcal{I}_1$ , i.e.,  $\nu_1(\alpha) \neq t$ . The argument for  $f$  is similar.

This result is, in fact, just another version of monotonicity; I will call it the *Monotonicity Lemma*.

Let  $\mathcal{I}_0$  be any arithmetical interpretation, with evaluation function  $\nu_0$ . We now define a transfinite sequence of arithmetical interpretations,

<sup>117</sup>Whether or not more follows with minimally inconsistent *LP* (see 7.6) is presently unknown. Another non-monotonic notion of inference also suggests itself here. According to this, the things that follow are the things that hold in all minimally inconsistent models where the arithmetic part is the standard model. Employing this would be appropriate if there were good reasons to believe that the only inconsistencies involve the truth predicate.



$\langle \mathcal{J}_i; i \in On \rangle$  ( $On$  is the class of ordinals). I will make the construction slightly more complex than necessary, for the benefits of the next section. It suffices to define the evaluation function  $\nu_i$  of each interpretation. If  $i > 0$  and  $n$  is not the code of a sentence, then  $\nu_i(Tn) = \nu_0(Tn)$ . We therefore need to consider only atomic formulas of the form  $T \langle \alpha \rangle$ . Let us say that  $\alpha$  is *eventually  $t$  by  $k$*  iff  $\exists i > 0 \forall j (i \leq j < k, \nu_j(\alpha) = t)$ . Similarly for  $f$ . Then for  $k \neq 0$ :

$$\begin{aligned} \nu_k(T \langle \alpha \rangle) &= t && \text{if } \alpha \text{ is eventually } t \text{ by } k \\ &= f && \text{if } \alpha \text{ is eventually } f \text{ by } k \\ &= b && \text{otherwise} \end{aligned}$$

We can now establish that if  $0 < i \leq k$  then  $\nu_i \preceq \nu_k$ . The proof is by transfinite induction. Suppose that the result holds for all  $j < k$ . We show it for  $k$ . Since the truth values of atomic formulas other than ones of the form  $T \langle \alpha \rangle$  are constant, we need consider only these. So suppose that  $\nu_i(T \langle \alpha \rangle) = t$ . Then  $\alpha$  is eventually  $t$  by  $i$ . In particular, for some  $0 < j < i$ ,  $\nu_j(\alpha) = t$ . By induction hypothesis, for all  $l$  such that  $j < l < k$ ,  $\nu_j \preceq \nu_l$ . Hence, by monotonicity  $\nu_l(\alpha) = t$ . Hence,  $\alpha$  is eventually  $t$  by  $k$ , i.e.,  $\nu_k(T \langle \alpha \rangle) = t$ . The case for  $f$  is similar.

What this lemma shows is that once  $i > 0$ , and increases, sentences of the form  $T \langle \alpha \rangle$  can change their truth value at most once. If they ever attain a classical value, they keep it. Since there is only a countable number of sentences of this form, there must be an ordinal,  $l$ , by which all the formulas that change value have done so. Hence  $\nu_l = \nu_{l+1}$ . Call  $\mathcal{J}_l, \mathcal{J}_*$ ; and its corresponding evaluation function  $\nu^*$ . Then if  $\nu^*(\alpha) = t$ ,  $\nu^*(T \langle \alpha \rangle) = t$ . Similarly for  $f$  and  $b$ . Hence  $\nu^*(\alpha) = \nu^*(T \langle \alpha \rangle)$ , and so  $\mathcal{J}_*$  is a model of  $T_1$ . For the same reason,  $\mathcal{J}_*$  also verifies the two-way rule:

$$\frac{\neg \alpha}{\neg T \langle \alpha \rangle}$$

Yet the theory is not trivial: anything false in the standard model of arithmetic is untrue in  $\mathcal{J}_*$ , and so  $T_1 \not\models \alpha$ .

It is not difficult to see that the construction used to define  $\mathcal{J}_*$  is, in fact, just a dualised form of Kripke's fixed point construction for a logic with truth value gaps using the strong Kleene three-valued logic.<sup>118</sup> (Provided we start with a suitable ground model, monotonicity is guaranteed from the beginning, and so we can just set  $\nu_k(T \langle \alpha \rangle)$  to  $t$  (or  $f$ ) if  $\alpha$  takes the value  $t$  (or  $f$ ) at some  $i < k$ .) Hence, if any sentence is *grounded* in Kripke's sense, it takes a classical value in  $\mathcal{J}_*$ . In particular, if  $\alpha$  is any false grounded sentence,  $T_1 \not\models \alpha$ .

---

<sup>118</sup>See the article on Semantics and the Liar Paradox in this *Handbook*. One of the first people to realise that the construction could be dualised for this end was Dowden [1984].

## 8.2 Adding a Conditional

Although  $T_1$  validates the two-way inferential  $T$ -schema, it does not validate the  $T$ -schema as formulated with a detachable conditional. This is for the simple reason that  $LP$  does not contain such a conditional. A natural thought is to augment the language with one to make this possible. Let the resulting language be  $L_{\rightarrow}$ . Not all conditionals are suitable here, however. This is due to Curry paradoxes. If the conditional satisfies the inference of contraction:  $\alpha \rightarrow (\alpha \rightarrow \beta) \models \alpha \rightarrow \beta$ , then the theory collapses into triviality. For consider the fixed-point formula,  $\gamma$ , of the form  $T\langle\gamma\rangle \rightarrow \perp$  (or if  $\perp$  is not present, just an arbitrary  $\beta$ ). The  $T$ -schema gives:  $T\langle\gamma\rangle \leftrightarrow (T\langle\gamma\rangle \rightarrow \perp)$ . Contraction gives us:  $T\langle\gamma\rangle \rightarrow \perp$  and then a couple of applications of *modus ponens* give  $\perp$ .<sup>119</sup>

This fact rules out the use of all the non-transitive logics we looked at (since they validate  $\alpha \leftrightarrow (\alpha \rightarrow \beta) \models \beta$ ), all the da Costa logics and discussive logic (using discussive implication for the  $T$ -schema), since these validate contraction, and those relevance logics that validate contraction, such as  $R$ .<sup>120</sup> A relevant logic without contraction can be used for the purpose.

Let  $T_2$  be as for  $T_0$ , except that the  $T$ -schema is formulated with  $\rightarrow$ , and the underlying logic is  $BX$  (see 5.5, 6.5).  $T_2$  is inconsistent, since it is obviously stronger than  $T_1$ . But it can be shown to be non-trivial. If we try to generalise the proof for  $T_1$  in simple ways, attempts are stymied by the failure of anything like monotonicity once  $\rightarrow$  is involved. However, there is a way of building on the proof.<sup>121</sup> This requires us to move from objectual semantics to simple evaluational semantics. For the purpose of this section (and this one only), an atomic formula will be any of the usual kind *or* any one of the form  $\alpha \rightarrow \beta$ . Clearly, any sentence of the language can be built up from atomic formulas using  $\wedge$ ,  $\vee$ ,  $\neg$ ,  $\exists$  and  $\forall$ . Call an evaluation of atomic formulas,  $\nu$ , *arithmetical* if it assigns to every identity its value in the standard model of arithmetic. Given an arithmetical evaluation, it is extended to an evaluation of all sentences by  $LP$  truth conditions, using substitutional quantification.

A quick induction shows that any arithmetical evaluation assigns  $t$  to all the arithmetic truths of the standard model (which do not contain  $\rightarrow$  or  $T$ ), and  $f$  to all the falsehoods. Moreover, for this notion of valuation, we do have the Monotonicity Lemma. Finally, given any such evaluation, we

<sup>119</sup>An argument of this kind first appeared in Curry [1942]. Different versions that employ close relatives of contraction, such as  $\vdash (\alpha \wedge (\alpha \rightarrow \beta)) \rightarrow \beta$  (but *not*  $\alpha \wedge (\alpha \rightarrow \beta) \vdash \beta$ ) can also be found in the literature. See, e.g., Meyer *et al.* [1979].

<sup>120</sup>For good measure, it also rules out using Rescher and Manor's non-adjunctive approach. Using this, every consistent sentence would follow, since if  $\beta$  is consistent, so is  $\alpha \leftrightarrow (\alpha \rightarrow \beta)$ .

<sup>121</sup>The following is taken from Priest [1991b], which simply modifies Brady's proof for set theory in [1989].

can construct a fixed point,  $\nu^*$ , such that  $\nu^*(\alpha) = \nu^*(T \langle \alpha \rangle)$ , as in 8.1. The construction is the same, except that in the definition of  $\nu_k$ , we set  $Ts$  to  $t$  if  $s$  is a (closed) term that evaluates to the code of  $\alpha$ , and  $\alpha$  is eventually  $t$  by  $k$ . Similarly for  $f$ . (The values of atoms of the form  $\alpha \rightarrow \beta$  do not change in the process.)

An induction shows that if  $\mu \preceq \nu$  then for all  $i$ ,  $\mu_i \preceq \nu_i$ . Suppose the result for all  $i < k$ . We show it for  $k$ . We need consider only those atomic formulas of the form  $Ts$  where  $s$  evaluates to the code of sentence  $\alpha$ .  $\mu_k(Ts) = t$  iff  $\alpha$  is eventually  $t$  by  $k$ , for  $\mu$ . By induction hypothesis, this implies that  $\alpha$  is eventually  $t$  by  $k$  for  $\nu$ . Hence,  $\nu_k(Ts) = t$ , as required. The case for  $f$  is similar. From this result it obviously follows that if  $\mu \preceq \nu$  then  $\mu^* \preceq \nu^*$ .

Let  $\Rightarrow$  be the conditional connective of *RM3* (see 5.4, identifying  $+1$ ,  $0$ , and  $-1$  with  $t$ ,  $b$ , and  $f$ , respectively). This also plays a role in the proof. Its relevant property is that if  $\mu \preceq \nu$  then if  $\alpha$  and  $\beta$  are formulas of  $L_{\rightarrow}$  and  $\mu(\alpha \Rightarrow \beta) = t$ ,  $\nu(\alpha \Rightarrow \beta) = t$ . For if  $\mu(\alpha \Rightarrow \beta) = t$  then  $\mu(\alpha) = f$  or  $\mu(\beta) = t$ . By monotonicity  $\nu(\alpha) = f$  or  $\nu(\beta) = t$ . Hence,  $\nu(\alpha \Rightarrow \beta) = t$ .

Let  $\nu_0$  be the arithmetical interpretation that assigns every sentence of the form  $Ts$  the value  $b$ . We now define a transfinite sequence of arithmetic valuations,  $\langle \nu_i; i \in On \rangle$ , as follows. (I write  $(\nu_j)^*$  as  $\nu_j^*$ .) For  $k \neq 0$ :

$$\begin{aligned} \nu_k(\alpha \rightarrow \beta) &= t && \text{if } \forall j < k, \nu_j^*(\alpha \Rightarrow \beta) = t \\ &= f && \text{if } \exists j < k, \nu_j^*(\alpha \Rightarrow \beta) = f \\ &= b && \text{otherwise} \end{aligned}$$

And where  $\alpha$  is of the form  $Ts$ , where  $s$  is any closed term which evaluates to the code of a sentence:

$$\begin{aligned} \nu_k(\alpha) &= t && \text{if } \exists i \forall j (i \leq j < k, \nu_j^*(\alpha) = t) \\ &= f && \text{if } \exists i \forall j (i \leq j < k, \nu_j^*(\alpha) = f) \\ &= b && \text{otherwise} \end{aligned}$$

We can now establish that if  $i \leq k$  then  $\nu_i \preceq \nu_k$ . The proof is by transfinite induction. Suppose that the result holds for all  $j < k$ . We need to consider cases where a formula is of the form  $\alpha \rightarrow \beta$  or  $Ts$ , where  $s$  is a term that evaluates to the code of a sentence. Take them in that order.

Suppose that  $\nu_i(\alpha \rightarrow \beta) = t$ . Then  $\nu_0^*(\alpha \Rightarrow \beta) = t$ . By induction hypothesis, for  $0 < j < k$ ,  $\nu_0 \preceq \nu_j$ . Thus,  $\nu_0^* \preceq \nu_j^*$ . Hence,  $\nu_j^*(\alpha \Rightarrow \beta) = t$ , by the observation concerning  $\Rightarrow$ . Thus,  $\nu_k(\alpha \rightarrow \beta) = t$ , as required. The case for  $f$  is trivial.

For the other case, suppose that  $\nu_i(\alpha) = t$ . Then  $\exists j < i, \nu_j^*(\alpha) = t$ . By induction hypothesis, if  $j \leq l < k$ ,  $\nu_j \preceq \nu_l$ , and hence  $\nu_j^* \preceq \nu_l^*$ . By monotonicity,  $\nu_l^*(\alpha) = t$ . Thus,  $\nu_k(\alpha) = t$ . The case for  $f$  is similar.

What this lemma shows, as before, is that we must eventually reach an  $l$  such that  $\nu_l = \nu_{l+1}$ . Let this evaluation be  $\vec{\nu}$ . Then  $\vec{\nu}$  is a model of all the extensional arithmetic apparatus. It also models the  $T$ -schema. For if  $i < l$ ,  $\nu_i^*(\alpha) = \nu_i^*(T \langle \alpha \rangle)$ , and so  $\nu_i^*(\alpha \leftrightarrow T \langle \alpha \rangle) = t$  or  $b$ , and  $\vec{\nu}(\alpha \leftrightarrow T \langle \alpha \rangle) = t$  or  $b$ . (For the same reason,  $\vec{\nu}$  models the contraposed form:  $\neg\alpha \leftrightarrow \neg T \langle \alpha \rangle$ . Since  $T \langle \neg\alpha \rangle \leftrightarrow \neg\alpha$  is an instance of the  $T$ -schema, it also models  $T \langle \neg\alpha \rangle \leftrightarrow \neg T \langle \alpha \rangle$ .)

It remains to check that  $\vec{\nu}$  models the axioms and respects the rules of inference of  $BX$ . This requires no little checking. Most of it is routine. Here, for example, is one of the harder propositional axioms:  $((\alpha \rightarrow \beta) \wedge (\alpha \rightarrow \gamma)) \rightarrow (\alpha \rightarrow ((\beta \wedge \gamma)))$ . Let the antecedent be  $\varphi$ , and the consequent be  $\psi$ . Then  $\vec{\nu}(\varphi \rightarrow \psi) = t$  or  $b$  iff for no  $i < l$ ,  $\nu_i^*(\varphi \Rightarrow \psi) = f$ . Now, suppose that  $\nu_i^*(\varphi \Rightarrow \psi) = f$ . Then one of:

$$\begin{aligned} \nu_i^*(\varphi) = t \text{ and } (\nu_i^*(\psi) = b \text{ or } \nu_i^*(\psi) = f) \\ \nu_i^*(\varphi) = b \text{ and } \nu_i^*(\psi) = f \end{aligned}$$

In the first case,  $\nu_i(\alpha \rightarrow \beta) = \nu_i(\alpha \rightarrow \gamma) = t$ . But then for all  $j < i$ ,  $\nu_j^*(\alpha \Rightarrow \beta) = \nu_j^*(\alpha \Rightarrow \gamma) = t$ , in which case  $\nu_j^*(\alpha \Rightarrow (\beta \wedge \gamma)) = t$ , and so  $\nu_i(\alpha \rightarrow (\beta \wedge \gamma)) = t$ , which is impossible. In the second case,  $\nu_i(\alpha \rightarrow \beta) = t$  or  $b$ , and  $\nu_i(\alpha \rightarrow \gamma) = t$  or  $b$ . But then for all  $j < i$ ,  $\nu_j^*(\alpha \Rightarrow \beta) = t$  or  $b$ , and  $\nu_j^*(\alpha \Rightarrow \gamma) = t$  or  $b$ , in which case  $\nu_j^*(\alpha \Rightarrow (\beta \wedge \gamma)) = t$  or  $b$ , and so  $\nu_i(\alpha \rightarrow ((\beta \wedge \gamma))) = t$  or  $b$ , which is also impossible.

For further details, see Brady [1989].<sup>122</sup> The construction shows that  $T_1$  is non-trivial, since if  $\alpha$  is any arithmetic sentence false in the standard model  $\vec{\nu}(\alpha) = f$ . (Indeed, as with the previous construction, which is incorporated in this, if  $\alpha$  is any false grounded sentence, the same is true.)

### 8.3 Advantages of a Paraconsistent Approach

What we have seen is that it is possible to have a theory containing all the machinery of arithmetic, plus a truth predicate which satisfies the  $T$ -schema for every sentence of the language—whether this is formulated as a material biconditional, a two-way rule of inference, or a detachable bi-conditional. It is inconsistent, but non-trivial; in fact, the inconsistencies do not spread

<sup>122</sup>Brady shows that the construction verifies propositional logics that are a good deal stronger than  $BX$ . His treatment of identity is different, though. To verify the substitutivity rule of 7.1, it suffices to show that if  $t_1 = t_2$  holds in an interpretation then  $\alpha(x/t_1)$  and  $\alpha(x/t_2)$  have the same truth value. A quick induction shows that if this is true for atomic  $\alpha$  it is true for all  $\alpha$ . Hence, we need consider only these. Next, show by induction that if this holds for  $\nu$  it holds for all evaluations in the construction of  $\nu^*$ , and so of  $\nu^*$  itself. Finally, we show by induction that it holds for every  $\nu_i$  in the hierarchy, and hence for  $\vec{\nu}$ .

into the arithmetic machinery.<sup>123</sup> Thus, it is possible to have a workable, if inconsistent, theory which respects the central intuition about truth.

It is not my aim here to discuss the shortcomings of other standard approaches to the theory of truth,<sup>124</sup> but none can match this. All restrict the  $T$ -schema in one way or another. The one that comes closest to having the full  $T$ -schema is Kripke's account of truth, which at least has it in the form of a two way rule of inference. However, this account has the singular misfortune of being self-referentially inconsistent. According to this account, if  $\alpha$  is the Liar sentence it is *neither* true nor false, and so not true, but the theory pronounces  $\neg T \langle \alpha \rangle$  itself neither true nor false. According to  $T_2$ ,  $\alpha$  is *both* true and false (i.e., has a true negation), and this is exactly what it proves:  $T \langle \alpha \rangle \wedge \neg T \langle \alpha \rangle$  entails  $T \langle \alpha \rangle \wedge T \langle \neg \alpha \rangle$ . It might also show that  $\alpha$  is *not* true (and so not both true and false). But paraconsistency shows you exactly how to live with this kind of contradiction.

This is not unconnected with the matter of "strengthened" paradoxes. If someone holds the Liar sentence to be neither true nor false, one can invite them to consider the sentence,  $\beta$ , 'This sentence is not true' (as opposed to false). Whether  $\beta$  is true, false or neither, a contradiction arises. It is sometimes suggested that a paraconsistent account of truth falls to the same problem, since  $\beta$  can have no consistent truth-value on this account either. It should be clear that this argument is just an *ignoratio*. A paraconsistent account does not *require* it to have a consistent truth-value. In fact, according to  $T_2$ ,  $T \langle \neg \alpha \rangle \leftrightarrow \neg T \langle \alpha \rangle$ ; if this is right, there is no distinction between the standard Liar and the "strengthened Liar" at all.<sup>125</sup>

Let me finish with a word of caution. We can construct non-trivial theories which incorporate the  $S$ -schema of satisfaction and the  $D$ -schema of denotation, in exactly the same way as we did the  $T$ -schema. If, however, we try to add descriptions to a theory with self-reference and the  $D$ -schema, trouble does arise.

Suppose that we have a description operator,  $\varepsilon$ , satisfying the Hilbertian principle:  $\exists x \alpha \vdash \alpha(x/\varepsilon x \alpha)$ . If  $t$  is any closed term,  $t = t$ , and so by the  $D$ -schema  $D \langle t \rangle t$ , and  $\exists x D \langle t \rangle x$ . Thus, by the description principle,  $D \langle t \rangle \varepsilon x D \langle t \rangle x$ , whence, by the  $D$ -schema again:

$$t = \varepsilon x D \langle t \rangle x$$

---

<sup>123</sup>Nor does the  $T$ -schema have to be taken as axiomatic. One can give truth conditions for atomic sentences and then prove the  $T$ -schema in the usual Tarskian fashion. See Priest [1987], ch. 7.

<sup>124</sup>For this, see Priest [1987], ch. 2.

<sup>125</sup>The advantages of a paraconsistent account of truth rub off onto any account of modal (deontic, doxastic, etc.) operators that treats them as predicates. For all such theories are just sub-theories of the theory of truth. See Priest [1991b]. We will have an application of this concerning provability in 9.6.

Now in arithmetic, just as for any formula,  $\alpha$ , with one free variable,  $x$ , we can find a sentence,  $\beta$ , of the form  $\alpha(x/\langle\beta\rangle)$ , so, for any term,  $t$ , with one free variable,  $x$ , we can find a closed term,  $s$ , such that  $s$  is  $t(x/\langle s\rangle)$ . If  $f$  is any one place function symbol, apply this fact to the term  $f\varepsilon yDxy$ , to obtain an  $s$  such that:

$$s = f\varepsilon yD\langle s\rangle y$$

Since  $s = \varepsilon yD\langle s\rangle y$ , it follows that  $s = fs$ : any function has a fixed point. This shows that the semantic machinery does have purely arithmetic consequences. In particular, for example,  $\exists x x = x + 1$ . Arithmetic statements like this can be kept under control, as we will see later in the next part, but worse is to come.

Let  $f$  be the parity function, i.e.:

$$\begin{aligned} fx &= 0 && \text{if } x \text{ is odd} \\ &= 1 && \text{if } x \text{ is even} \end{aligned}$$

We have  $fs = 0 \vee fs = 1$ . In the first case  $s = fs = 0$ , which is even, and so  $fs = 1$ . Thus,  $0 = 1$ . Similarly in the second case. This is unacceptable, even for someone who supposes that there are *some* inconsistent numbers.

Where to point the finger of suspicion is obvious enough. As we saw, the  $D$ -schema entails  $\exists xD\langle t\rangle x$ , for any closed term,  $t$ ; and there is no reason why someone who subscribes to a paraconsistent account of semantic notions must believe that every term has a denotation: in particular, in the vernacular, ‘ $s$ ’ is ‘a number that is 1 if it is even and 0 if it is odd’, which would certainly seem to have no denotation. This suggests that the  $D$ -schema should be subjected to the condition that  $\exists xD\langle t\rangle x$  in some suitable way. The behaviour of resulting theories is a particularly interesting unsolved problem.<sup>126</sup>

#### 8.4 Set Theory in $LP$

Let us now turn to the second theory that we will look at, set theory. This is a theory of sets governed by the full Comprehension schema. This schema is structurally very similar to the  $T$ -schema, and many of the considerations of previous subsections carry over to set theory in a straightforward manner. The major element of novelty concerns the other axiom, the Extensionality axiom.

Let us start with set theory in  $LP$ . The language here contains just the predicates  $=$  and  $\in$ , and the axioms are:

---

<sup>126</sup>For a further discussion of all of these issues, see Priest [1997a].

$$\begin{aligned} & \exists x \forall y (y \in x \equiv \alpha) \\ & \forall x (x \in y \equiv x \in z) \supset y = z \end{aligned}$$

where  $x$  does not occur free in  $\alpha$ . Call this theory  $S_0$ .  $S_0$  is inconsistent. For putting  $y \notin y$  for  $\alpha$ , and instantiating the quantifier we get:  $\forall y (y \in r \equiv y \notin y)$ , whence  $r \in r \equiv r \notin r$ . Cashing out  $\supset$  in terms of  $\neg$  and  $\vee$  gives  $r \in r \wedge r \notin r$ .

In constructing models of  $S_0$ , the following observation (due to Restall [1992]) is a useful one. First some definitions. Given two vectors of  $LP$  values,  $(g_m; m \in D)$ ,  $(h_m; m \in D)$ , the first *subsumes* the second iff for all  $m \in D$ ,  $g_m \supseteq h_m$ . Now consider a matrix of such values  $(e_{m,n}; m, n \in D)$ . This is said to *cover* the vector  $(g_m; m \in D)$  iff for some  $n \in D$ , the vector  $(e_{m,n}; m \in D)$  subsumes it. A vector indexed by  $D$  is *classical* iff all its members are  $t$  or  $f$ . (Recall that we are writing  $\{1\}$ ,  $\{1, 0\}$ ,  $\{0\}$  as  $t$ ,  $b$ ,  $f$ , respectively.)

Now the observation. Consider an  $LP$  interpretation,  $\langle D, d \rangle$ , and the matrix  $(e_{m,n}; m, n \in D)$ , where  $e_{m,n} = \nu(m \in n)$ . If this covers every classical vector indexed by  $D$  it verifies the Comprehension principle. For let  $\alpha$  be any formula not containing  $x$ , and consider the vector  $(\nu(\alpha(y/m)); m \in D)$ . This certainly subsumes some classical vector; choose one such, and let this be subsumed by  $(e_{m,n}; m \in D)$ . Now consider any formula of the form  $m \in n \equiv \alpha(y/m)$ . Where the two sides differ in value, one of them has the value  $b$ . Hence, the value of the biconditional is either  $t$  or  $b$ . Thus the same is true of  $\forall y (y \in n \equiv \alpha)$ , and  $\exists x \forall y (y \in x \equiv \alpha)$ .

Using this fact, it is easy to construct models for  $S_0$ . Consider an  $LP$  interpretation,  $\langle D, d \rangle$ , where  $D = \{m, n\}$ , and  $e_{m,n}$  is given by the following matrix:

$\in$	$m$	$n$
$m$	$b$	$t$
$n$	$b$	$t$

Each column is the membership vector of the appropriate member of  $D$ ; and since that of  $m$  subsumes every classical vector indexed by  $D$ , this verifies the Comprehension axiom. In the Extensionality axiom, if  $y$  and  $z$  are the same, the axiom is obviously true. If they are distinct, one is  $n$  and the other is  $m$ , and for each  $x$ , the value of  $x \in n \equiv x \in m$  is  $b$ . Hence,  $\forall x (x \in n \equiv x \in m)$  has the value  $b$  and Extensionality is verified. In this model,  $m \notin n$  and  $n \notin n$  have the value  $f$ , as, therefore do  $\exists y y \notin n$  and  $\forall x \exists y y \notin x$ . Hence,  $S_0$  is non-trivial.

A characterisation of what can be proved in  $S_0$  (and of what its minimally inconsistent consequences are) is still an open question. There are, however, certainly theorems of Zermelo Fraenkel set theory,  $ZF$ , that are not provable

in  $S_0$ . For example, in  $ZF$  there is provably no universal set:  $ZF \vdash \forall x \exists y y \notin x$ . But this is not a consequence of  $S_0$ , as we have just seen.<sup>127</sup>

The simple model of  $S_0$  that we have just used to prove non-triviality is obviously pathological in some sense. An interesting question is what the “intended” interpretations of  $S_0$  are like. Whilst unable to give an answer to this, I note that for any classical model of  $ZF$ ,  $\mathcal{M} = \langle D, d \rangle$ , there is a model of  $S_0$  which has  $\mathcal{M}$  as a substructure. Let  $a$  be some new object, let  $\mathcal{M}^+ = \langle D^+, d^+ \rangle$ , where  $D^+ = D \cup \{a\}$ ,  $d^+$  is the same as  $d$ , except that for every  $c \in D^+$ , the value of  $c \in a$  is  $b$ ; for every  $c \in D$ , the value of  $a \in c$  is  $f$ ; the value of  $a = a$  is  $t$ ; and for every  $c \in D$ , the value of  $a = c$  is  $f$ .  $\mathcal{M}$  is clearly a substructure of  $\mathcal{M}^+$ . The membership vector of  $a$  subsumes every classical vector, and hence  $\mathcal{M}$  is a model of Comprehension.

It remains to verify Extensionality:  $\forall x(x \in m \equiv x \in n) \supset m = n$ . If  $m$  and  $n$  are the same in  $\mathcal{M}^+$ , then the consequent is true, as is the conditional. So suppose that they are distinct. If they are both in  $D$ , then, by extensionality in  $\mathcal{M}$ , there is some  $c \in D$  such that  $c \in m$  is  $t$  and  $c \in n$  is  $f$ , or vice versa. Whichever of these is the case,  $c \in m \equiv c \in n$  is  $f$ , as is  $\forall x(x \in m \equiv x \in n)$ . Hence the conditional is  $t$ . Finally, suppose that  $m \in D$  and  $n$  is  $a$  (or vice versa, which is similar). Then if  $c \in D^+$ , every sentence of the form  $c \in n$  is  $b$ . Hence, every sentence of the form  $c \in m \equiv c \in n$  is  $b$ , as therefore is  $\forall x(x \in n \equiv x \in m)$ . Hence, the conditional is true.

### 8.5 Brady’s Non-triviality Proof

As a working set theory,  $S_0$  is rather weak. Since the Comprehension axiom is only a material one, we cannot infer that something is in a set from the fact that it satisfies its defining condition, and vice versa. This suggests strengthening the principle to a two-way rule of inference, as we did for truth theory. This, in turn, requires the addition of set abstracts to the language. So let us enrich the language with terms of the form  $\{x; \alpha\}$  for any variable,  $x$ , and formula,  $\alpha$ ; and trade in the Comprehension principle of  $S_0$  for the two-way rule:

$$\frac{x \in \{y; \alpha\}}{\alpha(y/x)}$$

Call this theory  $S_1$ .  $S_1$  is inconsistent. For let  $r$  be  $\{x; x \notin x\}$ . Then:

$$\frac{r \in r}{r \notin r}$$

The law of excluded middle then quickly gives us  $r \in r \wedge r \notin r$ .

<sup>127</sup> For this, and some further observations in this direction, see Restall [1992].



The non-triviality of  $S_1$  is presently an open question. But even though it is probably non-trivial, as a working set theory, it is still rather weak. This is because we have no useful way of establishing that two sets are identical. Even if we can show that  $\forall x(\alpha \equiv \beta)$ , and so that  $\forall x(x \in \{x; \alpha\} \equiv x \in \{x; \beta\})$ , we cannot infer that  $\{x; \alpha\} = \{x; \beta\}$  since Extensionality does not support a detachable inference.

We might hope to circumvent this problem by trading in the Extensionality principle for the corresponding rule:

$$\frac{\forall x(\alpha \equiv \beta)}{\{x; \alpha\} = \{x; \beta\}}$$

But if we do this, trouble arises.<sup>128</sup> For let  $r$  be as before. Then since  $r \in r$  must take the value  $b$  in any interpretation, we have, for any  $\alpha$ ,  $\forall x(\alpha \equiv r \in r)$ , and so  $\{x; \alpha\} = \{x; r \in r\}$ . Thus, for any  $\alpha$  and  $\beta$ ,  $\{x; \alpha\} = \{x; \beta\}$ ; which is rather too much.

The problem arises because the Extensionality rule of inference allows us to move from an equivalence that does not guarantee substitution ( $\alpha \equiv \beta, \beta \equiv \gamma \not\vdash_{LP} \alpha \equiv \gamma$ ) to one that does (identity). This suggests formulating Extensionality itself with a connective that legitimises substitution. So let us add a detachable connective to the language,  $\leftrightarrow$ , and formulate Extensionality as:

$$\frac{\forall x(\alpha \leftrightarrow \beta)}{\{x; \alpha\} = \{x; \beta\}}$$

The trouble then disappears.

And now that we have a detachable conditional connective at our disposal, it is natural to formulate the Comprehension principle as a detachable biconditional, as follows:

$$\forall y(y \in \{x; \alpha\} \leftrightarrow \alpha(x/y))$$

We have to be careful about the conditional connective here. As with truth, any conditional connective that satisfies contraction would give rise to triviality. For let  $c$  be  $\{x \in x \rightarrow \perp\}$ . Then an instance of Comprehension is  $y \in c \leftrightarrow (y \in y \rightarrow \perp)$ . Instantiating with  $c$ , we get  $c \in c \leftrightarrow (c \in c \rightarrow \perp)$ , and we can then proceed, as with truth, to obtain  $\perp$ . Even if the logic does not contain contraction, Curry-style paradoxes may still be forthcoming. For example, if we drop the contraction axiom from the relevant logic  $R$

---

<sup>128</sup>There are other cases where the full Comprehension principle by itself is alright, but throwing in extensionality causes problems; for example, set theory based on Lukasiewicz' continuum-valued logic. See White [1979].

and add the law of excluded middle, the Comprehension principle still gives triviality.<sup>129</sup> Again however, a relevant logic without contraction will do the job.

Consider the set theory with Extensionality and Comprehension formulated as just described, and based on the underlying logic  $BX$  (with free variables, so that these may occur in the schematic letters of Extensionality and Comprehension). Call this  $S_2$ . The first thing to note about  $S_2$  is that identity can be defined in it, in Russellian fashion. Writing  $x = y$  for  $\forall z(x \in z \leftrightarrow y \in z)$ ,  $x = x$  follows. Substituting  $\{w; \alpha\}$  for  $z$ , and using the Comprehension principle gives  $\alpha(w/x) \leftrightarrow \alpha(w/y)$ . Hence, we need no longer assume that  $=$  is part of the language.

Since the Comprehension principle of  $S_2$  gives the two-way deduction version of  $S_1$ ,  $S_2$  is inconsistent. It is also demonstrably non-trivial, as shown by Brady [1989].<sup>130</sup> To prove this, we repeat the proof for  $T_2$  of 8.2 with three modifications. The first, a minor one, is that we add two propositional constants  $t$  and  $f$  to the language; their truth values are always what the letters suggest. (This is necessary to kick-start the generation of the fixed point into motion. In the case of truth, this was done by the arithmetic sentences.) More substantially, in constructing  $\nu^*$  we replace the clause for  $T$  by:

$$\begin{aligned} \nu_k(s \in \{x; \alpha\}) &= t && \text{if } \alpha(x/s) \text{ is eventually } t \text{ by } k \\ &= f && \text{if } \alpha(x/s) \text{ is eventually } f \text{ by } k \\ &= b && \text{otherwise} \end{aligned}$$

where  $s$  is any closed term, and  $\alpha$  contains at most  $x$  free. The final modification is that in extending evaluations to all formulas, we use substitutional quantification with respect to the closed set abstracts.

Now,  $\vec{\nu}$  verifies all the theorems of  $S_2$ , in the sense that if  $\alpha$  is any closed substitution instance of a theorem, it receives the value  $t$  or  $b$  in  $\vec{\nu}$ . This is shown by an induction on the length of proofs. That the logical axioms have this property, and the logical rules of inference preserve this property, is shown as in 8.2. This leaves the set theoretic ones.

Given the construction of  $\vec{\nu}$ , it is not difficult to see that it verifies the Comprehension principle. It is not at all obvious that Extensionality preserves verification. What needs to be shown is that if  $\forall x(\alpha \leftrightarrow \beta)$  is verified, so is anything of the form  $a \in c \leftrightarrow b \in c$ , where  $a$  is  $\{x; \alpha\}$  and  $b$  is  $\{x; \beta\}$ . Let  $c$  be  $\{y; \gamma\}$ . Then, given Comprehension, what needs to be shown is that  $\gamma(y/a) \leftrightarrow \gamma(y/b)$  is verified. If this can be shown for atomic  $\gamma$ , the result will follow by induction. Given the premise of the inference and Comprehension, it is true if  $\gamma$  is of the form  $d \in y$ . If it is of the form  $y \in d$ , where

<sup>129</sup>See Slaney [1989]. Other classical principles are also known to give rise to triviality in conjunction with the Comprehension schema. See Bunder [1986].

<sup>130</sup>A modification of the proof shows that the theory based on the logic  $B$  is, in fact, consistent. See Brady [1983].

$d$  is  $\{z; \delta\}$ , we need to show that  $\delta(z/a) \leftrightarrow \delta(z/b)$  is verified. We obviously have a regress. In fact, the regression grounds out in a suitable way in the construction of  $\vec{v}$ . For details, see Brady [1989].<sup>131</sup>

The non-triviality of  $S_2$  is established since there are many sentences that are not verified by  $\vec{v}$ . It is easy to check, for example, that any sentence of the form  $c \in \{x; f\}$  takes the value  $f$ , as, therefore, does the formula  $\forall x \exists y y \in x$ .

A notable feature of Brady's proof is the following. As formulated, the Comprehension principle entails:  $\exists y \forall x (x \in y \leftrightarrow \alpha)$ , where  $y$  does not occur in  $\alpha$ . (The  $y$  in question is  $\{x; \alpha\}$  and so cannot be a subformula of  $\alpha$ .) If we relax the restriction, we get an absolutely unrestricted version of the principle. Brady's proof can be extended to verify this version too, by adding a fixed point operator to the language, and treating it suitably. Again, for details, see Brady [1989].

Finally, it is worth observing that the  $T$ -schema is interpretable in  $S_2$ . If  $\alpha$  is any closed formula, let us write  $\langle \alpha \rangle$  for  $\{z; \alpha\}$ , where  $z$  is some fixed variable. Define  $Tx$  to be  $a \in x$ , where  $a$  is any fixed term. Then  $T \langle \alpha \rangle = a \in \{z; \alpha\} \leftrightarrow \alpha$ . Moreover, the absolutely unrestricted Comprehension principle gives us fixed points of the kind required for self-reference. Let  $\alpha$  be any formula of one free variable,  $x$ . By the principle, there is a set,  $s$ , such that  $\forall x (x \in s \leftrightarrow \alpha(x/\{z; a \in s\}))$ . It follows that  $a \in s \leftrightarrow \alpha(x/\{z; a \in s\})$ . Thus, if  $\beta$  is  $a \in s$ , we have  $\beta \leftrightarrow \alpha(x/\langle \beta \rangle)$ .  $S_2$  (with the absolutely unrestricted Comprehension principle) therefore gives us a demonstrably non-trivial *joint* theory of truth, sethood and self-reference.

### 8.6 Paraconsistent Set Theory

Despite the strong structural similarities between semantics and set theory, there is an important historical difference. Set theory is a well developed mathematical theory in a way that semantics is not. In the case of set theory, it is therefore natural to ask how a paraconsistent theory such as  $S_2$  relates to this development.

To answer this question (at least to the extent that the answer is known), it will be useful to divide set theory into three parts. The first comprises that basic set-theory which all branches of mathematics use as a tool. The second is transfinite set theory, as it can be established in  $ZF$ . The third

---

<sup>131</sup> Brady's treatment of identity is slightly different from the one given here. He defines  $x = y$  as  $\forall z (z \in x \leftrightarrow z \in y)$ . Given Comprehension, this delivers the version of Extensionality used here straight away. What is lost is the substitution principle  $x = y, \alpha(w/x) \vdash \alpha(w/y)$ . Given the Comprehension principle, this can be reduced to  $x = y, x \in z \vdash y \in z$  (which follows from our definition of identity). Brady takes something stronger than this as his substitutivity axiom:  $\vdash (x = y \wedge z = z) \rightarrow (x \in z \leftrightarrow y \in z)$ . Hence, his construction certainly verifies the weaker principle. It is worth noting that the construction does not validate the simpler  $\vdash x = y \rightarrow (x \in z \leftrightarrow y \in z)$ , which, in any case, is known to be a Destroyer of Relevance. See Routley [1980b], sect. 7.

concerns results about sets, like Russell's set and the universal set, that do not exist in  $ZF$ . Let us take these matters in turn.

$S_2$  is able to provide for virtually all of bread-and-butter set theory (Boolean operations on sets, power sets, products, functions, operations on functions, etc.), and so provide for the needs of working mathematics.<sup>132</sup> For example, if we define the Boolean operators,  $x \cap y$ ,  $x \cup y$  and  $\bar{x}$  as  $\{z; z \in x \wedge z \in y\}$ ,  $\{z; z \in x \vee z \in y\}$  and  $\{z; z \notin x\}$ , respectively, and  $x \subseteq y$  as  $\forall z(z \in x \rightarrow z \in y)$ , then we can establish the usual facts concerning these notions. Some care needs to be taken over defining a universal set,  $U$ , and empty set,  $\phi$ , though. If we define  $\phi$ , as  $\{x; x \neq x\}$ , we cannot show that for all  $y$ ,  $\phi \subseteq y$ , since the underlying logic is relevant and cannot prove  $x \neq x \rightarrow \alpha$  for arbitrary  $\alpha$ . (Dually for  $U$ .) If we define  $\phi$  as  $\{x; \forall z x \in z\}$ , this problem is solved, since  $\forall z x \in z \rightarrow x \in y$ . (Dually for  $U$ .)

The reason for the qualification 'virtually' in the first sentence of the last paragraph, is as follows. The sets, as structured by union, intersection and complementation, are not a Boolean algebra, but a De Morgan algebra with maximum and minimum elements. Though we can show that  $\forall y y \notin x \cap \bar{x}$ , we cannot show that  $x \cap \bar{x} \subseteq \phi$ , since, relevantly,  $(\alpha \wedge \neg\alpha) \rightarrow \beta$  fails. (Dually for  $U$ .) There are, in a sense, more than one universal and empty sets. Moreover, this is essential. If we had  $x \cap \bar{x} \subseteq \phi$  then, taking  $\{z; \alpha\}$  for  $x$ , we get  $(\alpha \wedge \neg\alpha) \rightarrow \forall y z \in y$ . Now take  $\{z; \beta\}$  for  $y$ , and we get  $(\alpha \wedge \neg\alpha) \rightarrow \beta$ ; paraconsistency fails. In fact, Dunn [1988] shows that if the principles that there is a unique universal set, and a unique empty set, are added to any set theory such as  $S_2$ , full classical logic falls out.

Turning to the second area, the question of how much of the usual transfinite set theory can be established in  $S_2$  is one to which the answer is currently unknown. What can be said is that the *standard* proofs of a number of results break down. This is particularly the case for results that are proved by *reductio*, such as Cantor's Theorem. Where  $\alpha$  is an assumption made for the purpose of *reductio*, we may well be able to establish that  $(\alpha \wedge \beta) \rightarrow (\gamma \wedge \neg\gamma)$ , for some  $\gamma$ , where  $\beta$  is the conjunction of other facts appealed to in deducing the contradiction (such as instances of the Comprehension principle). But contraposing and detaching will give us only  $\neg\alpha \vee \neg\beta$ , and we can get no further.<sup>133</sup>

Lastly, the third area: reasoning in  $S_2$ , one can prove various results about sets that are impossible in  $ZF$ . For example, as usual, let  $\{x\}$  be  $\{y; y = x\}$ ,  $\{x, y\}$  be  $\{x\} \cup \{y\}$  and  $\bigcup x$  be  $\{z; \exists y \in x, z \in y\}$ .  $r = \{x; x \notin x\}$ , and we know that  $r \in r$  and  $r \notin r$ . Then:<sup>134</sup>

- (1) If  $x \in r$  then  $\{x\} \in r$ . For  $\{x\} \in \{x\}$  or  $\{x\} \notin \{x\}$ . In the first case,

<sup>132</sup>Much of this is spelled out in Routley [1980b], sect. 8.

<sup>133</sup>Interesting enough, however, it is possible to prove a version of the Axiom of Choice using the completely unrestricted version of the Comprehension principle. See Routley [1980b], sect. 8.

<sup>134</sup>The following is taken from Arruda and Batens [1982].

$\{x\} = x$ , and so  $\{x\} \in r$ . In the second case,  $\{x\} \in r$  by definition.

(2) If  $x, y \in r$  then  $\{x, y\} \in r$ . For  $\{x, y\} \in \{x, y\}$  or  $\{x, y\} \notin \{x, y\}$ . In the first case,  $\{x, y\} = x$  or  $\{x, y\} = y$ , and so  $\{x, y\} \in r$ . In the second case,  $\{x, y\} \in r$  by definition.

(3)  $\{\{x, r\}\} \in r$ . For  $\{\{x, r\}\} \in \{\{x, r\}\}$  or  $\{\{x, r\}\} \notin \{\{x, r\}\}$ . In the first case,  $\{\{x, r\}\} = \{x, r\}$ , hence,  $x = \{x, r\} = r$ . But then  $x, r \in r$  so  $\{x, r\} \in r$ , by (2), and,  $\{\{x, r\}\} \in r$ , by (1). In the second case,  $\{\{x, r\}\} \in r$  by definition.

(4)  $\forall x x \in \bigcup r$ . For suppose that  $\{x, r\} \in \{x, r\}$ . Then  $\{x, r\} = x$  or  $\{x, r\} = r$ . In the first case,  $\{x\} = \{\{x, r\}\}$ , so  $\{x\} \in r$ , by (3). In the second,  $\{x, r\} \in r$ . In either case  $x \in \bigcup r$ . Suppose, on the other hand, that  $\{x, r\} \notin \{x, r\}$ . Then  $\{x, r\} \in r$ , by definition, and so  $x \in \bigcup r$ .

That  $\bigcup r$  is universal, is hardly a profound result. But it at least illustrates the fact that there are possibilities which transcend  $ZF$ .

Let me end this section with a speculative comment on what all this shows. The discussion of this section, and especially the part concerning the non-Boolean properties of sets in  $S_2$ , shows that it is impossible to recapture standard set theory in its entirety in this theory. Sets are extensional entities *par excellence*; using an intensional connective in their identity conditions is bound to gum up the works. In fact, it seems to me that the most plausible way of viewing  $S_2$  is as a theory of properties, where intensional identity conditions are entirely appropriate. But what you call these entities does not really matter here. The important fact is that they are not the sets of standard modern mathematical practice.

If we want a theory of such entities, the appropriate identity conditions must employ  $\equiv$ , and this means that we are back with the proof-theoretically weak  $S_0$  (or  $S_1$ ). Since this does not contain  $ZF$ , how should someone who subscribes to a paraconsistent theory of such sets view modern mathematical practice?

One answer is as follows. The standard model of  $ZF$  is the cumulative hierarchy. As we saw in 8.2, there are models of  $S_0$  which contain this hierarchy. We may thus take it that the intended interpretation of  $S_0$  is a model of this kind (or if there are more than one, that they are all models of this kind). The cumulative hierarchy is therefore a (consistent) fragment of the set-theoretic universe, and modern set theory provides a description of it. There is, however, more to the universe than this fragment. A classical logician may well agree with that claim. For example, they may think that there are also non-well-founded sets. The paraconsistent logician agrees with this: after all,  $r$  is not well-founded; but they will think that sets outside the hierarchy may have even more remarkable properties: some of them are inconsistent.

## 9 ARITHMETIC AND ITS METATHEORY

In this part I want to look at the application of paraconsistent logic to another important mathematical theory: arithmetic. The situation concerning arithmetic is rather different from that concerning set theory and semantics. There are no apparently obvious and intrinsically arithmetical principles that give rise to contradiction, in the way that the Comprehension principle and the  $T$ -schema do—or if there are, this fact has not yet been discovered. In the first instance, the paraconsistent interest in arithmetic arises because there is a class of inconsistent models of arithmetic. (It might be more accurate to say ‘models of inconsistent arithmetic’.) It may be supposed that these models are pathological in some sense.<sup>135</sup> I will come back to this matter later. But even if it is so, the models nevertheless have an interesting and important *mathematical* structure, as do the classical non-standard models of arithmetic—which are, in fact, just a special case, as we will see. And just as one does not have to be an intuitionist to find intuitionistic structures of intrinsic mathematical interest, so one does not have to be a dialetheist for the same to be true of inconsistent structures. One thing this part illustrates, therefore, is the existence of a new branch of mathematics which concerns the investigation of just such structures.<sup>136</sup>

The existence of inconsistent models of arithmetic bears, as might be expected, on the limitative theorems of Metamathematics. And whatever the status of the inconsistent models themselves, many have held that these theorems have important philosophical implications. This part will also look at the connection between the inconsistent models and the limitative theorems, and I will comment on the significance of this for the philosophical implications of Gödel’s incompleteness theorem.

### 9.1 *The Collapsing Lemma*

Let us start with a theorem about  $LP$  on which much of the following depends: the Collapsing Lemma.<sup>137</sup>

Let  $\mathcal{I} = \langle D, d \rangle$  be any interpretation for  $LP$ . Let  $\sim$  be any equivalence relation on  $D$ , that is also a congruence relation on the denotations of the function symbols in the language (i.e., if  $g$  is such a denotation, and  $d_i \sim e_i$  for all  $1 \leq i \leq n$ , then  $g(d_1, \dots, d_n) \sim g(e_1, \dots, e_n)$ ). If  $d \in D$  let  $[d]$  be the

<sup>135</sup>Though this claim has certainly been queried. See Priest [1994].

<sup>136</sup>On this, see further, Mortensen [1995]. Perhaps surprisingly, the first person to investigate an inconsistent arithmetic was Nelson [1959], who gave a realisability-style semantics for the language of arithmetic, according to which the set of formulas realised was inconsistent (and closed under a logic somewhat weaker than intuitionist logic).

<sup>137</sup>The theorem works equally well for  $FDE$ , but we will be concerned primarily with models of theories that contain the law of excluded middle, and so where there are no truth-value gaps.

equivalence class of  $d$  under  $\sim$ . Define an interpretation,  $\mathcal{I}_\sim = \langle D_\sim, d_\sim \rangle$ , to be called the *collapsed interpretation*, where  $D_\sim = \{[d]; d \in D\}$ ; if  $c$  is a constant,  $d_\sim(c) = [d(c)]$ ; if  $f$  is an  $n$ -place function symbol:

$$d_\sim(f)([d_1], \dots, [d_n]) = [d(f)(d_1, \dots, d_n)]$$

(this is well defined, since  $\sim$  is a congruence relation); and if  $P$  is an  $n$ -place predicate, its extension and anti-extension in  $\mathcal{I}_\sim$ ,  $E_P^\sim$  and  $A_P^\sim$ , are defined by:

$$\begin{aligned} \langle [d_1], \dots, [d_n] \rangle \in E_P^\sim &\text{ iff for all } 1 \leq i \leq n, \exists e_i \sim d_i, \langle e_1, \dots, e_n \rangle \in E_P \\ \langle [d_1], \dots, [d_n] \rangle \in A_P^\sim &\text{ iff for all } 1 \leq i \leq n, \exists e_i \sim d_i, \langle e_1, \dots, e_n \rangle \in A_P \end{aligned}$$

where  $E_P$  and  $A_P$  are the extension and anti-extension of  $P$  in  $\mathcal{I}$ . It is easy to check that  $E_\sim^\sim$  is  $\{\langle [d], [d] \rangle; d \in D\}$ , as required for an  $LP$  interpretation.

The collapsed interpretation, in effect, identifies all members of an equivalence class to produce a composite individual that has the properties of all of its members. It may, of course, be inconsistent, even if its members are not.

A swift induction confirms that for any closed term,  $t$ ,  $d_\sim(t) = [d(t)]$ . Hence:

$$\begin{aligned} 1 \in \nu(Pt_1 \dots t_n) &\Rightarrow \langle d(t_1), \dots, d(t_n) \rangle \in E_P \\ &\Rightarrow \langle [d(t_1)], \dots, [d(t_n)] \rangle \in E_P^\sim \\ &\Rightarrow \langle d_\sim(t_1), \dots, d_\sim(t_n) \rangle \in E_P^\sim \\ &\Rightarrow 1 \in \nu_\sim(Pt_1 \dots t_n) \end{aligned}$$

Similarly for 0 and anti-extensions. Monotonicity then entails that for any formula,  $\alpha$ ,  $\nu(\alpha) \subseteq \nu_\sim(\alpha)$ . This is the *Collapsing Lemma*.<sup>138</sup>

The Collapsing Lemma assures us that if an interpretation is a model of some set of sentences, then any interpretation obtained by collapsing it will also be a model. This gives us an important way of constructing inconsistent models. In particular, if the language contains no function symbols, and  $\mathcal{I}$  is a model of some set of sentences, then, by appropriate choice of equivalence relation, we can collapse it down to a model of *any* smaller size. Thus we have a very strong downward Löwenheim-Skolem Theorem: If a theory in a language without function symbols has a model, it has a model of all smaller cardinalities.

I note that, since monotonicity holds for second order  $LP$  (section 7.2), the Collapsing Lemma extends to second order  $LP$ . Details are left as an exercise.

---

<sup>138</sup>The result is proved in Priest [1991a]. A similar result was proved by Dunn [1979].

## 9.2 Collapsed Models of Arithmetic

From now on, let  $L$  be the standard language of first-order arithmetic: one constant,  $\mathbf{0}$ , function symbols for successor, addition and multiplication,  $'$ ,  $+$ , and  $\times$ , respectively, and one predicate symbol,  $=$ . If  $\mathcal{I}$  is any interpretation, let  $Th(\mathcal{I})$  (the *theory* of  $\mathcal{I}$ ) be the set of all sentences true in  $\mathcal{I}$ . Let  $\mathcal{N}$  be the standard model of arithmetic, and  $A = Th(\mathcal{N})$ . Let  $\mathcal{M} = \langle M, d \rangle$  be any classical model of  $A$ —which is just special cases of an *LP* model. (As is well known, there are many of these other than  $\mathcal{N}$ .<sup>139</sup>) I will refer to the denotations of  $'$ ,  $+$ , and  $\times$  as the arithmetic operations of  $\mathcal{M}$ , and since no confusion is likely, use the same signs for them.<sup>140</sup>

Let  $\sim$  be an equivalence relation on  $M$ , that is also a congruence relation with respect to the interpretations of the function symbols. Then we may construct the collapsed interpretation,  $\mathcal{M}_\sim$ . By the Collapsing Lemma,  $\mathcal{M}_\sim$  is a model of  $A$ . Provided that  $\sim$  is not the trivial equivalence relation, that relates each thing only to itself, then  $\mathcal{M}_\sim$  will model inconsistencies. For suppose that  $\sim$ , relates the distinct members of  $M$ ,  $n$  and  $m$ , then in  $\mathcal{M}_\sim$ ,  $[n] = [m]$  and so  $\langle [n], [m] \rangle$  is in the extension of  $=$ . But since  $n \neq m$  in  $\mathcal{M}$ ,  $\langle [n], [m] \rangle$  is in the anti-extension too. Thus,  $\exists x(x = x \wedge x \neq x)$  holds in  $\mathcal{M}_\sim$ .

As an illustration of constructing an inconsistent model of  $A$  using the Collapsing Lemma, suppose that we partition  $M$  into  $n+1$  successive blocks,  $C_0, \dots, C_{n+1}$ , such that if  $x, z \in C_i$  and  $x < y < z$  then  $y \in C_i$ . And suppose that for  $0 < i \leq n+1$ ,  $C_i$  is closed under the arithmetic operations of  $\mathcal{M}$ . (The existence of such a partition follows from a standard result in the study of classical models of arithmetic. See Kaye [1991], sect. 6.1.) Let  $1 \leq k \in C_0 \cup C_1$  and define  $x \sim y$  as:

$$(x, y \in C_0 \text{ and } x = y) \text{ or} \\ \text{for some } 0 < i \leq n+1, x, y \in C_i \text{ and } x = y \text{ mod } k$$

where ' $x = y \text{ mod } k$ ' means that for some  $j \in M$ ,  $x + j \times k = y$ , in  $\mathcal{M}$ .

It is not difficult to check that  $\sim$  is an equivalence relation on  $M$ , and, moreover, that it is a congruence relation on the arithmetic operations of  $\mathcal{M}$ . Hence, we may use it to give a collapsed model. In this,  $C_0$  collapses into an initial tail of numbers, and each  $C_i$  ( $0 < i \leq n+1$ ) collapses into a block of period  $k$ . For example, if  $\mathcal{M}$  is the standard model,  $n = 1$  and  $C_0 = \emptyset$ , the collapsed model is a simple cycle of period  $k$ . The successor function in the model may be depicted as follows:

<sup>139</sup>See, e.g., Kaye [1991].

<sup>140</sup>For a more detailed discussion of the material in this section, see Priest [1997a].



$$\begin{array}{ccccccc}
 0 & \rightarrow & 1 & \rightarrow & \dots & \rightarrow & i \\
 \uparrow & & & & & & \downarrow \\
 k-1 & \leftarrow & & \dots & & \leftarrow & i+1
 \end{array}$$

I will call such models *cycle models*. They were, in fact, the first inconsistent models to be discovered.<sup>141</sup> If  $\mathcal{M}$  is any model,  $n = 1$ , and  $k = 1$ , we have a tail isomorphic to  $C_0$ , and then a degenerate single-point cycle. In particular, if  $\mathcal{M}$  is a non-standard model and  $C_0$  comprises the standard numbers, we have the natural numbers with a “point at infinity”,  $\Omega$ :

$$0 \rightarrow 1 \rightarrow \dots \rightarrow \Omega \leftarrow$$

### 9.3 Inconsistent Models of Arithmetic

Now that we have seen the existence of inconsistent models of arithmetic, let us look at their general structure.

Take any *LP* model of arithmetic,  $\mathcal{M} = \langle M, d \rangle$ . I will call the denotations of the numerals *regular* numbers. Let  $x \leq y$  be defined in the usual way, as  $\exists z \ x + z = y$ . It is easy to check that  $\leq$  is transitive. For if  $i \leq j \leq k$  then for some  $x, y, i + x = j$  and  $j + y = k$ . Hence  $(i + x) + y = k$ . But  $(i + x) + y = i + (x + y)$  (since it is a model of arithmetic). The result follows.

If  $i \in M$ , let  $N(i)$  (the *nucleus* of  $i$ ) be  $\{x \in M; i \leq x \leq i\}$ . In a classical model,  $N(i) = \{i\}$ , but this need not be the case in an inconsistent model. For example, in a cycle model the members of the cycle constitute a nucleus. If  $j \in N(i)$  then  $N(i) = N(j)$ . For if  $x \in N(j)$  then  $i \leq j \leq x \leq j \leq i$ , so  $x \in N(i)$ , and similarly in the other direction. Thus, every member of a nucleus defines the same nucleus.

Now, if  $N_1$  and  $N_2$  are nuclei, define  $N_1 \preceq N_2$  to mean that for some (or all, it makes no difference)  $i \in N_1$  and  $j \in N_2, i \leq j$ . It is not difficult to check that  $\preceq$  is a partial ordering. Moreover, since for any  $i$  and  $j, i \leq j$  or  $j \leq i$ , it is a linear ordering. The least member of the ordering is  $N(0)$ . If  $N(1)$  is distinct from this, it is the next (since for any  $x, x \leq 0 \vee x \geq 1$ ), and so on for all regular numbers.

Say that  $i \in M$  has period  $p \in M$  iff  $i + p = i$ . In a classical model every number has period 0 and only 0. But again, this need not be the case in an inconsistent model, as the cycle models demonstrate. If  $i \leq j$  and  $i$  has period  $p$  so does  $j$ . For  $j = i + x$ , so  $p + j = p + i + x = i + x = j$ . In particular, if  $p$  is a period of some member of a nucleus, it is a period of

---

<sup>141</sup>This was by Meyer [1978]. Things are spelled out in Meyer and Mortensen [1984]. The idea of collapsing non-standard classical models is to be found in Mortensen [1987]. Different structures can be collapsed to provide inconsistent models of other kinds of number, e.g., real numbers. See Mortensen [1995].

every member. We may thus say that  $p$  is a period of the nucleus itself. It also follows that if  $N_1 \preceq N_2$  and  $p$  is a period of  $N_1$  it is a period of  $N_2$ .

If a nucleus has a regular non-zero period,  $m$ , then it must have a minimum (in the usual sense) non-zero period, since the sequence  $0, 1, 2, \dots, m$  is finite. If  $N_1 \preceq N_2$  and  $N_1$  has minimum regular non-zero period,  $p$ , then  $p$  is a period of  $N_2$ . Moreover, the minimum non-zero period of  $N_2$ ,  $q$ , must be a divisor (in the usual sense) of  $p$ . For suppose that  $q < p$ , and that  $q$  is not a divisor of  $p$ . For some  $0 < k < q$ ,  $p$  is some finite multiple of  $q$  plus  $k$ . So if  $x \in N_2$ ,  $x = x + q = x + p + \dots + p + k$ . Hence  $x = x + k$ , i.e.,  $k$  is a period of  $N_2$ , which is impossible.

If a nucleus has period  $p \geq 1$ , I will call it *proper*. Every proper nucleus is closed under successors. For suppose that  $j \in N$  with period  $p$ . Then  $j \leq j' \leq j + p = j$ . Hence,  $j' \in N$ . In an inconsistent model, a number may have more than one predecessor, i.e., there may be more than one  $x$  such that  $x' = j$ . (Although  $x' = y' \supset x = y$  holds in the model, we cannot necessarily detach to obtain  $x = y$ .)<sup>142</sup> But if  $j$  is in a proper nucleus,  $N$ , it has a unique predecessor in  $N$ . For let the period of  $N$  be  $q'$ . Then  $(j + q)' = j + q' = j$ . Hence,  $j + q$  is a predecessor of  $j$ ; and  $j \leq j + q' = j$ . Hence,  $j + q \in N$ . Next, suppose that  $x$  and  $y$  are in the nucleus, and that  $x' = y' = j$ . We have that  $x \leq y \vee y \leq x$ . Suppose, without loss of generality, the first disjunct. Then for some  $z$ ,  $x + z = y$ ; so  $j + z = j$ , and  $z$  is a period of the nucleus. But then  $x = x + z = y$ . I will write the unique predecessor of  $j$  in the nucleus as  $'j$ .

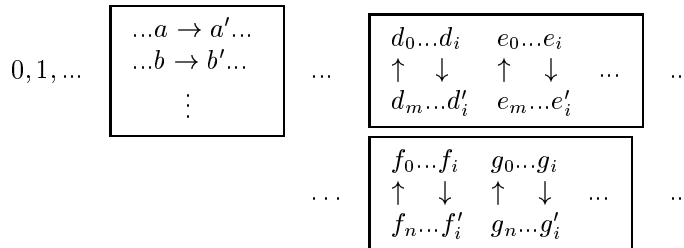
Now let  $N$  be any proper nucleus, and  $i \in N$ . Consider the sequence  $\dots, ''i, 'i, i, i, 'i, i'' \dots$ . Call this the *chromosome* of  $i$ . Note that if  $i, j \in N$ , the chromosomes of  $i$  and  $j$  are identical or disjoint. For if they have a common member,  $z$ , then all the finite successors of  $z$  are identical, as are all its finite predecessors (in  $N$ ). Thus they are identical. Now consider the chromosome of  $i$ , and suppose that two members are identical. There must be members where the successor distance between them is a minimum. Let these be  $j$  and  $j' \dots'$  where there are  $n$  primes. Then  $j = j + n$ , and  $n$  is a period of the nucleus—in fact, its minimum non-zero period—and the chromosome of every member of the nucleus is a successor cycle of period  $n$ .

Hence, any proper nucleus is a collection of chromosomes, all of which are either successor cycles of the same finite period, or are sequences isomorphic to the integers (positive and negative). Both sorts are possible in an inconsistent model. Just consider the collapse of a non-standard model, of the kind given in the last section, by an equivalence relation which leaves all the standard numbers alone and identifies all the others modulo  $p$ . If  $p$  is standard, the non-standard numbers collapse into a successor cycle; if it

<sup>142</sup>In fact, it is not difficult to show that there is at most one number with multiple predecessors; and this can have only two.

is non-standard, the nucleus generated is of the other kind.

To summarise so far, the general structure of a model is a linear sequence of nuclei. There are three segments (any of which may be empty). The first contains only improper nuclei. The second contains proper nuclei with linear chromosomes. The final segment contains proper nuclei with cyclical chromosomes of finite period. A period of any nucleus is a period of any subsequent nucleus; and in particular, if a nucleus in the third segment has minimum non-zero period,  $p$ , the minimum non-zero period of any subsequent nucleus is a divisor of  $p$ . Thus, we might depict the general structure of a model as follows (where  $m + 1$  is a multiple of  $n + 1$ ):



Another obvious question is what possible orderings the proper nuclei can have. For a start, they can have the order-type of any ordinal. To prove this, one establishes by transfinite induction that for any ordinal,  $\alpha$ , there is a classical model of arithmetic in which the non-standard numbers can be partitioned into a collection of disjoint blocks with order-type  $\alpha$ , closed under arithmetic operations. One then collapses this interpretation in such a way that each block collapses into a nucleus.

The proper nuclei need not be discretely ordered. They can also have the order-type of the rationals. To prove this, one considers a classical non-standard model of arithmetic, where the order-type of the non-standard numbers is that of the rationals. It is possible to show that these can be partitioned into a collection of disjoint blocks, closed under arithmetic operations, which themselves have the order-type of the rationals. One can then collapse this model in such a way that each of the blocks collapses into a proper nucleus, giving the result. This proof can be extended to show that any order-type that can be embedded in the rationals in a certain way, can also be the order-type of the proper nuclei. This includes  $\omega^*$  (the reverse of  $\omega$ ) and  $\omega^* + \omega$ , but *not*  $\omega + \omega^*$ . For details of all this, see Priest [1997b].

What other linear order-types proper nuclei may or may not have, is still an open question.

### 9.4 Finite Models of Arithmetic

First-order arithmetic has many classical nonstandard models, but none of these is finite. One of the intriguing features of *LP* is that it permits

finite models of arithmetic, e.g., the cycle models. For these, a complete characterisation is known.

Placing the constraint of finitude on the results of the previous section, we can infer as follows. The sequence of improper nuclei is either empty or is composed of the singletons of  $0, 1, \dots, n$ , for some finite  $n$ . There must be a finite collection of proper nuclei,  $N_1 \preceq \dots \preceq N_m$ ; each  $N_i$  must comprise a finite collection of successor cycles of some minimum non-zero finite period,  $p_i$ . And if  $1 \leq i \leq j \leq m$ ,  $p_j$  must be a divisor of  $p_i$ .<sup>143</sup>

Moreover, there are models of any structure of this form. To show this, we can generalise the construction of 9.2. Take any non-standard classical model of arithmetic. This can be partitioned into the finite collection of blocks:

$$C_0, C_{1_0}, \dots, C_{1_{k(1)}}, \dots, C_{i_0}, \dots, C_{i_{k(i)}}, \dots, C_{m_0}, \dots, C_{m_{k(m)}}$$

where  $C_0$  is either empty or is of the form  $\{0, \dots, n\}$ , each subsequent block is closed under arithmetic operations, and there are  $k(i)$  successor cycles in  $N_i$ . We now define a relation,  $x \sim y$ , as follows:

$$\begin{aligned} &(x, y \in C_0 \text{ and } x = y) \text{ or} \\ &\text{for some } 1 \leq i \leq m: \\ &\quad (\text{for some } 0 < j < k(i), x, y \in C_{i_j}, \text{ and } x = y \bmod p_i) \text{ or} \\ &\quad (x, y \in C_{i_0} \cup C_{i_{k(i)}} \text{ and } x = y \bmod p_i) \end{aligned}$$

One can check that  $\sim$  is an equivalence relation, and also that it is a congruence relation on the arithmetic operations. Hence we can construct the collapsed model.  $\sim$  leaves all members of  $C_0$  alone. For every  $i$  it collapses every  $C_{i_j}$  into a successor cycle of period  $p_i$ , and it identifies the blocks  $C_{i_0}$  and  $C_{i_{k(i)}}$ . Thus, the sequence  $C_{i_0}, \dots, C_{i_{k(i)}}$  collapses into a nucleus of size  $k(i)$ . The collapsed model therefore has exactly the required structure.<sup>144</sup>

There are many interesting questions about inconsistent models, even the finite ones, whose answer is not known. For example: how many models of each structure are there? (The behaviour of the successor function in a model does not determine the behavior of addition and multiplication, except in the tail.) Perhaps the most important question is as follows. Not all inconsistent model of arithmetic are collapses of classical models. Let  $\mathcal{M}$  be any model of arithmetic; if  $\mathcal{M}'$  is obtained from  $\mathcal{M}$  by adding extra pairs to the anti-extension of  $=$ , call  $\mathcal{M}'$  an *extension* in  $\mathcal{M}$ . If  $\mathcal{M}'$  is an extension of  $\mathcal{M}$ , monotonicity ensures that it is a model of arithmetic. Now, consider the extension of the standard model obtained by adding  $\langle 0, 0 \rangle$  to the anti-extension of  $=$ . This is not a collapsed model, since, if it were, 0 would have to have been identified with some  $x > 0$ . But then 1 would have

<sup>143</sup>It is also possible to show that each nucleus is closed under addition and multiplication.

<sup>144</sup>For further details, see Priest [1997a].

been identified with  $x' > 1$ . Hence,  $\mathbf{0}' \neq \mathbf{0}'$  would also be true in the model, which it is not. Maybe, however, each inconsistent model is the extension of a collapsed classical model. If this conjecture is correct, collapsed models can be investigated via an analysis of the classical models of arithmetic and their congruence relations.

### 9.5 *The Limitative Theorems of Metamathematics*

Let us now turn to the limitative theorems of Metamathematics in the context of  $LP$ . These are the theorems of Löwenheim-Skolem, Church, Tarski and Gödel. I will take them in that order.<sup>145</sup> In what follows, let  $P$  be the set of theorems of classical Peano Arithmetic, and let  $Q$  be any non-trivial theory that contains  $P$ .

According to the classical Löwenheim-Skolem,  $Q$  has models of every infinite cardinality but has no finite models. Moving to  $LP$  changes the situation somewhat.  $Q$  still has a model of every infinite cardinality.<sup>146</sup> But it has models of finite size too: any inconsistent model may be collapsed to a finite model merely by identifying all numbers greater than some cut-off.<sup>147</sup>

The situation with second order  $P$  is again different in  $LP$ . Classically, this is known to be categorical, having the standard model as its only interpretation. But as I noted in 9.1, the Collapsing Lemma holds for second order  $LP$ . Hence, second order  $P$  is not categorical in  $LP$ . For example, it has finite models.

Turning to Church's theorem, this says, classically, that  $Q$  is undecidable. In  $LP$ , extensions of  $Q$  may be decidable. For example, let  $\mathcal{M}$  be any finite model of  $A$  ( $= Th(\mathcal{N})$ ), and let  $Q$  be  $Th(\mathcal{M})$ . Then  $Q$  is a theory that contains  $P$ . Yet  $Q$  is decidable, as is the theory of any finite interpretation. In the language of  $\mathcal{M}$  there is only a finite number of atomic sentences; their truth values can be listed. The truth values of truth functions of these can be computed according to ( $LP$ ) truth tables, and the truth values of quantified sentences can be computed, since  $\exists x\alpha$  has the same truth value

<sup>145</sup>For a statement of these in the classical context, see Boolos and Jeffrey [1974]. This section expands on the appendix of Priest [1994].

<sup>146</sup>The standard classical proof of this adds a new set of constants,  $\{c_i, i \in I\}$ , to the language, and all sentences of the form  $c_i \neq c_j, i \neq j$ , to  $Q$ . It then uses the compactness theorem. Things are more complex in  $LP$ , since the fact that  $c_i \neq c_j$  holds in an interpretation does not mean that the denotations of these constants are distinct. After extending the language, we observe that  $c_i = c_j$  cannot be proved. We then construct a prime theory in the manner of 4.3, keeping things of this form *out*. This is then used to define an appropriate interpretation.

<sup>147</sup>Let us say that  $M$  is an *exact model* of a theory iff the truths of  $M$  are exactly the members of the theory. Classically, for complete theories, there is no difference between modelling and exact modelling. The situation for  $LP$  is more complex. It can be shown that if  $Q$  has an infinite exact model it has exact models of every greater cardinality. On the other hand, if  $Q$  has a finite model,  $\mathcal{M}$ , in which every number is denoted by a numeral,  $\mathcal{M}$  can be shown to be the only exact model of  $Q$  (up to isomorphism).

as the disjunction of all formulas of the form  $\alpha(x/a)$ , where  $a$  is in the domain of  $\mathcal{M}$ ; dually for  $\forall$ .

Tarski's Theorem: this says that  $Q$  cannot contain its own truth predicate, in the sense that even if  $Q$  is a theory in an extended language, there is no formula,  $\beta$ , of one free variable,  $x$ , such that  $\beta(x/\langle\alpha\rangle) \equiv \alpha \in Q$ , for all closed formulas,  $\alpha$ , of the language. This, too, fails for  $LP$ . Let  $\mathcal{M}$  be any (classical) model of  $P$ , let  $\mathcal{M}'$  be any finite collapse of  $\mathcal{M}$ , and let  $Q$  be  $Th(\mathcal{M}')$ . By the Collapsing Lemma,  $Q$  contains  $P$ . Since  $Q$  is decidable, it is representable in (classical)  $P$  by a formula,  $\beta$ , of one free variable,  $x$ . That is, we have :

$$\begin{aligned} \text{If } \alpha \in Q \text{ then } \beta(x/\langle\alpha\rangle) \in P \\ \text{If } \alpha \notin Q \text{ then } \neg\beta(x/\langle\alpha\rangle) \in P \end{aligned}$$

By the Collapsing Lemma, ' $P$ ' may be replaced by ' $Q$ '. If  $\alpha \in Q$ ,  $\beta(x/\langle\alpha\rangle) \in Q$ , and so  $\beta(x/\langle\alpha\rangle) \equiv \alpha \in Q$  (since  $\gamma, \delta \models_{LP} \gamma \equiv \delta$ ); and if  $\alpha \notin Q$ ,  $\neg\beta(x/\langle\alpha\rangle) \in Q$ , and so  $\beta(x/\langle\alpha\rangle) \equiv \alpha \in Q$  (since  $\neg\alpha \in Q$  and  $\neg\gamma, \neg\delta \models_{LP} \gamma \equiv \delta$ ).

There is no guarantee that  $\alpha$  and  $\beta(x/\langle\alpha\rangle)$  have the same truth value in  $\mathcal{M}'$ . In particular, then,  $\beta$  may not satisfy the  $T$ -schema in the form of a two-way rule of inference. So it might be said that  $\beta$  is not really a truth predicate. Whether or not this is so, we have already seen that there are  $Q$ s where there is a predicate satisfying this condition (though this has to be added to the language of arithmetic): the theory  $T_1$  of section 8.1.<sup>148</sup> Finally, let us turn to Gödel's undecidability theorems. A statement of the first of these is that if  $Q$  is axiomatisable then there are sentences true in the standard model that are not in  $Q$ . It is clear that this may fail in  $LP$ . Let  $\mathcal{M}$  be any finite model of arithmetic. Then if  $Q$  is  $Th(\mathcal{M})$ ,  $Q$  contains all of the sentences true in the standard model of arithmetic, but is decidable, as we have noted, and hence axiomatisable (by Craig's Theorem).

It is worth asking what happens to the "undecidable" Gödel sentence in such a theory. Let  $\beta$  be any formula that represents  $Q$  in  $Q$ . (There are such formulas, as we just saw.) Then a Gödel sentence is one,  $\alpha$ , of the form  $\neg\beta(x/\langle\alpha\rangle)$ . If  $\alpha \in Q$  then  $\neg\beta(x/\langle\alpha\rangle) \in Q$ , but  $\beta(x/\langle\alpha\rangle) \in Q$  by representability. If  $\alpha \notin Q$  then  $\neg\beta(x/\langle\alpha\rangle) \in Q$  by representability, i.e.,  $\alpha \in Q$ , so  $\beta(x/\langle\alpha\rangle) \in Q$  by representability. In either case, then,

<sup>148</sup>The construction of 8.1 can be applied to any model of arithmetic—not just the standard model—as the ground model. However, if we apply it to a finite model care needs to be exercised. The construction will not work as given, since different formulas may be coded by the same number in the model, which renders the definition of the sequence of interpretations illicit. We can switch to evaluational semantics, as in 8.2, though the construction then no longer validates the substitutivity of identicals. Alternatively, we can refrain from using numbers as names, but just augment the language with names for all sentences.

$\alpha \wedge \neg\alpha \in Q$ .

Gödel's second undecidability theorem says that the statement that canonically asserts the consistency of  $Q$  is not in  $Q$ ; this statement is usually taken to be  $\neg\beta(x/\langle\alpha_0\rangle)$ , where  $\alpha_0$  is  $\mathbf{0} = \mathbf{0}'$ , and  $\beta$  is the canonical proof predicate of  $Q$ . This also fails in  $LP$ .<sup>149</sup> Let  $Q$  be as in the previous two paragraphs. Then  $Q$  is not consistent. However, it is still the case that  $\alpha_0 \notin Q$  (provided that the collapse is not the trivial one). Consider the relationship:  $n$  is (the code of) a proof of formula (with code)  $m$  in  $Q$ . Since this is recursive, it is represented in  $A$  by a formula  $Prov(x, y)$ . If  $\alpha$  is provable in  $Q$  then for some  $n$ ,  $Prov(\mathbf{n}, \langle\alpha\rangle) \in A$  (where  $\mathbf{n}$  is the numeral for  $n$ ); thus,  $\exists xProv(x, \langle\alpha\rangle) \in A$  and so  $Q$ . If  $\alpha$  is not provable in  $Q$  then for all  $n$ ,  $\neg Prov(\mathbf{n}, \langle\alpha\rangle) \in A$ ; thus,  $\forall x\neg Prov(x, \langle\alpha\rangle) \in A$  (since  $A$  is  $\omega$ -complete) and  $\neg\exists xProv(x, \langle\alpha\rangle) \in A$  and so  $Q$ . Thus,  $\exists xProv(x, y)$  represents  $Q$  in  $Q$ . In particular, since  $\alpha_0 \notin Q$ ,  $\neg\exists xProv(x, \langle\alpha_0\rangle) \in Q$ , as required.

### 9.6 The Philosophical Significance of Gödel's Theorem

People have tried to make all sorts of philosophical capital out of the negative results provided by the limitative theorems of classical Metamathematics. As we have seen, all of these, save the Löwenheim-Skolem Theorem, fail for arithmetic based on a paraconsistent logic. Setting this theorem aside, then, nothing can be inferred from these negative results unless one has reason to rule out paraconsistent theories. At the very least, this adds a whole new dimension to the debates in question.

This is not the place to discuss all the philosophical issues that arise in this context, but let me say a little more about one of the theorems by way of illustration. Doubtless, the incompleteness result that has provoked most philosophical rumination is Gödel's first incompleteness theorem: usually in a form such as: for any axiomatic theory of arithmetic (with sufficient strength, etc.), which we can recognise to be sound, there will be an arithmetic truth—viz., its Gödel sentence—not provable in it, but which we can establish as true.<sup>150</sup> This is just false, paraconsistently. If the theory is inconsistent, the Gödel sentence may well be provable in the theory, as we have seen.

An obvious thought at this point is that if we can recognise the theory to be sound then it can hardly be inconsistent. But unless one closes the question prematurely, by a refusal to consider the paraconsistent possibility, this is by no means obvious. What *is* obvious to anyone familiar with the subject, is that at the heart of Gödel's theorem, is a paradox. The paradox concerns the sentence,  $\gamma$ , 'This sentence is not provable', where 'provable'

<sup>149</sup>It is worth noting that there are *consistent* arithmetics based on some relevant logics, notably  $R$ , for which the statement of consistency is in the theory. See Meyer [1978].

<sup>150</sup>For example, the theorem is stated in this form in Dummett [1963]; it also drives Lucas' notorious [1961], though it is less clearly stated there.

is not to be understood to mean being the theorem of some axiom system or other, but as meaning ‘demonstrated to be true’. If  $\gamma$  is provable, then it is true and so not provable. Thus we have proved  $\gamma$ . It is therefore true, and so unprovable. Contradiction. The argument can be formalised with one predicate,  $B$ , satisfying the conditions:

$$\begin{aligned} &\vdash B \langle \alpha \rangle \rightarrow \alpha \\ &\text{If } \vdash \alpha \text{ then } \vdash B \langle \alpha \rangle \end{aligned}$$

for all closed  $\alpha$ —including sentences containing  $B$ . For if  $\gamma$  is of the form  $\neg B \langle \gamma \rangle$ , then, by the first,  $\vdash B \langle \gamma \rangle \rightarrow \neg B \langle \gamma \rangle$ , and so  $\vdash \neg B \langle \gamma \rangle$ , i.e.,  $\vdash \gamma$ . Hence,  $\vdash B \langle \gamma \rangle$ , by the second.

And we *do* recognise these principles to be sound. Whatever is provable is true, by definition; and demonstrating  $\alpha$  shows that  $\alpha$  is provable, and so counts as a demonstration of this fact.<sup>151</sup>

$B$  is a predicate of numbers, but we do not have to assume that  $B$  is definable in terms of  $'$ ,  $+$  and  $\times$  using truth functions and quantifiers. The argument could be formalised in a language with  $B$  as primitive. As we saw in the previous part in connection with truth, it is quite possible to have an inconsistent theory with a predicate of this kind, where the sentences definable in terms of  $'$ ,  $+$  and  $\times$  using truth functions and quantifiers behave quite consistently.

Of course, if  $B$  is so definable, which it will be if the set of things we can prove is axiomatic, then the set of things that hold in this language *is* inconsistent. And there are reasons for supposing that this is indeed the case.<sup>152</sup> Even this does not necessarily mean that the familiar natural numbers behave strangely, however. As the model with the “point at infinity” of 9.2 showed, it is quite possible for inconsistent models to have the ordinary natural numbers as a substructure.<sup>153</sup> There are just more possibilities in Heaven and Earth than are dreamt of in a consistent philosophy.

## 10 PHILOSOPHICAL REMARKS

In previous parts I have touched occasionally on the philosophical aspects of paraconsistency. In this section I want to take up a few of the philosophical implications of paraconsistency at slightly greater length. Its major

<sup>151</sup>The paradox is structurally the same as a paradox often called the ‘Knower paradox’. In this,  $B$  is interpreted as ‘It is known that’. For references and discussion of this paradox and others of its kind, see Priest [1991b].

<sup>152</sup>See Priest [1987], ch. 3. This chapter discusses the connection between Gödel’s theorem, the paradoxes of self-reference and dialetheism at greater length.

<sup>153</sup>Though whether the theory of that particular model is axiomatisable is currently unknown.



implication is very simple. As I noted in 3.1, the absolute unacceptability of inconsistency has been deeply entrenched in Western philosophy. It is an assumption that has hardly been questioned since Aristotle. Whilst the law of non-contradiction is a traditional statement of this fact, it is *ECQ* which expresses the real *horror contradictionis*: contradictions explode into triviality. Paraconsistency challenges exactly this, and so questions any philosophical claim based on this supposed unacceptability. This does not mean that consistency cannot play a regulative function: it may still be an expected norm, departure from which requires a justification; but it can no longer provide a constraint of absolute nature. Given the centrality of consistency to Western thought, the philosophical ramifications of paraconsistency are bound to be profound, and this is hardly the place to take them all—or even some—up at great length. What I will do here is consider various objections to employing a paraconsistent logic, and explore a little some of the philosophical issues that arise in this context. In the process we will need to consider not only the purposes of logic, but also the natures of negation, denial, rational belief and belief revision.<sup>154</sup>

### 10.1 *Instrumentalism and Information*

Why, then, might one object to paraconsistent logic? Logic has many uses, and any objection to the use of a paraconsistent logic must depend on what it is supposedly being used for. One thing one may want a logic for is to draw out consequences of some information in a purely instrumental way. In such circumstances one may use any logic one likes provided that it gives appropriate results. And if the information is inconsistent, an explosive logic is hardly likely to do this.

Referring back to the list of motivations for the use of a paraconsistent logic in 2.2, drawing inferences from a scientific theory would fall into this category if one is a scientific instrumentalist. Drawing inferences from the information in a computer data base could also fall into this category. If the logic gives the right results—or at least, does not give the wrong results—use it.

The only objection that there is likely to be to the use of a paraconsistent logic in this context is that it is too weak to be of any serious use. One might note, for example, that most paraconsistent logics invalidate the disjunctive syllogism, a special case of resolution, on the basis of which many theorem-provers work.<sup>155</sup> This objection carries little weight, however. Theorem-

<sup>154</sup>Other philosophical aspects of paraconsistency are discussed in numerous places, e.g., da Costa [1982], Priest [1987], Priest *et al.* [1989], ch. 18.

<sup>155</sup>It is worth noting, however, that some theorem-provers that use resolution are not complete with respect to classical semantics. For example, to determine whether  $\alpha$  follows from the information in a data base, some theorem-provers employ a heuristic that requires them resolve  $\neg\alpha$  with something on the data base, and so on recursively. Em-

provers can certainly be based on other mechanisms.<sup>156</sup> Moreover, the inferential moves of the standard programming language PROLOG can all be interpreted validly in many paraconsistent logics (when ‘:-’ is interpreted as ‘-’).

One will often require a logic for something other than merely instrumental use. This does not mean that one is necessarily interested in truth-preservation, however. One might, for example, require a logic whose valid inferences preserve, not truth, but information. The computer case could also be an example of this. Other natural examples of this in the list of 2.2 are the fictional and counterfactual situations. By definition, truth is not at issue here.<sup>157</sup>

Information-preservation implies truth preservation, presumably, but the converse is not at all obvious, and not even terribly plausible. The information that the next flight to Sydney leaves at 3.45 and does not leave at 3.45 would hardly seem to contain the information that there is life on Mars. A paraconsistent logic is therefore a plausible one in this context.

What information, and so information-preservation, are, is an issue that is currently much discussed. One popular approach is based on the situation semantics of Barwise and Perry [1983].<sup>158</sup> This takes a unit of information (an *infor*) to be something of the form  $\langle R, a_1, \dots, a_n, s \rangle$ , where  $R$  is an  $n$ -place relation, the  $a_i$ 's are objects, and  $s$  is a sign-bit (0 or 1). A situation is a set of infons. The situations in question do not have to be veridical in any sense. In particular, they may be both inconsistent and incomplete. In fact, it is easy to see that a situation, so characterised, is just a relational *FDE* evaluation. This approach to information therefore naturally incorporates a paraconsistent logic, which may be thought of as a logic of information preservation.<sup>159</sup>

## 10.2 Negation

Another major use of logic (perhaps the one that many think of first) is in contexts where we want inference to be truth-preserving; for example,

---

employing this procedure when the data base is  $\{p, \neg p\}$  and the query is  $q$  will result in a negative answer. Such inference engines are therefore paraconsistent, though they do not answer to any principled semantics that I am aware of.

<sup>156</sup>For details of some automated paraconsistent logics, see, e.g., Blair and Subrahmanian [1988], Thistlewaite *et al.* [1988].

<sup>157</sup>One might also take the other example on that list, constitutions and other legal documents, to be an example of this. Such documents certainly contain information. And one might doubt that this information is the sort of thing that is true or false: it can, after all, be brought into effect by fiat—and may be inconsistent. However, if it is that sort of thing, legal reasoning concerning it would seem to require truth-preservation.

<sup>158</sup>See, e.g., Devlin [1991].

<sup>159</sup>It is worth noting that North American relevant logicians have very often—if not usually—thought of the *FDE* valuations information-theoretically, as *told* true and *told* false. See, e.g., Anderson *et al.* [1992], sect. 81.

where we are investigating the veridicality of some theory or other. And here, it is very natural to object to the use of a paraconsistent logic. Since truth is never inconsistent a paraconsistent logic is not appropriate.

A paraconsistent logician who thinks that truth is consistent may agree with this, in a sense. We have already seen in 7.6 how a paraconsistent logic, applied to a consistent situation, may give classical reasoning. However, a dialetheist *will* object; not to the need for truth preservation, but to the claim that truth is consistent: some contradictions are true: dialetheias.

This is likely to provoke the fiercest objections. Let me start by dividing these into two kinds: local and global. Global objections attack the possibility of dialetheias on completely general grounds. Local objections, by contrast, attack the claim that some particular claims are dialethic on grounds specific to the situation concerned.

Let us take the global objections first. Why might one think that dialetheias can be ruled out quite generally, independently of the considerations of any particular case? A first argument is to the effect that a contradiction cannot be true, since contradictions entail everything, and not everything is true. It is clear that in the context where the use of a paraconsistent logic is being defended, this simply begs the question.

Of more substance is the following objection. The truth of contradictions is ruled out by the (classical) account of negation, which is manifestly correct. The amount of substance is only slightly greater here, though: the claim that the classical account of negation is manifestly correct is just plain false.

An account of negation is a *theory* concerning the behaviour of something or other. It is sometimes suggested that it is an account of how the particle ‘not’, and similar particles in other languages, behaves. This is somewhat naive. Inserting a ‘not’ does not necessarily negate a sentence. (The negation of ‘All logicians do believe the classical account of negation’ is not ‘All logicians do not believe the classical account of negation’.) And ‘not’ may function in ways that have nothing to do with negation at all. Consider, e.g.: ‘I’m not a Pom; I’m English’, where it is connotations of what is said that are being rejected, not the literal truth.

It seems to me that the most satisfactory understanding of an account of negation is to regard it as a theory of the relationship between statements that are contradictories. Note that this by no means rules out a paraconsistent account of negation.<sup>160</sup> Even supposing that we characterise contradictories as pairs of formulas such that at least one must be true and at most one can be true—with which an intuitionist would certainly disagree—it is quite possible to have both  $\Box(\alpha \vee \neg\alpha)$  and  $\neg\Diamond(\alpha \wedge \neg\alpha)$  valid in a paraconsistent logic, as we saw in 7.3.

Anyway, whatever we take a theory of negation to be a theory *of*, it is but

---

<sup>160</sup> As Slater [1995] claims.

a *theory*. And different theories are possible. As we have already observed, Aristotle gave an account of this relationship that was quite different from the classical account as it developed after Boole and Frege. And modern intuitionists, too, give a quite different account. Which account is correct is to be determined by the usual criteria for the rational acceptability of theories. (I will say a little more about this later.) The matter is not at all obvious.

Quine is well known for his objection to non-classical logic in general, and paraconsistent logic in particular, on the ground that changing the logic (from the classical one) is ‘changing the subject’, i.e., succeeds only in giving an account of something else ([1970], p. 81). This just confuses logic, *qua* theory, with logic, *qua* object of theory. Changing one’s theory of logic no more changes what it is one is theorising about—in this case, relationships grounding valid reasoning—than changing one’s theoretical geometry changes the geometry of the cosmos. Nor does it help to suppose that logic, unlike geometry, is analytic (i.e., true solely in virtue of meanings). Whether or not, e.g., ‘There will be a sea battle tomorrow or there will not’ is analytic in this sense, is no more obvious than is the geometry of the cosmos. And changing from one theory, according to which it is analytic, to another, according to which it is not, does not change the facts of meaning.<sup>161</sup>

How plausible a paraconsistent account of negation is depends, of course, on which paraconsistent account of negation is given. As we saw in part 4, there are many. One of the simplest and most natural is provided by the relational semantics of 4.5. This is just the classical account, except that classical logic makes the *extra* assumption that all statements have *exactly* one truth value. And logicians as far back as Aristotle have questioned that assumption.<sup>162</sup>

### 10.3 Denial

Another global objection to dialetheism goes by way of a supposed connection between negation and denial. It is important to be clear about the distinction between these two things for a start. Negation is a syntactic and/or semantic feature of language. Denial is a feature of language use: it is one particular kind of force that an utterance can have, one kind of illocutionary act, as Austin put it. Specifically, it is to be contrasted with assertion.<sup>163</sup> Typically, to assert something is to express one’s belief in, or acceptance of, it (or some Gricean sophistication thereof). Typically, to deny something is to express one’s rejection of it, that is, one’s refusal to ac-

<sup>161</sup> The analogy between logic and geometry is discussed further in Priest [1997a].

<sup>162</sup> The topics of this section and the next are discussed at greater length in Priest [1999].

<sup>163</sup> Traditional logic usually drew the distinction, not in terms of saying, but in terms of judging. It can be found in these terms, for example, in the *Port-Royal Logic* of Arnauld and Nicole.

cept it (or some Gricean sophistication thereof). Clearly, if one is uncertain about a claim, one may wish neither to assert nor to deny it.

Although assertion and denial are clearly different kinds of speech act, Frege argued, and many now suppose, that denial may be reduced to the assertion of negation.<sup>164</sup> If this is correct, then dialetheism faces an obvious problem. Even if some contradictions were true, no one could ever endorse a contradiction, since they could not express an acceptance of one of the contradictories without expressing a rejection of the other.<sup>165</sup>

Frege's reduction has no appeal if we take the negation of a statement simply to be its contradictory. In asserting 'Some men are not mortal', I am not denying 'All men are mortal'. I might not even realise that these are contradictories, and neither might anyone else. And if this does not seem plausible in this simple case, just make the example more complex, and recall that there is no decision procedure for contradictories.

The reduction takes on more plausibility if we identify the negation of a sentence as that sentence prefixed by 'It is not the case that'. But even in this case, the claim that to assert a negation is invariably to deny the sentence negated appears to be false. Dialetheists who asserts both, e.g., 'The Liar sentence is true' and 'It is not the case that the Liar sentence is true', are not expressing the rejection of the former with the latter: they are simply expressing their acceptance of a certain negated sentence.<sup>166</sup> It may well be retorted that this reply just begs the question, since what is at issue is whether a dialetheist *can* do just this. This may be so; but it may now be fairly pointed out that the original objection just begs the question against the dialetheist too.

In any case, there would appear to be plenty of other examples where to assert a negation is not to deny. For example, we may be brought to see that our views are inconsistent by being questioned in Socratic fashion and thus made to assert an explicit contradiction. When this happens, we are not expressing the rejection of any view. What the questioning exposes is exactly our *acceptance* of contradictory views. We may, in the light of the questioning, come to reject one of the contradictories, and so revise our views, but that is another matter.<sup>167</sup>

To assert a negated sentence is not, then, *ipso facto* to deny the sentence negated. Some, having taken this point to heart, object on the other side of the street: dialetheists have no way of expressing some of their views, specifically their rejection of certain claims: we need take nothing a dialetheist

---

<sup>164</sup>See Frege [1919].

<sup>165</sup>For an objection along these lines, see Smiley in Priest and Smiley [1993].

<sup>166</sup>And even those who take negation to express denial must hold that there is more to the meaning of negation than this. It cannot, for example, perform that function when it occurs attached to *part* of a sentence.

<sup>167</sup>Some non-dialetheists have even argued that it may not even be rational to revise our views in some contexts. See, e.g., Prior on the paradox of the preface [1971], pp. 84f.

says as a denial.<sup>168</sup>

This objection is equally flawed. For a start, even if to assert a negated sentence is to deny it, it is certainly not the only way to deny it. One can do so by a certain shake of the head, or by the use of some other body language. A dialetheist may deny in this way. Moreover, just because the assertion of a negated sentence by a dialetheist (or even a non-dialetheist, as we have seen) may not be a denial, it does not follow that it is not. In denial, a person aims to communicate to a listener a certain mental state, that of rejection; and asserting a negated sentence with the right intonation, and in the right context, may well do exactly that—even if the person is a dialetheist.<sup>169</sup>

#### 10.4 *The Rational Acceptability of Contradictions*

This does not exhaust the possible global objections to dialetheism,<sup>170</sup> but let us move on to the local ones. These do not object to the possibility of dialetheism in general, but to particular (supposed) cases of it. We noted in 2.2 that a number of these have been proposed, which include legal dialetheias, descriptions of states of change, borderline cases of vague predications and the paradoxes of self-reference. Though the detailed reasons for endorsing dialetheism in each case are different, their general form is the same: a dialethic account of the phenomenon in question provides the most satisfactory way of handling the problems it poses. A local objection may therefore be provided by producing a consistent account of the phenomenon, and arguing this is rationally preferable. The precise issues involved here will, again, depend on the topic in question; but let us examine one issue in more detail. This will allow the illustration of a number of more general points.

The case we will look at is that of the semantic paradoxes. The background to this needs no long explanation, since a logician or philosopher who does not know it may fairly be asked where they have been this century. Certain arguments such as the Liar paradox, and many others discovered in the middle ages and around the turn of this century, appear to be sound arguments to the effect that certain contradictions employing self-reference and semantic notions are true. A dialethic account simply endorses the semantic principles in question, and thus the contradictions to which these give rise. A consistent account must find some way of rejecting the reasoning, notably by giving a different account of how the semantic apparatus

<sup>168</sup>Objections along these lines can be found in Batens [1990], and Parsons [1990]. A reply can be found in Priest [1995].

<sup>169</sup>That the same sentence may have different forces in different contexts is hardly a novel observation. For example, an utterance of 'Is the door closed', can be a question, a request or a command, depending on context, intonation, power-relationships, etc.

<sup>170</sup>Some others, together with appropriate discussion, can be found in Sainsbury [1995], ch. 6.

functions. This account must both do justice to the data, and avoid the contradictions.

Many such accounts have, of course, been offered. But they are all well known to suffer from various serious problems. For example, they may provide no independent justification for the restrictions on the semantic principles involved, and so fail to explain why we should be so drawn to the general and contradiction-producing principles. They are often manifestly contrived and/or fly in the face of other well established views. Perhaps most seriously, none of them seems to avoid the paradoxes: all seem to be subject to extended paradoxes of one variety or another.<sup>171</sup> If the global objections to dialetheism have no force, then, the dialethic position here seems manifestly superior.<sup>172</sup>

It might be said that the inconsistency of the theory is at least a *prima facie* black mark against it. This may indeed be so; but even if one of the consistent theories could find plausible replies to its problems, as long as the theory is complex and fighting a rearguard action, the dialethic account may still have a simplicity, boldness and mathematical elegance that makes it preferable.

As orthodox philosophy of science realised a long time ago, there are many criteria which are good-making for theories: simplicity, adequacy to the data, preservation of established problem-solutions, etc.; and many which are bad-making: being contrived, handling the data in an *ad hoc* way, and, let us grant, being inconsistent, amongst others.<sup>173</sup> These criteria are usually orthogonal, and may even pull in opposite directions. But when applied to rival theories, the combined effect may well be to render an inconsistent theory rationally preferable to its consistent rival.

General conclusion: a theory in some area is to be rationally preferred to its rivals if it best satisfies the standard criteria of theory choice, familiar from the philosophy of science. An inconsistent theory may be the only viable theory; and even if it is not, it may still, on the whole, be rationally preferable.<sup>174</sup>

---

<sup>171</sup> All this is documented in Priest [1987], ch. 2.

<sup>172</sup> One strategy that may be employed at this point is to argue that a dialethic theory is trivial, and hence that any other theory, even one with problems, is better. As we have seen, dialethic truth-theory is non-trivial, but one might nonetheless hope to prove that it is trivial when conjoined with other unobjectionable apparatus. Such arguments have been put forward by Denyer [1989], Smiley, in Priest and Smiley [1993], and Everett [1995] and elsewhere. Replies can be found in, respectively, Priest [1989b], Priest and Smiley [1993], and Priest [1996]. Since my aim here is to illustrate general features of the situation, I will not discuss these arguments.

<sup>173</sup> Though one might well challenge the last of these as a universal rule. There might be nothing wrong with *some* contradictions at all. See Priest [1987], sect. 13.6, and Sylvan [1992], sect. 2.

<sup>174</sup> For a longer discussion of the relationship between paraconsistency and rationality, see Priest [1987], ch. 7.

### 10.5 Boolean Negation

Another sort of local objection to some dialethic theories is based on the claim that, whatever one says about negation, there is certainly an operator that behaves in the way that Boolean negation does—call it what you like. Some paraconsistent logicians may even agree with this. (As we saw in 5.3, such an operator is definable in some of the da Costa systems.) And if the point is correct, it suffices to dispose of any dialethic account of the semantic paradoxes which endorses the *T*-schema; similarly, any account of set theory that endorses the Comprehension schema. For as I observed in the introduction to part 8, these schemas will then generate Boolean contradictions, and so entail triviality.

Someone who endorses such an account of semantics or set theory must therefore object to the claim that there is an operator that behaves as does Boolean negation. Why, after all, should we suppose this?<sup>175</sup> It might be suggested that we can simply define an operator,  $-$ , satisfying the proof theoretic principles of Boolean negation, and in particular:  $\alpha, -\alpha \vdash \beta$ . Such a suggestion would fail: the reason is simply that there is no guarantee that a connective, so characterised, has any determinate sense. The point was made by Prior [1960], who illustrated it with the operator “tonk”,  $*$ , supposedly characterised by the rules  $\alpha \vdash \alpha * \beta$ ,  $\alpha * \beta \vdash \beta$ . Such an operator induces triviality and can make no sense. Similarly, a paraconsistent logician who endorses the *T*-schema may fairly point out that the supposition that there is an operator satisfying the proof-theoretic conditions of Boolean negation induces triviality, and so makes no sense.<sup>176</sup>

The claim is theory laden, in the sense that it presupposes that the *T*-schema is correct. (The addition of such an operator need not produce triviality if only more limited machinery is present.) But any claim about what makes sense is bound to be theory-laden in a similar way. Prior’s argument, for example, presupposes the transitivity of deducibility, which may be questioned, as we have seen. The thought that Boolean negation is meaningless may initially be somewhat shocking. But the point has been argued by intuitionist logicians for many years. And though the grounds are quite different,<sup>177</sup> the paraconsistent logician sides with the intuitionist against the classical logician on this occasion.

Can we not, though, characterise Boolean negation semantically, and so show that it is a meaningful connective? The answer is, again, no; not without begging the question. How one attempts to characterise Boolean negation semantically will depend, of course, on one’s preferred sort of se-

<sup>175</sup>The following material is covered in more detail in Priest [1990].

<sup>176</sup>There may, of course, be operators that behave like Boolean negation in a limited domain. That is another matter.

<sup>177</sup>The intuitionist reason is that meaningful logical operators cannot generate statements with recognition-transcendent truth conditions, which Boolean negation does. See, e.g., Dummett [1975].



mantics. Let me illustrate the matter with the Dunn semantics. Similar considerations apply to others. With these semantics, the natural attempt to characterise Boolean negation is:

$-\alpha\rho 1$  iff it is not the case that  $\alpha\rho 1$   
 $-\alpha\rho 0$  iff  $\alpha\rho 1$

And such a characterisation makes perfectly good semantic sense. However, it does not entail that  $-$  satisfies the Boolean proof-theoretic principles. Why should one suppose, crucially, that it validates  $\alpha, -\alpha \models \beta$ ? From the characterisation, it certainly follows that for all  $\rho$ , it is not the case that  $\alpha\rho 1$  and  $-\alpha\rho 1$ ; but to infer from this that for all  $\rho$ , if  $\alpha\rho 1$  and  $-\alpha\rho 1$  then  $\beta\rho 1$  (which states that the inference is valid), just employs the principle of inference that a conditional is true if the negation of its antecedent is. And no sensible paraconsistent conditional validates this.

In other words, to insist that  $-$ , so characterised, is explosive, just begs the question against the paraconsistentist. And if it is claimed that the negation in the statement of the truth conditions is itself Boolean, and so the inference is acceptable, this again begs the question: whether there is a connective satisfying the Boolean proof-theoretic conditions is exactly what is at issue.<sup>178</sup>

### 10.6 *Logic as an Organon of Criticism*

We have now noted three reasons why one might employ a logic: as a purely instrumental means of generating consequences, as an organon of information preservation, and as an organon of truth preservation. This does not exhaust the uses for which one might employ a logic. Another very traditional one is as an organon of criticism, to force others to revise their views. One may object to the use of a paraconsistent logic in this context as follows. If one subscribes to a paraconsistent logic, then there is nothing to stop a person from accepting any inconsistency to which their views lead. Hence, paraconsistency renders logic useless in this context.<sup>179</sup>

The move from the premise that contradictions do not entail everything to the claim that there is nothing to stop a person subscribing to a contradiction is a blatant *non-sequitur*. The threat of triviality may be a reason

<sup>178</sup> I have sometimes heard it said that the logic of a metatheory must be classical. This is just false, as the existence of intuitionist metatheories serves to remind. For certain purposes a dialetheist may, in any case, use a classical metatheory. If, for example, we are trying to show a certain theory to be non-trivial, it suffices to show all the theorems have some property which not all sentences have. This might well be shown using *ZF*. As we saw in 8.6, *ZF* makes perfectly good dialethic sense.

<sup>179</sup> An objection of this kind is to be found in Popper [1963], pp. 316-7. The following is discussed at greater length in Priest [1987], ch. 7.

for revision; it is not the only reason. This is quite obvious in the case of non-dialethic paraconsistency. If a contradiction is entailed by one's views, then even though they do not explode into triviality, they are still not true. One will still, therefore, wish to revise. One may not, as in the classical case, have to revise immediately. It may not be at all clear how to revise; and in the meantime, an inconsistent but non-trivial belief set is better than no belief set at all. But the pressure will still be there to revise in due course.

The situation may be thought to change if one brings dialetheism into the picture. For the contradiction may then be true, and the pressure to revise is removed. Again, however, the conclusion is too swift. It is certainly true that showing that a person's views are inconsistent may not necessarily force a dialetheist to revise, but other things may well do so. For example, if a person is committed to something of the form  $\alpha \rightarrow \perp$ , and their views are shown to entail  $\alpha$ , there will be pressure to revise, for exactly the classical reason.<sup>180</sup>

Even if a dialetheist's views do not collapse into triviality, the inference to the claim that there is no pressure to revise is still too fast. The fact that there is no *logical* objection to holding on to a contradiction does not show there are no other kinds of objection. There is a lot more to rationality than consistency. Even those who hold consistency to be a constraint on rationality hold that there are many other such constraints. In fact, consistency is a rather weak constraint. That the earth is flat, that Elvis is alive and living in Melbourne, or, indeed, that one is Kermit the Frog, are all views that can be held consistently if one is prepared to make the right kinds of move elsewhere; but these views are manifestly irrational. For a start, there is no evidence for them; moreover, to make things work elsewhere one has to make all kinds of *ad hoc* adjustments to other well-supported views. And whatever constraints there are on rational belief—other than consistency—these work just as much on a dialetheist, and may provide pressure to revise. Not, perhaps, pressure of the *stand 'em up - knock 'em down* kind. But such would appear to be illusory in any case. As the history of ideas has shown, rational debates may be a long and drawn out business. There is no magic strategy that will always win the debate—other than employing (or at least showing) the instruments of torture.<sup>181</sup>

---

<sup>180</sup> Provided that one is not a person who believes that everything is true, then asserting  $\alpha \rightarrow \perp$  is a way of denying  $\alpha$ . A dialetheist might do this for a whole class of sentences, and so rule out contradictions occurring in certain areas, wholesale.

<sup>181</sup> Avicenna, apparently, realised this. According to Scotus, he wrote that those who deny the law of non-contradiction 'should be flogged or burned until they admit that it is not the same thing to be burned and not burned, or whipped and not whipped'. (*The Oxford Commentary on the Four Books of the Sentences*, Bk. I, Dist. 39. Thanks to Vann McGee for the reference.)

## 11 CONCLUSION

Let me conclude this essay by trying to put a little perspective into the development of paraconsistent logic. Paraconsistent and explosive accounts of validity are both to be found in the history of logic. The revolution in logic that started around the turn of the century, and which was constituted by the development and application of novel and powerful mathematical techniques, entrenched explosion on the philosophical scene. The application of the same techniques to give paraconsistent logics had to wait until after the second world war.

The period from then until about the late 1970s saw the development of many paraconsistent logics, their proof theories and semantics, and an initial exploration of their possible applications. Though there are still many open problems in these areas, as I have indicated in this essay, the subject was well enough developed by that time to permit the beginning of a second phase: the investigation of inconsistent mathematical theories and structures in their own rights. Whereas the first period was dominated by a negative metaphor of paraconsistency as damage control, the second has been dominated by a more positive attitude: let us investigate inconsistent mathematical structures, both for their intrinsic interest and to see what problems—philosophical, mathematical, or even empirical—they can be used to solve.<sup>182</sup>

Where this stage will lead is as yet anyone's guess. But let me speculate. Traditional wisdom has it that there have been three foundational crises in the history of mathematics. The first arose around the Fourth Century BC, with the discovery of irrational numbers, such as  $\sqrt{2}$ . It resulted in the overthrow of the Pythagorean doctrine that mathematical truths are exhausted by the domain of the whole numbers (and the rational numbers, which are reducible to these); and eventually, in the development of an appropriate mathematics. The second started in the Seventeenth Century with the discovery of the infinitesimal calculus. The appropriate mathematics came a little faster this time; and the result was the overthrow of the Aristotelian doctrine that truth is exhausted by the domain of the finite (or at least the potential infinite, which is a species of the finite). The third crisis started around the turn of this century, with the discovery of apparently inconsistent entities (such as the Russell set and the Liar sentence) in the foundations of logic and set theory—or at least, with the realisation that such entities could not be regarded as mere curiosities. This provided a major—perhaps the major—impetus for the development of paraconsistent logic and mathematics (as far as it has got). And the philosophical result may be the overthrow of another Aristotelian doctrine: that truth is

---

<sup>182</sup>It must be said that both stages have been pursued in the face of an attitude sometimes bordering on hostility from certain sections of the establishment logico-philosophical, though things are slowly changing.

exhausted by the domain of the consistent.<sup>183</sup>

*University of Queensland, Australia.*

## BIBLIOGRAPHY

- [Anderson and Belnap, 1975] A. Anderson and N. Belnap. *Entailment: the Logic of Relevance and Necessity*, Vol. I, Princeton University Press, Princeton, 1975.
- [Anderson et al., 1992] A. Anderson, N. Belnap and J. M. Dunn. *Entailment: the Logic of Relevance and Necessity*, Vol. II, Princeton University Press, Princeton, 1992.
- [Avron, 1990] A. Avron. Relevance and Paraconsistency—a New Approach. *Journal of Symbolic Logic*, **55**, 707–732, 1990.
- [Arruda, 1977] A. Arruda. On the Imaginary Logic of N. A. Vasil'ev', In [Arruda et al., 1977, pp. 3–24].
- [Arruda, 1980] A. Arruda. The Paradox of Russell in the System  $NF_n$ . In [Arruda et al., 1980b, pp. 1–13].
- [Arruda and Batens, 1982] A. Arruda and D. Batens. Russell's Set versus the Universal Set in Paraconsistent Set Theory. *Logique et Analyse*, **25**, 121–133, 1982.
- [Arruda et al., 1977] A. Arruda, N. da Costa and R. Chuaqui, eds. *Non-classical Logic, Model Theory and Computability*, North Holland, Amsterdam, 1977.
- [Arruda et al., 1980a] A. Arruda, R. Chuaqui and N. da Costa, eds. *Mathematical Logic in Latin America*, North Holland, Amsterdam, 1980.
- [Arruda et al., 1980b] A. Arruda, N. da Costa and A. Sette. *Proceedings of the Third Brazilian Conference on Mathematical Logic*, Sociedade Brasileira de Lógica, São Paulo, 1980.
- [Asenjo, 1966] F. G. Asenjo. A Calculus of Antimonies. *Notre Dame Journal of Formal Logic*, **16**, 103–105, 1966.
- [Asenjo and Tamburino, 1975] F. G. Asenjo and J. Tamburino. Logic of Antimonies. *Notre Dame Journal of Formal Logic*, **7**, 272–278, 1975.
- [Baaz, 1986] M. Baaz. A Kripke-type Semantics for da Costa's Paraconsistent Logic,  $C_\omega$ . *Notre Dame Journal of Formal Logic*, **27**, 523–527, 1986.
- [Barwise and Perry, 1983] J. Barwise and J. Perry. *Situations and Attitudes*, Bradford Books, MIT Press, Cambridge, MA, 1983.
- [Batens, 1980] D. Batens. Paraconsistent Extensional Propositional Logics. *Logique et Analyse*, **23**, 195–234, 1980.
- [Batens, 1989] D. Batens. Dynamic Dialectical Logics. In [Priest et al., 1989, Chapter 6].
- [Batens, 1990] D. Batens. Against Global Paraconsistency. *Studies in Soviet Thought*, **39**, 209–229, 1990.
- [Beziau, 1990] J.-Y. Beziau. Logiques Construites Suivant les Méthodes de da Costa, I: Logiques Paraconsistentes, Paracomplètes, Non-Aletheic, Construites Suivant la Première Méthode de da Costa. *Logique et Analyse*, **33**, 259–272, 1990.
- [Blair and Subrahmanian, 1988] H. A. Blair and V. S. Subrahmanian. Paraconsistent Foundations of Logic Programming. *Journal of Non-Classical Logic*, **5**, 45–73, 1988.
- [Błaszczuk, 1984] J. Błaszczuk. Some Paraconsistent Sentential Calculi. *Studia Logica*, **43**, 51–61, 1984.
- [Boolos and Jeffrey, 1974] G. Boolos and R. Jeffrey. *Computability and Logic*, Cambridge University Press, Cambridge, 1974.

---

<sup>183</sup>I am very grateful to those who read the first draft of this essay, and gave me valuable comments and suggestions: Diderik Batens, Ross Brady, Otávio Bueno, Newton da Costa, Chris Mortensen, Nicholas Rescher, Richard Sylvan, Koji Tanaka, Neil Tennant and Matthew Wilson.

- [Brady, 1982] R. Brady. Completeness Proofs for the Systems *RM3* and *BN4*. *Logique et Analyse*, **25**, 9–32, 1982.
- [Brady, 1983] R. Brady. The Simple Consistency of Set Theory Based on the Logic *CSQ*. *Notre Dame Journal of Formal Logic*, **24**, 431–439, 1983.
- [Brady, 1989] R. Brady. The Non-Triviality of Dialectical Set Theory. In [Priest *et al.*, 1989, Chapter 16].
- [Brady, 1991] R. Brady. Gentzenization and Decidability of Some Contraction-Less Relevant Logics. *Journal of Philosophical Logic*, **20**, 97–117, 1991.
- [Brady and Routley, 1989] R. Brady and R. Routley. The Non-Triviality of Extensional Dialectical Set Theory. In [Priest *et al.*, 1989, Chapter 15].
- [Brink, 1988] C. Brink. Multisets and the Algebra of Relevant Logics. *Journal of Non-Classical Logic*, **5**, 75–95, 1988.
- [Brown, 1993] B. Brown. Old Quantum Theory: a Paraconsistent Approach. *Proceedings of the Philosophy of Science Association*, **2**, 397–441, 1993.
- [Bunder, 1984] M. Bunder. Some Definitions of Negation Leading to Paraconsistent Logics. *Studia Logica*, **43**, 75–78, 1984.
- [Bunder, 1986] M. Bunder. Tautologies that, with an Unrestricted Comprehension Schema, Lead to Triviality. *Journal of Non-Classical Logic*, **3**, 5–12, 1986.
- [Carnap, 1950] R. Carnap. *Logical Foundations of Probability*. University of Chicago Press, Chicago, 1950.
- [Chang, 1963] C. Chang. Logic with Positive and Negative Truth Values. *Acta Philosophica Fennica*, **16**, 19–38, 1963.
- [Chellas, 1980] B. Chellas. *Modal Logic: an Introduction*, Cambridge University Press, Cambridge, 1980.
- [Curry, 1942] H. Curry. The Inconsistency of Certain Formal Logics. *Journal of Symbolic Logic*, **7**, 115–117, 1942.
- [Da Costa, 1974] N. Da Costa. On the Theory of Inconsistent Formal Systems. *Notre Dame Journal of Formal Logic*, **15**, 497–510, 1974.
- [Da Costa, 1982] N. Da Costa. The Philosophical Import of Paraconsistent Logic. *Journal of Non-Classical Logic*, **1**, 1–19, 1982.
- [Da Costa, 1986] N. Da Costa. On Paraconsistent Set Theory. *Logique et Analyse*, **29**, 361–371, 1986.
- [Da Costa and Alves, 1977] N. Da Costa and E. Alves. A Semantical Analysis of the Calculi  $C_n$ . *Notre Dame Journal of Formal Logic*, **18**, 621–630, 1977.
- [Da Costa and Guillaume, 1965] N. Da Costa and M. Guillaume. Négations Composées et la Loi de Peirce dans les Systems  $C_n$ . *Portugaliae Mathematica*, **24**, 201–209, 1965.
- [Da Costa and Marconi, 1989] N. Da Costa and D. Marconi. An Overview of Paraconsistent Logic in the 80s. *Journal of Non-Classical Logic* **5**, 45–73, 1989.
- [Denyer, 1989] N. Denyer. Dialetheism and Trivialisation. *Mind*, **98**, 259–268, 1989.
- [Devlin, 1991] K. Devlin. *Logic and Information*, Cambridge University Press, Cambridge, 1991.
- [D'Ottaviano and da Costa, 1970] I. D'Ottaviano and N. da Costa. Sur un Problème de Jaśkowski. *Comptes Rendus Hebdomadaires de l'Académie des Sciences, Paris*, **270A**, 1349–1353, 1970.
- [Dowden, 1984] B. Dowden. Accepting Inconsistencies from the Paradoxes. *Journal of Philosophical Logic*, **13**, 125–130, 1984.
- [Dummett, 1963] M. Dummett. The Philosophical Significance of Gödel's Theorem. *Ratio*, **5**, 140–55, 1963. Reprinted as [Dummett, 1978, Chapter 12].
- [Dummett, 1975] M. Dummett. The Philosophical Basis of Intuitionist Logic. In H. Rose and J. Shepherdson, eds. *Logic Colloquium '73*, North Holland, Amsterdam, 1975. Reprinted as [Dummett, 1978, Chapter 14].
- [Dummett, 1977] M. Dummett. *Elements of Intuitionism*, Oxford University Press, Oxford, 1977.
- [Dummett, 1978] M. Dummett. *Truth and Other Enigmas*, Duckworth, London, 1978.
- [Dunn, 1976] J. M. Dunn. Intuitive Semantics for First Degree Entailment and 'Coupled Trees'. *Philosophical Studies*, **29**, 149–168, 1976.
- [Dunn, 1979] J. M. Dunn. A theorem in 3-valued model theory with connections to number theory, type theory and relevance. *Studia Logica*, **38**, 149–169, 1979.

- [Dunn, 1980] J. M. Dunn. A Sieve for Entailments. *Journal of Philosophical Logic*, **9**, 41–57, 1980.
- [Dunn, 1988] J. M. Dunn. The Impossibility of Certain Second-Order Non-Classical Logics with Extensionality. In *Philosophical Analysis*, D. F. Austin, ed., pp. 261–79. Kluwer Academic Publishers, Dordrecht, 1988.
- [Everett, 1995] A. Everett. Absorbing Dialetheias. *Mind*, **103**, 414–419, 1995.
- [Frege, 1919] G. Frege. Negation. *Beiträge zur Philosophie des Deutschen Idealismus*, **1**, 143–157, 1919. Reprinted in translation in *Translations from the Philosophical Writings of Gottlob Frege*, P. Geach and M. Black, eds., pp. 117–135. Basil Blackwell, Oxford, 1960.
- [Goodman, 1981] N. D. Goodman. The Logic of Contradiction. *Zeitschrift für Mathematische Logik und Grundlagen der Mathematik*, **27**, 119–126, 1981.
- [Goodship, 1996] L. Goodship. On Dialethism. *Australasian Journal of Philosophy*, **74**, 153–161, 1996.
- [Haack, 1974] S. Haack. *Deviant Logic*, Cambridge University Press, Cambridge, 1974.
- [Halpern, 1986] J. Y. Halpern, ed. *Theoretical Aspects of Reasoning about Knowledge*, Morgan Kaufmann, Los Altos, 1986.
- [Jaśkowski, 1969] S. Jaśkowski. Propositional Calculus for Contradictory Deductive Systems. *Studia Logica*, **24**, 143–157, 1969.
- [Kaye, 1991] R. Kaye. *Models of Peano Arithmetic*, Clarendon Press, Oxford, 1991.
- [Kotas and da Costa, 1978] J. Kotas and N. da Costa. On the Problem of Jaśkowski and the Logics of Lukasiewicz. In [Arruda *et al.*, 1977, pp. 127–139].
- [Kotas and da Costa, 1989] J. Kotas and N. da Costa. Problems of Modal and Discussive Logic. In [Priest *et al.*, 1989, Chapter 8].
- [Loparić, 1986] A. Loparić. A Semantical Study of some Propositional Calculi. *Journal of Non-Classical Logic*, **3**, 73–95, 1986.
- [Loparić and da Costa, 1984] A. Loparić and N. da Costa. Paraconsistency, Paracompleteness and Valuations. *Logique et Analyse*, **27**, 119–131, 1984.
- [Lucas, 1961] J. R. Lucas. Minds, Machines and Gödel. *Philosophy*, **36**, 112–127, 1961. Reprinted in *Minds and Machines*, A. Anderson, ed., pp. 43–59. Prentice Hall, Englewood Cliffs, 1964.
- [Lukasiewicz, 1971] J. Lukasiewicz. On the Principle of Contradiction in Aristotle. *Review of Metaphysics*, **24**, 485–509, 1971.
- [Marconi, 1984] D. Marconi. Wittgenstein on Contradiction and the Philosophy of Paraconsistent Logic. *History of Philosophy Quarterly*, **1**, 333–352, 1984.
- [Martin, 1986] C. Martin. William's Machine. *Journal of Philosophy*, **83**, 564–572, 1986.
- [Meyer, 1978] R. K. Meyer. Relevant Arithmetic. *Bulletin of the Section of Logic, Polish Academy of Sciences*, **5**, 133–137, 1978.
- [Meyer and Mortensen, 1984] R. K. Meyer and C. Mortensen. Inconsistent Models for Relevant Arithmetics. *Journal of Symbolic Logic*, **49**, 917–929, 1984.
- [Meyer and Routley, 1972] R. K. Meyer and R. Routley. Algebraic Analysis of Entailment, I. *Logique et Analyse*, **15**, 407–428, 1972.
- [Meyer *et al.*, 1979] R. K. Meyer, R. Routley and J. M. Dunn. Curry's Paradox. *Analysis*, **39**, 124–128, 1979.
- [Mortensen, 1980] C. Mortensen. Every Quotient Algebra for  $C_1$  is Trivial. *Notre Dame Journal of Formal Logic*, **21**, 694–700, 1980.
- [Mortensen, 1984] C. Mortensen. Aristotle's Thesis in Consistent and Inconsistent Logics. *Studia Logica*, **43**, 107–116, 1984.
- [Mortensen, 1987] C. Mortensen. Inconsistent Nonstandard Arithmetic. *Journal of Symbolic Logic*, **52**, 512–518, 1987.
- [Mortensen, 1989] C. Mortensen. Anything is Possible. *Erkenntnis*, **30**, 319–337, 1989.
- [Mortensen, 1995] C. Mortensen. *Inconsistent Mathematics*, Kluwer Academic Publishers, Dordrecht, 1995.
- [Nelson, 1959] D. Nelson. Negation and Separation of Concepts in Constructive Systems. In *Constructivity in Mathematics*, A. Heyting, ed., pp. 208–225 North Holland, Amsterdam, 1959.
- [Parsons, 1990] T. Parsons. True Contradictions. *Canadian Journal of Philosophy*, **20**, 335–353, 1990.

- [Peña, 1984] L. Peña. Identity, Fuzziness and Noncontradiction. *Noûs*, **18**, 227–259, 1984.
- [Peña, 1989] L. Peña. Verum et Ens Convertuntur. In [Priest *et al.*, 1989, Chapter 20].
- [Popper, 1963] K. R. Popper. *Conjectures and Refutations*, Routledge and Kegan Paul, London, 1963.
- [Prawitz, 1965] D. Prawitz. *Natural Deduction*, Almqvist & Wiksell, Stockholm, 1965.
- [Priest, 1979] G. Priest. Logic of Paradox. *Journal of Philosophical Logic*, **8**, 219–241, 1979.
- [Priest, 1980] G. Priest. Sense, Truth and *Modus Ponens*. *Journal of Philosophical Logic*, **9**, 415–435, 1980.
- [Priest, 1982] G. Priest. To Be and Not to Be: Dialectical Tense Logic. *Studia Logica*, **41**, 249–268, 1982.
- [Priest, 1987] G. Priest. *In Contradiction*, Martinus Nijhoff, the Hague, 1987.
- [Priest, 1989a] G. Priest. Dialectic and Dialetheic. *Science and Society*, **53**, 388–415, 1989.
- [Priest, 1989b] G. Priest. Denyer's  $\$$  Not Backed by Sterling Arguments. *Mind*, **98**, 265–268, 1989.
- [Priest, 1990] G. Priest. Boolean Negation and All That. *Journal of Philosophical Logic*, **19**, 201–215, 1990.
- [Priest, 1991a] G. Priest. Minimally Inconsistent *LP*. *Studia Logica*, **50**, 321–331, 1991.
- [Priest, 1991b] G. Priest. Intensional Paradoxes. *Notre Dame Journal of Formal Logic*, **32**, 193–211, 1991.
- [Priest, 1992] G. Priest. What is a Non-Normal World? *Logique et Analyse*, **35**, 291–302, 1992.
- [Priest, 1994] G. Priest. Is Arithmetic Consistent? *Mind*, **103**, 321–331, 1994.
- [Priest, 1995] G. Priest. Gaps and Gluts: Reply to Parsons. *Canadian Journal of Philosophy*, **25**, 57–66, 1995.
- [Priest, 1996] G. Priest. Everett's Trilogity. *Mind*, **105**, 631–647, 1996.
- [Priest, 1997a] G. Priest. On a Paradox of Hilbert and Bernays. *Journal of Philosophical Logic*, **26**, 45–56, 1997.
- [Priest, 1997b] G. Priest. Inconsistent Models of Arithmetic: Part I, Finite Models. *Journal of Philosophical Logic*, **26**, 223–235, 1997.
- [Priest, 1998] G. Priest. To be *and* Not to Be — That is the Answer. On Aristotle on the Law of Non-contradiction. *Philosophiegeschichte und Logische analyse*, **1**, 91–130, 1998.
- [Priest, 1999] G. Priest. What Not? A Defence of a Dialetheic Theory of Negation. In *Negation*, D. Gabbay and H. Wansing, eds., pp. 101–120. Kluwer Academic Publishers, Dordrecht, 1999.
- [Priest, forthcoming:a] G. Priest. On Alternative Geometries, Arithmetics and Logics; a Tribute to Lukasiewicz', In *Proceedings of the Conference Lukasiewicz in Dublin*, M. Baghrarian, ed., to appear.
- [Priest, forthcoming:b] G. Priest. Inconsistent Models of Arithmetic: Part II, the General Case. *Journal of Symbolic Logic*, to appear.
- [Priest *et al.*, 1989] G. Priest, R. Routley and G. Norman, eds. *Paraconsistent Logic: Essays on the Inconsistent*, Philosophia Verlag, Munich, 1989.
- [Priest and Smiley, 1993] G. Priest and T. Smiley. Can Contradictions be True? *Proceedings of the Aristotelian Society, Supplementary Volume*, **65**, 17–54, 1993.
- [Priest and Sylvan, 1992] G. Priest and R. Sylvan. Simplified Semantics for Basic Relevant Logics. *Journal of Philosophical Logic*, **21**, 217–232, 1992.
- [Prior, 1960] A. Prior. The Runabout Inference Ticket. *Analysis*, **21**, 38–39, 1960. Reprinted in *Philosophical Logic*, P. Strawson, ed. Oxford University Press, Oxford, 1967.
- [Prior, 1971] A. Prior. *Objects of Thought*, Oxford University Press, Oxford, 1971.
- [Pynko, 1995a] A. Pynko. Characterising Belnap's Logic via De Morgan Laws. *Mathematical Logic Quarterly*, **41**, 442–454, 1995.
- [Pynko, 1995b] A. Pynko. Priest's Logic of Paradox. *Journal of Applied and Non-Classical Logic*, **2**, 219–225, 1995.

- [Quine, 1970] W. V. O. Quine. *Philosophy of Logic*, Prentice-Hall, Englewood Cliffs, 1970.
- [Rescher, 1964] N. Rescher. *Hypothetical Reasoning*, North Holland Publishing Company, Amsterdam, 1964.
- [Rescher, 1969] N. Rescher. *Many-valued Logic*, McGraw-Hill, New York, 1969.
- [Rescher and Brandom, 1980] N. Rescher and R. Brandom. *The Logic of Inconsistency*, Basil Blackwell, Oxford, 1980.
- [Rescher and Manor, 1970–71] N. Rescher and R. Manor. On Inference from Inconsistent Premises. *Theory and Decision*, **1**, 179–217, 1970–71.
- [Restall, 1992] G. Restall. A Note on Naive Set Theory in *LP*. *Notre Dame Journal of Formal Logic*, **33**, 422–432, 1992.
- [Restall, 1993] G. Restall. Simplified Semantics for Relevant Logics (and Some of Their Rivals). *Journal of Philosophical Logic*, **22**, 481–511, 1993.
- [Restall, 1995] G. Restall. Four-Valued Semantics for Relevant Logics (and Some of Their Rivals). *Journal of Philosophical Logic*, **24**, 139–160, 1995.
- [Rosser and Turquette, 1952] J. B. Rosser and A. R. Turquette. *Many-valued Logics*, North Holland, Amsterdam, 1952.
- [Routley, 1978] R. Routley. Semantics for Connexive Logics, I. *Studia Logica*, **37**, 393–412, 1978.
- [Routley, 1979] R. Routley. Dialectical Logic, Semantics and Metamathematics. *Erkenntnis*, **14**, 301–331, 1979.
- [Routley, 1980a] R. Routley. Problems and Solutions in Semantics of Quantified Relevant Logics. In [Arruda *et al.*, 1980a, pp. 305–340].
- [Routley, 1980b] R. Routley. Ultralogic as Universal. Appendix I of *Exploring Meinong's Jungle and Beyond*, Research School of Social Sciences, Australian National University, Canberra, 1980.
- [Routley, 1984] R. Routley. The American Plan Completed; Alternative Classical-Style Semantics, without Stars, for Relevant and Paraconsistent Logics. *Studia Logica*, **43**, 131–158, 1984.
- [Routley, 1989] R. Routley. Philosophical and Linguistic Inroads: Multiply Intensional Relevant Logics. In *Directions in Relevant Logic*, J. Norman and R. Sylvan, eds. Ch. 19. Kluwer Academic Publishers, Dordrecht, 1989.
- [Routley and Loparić, 1978] R. Routley and A. Loparić. A Semantical Analysis of Arruda-da Costa *P* Systems and Adjacent Non-Replacement Systems. *Studia Logica*, **37**, 301–320, 1978.
- [Routley and Loparić, 1980] R. Routley and A. Loparić. Semantics for Quantified and Relevant Logics without Replacement. In [Arruda *et al.*, 1980b, pp. 263–280].
- [Routley and Meyer, 1973] R. Routley and R. K. Meyer. The Semantics of Entailment, I. In *Truth, Syntax and Modality*, H. Leblanc, ed. North Holland, Amsterdam, 1973.
- [Routley *et al.*, 1982] R. Routley, V. Plumwood, R. K. Meyer and R. Brady. *Relevant Logics and Their Rivals*, Vol. I, Ridgeview, Atascadero, 1982.
- [Routley and Routley, 1972] R. Routley and V. Routley. The Semantics of First Degree Entailment. *Noûs*, **6**, 335–359, 1972.
- [Routley and Routley, 1989] R. Routley and V. Routley. Moral Dilemmas and the Logic of Deontic Notions. In [Priest *et al.*, 1989, Chapter 23].
- [Sainsbury, 1995] M. Sainsbury. *Paradoxes*, (2nd ed.), Cambridge University Press, Cambridge, 1995.
- [Schotch and Jennings, 1980] P. Schotch and R. Jennings. Inference and Necessity. *Journal of Philosophical Logic*, **9**, 327–340, 1980.
- [Slaney, 1989] J. Slaney. *RWX* is not Curry Paraconsistent. In [Priest *et al.*, 1989, Chapter 17].
- [Slater, 1995] B. Slater. Paraconsistent Logics? *Journal of Philosophical Logic*, **24**, 451–454, 1995.
- [Smiley, 1959] T. J. Smiley. Entailment and Deducibility. *Proceedings of the Aristotelian Society*, **59**, 233–254, 1959.
- [Sylvan, 1992] R. Sylvan. Grim Tales Retold: How to Maintain Ordinary Discourse about—and Despite—Logically Embarrassing Notions and Totalities. *Logique et Analyse*, **35**, 349–374, 1992.



- [Sylvan, 2000] R. Sylvan. A preliminary western history of sociative logics. In *Sociative Logics and their Applications: Essays by the late Richard Sylvan*, D. Hyde and G. Priest, eds. Ashgate Publishers, Aldershot, 2000.
- [Tennant, 1980] N. Tennant. A Proof-Theoretic Approach to Entailment. *Journal of Philosophical Logic*, **9**, 185–209, 1980.
- [Tennant, 1984] N. Tennant. Perfect Validity, Entailment and Paraconsistency. *Studia Logica*, **43**, 181–200, 1984.
- [Tennant, 1987] N. Tennant. *Anti-Realism and Logic: Truth as Eternal*, Clarendon Press, Oxford, 1987.
- [Tennant, 1992] N. Tennant. *Autologic*, Edinburgh University Press, Edinburgh, 1992.
- [Thistlewaite *et al.*, 1988] P. Thistlewaite, M. McRobbie and R. K. Meyer. *Automated Theorem-Proving in Non-Classical Logics*, John Wiley & Sons, New York, 1988.
- [Urbas, 1989] I. Urbas. Paraconsistency and the *C*-systems of da Costa. *Notre Dame Journal of Formal Logic*, **30**, 583–597, 1989.
- [Urbas, 1990] I. Urbas. Paraconsistency. *Studies in Soviet Thought*, **39**, 343–354, 1990.
- [Urquhart, 1984] A. Urquhart. The Undecidability of Entailment and Relevant Implication. *Journal of Symbolic Logic*, **49**, 1059–1073, 1984.
- [White, 1979] R. White. The Consistency of the Axiom of Comprehension in the Infinite-Valued Predicate Logic of Lukasiewicz. *Journal of Philosophical Logic*, **8**, 509–534, 1979.
- [Wittgenstein, 1975] L. Wittgenstein. *Philosophical Remarks*, Basil Blackwell, Oxford, 1975.



# INDEX

- $F(\tau)$ , 270
- $G(\tau)$ , 270
- $HRM(i_1, \dots, i_n)$ , 261
- $PRM(i_1, \dots, i_n)$ - $\lambda$ -terms, 265
- $R_{\rightarrow}$ , 229
- $T_{\rightarrow}$ , 229
- $[Y/x]X$ , 241
- $\Lambda^-$ , 268
- !, modality in linear logic, 118
- $\beta$ -normal form, 242
- $\beta$ -redex, 242
- $\beta$ -reduction, 241, 242
- $\beta\eta$ -normal form, 242
- $\beta\eta$ -reduction, 242
- o, combination of information, 63
- o, fusion, 12
- $\eta$ -normal form, 243
- $\eta$ -redex, 242
- $\eta$ -reduction, 242
- $\forall$ -elimination, 14
- $\forall$ -introduction, 14
- $\gamma$ , *see* disjunctive syllogism
- $\lambda I$ -calculus, 19
- $\lambda$ -depth, 270
- $\lambda$ -terms, 240
- $\rightarrow$  as  $\leq$ , 326
- $\Box$ , necessity, 13
- $\sim$ , relational converse, 115
- $dapn(\tau)$ , 270
- $dcp(\tau)$ , 270
- $dn(\tau)$ , 270
- $do(\tau)$ , 270
- Łukasiewicz generalisations, 307
- Łukasiewicz, J., 295, 306
  
- abstraction, 240
- accessibility relation, 138–141, 146, 190, 194, 200
- accessible, 139, 146
- Ackermann, W., 16
- adjunction, 10
- algebraic, 147, 190
- algebraic BZL-realization, 199
- algebraic logics, 336
- algebraic realization, 137, 141, 143, 145, 147, 148, 174, 191, 194
- algebraic realization for first-order **OL**, 172
- algebraic semantics, 137, 147, 148, 152, 153, 176, 189, 195, 198, 312
- algebraically adequate, 145
- algebraically complete realization, 162
- Amemiya, I., 164
- Amemiya–Halperin Theorem, 165
- Amemiya–Halperin’s Theorem, 166
- Anderson, A. R., 1–5, 7, 9, 11, 14–17, 86, 88
- Anderson, C. A., 295
- angle bisecting condition, 171
- application, 231
- arithmetic, 366
  - Peano, 42
  - relevant, 41–44
- Arruda, A., 295
- Asenjo, F. G., 295
- assertion, 8
- atom, 167
- atomic, 170
- atomic types, 234
- atomicity, 170
- attribute, 172
- Austin, J. L., 4

- Avron, A., 18  
 axiom, 86  
 axiomatizable logic, 158
- B**, the logic, 76–78, 82  
**B**, 155  
**B**-realization, 155  
**B**-realization, 156, 157  
**B**<sup>+</sup>, the logic, 12, 75, 77  
 Barwise, J., 114  
 basic logic, 193, 213, 214, 216, 222  
 basic orthologic, 219, 222  
 Batens, D., 289, 349  
 Battilotti, G., 213, 214, 224, 226  
 Bell, J. L., 164, 165  
 Belnap extension lemma, 40  
 Belnap, N. D., 1–5, 7, 9, 11, 14–17, 86, 88, 100  
 Beltrametti, E., 135, 179  
 Bennett, M. K., 184  
 Birkhoff, G., 129–132, 134  
 Birkhoff, S., 179  
**B**<sup>o</sup>-realization, 157  
 Bohr, N., 134  
 Boolean algebra, 130, 148, 153, 187  
 Boolean negation, 384  
 Boolean-valued models, 177, 178  
 Born probability, 180  
 Born rule, 180  
 bound variables, 241  
**BQ**, the logic, 82  
 bracket abstraction, 233  
 Brady, R. T., 100  
 branching counter machine, 94  
 Brouwer–Zadeh lattice, 183  
 Brouwer–Zadeh poset, 183  
 Brouwer–Zadeh logics, 193  
 Bugajski, S., 179  
 Buridan, J., 294  
 Busch, P., 179  
**BZ**<sup>3</sup>-lattice, 202  
 BZ-frame, 195  
 BZ-lattice, 193, 195
- BZ**-poset, 183
- calculus for orthologic, 220  
 canonical model, 160, 162, 163, 170, 200, 208  
 canonical realization, 163  
 Cassinelli, G., 135  
 Cattaneo, G., 179, 183, 196  
 Cauchy sequences, 131  
 Chang, C. C., 185, 210  
 Chovanec, F., 184  
 Church, A., 7, 8, 16, 19  
 classical relevance logics, 7  
 Clinton, Bill, 118  
 closed subspace, 131–133, 135–137, 146, 176, 179, 203  
 closed term, 241  
 coincidence lemma, 174  
 collinearity, 102  
 combinators, 229  
 combinatory completeness, 233  
 combinatory logic, 232  
 compatibility, 148, 160  
 compatible, 148, 167, 170  
 complete, 154, 170  
 complete Boolean algebra, 177  
 complete BZ-lattice, 198  
 complete extension, 153  
 complete involutive bounded lattice, 182  
 complete ortholattice, 165, 166  
 complete orthomodular lattice, 163, 177  
 completeness, 175  
 completeness theorem, 160, 161, 175, 191, 196, 200, 208  
 completion by cuts, 166  
 conditional, 2, 354  
   counterfactual, 6  
 conditional connectives, 315  
 configuration, 158  
 confinement, 14  
 confusion  
   about ‘o’, 112

- about the logic **B**, 76
- between kinds of consequence, 74
- use-mention, 3
- connexive logic, 32
- consecution calculus, 86
- consequence, 141, 172, 194, 195, 210
- consequence in a realization, 138, 141, 174
- conservative extension, 34
- consistency, 159
- consequence in a given realization, 205
- constant domain, 173
- context, 213, 221
- continuous lattice, 60
- contraction, 86, 87, 100, 218, 219
- contradiction, 7
- converse
  - of a relation, 115
- Correspondence Thesis, 4
- Costa, N. C. A. da, 289
- counterfactual conditional, 149
- covering property, 170
- covers, 170
- cube, 217
- cube of logics, 213, 218
- Curry's Lemma, 88
- Curry-Howard isomorphism, 237
- cut rule, 47, 87
- cut-elimination, 87, 213, 221, 222
- cut-elimination theorem, 213
- cut-formula, 221
- Cutland, N., 213, 223
  
- Da Costa's  $C$ -system, 318
- da Costa's  $C_1$ , 305
- da Costa's  $C_\omega$ , 304
- da Costa, N. C. A., 178
- Dalla Chaiara, M. L., 204
- Dalla Chiara, M. L., 155, 178, 184
- Davies, E. B., 179
- de Morgan monoid, 60
  
- word problem, 101
- decidability, 100, 157
- decidability (of paraconsistent logic), 328
- decidable, 196
- deduction
  - 'Official', 17
  - relevant, 15
- deduction theorem, 8, 14, 15, 153, 161
  - modal, 17
  - modal relevant, 18
  - relevant, 16
- deductive closure, 159
- definition of the canonical model, 175
- demodaliser, 9
- denial, 380
- derivability, 159, 207
- derivable, 159
- derivation, 159, 207
- description operator, 204
- dialetheism, 291
- Dickson's theorem, 92
- difference poset, 184
- discursive implication, 317
- Dishkant, H., 138
- disjunctive syllogism ( $\gamma$ )
  - admissibility in **R**, 36
- disjunctive syllogism ( $\gamma$ ), 1, 7, 31, 33, 47
  - admissibility **RQ**, 39
  - admissibility in  $\mathbf{R}^\omega$ , 41
  - admissibility in  $\mathbf{R}^{\#\#}$ , 43
  - admissibility in **R**, 34, 35, 39, 45
  - and Boolean negation, 82
  - in the metalanguage, 36
  - inadmissibility in  $\mathbf{R}^{\#}$ , 43
  - the Lewis 'Proof', 32
- Disjunctive syllogism (DS), 294
- display logic, 100
- distribution

- of conjunction over disjunction, 10, 29, 37, 66, 93, 95, 97, 118
- distributive laws, 137
- distributivity, 133, 153
- domain, 172, 173
- domain of individuals, 173
- domains of certainty, 196
- double negation, 12
- Dummett, M., 213, 223
- Dunn, J. M., 46, 48, 196
- Dvurečenskij, A., 184
- Dwyer, R., 10
  
- E**, the logic, 1, 2, 5, 7, 9, 11–13, 16, 19–21, 23, 25–27, 30, 31, 34–36, 39, 47–49, 75, 77, 80, 84–86, 92, 93, 100, 102, 108, 109, 113, 116, 117
- E**, the logic
  - undecidability, 102
- $E_{\rightarrow}$ , the logic, 9–11, 17–19
- E<sub>fde</sub>**, the logic, 27, 54
- effect, 190, 203
- effect algebra, 184, 189, 205
- effect algebras, 204
- effects, 180, 181, 192, 209
- Einstein, A., 134
- elementary (first-order) property, 162
- elementary class, 163
- elementary substructure, 164
- elimination of the cuts, 220
- elimination theorem, 87
- entailment, 2, 3, 151
  - non-transitive, 32
- entailment in the algebraic semantics, 151
- entailment-connective, 151
- epistemic logics, 155
- EQ**, the logic, 13, 47, 48, 82
- event-state systems, 135
- events, 130
  
- ex contradictione quodlibet* (ECQ), 288
- excluded middle principle, 130
- existential quantifier, 14
- expansion, 13
- experimental proposition, 130
- experimental proposition, 129, 131, 132, 136, 179
- exponentials, 224
  
- f*, the false constant, 12
- Faggian, C., 213, 226
- FDE, 333
- filtration, 297
- Finch, P. D., 149
- Fine, K., 66, 163
- finite model property, 157, 196
- Finkelstein, A., 135
- first-degree entailment, 60
- formula
  - signed, 100
- formulas-as-types isomorphism, 237
- Foulis, D. J., 135, 136, 179, 184
- free variable, 241
- full cut, 223
- functional character, 229
- fusion, 12
- fuzzy, 180
- fuzzy accessibility relation, 194
- fuzzy complement, 183, 193, 197
- fuzzy intuitionistic logics, 193
- fuzzy negation, 182, 196, 200
- fuzzy-accessible, 194
- fuzzy-intuitionistic semantics, 196
- fuzzy-like negation, 193
- fuzzy-orthogonal set, 194
  
- Gödel's Theorem, 375
- generalisation, 14
- generalized complement, 205
- Gentzen system, 86, 88, 89, 92
- Gentzen, G., 158, 220
- Gibbins, P., 213, 223
- Girard's linear logic, 193

- Girard's linear negation, 213  
 Girard's negation, 221  
 Girard, J.-Y., 118, 213, 219  
 Giuntini, R., 183–185, 189, 194, 202, 204  
 Goldblatt, R. H., 155, 158, 162, 163, 165  
 grammar, 3  
 Greechie diagram, 167  
 Greechie, R. J., 135, 166, 167, 182  
 Grice, H. P., 6  
 Gudder, S. P., 135, 182, 189  
 guinea pigs, 5  
  
**H**, the logic, 7, 12, 77  
**H**<sup>+</sup>, the logic, 12  
**H**<sub>→</sub>, the logic, 10, 15, 17, 22  
 Haack, S., 6  
 Halperin, I., 164  
 Hardegree, G. H., 149  
 Hegel, G., 292  
 Henkin constructions, 302  
 Heraklitus, 292  
 Herbrand–Tarski, 153, 161  
 heredity, 64  
 hereditary right maximal terms, 261  
 Heyting algebra, 314  
 Hilbert lattice, 171  
 Hilbert quantum logic, 166  
 Hilbert space, 131, 135–137, 145, 150, 165, 170, 180, 190, 203, 209  
 Hilbert space lattice, 135  
 Hilbert systems, 7  
 Hilbert–Bernays, 158  
 Hilbert-space realizations, 190  
 Hilbertian space, 171  
  
 identity and function symbols (paraconsistent logic), 338  
 identity axiom, 86  
 implication, 2  
 implication-connective, 146–148, 150  
 import-export condition, 148  
  
 incompatible quantities, 133, 136  
 indistinguishability relation, 178  
 individual, 173  
 individual concept, 173, 176  
 inductive extension, 174  
 inference ticket, 9  
 infimum, 136, 137, 172, 182  
 infinitary conjunction, 197  
 infinitary disjunction, 198  
 information, 63  
 inhabitant, 234, 243  
 inner product, 131, 132, 146, 164  
 interpretation, 172, 177  
 intuitionistic accessibility relation, 194  
 intuitionistic complement, 183, 197  
 intuitionistic logic, 12, 139, 155, 213  
 intuitionistic logic (**H**), 7  
 intuitionistic orthogonal set, 195  
 intuitionistically-accessible, 194  
 involutive bounded lattice, 189, 191  
 involutive bounded poset, 183, 188  
 involutive bounded poset (lattice), 181  
 irreducibility, 170  
 irreducible, 170  
  
 Jauch, J., 135  
 Jaśkowski, S., 295  
  
**K**, the logic, 45–47  
 Kôpka, F., 184  
 Kalmbach, G., 147, 166  
**KB**, 191  
 Keating, P., 115  
 Keller, H. A., 171  
 Klaua, D., 196  
 Kleene condition, 181  
 Kleene, S. C., 306  
**KR**, the logic, 102–104, 107, 108  
**KR**, the logic  
     undecidability, 102  
 Kraus, S., 179

- Kripke, S., 86  
 Kripke-style semantics, 138  
 Kripkean models, 155  
 Kripkean realization, 139–143, 145,  
     149, 174, 194  
 Kripkean realization for (first-order)  
     **OL**, 173  
 Kripkean realizations, 155  
 Kripkean semantics, 138, 139, 143,  
     147, 148, 150, 152, 156,  
     160, 190, 191, 194–197
- Lahti, P., 179  
 lambda calculus, 229  
 lambda reductions, 247  
 lambda terms, 240  
 lattice, 182  
     de Morgan, 59  
**LC**, the logic, 79  
 Leibniz' principle of indiscernibles,  
     179  
 Leibniz-substitutivity principle, 178  
 lemma of the canonical model, 161,  
     176, 201  
 Lemmon, E. J., 22  
 length of a lattice, 170  
 Lewis, C. I., 293  
 Lewis, D. K., 6  
 Lindenbaum, 153  
 Lindenbaum property, 154  
 line, 102  
 linear combinations, 131  
 linear logic, 46, 118, 222, 224  
 linear orthologic, 213  
 Inf, *see* long normal form  
 logic  
     deviant, 6  
     the One True, 15  
 logic *R*, 324  
 logical consequence, 138, 141, 143–  
     145, 152, 156, 172, 174,  
     194, 195, 205, 210  
 logical theorem, 159  
 logical truth, 138, 141, 143, 156,  
     172, 174, 194, 195, 205,  
     210  
 long, 270  
 long normal form, 246  
**LP**, 333  
**LR**, the logic, 93–95, 99  
**LR<sup>+</sup>**, the logic, 93  
**LRW**, the logic, 99, 118  
 Ludwig, G., 179  
 Łukasiewicz axiom, 187  
 Łukasiewicz quantum logic, 209  
 Łukasiewicz' infinite many-valued  
     logic, 193  
 Łukasiewicz' infinite-many-valued  
     logic, 185
- Mackey, G., 134  
 MacNeille completion, 163, 166,  
     182, 183, 190  
 Mangani, P., 186  
 many-valued conditionals, 319  
 many-valued logics, 332  
 Martin, C., 294  
 Martin, Errol, 10, 48  
 maximal Boolean subalgebras, 167  
 maximal filter, 153  
 McLaren, 166  
 metavaluation, 37  
 metrically complete, 131  
 metrically complete, 165  
 Meyer, R. K., 10, 46, 48, 86, 101  
 Minari, P., 144  
 mingle, 13  
 minimal quantum logic, 136  
 minimally inconsistent, 349  
 minimally inconsistent, 349  
 Mittelstaedt, 179  
 Mittelstaedt, P., 135, 149  
 mix rule, 87  
 mixtures, 135  
 modal deduction theorem, 17  
 modal interpretation, 155, 157  
 modal operators, 339



- modal relevant deduction theorem, 18  
 modal translation, 156, 191  
 modality, 2  
 model, 138, 141, 152  
 modular lattice, 105  
*modus ponens*, 20, 32  
*modus ponens*, 15–17, 19, 31, 37–39  
     for the material conditional, 47  
 Moh Shaw-Kwei, 8, 16  
 monotonicity lemma, 352  
 Morash, R.P., 171  
 Mortensen, C., 288  
 multisets, 214, 215  
 MV-algebra, 185–188, 210, 211  
  
 natural deduction, 21, 158  
 necessity, 2, 11, 13  
 necessity operator, 198  
 negation, 7, 12, 378  
     Boolean, 12, 81, 82, 102  
     minimal, 12  
 negative domain, 197, 203  
 Nicholas of Cusa, 292  
 Nishimura, H., 213  
 Nisticò, G., 183, 196  
 non-adjunction, 299  
 non-adjunctive logics, 330  
 normal form, 239  
 normal worlds, 322  
**NR**, the natural deduction system, 21  
 null space, 133  
  
 observable, 146  
 ‘Official deduction’, 17  
 Once  $(j_1, \dots, j_k)$ , 255  
 Once<sup>+</sup>  $(j_1, \dots, j_k)$ , 256  
 Once<sup>-</sup>  $(j_1, \dots, j_k)$ , 255  
 One True Logic, 15  
 operational, 214  
 operational rule, 86  
  
 Organon, 293  
 Orlov, 295  
 Orlov, I., 7  
 ortho-valued universe, 177  
 ortho-pair realization, 197, 203  
 ortho-pair semantics, 199  
 ortho-valued (set-theoretical) universe  
     V, 177  
 ortho-valued models, 177  
 orthoalgebra, 184, 185, 204, 205  
 orthoarguesian law, 167  
 orthocomplement, 132, 137, 139, 213  
 orthocomplemented orthomodular lattice, 133  
 orthoframe, 139, 140, 150, 155, 156, 165, 166, 194, 197, 198, 201  
 orthogonal complement, 132  
 ortholattice, 137, 142, 147, 153, 162, 164, 166, 182, 191, 197  
 orthologic, 136, 189, 205, 213, 219, 221, 223, 224  
 orthomodular, 143, 144, 148, 165, 174  
 orthomodular canonical model, 162  
 orthomodular lattice, 142, 143, 147, 148, 153, 154, 162, 167, 176, 178  
 orthomodular poset, 181, 204, 205  
 orthomodular property, 191  
 orthomodular quantum logic, 136  
 orthomodular realization, 157, 161  
 orthomodularity, 162  
 orthopair semantics, 199  
 orthopairproposition, 197, 198, 200, 201, 203  
 orthopairpropositional conjunction, 197  
 orthopairpropositional disjunction, 197  
 orthoposet, 181  
 Oxford English Dictionary, 5

- P–W**, the logic, 10  
**PA**, Peano arithmetic, 42  
**PA<sup>+</sup>**, positive Peano arithmetic, 42  
 paraconsistent modal operators, 342  
 paraconsistent quantum logic, 189, 219, 220  
 paraconsistent set theory, 363  
 permutation, 7, 86  
     total, 102  
 Perry, J., 114  
 phase-space, 129–131  
 Philo-law, 147  
 physical event, 179  
 physical property, 179  
 physical qualities, 130  
 Piron, C., 135  
 Piron–McLaren’s coordinationization theorem, 170  
 Piron–McLaren’s Theorem, 171  
 point, 102  
 polarities, 59  
 polynomial conditionals, 147, 150  
 positive conditionals, 146  
 positive domain, 197, 203  
 positive laws, 149  
 positive logic, 139, 146, 153  
 positive paradox, 7, 10, 22  
 positive-plus logics, 331  
 possibility operator, 198  
 possible worlds, 138  
 Powers, L., 10  
 Pratt, V., 222  
 Prawitz, D., 66  
 pre-Hilbert space, 164  
 predicate-concept, 173  
 prefixing, 7  
 premisses, 214  
 Priest, G., 349  
 principal formula, 215, 221  
 principal type scheme (PTS), 235  
 principle quasi-ideals, 141  
 principal ideal, 182  
 Prior Analytics, 293  
 probability measure, 135  
 probability, 344  
 projection, 133, 150, 179, 180, 193  
 projective space, 102  
 projection, 183  
 proof, 15  
 proof reductions, 247  
 proper filter, 153  
 proposition, 138, 139, 141, 144, 146, 149, 151, 164, 165, 179, 195  
 propositional quantification, 11  
 pseudo canonical realization, 163  
 pseudocomplemented lattice, 147, 148  
 Pulmannová, S., 184  
 pure state, 130, 132, 133, 145, 146, 150, 190  
 pure states, 129, 131, 146, 203  
  
*Q*-combinators, 237  
*Q*-combinatory logic, 238  
*Q*-definable, 252  
*Q*-logic, 238  
*Q*-terms, 237  
**Q<sup>+</sup>**, the logic, 102  
 Q-translation algorithm, 249  
 QMV-algebra, 185, 187, 188, 209, 210  
 quantification  
     propositional, 11  
 quantifier, 14  
 quantifiers (in paraconsistent logic), 329  
 quantum events, 132  
 quantum logical approach, 135, 171  
 quantum logical implication, 158  
 quantum MV-algebra, 185  
 quantum proposition, 157  
 quantum set theory, 178  
 quantum-logical natural numbers, 177  
 quantum-sets, 178  
 quasets, 178

- quasi-consequence, 152
- quasi-ideal, 145, 147, 165
- quasi-linear, 188
- quasi-linear QMV-algebra, 189, 209
- quasi-model, 152, 154
- quasiset, 178
- Quesada, M., 288
  
- $\mathbf{R}^\sharp$ , the arithmetic, 41–44
- $\mathbf{R}^{\sharp\sharp}$ , the arithmetic, 41, 43, 44
- $\mathbf{R}$ , the logic, 2, 5, 7, 9–14, 17–21, 23, 25–27, 30, 31, 34–39, 41, 42, 49, 54, 55, 57, 58, 60–63, 65–82, 84–89, 92, 93, 95–102, 104, 108, 109, 111–114, 116–118
- $\mathbf{R}$ , the logic
  - undecidability, 102
- $\mathbf{R}^+$ , the logic, 10–12
- $\mathbf{R}^\square$ , the logic, 13
- $\mathbf{R}_{\rightarrow}$ , the logic
  - decidability, 86
- $\mathbf{R}_{\rightarrow}$ , the logic, 7–10, 16–22, 64, 65, 86, 87
- $\mathbf{R}_{\neg}$ , the logic, 45
- $\mathbf{R}_{\rightarrow, \wedge}$ , the logic
  - complexity, 93
- $\mathbf{R}_{\text{fde}}$ , the logic, 27–29, 54, 59
- $\mathbf{R}_{\rightarrow, \wedge, \vee}$ , the logic, 10
- $\mathbf{R}_{\rightarrow, \wedge}$ , the logic, 93, 94
- $\mathbf{R}$ -theory, 36
- $\mathbf{RA}$ , the logic, 116, 117
- Ramsey-test, 290
- Randall, C., 135, 136, 179
- rational acceptability of contradictions, 382
- realizability, 152
- realizable, 154
- realization, 209
- realization for, 205
- realization for a strict-implication language, 151
- reassurance, 349
- reductio, 12
  
- reduction, 231
- reference, 173, 175
- reflection, 215
- regular, 182, 191
- regular involutive bounded poset, 183
- regular paraconsistent quantum logic, 191, 195
- regular symmetric frame, 194
- regularity, 181
- regularity property, 190
- relation
  - binary, 80
  - ternary, 68, 78, 80, 102
- relation algebra, 115
- relational valuations, 308
- relevance, 2, 6
- relevant arithmetic, 41–44
- relevant arrows, 321
- relevant deduction, 15
- relevant implication, 2
- relevant logics, 193, 335
- relevant predication, 118
- representation theorem, 135, 165
- residuation, 12
- resource consciousness, 18
- Restall, G., 100
- restricted-assertion, 8, 9
- rigid, 173
- rigid individual concepts, 175
- $\mathbf{RM}$ , the logic, 7, 13, 18, 78–80, 101
- $\mathbf{RM3}$ , the logic, 95
- $\mathbf{RMO}_{\rightarrow}$ , the logic, 18
- Routley interpretation, 310
- Routley, R., 63, 292
- Routley, V., 292
- $\mathbf{RQ}$ , the logic, 13, 14, 39–42, 70, 82, 83
- rule, 158, 195, 199, 205, 214, 215
  
- $\mathbf{S}_5$ , 196, 202
- $\mathbf{S4}$ , 155
- $\mathbf{S}_{\rightarrow}$ , the logic, 10

- S4**, the logic, 9, 13, 17, 21, 27, 30, 77, 79  
**S5**, the logic, 48, 78, 79  
 Sambin, G., 213, 214  
 Sasaki projection, 167  
 Sasaki-hook, 149  
 satisfication, 173  
 Schütte, K., 46, 47  
 Schrödinger, E., 178  
 Scott, D., 60  
 Scotus, 294  
 second-order (paraconsistent) logic, 338  
 secondary formula, 215  
 self-distribution, 8  
     permuted, 9  
 self-implication, 7  
 semantics, 48  
     algebraic, 49  
     operational, 63  
     Routley–Meyer, 6, 57, 118  
     semi-lattice, 64, 66  
     set-theoretical, 49, 55, 63  
 semantics and set theory, 350  
 semi-transparent effect, 180, 190  
 separable, 131  
 separable Hilbert space, 165  
 separation theorem, 86  
 sequent, 46  
     contraction of, 87  
     right handed, 46  
 sequent calculus, 158, 213, 214  
 sequents, 214  
 set theory in  $LP$ , 358  
 set-theory, 177  
 set-up, 57, 68  
 sharp, 179, 180, 190  
 sharp property, 193, 203  
 Sheffield Shield, The, 114  
 $\sigma$ -complete orthomodular lattice, 135, 170  
 signed formula, 100  
 siilarity logics, 155  
 situation theory, 114  
 Slomson, A. B., 164, 165  
 Socrates, 118  
 Solèr, M. P., 171  
 soundness, 175, 196, 208  
 soundness theorem, 160, 161, 199, 200  
 Stalnaker, , 149  
 Stalnaker, R., 6  
 Stalnaker-function, 149  
 standard orthomodular Kripkean realization, 162  
 standard quantum logic, 136  
 standard realization, 162  
 statistical operator, 180, 181  
 statistical operators, 135  
 Stone theorem, 153  
 strict implication, 150, 151  
 strict-implication operation, 150  
 strong Brouwer–Zadeh logic, 193  
 strong partial quantum logic, 204, 205  
 structural, 214  
 structural rule, 86, 216, 217, 223  
 subspaces, 154, 166  
 substructural logics, 1, 223, 229  
 substructure, 164  
 suffixing, 8  
 Sugihara countermodel, 307  
 Sugihara generalisation, 307  
 Sugihara Matrix, 101  
 summetry, 215  
 superposition principle, 131  
 Suppes, P., 135  
 supremum, 133, 136, 137, 172, 182, 207  
 Sylvan, R., 293  
 symmetric frame, 190, 191  
 symmetry, 221  
 syntactical compactness, 159  
 syntactical compatibility, 159  
 syntactically compatible, 175  
 $T$ , the true constant, 12

- T**, the logic, 11, 19, 77, 78, 92, 108, 109  
*t*, the true constant, 11  
**t**, the logic, 77  
**T–W**, the logic, 10  
**T<sup>+</sup>**, the logic, 11  
**T<sub>→</sub>**, the logic, 9–11  
 Takeuti, G., 177  
 Tamura, S., 213  
 Tennant, N., 32  
 ternary relation, *see* relation, ternary  
 the classical recapture, 347  
 the limitative theorems of meta-mathematics, 373  
 theorems, 37  
 theory of descriptions, 176  
 theories of quasiset, 178  
 theory, 35  
 thinning, 86  
 Thomason, R., 6  
 title, 1  
 told values, 60  
 Toraldo di Francia, G., 178  
 total space, 133  
**TQ**, the logic, 82  
 transfer rule, 219  
 translations, 229  
 true, 177  
 truncated sum, 186  
 truth, 138, 141, 143, 156, 172, 174, 194, 195, 205  
**TW**, the logic, 77, 78, 106–108  
**TW<sup>+</sup>**, the logic, 12, 77  
**TW<sub>→</sub>**, the logic, 10  
**TWQ**, the logic, 82  
 type, 229  
 type assignment, 234, 243  
  
 uncertainty principles, 133  
 undecidability, 102  
 unitary vector, 131, 145  
 universal quantifier, 14  
 unsharp, 190  
 unsharp approach, 180  
  
 unsharp orthoalgebra, 184  
 unsharp partial quantum logic, 204  
 unsharp physical properties, 183  
 unsharp property, 193  
 Urbas, I., 289  
 urelements, 178  
 Urquhart, A., 63, 93, 102  
  
 vacuous quantification, 14  
 validity, 64  
 valuation, 173  
 valuation-function, 137, 172  
 Varadarajan, V. S., 135, 164, 170  
 verifiability, 152, 153  
 verifiable, 154  
 verification, 173  
 visibility, 213, 215, 221  
 von Neumann's collapse of the wave function, 150  
 von Neumann, J., 129–132, 134, 179  
  
**W**, the logic, 10, 99  
 wave functions, 131  
 Way Down Lemma, 35  
 Way Up Lemma, 35  
 weak Brouwer–Zadeh logic, 193  
 weak consequence, 152  
 weak equality, 232  
 weak implication calculus, 7  
 weak import-export, 149  
 weak Lindenbaum theorem, 160, 161, 174  
 weak non monotonic behaviour, 150  
 weak partial quantum logic, 204, 205, 207  
 weakening, 86, 218, 219  
 weaker sets of combinators, 229  
 weakly linear, 188  
 Wittgenstein, L., 287, 292  
 Wolf, R. G., 1  
 world valuation, 173  
  
**X**, the logic, 14, 15, 77

Zermelo–Fraenkel, 178  
zero degree entailment, 309

---

# Handbook of Philosophical Logic

2nd Edition

Volume 7

edited by Dov M. Gabbay and F. Guentner





## CONTENTS

Editorial Preface	vii
<b>Dov M. Gabbay</b>	
Basic Tense Logic	1
<b>John P. Burgess</b>	
Advanced Tense Logic	43
<b>M. Finger, D. Gabbay and M. Reynolds</b>	
Combinations of Tense and Modality	205
<b>Richmond H. Thomason</b>	
Philosophical Perspectives on Quantification in Tense and Modal Logic	235
<b>Nino B. Cocchiarella</b>	
Tense and Time	277
<b>Steven T. Kuhn and Paul Portner</b>	
Index	347



## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London

Logic	IT			
	Natural language processing	Program control specification, verification, concurrency	Artificial intelligence	Logic programming
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs

<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical fragments</b>	Basic back-ground language	Program synthesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely useful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calculus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syntax	Modules. Combining languages	Logics of space and time	Combining features
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game semantics gaining ground		
<b>Object level/metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action-revision models</b>			ditto	



	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model



## BASIC TENSE LOGIC

## 1 WHAT IS TENSE LOGIC?

We approach this question through an example:

- (1) *Smith:* Have you heard? Jones is going to Albania!  
*Smythe:* He won't get in without an extra-special visa.  
 Has he remembered to apply for one?  
*Smith:* Not yet, so far as I know.  
*Smythe:* Then he'll have to do so soon.

In this bit of dialogue the argument, such as it is, turns on issues of temporal order. In English, as in all Indo-European and many other languages, such order is expressed in part through changes in verb-form, or tenses. How should the logician treat such tensed arguments?

A solution that comes naturally to mathematical logicians, and that has been forcefully advocated in [Quine, 1960], is to regiment ordinary tensed language to make it fit the patterns of classical logic. Thus Equation 1 might be reduced to the quasi-English Equation 1 below, and thence to the 'canonical notation' of Equation 3:

- (2) Jones/visits/Albania at some time later than the present.

At any time later than the present, if Jones/visits/Albania then, then at some earlier time Jones/applies/for a visa.

At no time earlier than or equal to the present it is the case that Jones/applies/for a visa.

Therefore, Jones/applies/for a visa at some time later than the present.

- (3)  $\exists t(c < t \wedge P(t))$   
 $\forall t(c < t \wedge P(t) \rightarrow \exists u(u < t \wedge Q(u)))$   
 $\neg \exists t((t < c \vee t = c) \wedge Q(t))$   
 $\therefore \exists t(c < t \wedge Q(t)).$

Regimentation involves introducing quantification over instants  $t, u, \dots$  of time, plus symbols of the present instant  $c$  and the earlier-later relation  $<$ . Above all, it involves treating such a linguistic item as 'Jones is visiting Albania' *not* as a complete sentence expressing a proposition and having a truth-value, to be symbolised by a sentential variable  $p, q, \dots$ , but rather as a predicate expressing a property on instants, to be symbolised by a one-place predicate variable  $P, Q, \dots$ . Regimentation has been called *detensing*

since the verb in, say, ‘Jones/visits/Albania at time  $t$ ’, written here in the grammatical present tense, ought really to be regarded as *tenseless*; for it states not a present fact but a timeless or ‘eternal’ property of the instant  $t$ . Bracketing is one convention for indicating such tenselessness. The knack for regimenting or detensing, for reducing something like Equation 1 to something like Equation 3, is easily acquired. The analysis, however, cannot stop there. For a tensed argument like that above must surely be regarded as an *enthymeme*, having as unstated premises certain assumptions about the structure of Time. Smith and Smythe, for instance, probably take it for granted that of any two distinct instants, one is earlier than the other. And if this assumption is formalised and added as an extra premise, then Equation 3, invalid as it stands, becomes valid.

Of course, it is the job of the cosmologist, not the logician, to judge whether such an assumption is physically or metaphysically correct. What *is* the logician’s job is to formalise such assumptions, correct or not, in logical symbolism. Fortunately, most assumptions people make about the structure of Time go over readily into first- or, at worst, second-order formulas.

### 1.1 Postulates for Earlier-Later

(B0)	Antisymmetry	$\forall x \forall y \neg(x < y \wedge y < x)$
(B1)	Transitivity	$\forall x \forall y \forall z (x < y \wedge y < z \rightarrow x < z)$
(B2)	Comparability	$\forall x \forall y (x < y \vee x = y \vee y < x)$
(B3)	(a) Maximum	$\exists x \forall y (y < x \vee y = x)$
	(b) Minimum	$\exists x \forall y (x < y \vee x = y)$
(B4)	(a) No Maximals	$\forall x \exists y (x < y)$
	(b) No Minimals	$\forall x \exists y (y < x)$
(B5)	Density	$\forall x \forall y (x < y \rightarrow \exists z (x < z \wedge z < y))$
(B6)	(a) Successors	$\forall x \exists y (x < y \wedge \neg \exists z (x < z \wedge z < y))$
	(b) Predecessors	$\forall x \exists y (y < x \wedge \neg \exists z (y < z \wedge z < x))$
(B7)	Completeness	$\forall U ((\exists x U(x) \wedge \exists x \neg U(x) \wedge$ $\forall x \forall y (U(x) \wedge$ $\wedge \neg U(y) \rightarrow x < y)) \rightarrow$ $(\exists x (U(x) \wedge$ $\wedge \forall y (x < y \rightarrow \neg U(y))) \vee$ $\exists x (\neg U(x) \wedge$ $\wedge \forall y (y < x \rightarrow U(y))))$
(B8)	Wellfoundedness	$\forall U (\exists x U(x) \rightarrow \exists x (U(x) \rightarrow$ $\wedge \forall y (y < x \rightarrow \neg U(y)))$
(B9)	(a) Upper Bounds	$\forall x \forall y \exists z (x < z \wedge y < z)$
	(b) Lower Bounds	$\forall x \forall y \exists z (z < x \wedge z < y).$

For more on the development of the logic of time as a branch of applied first- and second-order logic, see [van Benthem, 1978].

The alternative to regimentation is the development of an autonomous *tense logic* (also called *temporal logic* or *chronological logic*), first undertaken in [Prior, 1957] (though several precursors are cited in [Prior, 1967]). Tense logic takes seriously the idea that items like ‘Jones is visiting Albania’ are already complete sentences expressing propositions and having truth-values, and that they should therefore be symbolised by sentential variables  $p, q, \dots$ . Of course, the truth-value of a sentence in the present tense may well differ from that of the corresponding sentence in the past or future tense. Hence, tense logic will need some way of symbolising the relations between sentences that differ only in the tense of the main verb. At its simplest, tense logic adds for this purpose to classical truth-functional sentential logic just two one-place connectives: the future-tense or ‘will’ operator  $F$  and the past-tense or ‘was’ operator  $P$ . Thus, if  $p$  symbolises ‘Jones is visiting Albania’, then  $Fp$  and  $Pp$  respectively symbolise something like ‘Jones is sooner or later going to visit Albania’ and ‘Jones has at least once visited Albania’. In reading tense-logical symbolism aloud,  $F$  and  $P$  may be read respectively as ‘it will be the case that’ and ‘it was the case that’. Then  $\neg F\neg$ , usually abbreviated  $G$ , and  $\neg P\neg$ , usually abbreviated  $H$ , may be read respectively as ‘it is always going to be the case that’ and ‘it has always been the case that’. Actually, for many purposes it is preferable to take  $G$  and  $H$  as primitive, defining  $F$  and  $P$  as  $\neg G\neg$  and  $\neg H\neg$  respectively. Armed with this notation, the tense-logician will reduce Equation 1 above to the stylised Equation 1.1 and then to the tense-logical Equation 5:

(4) Future-tense (Jones visits Albania)

Not future-tense (Jones visits Albania and not past-tense (Jones applies for a visa)).

Not past-tense (Jones applies for a visa) and not Jones applies for a visa.

Therefore, future-tense (Jones applies for a visa)

(5)  $Fp$   
 $\neg F(p \wedge \neg Pq)$   
 $\neg Pq \wedge \neg q$   
 $\therefore Fq.$

Of course, we will want some axioms and rules for the new temporal operators  $F, P, G, H$ . All the axiomatic systems considered in this survey will share the same standard format.

## 1.2 Standard Format

We start from a stock of sentential *variables*  $p_0, p_2, p_2, \dots$ , usually writing  $p$  for  $p_0$  and  $q$  for  $p_1$ . The (well-formed) *formulas* of tense logic are built

up from the variables using negation ( $\neg$ ), and conjunction ( $\wedge$ ), and the strong future ( $G$ ) and strong past ( $H$ ) operators. The *mirror image* of a formula is the result of replacing each occurrence of  $G$  by  $H$  and vice versa. Disjunction ( $\vee$ ), material conditional ( $\rightarrow$ ), material biconditional ( $\leftrightarrow$ ), constant true ( $\top$ ), constant false ( $\perp$ ), weak future ( $F$ ), and weak past ( $P$ ) can be introduced as abbreviations.

As *axioms* we take all substitution instances of truth-functional tautologies. In addition, each particular system will take as axioms all substitution instances of some finite list of extra axioms, called the *characteristic* axioms of the system. As *rules* of inference we take Modus Ponens (MP) plus the specifically tense-logical:

Temporal Generalisation(TG): From  $\alpha$  to infer  $G\alpha$  and  $H\alpha$

The *theses* of a system are the formulas obtainable from its axioms by these rules. A formula is *consistent* if its negation is not a thesis; a set of formulas is *consistent* if the conjunction of any finite subset is. These notions are, of course, relative to a given system.

The systems considered in this survey will have characteristic axioms drawn from the following list:

### 1.3 Postulates for a Past-Present-Future

- |      |  |     |  |
|------|--|-----|--|
| (A0) | (a) $G(p \rightarrow q) \rightarrow (Gp \rightarrow Gq)$                             | (b) | $H(p \rightarrow q) \rightarrow (Hp \rightarrow Hq)$ |
|      | (c) $p \rightarrow Gpp$  | (d) | $p \rightarrow HFp$                                  |
| (A1) | (a) $Gp \rightarrow GGp$   | (b) | $Hp \rightarrow HHp$                                 |
| (A2) | (a) $Pp \wedge Fq \rightarrow F(p \wedge Fq) \vee F(p \wedge q) \vee F(Fp \wedge q)$ |     |  |
|      | (b) $Pp \wedge Pq \rightarrow P(p \wedge Pq) \vee P(p \wedge q) \vee P(Pp \wedge q)$ |     |  |
| (A3) | (a) $G\perp \vee FG\perp$  | (b) | $H\perp \vee PH\perp$                                |
| (A4) | (a) $Gp \rightarrow Fp$  | (b) | $Hp \rightarrow Pp$                                  |
| (A5) | (a) $Fp \rightarrow FFp$   | (b) | $Pp \rightarrow PPp$                                 |
| (A6) | (a) $p \wedge Hp \rightarrow FHp$  | (b) | $p \wedge Gp \rightarrow PGp$                        |
| (A7) | (a) $Fp \wedge FG\neg p \rightarrow F(HFp \wedge G\neg p)$                           |     |  |
|      | (b) $Pp \wedge PH\neg p \rightarrow P(GPp \wedge H\neg p)$                           |     |  |
| (A8) | (a) $H(Hp \rightarrow p) \rightarrow Hp$   |     |  |
| (A9) | (a) $FGp \rightarrow GFp$  | (b) | $PHp \rightarrow HPp$                                |

A few definitions are needed before we can state precisely the basic problem of tense logic, that of finding characteristic axioms that ‘correspond’ to various assumptions about Time.

### 1.4 Formal Semantics

A *frame* is a nonempty set  $C$  equipped with a binary relation  $R$ . A *valuation* in a frame  $(X, R)$  is a function  $V$  assigning each variable  $p_i$  a subset of  $X$ . Intuitively,  $X$  can be thought of as representing the set of instants of time,  $R$

the earlier-later relation,  $V$  the function telling us *when* each  $p_i$  is the case. We extend  $V$  to a function defined on *all* formulas, by abuse of notation still called  $V$ , inductively as follows:

$$\begin{aligned} V(\neg\alpha) &= X - V(\alpha) \\ V(\alpha \wedge \beta) &= V(\alpha) \cap V(\beta) \\ V(G\alpha) &= \{x \in X : \forall y \in X(xRy \rightarrow y \in V(\alpha))\} \\ V(H\alpha) &= \{x \in X : \forall y \in X(yRx \rightarrow y \in V(\alpha))\}. \end{aligned}$$

(Some writers prefer a different notion. Thus, what we have expressed as  $x \in V(\alpha)$  may appear as  $\|\alpha\|_x^V = \text{TRUE}$  or as  $(X, R, V) \models \alpha[x]$ .) A formula  $\alpha$  is *valid* in a frame  $(X, R)$  if  $V(\alpha) = X$  for every valuation  $V$  in  $(X, R)$ , and is *satisfiable* in  $(X, R)$  if  $V(\alpha) \neq \emptyset$  for some valuation  $V$  in  $(X, R)$ , or equivalently if  $\neg\alpha$  is not valid in  $(X, R)$ . Further,  $\alpha$  is *valid* over a class  $\mathcal{K}$  of frames if it is valid in every  $(X, R) \in \mathcal{K}$ , and is *satisfiable* over  $\mathcal{K}$  if it is satisfiable in some  $(X, R) \in \mathcal{K}$ , or equivalently if  $\neg\alpha$  is not valid over  $\mathcal{K}$ . A system  $\mathbf{L}$  in standard format is *sound* for  $\mathcal{K}$  if every thesis of  $\mathbf{L}$  is valid over  $\mathcal{K}$ , and a sound system  $\mathbf{L}$  is *complete* for  $\mathcal{K}$  if conversely every formula valid over  $\mathcal{K}$  is a thesis of  $\mathbf{L}$ , or equivalently, if every formula consistent with  $\mathbf{L}$  is satisfiable over  $\mathcal{K}$ . Any set (let us say, finite)  $\Phi$  of first- or second-order axioms about the earlier-later relation  $<$  determines a class  $\mathcal{K}(\Phi)$  of frames, the class of its *models*. The basic correspondence problem of tense logic is, given  $\Phi$  to find characteristic axioms for a system  $\mathbf{L}$  that will be sound and complete for  $\mathcal{K}(\Phi)$ . The next two sections of this survey will be devoted to representing the solution to this problem for many important  $\Phi$ .

### 1.5 Motivation

But first it may be well to ask, why bother? Several classes of motives for developing an autonomous tense logic may be cited:

(a) *Philosophical* motives were behind much of the pioneering work of A. N. Prior, to whom the following point seemed most important: whereas our ordinary language is tensed, the language of physics is mathematical and so untensed. Thus, there arise opportunities for confusions between different ‘terms of ideas’. Now working in tense logic, what we learn is precisely how to avoid confusing the tensed and the tenseless, and how to clarify their relations (e.g. we learn that essentially the same thought can be formulated tenselessly as, ‘Of any two distinct instants, one /is/ earlier and the other /is/ later’, and tensedly as, ‘Whatever is going to have been the case either already has been or now is or is sometime going to be the case’). Thus, the study of tense logic can have at least a ‘therapeutic’ value. Later writers have stressed other philosophical applications, and some of these are treated elsewhere in this *Handbook*.

(b) *Exegetical* applications again interested Prior (see his [Prior, 1967, Chapter 7]). Much was written about the logic of time (especially about future contingents) by such ancient writers as Aristotle and Diodoros Kronos (whose works are unfortunately lost) and by such mediaeval ones as William of Ockham or Peter Auriolo. It is tempting to try to bring to bear insights from modern logic to the interpretation of their thought. But to pepper the text of an Aristotle or an Ockham with such regimenters' phrases as 'at time  $t$ ' is an almost certain guarantee of misunderstanding. For these earlier writers thought of such an item as 'Socrates is running' as being already complete as it stands, *not* as requiring supplementation before it could express a proposition or have a truth-value. Their standpoint, in other words, was like that of modern tense logic, whose notions and notations are likely to be of most use in interpreting their work, if any modern developments are.

(c) *Linguistic* motivations are behind much recent work in tense logic. A certain amount of controversy surrounds the application of tense logic to natural language. See, e.g. van Benthem [1978; 1981] for a critic's views. To avoid pointless disputes it should be emphasised from the beginning that tense logic does not attempt the faithful replication of every feature of the deep semantic structure (and still less of the surface syntax) of English or any other language; rather, it provides an idealised model giving the sympathetic linguist food for thought. an example: in tense logic,  $P$  and  $F$  can be iterated indefinitely to form, e.g.  $PPPFp$  or  $FPFPp$ . In English, there are four types of verbal modifications indicating temporal reference, each applicable at most *once* to the main verb of a sentence: Progressive (be + ing), Perfect (have + en), Past (+ ed), and Modal auxiliaries (including will, would). Tense logic, by allowing unlimited iteration of its operators, departs from English, to be sure. But by doing so, it enables us to raise the question of whether the multiple compounds formable by such iteration are really all distinct in meaning; and a theorem of tense logic (see Section 3.5 below) tells us that on reasonable assumptions they are not, e.g.  $PPPFp$  and  $FPFPp$  both collapse to  $PFp$  (which is equivalent to  $PPp$ ). and this may suggest *why* English does not *need* to allow unlimited iteration of its temporal verb modifications.

(d) *Computer Science*: Both tense logic itself and, even more so, the closely related so-called *dynamic logic* have recently been the objects of much investigation by theorists interested in program verification. temporal operators have been used to express such properties of programs as termination, correctness, safety, deadlock freedom, clean behaviour, data integrity, accessibility, responsiveness, and fair scheduling. These studies are mainly concerned only with *future* temporal operators, and so fall technically within the province of *modal* logic. See Harel *et al.*'s chapter on dynamic logic in Volume 4 of this *Handbook*, Pratt [1980] among other items in our bibliog-



raphy.

(e) *Mathematics*: Some taste of the purely mathematical interest of tense logic will, it is hoped, be apparent from the survey to follow. Moreover, tense logic is not an isolated subject within logic, but rather has important links with modal logic, intuitionistic logic, and (monadic) second-order logic.

Thus, the motives for investigating tense logic are many and varied.

## 2 FIRST STEPS IN TENSE LOGIC

Let  $\mathbf{L}_0$  be the system in standard format with characteristic axioms (A0a, b, c, d). Let  $\mathcal{K}_0$  be the class of *all* frames. We will show that  $\mathbf{L}_0$  is (sound and) complete for  $\mathcal{L}_0$ , and thus deserves the title of *minimal* tense logic. The method of proof will be applied to other systems in the next section. Throughout this section, thesishood and consistency are understood relative to  $\mathbf{L}_0$ , validity and satisfiability relative to  $\mathcal{K}_0$ .

**THEOREM 1** (Soundness Theorem).  $\mathbf{L}_0$  is sound for  $\mathcal{K}_0$ .

**Proof.** We must show that any thesis (of  $\mathbf{L}_0$ ) is valid (over  $\mathcal{K}_0$ ). for this it suffices to show that each axiom is valid, and that each rule preserves validity. the verification that tautologies are valid, and that substitution and MP preserves validity is a bit tedious, but entirely routine.

To check that (A0a) is valid, we must show that for all relevant  $X, R, V$  and  $x$ , if  $x \in V(G(p \rightarrow q))$  and  $x \in V(Gp)$ , then  $x \in V(Gq)$ . Well, the hypotheses here mean, first that whenever  $xRy$  and  $y \in V(p)$ , then  $y \in V(q)$ ; and second that whenever  $xRy$ , then  $y \in V(p)$ . The desired conclusion is that whenever  $xRy$ , then  $y \in V(q)$ ; which follows immediately. Intuitively, (A0a) says that if  $q$  is going to be the case whenever  $p$  is, and  $p$  is always going to be the case, then  $q$  is always going to be the case. The treatment of (A0b) is similar.

To check that (A0c) is valid, we must show that for all relevant  $X, R, V$ , and  $x$ , if  $x \in V(p)$ , then  $x \in V(GPp)$ . Well, the desired conclusion here is that for every  $y$  with  $xRy$  there is a  $z$  with  $zRy$  and  $z \in V(p)$ . It suffices to take  $z = x$ . Intuitively, (A0c) says that whatever is now the case is always going to have been the case. The treatment of (A0d) is similar.

To check that TG preserves validity, we must show that if for all relevant  $X, R, V$ , and  $x$  we have  $x \in V(\alpha)$ , then for all relevant  $X, R, V$ , and  $x$  we have  $x \in V(H\alpha)$  and  $x \in V(G\alpha)$ , in other words, that whenever  $yRx$  we have  $y \in V(\alpha)$  and whenever  $xRy$  we have  $y \in V(\alpha)$ . But this is immediate. Intuitively, TG says that if something is now the case *for logical reasons alone*, then for logical reasons alone it always has been and is always going to be the case: logical truth is eternal. ■

In future, verifications of soundness will be left as exercises for the reader. Our proof of the completeness of  $\mathbf{L}_0$  for  $\mathcal{K}_0$  will use the method of maximal

consistent sets, first developed for first-order logic by L. Henkin, systematically applied to tense logic by E. J. Lemmon and D. Scott (in notes eventually published as [Lemmon and Scott, 1977]), and refined [Gabbay, 1975].

The completeness of  $\mathbf{L}_0$  for  $\mathcal{K}_0$  is due to Lemmon. We need a number of preliminaries.

**THEOREM 2** (Derived rules). *The following rules of inference preserve thesishood:*

1. from  $\alpha_1, \alpha_2, \dots, \alpha_n$  to infer any truth-functional consequence  $\beta$
2. from  $\alpha \rightarrow \beta$  to infer  $G\alpha \rightarrow G\beta$  and  $H\alpha \rightarrow H\beta$
3. from  $\alpha \leftrightarrow \beta$  and  $\theta(\alpha/p)$  to infer  $\theta(\beta/p)$
4. from  $\alpha$  to infer its mirror image.

**Proof.**

1. To say that  $\beta$  is a truth-functional consequence of  $\alpha_1, \alpha_2, \dots, \alpha_n$  is to say that  $(\alpha_1 \wedge \alpha_2 \wedge \dots \wedge \alpha_n \rightarrow \beta)$  or equivalently  $\alpha_1 \rightarrow (\alpha_2 \rightarrow (\dots (\alpha_n \rightarrow \beta) \dots))$  is an instance of a tautology, and hence is an axiom. We then apply MP.
2. From  $\alpha \rightarrow \beta$  we first obtain  $G(\alpha \rightarrow \beta)$  by TG, and then  $G\alpha \rightarrow G\beta$  by A0a and MP. Similarly for  $H$ .
3. Here  $(\alpha/p)$  denotes substitution of  $\alpha$  for the variable  $p$ . It suffices to prove that if  $\alpha \rightarrow \beta$  and  $\beta \rightarrow \alpha$  are theses, then so are  $\theta(\alpha/p) \rightarrow \theta(\beta/p)$  and  $\theta(\beta/p) \rightarrow \theta(\alpha/p)$ . This is proved by induction on the complexity of  $\theta$ , using part (2) for the cases  $\theta = G\chi$  and  $\theta = H\chi$ . In particular, part (3) allows us to insert and remove double negations freely. We write  $\alpha \approx \beta$  to indicate that  $\alpha \leftrightarrow \beta$  is a thesis.
4. This follows from the fact that the tense-logical axioms of  $\mathbf{L}_0$  come in mirror-image pairs, (A0a, b) and (A0c, d). Unlike parts (1)–(3), part (4) will *not* necessarily hold for every extension of  $\mathbf{L}_0$ . ■

**THEOREM 3** (Theses). *Items (a)–(h) below are theses of  $\mathbf{L}_0$ .*

**Proof.** We present a deduction, labelling some of the lines as theses for future reference:

- |          |  |                          |
|----------|--|--------------------------|
| (1)      | $G(p \rightarrow q) \rightarrow G(\neg q \rightarrow \neg p)$            | from a tautology by 1.2b |
| (2)      | $G(\neg q \rightarrow \neg p) \rightarrow (G\neg q \rightarrow G\neg p)$ | (A0a)                    |
| (a) (3)  | $G(p \rightarrow q) \rightarrow (Fp \rightarrow Fq)$                     | from 1,2 by 1.2a         |
| (4)      | $Gp \rightarrow G(q \rightarrow p \wedge q)$                             | from a tautology by 1.2b |
| (5)      | $G(q \rightarrow p \wedge q) \rightarrow (Fq \rightarrow F(p \wedge q))$ | 3                        |
| (b) (6)  | $Gp \wedge Fq \rightarrow F(p \wedge q)$                                 | from 4, 5 by 1.2a        |
| (7)      | $p \rightarrow GPP$  | (A0c)                    |
| (8)      | $GPP \wedge Fq \rightarrow F(Pp \wedge q)$                               | 6                        |
| (c) (9)  | $p \wedge Fq \rightarrow F(Pp \wedge q)$                                 | from 7, 8 by 1.2a        |
| (10)     | $G(p \wedge q) \rightarrow Gp$   |                          |
|          | $G(p \wedge q) \rightarrow Gq$   | from tautologies by 1.2b |
| (11)     | $G(q \rightarrow p \wedge q) \rightarrow (Gq \rightarrow G(p \wedge q))$ | (A0a)                    |
| (d) (12) | $Gp \wedge Gq \leftrightarrow G(p \wedge q)$                             | 12                       |
| (14)     | $G\neg p \wedge G\neg q \rightarrow G\neg(p \vee q)$                     | from 13 by 1.3c          |
| (e) (15) | $Fp \vee Fq \leftrightarrow F(p \vee q)$                                 | from 14 by 1.2a          |
| (16)     | $Gp \rightarrow G(p \vee q)$   |                          |
|          | $Gq \rightarrow G(p \vee q)$   | from tautologies by 1.2b |
| (f) (17) | $Gp \vee Gq \rightarrow G(p \vee q)$                                     | from 16 by 1.2a          |
| (18)     | $G\neg q \vee G\neg q \rightarrow G(\neg p \vee \neg q)$                 | 17                       |
| (19)     | $G\neg p \vee G\neg q \rightarrow G\neg(p \wedge q)$                     | from 18 by 1.2c          |
| (g) (20) | $F(p \wedge q) \rightarrow Fp \wedge Fq$                                 | from 19 by 1.2a          |
| (21)     | $\neg p \rightarrow HF\neg p$  | (A0d)                    |
| (22)     | $\neg p \rightarrow H\neg Gp$  | from 21 by 1.2c          |
| (h) (23) | $PGp \rightarrow p$  | from 22 by 1.2a          |

Also the mirror images of 1.3a–h are theses by 1.2d. ■

We assume familiarity with the following:

LEMMA 4 (Lindenbaum’s Lemma). *Any consistent set of formulas can be extended to a maximal consistent set.*

LEMMA 5. *Let  $Q$  be a maximal consistent set of formulas. For all formulas we have:*

1. *If  $\alpha_1, \dots, \alpha_n \in A$  and  $\alpha_1 \wedge \dots \wedge \alpha_n \rightarrow \beta$  is a thesis, then  $\beta \in A$ .*
2.  *$\neg\alpha \in A$  iff  $\alpha \notin A$*
3.  *$(\alpha \wedge \beta) \in A$  iff  $\alpha \in A$  and  $\beta \in A$*
4.  *$(\alpha \vee \beta) \in A$  iff  $\alpha \in A$  or  $\beta \in A$ .*

They will be used tacitly below.

Intuitively, a maximal consistent set—henceforth abbreviated MCS—represents a full description of a possible state of affairs. For MCSs  $A, B$  we say that  $A$  is *potentially followed* by  $B$ , and write  $A \dashv\vdash B$ , if the conditions

of Lemma 6 below are met. Intuitively, this means that a situation of the sort described by  $A$  could be followed by one of the sort described by  $B$ .

LEMMA 6. *For any MCSs  $A, B$ , the following are equivalent:*

1. *whenever  $\alpha \in A$ , we have  $P\alpha \in B$ ,*
2. *whenever  $\beta \in B$ , we have  $F\beta \in A$ ,*
3. *whenever  $G\gamma \in A$ , we have  $\gamma \in B$ ,*
4. *whenever  $H\delta \in B$ , we have  $\delta \in A$ .*

**Proof.** To show (1) implies (3): assume (1) and let  $G\gamma \in A$ . Then  $PG\gamma \in B$ , so by Thesis 3(h) we have  $\gamma \in B$  as required by (3).

To show (3) implies (2): assume (3) and let  $\beta \in B$ . then  $\neg\beta \notin B$ , so  $G\neg\beta \notin A$ , and  $F\beta = \neg G\neg\beta \in A$  as required by (2).

Similarly (2) implies (4) and (4) implies (1). ■

LEMMA 7. *Let  $C$  be an MCS,  $\gamma$  any formula:*

1. *if  $F\gamma \in C$ , then there exists an MCS  $B$  with  $C \rightarrow B$  and  $\gamma \in B$ ,*
2. *if  $P\gamma \in C$ , then there exists an MCS  $A$  with  $A \rightarrow C$  and  $\gamma \in A$ .*

**Proof.** We treat (1): it suffices (by the criterion of Lemma 6(a)) to obtain an MCS  $B$  containing  $B_0 = \{P\alpha : \alpha \in C\} \cup \{\gamma\}$ . For this it suffices (by Lindenbaum's Lemma) to show that  $B_0$  is consistent. For this it suffices (by the closure of  $C$  under conjunction plus the mirror image of Theorem 3(g)) to show that for any  $\alpha \in C$ ,  $P\alpha \wedge \gamma$  is consistent. For this it suffices (since TG guarantees that  $\neg F\delta$  is a thesis whenever  $\neg\delta$  is) to show that  $F(P\alpha \wedge \gamma)$  is consistent. And for this it suffices to show that  $F(P\alpha \wedge \gamma)$  belongs to  $C$ —as it must by 3(c). ■

DEFINITION 8. A *chronicle* on a frame  $(X, R)$  is a function  $T$  assigning each  $x \in X$  an MCS  $T(x)$ . Intuitively, if  $X$  is thought of as representing the set of instants, and  $R$  the earlier-later relation,  $T$  should be thought of as providing a complete description of what goes on at each instant.  $T$  is *coherent* if we have  $T(x) \rightarrow T(y)$  whenever  $xRy$ .  $T$  is *prophetic* (resp. *historic*) if it is coherent and satisfies the first (resp. second) condition below:

1. *whenever  $F\gamma \in T(x)$  there is a  $y$  with  $xRy$  and  $\gamma \in T(y)$ ,*
2. *whenever  $P\gamma \in T(x)$  there is a  $y$  with  $yRx$  and  $\gamma \in T(y)$ .*

$T$  is *perfect* if it is both prophetic and historic. Note that  $T$  is coherent iff it satisfies the two following conditions:

3. whenever  $\gamma \in T(x)$  and  $xRy$ , then  $\gamma \in T(y)$ ,
4. whenever  $H\gamma \in T(x)$ , and  $yRx$ , then  $\gamma \in T(y)$ .

If  $V$  is a valuation in  $(X, R)$ , the *induced* chronicle  $T_V$  is defined by  $T_V(x) = \{\gamma : x \in V(\gamma)\}$ ;  $T_V$  is always perfect. If  $T$  is a perfect chronicle on  $(X, R)$ , the *induced* valuation  $V_T$  is defined by  $V_T(p_i) = \{x : p_i \in T(x)\}$ . We have:

LEMMA 9 (Chronicle Lemma). *Let  $T$  be a perfect chronicle on a frame  $(X, R)$ . If  $V = V_T$  is the valuation induced by  $T$ , then  $T = T_V$  the chronicle induced by  $V$ . In other words, for all formulas  $\gamma$  we have:*

$$(+) \quad V(\gamma) = \{x : \gamma \in T(x)\}$$

*In particular, any member of any  $T(x)$  is satisfiable in  $(X, R)$ .*

**Proof.** (+) is proved by induction on the complexity of  $\gamma$ . As a sample, we treat the induction step for  $G$ : assume (+) for  $\gamma$ , to prove it for  $G\gamma$ :

On the one hand, if  $G\gamma \in T(x)$ , then by Definition 8(3), whenever  $xRy$  we have  $\gamma \in T(y)$  and by induction hypothesis  $y \in V(\gamma)$ . This shows  $x \in V(G\gamma)$ .

On the other hand, if  $G\gamma \notin T(x)$ , then  $F\neg\gamma \approx \neg G\gamma \in T(x)$ , so by Definition 8(1) for some  $y$  with  $xRy$  we have  $\neg\gamma \in T(y)$  and  $\gamma \notin T(y)$ , whence by induction hypothesis,  $y \notin V(\gamma)$ . This shows  $x \notin V(G\gamma)$ . ■

To prove the completeness of  $\mathbf{L}_0$  for  $\mathcal{K}_0$  we must show that every consistent formula  $\gamma_0$  is satisfiable. Now Lemma 9 suggests an obvious strategy for proving  $\gamma_0$  satisfiable, namely to construct a perfect chronicle  $T$  on some frame  $(X, R)$  containing an  $x_0$  with  $\gamma_0 \in T(x_0)$ . We will construct  $X, R$ , and  $T$  piecemeal.

DEFINITION 10. Fix a denumerably infinite set  $W$ . Let  $M$  be the set of all triples  $(X, R, T)$  such that :

1.  $X$  is a nonempty finite subset of  $W$ ,
2.  $R$  is an antisymmetric binary relation on  $X$ ,
3.  $T$  is a coherent chronicle on  $(X, R)$ .

For  $\mu = (X, R, T)$  and  $\mu' = (X', R', T')$  in  $M$  we say  $\mu'$  *extends*  $\mu$  if (when relations and functions are identified with sets of ordered pairs) we have:

- 1'.  $X \subseteq X'$
- 2'.  $R = R' \cap (X \times X)$
- 3'.  $T \subseteq T'$ .

A conditional requirement of form 8(1) or (2) will be called *unborn* for  $\mu = (X, R, T) \in M$  if its antecedent is not fulfilled; that is, if  $x \notin X$  or if  $x \in X$  but  $F\gamma$  or  $P\gamma$  a the case may be does not belong to  $T(x)$ . It will be called *alive* for  $\mu$  if its antecedent is fulfilled but its consequent is not; in other words, there is no  $y \in X$  with  $xRy$  or  $yRx$  as the case may be and  $\gamma \in T(y)$ . It will be called *dead* for  $\mu$  if its consequent is fulfilled.

Perhaps no member of  $M$  is perfect; but any imperfect member of  $M$  can be improved:

LEMMA 11 (Killing Lemma). *Let  $\mu = (X, R, T) \in M$ . For any requirement of form 8(1) or (2) which is alive for  $\mu$ , there exists an extension  $\mu' = (X', R', T') \in M$  of  $\mu$  for which that requirement is dead.*

**Proof.** We treat a requirement of form 8(1). If  $x \in X$  and  $F\gamma \in T(x)$ , by 7(1) there is an MCS  $B$  with  $T(x) \rightarrow B$  and  $\gamma \in B$ . It therefore suffices to fix  $y \in W - X$  and set

1.  $X' = X \cup \{y\}$
2.  $R' = R \cup \{(x, y)\}$
3.  $T' = T \cup \{(y, B)\}$ . ■

THEOREM 12 (Completeness Theorem).  $\mathbf{L}_0$  is complete for  $\mathcal{K}_0$ .

**Proof.** Given a consistent formula  $\gamma_0$ , we wish to construct a frame  $(X, R)$  and a perfect chronicle  $T$  on it, with  $\gamma_0 \in t(x_0)$  for some  $x_0$ . To this end we fix an enumeration  $x_0, x_1, x_2, \dots$  of  $W$ , and an enumeration  $\gamma_0, \gamma_1, \gamma_2, \dots$  of all formulas. To the requirement of form 8(1) (resp. 8(2)) for  $x = x_i$  and  $\gamma = \gamma_j$  we assign the *code number*  $2 \cdot 5^i \cdot 7^j$  (resp.  $3 \cdot 5^i \cdot 7^j$ ). Fix an MCS  $C_0$  with  $\gamma_0 \in C_0$ , and let  $\mu_0 = (X_0, R_0, T_0)$  where  $X_0 = \{x_0\}$ ,  $R_0 = \emptyset$ , and  $T_0 = \{(x_0, C_0)\}$ . If  $\mu_n$  is defined, consider the requirement, which among all those which are alive for  $\mu_n$ , has the least code number. Let  $\mu_{n+1}$  be an extension of  $\mu_n$  for which that requirement is dead, as provided by the Killing Lemma. Let  $(X, R, T)$  be the union of the  $\mu_n = (X_n, R_n, T_n)$ ; more precisely, let  $X$  be the union of the  $X_n$ ,  $R$  of the  $R_n$ , and  $T$  of the  $T_n$ . It is readily verified that  $T$  is a perfect chronicle on  $(X, R)$ , as required. ■

The observant reader may be wondering why in Definition 10(2) the relation  $R$  was required to be antisymmetric. the reason was to enable us to make the following remark: our proof actually shows that every thesis of  $\mathbf{L}_0$  is valid over the class  $\mathcal{K}_0$  of all frames, and that every formula consistent with  $\mathbf{L}_0$  is satisfiable over the class  $\mathcal{K}_{\text{anti}}$  of antisymmetric frames. Thus,  $\mathcal{K}_0$  and  $\mathcal{K}_{\text{anti}}$  give rise to the same tense logic; or to put the matter differently, there is no characteristic axiom for tense logic which ‘corresponds’ to the assumption that the earlier-later relation on instants of time is antisymmetric.

In this connection a remark is in order: suppose we let  $X$  be the set of all MCSs,  $R$  the relation  $\rightarrow$ ,  $V$  the valuation  $V(p_i) = \{x : p_i \in x\}$ . Then using Lemmas 6 and 7 it can be checked that  $V(\gamma) = \{x : \gamma \in x\}$  for all  $\gamma$ . In this way we get a quick proof of the completeness of  $\mathbf{L}_0$  for  $\mathcal{K}_0$ . However, this  $(X, R)$  is not antisymmetric. Two MCSs  $A$  and  $B$  may be *clustered* in the sense that  $A \rightarrow B$  and  $B \rightarrow A$ . There is a trick, known as ‘bulldozing’, though, for converting nonantisymmetric frames to antisymmetric ones, which can be used here to give an alternative proof of the completeness of  $\mathbf{L}_0$  for  $\mathcal{K}_{\text{anti}}$ . See Bull and Segerberg’s chapter in Volume 3 of this *Handbook* and [Segerberg, 1970].

### 3 A QUICK TRIP THROUGH TENSE LOGIC

The material to be presented in this section was developed piecemeal in the late 1960s. In addition to persons already mentioned, R. Bull, N. Cocchiarella and S. Kripke should be cited as important contributors to this development. Since little was published at the time, it is now hard to assign credits.

#### 3.1 Partial Orders

Let  $\mathbf{L}_1$  be the extension for  $\mathbf{L}_0$  obtained by adding (A1a) as an extra axiom. Let  $\mathcal{K}_1$  be the class of *partial orders*, that is, of antisymmetric, transitive frames. We claim  $\mathbf{L}_1$  is (sound and) complete for  $\mathcal{K}_1$ . Leaving the verification of soundness as an exercise for the reader, we sketch the modifications in the work of the preceding section needed to establish completeness.

First of all, we must now understand the notions of thesishood and consistency and, hence, of MCS and chronicle, as relative to  $\mathbf{L}_0$ . Next, we must revise clause 10(2) in the definition of M to read:

2<sub>1</sub>.  $R$  is a partial order on  $X$ .

This necessitates a revision in clause 11(2) in the proof of the Killing Lemma. Namely, in order to guarantee that  $R'$  will be a partial order on  $X'$ , that clause must now read:

2<sub>1</sub>.  $R' = R \cup \{(x, y)\} \cup \{(v, y) : vRx\}$ .

But now it must be checked that  $T'$ , as defined by clause 11(3), remains a coherent chronicle under the revised definition of  $R'$ . Namely, it must be checked that if  $vRx$ , then  $T(v) \rightarrow B$ . To show this (and so complete the proof) the following suffices:

LEMMA *Let  $A, C, B$  be MCSs. If  $A \rightarrow C$  and  $C \rightarrow B$ , then  $A \rightarrow B$ .*

**Proof.** We use criterion 6(3) for  $\neg\exists$ : assume  $G\gamma \in A$ , to prove  $\gamma \in B$ . Well, by the new axiom (A1a) we have  $GG\gamma \in A$ . Then since  $A \neg\exists C$ , we have  $G\gamma \in C$ , and since  $C \neg\exists B$ , we have  $\gamma \in B$ . ■

It is worth remarking that the mirror image (A1b) of (A1a) is equally valid over partial orders, and must thus by the completeness theorem be a thesis of  $\mathbf{L}_0$ . To find a deduction of it is a nontrivial exercise.

### 3.2 Total Orders

Let  $\mathbf{L}_2$  be the extension of  $\mathbf{L}_1$  obtained by adding (A2a, b) as extra axioms. Let  $\mathcal{K}_1$  be the class of *total orders*, or frames satisfying antisymmetry, transitivity, and comparability. Leaving the verification of soundness to the reader, we sketch the modifications in the work of Section 3.1 above, beyond simply understanding thesishood and related notions as relative to  $\mathbf{L}_2$ , needed to show  $\mathbf{L}_2$  complete for  $\mathcal{K}_2$ .

To begin with, we must revise clause 10(2) in the definition of  $M$  to read:

2<sub>2</sub>.  $R$  is a partial order on  $X$ .

This necessitates revisions in the proof of the Killing Lemma, for which the following will be useful:

**LEMMA** *Let  $A, B, C$  be MCSs. If  $A \neg\exists B$  and  $A \neg\exists C$ , then either  $B = C$  or  $B \neg\exists C$  or  $C \neg\exists B$ .*

**Proof.** Suppose for contradiction that the two hypotheses hold but none of the three alternatives in the conclusion holds. Using criterion 6(2) for  $\neg\exists$ , we see that there must exist a  $\gamma_0 \in C$  with  $F\gamma_0 \notin b$  (else  $B \neg\exists C$ ) and a  $\beta_0 \in B$  with  $F\beta_0 \notin C$  (else  $C \neg\exists B$ ). Also there must exist a  $\delta$  with  $\delta \in B, \delta \notin C$  (else  $B = C$ ). Let  $\beta = \beta_0 \wedge \neg F\gamma_0 \wedge \delta \in B, \gamma = \gamma_0 \wedge \neg F\beta_0 \wedge \neg\delta \in C$ . We have  $F\beta \in A$  (since  $A \neg\exists B$ ) and  $F\gamma \in A$  (since  $A \neg\exists C$ ). hence, by A2a, one of  $F(\beta \wedge F\gamma), F(F\beta \wedge \gamma), F(\beta \wedge \gamma)$  must belong to  $A$ . But this is impossible since all three are easily seen (using 3(7)) to be inconsistent. ■

Turning now to the Killing Lemma, consider a requirement of form 8(1) which is alive for a certain  $\mu = (X, R, T) \in M$ . We claim there is an extension  $\mu' = (X', R', T')$  for which it is dead. This is proved by induction on the number  $n$  of successors which  $x$  has in  $(X, R)$ . We fix an MCS  $B$  with  $T(x) \neg\exists B$  and  $\gamma \in B$ . If  $n = 0$ , it suffices to define  $\mu'$  as was done in Section 3.1 above.

If  $n > 0$ , let  $x'$  be the immediate successor of  $x$  in  $(X, R)$ . We cannot have  $\gamma \in T(x')$  or else our requirement would already be dead for  $\mu$ . If  $F\gamma \in T(x')$ , we can reduce to the case  $n - 1$  by replacing  $x$  by  $x'$ . So suppose  $F\gamma \notin T(x')$ . Then we have neither  $B = T(x')$  nor  $T(x') \neg\exists B$ .



Hence, by the Lemma, we must have  $B \rightarrow \neg T(x')$ . Therefore it suffices to fix  $y \in W - X$  and set:

$$\begin{aligned} X' &= X \cup \{y\} \\ R' &= R \upharpoonright \text{cup}\{(x, y), (y, x')\} \cup \{(v, y) : vRx\} \cup \{(y, v) : (x'Rv)\} \\ I' &= T \cup \{(y, B)\}. \end{aligned}$$

In other words, we insert a point between  $x$  and  $x'$ , assigning it the set  $B$ . Requirements of form 8(2) are handled similarly, using a mirror image of the Lemma, proved using (A2b). No further modifications in the work of Section 3.1 above are called for.

The foregoing argument also establishes the following: let  $\mathbf{L}_{\text{tree}}$  be the extension of  $\mathbf{L}_1$  obtained by adding (A2b) as an extra axiom. Let  $\mathcal{K}_{\text{tree}}$  be the class of *trees*, defined for present purposes as those partial orders in which the predecessors of any element are totally ordered. Then  $\mathbf{L}_{\text{tree}}$  is complete for  $\mathcal{K}_{\text{tree}}$ .

It is worth remarking that the following are valid over total orders:

$$FPp \rightarrow Pp \vee p \vee Fp, \quad PFP \rightarrow Pp \vee p \vee Fp.$$

To find deductions of them in  $\mathbf{L}_2$  is a nontrivial exercise. As a matter of fact, these two items could have been used instead of (A2a, b) as axioms for total orders. One could equally well have used their contrapositives:

$$Hp \wedge p \wedge Gp \rightarrow GHp, \quad Hp \wedge p \wedge Gp \rightarrow HGP.$$

The converses of these four items are valid over partial orders.

### 3.3 No Extremals (No Maximals, No Minimals)

Let  $\mathbf{L}_3$  (resp.  $\mathbf{L}_4$ ) be the extension of  $\mathbf{L}_2$  obtained by adding (A3a, b) (resp. (A4a, b)) as extra axioms. Let  $\mathcal{K}_3$  (resp.  $\mathcal{K}_4$ ) be the class of total orders having (resp. not having) a maximum and a minimum. Beyond understanding the notions of consistency and MCS relative to  $\mathbf{L}_3$  or  $\mathbf{L}_4$  as the case may be, no modification in the work of Section 3.2 above is needed to prove  $\mathbf{L}_3$  complete for  $\mathcal{K}_3$  and  $\mathbf{L}_4$  for  $\mathcal{K}_4$ . The following observations suffice:

On the one hand, understanding consistency and MCS relative to  $\mathbf{L}_3$ , if  $(X, R)$  is any total order and  $T$  any perfect chronicle on it, then for any  $x \in X$ , either  $G\perp \in T(x)$  itself, or  $FG\perp \in T(x)$  and so  $G\perp \in t(y)$  for some  $y$  with  $xRy$ —this by (A3a). But if  $G\perp \in T(z)$ , then with  $w$  with  $zRw$  would have to have  $\perp \in T(w)$ , which is impossible so  $z$  must be the maximum of  $(X, R)$ . Similarly, A3b guarantees the existence of a minimum in  $(X, R)$ .

On the other hand, understanding consistency and MCS relative to  $\mathbf{L}_4$ , if  $(X, R)$  is any total order and  $T$  any perfect chronicle on it, then for any



Table 1.

$GGHp \approx GHp$	$FGHp \approx GHp$
$GFHp \approx GHp$	$FFHp \approx FHp$
$GPGp \approx Gp$	$FPGp \approx FGp$
$GPHp \approx PHp$	$FPHp \approx PHp$
$GFGp \approx FGp$	$FFGp \approx FGp$
$GHPp \approx HPp$	$FHPp \approx HPp$
$GGFp \approx GFp$	$FGFp \approx GFp$
$GGPp \approx GPp$	$FGPp \approx FPp$
$GHFp \approx GFp$	$FHFp \approx Fp$
$GFPp \approx FPp$	$FFPp \approx FPp$

Similarly, the extension  $\mathbf{L}_Q$  of  $\mathbf{L}_2$  obtained by adding (A4a, b) and (A5a) is complete for the class of dense total orders without maximum or minimum. A famous theorem tells us that any *countable* order of this class is isomorphic to the rational numbers in their usual order. Since our method of proof always produces a *countable* frame, we can conclude that  $\mathbf{L}_Q$  is the tense logic of the rationals. The accompanying diagram (1) indicates some implications that are valid over dense total orders without maximum or minimum, and hence theses of  $\mathbf{L}_Q$ ; no further implications among the formulas considered are valid. A theorem of C. L. Hamblin tells us that in  $\mathbf{L}_Q$  any sequence of  $G$ s,  $H$ s,  $F$ s and  $P$ s prefixed to the variable  $p$  is provably equivalent to one of the 15 formulas in our diagram. It obviously suffices to prove this for sequences of length three. The reductions listed in the accompanying Table 1 together with their mirror images, suffice to prove this. It is a pleasant exercise to verify all the details.

### 3.5 Discreteness

The extension  $\mathbf{L}_6$  of  $\mathbf{L}_2$  obtained by adding (A6a, b) is complete for the class  $\mathcal{K}_6$  of total orders in which every element has an immediate successor and an immediate predecessor. The proof involves quite a few modifications in the work of Section 3.2 above, beginning with:

LEMMA *For any MCS  $A$  there exists an MCS  $B$  such that:*

1. *whenever  $F\gamma \in A$  then  $\gamma \vee F\gamma \in B$ .*

*Moreover, any such MCS further satisfies:*

2. *whenever  $P\delta \in B$ , then  $\delta \vee P\delta \in A$ ,*

3. whenever  $A \rightarrow C$ , then either  $B = C$  or  $B \rightarrow C$ ,
4. whenever  $C \rightarrow B$ , then either  $A = C$  or  $C \rightarrow A$ .

**Proof.**

1. The problem quickly reduces to proving the consistency of any finite set of formulas of the forms  $P\alpha$  for  $\alpha \in A$  and  $\gamma \vee F\gamma$  for  $F\gamma \in A$ . To establish this, one notes that the following is valid over total orders, hence a thesis of  $\mathbf{L}_2$  and *a fortiori* of  $\mathbf{L}_6$ :

$$Fp_0 \wedge Fp_1 \wedge \dots \wedge Fp_n \rightarrow \\ F((p_0 \vee Fp_0) \wedge (p_1 \vee Fp_1) \wedge \dots \wedge (p_n \vee Fp_n))$$

2. We prove the contrapositive. Suppose  $\delta \vee P\delta \notin A$ . By (A6a),  $FH\neg\delta \in A$ . by part (1),  $H\neg\delta \vee FH\neg\delta \in B$ . But  $FHp \rightarrow Hp$  is valid over total orders, hence a thesis of  $\mathbf{L}_2$  and *a fortiori* of  $\mathbf{L}_6$ . So  $H\neg\delta \in B$  and  $P\delta \notin B$  as required.
3. Assume for contradiction that  $A \rightarrow C$  but neither  $B = C$  nor  $B \rightarrow C$ . Then there exist a  $\gamma_0 \in C$  with  $\gamma_0 \notin B$  and a  $\gamma_1 \in C$  with  $F\gamma_1 \notin B$ . Let  $\gamma = \gamma_0 \wedge \gamma_1$ . Then  $\gamma \in C$  and since  $A \rightarrow C$ ,  $F\gamma \in A$ . but  $\gamma \vee F\gamma \notin B$ , contrary to (1).
4. Similarly follows from (2). ■

We write  $A \rightarrow' B$  to indicate that  $A, B$  are related as in the above Lemma. Intuitively this means that a situation of the sort described by  $A$  could be *immediately* followed by one of the sort described by  $B$ .

We now take  $M$  to be the set of *quadruples*  $(X, R, S, T)$  where on the one hand, as always  $X$  is a nonempty finite subset of  $W$ ,  $R$  a total order on  $X$ , and  $T$  a coherent chronicle on  $(X, R)$ ; while on the other hand, we have:

4. whenever  $xSy$ , then  $y$  immediately succeeds  $x$  in  $(X, R)$ ,
5. whenever  $xSy$ , then  $T(x) \rightarrow' T(y)$ ,

Intuitively  $xSy$  means that no points are ever to be added between  $x$  and  $y$ . We say  $(X', R', S', T')$  *extends*  $(X, R, S, T)$  if on the one hand, as always, Definition 10(1', 2', 3') hold; while on the other hand,  $S \subseteq S'$ . In addition to requirements of the form 8(1, 2) we need to consider requirements of the form:

5. there exists a  $y$  with  $xSy$ ,
4. there exists a  $y$  with  $ySx$ .

To ‘kill’ a requirement of form (5), take an MCS  $B$  with  $T(x) \rightarrow 3' B$ . If  $x$  is the maximum of  $(X, R)$  it suffices to fix  $z \in W - X$  and set:

$$\begin{aligned} X' &= X \cup \{z\}, & R' &= R \cup \{(x, z)\} \cup \{(v, z) : vRx\}, \\ S' &= S \cup \{(x, z)\}, & T' &= T \cup \{(z, B)\} \end{aligned}$$

Otherwise, let  $y$  immediately succeed  $x$  in  $(X, R)$ . If  $B = T(y)$  set:

$$\begin{aligned} X' &= X, & R' &= R, \\ S' &= S \cup \{(x, y)\} & T' &= T. \end{aligned}$$

Otherwise, we have  $B \rightarrow 3T(y)$ , and it suffices to fix  $z \in W - X$  and set:

$$\begin{aligned} X' &= X, & R' &= R \cup \{(x, z), (z, y)\} \cup \\ & & & \cup \{(v, z) : vRx\} \cup \{(z, v) : yRv\}, \\ S' &= S \cup \{(x, z)\}, & T' &= T \cup \{(z, B)\} \end{aligned}$$

Similarly, to kill a requirement of form (6) we use the mirror image of the Lemma above, proved using (A6b).

It is also necessary to check that when  $xSy$  we never need to insert a point between  $x$  and  $y$  in order to kill a requirement of form 8(1) or (2). Reviewing the construction of Section 3.2 above, this follows from parts (3), (4) of the Lemma above. The remaining details are left to the reader.

A total order is *discrete* if every element but the maximum (if any) has an immediate successor, and every element but the minimum (if any) has an immediate predecessor. The foregoing argument establishes that we get a complete axiomatisation for the tense logic of discrete total orders by adding to  $\mathbf{L}_2$  the following weakened versions of (A6a, b):

$$p \wedge Hp \rightarrow G \perp \vee FHp, \quad p \wedge Gp \rightarrow H \perp \vee PGp.$$

A total order is *homogeneous* if for any two of its points  $x, y$  there exists an automorphism carrying  $x$  to  $y$ . Such an order cannot have a maximum or minimum and must be either dense or discrete. In Burgess [1979] it is indicated that a complete axiomatisation of the tense logic of homogeneous orders is obtainable by adding to  $\mathbf{L}_4$  the following which should be compared with (A5a) and (A6a, b):

$$(Fp \rightarrow FFp) \vee [(q \wedge Hq \rightarrow FHq) \wedge (q \wedge Gq \rightarrow PGq)].$$

### 3.6 Continuity

A *cut* in a total order  $(X, R)$  is a partition  $(Y, Z)$  of  $X$  into two nonempty pieces, such that whenever  $y \in Y$  and  $z \in Z$  we have  $yRz$ . A *gap* is a cut  $(Y, Z)$  such that  $Y$  has no maximum and  $Z$  no minimum.  $(X, R)$  is complete if it has no gaps. The *completion*  $(X^+, R^+)$  of a total order  $(X, R)$  is the complete total order obtained by inserting, for each gap  $(Y, Z)$  in  $(X, R)$ ,

an element  $w(Y, Z)$  after all elements of  $Y$  and before all elements of  $Z$ . For example, the completion of the rational numbers in their usual order is the real numbers in their usual order. The extension  $\mathbf{L}_7$  of  $\mathbf{L}_2$  obtained by adding (A7a, b) is complete for the class  $\mathcal{K}_7$  of complete total orders. The proof requires a couple of Lemmas:

**LEMMA** *Let  $T$  be a perfect chronicle on a total order  $(X, R)$ , and  $(Y, Z)$  a gap in  $(X, R)$ . Then if  $G\alpha \in T(z)$  for all  $z \in Z$ , then  $G\alpha \in T(y)$  for some  $y \in Y$ .*

**Proof.** Suppose for contradiction that  $G\alpha \in T(z)$  for all  $z \in Z$  but  $F\neg\alpha \approx \neg G\alpha \in T(y)$  for all  $y \in Y$ . For any  $y_0 \in Y$  we have  $F\neg\alpha \wedge FG\alpha \in T(y_0)$ . Hence, by A7a,  $F(G\alpha \wedge HF\neg\alpha) \in T(y_0)$ , and there is an  $x$  with  $y_0Rx$  and  $G\alpha \in HF\neg\alpha \in T(x)$ . But this is impossible, since if  $x \in Y$  then  $G\alpha \notin T(x)$ , while if  $x \in Z$  then  $HF\neg\alpha \notin T(x)$ . ■

**LEMMA** *Let  $T$  be a perfect chronicle on a total order  $(X, R)$ . Then  $T$  can be extended to a perfect chronicle  $T^+$  on its completion  $(X^+, R^+)$ .*

**Proof.** For each gap  $(Y, Z)$  in  $(X, R)$ , the set:

$$C(Y, Z) = \{P\alpha : \exists y \in Y(\alpha \in T(y))\} \cup \{F\alpha : \exists z \in Z(\alpha \in T(z))\}$$

is consistent. This is because any finite subset, involving only  $y_1, \dots, y_m$  from  $Y$  and  $z_1, \dots, z_n$  from  $Z$  will be contained in  $T(x)$  where  $x$  is any element of  $Y$  after all the  $y_i$  or any element of  $Z$  before all the  $z_j$ . Hence, we can define a coherent chronicle  $T^+$  on  $(X^+, R^+)$  by taking  $T^+(w(Y, Z))$  to be some MCS extending  $C(Y, Z)$ . Now if  $F\alpha \in T^+(w(Y, Z))$ , we claim that  $F\alpha \in T(z)$  for some  $z \in Z$ . For if not, then  $G\neg\alpha \in T(z)$  for all  $z \in Z$ , and by the previous Lemma,  $G\neg\alpha \in T(y)$  for some  $y \in Y$ . But then  $PG\neg\alpha$ , which implies  $\neg F\alpha$ , would belong to  $C(Y, Z) \subseteq T^+(w(Y, Z))$ , a contradiction. It hardly needs saying that if  $F\alpha \in T(z)$ , then there is some  $x$  with  $zRx$  and *a fortiori*  $w(Y, Z)R^+x$  having  $\alpha \in T(x)$ . This shows  $T^+$  is prophetic. Axiom (A7b) gives us a mirror image to the previous Lemma, which can be used to show  $T^+$  historic. ■

To prove the completeness of  $\mathbf{L}_7$  for  $\mathcal{K}_7$ , given a consistent  $\gamma_0$  use the work of Section 2.2 above to construct a perfect chronicle  $T$  on a frame  $(X, R)$  such that  $\gamma_0 \in T(x_0)$  for some  $x_0$ . Then use the foregoing Lemma to extend to a perfect chronicle on a complete total order, as required to prove satisfiability. ■

Similarly,  $\mathbf{L}_R$ , the extension of  $\mathbf{L}_2$  obtained by adding (A4a, b) and (A5a) and (A7a, b) is complete for the class of complete dense total orders without maximum or minimum, sometimes called *continuous* orders. As a matter of fact, our construction shows that any formula consistent with this theory is satisfiable in the completion of the rationals, that is, in the reals. Thus  $\mathbf{L}_R$  is the tense logic of real time and, hence, of the time of classical physics.

### 3.7 Well-Orders

The extension  $\mathbf{L}_8$  of  $\mathbf{L}_2$  obtained by adding (A8) is complete for the class  $\mathcal{K}_8$  of all well-orders. For the proof it is convenient to introduce the abbreviations  $Ip$  for  $Pp \vee p \vee Fp$  or ‘ $p$  sometime’, and  $Bp$  for  $p \wedge \neg Pp$  or ‘ $p$  for the first time’. an easy consequence of (A8) is  $Ip \rightarrow IBp$ : if something *ever* happens, then there is a *first* time when it happens the reader can check that the following are valid over total orders; hence, theses of ( $\mathbf{L}_2$  and *a fortiori* of  $\mathbf{L}_9$ ):

1.  $Ip \wedge Iq \rightarrow I(Pp \wedge q) \vee I(p \wedge q) \vee I(p \wedge Pq)$ ,
2.  $I(q \wedge Fr) \wedge I(PBp \wedge Bq) \rightarrow I(p \wedge Fr)$ .

Now, understanding consistency, MCS, and related notions relative to  $\mathbf{L}_8$ , let  $\delta_0$  be any consistent formula and  $D_0$  any MCS containing it. Let  $\delta_1, \dots, \delta_k$  be all the proper subformulas of  $\delta_0$ . Let  $\Gamma$  be the set of formulas of form

$$(\neg)\delta_0 \wedge (\neg)\delta_1 \wedge \dots \wedge (\neg)\delta_k$$

where each  $\delta_i$  appears once, plain or negated. Note that distinct elements of  $\Gamma$  are truth-functionally inconsistent. Let  $\Gamma' = \{\gamma \in \Gamma : I\gamma \in D_0\}$ . Note that for each  $\gamma \in \Gamma'$  we have  $IB\gamma \in D_0$ , and that for distinct  $\gamma, \gamma' \in \Gamma'$  we must by (1) have either  $I(PB\gamma \wedge B\gamma')$  or  $I(PB\gamma' \wedge B\gamma)$  in  $D_0$ . Enumerate the elements of  $\Gamma'$  as  $\gamma_0, \gamma_1, \dots, \gamma_n$  so that  $I(PB\gamma_i \wedge B\gamma_j) \in D_0$  iff  $i < j$ . We write  $i \triangleleft j$  if  $I(\gamma_i \wedge F\gamma_j) \in D_0$ . This clearly holds whenever  $i < j$ , but may also hold in other cases. A crucial observation is:

$$(+) \quad \text{If } i < j \leq k \text{ and } k \triangleleft i, \text{ then } j \triangleleft i$$

This follows from (2). These tedious preliminaries out of the way, we will now define a set  $X$  of ordinals and a function  $t$  from  $X$  to  $\Gamma'$ . Let  $a, b, c, \dots$  range over *positive* integers:

- We put  $0 \in X$  and set  $t(0) = \gamma_0$ .
- If  $0 \triangleleft 0$  we also put each  $a \in X$  and set  $t(a) = \gamma_0$ .
- We put  $\omega \in X$  and set  $t(\omega) = \gamma_1$ .
- If  $1 \triangleleft 1$  we also put each  $\xi = \omega \cdot b \in X$  and set  $t(\xi) = \gamma_1$ .
- If  $1 \triangleleft 0$  we also put each  $\xi = \omega \cdot b + a \in X$  and set  $t(\xi) = \gamma_0$ .
- We put  $\omega^2 \in X$  and set  $t(\omega^2) = \gamma_2$ .
- If  $2 \triangleleft 2$  we also put each  $\xi = \omega^2 \cdot c \in X$  and set  $t(\xi) = \gamma_2$ .
- If  $2 \triangleleft 1$  we also put each  $\xi = \omega^2 \cdot c + \omega \cdot b \in X$ , and set  $t(\xi) = \gamma_1$ .
- If  $2 \triangleleft 0$  we also put each  $\xi = \omega^2 \cdot c + \omega \cdot b + a \in X$  and set  $t(\xi) = \gamma_0$ .
- and so on.

Using (+) one sees that whenever  $\xi, \eta \in X$  and  $\xi < \eta$ , then  $i \triangleleft j$  where  $t(\xi) = \gamma_i$  and  $t(\eta) = \gamma_j$ . Conversely, inspection of the construction shows that:

1. whenever  $\xi \in X$  and  $t(\xi) = \gamma_j$  and  $j \triangleleft k$ , then there is an  $\eta \in X$  with  $\xi < \eta$  and  $t(\eta) = \gamma_k$
2. whenever  $\xi \in X$  and  $t(\xi) = \gamma_j$  and  $i < j$ , then there is an  $\eta \in X$  with  $\eta < \xi$  and  $t(\eta) = \gamma_i$ .

For  $\xi \in X$  let  $T(\xi)$  be the set of conjuncts of  $t(\xi)$ . Using (1) and (2) one sees that  $T$  satisfies all the requirements 8(1,2,3,4) for a perfect chronicle, *so far as these pertain to subformulas of  $\delta_0$* . Inspection of the proof of Lemma 9 then shows that this suffices to prove  $\delta_0$  satisfiable in the well-order  $(X, <)$ . ■

Without entering into details here, we remark that variants of  $\mathbf{L}_8$  provide axiomatisations of the tense logics of the integers, the natural numbers, and of finite total orders. In particular, for the natural numbers one uses  $\mathbf{L}_\omega$ , the extension of  $\mathbf{L}_2$  obtained by adding (A8) and  $p \wedge Gp \rightarrow H\perp \vee PGp$ .  $\mathbf{L}_\omega$  is the tense logic of the notion of time appropriate for discussing the working of a digital computer, or of the mental mathematical constructions of Brouwer's 'creative subject'.

### 3.8 Lattices

The extension  $\mathbf{L}_9$  of  $\mathbf{L}_1$  obtained by adding (A4a, b) and (A9a, b) is complete for the class  $\mathcal{K}_9$  of partial orders without maximal or minimal elements in which any two elements have an upper and a lower bound. We sketch the modifications in the work of Section 3.2 above needed to prove this:

To begin with, we must revise clause 10(2) in the definition of  $M$  to read:

- 2g.  $R$  is a partial order on  $X$  having a maximum and a minimum.

This necessitates revisions in the proof of the Killing Lemma, for which the following will be useful:

**LEMMA** *Let  $A, B, C$  be MCSs. If  $A \rightarrow B$  and  $A \rightarrow C$ , then there exists an MCS  $D$  such that  $B \rightarrow D$  and  $C \rightarrow D$ .*

**Proof.** The problem quickly reduces to showing  $\{\beta : G\beta \in B\} \cup \{\gamma : G\gamma \in C\}$  consistent. For this it suffices (using 3(4)) to show that  $\beta \wedge \gamma$  is consistent whenever  $G\beta \in B, G\gamma \in C$ . Now in that case we have  $FG\beta, FG\gamma \in A$ , since  $A \rightarrow B, C$ . By A9a, we then have  $GF\beta \in A$ , and by 3(2) we then have  $F(F\beta \wedge G\gamma) \in A$  and  $FF(\beta \wedge \gamma) \in A$ , which suffices to prove  $\beta \wedge \gamma$  consistent as required. ■

Turning now to the Killing Lemma, trouble arises when for a given  $(X, R, T) \in M$  a requirement of form Definition 8(1) is said to be 'killed' for some  $x$  other than the maximum  $y$  of  $(X, R)$  and some  $F\gamma \in T(x)$ .



Fixing an MCS  $B$  with  $T(x) \rightarrow B$  and  $\gamma \in B$ , and  $az \in W - X$ , we would like to add  $z$  to  $x$  placing it after  $x$  and assigning it the MCS  $B$ . But we cannot simply do this, else the resulting partial order would have no maximum. (For  $y$  and  $z$  would be incomparable.) So we apply the Lemma (with  $A = T(x), C = T(y)$ ) to obtain an MCS  $D$  with  $B \rightarrow D$  and  $T(y) \rightarrow D$ . We fix a  $w \in W - X$  distinct from  $z$ , and set:

$$\begin{aligned} X' &= X \cup \{z, w\}, \\ R' &= R \cup \{(x, z), (z, w)\} \cup \{(v, z) : vRx\} \cup \{(v, w) : v \in X\}. \\ T' &= T \cup \{(z, B), (w, D)\}. \end{aligned}$$

Similarly, a requirement of form 8(2) involving an element other than the minimum is treated using the mirror image of the Lemma above, proved using (A9b).

Now given a formula  $\gamma_0$  consistent with  $\mathbf{L}_9$ , the construction of Definition 10 above produces a perfect chronicle  $T$  on a partial order  $(X, R)$  with  $\gamma_0 \in T(x_0)$  for some  $x_0$ . The work of Section 2.4 above shows that  $(X, R)$  will have no maximal or minimal elements. Moreover,  $(X, R)$  will be a union of partial orders  $(X_n, R_n)$  satisfying (2g). Then any  $x, y \in X$  will have an  $R$ -upper bound and an  $R$ -lower bound, namely the  $R_n$ -maximum and  $R_n$ -minimum elements of any  $X_n$  containing them both. Thus,  $(X, R) \in \mathcal{K}_9$  and  $\gamma_0$  is satisfiable over  $\mathcal{K}_9$ . ■

A *lattice* is a partial order in which any two elements have a *least* upper bound and a *greatest* lower bound. Actually, our proof shows that  $\mathbf{L}_9$  is complete for the class of lattices without maximum or minimum. It is worth mentioning that (A9a, b) could have been replaced by:

$$Fp \wedge Fq \rightarrow F(Pp \wedge Pq), \quad Pp \wedge Pq \rightarrow P(Fp \wedge Fq).$$

Weakened versions of these axioms can be used to give an axiomatisation for the tense logic of arbitrary lattices.

#### 4 THE DECIDABILITY OF TENSE LOGICS

All the systems of tense logic we have considered so far are recursively decidable. Rather than give an exhaustive (and exhausting) survey, we treat here two examples, illustrating the two basic methods of proving decidability: one method, borrowed from modal logic, is that of using so-called *filtrations* to establish what is known as the *finite model property*. The other, borrowed from model theory, is that of using so-called *interpretations* in order to be able to exploit a powerful theorem of [Rabin, 1966].

**THEOREM 13.**  $\mathbf{L}_9$  is decidable.

**Proof.** Let  $\mathcal{K}$  be the class of models of (B1) and (B9a, b); thus  $\mathcal{K}$  is like  $\mathcal{K}_9$  except that we do *not* require antisymmetry. Let  $\mathcal{K}'$  be the class of finite

elements of  $\mathcal{K}$ . It is readily verified that  $\mathbf{L}_9$  is sound for  $\mathcal{K}$  and *a fortiori* for  $\mathcal{K}'$ . We claim that  $\mathbf{L}_9$  is complete for  $\mathcal{K}'$ . This provides an effective procedure for testing whether a given formula  $\alpha$  is a thesis of  $\mathbf{L}_9$  or not, as follows: search simultaneously through all deductions in the system  $\mathbf{L}_9$  and through all members of  $\mathcal{K}'$ —or more precisely, of some nice countable subclass of  $\mathcal{K}'$  containing at least one representative of each isomorphism-type. Eventually one either finds a deduction of  $\alpha$ , in which case  $\alpha$  is a thesis, or one finds an element of  $\mathcal{K}'$  in which  $\neg\alpha$  is satisfiable, in which case by our completeness claim,  $\alpha$  is not a thesis.

To prove our completeness claim, let  $\gamma_0$  be consistent with  $\mathbf{L}_9$ . We showed in Section 2.9 above how to construct a perfect chronicle  $T$  on a frame  $(X, R) \in \mathcal{K}_9 \subseteq \mathcal{K}$  having  $\gamma_0 \in T(x_0)$  for some  $x_0$ . For  $x \in X$ , let  $t(x)$  be the set of subformulas of  $\gamma_0$  in  $T(x)$ . Define an equivalence relation on  $X$  by:

$$x \leftrightarrow y \text{ iff } t(x) = t(y).$$

Let  $[x]$  denote the equivalence class of  $x$ ,  $X'$  the set of all  $[x]$ . Note that  $X'$  is finite, having no more than  $2^k$  elements, where  $k$  is the number of subformulas of  $\gamma_0$ . Consider the relations on  $X'$  defined by:

$$\begin{aligned} aR^+b & \text{ iff } xRy \text{ for some } x \in a \text{ and } y \in b, \\ aR'b & \text{ iff for some finite sequence } a = c_0, c_1, \dots, c_{n-1}, c_n = b \\ & \text{ we have } c_iR^+c_{i+1} \text{ for all } i < n. \end{aligned}$$

Clearly  $R'$  is transitive, while  $R^+$  and, hence,  $R'$  inherit from  $R$  the properties expressed by B9a, b. Thus  $(X', R') \in \mathcal{K}'$ . Define a function  $t'$  on  $X'$  by letting  $t'(a)$  be the common value of  $t(x)$  for all  $x \in a$ . In particular for  $a_0 = [x_0]$  we have  $\gamma_0 \in t'(a_0)$ . We claim that  $t'$  satisfies clauses 8(1, 2, 3, 4) of the definition of a perfect chronicle *so far as these pertain to subformulas of  $\gamma_0$* . As remarked in Section 3.8 above, this suffices to show  $\gamma_0$  satisfiable in  $(X', R')$  and, hence, satisfiable over  $\mathcal{K}'$  as required.

In connection with Definition 8(1), what we must show is:

$$1. \text{ whenever } F\gamma \in t(a) \text{ there is a } b \text{ with } aR'b \text{ and } \gamma \in t(b)$$

Well, let  $a = [x]$ , so  $F\gamma \in t(x) \subseteq T(x)$ . There is a  $y$  with  $xRy$  and  $\gamma \in t(y)$  since  $T$  is prophetic. Letting  $b = [y]$  we have  $aR^+b$  and so  $aR'b$ .

In connection with Definition 8(3) what we must show is:

$$3'. \text{ whenever } G\gamma \in t(a) \text{ and } aR'b, \text{ then } \gamma \in t(b).$$

For this it clearly suffices to show:

$$3^+ \text{ whenever } G\gamma \in t(a) \text{ and } aR^+b, \text{ then } \gamma \in t(b) \text{ and } G\gamma \in t(b).$$

To show this, assuming the two hypotheses, fix  $x \in a$  and  $y \in b$  with  $xRy$ . We have  $G\gamma \in t(x) \subseteq T(x)$ , so by (A1a),  $GG\gamma \in T(x)$ . Hence,  $\gamma \in t(y)$  and  $G\gamma \in t(y)$ , since  $T$  is coherent—which completes the proof.

Definitions 8(2, 4) are treated similarly. ■

**THEOREM 14.**  $\mathbf{L}_R$  is decidable.

**Proof.** We introduce an alternative definition of *validity* which is useful in other contexts. To each tense-logical formula  $\alpha$  we associate a first-order formula  $\hat{\alpha}$  as follows: for a sentential variable  $p_i$  we set  $\hat{p}_i = P_i(x)$  where  $P_i$  is a one-place predicate variable. We then proceed inductively:

$$\begin{aligned} (\neg\alpha)^\wedge &= \neg\hat{\alpha}, \\ (\alpha \wedge \beta)^\wedge &= \hat{\alpha} \wedge \hat{\beta} \\ (G\alpha)^\wedge &= \forall y(x < y \rightarrow \hat{\alpha}(y/x)), \\ (H\alpha)^\wedge &= \forall y(y < x \rightarrow \hat{\alpha}(y/x)). \end{aligned}$$

Here  $(y/x)$  represents the result of substituting for  $x$  the alphabetically first variable  $y$  not occurring yet. Given a valuation  $V$  in a frame  $(X, R)$  we have an interpretation in the sense of first-order model theory, in which  $R$  interprets the symbol  $<$  and  $V(p_i)$  the symbol  $P_i$ . Unpacking the definitions it is entirely trivial that we always have:

$$(*) \quad a \in V(\alpha) \text{ iff } (X, R, V(p_0), V(p_1), V(p_2), \dots) \models \hat{\alpha}(x),$$

where  $\models$  is the usual satisfaction relation of model theory. We now further define:

$$a^+ = \forall P_0 \forall P_1, \dots, \forall P_k \forall x \hat{\alpha}(x),$$

where  $p_0, p_1, \dots, p_k$  include all the variables occurring in  $\alpha$ . Note that  $\alpha^+$  is a second-order formula of the simplest kind: it is *monadic* (all its second-order variables are *one-* place predicate variables) and *universal* (consisting of a string of universally-quantified second-order variables prefixed to a first-order formula). It is entirely trivial that:

$$(+)$$

$$\alpha \text{ is valid in } (X, R) \text{ iff } (X, R) \models \alpha^+$$

It follows that to prove the decidability of the tense logic of a given class  $\mathcal{K}$  of frames it will suffice to prove the decidability of the set of universal monadic (second-order) formulas true in all members of  $\mathcal{K}$ .

Let  $2^{<\omega}$  be the set of all finite 0,1-sequences. Let  $*0$  be the function assigning the argument  $s = (i_0, i_1, \dots, i_n) \in 2^{<\omega}$  the value  $s * 0 = (i_0, i_1, \dots, i_n, 0)$ , and similarly for  $*1$ . Rabin proves the decidability of the set  $S2S$  of monadic (second order) formulas true in the structure  $(2^{<\omega}, *0, *1)$ . He deduces as an easy corollary the decidability of the set of monadic formulas true in the frame  $(\mathbb{Q}, <)$  consisting of the rational numbers with their usual order. This immediately yields the decidability of the system  $\mathbf{L}_Q$  of Section 2.5 above. Further corollaries relevant to tense logic are the decidability of the set of monadic formulas true in all countable total orders, and similarly for countable well-orders.

It only remains to reduce the decision problem for  $\mathbf{L}_R$  to that for  $\mathbf{L}_Q$ . The work of 2.7 above shows that a formula  $\alpha$  is satisfiable in the frame  $(\mathbb{R}, <)$  consisting of the real numbers with their usual order, iff it is satisfiable in the frame  $(\mathbb{Q}, <)$  by a valuation  $V$  with the property:

1.  $V(\alpha) = \mathbb{Q}$  for every substitution instance  $\alpha$  of (A7a or b).

Inspection of the proof actually shows that it suffices to have:

2.  $V(\alpha') = \mathbb{Q}$  where  $\alpha'$  is the conjunction of all instances of (A7a or b) obtainable by substituting subformulas of  $\alpha$  for variables.

A little thought shows that this amounts to demanding:

3.  $V(\alpha \wedge GH\alpha') \neq \emptyset$ .

In other words,  $\alpha$  is satisfiable in  $(\mathbb{R}, <)$  iff  $\alpha \wedge GH\alpha'$  is satisfiable in  $(\mathbb{R}, <)$ , which effects the desired reduction. For the lengthy original proof see [Bull, 1968]. Other applications of Rabin's theorem are in [Gabbay, 1975]. Rabin's proof uses automata-theoretic methods of Büchi; these are avoided by [Shelah, 1975]. ■

## 5 TEMPORAL CONJUNCTIONS AND ADVERBS

### 5.1 *Since, Until, Uninterruptedly, Recently, Soon*

All the systems discussed so far have been based on the primitives  $\neg, \wedge, G, H$ . It is well-known that any truth function can be defined in terms of  $\neg, \wedge$ . Can we say something comparable about temporal operators and  $G, H$ ? When this question is formulated precisely, the answer is a resounding NO.

**DEFINITION 15.** Let  $\varphi$  be a first-order formula having one free variable  $x$  and no nonlogical symbols but the two-place predicate  $<$  and the one-place predicates  $P_1, \dots, P_n$ . corresponding to  $\varphi$  we introduce a new  $n$ -place connective, the (first-order, one-dimensional) temporal operator  $O(\varphi)$ . We describe the formal semantics of  $O(\varphi)$  in terms of the alternative approach of Theorem 14 above: we add to the definition of  $\hat{\phantom{a}}$  the clause:

$$(O(\varphi)(\alpha_1, \dots, \alpha_n))^\wedge = \varphi(\hat{\alpha}_1/P_1, \dots, \hat{\alpha}_n/P_n).$$

Here  $\hat{\alpha}/P$  denotes substitution of the formula  $\hat{\alpha}$  for the predicate variable  $P$ . We then let formula (\*) of Theorem 14 above *define*  $V(\alpha)$  for formulas  $\alpha$  involving  $O(\varphi)$ . Examples 16 below illustrate this rather involved definition. If  $\mathcal{O} = \{O(\varphi_1), \dots, O(\varphi_k)\}$  is a set of temporal operators, an  $\mathcal{O}$ -*formula* is one built up from sentential variables using  $\neg, \wedge$ , and elements of  $\mathcal{O}$ . A temporal operator  $O(\varphi)$  is  $\mathcal{O}$ -*definable* over a class  $\mathcal{K}$  of frames if there is an  $\mathcal{O}$ -formula  $\alpha$  such that  $O(\varphi)(p_1, \dots, p_n) \leftrightarrow \alpha$  is valid over  $\mathcal{K}$ .  $\mathcal{O}$  is

*temporally complete* over  $\mathcal{K}$  if every temporal operator is  $\mathcal{O}$ -definable over  $\mathcal{K}$ . Note that the smaller  $\mathcal{K}$  is—it may consist of a single frame—the easier it is to be temporally complete over it.

EXAMPLES 16.

1.  $\forall y(x < y \rightarrow P_1(y))$
2.  $\forall y(y < x \rightarrow P_1(y))$
3.  $\exists y(x < y \wedge \forall z(x < z \wedge z < y \rightarrow P_1(z)))$
4.  $\exists y(y < x \wedge \forall z(y < z \wedge z < x \rightarrow P_1(z)))$
5.  $\exists y(x < y \wedge P_1(y) \wedge \forall z(y < z \wedge z < x \rightarrow P_1(z)))$

For (1),  $O(\varphi)$  is just  $G$ . For (2),  $O(\varphi)$  is just  $H$ . For (3),  $O(\varphi)$  will be written  $G'$ , and may be read ‘ $p$  is going to be uninterruptedly the case for some time’. For (4),  $O(\varphi)$  will be written  $H'$ , and may be read ‘ $p$  has been uninterruptedly the case for some time’. For (5),  $O(\varphi)$  will be written  $U$ , and  $U(p, q)$  may be read ‘until  $p, q$ ’; it predicts a future occasion of  $p$ ’s being the case, up until which  $q$  is going to be uninterruptedly the case. For (6),  $O(\varphi)$  will be written  $S$ , and  $S(p, q)$  may be read ‘since  $p, q$ ’. In terms of  $G'$  we define  $F' = \neg G; \neg$ , read ‘ $p$  is going to be the case arbitrarily soon’. In terms of  $H'$  we define  $P' = \neg H' \neg$ , read ‘ $p$  has been the case arbitrarily recently’. Over all frames,  $Gp$  is definable as  $\neg U(\neg p, \top)$ , and  $G'$  as  $U(\top, p)$ . Similarly,  $H$  and  $H'$  are definable in terms of  $S$ . The following examples are due to H. Kamp:

PROPOSITION 17.  $G'$  is not  $G, H$ -definable over the frame  $(\mathbb{R}, <)$ .

**Sketch of Proof.** Define two valuations over that frame by:

$$V(p) = \{0, \pm 1, \pm 2, \pm 3, \dots\} \quad W(p) = V(p) \cup \{\pm \frac{1}{2}, \pm \frac{1}{4}, \pm \frac{1}{8}, \dots\}$$

Then intuitively it is plausible, and formally it can be proved that for any  $G, H$ -formula  $\alpha$  we have  $0 \in V(\alpha)$  iff  $0 \in W(\alpha)$ . But  $0 \in V(G'p) - W(G'p)$ . ■

PROPOSITION 18.  $U$  is not  $G, H, G', H'$ -definable over the frame  $(\mathbb{R}, <)$ .

**Sketch of Proof.** Define two valuations by:

$$\begin{aligned} V(p) &= \{\pm 1, \pm 2, \pm 3, \pm 4, \dots\} & W(p) &= \{\pm 2, \pm 3, \pm 4, \dots\} \\ V(q) &= W(q) = \text{the union of the open intervals} \\ &\dots, (-5, -4), (-3, -2), (-1, +1), \\ &(+2, +3), (+4, +5), \dots \end{aligned}$$

Then intuitively it is plausible, and formally it can be proved that for any  $G, H, G', H'$ -formula  $\alpha$  we have  $0 \in V(\alpha)$  iff  $0 \in W(\alpha)$ . But  $0 \in V(U(p, q) - W(U(p, q)))$ . ■

Such examples might inspire pessimism, but [Kamp, 1968] proves:

**THEOREM 19.** *The set  $\{U, S\}$  is temporally complete over continuous orders.*

We will do no more than outline the difficult proof (in an improved version due to Gabbay): Let  $\mathcal{O}$  be a set of temporal operators,  $\mathcal{K}$  a class of frames. An  $\mathcal{O}$ -formula  $\alpha$  is *purely past* over  $\mathcal{K}$  if whenever  $(X, R) \in \mathcal{K}$  and  $x \in K$  and  $V, W$  are valuations in  $(X, R)$  agreeing before  $x$  (so that for all  $i$ ,  $V(p_i) \cap \{y : yRx\} = W(p_i) \cap \{y : yRx\}$ ) then  $x \in V(\alpha)$  iff  $x \in W(\alpha)$ . Similarly, one defines *purely present* and *purely future*, and one defines *pure* to mean purely past, or present, or future. Note that  $Hp, H'p, S(p, q)$ , are purely past, their mirror images purely future, and any truth-functional compound of variables purely present.  $\mathcal{O}$  has the *separation property* over  $\mathcal{K}$  if for every  $\mathcal{O}$ -formula  $\alpha$  there exists a truth-functional compound  $\beta$  of  $\mathcal{O}$ -formulas pure over  $\mathcal{K}$  such that  $\alpha \leftrightarrow \beta$  is valid over  $\mathcal{K}$ .  $\mathcal{O}$  is *strong* over  $\mathcal{K}$  if  $G, H$  are  $\mathcal{O}$ -definable over  $\mathcal{K}$ . Gabbay [1981a] proves:

**Criterion 20.** *Over any given class  $\mathcal{K}$  of total orders, if  $\mathcal{O}$  is strong and has the separation property, then it is temporally complete.*

A full proof being beyond the scope of this survey (see, however, the next chapter ‘Advanced Tense Logic’), we offer a sketch: we wish to find for any first-order formula  $\varphi(x, <, P_1, \dots, P_n)$  an  $\mathcal{O}$ -formula  $\alpha(p_1, \dots, p_n)$  *representing* it in the sense that for any  $(X, R) \in \mathcal{K}$  and any valuation  $V$  and any  $a \in X$  we have:

$$a \in V(\alpha) \text{ iff } (X, R, V(p_1), \dots, V(p_n)) \models \varphi(a/x).$$

The proof proceeds by induction on the depth of nesting of quantifiers in  $\varphi$ , the key step being  $\varphi(x) = \exists y \psi(x, y)$ . In this case, the atomic subformulas of  $\psi$  are of the forms  $P_i(x), P_i(z), z < x, z = x, x < z, z = w, z < w$ , where  $z$  and  $w$  are variables other than  $x$ . Actually, we may assume there are no subformulas of the form  $P_i(x)$  since these can be brought outside the quantifier  $\exists y$ . We introduce new singulary predicates  $Q^-, Q^0, Q^+$  and replace the subformulas of  $\psi$  of forms  $z < x, z = x, x < z$  by  $Q^-(z), Q^0(z), Q^+(z)$ , to obtain a formula  $\vartheta(y, <, P_1, \dots, P_n, Q^-, Q^0, Q^+)$  to which we can apply our induction hypothesis, obtaining an  $\mathcal{O}$ -formula  $\delta(p_1, \dots, p_n, q^-, q^0, q^+)$  representing it. Let  $\gamma(p_1, \dots, sp_n) = \delta(p-1, \dots, p_n, Fq, q, Pq)$ , and  $\beta = P\gamma \vee \gamma \vee F\gamma$ . It is readily verified that for any  $(X, R) \in \mathcal{K}$  and any  $a, b \in X$  and any valuation  $V$  with  $V(q) = \{a\}$  that we have:

$$\begin{aligned} b \in V(\gamma) & \text{ iff } (X, R, V(p_1), \dots, V(p_n)) \models \psi(a/x, b/y), \\ a \in V(\beta) & \text{ iff } (X, R, V(p_1), \dots, V(p_n)) \models \varphi(a/x). \end{aligned}$$

By hypothesis,  $\beta$  is equivalent over  $\mathcal{K}$  to a truth-functional compound of purely past formulas  $\beta_i^-$ , purely present ones  $\beta_j^0$ , and purely future ones

$\beta_k^+$ . In each  $\beta_i^-$  (resp.  $\beta_j^0$ ) (resp.  $\beta_k^+$ ) replace  $q$  by  $\perp$  (resp.  $\top$ ) (resp.  $\perp$ ) to obtain an  $\mathcal{O}$ -formula  $\alpha$ . It is readily verified that  $\alpha$  represents  $\varphi$ .

It ‘only’ remains to show:

LEMMA 21. *The set  $\{U, S\}$  has the separation property over complete orders.*

Though a full proof is beyond the scope of this survey, we sketch the method for achieving the separation for a formula  $\alpha$  in which there is a single occurrence of an  $S$  within the scope of a  $U$ . This case (and its mirror image) is the first and most important in a general inductive proof.

To begin with, using conjunctive and disjunctive normal forms and such easy equivalences as:

$$\begin{aligned} U(p \vee q, t) &\leftrightarrow U(p, t) \vee U(q, t), \\ U(p, q \wedge r) &\leftrightarrow U(p, q) \wedge U(p, r), \\ \neg S(q, r) &\leftrightarrow S(\neg r, \neg q) \vee P'\neg r, \end{aligned}$$

we can achieve a reduction to the case where  $\alpha$  has one of the forms:

1.  $U(p \wedge S(q, r), t)$
2.  $U(p, q \wedge S(r, t))$

For (1), an equivalent which is a truth-functional compound of pure formulas is provided by :

$$1'. \quad [(S(q, r) \vee q) \wedge U(p, r \wedge t)] \vee U(q \wedge U(p, r \wedge t), t)$$

For (2) we have:

$$2'. \quad \{[(S(r, t) \wedge t) \vee r] \wedge [U(p, t) \vee U(\beta, t)]\} \vee \beta$$

where  $\beta$  is:  $F'\neg t \wedge U(p, q \vee S(r, t))$ . This, despite its complexity, is purely future. The observant reader should be able to see how completeness is needed for the equivalence of (2) and  $w'$ .

Unfortunately,  $U$  and  $S$  take us no further, for Kamp proves:

PROPOSITION 22. *The set  $\{U, S\}$  is not temporally complete over  $(\mathbb{Q}, <)$ .*

Without entering into details, we note that one undefinable operator is  $O(\varphi)$  where  $\varphi$  says:

$$\begin{aligned} \exists y(x < y \wedge \forall z(x < z \wedge z < y \rightarrow \\ (\forall w(x < w \wedge w < z \rightarrow P_1((w))) \vee \forall w(z < w \wedge w < y \rightarrow P_2(w)))) \end{aligned}$$

Over complete orders  $O(\varphi)(p, q)$  amounts to  $U(G'q \wedge (p \vee q), p)$ .

J. Stavi has found two new operators  $U', S'$  and proved:

THEOREM 23. *The set  $\{U, S, U', S'\}$  is temporally complete over total orders.*

Gabbay has greatly simplified the proof: the idea is to try to prove the separation property over arbitrary total orders, and see what operators one needs. One quickly hits on the right  $U', S'$ . The combinatorial details cannot detain us here.

What about axiomatisability for  $U, S$ -tense logic? Some years ago Kamp announced (but never published) finite axiomatisability for various classes of total orders. Some are treated in [Burgess, 1982], where the system for dense orders takes a particularly simple form: we depart from standard format only to the extent of taking  $U, S$  as our primitives. As characteristic axioms, it suffices to take the following and their mirror images:

$$\begin{aligned} G(p \rightarrow q) &\rightarrow (U(p, r) \rightarrow U(q, r)) \wedge ((U(r, p) \rightarrow U(r, q)) \\ p \wedge U(q, r) &\rightarrow U(q \wedge S(p, r), r), \\ U(p, q) &\leftrightarrow U(p, q \wedge U(p, q)) \leftrightarrow U(q \wedge U(p, q), q), \\ U(\cdot, q) \wedge \neg U(p, r) &\rightarrow U(q \wedge \neg r, q), \\ U(p, q) \wedge U(r, s) &\rightarrow U(p \wedge r, q \wedge s \vee U(p \wedge s, q \wedge s) \vee U(q \wedge r, q \wedge s). \end{aligned}$$

A particularly important axiomatisability result is in [Gabbay *et al.*, 1980].

What about decidability? Rabin's theorem applies in most cases, the notable exceptions being complete orders, continuous orders, and  $(\mathbb{R}, <)$ . Here techniques of monadic second-order logic are useful. Decidability for the cases of complete and continuous orders is established in [Gurevich, 1977, Appendix]; and for  $(\mathbb{R}, <)$  in [Burgess and Gurevich, 1985]. A fact (due to Gurevich) from the latter paper worth emphasising is that the  $U, S$ -tense logics of  $(\mathbb{R}, <)$  and of arbitrary continuous orders are *not* the same.

## 5.2 *Now, Then*

We have seen that simple  $G, H$ -tense logic is inadequate to express certain temporal operators expressible in English. Indeed it turns out to be inadequate to express even the shortest item in the English temporal vocabulary, the word 'now'. Just what role this word plays is unclear—some incautious writers have even claimed it is semantically redundant—but [Kamp, 1971] gives a thorough analysis. Let us consider some examples:

0. The seismologist predicted that there would be an earthquake.
1. The seismologist predicted that there would be an earthquake *now*.
2. The seismologist predicted that there would already have been an earthquake *before now*.
3. The seismologist predicted that there would be an earthquake, but not till *after now*.



As Kamp says:

The function of the word ‘now’ in (1) is to make the clause to which it applies—i.e. ‘there would be an earthquake’—refer to the moment of utterance of (1) and not to the moment of moments (indicated by other temporal modifiers that occur in the sentence) to which the clause would refer (as it does in (0)) if the word ‘now’ were absent.

### 5.3 Formal Semantics

To formalise this observation, we introduce a new one-place connective  $J$  (for *jetzt*). We define a *pointed frame* to be a frame with a designated element. A *valuation* in a pointed frame  $(X, R, x_0)$  is just a valuation in  $(X < R)$ . We extend the definition of 0.4 above to  $G, H, J$ -formulas by adding the clause:

$$V(J\alpha) = X \text{ if } x_0 \in V(\alpha), \emptyset \text{ if } x_0 \notin V(\alpha)$$

is *valid* in  $(X, R, x_0)$  if  $x_0 \in V(\alpha)$  for all valuations  $V$ .

An alternative approach is to define a *2-valuation* in a frame  $(X, R)$  to be a function assigning each  $p_i$  a subset of the Cartesian product  $X^2$ . Parallel to 1.4 above we have the following inductive definition:

$$\begin{aligned} V(\neg\alpha) &= X^2 - V(\alpha), \\ V(\alpha \wedge \beta) &= V(\alpha) \cap V(\beta), \\ V(G\alpha) &= \{(x, y) : \forall x'(xRx' \rightarrow (x', y) \in V(\alpha))\}, \\ V(H\alpha), & \text{ similarly,} \\ V(J\alpha) &= \{(x, y) : (y, y) \in V(\alpha)\} \end{aligned}$$

$\alpha$  is *valid* in  $(X, R)$  if  $\{(y, y) : y \in X\} \subseteq V(\alpha)$  for all 2-valuations  $V$ .

The two alternatives are related as follows: Given a 2-valuation  $V$  in the frame  $(X, R)$ , for each  $y \in X$  consider the valuation  $V_y$  in the pointed frame  $(X < R, y)$  given by  $V_y(p_i) = \{x : (x, y) \in V(p_i)\}$ . Then we always have  $(y, y) \in V(\alpha)$  iff  $y \in V_y(\alpha)$ .

The second approach has the virtue of making it clear that though  $J$  is not a temporal operator in the sense of the preceding section, it is in a sense that can be made precise a *two-dimensional* tense operator. This suggests the project of investigating two- and multi-dimensional operators generally. Some such operators, for instance the ‘then’ of [Vlach, 1973], have a natural reading in English. Among other items in our bibliography, [Gabbay, 1976] and [Gabbay and Guenther, 1982] contain much information on this topic.

Using  $J$  we can express (0)–(3) as follows:

- 0'.  $P$  (seismologist says:  $F$  (earthquake occurs)),
- 1'.  $P$  (seismologist says:  $J$  (earthquake occurs)),

2'.  $P$  (seismologist says:  $JP$  (earthquake occurs)),

3'.  $P$  (seismologist says:  $JF$  (earthquake occurs)).

The observant reader will have noted that (0')–(3') are not really representable by  $G, H, J$ -formulas since they involve the notion of 'saying' or 'predicting'), a *propositional attitude*. Gabbay, too, gives many examples of uses of 'now' and related operators, and on inspection these, too, turn out to involve propositional attitudes. That this is no accident is shown by the following result of Kamp:

**THEOREM 24** (Eliminability theorem). *For any  $G, H, J$ -formula  $\alpha$  there is a  $G, H$ -formula  $\alpha^*$  equivalent over all pointed frames.*

**Proof.** Call a formula *reduced* if it contains no occurrence of a  $J$  within the scope of a  $G$  or an  $H$ . Our first step is to find for each formula  $\alpha$  an equivalent reduced formula  $\alpha_R$ . This is done by induction on the complexity of  $\alpha$ , only the cases  $\alpha = G\beta$  or  $\alpha = H\beta$  being nontrivial. In, for instance, the latter case, we use the fact that any truth-function can be put into disjunctive normal form, plus the following valid equivalence:

$$(R) \quad H((Jp \wedge q) \wedge r) \leftrightarrow ((Jp \wedge H(q \vee r)) \vee (\neg Jp \wedge Hr))$$

Details are left to the reader. Our second step is to observe that if  $\beta$  is reduced, then it is equivalent to the result  $\beta^-$  of dropping all its occurrences of  $J$ . It thus suffices to set  $\alpha^* = (\alpha_R)^-$ . ■

The foregoing theorem says that in the presence of truth-functions and  $G$  and  $H$ , the operator  $J$  is, in a sense, redundant. By contrast, examples (0)–(3) suggest that in contexts with propositional attitudes,  $J$  is *not* redundant; the lack of a generally-accepted formalisation of the logic of propositional attitudes makes it impossible to turn this suggestion into a rigorous theorem. But in contexts with *quantifiers*, Kamp *does* prove rigorously that  $J$  is irredundant. Consider:

4. The Academy of Arts rejected an applicant who was to become a terrible dictator and start a great war.

5. The Academy of arts has rejected an applicant who is to become a terrible dictator and start a great war.

The following formalisations suggest themselves:

4'.  $P(\exists x(R(x) \wedge FD(x)))$

5'.  $P(\exists x(R(x) \wedge JFD(x)))$ ,

the difference between (4) and (5) lying precisely in the fact that the latter, unlike the former, definitely places the dictatorship and war in the hearer's future. What Kamp proves is that (5') cannot be expressed by a  $G, H$ -formula with quantifiers.

Returning to sentential tense logic, Theorem 24 obviously reduces the *decision* problem for  $G, H, J$ -tense logic to that for  $G, H$ -tense logic. As for *axiomatisability*, obviously we cannot adopt the standard format of  $G, H$ -tense logic, since the rule TG does not preserve validity for  $G, H, J$ -formulas. For instance:

$$(D0) \quad p \leftrightarrow Jp$$

is valid, but  $G(p \leftrightarrow Jp)$  and  $H(p \leftrightarrow Jp)$  are not. Kamp overcomes this difficulty, and shows how, in very general contexts, to obtain from a complete axiomatisation of a logic without  $J$ , a complete axiomatisation of the same logic with  $J$ . For the sentential  $G, H, J$ -tense logic of total orders, the axiomatisation takes a particularly simple form: take as sole rule MP. Let  $Lp$  abbreviate  $Hp \wedge p \wedge Gp$ . Take as axioms all substitution instances of tautologies, of (D0) above, and of  $L\alpha$ , where  $\alpha$  may be any item on the lists (D1), (D2) below, or the mirror image of such an item:

$$(D1) \quad \begin{aligned} G(p \rightarrow q) &\rightarrow (Gp \rightarrow Gq) \\ p &\rightarrow GPp \\ Gp &\leftrightarrow GGp \\ Lp &\leftrightarrow GHp \end{aligned}$$

$$(D2) \quad \begin{aligned} J\neg p &\leftrightarrow \neg Jp \\ J(p \wedge q) &\leftrightarrow Jp \wedge Jq \\ \neg L\neg Jp &\leftrightarrow LJp \\ Lp &\rightarrow Jp. \end{aligned}$$

(In outline, the proof of completeness runs thus: using (D1) one deduces  $Lp \rightarrow LLp$ . It follows that the class of theses deducible without use of (D0) is closed under TG. Our work in Section 3.2 shows that we then get the complete  $G, H$ -tense logic of total orders. We then use (D2) to prove the equivalence (R) in the proof of Theorem 24 above. More generally, for any  $\alpha$ ,  $\alpha \leftrightarrow \alpha_R$  is deducible without using (D0). Moreover, using D0,  $\beta \leftrightarrow \beta^-$  is deducible for any reduced formula  $\beta$ . Thus in general  $\alpha \leftrightarrow \alpha^*$  is a thesis, completing the proof.)

## 6 TIME PERIODS

The geometry of Space can be axiomatised taking unextended *points* as basic entities, but it can equally well be axiomatised by taking as basic certain regular open solid regions such as *spheres*. Likewise, the order of

Time can be described either (as in Section 1.1) in terms of *instants* in terms of *periods* of non zero duration. Recently it has become fashionable to try to redo tense logic, taking periods rather than instants as basic. Humberstone [1979] seems to be the first to have come out in print with such a proposal. This approach has become so popular that we must give at least a brief account of it; further discussion can be found in [van Benthem, 1991]. (See also Kuhn's discussion in the last chapter of this Volume of the *Handbook*.)

In part, the switch from instants to periods is motivated by a desire to model certain features of natural language. One of these is *aspect*, the verbal feature which indicates whether we are thinking of an occurrence as an *event* whose temporal stages (if any) do not concern us, or as a protracted *process*, forming, perhaps the backdrop for other occurrences. These two ways of looking at death (a popular, if morbid, example) are illustrated by:

When Queen Anne died, the Whigs brought in George.

While Queen Anne was dying, the Jacobites hatched treasonable plots.

Another feature of linguistic interest is the peculiar nature of *accomplishment* verbs, illustrated by:

1. The Amalgamated Conglomerate Building was built during the period March–August 1972.
- 1'. The ACB was built during the period April–July, 1972.
2. The ACB was being built (i.e. was under construction) during the period March–August, 1972.
- 2'. The ACB was under construction during the period April– July, 1972.

Note that (1) and (1') are inconsistent, whereas (2) implies (2')!

In part, the switch is motivated by a philosophical belief that periods are somehow more basic than instants. This motivation would be more convincing were 'periods' not assumed (as they are in too many recent works) to have sharply-defined (i.e. *instantaneous*) beginnings and ends. It may also be remarked that at the level of *experience* some occurrences do *appear* to be instantaneous (i.e. we don't *discern* stages in them). Thus 'bubbles when they burst' *seem* to do so 'all at once and nothing first'. While at the level of *reality*, some occurrences of the sort studied in quantum physics may well take place instantaneously, just as some elementary particles may well be pointlike. Thus the philosophical belief that every occurrence takes some time (period) to occur is not *obviously* true on any level.

Now for the mechanics of the switch: for any frame  $(X, R)$  we consider the set  $I(X, R)$  of nonempty bounded open intervals of form  $\{z : xRz \wedge zRy\}$ .

Among the many relations on this set that could be defined in terms of  $R$  we single out two:

$$\begin{aligned} \text{Inclusion: } a \subseteq b & \text{ iff } \forall x(x \in a \rightarrow x \in b), \\ \text{Order: } a \triangleleft b & \text{ iff } \forall x \forall y(x \in a \wedge y \in b \rightarrow xRy). \end{aligned}$$

To any class  $\mathcal{K}$  of frames we associate the class  $\mathcal{K}'$  of those structures of form  $(I(X, R), \subseteq, \triangleleft)$  with  $(X, R) \in \mathcal{K}$ , and the class  $\mathcal{K}^+$  of those structures  $(Y, S, T)$  that are isomorphic to elements of  $\mathcal{K}'$ .

A first problem in switching from instants to periods as the basis for the logic of time is to find each important class  $\mathcal{K}$  of frames a set of postulates whose models will be precisely the structures in  $\mathcal{K}^+$ . For the case of dense total orders without extrema, and for some other cases, suitable postulate sets are known, though none is very elegant. Of course this first problem is not yet a problem of *tense* logic; it belongs rather to applied first- and second-order logic.

To develop a period-based *tense* logic we define a *valuation* in a structure  $(Y, S, T)$ —where  $S, T$  are binary relations on  $Y$ —to be a function  $V$  assigning each  $p_i$  a subset of  $Y$ . Then from among all possible connectives that could be defined in terms of  $S$  and  $T$ , we single out the following:

$$\begin{aligned} V(\neg\alpha) &= Y - V(\alpha) \\ V(\alpha \wedge \beta) &= V(\alpha) \cap V(\beta) \\ V(\nabla\alpha) &= \{a : \forall b(bSa \rightarrow b \in V(\alpha))\} \\ V(\Delta\alpha) &= \{a : \forall b(aSb \rightarrow b \in V(\alpha))\} \\ V(F\alpha) &= \{a : \exists b(aTb \wedge b \in V(\alpha))\} \\ V(P\alpha) &= \{a : \exists b(bTa \wedge b \in V(\alpha))\}. \end{aligned}$$

The main *technical* problem now is, given a class  $\mathbf{L}$  of structures  $(Y, S, T)$ —for instance, one of form  $\mathbf{L} = \mathcal{K}^+$  for some class  $\mathcal{K}$  of frames—to find a sound and complete axiomatisation for the tense logic of  $\mathbf{L}$  based on the above connectives. Some results along these lines have been obtained, but none as definitive as those of instant-based tense logic reported in Section 3. Indeed, the choice of relations ( $\subseteq$  and  $\triangleleft$ ), and of admissible classes  $\mathbf{L}$  (should we only consider classes of form  $\mathcal{K}^+$ ?), and of connectives ( $\neg, \wedge, \Delta, \nabla, F, P$ ), and of admissible valuations (should we impose restrictions, such as requiring  $b \in V(p_i)$  whenever  $a \in V(p_i)$  and  $b \subseteq a$ ?) are all matters of controversy.

The main problem of *interpretation*—one to which advocates of period-based tense logic have perhaps not devoted sufficient attention—is how to make intuitive sense of the notion  $a \in V(p)$  of a sentence  $p$  being true *with respect to* a time-period  $a$ . One proposal is to take this as meaning that  $p$  is true *throughout*  $a$ . Now given a valuation  $W$  in a frame  $(X, R)$ , we can define a valuation  $I(W)$  in  $I(X, R)$  by  $I(W)(p_i) = \{a : a \subseteq W(p_i)\}$ . When and *only* when  $V$  has the form  $I(W)$  is ‘ $p$  is true throughout  $a$ ’ a tenable reading of  $a \in V(p)$ . It is not, however, easy to characterise intrinsically

those  $V$  that admit a representation in the form  $V = I(W)$ . Note that even in this case,  $a \in V(\neg p)$  does not express ‘ $(\neg p)$  is true throughout  $a$ ’ (but rather ‘ $\neg(p)$  is true throughout  $a$ ’). Nor does  $a \in V(p \vee q)$  express ‘ $(p \vee q)$  is true throughout  $a$ ’.

Another proposal, originating in [Burgess, 1982] is to read  $a \in V(p)$  as ‘ $+$  is *almost* always true during  $a$ ’. This reading is tenable when  $V$  has the form  $J(W)$  for some valuation  $W$  in  $(X, R)$ , where  $J(W)(p_i)$  is by definition  $\{a : a - W(p_i) \text{ is nowhere dense in the order topology on } (XR)\}$ . In this case, ‘ $(\neg p)$  is almost always true during  $a$ ’ is expressible by  $a \in V(\nabla \neg p)$ , and ‘ $(p \vee q)$  is almost always true during  $a$ ; by  $a \in V(\nabla \neg \nabla \neg (p \vee q))$ . But the whole problem of interpretation for period-based tense logic deserves more careful thought.

There have been several proposals to redo tense logic on the basis of 3- or 4- of multi-valued truth-functional logic. It is tempting, of instance, to introduce a truth-value ‘unstatable’ to apply to, say, ‘Bertrand Russell is smiling’ in 1789. In connection with the switch from instants to periods, some have proposed introducing new truth-values ‘changing from true to false’ and ‘changing from false to true’ to apply to, say, ‘the rocket is at rest’ at take-off and landing times. Such proposals, along with proposals to combine, say, tense logic and intuitionistic logic, lie beyond the scope of this survey.

## 7 GLIMPSES AROUND

### 7.1 Metric Tense Logic

In *metric* tense logic we assume Time has the structure of an ordered Abelian group. We introduce variables  $x, y, z, \dots$  ranging over group elements, and simples  $0, +, <$  for the group identity, addition, and order. We introduce operators  $\mathcal{F}, \mathcal{P}$  joining terms for group elements with formulas. Here, for instance,  $\mathcal{F}(x + y)(p \wedge q)$  means that it will be the case  $(x + y)$  time-units hence that  $p$  and  $q$ . Metric tense logic is intended to reflect such ordinary-language *quantitative* expressions as ‘10 years from now’ or ‘tomorrow about this time’ or ‘in less than five minutes’. The *qualitative*  $F, P$  of nonmetric tense logic can be recovered by the definitions  $Fp \leftrightarrow \exists x > 0 \mathcal{F}xp, Pp \leftrightarrow \exists x > 0 \mathcal{P}xp$ . Actually, the ‘ago’ operator  $\mathcal{P}$  is definable in terms of the ‘hence’ operator  $\mathcal{F}$  since  $\mathcal{P}xp$  is equivalent to  $\mathcal{F} - xp$ . It is not hard to write down axioms for metric tense logic whose completeness can be proved by a Henkin-style argument.

But decidability is lost: the decision problem for metric tense logic is easily seen to be equivalent to that for the set of all universal monadic (second- order) formulas true in all ordered Abelian groups. We will show that the decision problem for the validity of first-order formulas involving

a single two-place predicate  $\in$ —which is well known to be unsolvable—is reducible to the latter: given a first-order  $\in$ -formula  $\varphi$ , fix two one-place predicate variables  $U, V$ . Let  $\varphi_0$  be the result of restricting all quantifiers in  $\varphi$  to  $U$  (i.e.  $\forall x$  is replaced  $\forall x(U(x) \rightarrow \dots)$  and  $\exists x$  by  $\exists x(U(x) \wedge \dots)$ .) Let  $\varphi_1$  be the result of replacing each atomic subformula  $x \in y$  of  $\varphi_0$  by  $\exists z(V(z) \wedge V(z+x) \wedge V(z+x+y))$ . Let  $\varphi_2$  be the universal monadic formula  $\forall U \forall V (\exists x U(x) \rightarrow \varphi_1)$ . Clearly if  $\varphi$  is logically valid, then so is  $\varphi_2$  and, in particular, the latter is true in all ordered Abelian groups. If  $\varphi$  is not logically valid, it has a countermodel consisting of the positive integers equipped with a binary relation  $E$ . Consider the product  $\mathbb{Z} \times \mathbb{Z}$  where  $\mathbb{Z}$  is the additive group of integers; addition in this group is defined by  $(x, y) + (x', y') = (x+x', y+y')$ ; the group is orderable by  $(x, y) < (x', y')$  iff  $x < x'$  or  $(x = x'$  and  $y < y')$ . Interpret  $U$  in this group as  $\{(n, 0) : n > 0\}$ ; interpret  $V$  as the set consisting of the  $(2^m 3^n, 0)$ ,  $(2^m 3^n, m)$  and  $(2^m 3^n, m+n)$  for those pairs  $(m, n)$  with  $mEn$ . This gives a countermodel to the truth of  $\varphi_2$  in  $\mathbb{Z} \times \mathbb{Z}$ . Thus the desired reduction of decision problems has been effected.

Metric tense logic is, in a sense, a hybrid between the ‘regimentation’ and ‘autonomous tense logic’ approaches to the logic of time. Other hybrids of a different sort—not easy to describe briefly—are treated in an interesting paper of [Bull, 1978].

## 7.2 Time and Modality

As mentioned in the introduction, Prior attempted to apply tense logic to the exegesis of the writings of ancient and mediaeval philosophers and logicians (and for that matter of modern ones such as C. S. Peirce and J. Łukasiewicz) on future contingents. The relations between tense and mode or modality is properly the topic of Richmond H. Thomason’s chapter in this volume.

We can, however, briefly consider here the topic of so-called *Diodorean* and *Aristotelian* modal fragments of a tense logic  $L$ . The former is the set of modal formulas that become theses of  $L$  when  $\Box p$  is defined as  $p \wedge Gp$ ; the latter is the set of modal formulas that becomes theses of  $L$  when  $\Box p$  is defined as  $Hp \wedge p \wedge Gp$ . Though these seem far-fetched definitions of ‘necessity’, the attempt to isolate the modal fragments of various tense logics undeniable was an important stimulus for the earlier development of our subject. Briefly the results obtained can be tabulated as follows. It will be seen that the modal fragments are usually well-known C. I. Lewis systems.

Class of frames	Tense logic	Diodorean fragment	Aristotelian fragment
All frames	$\mathbf{L}_0$	$\mathbf{T}(=\mathbf{M})$	$\mathbf{B}$
Partial orders	$\mathbf{L}_1$	$\mathbf{S4}$	$\mathbf{B}$
Lattices	$\mathbf{L}_0$	$\mathbf{S4.2}$	$\mathbf{B}$
Total orders	$\mathbf{L}_2, \mathbf{L}_5$	$\mathbf{S4.3}$	$\mathbf{S5}$
Dense orders			

The Diodorean fragment of the tense logic  $\mathbf{L}_6$  of discrete orders has been determined by M. Dummett; the Aristotelian fragment of the tense logic of trees has been determined by G. Kessler. See also our comments below on R. Goldblatt's work.

### 7.3 *Relativistic Tense Logic*

The *cosmic* frame is the set of all point-events of space-time equipped with the relation of *causal accessibility*, which holds between  $u$  and  $v$  if a signal (material or electromagnetic) could be sent from  $u$  to  $v$ . The  $(n+1)$ -dimensional *Minkowski* frame is the set of  $(n+1)$ -tuples of real numbers equipped with the relation which holds between  $(a_0, a_1, \dots, a_n)$  and  $(b_0, b_1, \dots, b_n)$  iff:

$$\sum_{i=1}^n (b_i - a_i)^2 - (b_0 - a_0)^2 > 0 \text{ and } b_0 > a_0.$$

For present purposes, the content of the *special theory of relativity* is that the cosmic frame is isomorphic to the 4-dimensional Minkowski frame.

A little calculating shows that any Minkowski frame is a lattice without maximum or minimum, hence the tense logic of special relativity should at least include  $\mathbf{L}_0$ . Actually we will also want some axioms to express the density and continuity of a Minkowski frame. A surprising discovery of Goldblatt [1980] is that the *dimension* of a Minkowski frame influences its tense logic. Indeed, he shows that for each  $n$  there is a formula  $\gamma_{n+1}$  which is valid in the  $(m+1)$ -dimensional Minkowski frame iff  $m < n$ . For example, writing  $Ep$  for  $p \wedge Fp$ ,  $\gamma_2$  is:

$$Ep \wedge Eq \wedge Er \wedge \neg E(p \wedge q) \wedge \neg E(p \wedge r) \wedge \neg E(q \wedge r) \rightarrow E((Ep \wedge Eq) \vee (Ep \wedge Er) \vee (Eq \wedge Er)).$$

On the other hand, he also shows that the dimension of a Minkowski frame does *not* influence the diodorean modal fragment of its tense logic: the Diodorean modal logic of special relativity is the same as that of arbitrary lattices, namely  $\mathbf{S4.2}$ . Combining Goldblatt's argument with the 'trousers world' construction in general relativity, should produce a proof that the Diodorean modal fragment of the latter is the same as that of arbitrary partial orders, namely  $\mathbf{S4}$ .



Despite recent advances, the tense logic of special relativity has not yet been completely worked out; that of general relativity is even less well understood. Burgess [1979] contains a few additional philosophical remarks.

#### 7.4 *Thermodynamic Time*

One of the oldest metaphysical concepts (found in Hindu theology and pre-Socratic philosophy, and in modern psychological dress in Nietzsche and celestial mechanical dress in Poincaré) is that everything that has ever happened is destined to be repeated over and over again. This leads to a degenerate tense logic containing the principles  $Gp \rightarrow Hp$  and  $Gp \rightarrow p$  among others.

An antithetical view is that traditionally associated with the Second Law of Thermodynamics, according to which irreversible change are taking place that will eventually drive the Universe to a state of ‘heat-death’, after which no further change on a macroscopically observable level will take place. The tense logic of this view, which raises several interesting technical points, has been investigated by S. K. Thomason [1972]. The first thing to note is that the principle:

$$(A10) \quad GFp \rightarrow FGp$$

is acceptable for  $p$  expressing propositions about macroscopically observable states of affairs provided these do not contain hidden time references; e.g.  $p$  could be ‘there is now no life on Earth’, but not ‘particle  $\kappa$  currently has a momentum of precisely  $k$  gram- meters/second’ or ‘it is now an even number of days since the Heat Death occurred’. For the antecedent of (A20) says that arbitrarily far in the future there will be times when  $p$  is the case. But for the  $p$  that concern us, the truth-value of  $p$  is never supposed to change after the Heat Death. So in that case, there will come a time after which  $p$  is always going to be true, in accordance with the consequent of (A10).

The question now arises, how can we *formalise* the restriction of  $p$  to a special class of sentences? In general, propositions are represented in the formal semantics of tense logic by subsets of  $X$  in a frame  $(X, R)$ . A restricted class of propositions could thus be represented by a distinguished family  $\mathcal{B}$  of subsets of  $X$ . This motivates the following definition: an *augmented* frame is a triple  $(X, R, \mathcal{B})$  where  $(X, R)$  is a frame,  $\mathcal{B}$  a subset of the lower set  $\mathcal{B}(X)$  of  $X$  closed under complementation, finite intersection, and the operations:

$$\begin{aligned} gA &= \{x \in X : \forall y \in X (xRy \rightarrow y \in A)\} \\ hA &= \{x \in X : \forall y \in X (yRx \rightarrow y \in A)\}. \end{aligned}$$

A *valuation* in  $(X, R, \mathcal{B})$  is a function  $V$  assigning each variable  $p_i$  an element of  $\mathcal{B}$ . The closure conditions on  $\mathcal{B}$  guarantee that we will then have  $V(\alpha) \in \mathcal{B}$

for *all* formulas  $\alpha$ . It is now clear how to define validity. Note that if  $\mathcal{B} = \mathcal{P}(X)$ , then the validity in  $(X, R, \mathcal{B})$  reduces to validity in  $(X, R)$ ; otherwise more formulas may be valid in the former than the latter.

It turns out that the extension  $\mathbf{L}_{10}$  of  $\mathbf{L}_{\mathbf{Q}}$  obtained by adding (A10) is (sound and) complete for the class of augmented frames  $(X, R, \mathcal{B})$  in which  $(X, R)$  is a dense total order without maximum or minimum and:

$$\forall B \in \mathcal{B} \exists x (\forall y (xRy \rightarrow y \in B) \vee \forall y (xRy \rightarrow y \notin B)).$$

We have given complete axiomatisations for many intuitively important classes of frames. We have not yet broached the questions: when does the tense logic of a given class of frames admit a complete axiomatisation? When does a given axiomatic system of tense logic correspond to some class of frames in the sense of being complete for that class? For information on these large questions, and for bibliographical references, we refer the reader to Johan van Benthem's chapter in Volume 3 of this edition of the *Handbook* on so-called 'Correspondence Theory'. Suffice it to say here that positive general theorems are few, counterexamples many. The thermodynamic tense logic  $\mathbf{L}_{10}$  exemplifies one sort of pathology. Though it is not inconsistent, there is *no* (unaugmented) frame in which all its theses are valid!

### 7.5 Quantified Tense Logic

The interaction of temporal operators with universal and existential quantifiers raises many difficult issues, both philosophical (over identity through changes, continuity, motion and change, reference to what no longer exists or does not exist, essence, and many, many more) and technical (over undecidability, nonaxiomatisability, undefinability or multi-dimensional operators, and so forth) that it is pointless to attempt even a survey of the subject in a paragraph or two. We therefore refer the reader to Nino Cocchiarella's chapter in this volume and James W. Garson's chapter in Volume 3 of this edition of the *Handbook*, both on this subject.

*Princeton University, USA.*

## BIBLIOGRAPHY

- [Åqvist, 1975] L. Åqvist. Formal semantics for verb tenses as analysed by Reichenbach. In *Pragmatics of Language and Literature*. T. A. van Dijk, ed. pp. 229–236. North Holland, Amsterdam, 1975.
- [Åqvist and Guentner, 1978] L. Åqvist and F. Guentner. Fundamentals of a theory of verb aspect and events within the setting of an improved tense logic. In *Studies in Formal Semantics*. F. Guentner and C. Rohrer, eds. pp. 167–20. North Holland, Amsterdam, 1978.
- [Åqvist and Guentner, 1977] L. Åqvist and F. Guentner, eds. Tense logic (= *Logique et Analyse*, **80**), 1977.

- [Bull, 1968] R. A. Bull. An algebraic study of tense logic with linear time. *Journal of Symbolic Logic*, **33**, 27–38, 1968.
- [Bull, 1978] R. A. Bull. An approach to tense logic. *Theoria*, **36**, 1978.
- [Burgess, 1979] J. P. Burgess. Logic and time. *Journal of Symbolic Logic*, **44**, 566–582, 1979.
- [Burgess, 1982] J. P. Burgess. Axioms for tense logic. *Notre Dame Journal of Formal Logic*, **23**, 367–383, 1982.
- [Burgess and Gurevich, 1985] J. P. Burgess and Y. Gurevich. The decision problem for linear temporal logic. *Notre Dame Journal of Formal Logic*, **26**, 115–128, 1985.
- [Gabbay, 1975] D. M. Gabbay. Model theory for tense logics and decidability results for non-classical logics. *Ann. Math. Logic*, **8**, 185–295, 1975.
- [Gabbay, 1976] D. M. Gabbay. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*, Reidel, Dordrecht, 1976.
- [Gabbay, 1981a] D. M. Gabbay. Expressive functional completeness in tense logic (Preliminary report). In *Aspects of Philosophical Logic*, U. Mönnich, ed. pp.91–117. Reidel, Dordrecht, 1981.
- [Gabbay, 1981b] D. M. Gabbay. An irreflexivity lemma with applications of conditions on tense frames. In *Aspects of Philosophical Logic*, U. Mönnich, ed. pp. 67–89. Reidel, Dordrecht, 1981.
- [Gabbay and Guenther, 1982] D. M. Gabbay and F. Guenther. A note on many-dimensional tense logics. In *Philosophical Essays Dedicated to Lennart Åqvist on his Fiftieth Birthday*, T. Pauli, ed. pp. 63–70. University of Uppsala, 1982.
- [Gabbay et al., 1980] D. M. Gabbay, A. Pnueli, S. Shelah and J. Stavi. On the temporal analysis of fairness. *proc. 7th ACM Symp. Principles Prog. Lang.*, pp. 163–173, 1980.
- [Goldblatt, 1980] R. Goldblatt. Diodorean modality in Minkowski spacetime. *Studia Logica*, **39**, 219–236, 1980.
- [Gurevich, 1977] Y. Gurevich. Expanded theory of ordered Abelian groups. *Ann. Math. Logic*, **12**, 192–228, 1977.
- [Humberstone, 1979] L. Humberstone. Interval semantics for tense logics. *Journal of philosophical Logic*, **8**, 171–196, 1979.
- [Kamp, 1968] J. A. W. Kamp. Tense logic and the theory of linear order. Doctoral Dissertation, UCLA, 1968.
- [Kamp, 1971] J. A. W. Kamp. Formal properties of ‘Now’. *Theoria*, **37**, 27–273, 1971.
- [Lemmon and Scott, 1977] E. J. Lemmon and D. S. Scott. *An Introduction to Modal Logic: the Lemmon Notes*. Blackwell, 1977.
- [McArthur, 1976] R. P. McArthur. *Tense Logic*. Reidel, Dordrecht, 1976.
- [Normore, 1982] C. Normore. Future contingents. In *The Cambridge History of Later Medieval Philosophy*. A. Kenny et al., eds. University Press, Cambridge, 1982.
- [Pratt, 1980] V. R. Pratt. Applications of modal logic to programming. *Studia Logica*, **39**, 257–274, 1980.
- [Prior, 1957] A. N. Prior. *Time and Modality*, Clarendon Press, Oxford, 1957.
- [Prior, 1967] A. N. Prior. *Past, Present and Future*, Clarendon Press, Oxford, 1967.
- [Prior, 1968] A. N. Prior. *Papers on Time and Tense*, Clarendon Press, Oxford, 1968.
- [Quine, 1960] W. V. O. Quine. *Word and Object*. MIT Press, Cambridge, MA, 1960.
- [Rabin, 1966] M. O. Rabin. Decidability of second order theories and automata on infinite trees. *Trans Amer Math Soc.*, **141**, 1–35, 1966.
- [Rescher and Urquhart, 1971] N. Rescher and A. Urquhart. *Temporal Logic*, Springer, Berlin, 1971.
- [Rohrer, 1980] Ch. Rohrer, ed. *Time, Tense and Quantifiers*. Max Niemeyer, Tübingen, 1980.
- [Seegerberg, 1970] K. Seegerberg. Modal logics with linear alternative relations. *Theoria*, **36**, 301–322, 1970.
- [Seegerberg, 1980] K. Seegerberg, ed. Trends in modal logic. *Studia Logica*, **39**, No. 3, 1980.
- [Shelah, 1975] S. Shelah. The monadic theory of order. *Ann Math*, **102**, 379–419, 1975.
- [Thomason, 1972] S. K. Thomason. Semantic analysis of tense logic. *Journal of Symbolic Logic*, **37**, 150–158, 1972.
- [van Benthem, 1978] J. F. A. K. van Benthem. Tense logic and standard logic. *Logique et analyse*, **80**, 47–83, 1978.

- [van Benthem, 1981] J. F. A. K. van Benthem. Tense logic, second order logic, and natural language. In *Aspects of Philosophical Logic*, U. Mönnich, ed. pp. 1-20. Reidel, Dordrecht, 1981.
- [van Benthem, 1991] J. F. A. K. van Benthem. *The Logic of Time*, 2nd Edition. Kluwer Academic Publishers, Dordrecht, 1991.
- [Vlach, 1973] F. Vlach. Now and then: a formal study in the logic of tense anaphora. Doctoral dissertation, UCLA, 1973.

## ADVANCED TENSE LOGIC

### 1 INTRODUCTION

In this chapter we consider the tense (or temporal) logic with until and since connectives over general linear time. We will call this logic  $US/LT$ . This logic is an extension of Prior's original temporal logic of  $F$  and  $P$  over linear time [Prior, 1957], via the introduction of the more expressive connectives of Kamp's  $U$  for "until" and  $S$  for "since" [Kamp, 1968b].  $U$  closely mimics the natural language construct "until" with  $U(A, B)$  holding when  $A$  is constantly true from now up until a future time at which  $B$  holds.  $S$  is similar with respect to the past. We will see that  $U$  and  $S$  do indeed extend the expressiveness of the temporal language.

In the chapter we will also be looking at other related temporal logics. The logics differ from each other in two respects. Logics may differ in the kinds of structures which they are used to describe. Structures vary in terms of their underlying model of time (or frame): this can be like the natural numbers, or like the rationals or like the reals or some other linear order or some non-linear branching or multi-dimensional shape. Logics are defined with respect to a class of structures. Considering a logic defined by the class of all linear structures is a good base from which to begin our exploration. Temporal logics also vary in their language. For various purposes, until and since may be not expressive enough. For example, if we want to be able to reason about alternative avenues of development then we may want to allow branches in the flow of time and, in order to represent directly the fact of alternative possibilities, we may need to add appropriate branching connectives. Equally, until and since may be too strong: for simple reasoning about the forward development of a mechanical system, using since may not only be unnecessary, but may require additional axioms and complexity of a decision procedure.

In this chapter we will not be looking at temporal logics based on branching. See the handbook chapter by Thomason for these matters. We will also avoid consideration of temporal logics incorporating quantification. Instead, see the handbook chapter by Garson for a discussion of predicate temporal and modal logics and see the reference [Gabbay *et al.*, 1994] for a discussion of temporal logics incorporating quantification over propositional atoms.

So we will begin with a tour of the many interesting results concerning  $US/LT$  including axiom systems, related logics, decidability and complexity. In section 3 we sketch a proof of the expressive completeness of the logic.

Then, in section 4 we investigate combinations of logics with a temporal element. In section 5, we develop the proof theory for temporal logic within the framework of labelled deductive systems. In section 6, we show how temporal reasoning can be handled within logic programming. In section 7, we survey the much studied temporal logic of the natural numbers and consider the powerful automata technique for reasoning about it. Finally, in section 8, we consider the possibility of treating temporal logic in an imperative way.

## 2 $U, S$ LOGIC OVER GENERAL LINEAR TIME

Here we have a close look at the  $US$  logic over arbitrary linear orders.

### 2.1 *The logic*

Frames for our logic are linear. Thus we have a non-empty set  $T$  and a binary relation  $< \subseteq T \times T$  which is:

1. *irreflexive*, i.e.  $\forall t \in T$ , we do not have  $t < t$ ;
2. *total*, i.e.  $\forall s, t \in T$ , either  $s < t$ ,  $s = t$  or  $t < s$ ;
3. *transitive*, i.e.  $\forall s, t, u \in T$ , if  $s < t$  and  $t < u$  then  $s < u$ .

The underlying model of time for a temporal logic is captured by the frame  $(T, <)$ .

Any use of a temporal logic will involve something happening over time. The simplest method of trying to capture this formally is to use a propositional temporal logic. So we fix a countable set  $\mathcal{L}$  of atoms. The truth of a particular atom will vary in time. For example, points of time (i.e.  $t \in T$ ) may correspond to days and the truth of the atom  $r$  on a particular day may correspond to the event of rain on that day.

A structure is a particular history of the truth of all the atoms over the full extent of time. Structures  $(T, <, h)$  are linear so we have a linear frame  $(T, <)$  and we have a valuation  $h$  for the atoms, i.e. for each atom  $p \in \mathcal{L}$ ,  $h(p) \subseteq T$ . The set  $h(p)$  is the set of all time points at which  $p$  is true.

The language  $L(U, S)$  is generated by the 2-place connectives  $U, S$  along with classical  $\neg$  and  $\wedge$ . That is, we define the set of formulas recursively to contain the atoms and  $\top$  (i.e. truth) and for formulas  $A$  and  $B$  we include  $\neg A$ ,  $A \wedge B$ ,  $U(A, B)$  and  $S(A, B)$ . We read  $U(A, B)$  as “until  $A, B$ ” corresponding to  $B$  being true until  $A$  is. Similarly  $S$  is read as “since”.

Formulas are evaluated at points in structures. We write  $\mathcal{T}, x \models A$  when  $A$  is true at the point  $x \in T$ . This is defined recursively as follows. Suppose that we have defined the truth of formulas  $A$  and  $B$  at all points of  $\mathcal{T}$ . Then for all points  $x$ :

$\mathcal{T}, x \models p$	iff	$x \in h(p)$ , for $p$ atomic;
$\mathcal{T}, x \models \top$ ;		
$\mathcal{T}, x \models \neg A$	iff	$\mathcal{T}, x \not\models A$ ;
$\mathcal{T}, x \models A \wedge B$	iff	both $\mathcal{T}, x \models A$ and $\mathcal{T}, x \models B$ ;
$\mathcal{T}, x \models U(A, B)$	iff	there is a point $y > x$ in $T$ such that $\mathcal{T}, y \models A$ and for all $z \in T$ such that $x < z < y$ we have $\mathcal{T}, z \models B$ ;
$\mathcal{T}, x \models S(A, B)$	iff	there is a point $y < x$ in $T$ such that $\mathcal{T}, y \models A$ and for all $z \in T$ such that $y < z < x$ we have $\mathcal{T}, z \models B$ ;

Often definitions and results involving  $S$  can be given by simply exchanging  $U$  and  $S$  and swapping  $<$  and  $>$ . In that situation we just mention that a *mirror image* case exists and do not go into details.

There are many abbreviations that are commonly used in the language. As well as the classical  $\perp$  (i.e.  $\neg\top$  for falsity),  $\vee$ ,  $\rightarrow$  and  $\leftrightarrow$ , we have the following temporal abbreviations:

$FA$	$= U(A, \top)$	$A$ will happen (sometime);
$GA$	$= \neg F\neg A$	$A$ will always hold;
$PA$	$= S(A, \top)$	$A$ was true (sometime);
$HA$	$= \neg P\neg A$	$A$ was always true;
$K^+(A)$	$= \neg U(\top, \neg A)$	$A$ will be true arbitrarily soon;
$K^-(A)$	$= \neg S(\top, \neg A)$	$A$ was true arbitrarily recently.

Notice that Prior's original connectives  $F$  and  $P$  appear as abbreviations in this logic. The reader should check that their original semantics (see [Burgess, 2001]) are not compromised.

A formula  $\phi$  is *satisfiable* if it has a model: i.e. there is a structure  $\mathcal{T} = (T, <, h)$  and  $x \in T$  such that  $\mathcal{T}, x \models \phi$ . A formula is *valid* iff it is true at all points of all structures. We write  $\models A$  iff  $A$  is a validity. Of course, a formula is valid iff its negation is not satisfiable.

We can also define (semantic) consequence in the logic. Suppose that  $\Gamma$  is a set of formulas and  $A$  a formula. We say that  $A$  is a consequence of  $\Gamma$  and write  $\Gamma \models A$  iff whenever we have  $\mathcal{T}, t \models C$  for all  $C \in \Gamma$ , for some point  $t$  from some structure  $\mathcal{T}$ , then we also have  $\mathcal{T}, t \models A$ .

### *First-Order Monadic Logic of Order*

For many purposes such as assessing the expressiveness of temporal languages or establishing their decidability, it is useful to be able to move from the internal tensed view of the world to an external untensed view. In doing

so we can also make use of logics with more familiar syntax. In the case of our linear temporal logics we find it convenient to move to the first-order monadic logic of linear order which is a sub-logic of the full second-order monadic logic of linear order.

The language of the full second-order monadic logic of linear order has formulas built from  $<$ ,  $=$ , quantification over individual variable symbols and quantification over monadic (i.e. 1-ary) predicate symbols. To be more formal, suppose that  $X = \{x_0, x_1, \dots\}$  is our set of individual variable symbols and  $Q = \{P_0, P_1, \dots\}$  is our set of monadic predicates. The formulas of the language are  $x_i < x_j$ ,  $x_i = x_j$ ,  $P_i(x_j)$ ,  $\neg\alpha$ ,  $\alpha \wedge \beta$ ,  $\exists x_i\alpha$ , and  $\exists P_j\alpha$  for any  $i, j < \omega$  and any formula  $\alpha$ . We use the usual abbreviations  $x_i > x_j$ ,  $x_i \leq x_j$ ,  $x_i < x_j < x_k$ ,  $\forall x_i\alpha$  and  $\forall P_i\alpha$  etc.

As usual we define the concept of a free individual variable symbol in a formula. We similarly define the set of free monadic variables of a formula. Write  $\phi(x_1, \dots, x_m, P_1, \dots, P_n)$  to indicate that all the free variables (of both sorts) in the formula  $\phi$  are contained in the lists  $x_1, \dots, x_m$  and  $P_1, \dots, P_n$ .

The language is used to describe linear orders. Suppose that  $(T, <)$  is a linear order. As individual variable symbols we will often use  $t, s, r, u$  etc, instead of  $x_1, x_2, \dots$ .

An individual variable assignment  $V$  is a mapping from  $X$  into  $T$ . A predicate variable assignment  $W$  is a mapping from  $Q$  into  $\wp(T)$  (the set of subsets of  $T$ ). For an individual variable assignment  $V$ , an individual variable symbol  $x \in X$  and an element  $t \in T$ , we define the individual variable assignment  $V[x \mapsto t]$  by:

$$V[x \mapsto t](y) = \begin{cases} V(y) & y \neq x \\ t & y = x. \end{cases}$$

Similarly for predicate variable assignments and subsets of  $T$ .

For a formula  $\phi$ , variable assignments  $V$  (individual) and  $W$  (predicate), we define whether (or not resp.)  $\phi$  under  $V$  and  $W$  is true in  $(T, <)$ , written  $(T, <), V, W \models \phi$  by induction on the quantifier depths of  $\phi$ .

Given some  $\phi$ , suppose that for all its subformulas  $\psi$ , for all variable assignments  $V$  and  $W$ , we have defined whether or not  $(T, <), V, W \models \psi$ . For variable assignments  $V$  and  $W$  we define:



$$\begin{aligned}
 (T, <), V, W \models x_i < x_j & \text{ iff } V(x_i) < V(x_j); \\
 (T, <), V, W \models x_i = x_j & \text{ iff } V(x_i) = V(x_j); \\
 (T, <), V, W \models P_i(x_j) & \text{ iff } V(x_j) \in W(P_i); \\
 (T, <), V, W \models \neg\alpha & \text{ iff } (T, <), V, W \not\models \alpha; \\
 (T, <), V, W \models \alpha \wedge \beta & \text{ iff } (T, <), V, W \models \alpha \text{ and } (T, <), V, W \models \beta; \\
 (T, <), V, W \models \exists x_i \alpha & \text{ iff there exists some } t \in T \text{ such that} \\
 & (T, <), V[x_i \mapsto t], W \models \alpha. \\
 (T, <), V, W \models \exists P_j \alpha & \text{ iff there is some } S \subseteq T \text{ such that} \\
 & (T, <), V, W[P_j \mapsto S] \models \alpha.
 \end{aligned}$$

This is standard second-order semantics. Note that it is easy to show that the truth of a formula does not depend on the assignment to variables which do not appear free in the formula.

Mostly we will be interested in fragments of the full second-order monadic logic. In particular, we refer to the first-order monadic logic of linear order which contains just those formulas with no quantification of predicate variables. We will also mention the universal second-order monadic logic of linear order which contains just those formulas which consist of a first-order monadic formula nested under zero or more universal quantifications of predicate variables.

The important correspondence for us is that between temporal logics such as *US/LT* and the first-order monadic logic. Most of the temporal logics which we will consider allow a certain equivalence between their formulas and first-order monadic formulas. To define this we need to use a fixed one-to-one correspondence between the propositional atoms of the temporal language and monadic predicate variables. Let us suppose that  $p_i \in \mathcal{L}$  corresponds to  $P_i \in \mathcal{Q}$ .

The translation will propagate upwards through the full temporal language provided that each of the connectives have a first-order translation. In particular we require for any  $n$ -ary temporal connective  $C$  some first-order monadic formula  $\phi_C(t, P_1, \dots, P_n)$  which corresponds to  $C(p_1, \dots, p_n)$ . We say that  $\phi_C$  is the (*first-order*) *table* of  $C$  iff for every linear order  $(T, <)$ , for all  $h: \mathcal{L} \rightarrow \wp(T)$ , for all  $t_0 \in T$ , for all variable assignments  $V$  and  $W$ ,

$$\begin{aligned}
 (T, <, h), t_0 \models C(p_1, \dots, p_n) \\
 \text{iff } (T, <), V[t \mapsto t_0], W[P_1 \mapsto h(p_1), \dots, P_n \mapsto h(p_n)] \models \phi_C.
 \end{aligned}$$

$U$  and  $S$  have first-order tables as follows: the table of  $U$  is  $\phi_U = \exists s((t < s) \wedge P_1(s) \wedge \forall r((t < r \wedge r < s) \rightarrow P_2(r)))$ .

The table of  $S$  is the mirror image.

If we have a temporal logic with first-order tables for its connectives then it is straightforward to define a meaning-preserving translation (to the first-order monadic language) of all formulas in the language. The translation is

as follows:

$$\begin{aligned}
* p_i &= P_i(t) \\
* \top &= (t = t) \\
* \neg A &= \neg * A \\
* (A \wedge B) &= * A \wedge * B \\
* U(A, B) &= \phi_U[*A/P_1][*B/P_2] \\
* S(A, B) &= \phi_S[*A/P_1][*B/P_2]
\end{aligned}$$

Here, if  $\alpha$  is a first-order monadic formula with one free individual variable  $t$ , we use the notation  $\phi[\alpha/P]$  to mean the result of replacing every (free) occurrence of the predicate variable symbol  $P$  in the first-order monadic formula  $\phi$  by the monadic formula  $\alpha$ : i.e.  $P(x_i)$  gets replaced by  $\alpha[x_i/t]$ . This is a little complex as we must take care to avoid clashes of variable symbols. We will not go into details here.

For any temporal formula  $A$ ,  $*A$  is a first-order monadic formula with  $t$  being the only free individual variable symbol. If atom  $p_i$  appears in  $A$  then  $*A$  will also have a free predicate symbol  $P_i$ . There are no other predicate symbols in  $*A$ .

We then have:

LEMMA 1. *For each linear order  $(T, <)$ , for any  $t_0 \in T$ , for any temporal  $A$ , if  $A$  uses atoms from  $p_1, \dots, p_n$  then for all variable assignments  $V$  and  $W$ ,*

$$\begin{aligned}
(T, <, h), t_0 \models A \text{ iff } (T, <), V[t \mapsto t_0], W[P_1 \mapsto h(p_1)] \dots \\
[P_n \mapsto h(p_n)] \models *A.
\end{aligned}$$

**Proof.** By induction on the construction of  $A$ . ■

Below, when we discuss expressive power in section 3, we will consider whether there is a reverse translation from the first-order monadic language to the temporal language.

### *Uses of US/LT*

In the previous chapter [Burgess, 2001], we saw a brief survey of the motivations for developing a tense logic. There are particular reasons for concentrating on the logic *US/LT*. It is a basis for reasoning about events and states in general linear time or particularly dense or continuous time.

For example, it is certainly the case that most use of tense and aspect in natural language occur in the context of an assumed dense linear flow of time. Any kind of reasoning about the same sorts of situations, as in many branches of AI or cognitive science, also requires a formalism based on dense, continuous or general linear time.

In computer science the most widely used temporal logic (PTL which we will meet later) is based on a discrete natural numbers model of time. However, there has been much recent work on developing logics based on more general models for many applications. Particular applications include refinement, analogue devices, open systems, and asynchronous distributed or concurrent systems.

Refinement here concerns the process of making something more concrete or more specific. This can include making a specification for a system (or machine, or software system) less ambiguous, or more detailed, or less nondeterministic. This can also include making an algorithm, or program, or implementation, or design, more detailed, or more low level. There are several different ways of producing a refinement but one involves breaking up one step of a process into several smaller steps which accomplish the same overall effect. If a formal description of a less refined process assumes discrete steps of time then it is easy to see that it may be hard to relate it to a description of a more refined process. Extra points of time may have to be introduced in between the assumed ones. Using a dense model of time from the outset will be seen to avoid this problem. Comparing the formal descriptions of more and less refined processes is essential for checking the correctness of any refinement procedure.

It can also be seen that a general linear model of time will be useful in describing analogue devices (which might vary in state in a continuous way), open systems (which might be affected by an unbounded number of different environmental events happening at all sorts of times) and asynchronous and distributed systems (which may have processes going through changes in state at all sorts of times).

In general the logics are used to describe or specify, to verify, to model-check, to synthesize, or to execute. A useful description of these activities can be found in [Emerson, 1990]. Specification is the task of giving a complete, precise and unambiguous description of the behaviour expected of a system (or device). Verification is the task of checking, or proving, that a system does conform to a specification and this includes the more specific task of model-checking, or determining whether a given system conforms to a particular property. Synthesis is the act of more or less automatically producing a correct system from a specification (so this avoids the need for verification). Finally execution is the process of directly implementing a specification, treating the specification language as an implementation (or programming) language.

In many of these applications it is crucial to determine whether a formula is a consequence of a set of formulas. For example, we may have a large and detailed set of formulas exactly describing the behaviour of the system and we have a small and very interesting formula describing a crucial desired property of the system (e.g., “it will fly”) and we want to determine whether the latter follows from the former. We turn to this question now.

## 2.2 An Axiomatization for the Logic

We have seen the importance of the consequence relation in applications of temporal logic. Because of this there are good reasons to consider syntactical ways of determining this relation between sets of formulas and a single formula. One of the most widely used and widely investigated such approaches is via a Hilbert system. Here we look in detail at a Hilbert style axiom system for our  $US/LT$  logic.

The system, which we will call  $Ax(U, S)$ , was first presented in [Burgess, 1982] but was later simplified slightly in [Xu, 1988]. It has what are the usual inference rules for a temporal logic: i.e. modus ponens and two generalizations, temporal generalization towards the future and temporal generalization towards the past:

$$\frac{A, A \rightarrow B}{B} \quad \frac{A}{GA} \quad \frac{A}{HA}$$

Each rule has a list of formulas (or just one formula) as antecedents shown above the horizontal line and a single formula, the consequent below the line. An *instance* of a rule is got by choosing any particular  $L(U, S)$  formulas for the  $A$  and  $B$ . We describe the role of a rule below.

The axioms of  $Ax(U, S)$  are all substitution instances of truth-functional tautologies and the following temporal schemas:

- (1)  $G(A \rightarrow B) \rightarrow (U(A, D) \rightarrow U(B, D))$ ,
- (2)  $G(A \rightarrow B) \rightarrow (U(D, A) \rightarrow U(D, B))$ ,
- (3)  $A \wedge U(B, D) \rightarrow U(B \wedge S(A, D), D)$ ,
- (4)  $U(A, B) \rightarrow U(A, B \wedge U(A, B))$ ,
- (5)  $U(B \wedge U(A, B), B) \rightarrow U(A, B)$ ,
- (6)  $U(A, B) \wedge U(D, E) \rightarrow$   
 $(U(A \wedge D, B \wedge E) \vee U(A \wedge E, B \wedge E) \vee U(B \wedge D, B \wedge E))$ .

along with the mirror images of (1) to (6). Notice that (1) and (2) are closely related to the usual axioms for the modal logic  $K$ , (3) relates the mirror image connectives, (4) and (5) have something to do with transitivity and (6) captures an aspect of linearity.

We say that a formula  $B$  follows from a list  $A_1, \dots, A_n$  of formulas by one of the rules of inference iff there is an instance of the rule with  $A_1, \dots, A_n$  as the antecedents and  $B$  as the consequent. A deduction in  $Ax(U, S)$  is a finite sequence of formulas with each being either an instance of one of the axioms or following from a list of previous formulas in the sequence by one of the rules of inference. Any formula which appears as the last element in a derivation is called a *thesis* of the system. If  $A$  is a thesis then we write  $\vdash A$ .

We say that a formula  $D$  is a (*syntactic*) *consequence* of a set  $\Gamma$  of formulas iff there are  $C_1, \dots, C_n \in \Gamma$  such that  $\vdash (\bigwedge_{i=1}^n C_i) \rightarrow D$ . In that case we write  $\Gamma \vdash D$ . Note that in some other logics it is possible to use a Hilbert system to define consequence via the idea of a deduction from hypotheses. In general in temporal logic, we do not do so because of problems with the generalization rules.

Note that alternative presentations of such a system may use what is known as the substitution rule:

$$\frac{A}{A[B/q]}$$

We have a whole collection of instances of the substitution rule: one for each ( formula, atom , formula) triple. If  $B$  is a formula and  $q$  is an atom then we define the substitution  $A[B/q]$  of  $q$  by  $B$  in a formula  $A$  by induction on the construction of  $A$ . We simply have  $q[B/q] = B$  and  $p[B/q] = p$  for any atom  $p$  other than  $q$ . The induction then respects every other logical operator, e.g.,  $(A_1 \wedge A_2)[B/q] = (A_1[B/q]) \wedge (A_2[B/q])$ . If we include the substitution rule in an axiom system then axioms can be given in terms of particular atoms. For example, we could have an axiom  $G(p \rightarrow q) \rightarrow (U(p, r) \rightarrow U(q, r))$ .

*Soundness of The System*

We will now consider the relation between syntactic consequence and semantic consequence.

We say that an axiom system is *sound* (with respect to a semantically defined logic) iff any syntactic consequence ( of the axiom system) is also a semantic consequence. This can readily be seen to be equivalent to the property of every thesis being valid.

We can show that:

LEMMA 2. *The system  $Ax(U, S)$  is sound for  $US/LT$ .*

**Proof.** Via a simple induction it is enough to show that every axiom is valid and that each rule of inference preserves validity, i.e. that the consequent is valid if all the antecedents are.

The axioms are straightforward to check individually using obvious semantic arguments.

The inference rules are equally straightforward to check. For example, let us look at temporal generalization towards the future. Suppose that  $A$  is valid. We are required to show that  $GA$  is valid. So suppose that  $(T, <, g)$  is a linear structure and  $t \in T$ . Consider any  $s > t$ . By the validity of  $A$  we have  $(T, <, g), s \models A$ . This establishes that  $(T, <, g), t \models GA$  as required. ■

### *Completeness of the System*

An axiom system is said to be (sound and) *complete* for a logic iff syntactic consequence exactly captures semantic consequence.

In fact, this is sometimes called *strong completeness* because there is a weaker notion of completeness which is still useful. We say that an axiom system is *weakly complete* for a logic iff it is sound and every validity is a thesis. Clearly weak completeness follows strong completeness. As we will see, there are temporal logics for which we can only obtain weakly complete axiom systems.

However,

**THEOREM 3.**  *$Ax(U, S)$  is strongly complete for  $US/LT$ .*

**Proof.** We sketch the proof from [Burgess, 1982].

We need some common definitions in order to proceed. We say that a set of formulas is *consistent* (in the context of an axiom system) iff it is not the case that  $\Gamma \vdash \perp$ . If  $\Gamma$  is maximal in being consistent, i.e. the addition to  $\Gamma$  of any other formula in the language would result in inconsistency, then we say that  $\Gamma$  is a maximal consistent set (MCS). A useful result, due to Lindenbaum (see [Burgess, 2001]), gives us,

**LEMMA 4.** *If  $\Delta$  is a consistent set then there is an MCS  $\Gamma \supseteq \Delta$ .*

There are many useful properties of MCSs, e.g.,

**LEMMA 5.** *If  $\Gamma$  is an MCS then  $A \in \Gamma$  iff  $\neg A \notin \Gamma$ .*

In order to show the completeness of the axiom system we need only show that each MCS is satisfiable. To see this suppose that  $\Gamma \models A$ . Thus  $\Gamma \cup \{\neg A\}$  is unsatisfiable. It can not be consistent as then by lemma 4 it could be extended to an MCS which we know would be satisfiable. Thus we must be able to derive  $\perp$  from  $\Gamma \cup \{\neg A\}$ , say  $\bigwedge_{i=1}^n C_i \wedge \neg A \vdash \perp$ . It is a simple matter to show that then we have  $\vdash \bigwedge_{i=1}^n C_i \rightarrow A$ , i.e. that  $A$  is a syntactic consequence of  $\Gamma$ .

So suppose that  $\Gamma_0$  is an MCS: we want to show that it is satisfiable. We use the rationals as a base board on which we successively place whole maximal consistent sets of formulas as points which will eventually make up a flow of time. So, at each stage, we will have a subset  $T \subseteq \mathbb{Q}$  and an MCS  $\Gamma(t)$  for each  $t \in T$ .

Starting with our given maximal consistent set placed at zero, say, we look for counter-examples to either of the following rules:

1. if  $U(A, B) \in \Gamma(t)$  then there should be some  $\Gamma(s)$  placed at  $s > t$  with  $A \in \Gamma(s)$  and so that theories placed in between  $t$  and  $s$  all contain  $B$  and
2. if  $\neg U(A, B) \in \Gamma(t)$  and  $A \in \Gamma(s)$  placed at  $s > t$  then there should be  $\Gamma(r)$  placed somewhere in between with  $\neg B \in \Gamma(r)$ .

By carefully choosing a single maximal consistent set to right the counter-example and satisfy some other stringent conditions kept holding throughout the construction, we can ensure that the particular tuple  $(t, U(A, B))$  or  $(t, s, \neg U(A, B))$  never again forms a counter-example. In order to do so we need to record, for each interval between adjacent sets, which formulas must be belong to any set subsequently placed in that interval. Because there are only countable numbers of points and formulas involved, in the limit we can effect that we end up with a counter-example-free arrangements of sets. This is so nice that if we define a valuation  $h$  on the final  $T \subseteq \mathbb{Q}$  via  $t \in h(p)$  iff  $p \in \Gamma(t)$ , then for all  $t \in T$ , for all  $A \in L(U, S)$ ,

$$(T, <, h), t \models A \text{ iff } A \in \Gamma(t).$$

This is thus our model. ■

*The IRR rule*

Here we will examine a powerful alternative approach to developing Hilbert systems for temporal logics like  $US/LT$ . It is based on the use of rules such as the IRR (or irreflexivity) rule of [Gabbay, 1981]. Recall that a binary relation  $<$  on a set  $T$  is irreflexive if we do not have  $t < t$  for any  $t \in T$ . The IRR rule allows

$$\frac{q \wedge H(\neg q) \rightarrow A}{A} \quad \text{provided that the atom } q \text{ does not appear in the formula } A.$$

A short proof (see for example [Gabbay and Hodkinson, 1990], Proposition 2.2.1) establishes that,

LEMMA 6. *if  $\mathcal{I}$  is a class of (irreflexive) linear orders, then IRR is a valid rule in the class of all structures whose underlying flow of time comes from  $\mathcal{I}$ .*

**Proof.** Suppose that  $q \wedge H(\neg q) \rightarrow A$  is valid. Consider any structure  $(T, <, h)$  with  $(T, <)$  from  $\mathcal{I}$  and any  $t \in T$ . Let  $g$  be a valuation of the atoms on  $T$  which is like  $h$  but differs only in that  $g(q) = \{t\}$ , i.e.  $g = h[q \mapsto \{t\}]$ . By the assumed validity,  $(T, <, g) \models q \wedge H(\neg q) \rightarrow A$ . But notice that also  $(T, <, g) \models q \wedge H(\neg q)$ . Thus  $(T, <, g) \models A$ . Clearly, since  $q$  does not appear in  $A$ ,  $(T, <, h) \models A$ . Thus  $A$  is valid in the logic. ■

The original motivation for the use of this rule concerned the impossibility of writing an axiom to enforce irreflexivity of flows (see for example [van Benthem, 1991]). The usual technique in a completeness proof is to construct some model of a consistent formula and then turn it into an irreflexive model. IRR allows immediate construction of an irreflexive model. This is because it is always consistent to posit the truth of  $q \wedge H(\neg q)$  (for some ‘new’ atom  $q$ ) at any point as we do the construction.

The benefits of this rule for doing a completeness proof are enormous. Much use of it is made in [Gabbay *et al.*, 1994]. Venema [Venema, 1993] gives a long list of results proved using IRR or similar rules: examples include branching time logics [Zanardo, 1991] and two-dimensional modal logics [Kuhn, 1989]. In fact, the major benefit of IRR is a side-effect of its purpose. Not only can we construct a model which is irreflexive but we can construct a model in which each point has a unique name (as the first point where a certain atom holds).

As an example, consider our logic  $US/LT$ . We can, in fact mostly just use the standard axiomatization for  $F$  and  $P$  over the class of all linear orders. This is because if you have a unique name of the form  $r \wedge H(\neg r)$  for each point then their axiom

$$(UU) \quad r \wedge H(\neg r) \rightarrow [U(p, q) \leftrightarrow F(p \wedge H[Pr \rightarrow q])]$$

and its mirror image (SS) essentially define  $U$  and  $S$  in terms of  $F$  and  $P$ .

So here is an axiom system for  $US/LT$  similar to those seen in [Gabbay and Hodkinson, 1990] and [Gabbay *et al.*, 1994]. Call it  $Z$ . The rules are modus ponens, two generalizations, substitution:

$$\frac{A, A \rightarrow B}{B} \quad \frac{A}{GA} \quad \frac{A}{HA} \quad \frac{A}{A[B/q]}$$

and the IRR:

$$\frac{q \wedge H(\neg q) \rightarrow A}{A} \quad \text{provided that the atom } q \text{ does not appear in the formula } A.$$

The axioms are:

1. all truth functional tautologies,
2.  $G(p \rightarrow q) \rightarrow (Gp \rightarrow Gq)$ ,
3.  $Gp \rightarrow GGp$ ,
4.  $G(p \wedge Gp \rightarrow q) \vee G(q \wedge Gq \rightarrow p)$ ,
5.  $r \wedge H\neg r \rightarrow (U(p, q) \leftrightarrow F(p \wedge H(Pr \rightarrow q)))$ ,

And mirror images of the above.

**THEOREM 7.** *The axiom system is sound and (weakly) complete for  $US/LT$ .*

Soundness is the usual induction on the lengths of proofs.

For completeness we have to do quite a bit of extra work. However this extra work is quite general and can form the basis of many and varied completeness proofs. The general idea is along the lines of the usual Henkin-style completeness proof for Prior's logic over linear time (see [Burgess, 2001]) but there is no need of bulldozing of clusters. It is as follows.



Say that a set is  $Z$ -consistent iff there is no conjunction  $A$  of its formulas such that  $Z$  derives  $A \rightarrow \perp$ . We are interested in maximal consistent sets of a particular sort which we call IRR theories. They contain a ‘name’ of the form  $\neg q \wedge Hq$  and also similarly name any other time referred to by some zig-zagging sequence of  $F$ s and  $P$ s.

**DEFINITION 8.** A theory  $\Delta$  is said to be an IRR-theory if (a) and (b) hold:

- (a) For some  $q$ ,  $\neg q \wedge Hq \in \Delta$ .
- (b) Whenever  $X = \diamond_1(A_1 \wedge \diamond_2(A_2 \wedge \dots \wedge \diamond_n A_n)) \dots \in \Delta$ ,  
then for some new atom  $q$ ,  
 $X(q) = \diamond_1(A_1 \wedge \diamond_2(A_2 \wedge \dots \wedge \diamond_n(A_n \wedge \neg q \wedge Hq))) \dots \in \Delta$ ,  
where  $\diamond_i$  is either  $P$  or  $F$ .

We say that a theory  $\Delta$  is complete iff for all formulas  $A$ ,  $A \in \Delta$  iff  $\neg A \notin \Delta$ . The following lemma plays the part of the Lindenbaum lemma in allowing us to work with maximal consistent IRR-theories.

**LEMMA 9.** *Let  $A$  be any  $Z$ -consistent formula. Then there exists a complete,  $Z$ -consistent IRR-theory,  $\Delta$ , such that  $A \in \Delta$ . In fact, if  $\Delta_0$  is any  $Z$ -consistent theory such that an infinite number of atomic propositions  $q_i$  do not appear in  $\Delta_0$ , then there exists an IRR  $Z$ -consistent and complete theory  $\Delta \supseteq \Delta_0$ .*

Note that we are only proving a weak completeness result as we need to have a large number of spare atoms.

The main work of the truth lemma of the completeness proof is done by the following lemma in which we say that  $\Delta < \Gamma$  iff for all  $GA \in \Delta$ ,  $A \in \Gamma$ .

**LEMMA 10.** *Let  $\Delta$  be a  $Z$ -consistent complete IRR-theory. Let  $FA \in \Delta$  ( $PA \in \Delta$  respectively). Then there exists a  $Z$ -consistent complete IRR-theory  $\Gamma$  such that  $A \in \Gamma$  and  $\Delta < \Gamma$  ( $\Delta > \Gamma$  respectively).*

We can finish the completeness proof by then constructing a model from the consistent complete IRR-theories in the usual way: i.e. as worlds in our structure we use all those theories which are connected by some finite amount of  $<$  zig-zagging with  $\Delta_0$  which contains our formula of interest. The frame of such sets under  $<$  is made a structure by making an atom  $p$  true at the point  $\Delta$  iff  $p \in \Delta$ . So far this is the usual Henkin construction as seen in [Burgess, 2001] for example. The frame will automatically be irreflexive because every set contains an atom which is true there for the first time. So there is no need for bulldozing. Transitivity and linearity follow from the appropriate axioms.

Now the IRR rule, as well as making the completeness proof easier, also arguably makes proving from a set of axioms easier. This is because, being able to consistently introduce names for points into an axiomatic proof

makes the temporal system more like the perhaps more intuitive first-order one. There are none of the problems such as with losing track of “now”.

So given all these recommendations one is brought back to the question of why is it useful to do away with IRR. The growing body of theoretical work (see for example Venema [1991; 1993]) trying to formalize conditions under which the orthodox (to use Venema’s term) system of rules needs to be augmented by something like IRR can be justified as follows:

- adding a new rule of inference to the usual temporal ones is arguably a much more drastic step than adding axioms and it is always important to question whether such additions are necessary;
- in making an unorthodox derivation one may need to go beyond the original language in order to prove a theorem, which makes such axiomatizations less attractive from the point of view of ‘resource awareness’;
- (as argued in [Venema, 1991]), using an atom to perform the naming task of an individual variable in predicate logic is not really in the spirit of temporal/modal logic;

and

- (also as mentioned in [Venema, 1991]), unorthodox axiomatizations do not have some of the nice mathematical properties that orthodox systems have.

### 2.3 Decidability of $US/LT$

We have seen that it is often useful to be able to approach the question of consequence in temporal logics in a syntactic way. For many purposes it is enough to be able to determine validity as this is equivalent to determining consequence between finite sets of formulas. A *decision procedure* for a logic is an algorithm for determining whether any given formula is valid or not. The procedure must give correct “yes” or “no” answers for each formula of the language. A logic is said to be *decidable* iff there exists such a decision procedure for its validities.

Notice that a decision procedure is able to tell us more about validities than a complete axiom system. The decision procedure can tell us when a formula is not a validity while an axiom system can only allow us to derive the validities. This is important for many applications.

For many of the most basic temporal logics some general results allow us to show decidability. The logic  $US/LT$  is such a logic.

A traditional way of showing decidability for temporal logics is via the so-called *finite model property*. We say that a logic has a finite model property iff any satisfiable formula is satisfiable in a finite model, a model with a

finite number of time points. A systematic search through all finite models coupled with a systematic search through all validities from a complete axiom system gives us a decision procedure. See [Burgess, 2001] for further details of the finite model approach to decidability: this can not be used with  $US/LT$  as there are formulas, such as  $U(\top, \perp) \wedge GU(\top, \perp)$  which are satisfiable in infinite models only.

The logic  $US/LT$  can be shown to be decidable using the translation  $*$  of section 2.1 into the first-order monadic logic of order. In [Ehrenfeucht, 1961] it was shown that the first-order logic of linear order is decidable. Other proofs in [Gurevich, 1964] and [Läuchli and Leonard, 1966] show that this also applies to the first-order monadic logic of linear order. The decidability of  $US/LT$  follows immediately via lemma 1 and the effectiveness of  $*$ .

An alternative proof uses the famous result in [Rabin, 1969] showing the decidability of a second-order monadic logic. The logic is  $S2S$ , the second-order logic of two successors. The language has two unary function symbols  $l$  and  $r$  as well as a countably infinite number of monadic predicate symbols  $P_1, P_2, \dots$ . The formulas are interpreted in the binary tree structure  $(T, l, r)$  of all finite sequences of zeros and ones with:

$$l(a) = a^{\wedge}0, r(a) = a^{\wedge}1 \text{ for all } a \in T.$$

As usual a sentence of the language is a formula with no free variables. Rabin shows that  $S2S$  is decidable: i.e. there is an algorithm which given a sentence of  $S2S$ , correctly decides whether or not the sentence is true of the binary tree structure. Proofs of Rabin's difficult result use tree automata.

Rabin's is a very powerful decidability result and much used in establishing the decidability of other logics. For example, Rabin uses it to show that the full monadic second-order theory of the rational order is decidable. That is, there is an algorithm to determine whether a formula in the full monadic second-order language of order (as defined in section 2.1) is true of the order  $(\mathbb{Q}, <)$ . A short argument via the downward Löwenheim-Skolem theorem (see [Gurevich, 1964] or [Gabbay *et al.*, 1994]) then establishes that the universal monadic second-order theory of the class of all linear flows of time is decidable. Thus  $US/LT$  is too.

Once a decision procedure is known to exist for a useful logic it becomes an interesting problem to develop an efficient decision procedure for it. That is an algorithm which gives the "yes" or "no" answers to formulas. We might want to know about the fastest possible such algorithms, i.e. the complexity of the decision problem. To be more precise we need to turn the decision problem for a logic into a question for a Turing machine. There is a particular question about the symbolic representation of atomic propositions since we allow them to be chosen from an infinite set of atoms. A careful approach (seen in a similar example in [Hopcroft and Ullman, 1979]) is to suppose (by renaming) that the propositions actually used in a particular formula are  $p_1, \dots, p_n$  and to code  $p_i$  as the symbol  $p$  followed by  $i$  written

in binary. Of course this means that the input to the machine might be a little longer than the length of the formula. In fact a formula of length  $n$  may correspond to an input of length about  $n \log_2 n$ . There is no problem with output: we either want a “1” for, “yes, the formula is a validity” or a “0” for, “no, the formula is not a validity”.

Once we have a rigorously defined task for a Turing machine then we can ask all the usual questions about which complexity classes e.g., P, NP, PSPACE, EXPTIME, etc the problem belongs to. A further, very practical question then arises in the matter of actually describing and implementing efficient decision procedures for the logic. We return briefly to such questions later in the chapter.

For  $US/LT$  the complexity is an open problem. A result in [Reynolds, 1999] shows that it is PSPACE-hard. Essentially we can encode the running of a polynomially space-bounded Turing Machine in the logic. It is also believed that the decision problem for the logic is in PSPACE but, so far, this is just conjecture. The procedures which are contained within the decidability proofs above are little help as the method in [Läuchli and Leonard, 1966] relies on an enumeration of the validities in the first-order logic (with no clear guide to its complexity) and the complexity of Rabin’s original procedure is non-elementary.

#### 2.4 Other flows of time

We have had a close look at the logic  $US/LT$ . There are many other temporal logics. We can produce other logics by varying our language and its semantics (as we will see in subsection 2.5 below) and we can produce other logics by varying the class of structures which we use to define validity. Let  $\mathcal{K}$  be some class of linear orders. We define the  $L(U, S)$  (or Kamp) logic over  $\mathcal{K}$  to have validities exactly those formulas  $A$  of  $L(U, S)$  which are true at all points  $t$  from any structure  $(T, <, h)$  where  $(T, <) \in \mathcal{K}$ .

For example, the formula  $\neg U(\top, \perp)$  is a validity of the  $L(U, S)$  logic over the class of all dense linear orders. Let us have a closer look at this logic. In fact we can completely axiomatize this logic by adding the following two axioms to our complete axiomatization of  $US/LT$ :

$$\begin{aligned} &\neg U(\top, \perp), \\ &\neg S(\top, \perp), \end{aligned}$$

Soundness is clear. To show completeness, strong completeness, assume that  $\Gamma$  is a maximally consistent set of formulas, and here, this means consistent with the new axioms system. However,  $\Gamma$  will also be consistent with  $Ax(U, S)$  and so will have a linear model by theorem 1. Say that  $(T, <)$  is linear,  $t_0 \in T$ , and for all  $C \in \Gamma$ ,  $(T, <, h), t_0 \models C$ . Now  $G\neg U(\top, \perp) \in \Gamma$  because it this formula is a theorem derivable by generalization from one

of the new axioms. Thus, it is clear that no point in the future of  $t_0$  has a discrete successor. Similarly there are no such discrete jumps in the past or immediately on either side of  $t_0$ . Thus  $(T, <, h)$  was a dense model all along.

Decidability of the  $L(U, S)$  logic over dense time follows almost directly from the decidability of  $L(U, S)$  over linear time: to decide  $A$  over dense time just decide

$$A \wedge \neg U(\top, \perp) \wedge G\neg U(\top, \perp) \wedge \neg S(\top, \perp) \wedge H\neg S(\top, \perp).$$

A more specific logic still is the  $L(U, S)$  logic of rational (numbers) time. Here we define validity via truth at all points in any structure  $(\mathbb{Q}, <, g)$  (where  $<$  is the usual irreflexive ordering on the rationals). Such a logic has uses in reasoning about events and states when it might be inconvenient to assume that time is Dedekind complete. For example, a well situated gap in time could save arguments about whether there is a last moment when a light is on or a first moment when it is off. To axiomatize this logic it is enough to add to the system for dense time axioms asserting that there is neither an end of time nor a beginning:

$$GF\top, HPT.$$

The system is clearly sound. There are two ways to see that it is complete, both using the fact that any countable dense ordering without end points is (isomorphic to) the rationals. One way is to notice that Burgess' construction for a model of a set of formulas consistent with  $Ax(U, S)$  does construct a countable one. The other way is to use the downward Löwenheim-Skolem theorem on the monadic translations of the temporal formulas.

The same sorts of moves give us decidability of the  $L(U, S)$  logic of the rationals via the decidability over general linear time.

Another useful specific dense logic is the  $L(U, S)$  logic over real (numbers) time, sometimes loosely called continuous time temporal logic. This is used in many applications as the real numbers seem to be the right model of time for many situations. Unfortunately the real numbers are not straightforward to describe with temporal axioms. The logic was first axiomatized in [Gabbay and Hodkinson, 1990] using a combination of techniques from [Läuchli and Leonard, 1966], [Burgess and Gurevich, 1985] and [Doets, 1989] to do with definable equivalence classes, the IRR approach to axiomatizing the  $L(U, S)$  logics over linear time and expressive completeness ideas which we will see in section 3 below.

The axiomatization in [Gabbay and Hodkinson, 1990] consists of the basic axiom system for  $L(U, S)$  logic over general linear time using the IRR rule (see above) plus:

$P\top \wedge F\top$	no end points,
$Fp \rightarrow FFp$	density,
$FGq \wedge F\neg q \rightarrow F(Gq \wedge \neg PGq)$	future Dedekind completeness,
$PHq \wedge P\neg q \rightarrow P(Hq \wedge \neg FHq)$	past Dedekind completeness,

and a new axiom,

$$\begin{aligned} & F(q \wedge F(q \wedge r \wedge H\neg r)) \wedge U(r, q \rightarrow \neg U(q, \neg q)) \\ \rightarrow & F(K^+q \wedge K^-q \wedge F(r \wedge H\neg r)) \end{aligned}$$

called the SEP rule.

The Dedekind completeness axioms are due to Prior and, as we will see, can be used with  $F$  and  $P$  logics to capture Dedekind completeness, the property of there being no gaps in the flow of time. In fact, these axioms just ensure definable Dedekind completeness, i.e. that there are no gaps in time in a structure which can be noticed by looking at the truth values of formulas.

The axiom SEP is interesting. Nothing like it is needed to axiomatize continuous temporal logic with only Prior's connectives as the property it captures is not expressible without  $U$  or  $S$  and hence without  $K^+$  or  $K^-$ . SEP is associated with the separability of  $\mathbb{R}$ , i.e. the fact that it has a dense countable suborder (e.g., the rationals). It says roughly that if a formula is densely true in an interval then there is a point at which the formula is true both arbitrarily soon before and afterwards. That SEP is necessary in the axiom system is shown in [Gabbay and Hodkinson, 1990] when a structure is built in which all substitution instances of the other axioms including the Prior ones are valid while SEP is not. Such structures also show that the  $L(U, S)$  logic over the reals is distinct from the amp logic over arbitrary continuous flows of time i.e. those that are dense, Dedekind complete and without end points.

The completeness proof only gives a weak completeness result: i.e. the axiom system allows derivation of all validities but it does not give us the general consequence relation between a possibly infinite set of formulas and a formula. In fact it is impossible to give a strongly complete axiom system for this logic because it is not *compact*: there is an infinite set of formulas which is inconsistent but every finite subset of it is consistent. Here is one example:

$$\Gamma = \{FG\neg p, G\neg K^-p, A_0, A_1, \dots\}$$

where  $A_0 = Fp$  and for each  $n$ ,  $A_{n+1} = FA_n$ .

The proof relies on building a not necessarily real flowed model  $M$  of a given satisfiable formula  $A$ , say, and then showing that for each  $n$ , there is a real flowed structure which satisfies the same monadic sentences to

quantifier depth  $n$  as  $M$  does. By choosing  $n$  to be one more than the depth of quantifiers in  $*A$  one can see that we have a model of  $A$  by reasoning about the satisfiability of the monadic sentence  $\exists x * A(x)$ .

The later axiom system in [Reynolds, 1992] is complete for the  $L(U, S)$  logic over the reals and does not use the IRR rule. Instead it adds the following axioms to  $Ax(U, S)$ :

$$\begin{array}{ll}
 K^+ \top, K^- \top, F \top, P \top & \text{as before,} \\
 U(\top, p) \wedge F \neg p \rightarrow U(\neg p \vee K^+ \neg p, p) & \text{Prior-U,} \\
 S(\top, p) \wedge P \neg p \rightarrow S(\neg p \vee K^- \neg p, p) & \text{Prior-S, and} \\
 K^+ p \wedge \neg K^+(p \wedge U(p, \neg p)) \rightarrow K^+(K^+ p \wedge K^- p) & \text{SEP2}
 \end{array}$$

Prior-U and its mirror image are just versions of the Dedekind completeness axioms and SEP2 is a neater version of SEP also developed by Ian Hodkinson. The proof of completeness is similar to that in [Gabbay and Hodkinson, 1990] but requires quite a bit more work as the “names” produced by the IRR rule during construction are not available to help reason about definable equivalence classes.

The decidability of the  $L(U, S)$  logic over real time is also not straightforward to establish. It was proven by two different methods in [Burgess and Gurevich, 1985]. One method uses a variant of a traditional approach: show that a formula that is satisfiable over the reals is also satisfiable over the rationals under a valuation which conforms to a certain definition of “niceness”, show that a formula satisfiable under a “nice” valuation on the rationals is satisfiable over the reals, and show that deciding satisfiability under nice valuations over the rationals is decidable. The other method uses arguments about definable equivalence relations as in the axiomatization above. Both methods use Rabin’s decidability result for  $S2S$  and Kamp’s expressive completeness result which we will see in a later section.

The complexity of the decision problem for the  $L(U, S)$  logic over the reals is an open problem.

Now let us consider the  $L(U, S)$  logics over discrete time. To axiomatize the  $L(U, S)$  logic over the integers it is not enough to add the following discreteness and non-endedness axioms to the Burgess system  $Ax(U, S)$ :

$$U(\top, \perp) \text{ and } S(\top, \perp).$$

In fact, we must add these and Prior-style Dedekind completeness axioms such as:

$$Fp \rightarrow U(p, \neg p)$$

and its mirror image. To prove weak completeness (–it is clear that this logic is not compact–) requires a watered down version of the mechanisms for real numbers time or other ways of finding an integer-flowed model from a model with a definably Dedekind complete valuation over some other countable,

discrete flow without end points. There is a proof in [Reynolds, 1994]. An alternative axiom system in the usual IRR style is probably straightforward to construct. The decidability of this logic follows from the decidability of the full monadic second-order theory of the integers which was proved in [Büchi, 1962]. Again the complexity of the problem is open.

When we turn to natural numbers time we find the most heavily studied temporal logics. This is because of the wide-ranging computer science applications of such logics. However, it is not the  $L(U, S)$  logic which is studied here but rather logics like PTL which concentrate on the future and which we will meet in section 7 below. The  $S$  connective can be shown to be unnecessary in expressing properties: to see this is a straightforward use of the separation property of the  $L(U, S)$  logic over the natural numbers (see section 3 below). Despite this it has been argued (e.g., in [Lichtenstein *et al.*, 1985]) that  $S$  can help in allowing natural expression of certain useful properties: it is not necessarily easy or efficient to re-express the property without using  $S$ . Thus, axiom systems for the  $L(U, S)$  logic over the natural numbers have been presented. In [Lichtenstein *et al.*, 1985], such a complete system is given which is in the style of the axiom systems for the logic PTL which we will meet in section 7 below (and so we will not describe it here). In [Venema, 1991] a different but still complete axiom system is given along with others for  $L(U, S)$  logics over general classes of well-orderings. This system is simply  $Ax(U, S)$  with axioms for discreteness, Dedekind completeness beginning and no end. Again the completeness proof is subtle because the logic is not compact and there are many different countable, discrete, Dedekind complete orderings with a beginning and no end. The  $L(U, S)$  logic over the natural numbers is known to be decidable via monadic logic arguments (via [Büchi, 1962]) and, in [Lichtenstein *et al.*, 1985], a PSPACE decision procedure is given and the problem is shown to be PSPACE-complete.

### 2.5 Other linear time logics

We have met a variety of temporal logics based on using Kamp's  $U$  and  $S$  connectives (on top of propositional logic) over various classes of linear orders. Basing a logic on other classes of not necessarily linear orders can also give us useful or interesting logics as we will see in section 4 below. However, there is another way of constructing other temporal logics. For various reasons it might be interesting to build a language using other temporal connectives. We may want the temporal language to more closely mimic a particular natural language with its own ways of representing tense or aspect. We may think that  $U$  and  $S$  do not allow us to express some important properties. Or we may think that  $U$  and  $S$  allow us to express too much and so the  $L(U, S)$  language is unnecessarily complex to reason with for our particular application.



In the next few sections we will consider temporal logics for reasoning about certain classes of linear flows of time based on a variety of temporal languages. By a temporal language here we will mean a language built on top of propositional logic via the recursive use of one or more temporal connectives. By a temporal connective we will mean a logical connective symbol with a first order table as defined above.

Some of the common connectives include, as well as  $U$  and  $S$ ,:

- $Fp$   $\exists s > tP(s)$ ,  
it will sometime be the case that  $p$ ;
- $Pp$   $\exists s < tP(s)$ ,  
it was sometime the case that  $p$ ;
- $Xp$   $\exists s > tP(s) \wedge \neg \exists r(t < r < s)$ ,  
there is a next instant and  $p$  will hold then;
- $Yp$   $\exists s < tP(s) \wedge \neg \exists r(s < r < t)$ ,  
there was a previous instant and  $p$  held then.

Note that some (all) of these connectives can be defined in terms of  $U$  and  $S$ .

A traditional temporal (or modal) logic is that with just the connective  $F$  over the class of all linear flows of time. This logic (often with the symbol  $\diamond$  used for  $F$ ) is traditionally known as K4.3 because it can be completely axiomatized by axioms from the basic modal system  $K$  along with an axiom known as 4 (for transitivity) and an axiom for linearity which is not called 3 but usually  $L$ . The system includes modus ponens substitution and (future) temporal generalization and the axioms:

$$\begin{aligned} G(p \rightarrow q) &\rightarrow (Gp \rightarrow Gq) \\ Gp &\rightarrow GGp \\ G(p \wedge Gp \rightarrow q) &\vee G(q \wedge Gq \rightarrow p) \end{aligned}$$

where  $GA$  is the abbreviation  $\neg F\neg A$  in terms of  $F$  in this language. The proof of (strong) completeness involves a little bit of rearranging of maximal consistent sets as can be seen in [Burgess, 2001] or [Bull and Segerberg, in this handbook]. The decidability and NP-completeness of the decision problem can be deduced from the result of [Ono and Nakamura, 1980] mentioned shortly.

Adding Prior's past connective  $P$  to the language, but still defining consequence over the class of all linear orders results in the basic linear  $L(F, P)$  logic which is well described in [Burgess, 2001]. A strongly complete axiom system can be obtained by adding mirror images of the rules and axioms in K4.3.

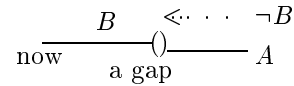
To see that the linear  $L(F, P)$  logic is decidable one could simply call on the decidability of the  $L(U, S)$  logic over linear time (as seen above). It is a trivial matter to see that a formula in the  $L(F, P)$  language can be translated

directly into an equivalent formula in the  $L(U, S)$  language. An alternative approach is to show that the  $L(F, P)$  logic has a finite model property: if  $A$  is satisfiable (in a linear structure) then  $A$  is satisfiable in a finite structure (of some type). As described in [Burgess, 2001], in combination with the complete axiom system, this gives an effective procedure for deciding the validity of any formula.

A third alternative is to use the result in [Ono and Nakamura, 1980] that if a  $L(F, P)$  formula of length  $n$  is satisfiable in a linear model then it is satisfiable in a finite connected, transitive, totally ordered but not necessarily anti-symmetric or irreflexive model containing at most  $n$  points. This immediately gives us a non-deterministic polynomial time decision procedure. Since propositional logic is NP-complete we conclude that the linear  $L(F, P)$  logic is too.

Another linear time logic has recently been studied in [Reynolds, 1999]. This is the linear time logic with just the connective  $U$ . It was studied because, despite the emerging applications of reasoning over general linear time, as we saw above, it is not known how computationally complex it is to decide validity in the linear  $L(U, S)$  logic. As a first step to solving this problem the result in this paper shows that the problem of deciding formulas with just  $U$  is PSPACE-complete. The proof uses new techniques based on the “mosaics” of [Németi, 1995]. A mosaic-based decision procedure consists in trying to establish satisfiability by guessing and checking a set of model pieces to see if they can be put together to form a model. Mosaics were first used in deciding a temporal logic in [Reynolds, 1998]. It is conjectured that similar methods may be used to show that deciding the  $L(U, S)$  logic is also PSPACE-complete.

The logics above have all been obviously not more expressive than the  $L(U, S)$  logic of linear time. Are there linear time temporal logics which are more expressive than the  $L(U, S)$  logic? We will see later that the answer is yes and that a completely expressive language (in a manner to be defined precisely) contains two more connectives along with Kamp’s. These are the Stavi connectives which were defined in [Gabbay *et al.*, 1980].  $U'(A, B)$  holds if  $B$  is true from now until a gap in time after which  $B$  is arbitrarily soon false but after which  $A$  is true for a while:  $U'(A, B)$  is as pictured



$S'$  is defined via the mirror image. Despite involving a gap,  $U'$  is in fact a first-order connective. Here is the first-order table for  $U'$ :

$$\begin{aligned}
 U'(p, q) \equiv & \\
 \exists s \quad & t < s \\
 \wedge \quad & \forall u \quad ( \\
 & \quad ( [ \quad \exists v(u < v \wedge \forall w(t < w < v \rightarrow q(w)) \quad ] \\
 & \quad \vee \quad [ \quad \forall v(u < v < s \rightarrow p(v)) \\
 & \quad \quad \wedge \quad \exists v(t < v < u \wedge \neg q(v)) \quad ] ) ) \\
 \wedge \quad & \exists u[t < u < s \wedge \neg q(u)] \\
 \wedge \quad & \exists u[t < u < s \wedge \forall v(t < v < u \rightarrow q(v))]
 \end{aligned}$$

Of course,  $S'$  has the mirror image table.

We will see in section 3 below that the logic with  $U$ ,  $S$ ,  $U'$  and  $S'$  is expressively complete for the class of structures with linear flow of time. There is no known complete axiom system for the logic with this rather complicated set of connectives. The decidability of the logic follows from the decidability of first-order monadic logic. However, the complexity of the decision problem is also open. If it was shown to be PSPACE-complete for example, then we would have the very interesting result that this temporal logic is far far easier to reason with than the equally expressive monadic logic (with one free variable).

Probably the most useful dense linear time temporal logics are those based on the real-numbers flow of time. Because it is expressively complete (as we will see), the  $L(U, S)$  logic over the reals is the most important such logic. The Stavi connectives are useless over the reals. We have seen that this logic can be axiomatized in several ways, and is decidable. However, the complexity of the decision procedure is not known. Other, less expressive, real-flowed temporal logics can be defined. Logics built with any combination of connectives from  $\{F, P, U, S\}$  will clearly be decidable. There is an axiomatization of the  $L(F, P)$  logic over the reals in [Burgess, 2001].

Various authors have studied a real-flowed temporal logic with a slightly unusual semantics. We say that a structure  $(\mathbb{R}, <, h)$  has *finite variability* iff in any bound interval of time, there are only finitely many points of time between which the truth values of all atoms are constant. A logic can be defined by evaluating  $U$  and  $S$  only on finitely variable structures. This allows the logic to be useful for reasoning about many situations but makes it amenable to the sorts of techniques which are used to reason about sequences of states and natural numbers time temporal logics. See [Kesten *et al.*, 1994] and [Rabinovich, 1998] for more details.

### 3 THE EXPRESSIVE POWER OF TEMPORAL CONNECTIVES

The expressivity of a language is always measured with respect to some other language. That is, when talking about expressivity, we are always comparing two or more languages. When measuring the expressivity of a large number of languages, it is usually more convenient to have a single

language with respect to which all other languages can be compared, if such a language is known to exist.

In the case of propositional one-dimensional temporal languages defined by the presence of a fixed number of *temporal connectives* (also called *temporal modalities*), the expressivity of those languages can be all measured against a fragment of first-order logic, namely the *monadic first-order language*. This is the fragment that contains a binary  $<$  (to represent the underlying temporal order),  $=$  (which we assume is always in the language) and a set of unary predicates  $Q_1(x), Q_2(x), \dots$  (which account for the interpretation of the propositional letters, that are interpreted as a subset of the temporal domain  $T$ ). Indeed, any one-dimensional temporal connective can be defined as a well-formed formula in such a fragment, known as the connective's *truth table*; one-dimensionality forces such truth tables to have a single free variable.

In the case of comparing the expressivity of temporal connectives, another parameter must be taken into account, namely the underlying flow of time. Two temporal languages may have the same expressivity over one flow of time (say, the integers) but may differ in expressivity over another (e.g. the rationals); see the discussion on the expressivity of the  $US$  connectives below.

Let us exemplify what we mean by those terms. Consider the connectives *since*( $S$ ), *until*( $U$ ), *future*( $F$ ), and *past*( $P$ ). Given a flow of time  $(T, <, h)$ , the truth value of each of the above connectives at a point  $t \in T$  is determined as follows:

$$\begin{aligned}
(T, <, h), t \models Fp & \quad \text{iff} \quad (\exists s > t)(T, <, h), s \models p, \\
(T, <, h), t \models Pp & \quad \text{iff} \quad (\exists s < t)(T, <, h), s \models p, \\
(T, <, h), t \models U(p, q) & \quad \text{iff} \quad (\exists s > t)((T, <, h), s \models p \wedge \\
& \quad \forall y(t < y < s \rightarrow (T, <, h), y \models q)), \\
(T, <, h), t \models S(p, q) & \quad \text{iff} \quad (\exists s < t)((T, <, h), s \models p \wedge \\
& \quad \forall y(s < y < t \rightarrow (T, <, h), y \models q))
\end{aligned}$$

If we assume that  $h(p)$  represents a first-order unary predicate that is interpreted as  $h(p) \subseteq T$ , then these truth values above can be expressed as first-order formulas. Thus:

- (a)  $(T, <, h), t \models Fq$  iff  $\chi_F(t, h(q))$  holds in  $(T, <)$ ,
- (b)  $(T, <, h), t \models Pq$  iff  $\chi_P(t, h(q))$  holds in  $(T, <)$ ,
- (c)  $(T, <, h), t \models U(q_1, q_2)$  iff  $\chi_U(t, h(q_1), h(q_2))$  holds in  $(T, <)$ , and
- (d)  $(T, <, h), t \models S(q_1, q_2)$  iff  $\chi_S(t, h(q_1), h(q_2))$  holds in  $(T, <)$ .

where

- (a)  $\chi_F(t, Q) = (\exists s > t)Q(s)$ ,
- (b)  $\chi_P(t, Q) = (\exists s < t)Q(s)$ ,
- (c)  $\chi_U(t, Q_1, Q_2) = (\exists s > t)(Q_1(s) \wedge \forall y(t < y < s \rightarrow Q_2(y)))$ ,
- (d)  $\chi_S(t, Q_1, Q_2) = (\exists s < t)(Q_1(s) \wedge \forall y(s < y < t \rightarrow Q_2(y)))$ .

$\chi_{\#}(t, Q_1, \dots, Q_n)$  is called the *truth table* for the connective  $\#$ . The number  $n$  of parameters in the truth table will be the number of places in the connective, e.g.  $F$  and  $P$  are one place connective, and their truth tables have a single parameter;  $S$  and  $U$  are two-place connectives, with truth tables having two parameters.

It is clear that in such a way, we start defining any number of connectives. For example consider  $\chi(t, Q) = \exists xy(t < x < y \wedge \forall s(x < s < y \rightarrow Q(s))$ ; then  $\chi(t, Q)$  means ‘There is an interval in the future of  $t$  inside which  $P$  is true.’ This is a table for a connective  $F_{\text{int}}$ :  $(T, <, h), t \models F_{\text{int}}(p)$  iff  $\chi(t, h(p))$  holds in  $(T, <)$ .

We are in condition of presenting a general definition of what a temporal connective is:

**DEFINITION 11.**

1. Any formula  $\chi(t, Q_1, \dots, Q_m)$  with one free variable  $t$ , in the monadic first-order language with predicate variable symbols  $Q_i$ , is called an  $m$ -place truth table (in one dimension).
2. Given a syntactic symbol  $\#$  for an  $m$ -place connective, we say it has a truth table  $\chi(t, Q_1, \dots, Q_m)$  iff for any  $T, h$  and  $t$ ,  $(*)$  holds:
  - $(*) : (T, <, h), t \models \#(q_1, \dots, q_m)$  iff  $(T, <) \models \chi(t, h(q_1), \dots, h(q_m))$ .

This way we can define as many connectives as we want. Usually, some connectives are definable using other connectives. For example, it is well known that  $F$  is definable using  $U$  as  $Fp \equiv U(p, \top)$ . As another example, consider a connective that states the existence of a “next” time point:  $\delta \equiv U(\top, \perp)$ .

The connective  $\delta$  is a nice example on how the definability of a connective by others depends on the class of flows of time being considered. For example, in a dense flow of time,  $\delta$  can be defined in terms of  $F$  and  $P$  — actually, since there are no “next” time points anywhere,  $\delta \equiv \perp$ . Similarly, in an integer-like flow of time,  $\delta$  is equivalent to  $\top$ .

On the other hand, consider the flow  $(T, <)$  of time with a single point without a “next time”:  $T = \{\dots - 2, -1, 0, 1, 2, \dots\} \cup \{(1/n) \mid n = 1, 2, 3, \dots\}$ , with  $<$  being the usual order; then  $\delta$  is not definable using  $P$  and  $F$ . To see that, suppose for contradiction that  $\delta$  is equivalent to  $A$  where  $A$  is written with  $P$  and  $F$  and, maybe, atoms. Replace all appearances of atoms by  $\perp$

to obtain  $A'$ . Since  $\delta \leftrightarrow A$  holds in the structure  $(T, <, h')$  with all atoms always false, in this structure  $\delta \leftrightarrow A'$  holds. As neither  $\delta$  nor  $A'$  contain atoms,  $\delta \leftrightarrow A'$  holds in all other  $(T, <, h)$  as well. Now  $A'$  contains only  $P$  and  $F$ ,  $\top$ , and  $\perp$  and the classical connectives. Since  $F\top \equiv P\top \equiv \top$  and  $F\perp \equiv P\perp \equiv \perp$ , at every point,  $A'$  must be equivalent (in  $(T, <)$ ) to either  $\top$  or  $\perp$  and so cannot equal  $\delta$  which is true at 1 and false at 0. As a consequence,  $\delta$  is not definable using  $P$  and  $F$  over linear time.

In general, given a family of connectives, e.g.  $\{F, P\}$  or  $\{U, S\}$ , we can build new connectives using the given ones. That these new connectives are connectives in the sense of Definition 11 follows from the following.

**LEMMA 12.** *Let  $\#_1(q_1, \dots, q_{m_1}), \dots, \#_n(q_1, \dots, q_{m_n})$  be  $n$  temporal connectives with tables  $\chi_1, \dots, \chi_n$ . Let  $A$  be any formula built up from atoms  $q_1, \dots, q_m$ , the classical connectives, and these connectives. Then there exists a monadic  $\psi_A(t, Q_1, \dots, Q_m)$  such that for all  $T$  and  $h$ ,*

$$(T, <, h), t \models A \text{ iff } (T, <) \models \psi_A(t, h(q_1), \dots, h(q_m)).$$

**Proof.** We construct  $\psi_A$  by induction on  $A$ . The simple cases are:  $\psi_{q_j} = Q_j(t)$ ,  $\psi_{\neg A} = \neg\psi_A$  and  $\psi_{A \wedge B} = \psi_A \wedge \psi_B$ .

For the temporal connective case, we construct the formula  $\psi_{\#_i(A_1, \dots, A_{m_i})} = \chi_i(t, \psi_{A_1}, \dots, \psi_{A_{m_i}})$ ; the right-hand side is a notation for the formula obtained by substituting  $\psi_{A_j}(x)$  in  $\chi_i$  wherever  $Q_j(x)$  appears, with the appropriate renaming of bound variables to avoid clashes. The induction hypothesis is applied over  $\psi_{A_1}, \dots, \psi_{A_{m_i}}$  and the result is simply obtained by truth table of the connective  $\#_i$ . ■

The formula  $\psi_A$  built above is called the *first-order translation* of a temporal formula  $A$ . An  $m$ -place connective  $\#$  with truth table  $\chi(t, Q_1, \dots, Q_m)$  is said to be *definable* from connectives  $\#_1, \dots, \#_n$  in a flow of time  $(T, <)$  if there exists a temporal formula  $A$  built from those connectives whose first order translation is  $\psi_A$  such that

$$(T, <) \models \psi_A \leftrightarrow \chi.$$

The *expressive power* of a family of connectives over a flow of time is measured by how many connectives it can express over the flow of time. If it can express any conceivable connective (given by a monadic formula), then that family of connectives is expressively complete.

**DEFINITION 13.** A temporal language with one-dimensional connectives is said to be *expressively complete* or, equivalently, *functionally complete*, in one dimension over a class  $\mathcal{T}$  of partial orders iff for any monadic formula  $\psi(t, Q_1, \dots, Q_m)$ , there exists an  $A$  of the language such that for any  $(T, <)$  in  $\mathcal{T}$ , for any interpretation  $h$  for  $q_1, \dots, q_m$ ,

$$(T, <) \models \forall t(\psi \leftrightarrow \psi_A)(t, h(q_1), \dots, h(q_m)).$$

In the cases where  $\mathcal{T} = \{(T, <)\}$  we talk of expressive completeness over  $(T, <)$ . For example, the language of Since and Until is expressively complete over integer time and real number flow of time, as we are going to see in Section 3.2; but they are not expressively complete over rational numbers time [Gabbay *et al.*, 1980].

DEFINITION 14. A flow of time  $(T, <)$  is said to be expressively complete (or functionally complete) (in one dimension) iff there exists a *finite* set of (one-dimensional) connectives which is expressively complete over  $(T, <)$ , in one dimension.

The qualification of one-dimensionality in the definitions above will be explained when we introduce the notion of H-dimension below.

These notions parallel the definability and expressive completeness of classical logic. We know that in classical logic  $\{\neg, \rightarrow\}$  is sufficient to define all other connectives. Furthermore, for any  $n$ -place truth table  $\psi : 2^n \rightarrow 2$  there exists an  $A(q_1, \dots, q_n)$  of classical logic such that for any  $h$ ,

$$h(A) = \psi(h(q_1), \dots, h(q_n)).$$

This is the expressive completeness of  $\{\neg, \rightarrow\}$  in classical logic.

The notion of expressive completeness leads us to formulate two questions:

- (a) Given a finite set of connectives and a class of flows of time, are these connectives expressively complete?
- (b) In case the answer to (a) is no, we would like to ask: given a class of flows, does there exist a finite set of one-dimensional connectives that is expressively complete?

These questions occupy us to the rest of this section. We show that the notion of expressive completeness is intimately related to the *separation property*.

The answer to question (b) is related to the notion of H-dimension, discussed in Section 3.3.

### 3.1 Separation and Expressive Completeness

The notion of separation involves partitioning a flow of time in disjoint parts (typically: present, past and future). A formula is separable if it is equivalent to another formula whose temporal connectives refer only to one of the partitions.

If every formula in a language is separable, that means that we have at least one connective that has enough expressivity over each of the partitions. So we might expect that that set of connectives is expressively complete over

a class of flows that admits such partitioning, provided the partitioning is also expressible by the connectives.

The notion of separation was initially analysed in terms of linear flows, where the notion of present, past and future most naturally applies. So we start our discussion with separation over linear time. We later extend separation to generic flows.

*Separation over linear time*

Consider a linear flow of time  $(T, <)$ . Let  $h, h'$  be two assignments and  $t \in T$ . We say that  $h, h'$  agree on the past of  $t$ ,  $h =_{<t} h'$ , iff for any atom  $q$  and any  $s < t$ ,

$$s \in h(q) \text{ iff } s \in h'(q).$$

We define  $h' =_{=t} h$  for *agreement of the present*, iff for any atom  $q$

$$t \in h(q) \text{ iff } t \in h'(q).$$

and  $h' =_{>t} h$ , for *agreement on the future*, iff for any atom  $q$  and any  $s > t$ ,

$$s \in h(q) \text{ iff } s \in h'(q).$$

Let  $\mathcal{T}$  be a class of linear flows of time and  $A$  be a formula in a temporal language over  $(T, <)$ . We say that  $A$  is a *pure past formula over  $\mathcal{T}$* , iff for all  $(T, <)$  in  $\mathcal{T}$ , for all  $t \in T$ ,

$$\forall h, h', (h =_{<t} h') \text{ implies that } (T, <, h), t \models A \text{ iff } (T, <, h'), t \models A.$$

Similarly, we define *pure future* and *pure present* formulas.

Such a definition of purity is a semantic one. In a temporal language containing only  $S$  and  $U$  there is also have a notion of syntactic purity as follows. A formula is a *Boolean combination* of  $\phi_1, \dots, \phi_n$  if it is built from  $\phi_1, \dots, \phi_n$  using only Boolean connectives. A *syntactically pure present* formula is a Boolean combination of atoms only. A *syntactically pure past* formula is a Boolean combination of formulas of the form  $S(A, B)$  where  $A$  and  $B$  are either pure present or pure past. Similarly, a *syntactically pure future* formula is a Boolean combination of formulas of the form  $U(A, B)$  where  $A$  and  $B$  are either pure present or pure future.

It is clear that if  $A$  is a syntactically pure past formula, then  $A$  is a pure past formula; similarly for pure present and pure future formulas. The converse, however, is not true. For example, from the semantical definition, all temporal temporally valid formulas are pure future (and pure past, and pure present), including those involving  $S$ .

We are now in a position to define the separation property.

**DEFINITION 15.** Let  $\mathcal{T}$  be a class of linear flows of time and  $A$  be a formula in a temporal language  $\mathbf{L}$ . We say  $A$  is *separable* in  $\mathbf{L}$  over  $\mathcal{T}$  iff there exists



a formula in  $\mathbf{L}$  which is a Boolean combination of pure past, pure future, and atomic formulas and is equivalent to  $A$  everywhere in any  $(T, <)$  from  $\mathcal{T}$ .

A set of temporal connectives is said to have the separation property over  $\mathcal{T}$  iff every formula in the temporal language of these connectives is separable in the language (over  $\mathcal{T}$ ).

We now show that separation implies expressive completeness.

**THEOREM 16.** *Let  $\mathbf{L}$  be a temporal language built from any number (finite or infinite) of connectives in which  $P$  and  $F$  are definable over a class  $\mathcal{T}$  of linear flows of time. If  $\mathbf{L}$  has the separation property over  $\mathcal{T}$  then  $\mathbf{L}$  is expressively complete over  $\mathcal{T}$ .*

**Proof.** If  $\mathcal{T}$  is empty,  $\mathbf{L}$  is trivially expressively complete, so suppose not. We have to show that for any  $\varphi(t, \overline{Q})$  in the monadic theory of linear order with predicate variable symbols  $\overline{Q} = (Q_1, \dots, Q_n)$ , there exists a formula  $A = A(q_1, \dots, q_n)$  in the temporal language such that for all flows of time  $(T, <)$  from  $\mathcal{T}$ , for all  $h, t$ ,  $(T, <, h), t \models A$  iff  $(T, <) \models \varphi(t, h(q_1), \dots, h(q_n))$ .

We denote this formula by  $A[\varphi]$  and proceed by induction on the depth  $m$  of nested quantifiers in  $\varphi$ . For  $m = 0$ ,  $\varphi(t)$  is quantifier free. Just replace each appearance of  $t = t$  by  $\top$ ,  $t < t$  by  $\perp$ , and each  $Q_j(t)$  by  $q_j$  to obtain  $A[\varphi]$ .

For  $m > 0$ , we can assume  $\varphi = \exists x \psi(t, x, \overline{Q})$  where  $\psi$  has quantifier depth  $\leq m$  (the  $\forall$  quantifier is treated as derived).

Assuming that we do not use  $t$  as a bound variable symbol in  $\psi$  and that we have replaced all appearances of  $t = t$  by  $\top$  and  $t < t$  by  $\perp$  then the atomic formulas in  $\psi$  which involve  $t$  have one of the following forms:  $Q_i(t)$ ,  $t < y$ ,  $t = y$ , or  $y < t$ , where  $y$  could be  $x$  or any other variable letter occurring in  $\psi$ .

If we regard  $t$  as fixed, the relations  $t < y$ ,  $t = y$ ,  $t > y$  become unary and can be rewritten, respectively, as  $R_{<}(y)$ ,  $R_{=}(y)$  and  $R_{>}(y)$ , where  $R_{<}$ ,  $R_{=}$  and  $R_{>}$  are new unary predicate symbols.

Then  $\psi$  can be rewritten equivalently as

$$\psi_0^t(x, \overline{Q}, R_{=}, R_{>}, R_{<}),$$

in which  $t$  appears only in the form  $Q_i(t)$ . Since  $t$  is free in  $\psi$ , we can go further and prove (by induction on the quantifier depth of  $\psi$ ) that  $\psi_0^t$  can be equivalently rewritten as

$$\psi^t = \bigvee_j [\alpha_j(t) \wedge \psi_j^t(x, \overline{Q}, R_{=}, R_{>}, R_{<})],$$

where

- $\alpha_j(t)$  is quantifier free,

- $Q_i(t)$  appear only in  $\alpha_j(t)$  and not at all in  $\psi_j^t$ ,
- and each  $\psi_j^t$  has quantifier depth  $\leq m$ .

By the induction hypothesis, there is a formula  $A_j = A_j(\bar{q}, r_=:, r_>, r_<)$  in the temporal language such that, for any  $h, x$ ,

$$(T, <, h), x \models A_j \text{ iff} \\ (T, <) \models \psi_j^t(x, h(q_1), \dots, h(q_n), h(r_=), h(r_>), h(r_<)).$$

Now let  $\diamond q$  be an abbreviation for a temporal formula equivalent (over  $\mathcal{T}$ ) to  $Pq \vee q \vee Fq$  whose existence in  $\mathbf{L}$  is guaranteed by hypothesis. Then let  $B(\bar{q}, r_=:, r_>, r_<) = \bigvee_j (A[\alpha_j] \wedge \diamond A_j)$ .  $A[\alpha_j]$  can be obtained from the quantifier free case.

In any structure  $(T, <)$  from  $\mathcal{T}$  for any  $h$  interpreting the atoms  $\bar{q}, r_=:, r_>$  and  $r_<$ , the following are straightforward equivalences

$$\begin{aligned} (T, <, h), t \models B \\ (T, <, h), t \models \bigvee_j (A[\alpha_j] \wedge \diamond A_j) \\ \bigvee_j ((T, <, h), t \models A[\alpha_j] \wedge (T, <, h), t \models \diamond A_j) \\ \bigvee_j (\alpha_j(t) \wedge \exists x ((T, <, h), x \models A_j)) \\ \bigvee_j (\alpha_j(t) \wedge \exists x \psi_j^t(x, h(q_1), \dots, h(q_n), h(r_=), h(r_>), h(r_<))) \\ \exists x \bigvee_j (\alpha_j(t) \wedge \psi_j^t(x, h(q_1), \dots, h(q_n), h(r_=), h(r_>), h(r_<))) \\ \exists x \psi_0^t(x, h(q_1), \dots, h(q_n), h(r_=), h(r_>), h(r_<)). \end{aligned}$$

Now provided we interpret the  $r$  atoms as the appropriate  $R$  predicates, i.e.:

- $h^*(r_=:) = \{t\}$ ,
- $h^*(r_<) = \{s \mid t < s\}$ , and
- $h^*(r_>) = \{s \mid s < t\}$ ,

we obtain

$$(T, <, h^*), t \models B \text{ iff } \exists x \psi(t, x, h^*(q_1), \dots, h^*(q_n)) \text{ iff} \\ \varphi(t, h^*(q_1), \dots, h^*(q_n)).$$

$B$  is almost the  $A[\varphi]$  we need except for one problem.  $B$  contains, besides the  $q_i$ , also three other atoms,  $r_=:, r_>$ , and  $r_<$ , and equation (\*) from Definition 9.1.1 above is valid for any  $h^*$  which is arbitrary on the  $q_i$  but very special on  $r_=:, r_>, r_<$ . We are now ready to use the separation property (which we haven't used so far in the proof). We use separation to eliminate  $r_=:, r_>, r_<$  from  $B$ . Since we have separation  $B$  is equivalent to a Boolean combination of atoms, pure past formulas, and pure future formulas.

So there is a Boolean combination  $\beta = \beta(\overline{p}_+, \overline{p}_-, p_0)$  such that

$$B \leftrightarrow \beta(\overline{B}_+, \overline{B}_-, B_0),$$

where  $B_0(\overline{q}, r_>, r_=:, r_<)$  is a combination of atoms,  $B_+(\overline{q}, r_>, r_=:, r_<)$  are pure future, and  $B_-(\overline{q}, r_>, r_=:, r_<)$  are pure past formulas.

Finally,  $B^* = \beta(B_+^*, B_-^*, B_0^*)$  where

- $B_0^* = B_0(\overline{q}, \perp, \top, \perp)$ ;
- $B_+^* = B_+(\overline{q}, \top, \perp, \perp)$ ;
- $B_-^* = B_-(\overline{q}, \perp, \perp, \top)$ .

Then we obtain for any  $h^*$ ,

$$\begin{aligned} (T, <, h^*), t \models B & \text{ iff } (T, <, h^*), t \models \beta(B_+, B_-, B_0) \\ & \text{ iff } (T, <, h^*), t \models \beta(B_+^*, B_-^*, B_0^*) \\ & \text{ iff } (T, <, h^*), t \models B^*. \end{aligned}$$

Hence

$$(T, <, h^*), t \models B^* \text{ iff } (T, <) \models \varphi(t, h^*(\overline{q})).$$

This equation holds for any  $h^*$  arbitrary on  $\overline{q}$ , but restricted on  $r_<, r_>, r_=:$ . But  $r_<, r_>, r_=:$  do not appear in it at all and hence we obtain that for any  $h$ ,  $(T, <, h), t \models B^*$  iff  $(T, <) \models \varphi(t, h^*(\overline{q}))$ . So make  $A[\varphi] = B^*$  and we are done.  $\blacksquare$

The converse is also true: expressive completeness implies separation over linear time. The proof involves using the first-order theory of linear time to first separate a first-order formula over linear time; a temporal formula is translated into the first-order language, where it is separated; expressive completeness is needed then to translate each separated first-order subformula into a temporal formula. Details are omitted, but can be found in [Gabbay *et al.*, 1994].

### Generalized Separation

The separation property is not restricted to linear flows of time. In this section we generalize the separation property over any class of flows of time and see that Theorem 16 has a generalised version.

The basic idea is to have some relations that will partition every flow of time in  $\mathcal{T}$ , playing the role of  $<$ ,  $>$  and  $=$  in the linear case.

**DEFINITION 17.** Let  $\varphi_i(x, y), i = 1, \dots, n$  be  $n$  given formulas in the monadic language with  $<$  and let  $\mathcal{T}$  be a class of flows of time. Suppose  $\varphi_i(x, y)$  partition  $\mathcal{T}$ , that is, for every  $t$  in each  $(T, <)$  in  $\mathcal{T}$  the sets  $T(i, t) = \{s \in T \mid \varphi_i(s, t)\}$  for  $i = 1, \dots, n$  are mutually exclusive and  $\bigcup_i T(i, t) = T$ .

In analogy to the way that  $F$  and  $P$  represented  $<$  and  $>$ , we assume that for each  $i$  there is a formula  $\beta_i(t, x)$  such that  $\varphi_i(t, x)$  and  $\beta_i(t, x)$  are equivalent over  $\mathcal{T}$  and  $\beta_i$  is a Boolean combination of some  $\varphi_j(x, t)$ . Also assume that  $<$  and  $=$  can be expressed (over  $\mathcal{T}$ ) as Boolean combinations of the  $\varphi_i$ .

Then we have the following series of definitions:

- For any  $t$  from any  $(T, <)$  in  $\mathcal{T}$ , for any  $i = 1, \dots, n$ , we say that truth functions  $h$  and  $h'$  agree on  $T(i, t)$  if and only if  $h(q)(s) = h'(q)(s)$  for all  $s$  in  $T(i, t)$  and all atoms  $q$ .
- We say that a formula  $A$  is *pure*  $\varphi_i$  over  $\mathcal{T}$  if for any  $(T, <)$  in  $\mathcal{T}$ , any  $t \in T$  and any two truth functions  $h$  and  $h'$  which agree on  $T(i, t)$ , we have

$$(T, <, h), t \models A \text{ iff } (T, <, h'), t \models A.$$

- The logic  $\mathbf{L}$  has the *generalized separation property* over  $\mathcal{T}$  iff every formula  $A$  of  $\mathbf{L}$  is equivalent over  $\mathcal{T}$  to a Boolean combination of pure formula.

**THEOREM 18** (generalized separation theorem). *Suppose the language  $\mathbf{L}$  can express over  $\mathcal{T}$  the 1-place connectives  $\#_i$ ,  $i = 1, \dots, n$ , defined by:*

$$(T, <, h), t \models \#_i(p) \quad \text{iff} \quad \exists s \quad \varphi_i(s, t) \text{ holds in } (T, <) \\ \text{and } (T, <, h), s \models p.$$

*If has the generalized separation property over a class  $\mathcal{T}$  of flows of time then  $\mathbf{L}$  is expressively complete over  $\mathcal{T}$ .*

A proof of this result appears in [Amir, 1985]. See also [Gabbay *et al.*, 1994].

The converse does not always hold in the general case, for it depends on the theory of the underlying class  $\mathcal{T}$ .

A simple application of the generalised separation theorem is the following. Suppose we have a first order language with the binary order predicates  $<$ ,  $>$ ,  $=$  with their usual interpretation, and suppose it also contains a *parallel* operator  $|$  defined by:

$$x|y =_{def} \neg[(x = y) \vee (x < y) \vee (y < x)].$$

Suppose we have a new temporal connective  $D$ , defined by

$$(T, <, h), t \models Dq \text{ iff } \exists x|t \text{ such that } (T, <, h), x \models q.$$

Finally,  $A$  is said to be *pure parallel* over a class  $\mathcal{T}$  of flows of time iff for all  $t$  from any  $(T, <)$  from  $\mathcal{T}$ , for all  $h =|_t h'$ ,

$$(T, <, h), t \models A \text{ iff } (T, <, h'), t \models A,$$

where  $h =_{|t} h'$  iff  $\forall x|t\forall q(x \in h(q) \leftrightarrow x \in h'(q))$ .

It is clear what separation means in the context of pure present, past, future, and parallel. It is simple to check that the  $<, >, =, |$  satisfy the general separation property and other preconditions for using the generalized separation theorem. Thus that theorem gives immediately the following.

**COROLLARY 19.** *Let  $\mathbf{L}$  be a language with  $F, P, D$  over any class of flows of time. If  $\mathbf{L}$  has a separation then  $\mathbf{L}$  is expressively complete.*

### 3.2 Expressive Completeness of Since and Until over Integer Time

As an example of the applications of separation to the expressive completeness of temporal language, we are going to sketch the proof of separation of the Since and Until-temporal logic containing over linear time. The full proof can be found in [Gabbay, 1989; Gabbay *et al.*, 1994]. With separation and Theorem 16 we immediately obtain that the connectives  $S$  and  $U$  are expressively complete over the integers; the original proof of the expressive completeness of  $S$  and  $U$  over the integers is due to Kamp [Kamp, 1968b].

The basic idea of the separation process is to start with a formula in which  $S$  and  $U$  may be nested inside each other and through several transformation steps we are going to systematically remove  $U$  from inside  $S$  and vice-versa. This gives us a syntactical separation which, obviously, implies separation.

As we shall see there are eight cases of nested occurrences of  $U$  within an  $S$  to worry about. It should be noted that all the results in the rest of this section have dual results for the *mirror images* of the formulas. The *mirror image* of a formula is the formula obtained by interchanging  $U$  and  $S$ ; for example, the mirror image of  $U(p \wedge S(q, r), u)$  is  $S(p \wedge U(q, r), u)$ .

We start dealing with Boolean connectives inside the scope of temporal operators, with some equivalences over integer flows of time. We say that a formula  $A$  is *valid* over a flow of time  $(T, <)$  if it is true at all  $t \in T$ ; notation:  $(T, <) \models A$

**LEMMA 20.** *The following formulas (and their mirror images) are valid over integer time:*

- $U(A \vee B, C) \leftrightarrow U(A, C) \vee U(B, C);$
- $U(A, B \wedge C) \leftrightarrow U(A, B) \wedge U(A, C);$
- $\neg U(A, B) \leftrightarrow G(\neg A) \vee U(\neg A \wedge \neg B, \neg A);$
- $\neg U(A, B) \leftrightarrow G(\neg A) \vee U(\neg A \wedge \neg B, B \wedge \neg A).$

**Proof.** Simply apply the semantical definitions. ■

We now show the eight separation cases involving simple nesting and atomic formulas only.

LEMMA 21. *Let  $p, q, A$ , and  $B$  be atoms. Then each of the formulas below is equivalent, over integer time, to another formula in which the only appearances of the until connective are as the formula  $U(A, B)$  and no appearance of that formula is in the scope of an  $S$ :*

1.  $S(p \wedge U(A, B), q)$ ,
2.  $S(p \wedge \neg U(A, B), q)$ ,
3.  $S(p, q \vee U(A, B))$ ,
4.  $S(p, q \vee \neg U(A, B))$ ,
5.  $S(p \wedge U(A, B), q \vee U(A, B))$ ,
6.  $S(p \wedge \neg U(A, B), q \vee U(A, B))$ ,
7.  $S(p \wedge U(A, B), q \vee \neg U(A, B))$ , and
8.  $S(p \wedge \neg U(A, B), q \vee \neg U(A, B))$ .

**Proof.** We prove the first case only; omitting the others. Note that  $S(p \wedge U(A, B), q)$  is equivalent to

$$\begin{aligned} & S(p, q) \wedge S(p, B) \wedge B \wedge U(A, B) \\ \vee & [A \wedge S(p, B) \wedge S(p, q)] \\ \vee & S(A \wedge q \wedge S(p, B) \wedge S(p, q), q). \end{aligned}$$

Indeed, the original formula holds at  $t$  iff there is  $s < t$  and  $u > s$  such that  $p$  holds at  $s$ ,  $A$  at  $u$ ,  $B$  everywhere between  $s$  and  $u$ , and  $q$  everywhere between  $s$  and  $t$ . The three disjuncts correspond to the cases  $u > t, u = t$ , and  $u < t$  respectively. Note that we make essential use of the linearity of time. ■

We now know the basic steps in our proof of separation. We simply keep pulling out  $U$ s from under the scopes of  $S$ s and vice versa until there are no more. Given a formula  $A$ , this process will eventually leave us with a *syntactically separated* formula, i.e. a formula  $B$  which is a Boolean combination of atoms, formulas  $U(E, F)$  with  $E$  and  $F$  built without using  $S$  and formulas  $S(E, F)$  with  $E$  and  $F$  built without using  $U$ . Clearly, such a  $B$  is separated.

We start dealing with more than one  $U$  inside an  $S$ . In this context, we call a formula in which  $U$  and  $S$  do not appear *pure*.

LEMMA 22. *Suppose that  $A$  and  $B$  are pure formulas and that  $C$  and  $D$  are such that any appearance of  $U$  is as  $U(A, B)$  and is not nested under any  $S$ s. Then  $S(C, D)$  is equivalent to a syntactically separated formula in which  $U$  only appears as the formula  $U(A, B)$ .*

**Proof.** If  $U(A, B)$  does not appear then we are done. Otherwise, by rearrangement of  $C$  and  $D$  into disjunctive and conjunctive normal form, respectively, and repeated use of Lemma 20 we can rewrite  $S(C, D)$  equivalently as a Boolean combination of formulas  $S(C_i, D_i)$  with no  $U$  appearing. Then the preceding lemma shows that each such Boolean constituent is equivalent to a Boolean combination of separated formulas. Thus we have a separated equivalent. ■

Next let us begin the inductive process of removing  $U$ s from more than one  $S$ . We present the separation in a crescendo. Each step introduces extra complexity in the formula being separated and uses the previous case as a starting point.

LEMMA 23. *Suppose that  $A, B$ , possibly subscripted, are pure formulas. Suppose  $C, D$ , possibly subscripted, contain no  $S$ . Then  $E$  has a syntactically separated equivalent if:*

- *the only appearance of  $U$  in  $E$  is as  $U(A, B)$ ;*
- *the only appearances of  $U$  in  $E$  are as  $U(A_i, B_i)$ ;*
- *the only appearances of  $U$  in  $E$  are as  $U(C_i, D_i)$ ;*
- *$E$  is any  $U, S$  formula.*

We omit the proof, referring to [Gabbay *et al.*, 1994, Chapter 10] for a detailed account. But note that since each case above uses the previous one as an induction basis, this process of separation tends to be highly exponential. Indeed, the separated version of a formula can be many times larger than the initial one. We finally have the main results.

THEOREM 24 (separation theorem). *Over the integer flow of time, any formula in the  $\{U, S\}$ -language is equivalent to a separated formula.*

**Proof.** This follows directly from the preceding lemma because, as we have already noted, syntactic separation implies separation. ■

THEOREM 25. *The language  $\{U, S\}$  is expressively complete over integer time.*

**Proof.** This follows from the separation theorem and Theorem 16. ■

Other known separation and expressive completeness results over linear time are [Gabbay *et al.*, 1994]:

- The language  $\{U, S\}$  is separable over real time. Indeed, it is separable over any Dedekind complete linear flow of time. As a consequence, it is also expressively complete over such flows.
- The language  $\{U, S\}$  is *not* separable over the rationals; as a result, it is not separable over the class of linear flows of time, nor is it expressively complete over such flows.

The problem of  $\{U, S\}$  over generic linear flows of time is that they may contain *gaps*. It is possible to define a first order formula that makes a proposition true up until a gap and false afterwards. Such formula, however, cannot be expressed in terms of  $\{U, S\}$ . So is there an extra set of connectives that is expressively complete over the rationals? The answer in this case is yes, and they are called the Stavi connectives. These are connectives whose truth value depends on the existence of gaps in the flow of time, and therefore are always false over integers or reals. For a detailed discussion on separation in the presence of gaps, please refer to [Gabbay *et al.*, 1994, Chapters 11 and 12].

We remain with the following generic question: given a flow of time, can we find a set of connectives that is expressively complete over it? This is the question that we investigate next.

### 3.3 *H-dimension*

The notion of *Henkin-* or H-dimension involves limiting the number of bound variables employed in first-order formulas. We will see that a *necessary* condition for there to exist a finite set of connectives which is expressively complete over a flow of time is that such flow of time have a finite H-dimension.

As for a *sufficient* condition for a finite expressively complete set of connectives, we will see that if *many-dimensional connectives* are allowed, than finite H-dimension implies the existence of such finite set of connectives. However, when we consider *one-dimensional connectives* such as Since and Until, finite H-dimension is no longer a sufficient condition.

In fact our approach in this discussion will be based on a *weak many-dimensional logic*. It is *many dimensional* because the truth value of a formula is evaluated at more than one time-point. It is *weak* because atomic formulas are evaluated only at a single time point (called the *evaluation point*), while all the other points are the *reference points*). Such weak many dimensionality allows us to define the truth table of many dimensional systems as formulas in the monadic first-order language, as opposed to a full  $m$ -dimensional system (in which atoms are evaluated at  $m$  time points) which would require an  $m$ -adic language.

An  *$m$ -dimensional table for an  $n$ -place connective* is a formula of the form  $\chi(x_1, \dots, x_m; R_1, \dots, R_n)$ , where  $\chi$  is a formula of the first-order predicate



language, written with symbols from  $\{<\} \cup \{R_1, \dots, R_n\}$ , where  $R_1, \dots, R_n$  are special  $m$ -place relation symbols. Without loss of expressivity, we will further assume that each term  $y_j$  occurring in  $R_i(y_1, \dots, y_m)$  is always a variable.

Fix a temporal system  $\mathcal{T}$  whose language contains atoms  $q_1, q_2, \dots$ , the classical connectives, and the special symbols  $\#_1, \dots, \#_j$ , standing for  $n_1, \dots, n_j$ -place connectives respectively. Let  $\chi_1, \dots, \chi_j$  be their  $m$ -dimensional  $n_1, \dots, n_j$ -place tables respectively.

**REMARK 26.** Since there are finitely many  $\chi_i$  to consider, we can further assume that there is  $b \geq m$  such that each  $\chi_i$  is written with variables  $x_1, \dots, x_b$  only.

The semantics of  $m$ -dimensional formulas is given by:

**DEFINITION 27.** Let  $(T, <)$  be a flow of time. Let  $h$  be an assignment into  $T$ , i.e. for any atom  $q$ ,  $h(q) \subseteq T$ . We define the truth value of each formula  $A$  of the language of  $\mathcal{T}$  at  $m$  indices  $a_1, \dots, a_{m-1}, t \in T$  under  $h$ , as follows:

1.  $(T, <, h), a_1, \dots, a_{m-1}, t \models q$  iff  $t \in h(q)$ ,  $q$  atomic.
2.  $(T, <, h), a_1, \dots, a_{m-1}, t \models A \wedge B$  iff  $(T, <, h), a_1, \dots, a_{m-1}, t \models A$  and  $(T, <, h), a_1, \dots, a_{m-1}, t \models B$ .
3.  $(T, <, h), a_1, \dots, a_{m-1}, t \models \neg A$  iff  $(T, <, h), a_1, \dots, a_{m-1}, t \not\models A$ .
4. For each  $i$  ( $1 \leq i \leq j$ ),  $(T, <, h), a_1, \dots, a_{m-1}, t \models \#_i(A_1, \dots, A_{n_i})$  iff  $T \models \chi_i(a_1, \dots, a_{m-1}, t, h(A_1), \dots, h(A_{n_i}))$ , where

$$h(A_k) =_{\text{def.}} \{(t_1, \dots, t_m) \in T^m \mid (T, <, h), t_1, \dots, t_m \models A_k\}.$$

Let  $L^M$  denote the monadic language with  $<$ , first-order quantifiers over elements, and an arbitrary number of monadic predicate symbols  $Q_i$  for subsets of  $T$ . We will regard the  $Q_i$  as predicate (subset) variables, implicitly associated with the atoms  $q_i$ . We define the translation of an  $m$ -dimensional temporal formula  $A$  into a monadic formula  $\delta A$ :

1. If  $A$  is an atom  $q_i$ , we set  $\delta A = (x_1 = x_1) \wedge \dots \wedge (x_{m-1} = x_{m-1}) \wedge Q_i(x_m)$ .
2.  $\delta(A \wedge B) = \delta A \wedge \delta B$ , and  $\delta(\neg A) = \neg \delta A$ .
3. Let  $A = \#_i(A_1, \dots, A_{n_i})$ , where  $\chi_i(x_1, \dots, x_m; R_1, \dots, R_{n_i})$  is the table of  $\#_i$ . Since we can always rewrite  $\chi$  such that all occurrences of  $R_k(y_1, \dots, y_m)$  in  $\chi$  are such that the terms  $y_i$  are variables, after a

suitable variable replacement we can write  $\delta A$  using only the variables  $x_1, \dots, x_b$  as:

$$\delta A = \chi_i(x_1, \dots, x_m, \delta A_1, \dots, \delta A_{n_i}).$$

Clearly, a simple induction gives us that:

$$(T, <, h), a_1, \dots, a_m \models B \text{ iff } T \models \delta B(a_1, \dots, a_m, h(q_1), \dots, h(q_k)).$$

such that  $\delta B(a_1, \dots, a_m, h(q_1), \dots, h(q_k))$  uses only the variables  $x_1, \dots, x_b$ .

Suppose that  $\mathcal{K}$  is a class of flows of time,  $\bar{x} = x_1, \dots, x_m$  are variables, and  $\bar{Q} = Q_1, \dots, Q_r$  are monadic predicates. If  $\alpha(\bar{x}, \bar{Q}), \beta(\bar{x}, \bar{Q})$  are formulas in  $L^M$  with free variables  $\bar{x}$  and free monadic predicates  $\bar{Q}$ , we say that  $\alpha$  and  $\beta$  are  $\mathcal{K}$ -equivalent if for all  $T \in \mathcal{K}$  and all subsets  $S_1, \dots, S_r \subseteq T$ ,

$$T \models \forall \bar{x} \left( \alpha(\bar{x}, S_1, \dots, S_r) \leftrightarrow \beta(\bar{x}, S_1, \dots, S_r) \right).$$

We say the temporal system  $\mathcal{T}$  is *expressively complete over  $\mathcal{K}$  in  $n$  dimensions* ( $1 \leq n \leq m$ ) if for any  $\alpha(x_1, \dots, x_n, \bar{Q})$  of  $L^M$  with free variables  $x_1, \dots, x_n$ , there exists a temporal formula  $B(\bar{q})$  of  $\mathcal{T}$  built up from the atoms  $\bar{q} = q_1, \dots, q_r$ , such that  $\alpha \wedge \bigwedge_{n < i \leq m} x_i = x_i$  and  $\delta B$  are equivalent in  $\mathcal{K}$ . In this case,  $\mathcal{K}$  is said to be *m-functionally complete* in  $n$  dimensions (symbolically,  $FC_n^m$ );  $\mathcal{K}$  is *functionally complete* if it is  $FC_1^m$  for some  $m$ .

Finally, we define the *Henkin or H-dimension*  $d$  of a class  $\mathcal{K}$  of flows as the smallest  $d$  such that:

- For any monadic formula  $\alpha(x_1, \dots, x_n, Q_1, \dots, Q_r)$  in  $L^M$  with free variables among  $x_1, \dots, x_n$  and monadic predicates  $Q_1, \dots, Q_r$  (with  $n, r$  arbitrary), there exists an  $L^M$ -formula  $\alpha'(x_1, \dots, x_n, Q_1, \dots, Q_r)$  that is  $\mathcal{K}$ -equivalent to  $\alpha$  and uses no more than  $d$  different bound variable letters.

We now show that for any class of flows, finite Henkin dimension is equivalent to functional completeness ( $FC_1^m$  for some  $m$ ).

**THEOREM 28.** *For any class  $\mathcal{K}$  of flows of time, if  $\mathcal{K}$  is functionally complete then  $\mathcal{K}$  has finite H-dimension.*

**Proof.** Let  $\sigma(\bar{Q})$  be any sentence of  $L^M$ . By functional completeness, there exists a  $B(\bar{q})$  of  $\mathcal{T}$  such that the formulas  $x_1 = x_1 \wedge \dots \wedge x_m = x_m \wedge \sigma(\bar{Q})$  and  $\delta B(x_1, \dots, x_m, \bar{Q})$  are  $\mathcal{K}$ -equivalent. We know that  $\delta B$  is written using variables  $x_1, \dots, x_b$  only. Hence the sentence  $\sigma^* = \exists x_1 \dots \exists x_m \delta B(x_1, \dots, x_m, \bar{Q})$  has at most  $b$  variables, and is clearly  $\mathcal{K}$ -equivalent to  $\sigma$ . So every sentence of  $L^M$  is  $\mathcal{K}$ -equivalent to one with at most  $b$  variables. This means that  $\mathcal{K}$  has H-dimension at most  $b$ , so it is finite.  $\blacksquare$

We now show the converse. That is, we assume that the class  $\mathcal{K}$  of flows of time has finite H-dimension  $m$ . Then we are going to construct a temporal logic that is expressively complete over  $\mathcal{K}$  and that is weakly  $m+1$ -dimensional (and that is why such proof does not work for 1-dimensional systems: it always constructs a logic of dimension at least 2).

Let us call this logic system  $\mathbf{d}$ . Besides atomic propositions  $q_1, q_2, \dots$  and the usual Boolean operators, this system has a set of constants (0-place operators)  $C_{i,j}^<$  and  $C_{i,j}^=$  and unary temporal connectives  $\Pi_i$  and  $\Box_i$ , for  $0 \leq i, j \leq m$ . If  $h$  is an assignment such that  $h(q) \subseteq T$  for atomic  $q$ , the semantics of  $\mathbf{d}$ -formulas is given by:

1.  $(T, <, h), x_0, \dots, x_m \models q$  iff  $x_0 \in h(q)$  for  $q$  atomic.
2. The tables for  $\neg, \wedge$  are the usual ones.
3.  $(T, <, h), x_0, \dots, x_m \models C_{i,j}^<$  iff  $x_i < x_j$ . Similarly we define the semantics of  $C_{i,j}^=$ .  $C_{i,j}^<, C_{i,j}^=$  are thus called *diagonal constants*.
4.  $(T, <, h), x_0, \dots, x_m \models \Pi_i A$  iff  $(T, <, h), x_i, \dots, x_i \models A$ . So  $\Pi_i$  “projects” the truth value on the  $i$ -th dimension.
5.  $(T, <, h), x_0, \dots, x_m \models \Box_i A$  iff  $(T, <, h), x_0, \dots, x_{i-1}, y, x_{i+1}, \dots, x_m \models A$  for all  $y \in T$ . So  $\Box_i$  is an “always” operator for the  $i$ -th dimension.

**LEMMA 29.** *Let  $\beta$  be a formula of  $L^M$  written only using the variable letters  $u_0, \dots, u_m$ , and having  $u_{i_1}, \dots, u_{i_k}$  free for arbitrary  $k \leq m$ . Then there exists a temporal formula  $A$  of  $\mathbf{d}$  such that for all  $h, t_0, \dots, t_m \in T$ ,*

$$(T, <, h), t_0, \dots, t_m \models A \text{ iff } \mathcal{K}, h \models \beta(t_{i_1}, \dots, t_{i_k}).$$

**Proof.** By induction on  $\beta$ . Assume first that  $\beta$  is atomic. If  $\beta$  is  $u_i < u_j$  let  $A = C_{i,j}^<$  if  $i \neq j$ , and  $\perp$  otherwise. Similarly for  $u_i = u_j$ . If  $\beta$  is  $Q(u_i)$ , let  $A$  be  $\Pi_i(q)$ .

The classical connectives present no difficulties. We turn to the case where  $\beta$  is  $\forall u_i \alpha(u_{i_1}, \dots, u_{i_k})$ . By induction hypothesis, let  $A$  be the formula corresponding  $\alpha$ ; then  $\Box_i A$  is the formula suitable for  $\beta$ . ■

We are now in a position to prove the converse of Theorem 28.

**THEOREM 30.** *For any class  $\mathcal{K}$  of flows of time, if  $\mathcal{K}$  has finite H-dimension then  $\mathcal{K}$  is functionally complete.*

**Proof.** Let  $\beta(u_0)$  be any formula of  $L^M$  with one free variable  $u_0$ . As  $\mathcal{K}$  has H-dimension  $m$ , we can suppose that  $\beta$  is written with variables  $u_0, \dots, u_m$ . By Lemma 29 there exists an  $A$  of  $\mathcal{T}$  such that for any  $T \in \mathcal{K}$ ,  $t \in T$ , and assignment  $h$  into  $T$ ,  $(T, <, h), t, \dots, t \models A$  iff  $\mathcal{K}, h \models \beta(t)$ . ■

As an application of the results above, we show that the class of partial orders is not functionally complete. For consider the formula corresponding to the statement *there are at least  $n$  elements in the order*:

$$\sigma_n = \exists x_1, \dots, x_n \bigwedge_{i \neq j} [(x_i \neq x_j) \wedge \neg(x_i < x_j)].$$

It can be shown that such formula cannot be written with less than  $n$  variables (e.g. [Gabbay *et al.*, 1994]). Since we are able to say that there are at least  $n$  elements in the order for any finite  $n$ , the class partial orders have infinite H-dimension and by Theorem 28 it is not functionally complete.

On the other hand, the reals and the integers must have finite H-dimension, for the  $\{U, S\}$  temporal logic is expressively complete over both. Indeed, [Gabbay *et al.*, 1994] shows that it has H-dimension at most 3, and so does the theory of linear order.

#### 4 COMBINING TEMPORAL LOGICS

There is a profusion of logics proposed in the literature for the modelling of a variety of phenomena, and many more will surely be proposed in the future. A great part of those logics deal only with “static” aspects, and the temporal evolution is left out. But eventually, the need to deal with the temporal evolution of a model appears. What we want to avoid is the so called *reinvention of the wheel*, that is, reworking from scratch the whole logic, its language, inference system and models, and reproving all its basic properties, when the temporal dimension is added.

We therefore show here several methods for combining logic systems and we study if the properties of the component systems are *transferred* to their combination. We understand a logic system  $\mathcal{L}_L$  as composed of three elements:

- (a) a language  $\mathcal{L}_L$ , normally given by a set of formation rules generating well formed formulas over a signature and a set of logical connectives.
- (b) An inference system, i.e. a relation,  $\vdash_L$ , between sets of formulas, represented by  $\Delta \vdash_L A$ . As usual,  $\vdash_L A$  indicates that  $\emptyset \vdash_L A$ .
- (c) The semantics of formulas over a class  $\mathcal{K}$  of model structures. The fact that a formulas  $A$  is true of or holds at a model  $\mathcal{M} \in \mathcal{K}$  is indicated by  $\mathcal{M} \models A$ .

Each method for combining logic systems proposes a way of generating the language, inference system and model structures from those of the component system.

The first method presented here adds a temporal dimension  $\mathbb{T}$  to a logic system  $L$ , called the *temporalisation* of a logic system  $\mathbb{T}(L)$ , with an automatic way of constructing:

- the language of  $\mathbb{T}(\mathbb{L})$ ;
- the inference system of  $\mathbb{T}(\mathbb{L})$ ; and
- the class of temporal models of  $\mathbb{T}(\mathbb{L})$ .

We do that in a way that the basic properties of soundness, completeness and decidability are *transferred* from the component logics  $\mathbb{T}$  and  $\mathbb{L}$  to the combined system  $\mathbb{T}(\mathbb{L})$ .

If the temporalised logic is itself a temporal logic, we have a two dimensional temporal logic  $\mathbb{T}(\mathbb{T}')$ . Such a logic is too weak, however, because, by construction, the temporal logic  $\mathbb{T}'$  cannot refer to the the logic system  $\mathbb{T}$ . We therefore present the *independent combination*  $\mathbb{T} \oplus \mathbb{T}'$  in which two temporal logics are symmetrically combined. As before, the language, inference systems and models of  $\mathbb{T} \oplus \mathbb{T}'$ , and show that the properties of soundness, completeness and decidability are transferred form  $\mathbb{T}$  and  $\mathbb{T}'$  to  $\mathbb{T} \oplus \mathbb{T}'$ .

The independent combination is not the strongest way to combine logics; in particular, the independent combination of two linear temporal logic does not necessarily produce a two-dimensional grid model. So we show how to produce the *full join* of two linear temporal logics  $\mathbb{T} \times \mathbb{T}'$ , such that all models will be two-dimensional grids. However, in this case we cannot guarantee that the basic properties of  $\mathbb{T}$  and  $\mathbb{T}'$  are transferred to  $\mathbb{T} \times \mathbb{T}'$ . In this sense, the independent combination  $\mathbb{T} \oplus \mathbb{T}'$  is a *minimal symmetrical combination* of logics that automatically transfers the basic properties. Any further interaction between the logics has to be separately investigated.

As a final way of combining logics, we present methods of combination that are motivated by the study of Labelled Deductive Systems (LDS) [Gabbay, 1996].

All temporal logics considered for combination here are assumed to be linear.

#### 4.1 Temporalising a Logic

The first of the combination methods, known as “adding a temporal dimension to a logic system” or simply “temporalising a logic system”, has been initially presented in [Finger and Gabbay, 1992].

Temporalisation is a methodology whereby an arbitrary logic system  $\mathbb{L}$  can be enriched with temporal features from a linear temporal logic  $\mathbb{T}$  to create a new, *temporalised* system  $\mathbb{T}(\mathbb{L})$ .

We assume that the language of temporal system  $\mathbb{T}$  is the *US* language and its inference system is an extensions of that of *US*/ $\mathcal{K}_{lin}$ , with its corresponding class of temporal linear models  $\mathcal{K} \subseteq \mathcal{K}_{lin}$ .

With respect to the logic  $\mathbb{L}$  we assume it is an extension of classical logic, that is, all propositional tautologies are valid in it. The set  $\mathcal{L}_{\mathbb{L}}$  is partitioned in two sets,  $BC_{\mathbb{L}}$  and  $ML_{\mathbb{L}}$ . A formula  $A \in \mathcal{L}_{\mathbb{L}}$  belongs to the set of *Boolean*

*combinations*,  $BC_L$ , iff it is built up from other formulas by the use of one of the Boolean connectives  $\neg$  or  $\wedge$  or any other connective defined only in terms of those; it belongs to the set of *monolithic formula*  $ML_L$  otherwise.

If  $L$  is not an extension of classical logic, we can simply “encapsulate” it in  $L'$  with a one-place symbol  $\#$  not occurring in either  $L$  or  $T$ , such that for each formula  $A \in \mathcal{L}_L$ ,  $\#A \in \mathcal{L}_{L'}$ ,  $\vdash_L A$  iff  $\vdash_{L'} \#A$  and the model structures of  $\#A$  are those of  $A$ . Note that  $ML_{L'} = \mathcal{L}_{L'}$ ,  $BC_{L'} = \emptyset$ .

The alphabet of the temporalised language uses the alphabet of  $L$  plus the two-place operators  $S$  and  $U$ , if they are not part of the alphabet of  $L$ ; otherwise, we use  $\bar{S}$  and  $\bar{U}$  or any other proper renaming.

**DEFINITION 31.** Temporalised formulas The set  $\mathcal{L}_{T(L)}$  of formulas of the logic system  $L$  is the smallest set such that:

1. If  $A \in ML_L$ , then  $A \in \mathcal{L}_{T(L)}$ ;
2. If  $A, B \in \mathcal{L}_{T(L)}$  then  $\neg A \in \mathcal{L}_{T(L)}$  and  $(A \wedge B) \in \mathcal{L}_{T(L)}$ ;
3. If  $A, B \in \mathcal{L}_{T(L)}$  then  $S(A, B) \in \mathcal{L}_{T(L)}$  and  $U(A, B) \in \mathcal{L}_{T(L)}$ .

Note that, for instance, if  $\Box$  is an operator of the alphabet of  $L$  and  $A$  and  $B$  are two formulas in  $\mathcal{L}_L$ , the formula  $\Box U(A, B)$  is *not* in  $\mathcal{L}_{T(L)}$ . The language of  $T(L)$  is independent of the underlying flow of time, but not its semantics and inference system, so we must fix a class  $\mathcal{K}$  of flows of time over which the temporalisation is defined; if  $\mathcal{M}_L$  is a model in the class of models of  $L$ ,  $\mathcal{K}_L$ , for every formula  $A \in \mathcal{L}_L$  we must have either  $\mathcal{M}_L \models A$  or  $\mathcal{M}_L \models \neg A$ . In the case that  $L$  is a temporal logic we must consider a “current time”  $o$  as part of its model to achieve that condition.

**DEFINITION 32.** *Semantics of the temporalised logic.*<sup>1</sup> Let  $(T, <) \in \mathcal{K}$  be a flow of time and let  $g : T \rightarrow \mathcal{K}_L$  be a function mapping every time point in  $T$  to a model in the class of models of  $L$ . A model of  $T(L)$  is a triple  $\mathcal{M}_{T(L)} = (T, <, g)$  and the fact that  $A$  is true in  $\mathcal{M}_{T(L)}$  at time  $t$  is written as  $\mathcal{M}_{T(L)}, t \models A$  and defined as:

$$\begin{aligned} \mathcal{M}_{T(L)}, t \models A, A \in ML_L & \text{ iff } g(t) = \mathcal{M}_L \text{ and } \mathcal{M}_L \models A. \\ \mathcal{M}_{T(L)}, t \models \neg A & \text{ iff it is not the case that } \mathcal{M}_{T(L)}, t \models A. \\ \mathcal{M}_{T(L)}, t \models (A \wedge B) & \text{ iff } \mathcal{M}_{T(L)}, t \models A \text{ and } \mathcal{M}_{T(L)}, t \models B. \\ \mathcal{M}_{T(L)}, t \models S(A, B) & \text{ iff there exists } s \in T \text{ such that } s < t \text{ and} \\ & \mathcal{M}_{T(L)}, s \models A \text{ and for every } u \in T, \text{ if} \\ & s < u < t \text{ then } \mathcal{M}_{T(L)}, u \models B. \end{aligned}$$

<sup>1</sup>We assume that the a model of  $T$  is given by  $(T, <, h)$  where  $h$  maps time points into sets of propositions (instead of the more common, but equivalent, mapping of propositions into sets of time points); such notation highlights that in the temporalised model each time point is associated to a model of  $L$ .

$\mathcal{M}_{\mathbb{T}(\mathbb{L})}, t \models U(A, B)$       iff there exists  $s \in T$  such that  $t < s$  and  
 $\mathcal{M}_{\mathbb{T}(\mathbb{L})}, s \models A$  and for every  $u \in T$ , if  
 $t < u < s$  then  $\mathcal{M}_{\mathbb{T}(\mathbb{L})}, u \models B$ .

The inference system of  $\mathbb{T}(\mathbb{L})/\mathcal{K}$  is given by the following:

**DEFINITION 33.** Axiomatisation for  $\mathbb{T}(\mathbb{L})$  An axiomatisation for the temporalised logic  $\mathbb{T}(\mathbb{L})$  is composed of:

- The axioms of  $\mathbb{T}/\mathcal{K}$ ;
- The inference rules of  $\mathbb{T}/\mathcal{K}$ ;
- For every formula  $A$  in  $\mathcal{L}_{\mathbb{L}}$ , if  $\vdash_{\mathbb{L}} A$  then  $\vdash_{\mathbb{T}(\mathbb{L})} A$ , i.e. all theorems of  $\mathbb{L}$  are theorems of  $\mathbb{T}(\mathbb{L})$ . This inference rule is called **Persist**.

**EXAMPLE 34.** Consider classical propositional logic  $\mathbb{PL} = \langle \mathcal{L}_{\mathbb{PL}}, \vdash_{\mathbb{PL}}, \models_{\mathbb{PL}} \rangle$ . Its temporalisation generates the logic system  $\mathbb{T}(\mathbb{PL}) = \langle \mathcal{L}_{\mathbb{T}(\mathbb{PL})}, \vdash_{\mathbb{T}(\mathbb{PL})}, \models_{\mathbb{T}(\mathbb{PL})} \rangle$ . It is not difficult to see that the temporalised version of  $\mathbb{PL}$  over any  $\mathcal{K}$  is actually the temporal logic  $\mathbb{T} = \mathbb{US}/\mathcal{K}$ .

If we temporalise over  $\mathcal{K}$  the one-dimensional logic system  $\mathbb{US}/\mathcal{K}$  we obtain the two-dimensional logic system  $\mathbb{T}(\mathbb{US}) = \langle \mathcal{L}_{\mathbb{T}(\mathbb{US})}, \vdash_{\mathbb{T}(\mathbb{US})}, \models_{\mathbb{T}(\mathbb{US})} \rangle = \mathbb{T}^2(\mathbb{PL})/\mathcal{K}$ . In this case we have to rename the two-place operators  $S$  and  $U$  of the temporalised alphabet to, say,  $\bar{S}$  and  $\bar{U}$ . Note, however, how weak this logic is, for  $\bar{S}$  and  $\bar{U}$  cannot occur within the scope of  $U$  and  $S$ .

In order to obtain a model for  $\mathbb{T}(\mathbb{US})$ , we must fix a “current time”,  $o_1$ , in  $\mathcal{M}_{\mathbb{US}} = (T_1, <_1, g_1)$ , so that we can construct the model  $\mathcal{M}_{\mathbb{T}(\mathbb{US})} = (T_2, <_2, g_2)$  as previously described. Note that, in this case, the flows of time  $(T_1, <_1)$  and  $(T_2, <_2)$  need not to be the same.  $(T_2, <_2)$  is the flow of time of the upper-level temporal system whereas  $(T_1, <_1)$  is the flow of time of the underlying logic which, in this case, happens to be a temporal logic. The satisfiability of a formula in a model of  $\mathbb{T}(\mathbb{US})$  needs two evaluation points,  $o_1$  and  $o_2$ ; therefore it is a *two-dimensional temporal logic*.

The logic system we obtain by temporalising  $US$ -temporal logic is the two-dimensional temporal logic described in [Finger, 1992].

This temporalisation process can be repeated  $n$  times, generating an  $n$  dimensional temporal logic with connectives  $U_i, S_i$ ,  $1 \leq i \leq n$ , such that for  $i < j$   $U_j, S_j$  cannot occur within the scope of  $U_i, S_i$ .

We analyse now the transfer of soundness, completeness and decidability from  $\mathbb{T}$  and  $\mathbb{L}$  to  $\mathbb{T}(\mathbb{L})$ ; that is, we are assuming the logics  $\mathbb{T}$  and  $\mathbb{L}$  have sound, complete and decidable axiomatisations with respect to their semantics, and we will analyse how such properties transfer to the combined system  $\mathbb{T}(\mathbb{L})$ . It is a routine task to analyse that if the inference systems of  $\mathbb{T}$  and  $\mathbb{L}$  are sound, so is  $\mathbb{T}(\mathbb{L})$ . So we concentrate on the proof of transference of completeness.

### Completeness

We prove the completeness of  $\mathbb{T}(\mathbb{L})/\mathcal{K}$  indirectly by transforming a consistent formula  $A$  of  $\mathbb{T}(\mathbb{L})$  into  $\varepsilon(A)$  and then mapping it into a consistent formula of  $\mathbb{T}$ . Completeness of  $\mathbb{T}/\mathcal{K}$  is used to find a  $\mathbb{T}$ -model for  $A^*$  that is used to construct a model for the original  $\mathbb{T}(\mathbb{L})$  formula  $A$ .

We first define the transformation and mapping. Given a formula  $A \in \mathcal{L}_{\mathbb{T}(\mathbb{L})}$ , consider the following sets:

$$\begin{aligned} Lit(A) &= Mon(A) \cup \{\neg B \mid B \in Mon(A)\} \\ Inc(A) &= \{\bigwedge \Gamma \mid \Gamma \subseteq Lit(A) \text{ and } \Gamma \vdash_{\mathbb{L}} \perp\} \end{aligned}$$

where  $Mon(A)$  is the set of maximal monolithic subformulae of  $A$ .  $Lit(A)$  is the set of literals occurring in  $A$  and  $Inc(A)$  is the set of inconsistent formulas that can be built with those. We transform  $A$  into  $A$  as:  $\varepsilon(A)$ :

$$\varepsilon(A) = A \wedge \bigwedge_{B \in Inc(A)} (\neg B \wedge G\neg B \wedge H\neg B)$$

The big conjunction  $in\varepsilon(A)$  is a theorem of  $\mathbb{T}(\mathbb{L})$ , so we have the following lemma.

LEMMA 35.  $\vdash_{\mathbb{T}(\mathbb{L})} \varepsilon(A) \leftrightarrow A$

If  $\mathbb{K}$  is a subclass of linear flows of time, we also have the following property (this is where linearity is used in the proof).

LEMMA 36. *Let  $\mathcal{M}_{\mathbb{T}}$  be a temporal model over  $\mathcal{K} \subseteq \mathcal{K}_{lin}$  such that for some  $o \in T$ ,  $\mathcal{M}_{\mathbb{T}}, o \models \sigma(\Box A)$ . Then, for every  $t \in T$ ,  $\mathcal{M}_{\mathbb{T}}, t \models \sigma(\Box A)$ .*

Therefore, if some subset of  $Lit(A)$  is inconsistent, the transformed formula  $\varepsilon(A)$  puts that fact in evidence so that, when it is mapped into  $\mathbb{T}$ , inconsistent subformulae will be mapped into falsity.

Now we want to map a  $\mathbb{T}(\mathbb{L})$ -formula into a  $\mathbb{T}$ -formula. For that, consider an enumeration  $p_1, p_2, \dots$ , of elements of  $\mathcal{P}$  and consider an enumeration  $A_1, A_2, \dots$ , of formulae in  $ML_{\mathbb{L}}$ . The *correspondence mapping*  $\sigma : \mathcal{L}_{\mathbb{T}(\mathbb{L})} \rightarrow \mathcal{L}_{\mathbb{T}}$  is given by:

$$\begin{aligned} \sigma(A_i) &= p_i \text{ for every } A_i \in ML_{\mathbb{L}}, i = 1, 2, \dots \\ \sigma(\neg A) &= \neg\sigma(A) \\ \sigma(A \wedge B) &= \sigma(A) \wedge \sigma(B) \\ \sigma(S(A, B)) &= S(\sigma(A), \sigma(B)) \\ \sigma(U(A, B)) &= U(\sigma(A), \sigma(B)) \end{aligned}$$

The following is the *correspondence lemma*.

LEMMA 37. *The correspondence mapping is a bijection. Furthermore if  $A$  is  $\mathbb{T}(\mathbb{L})$ -consistent then  $\sigma(A)$  is  $\mathbb{T}$ -consistent.*



LEMMA 38. *If  $A$  is  $\mathsf{T}(\mathsf{L})$ -consistent, then for every  $t \in T$ ,  $G_A(t) = \{B \in \mathit{Lit}(A) \mid \mathcal{M}_\mathsf{T}, t \models \sigma(B)\}$  is finite and  $\mathsf{L}$ -consistent.*

**Proof.** Since  $\mathit{Lit}(A)$  is finite,  $G_A(t)$  is finite for every  $t$ . Suppose  $G_A(t)$  is inconsistent for some  $t$ , then there exist  $\{B_1, \dots, B_n\} \subseteq G_A(t)$  such that  $\vdash_{\mathsf{L}} \bigwedge B_i \rightarrow \perp$ . So  $\bigwedge B_i \in \mathit{Inc}(A)$  and  $\Box \neg(\bigwedge B_i)$  is one of the conjuncts of  $\varepsilon(A)$ . Applying Lemma 36 to  $\mathcal{M}_\mathsf{T}, o \models \sigma(\varepsilon(A))$  we get that for every  $t \in T$ ,  $\mathcal{M}_\mathsf{T}, t \models \neg(\bigwedge \sigma(B_i))$  but by, the definition of  $G_A$ ,  $\mathcal{M}_\mathsf{T}, t \models \bigwedge \sigma(B_i)$ , which is a contradiction. ■

We are finally ready to prove the completeness of  $\mathsf{T}(\mathsf{L})/\mathcal{K}$ .

THEOREM 39 (Completeness transfer for  $\mathsf{T}(\mathsf{L})$ ). *If the logical system  $\mathsf{L}$  is complete and  $\mathsf{T}$  is complete over a subclass of linear flows of time  $\mathcal{K} \subseteq \mathcal{K}_{lin}$ , then the logical system  $\mathsf{T}(\mathsf{L})$  is complete over  $\mathcal{K}$ .*

**Proof.** Assume that  $A$  is  $\mathsf{T}(\mathsf{L})$ -consistent. By Lemma 38, we have  $(T, <) \in \mathcal{K}$  and associated to every time point in  $T$  we have a finite and  $\mathsf{L}$ -consistent set  $G_A(t)$ . By (weak) completeness of  $\mathsf{L}$ , every  $G_A(t)$  has a model, so we define the temporalised valuation function  $g$ :

$$g(t) = \{\mathcal{M}_\mathsf{L}^t \mid \mathcal{M}_\mathsf{L}^t \text{ is a model of } G_A(t)\}$$

Consider the model  $\mathcal{M}_{\mathsf{T}(\mathsf{L})} = (T, <, g)$  over  $\mathsf{K}$ . By structural induction over  $B$ , we show that for every  $B$  that is a subformula of  $A$  and for every time point  $t$ ,

$$\mathcal{M}_\mathsf{T}, t \models \sigma(B) \text{ iff } \mathcal{M}_{\mathsf{T}(\mathsf{L})}, t \models B$$

We show only the basic case,  $B \in \mathit{Mon}(A)$ . Suppose  $\mathcal{M}_\mathsf{T}, t \models \sigma(B)$ ; then  $B \in G_A(t)$  and  $\mathcal{M}_\mathsf{L}^t \models B$ , and hence  $\mathcal{M}_{\mathsf{T}(\mathsf{L})}, t \models B$ . Suppose  $\mathcal{M}_{\mathsf{T}(\mathsf{L})}, t \models B$  and assume  $\mathcal{M}_\mathsf{T}, t \models \neg\sigma(B)$ ; then  $\neg B \in G_A(t)$  and  $\mathcal{M}_\mathsf{L}^t \models \neg B$ , which contradicts  $\mathcal{M}_{\mathsf{T}(\mathsf{L})}, t \models B$ ; hence  $\mathcal{M}_\mathsf{T}, t \models \sigma(B)$ . The inductive cases are straightforward and omitted.

So,  $\mathcal{M}_{\mathsf{T}(\mathsf{L})}$  is a model for  $A$  over  $\mathsf{K}$  and the proof is finished. ■

Theorem 39 gives us sound and complete axiomatisations for  $\mathsf{T}(\mathsf{L})$  over many interesting classes of flows of time, such as the class of all linear flows of time,  $\mathcal{K}_{lin}$ , the integers,  $\mathbb{Z}$ , and the reals,  $\mathbb{R}$ . These classes are, in their  $\mathsf{T}$  versions, decidable and the corresponding decidability of  $\mathsf{T}(\mathsf{L})$  is dealt next.

Note that the construction above is finitistic, and therefore does not itself guarantee that compactness is transferred. However, an important corollary of the construction above is that the temporalised system is a *conservative extension* of both original systems, that is, no new theorem in the language of an original system is provable in the combined system. Formally,  $\mathsf{L}_1$  is a *conservative extension* of  $\mathsf{L}_2$  if it is an extension of  $\mathsf{L}_2$  such that if  $A \in \mathcal{L}_{\mathsf{L}_2}$ , then  $\vdash_{\mathsf{L}_1} A$  only if  $\vdash_{\mathsf{L}_2} A$ .

**COROLLARY 40.** *Let  $\mathsf{L}$  be a sound and complete logic system and  $\mathsf{T}$  be sound and complete over  $\mathcal{K} \subseteq \mathcal{K}_{lin}$ . The logic system  $\mathsf{T}(\mathsf{L})$  is a conservative extension of both  $\mathsf{L}$  and  $\mathsf{T}$ .*

**Proof.** Let  $A \in \mathcal{L}_{\mathsf{L}}$  such that  $\vdash_{\mathsf{T}(\mathsf{L})} A$ . Suppose by contradiction that  $\not\vdash_{\mathsf{L}} A$ , so by completeness of  $\mathsf{L}$ , there exists a model  $\mathcal{M}_{\mathsf{L}}$  such that  $\mathcal{M}_{\mathsf{L}} \models \neg A$ . We construct a temporalised model  $\mathcal{M}_{\mathsf{T}(\mathsf{L})} = (T, <, g)$  by making  $g(t) = \mathcal{M}_{\mathsf{L}}$  for all  $t \in T$ .  $\mathcal{M}_{\mathsf{T}(\mathsf{L})}$  clearly contradicts the soundness of  $\mathsf{T}(\mathsf{L})$  and therefore that of  $\mathsf{T}$ , so  $\vdash_{\mathsf{L}} A$ . This shows that  $\mathsf{T}(\mathsf{L})$  is a conservative extension of  $\mathsf{L}$ ; the proof of extension of  $\mathsf{T}$  is similar. ■

### *Decidability*

The transfer of decidability is also done using the correspondence mapping  $\sigma$  and the transformation  $\eta$ . Such a transformation is actually computable, as the following two lemmas state.

**LEMMA 41.** *For any  $A \in \mathcal{L}_{\mathsf{T}(\mathsf{L})}$ , if the logic system  $\mathsf{L}$  is decidable then there exists an algorithm for constructing  $\varepsilon(A)$ .*

**LEMMA 42.** *Over a linear flow of time, for every  $A \in \mathcal{L}_{\mathsf{T}(\mathsf{L})}$ ,*

$$\vdash_{\mathsf{T}(\mathsf{L})} A \text{ iff } \vdash_{\mathsf{T}} \sigma(\varepsilon(A)).$$

Decidability is a direct consequence of these two lemmas.

**THEOREM 43.** *If  $\mathsf{L}$  is a decidable logic system, and  $\mathsf{T}$  is decidable over  $\mathcal{K} \subseteq \mathcal{K}_{lin}$ , then the logic system  $\mathsf{T}(\mathsf{L})$  is also decidable over  $\mathcal{K}$ .*

**Proof.** Consider  $A \in \mathcal{L}_{\mathsf{T}(\mathsf{L})}$ . Since  $\mathsf{L}$  is decidable, by Lemma 41 there is an algorithmic procedure to build  $\varepsilon(A)$ . Since  $\sigma$  is a recursive function, we have an algorithm to construct  $\sigma(\varepsilon(A))$ , and due to the decidability of  $\mathsf{T}$  over  $\mathcal{K}$ , we have an effective procedure to decide if it is a theorem or not. Since  $\mathcal{K}$  is linear, by Lemma 42 this is also a procedure for deciding whether  $A$  is a theorem or not. ■

## *4.2 Independent Combination*

We now deal with the combination of two temporal logic systems. One of them will be called the *horizontal* temporal logic  $\mathsf{US}$ , while the other will be the *vertical* temporal logic  $\bar{\mathsf{U}}\bar{\mathsf{S}}$ . If we temporalise the horizontal logic with the vertical logic, we obtain a very weakly expressive system; if  $\mathsf{US}$  is the internal (horizontal) temporal logic in the temporalisation process ( $F$  is derived in  $\mathsf{US}$ ), and  $\bar{\mathsf{U}}\bar{\mathsf{S}}$  is the external (vertical) one ( $\bar{F}$  is defined in  $\bar{\mathsf{U}}\bar{\mathsf{S}}$ ), we cannot express that vertical and horizontal future operators commute,

$$F\bar{F}A \leftrightarrow \bar{F}FA.$$

In fact, the subformula  $F\bar{F}A$  is not even in the temporalised language of  $\bar{U}\bar{S}(US)$ , nor is the whole formula. In other words, the interplay between the two-dimensions is not expressible in the language of the temporalised  $\bar{U}\bar{S}(US)$ .

The idea is then to define a method for combining temporal logics that is symmetrical. As usual, we combine the languages, inference systems and classes of models.

**DEFINITION 44.** Let  $Op(L)$  be the set of non-Boolean operators of a generic logic  $L$ . Let  $\bar{T}$  and  $T$  be logic systems such that  $Op(T) \cap Op(\bar{T}) = \emptyset$ . The *fully combined language* of logic systems  $\bar{T}$  and  $T$  over the set of atomic propositions  $\mathcal{P}$ , is obtained by the union of the respective set of connectives and the union of the formation rules of the languages of both logic systems.

Let the operators  $U$  and  $S$  be in the language of  $US$  and  $\bar{U}$  and  $\bar{S}$  be in that of  $\bar{U}\bar{S}$ . Their fully combined language over a set of atomic propositions  $\mathcal{P}$  is given by

- every atomic proposition is in it;
- if  $A, B$  are in it, so are  $\neg A$  and  $A \wedge B$ ;
- if  $A, B$  are in it, so are  $U(A, B)$  and  $S(A, B)$ .
- if  $A, B$  are in it, so are  $\bar{U}(A, B)$  and  $\bar{S}(A, B)$ .

Not only are the two languages taken to be independent of each other, but the set of axioms of the two systems are supposed to be disjoint; so we call the following combination method the *independent combination* of two temporal logics.

**DEFINITION 45.** Let  $US$  and  $\bar{U}\bar{S}$  be two  $US$ -temporal logic systems defined over the same set  $\mathcal{P}$  of propositional atoms such that their languages are independent. The *independent combination*  $US \oplus \bar{U}\bar{S}$  is given by the following:

- The fully combined language of  $US$  and  $\bar{U}\bar{S}$ .
- If  $(\Sigma, \mathcal{I})$  is an axiomatisation for  $US$  and  $(\bar{\Sigma}, \bar{\mathcal{I}})$  is an axiomatisation for  $\bar{U}\bar{S}$ , then  $(\Sigma \cup \bar{\Sigma}, \mathcal{I} \cup \bar{\mathcal{I}})$  is an axiomatisation for  $US \oplus \bar{U}\bar{S}$ . Note that, apart from the classical tautologies, the set of axioms  $\Sigma$  and  $\bar{\Sigma}$  are supposed to be disjoint, but not the inference rules.
- The class of independently combined flows of time is  $\mathcal{K} \oplus \bar{\mathcal{K}}$  composed of biordered flows of the form  $(\tilde{T}, <, \bar{<})$  where the connected components of  $(\tilde{T}, <)$  are in  $\mathcal{K}$  and the connected components of  $(\tilde{T}, \bar{<})$  are in  $\bar{\mathcal{K}}$ , and  $\tilde{T}$  is the (not necessarily disjoint) union of the sets of time points  $T$  and  $\bar{T}$  that constitute each connected component.

A model structure for  $US \oplus \bar{U}\bar{S}$  over  $\mathcal{K} \oplus \bar{\mathcal{K}}$  is a 4-tuple  $(\tilde{T}, <, \bar{<}, g)$ , where  $(\tilde{T}, <, \bar{<}) \in \mathcal{K} \oplus \bar{\mathcal{K}}$  and  $g$  is an assignment function  $g: \tilde{T} \rightarrow 2^{\mathcal{P}}$ .

The semantics of a formula  $A$  in a model  $\mathcal{M} = (\tilde{T}, <, \bar{<}, g)$  is defined as the union of the rules defining the semantics of  $US/\mathcal{K}$  and  $\bar{U}\bar{S}/\bar{\mathcal{K}}$ . The expression  $\mathcal{M}, t \models A$  reads that the formula  $A$  is true in the (combined) model  $\mathcal{M}$  at the point  $t \in \tilde{T}$ . The semantics of formulas is given by induction in the standard way:

$$\begin{aligned} \mathcal{M}, t \models p & \quad \text{iff } p \in g(t) \text{ and } p \in \mathcal{P}. \\ \mathcal{M}, t \models \neg A & \quad \text{iff it is not the case that } \mathcal{M}, t \models A. \\ \mathcal{M}, t \models A \wedge B & \quad \text{iff } \mathcal{M}, t \models A \text{ and } \mathcal{M}, t \models B. \\ \mathcal{M}, t \models S(A, B) & \quad \text{iff there exists an } s \in \tilde{T} \text{ with } s < t \text{ and } \mathcal{M}, s \models A \\ & \quad \text{and for every } u \in \tilde{T}, \text{ if } s < u < t \text{ then } \\ & \quad \mathcal{M}, u \models B. \\ \mathcal{M}, t \models U(A, B) & \quad \text{iff there exists an } s \in \tilde{T} \text{ with } t < s \text{ and } \mathcal{M}, s \models A \\ & \quad \text{and for every } u \in \tilde{T}, \text{ if } t < u < s \text{ then } \\ & \quad \mathcal{M}, u \models B. \\ \mathcal{M}, t \models \bar{S}(A, B) & \quad \text{iff there exists an } s \in \tilde{T} \text{ with } s \bar{<} t \text{ and } \mathcal{M}, s \models A \\ & \quad \text{and for every } u \in \tilde{T}, \text{ if } s \bar{<} u \bar{<} t \text{ then } \mathcal{M}, u \models B. \\ \mathcal{M}, t \models \bar{U}(A, B) & \quad \text{iff there exists an } s \in \tilde{T} \text{ with } t \bar{<} s \text{ and } \mathcal{M}, s \models A \\ & \quad \text{and for every } u \in \tilde{T}, \text{ if } t \bar{<} u \bar{<} s \text{ then } \mathcal{M}, u \models B. \end{aligned}$$

The also independent combination of two logics appears in the literature under the names of *fusion* or *join*.

As usual, we will assume that  $\mathcal{K}, \bar{\mathcal{K}} \subseteq \mathcal{K}_{lin}$ , so  $<$  and  $\bar{<}$  are transitive, irreflexive and total orders; similarly, we assume that the axiomatisations are extensions of  $US/\mathcal{K}_{lin}$ .

The temporalisation process will be used as an inductive step to prove the transference of soundness, completeness and decidability for  $US \oplus \bar{U}\bar{S}$  over  $\mathcal{K} \oplus \bar{\mathcal{K}}$ . We define the *degree alternation* of a  $(US \oplus \bar{U}\bar{S})$ -formula  $A$  for  $US$ ,  $dg(A)$ :

$$\begin{aligned} dg(p) &= 0 \\ dg(\neg A) &= dg(A) \\ dg(A \wedge B) &= dg(S(A, B)) = dg(U(A, B)) = \max\{dg(A), dg(B)\} \\ dg(\bar{U}(A, B)) &= dg(\bar{S}(A, B)) = 1 + \max\{\bar{dg}(A), \bar{dg}(B)\} \end{aligned}$$

and similarly define  $\bar{dg}(A)$  for  $\bar{U}\bar{S}$ .

Any formula of the fully combined language can be seen as a formula of some finite number of alternating temporalisations of the form

$US(\bar{U}\bar{S}(US(\dots)))$ ; more precisely,  $A$  can be seen as a formula of  $US(L_n)$ , where  $dg(A) = n$ ,  $US(L_0) = US$ ,  $\bar{U}\bar{S}(L_0) = \bar{U}\bar{S}$ , and  $L_{n-2i} = \bar{U}\bar{S}(L_{n-2i-1})$ ,  $L_{n-2i-1} = US(L_{n-2i-2})$ , for  $i = 0, 1, \dots, \lceil \frac{n}{2} \rceil - 1$ .

Indeed, not only the language of  $US \oplus \bar{U}\bar{S}$  is decomposable in a finite number of temporalisation, but also its inferences, as the following important lemma indicates.

**LEMMA 46.** *Let  $US$  and  $\bar{U}\bar{S}$  be two complete logic systems. Then,  $A$  is a theorem of  $US \oplus \bar{U}\bar{S}$  iff it is a theorem of  $US(L_n)$ , where  $dg(A) = n$ .*

**Proof.** If  $A$  is a theorem of  $US(L_n)$ , all the inferences in its deduction can be repeated in  $US \oplus \bar{U}\bar{S}$ , so it is a theorem of  $US \oplus \bar{U}\bar{S}$ .

Suppose  $A$  is a theorem of  $US \oplus \bar{U}\bar{S}$ ; let  $B_1, \dots, B_m = A$  be a deduction of  $A$  in  $US \oplus \bar{U}\bar{S}$  and let  $n' = \max\{dg(B_i)\}$ ,  $n' \geq n$ . We claim that each  $B_i$  is a theorem of  $US(L_{n'})$ . In fact, by induction on  $m$ , if  $B_i$  is obtained in the deduction by substituting into an axiom, the same substitution can be done in  $US(L_{n'})$ ; if  $B_i$  is obtained by Temporal Generalisation from  $B_j$ ,  $j < i$ , then by the induction hypothesis,  $B_j$  is a theorem of  $US(L_{n'})$  and so is  $B_i$ ; if  $B_i$  is obtained by Modus Ponens from  $B_j$  and  $B_k$ ,  $j, k < i$ , then by the induction hypothesis,  $B_j$  and  $B_k$  are theorems of  $US(L_{n'})$  and so is  $B_i$ .

So  $A$  is a theorem of  $US(L_{n'})$  and, since  $US$  and  $\bar{U}\bar{S}$  are two complete logic systems, by Theorem 39, each of the alternating temporalisations in  $US(L_{n'})$  is a conservative extension of the underlying logic; it follows that  $A$  is a theorem of  $US(L_n)$ , as desired. ■

Note that the proof above gives conservativeness as a corollary. The transference of soundness, completeness and decidability also follows directly from this result.

**THEOREM 47 (Independent Combination).** *Let  $US$  and  $\bar{U}\bar{S}$  be two sound and complete logic systems over the classes  $\mathcal{K}$  and  $\bar{\mathcal{K}}$ , respectively. Then their independent combination  $US \oplus \bar{U}\bar{S}$  is sound and complete over the class  $\mathcal{K} \oplus \bar{\mathcal{K}}$ . If  $US$  and  $\bar{U}\bar{S}$  are complete and decidable, so is  $US \oplus \bar{U}\bar{S}$ .*

**Proof.** Soundness follows immediately from the validity of axioms and inference rules.

We only sketch the proof of completeness here. Given a  $US \oplus \bar{U}\bar{S}$ -consistent formula  $A$ , Lemma 46 is used to see that it is also consistent in  $US(L_n)$ , so a temporalised  $US(L_n)$ -model is built for it. Then, by induction on the degree of alternation of  $A$ , this  $US(L_n)$  is used to construct a  $US \oplus \bar{U}\bar{S}$ -model; each step of such construction preserves the satisfiability of formulas of a limited degree of alternation, so in the final model,  $A$ , is satisfiable; and completeness is proved. For details, see [Finger and Gabbay, 1996].

For decidability, suppose we want to decide whether a formula  $A \in US \oplus \bar{U}\bar{S}$  is a theorem. By Lemma 46, this is equivalent to deciding whether  $A \in US(L_n)$  is a theorem, where  $n = dg(A)$ . Since  $US/\mathcal{K}$  and  $\bar{U}\bar{S}/\bar{\mathcal{K}}$  are both

complete and decidable, by successive applications of Theorems 39 and 43, it follows that the following logics are decidable:  $US(\bar{U}\bar{S})$ ,  $\bar{U}\bar{S}(US(\bar{U}\bar{S})) = \bar{U}\bar{S}(L_2)$ ,  $\dots$ ,  $\bar{U}\bar{S}(L_{n-1}) = L_n$ ; so a the last application of Theorems 39 and 43 yields that  $US(L_n)$  is decidable.  $\blacksquare$

*The minimality of the independent combination*

The logic  $US \oplus \bar{U}\bar{S}$  is the minimal logic that conservatively extends both  $US$  and  $\bar{U}\bar{S}$ . This result was first shown for the independent combination of monomodal logics independently by [Kracht and Wolter, 1991] and [Fine and Schurz, 1991].

Indeed, suppose there is another logic  $T_1$  that conservatively extends both  $US$  and  $\bar{U}\bar{S}$  but some theorem  $A$  of  $US \oplus \bar{U}\bar{S}$  is not a theorem of  $T_1$ . But  $A$  can be obtained by a finite number of inferences  $A_1, \dots, A_n = A$  using only the axioms of  $US$  and  $\bar{U}\bar{S}$ . But any conservative extension of  $US$  and  $\bar{U}\bar{S}$  must be able to derive  $A_i$ ,  $1 \leq i \leq n$ , from  $A_1, \dots, A_{i-1}$ , and therefore it must be able to derive  $A$ ; contradiction.

Once we have this minimal combination between two logic systems, any other interaction between the logics must be considered on its own. As an example, consider the following formulas expressing the commutativity of future and past operators between the two dimensions are not generally valid over a model of  $US \oplus \bar{U}\bar{S}$ :

$$\begin{aligned} l_1 \quad & F\bar{F}A \leftrightarrow \bar{F}FA \\ l_2 \quad & F\bar{P}A \leftrightarrow \bar{P}FA \\ l_3 \quad & P\bar{F}A \leftrightarrow \bar{F}PA \\ l_4 \quad & P\bar{P}A \leftrightarrow \bar{P}PA \end{aligned}$$

Now consider the *product* of two linear temporal models, given as follows.

DEFINITION 48. Let  $(T, <) \in \mathcal{K}$  and  $(\bar{T}, \bar{<}) \in \bar{\mathcal{K}}$  be two linear flows of time. The *product* of those flows of time is  $(T \times \bar{T}, <, \bar{<})$ . A *product model* over  $\mathcal{K} \times \bar{\mathcal{K}}$  is a 4-tuple  $\mathcal{M} = (T \times \bar{T}, <, \bar{<}, g)$ , where  $g : T \times \bar{T} \rightarrow 2^{\mathfrak{A}}$  is a two-dimensional assignment. The semantics of the horizontal and vertical operators are independent of each other:

$$\begin{aligned} \mathcal{M}, t, x \models S(A, B) \quad & \text{iff} \quad \text{there exists } s < t \text{ such that } \mathcal{M}, s, x \models A \\ & \text{and for all } u, s < u < t, \mathcal{M}, u, x \models B. \\ \mathcal{M}, t, x \models \bar{S}(A, B) \quad & \text{iff} \quad \text{there exists } y \bar{<} x \text{ such that } \mathcal{M}, t, y \models A \\ & \text{and for all } z, y \bar{<} z \bar{<} x, \mathcal{M}, t, z \models B. \end{aligned}$$

Similarly for  $U$  and  $\bar{U}$ , the semantics of atoms and Boolean connectives remaining the standard one. A formula  $A$  is valid over  $\mathcal{K} \times \bar{\mathcal{K}}$  if for all models  $\mathcal{M} = (T, <, \bar{T}, \bar{<}, g)$ , for all  $t \in T$  and  $x \in \bar{T}$  we have  $\mathcal{M}, t, x \models A$ .

It is easy to verify that the formulas  $l_1$ – $l_4$  are valid over product models. We wonder if such product of logics transfers the properties we have investigated for the previous logics. The answer is: it depends. We have the following results.

PROPOSITION 49.

- (a) *There is a sound and complete axiomatisation for  $US \times \bar{U}\bar{S}$  over the classes of product models  $\mathcal{K}_{lin} \times \mathcal{K}_{lin}$ ,  $\mathcal{K}_{dis} \times \mathcal{K}_{dis}$ ,  $\mathbb{Q} \times \mathbb{Q}$ ,  $\mathcal{K}_{lin} \times \mathcal{K}_{dis}$ ,  $\mathcal{K}_{lin} \times \mathbb{Q}$  and  $\mathbb{Q} \times \mathcal{K}_{dis}$  [Finger, 1994].*
- (b) *There are no finite axiomatisations for the valid two-dimensional formulas over the classes  $\mathbb{Z} \times \mathbb{Z}$ ,  $\mathbb{N} \times \mathbb{N}$  and  $\mathbb{R} \times \mathbb{R}$  [Venema, 1990].*

Note that the all the component one-dimensional mentioned above logic systems are complete and decidable, but their product sometimes is complete, sometimes not. Also, the logics in (a) are all decidable and those in (b) are undecidable.

This is to illustrate the following idea: given an independent combination of two temporal logics, the addition of extra axioms, inference rules or an extra condition on its models has to be studied on its own, just as adding a new axiom to a modal logic or imposing a new property on its accessibility relation has to be analysed on its own.

#### *Combinations of logics in the literature*

The work on combining temporal logics presented here has first appeared in the literature in [Finger and Gabbay, 1992; Finger and Gabbay, 1996].

General combinations of logics have been addressed in the literature in various forms. Combinations of tense and modality were discussed in the next chapter in this volume, (which reproduces [Thomason, 1984]), without explicitly providing a general methodology for doing so. A methodology for constructing logics of belief based on existing deductive systems is the *deductive model* of Konolige [Konolige, 1986]; in this case, the language of the original system was the base for the construction of a new modal language, and the modal logic system thus generated had its semantics defined in terms of the inferences of the original system. This is a methodology quite different from the one adopted here, in which we separately combine language, inference systems and class of models.

Combination of two monomodal logics and the transference of properties have been studied by Kracht and Wolter [1991] and Fine and Schurz [1991]; the latter even considers the transference of properties through the combination of  $n$ -monomodal logics. These works differ from the combination of temporal logics in several ways: their modalities have no interaction whatsoever (unlike  $S$  and  $U$ , which actually interact with each other); they only

consider one-place modalities ( $\Box$ ); and their constructions are not a recursive application of the temporalisation (or any similar external application of one logic to another).

A stronger combination of logics have been investigated by Gabbay and Shehtman [Gabbay and Shehtman, 1998], where the starting point is the product of two Kripke frames, generating the *product* of the two monomodal logics. It shows that the transference of completeness and decidability can either succeed or fail for the product, depending on the properties of the component logics. The failure of transference of decidability for temporal products in  $FP/\mathcal{K}_{lin} \times FP/\mathcal{K}_{lin}$  has been shown in [Marx and Reynolds, 1999], and fresh results on the products of logics can be found in [Reynolds and Zakharyashev, 2001].

The transference of soundness, completeness and decidability are by no means the only properties to study. Kracht and Wolter [Kracht and Wolter, 1991] study the transference of interpolation between two monomodal logics. The complexity of the combination of two monomodal logics is studied in [Spaan, 1993]; the complexity of products are studied in [Marx, 1999]. Gabbay and Shehtman [Gabbay and Shehtman, 1998] report the failure of transference of the finite model property for their product of modal logics. With respect to specific temporal properties, the transference of the separation property is studied in [Finger and Gabbay, 1992].

For a general combining methodology, see [Gabbay, 1998].

## 5 LABELLED DEDUCTION PRESENTATION OF TEMPORAL LOGICS

### 5.1 Introducing LDS

This section develops proof theory for temporal logic within the framework of labelled deductive systems [Gabbay, 1996].

To motivate our approach consider a temporal formula  $\alpha = FA \wedge PB \wedge C$ . This formula says that we want  $A$  to hold in the future (of now),  $B$  to hold in the past and  $C$  to hold now. It represents the following temporal configuration:

- $t < d < s, t \vDash B, d \vDash C$  and  $s \vDash A$

where  $d$  is now and  $t, s$  are temporal points.

Suppose we want to be very explicit about the temporal configuration and say that we want another instance of  $B$  to hold in the past of  $s$  but not in the past or future of  $d$ , i.e. we want an additional point  $r$  such that:

- $r < s, \sim (r = d \vee r < d \vee d < r)$  and  $r \vDash B$ .



The above cannot be expressed by a formula. The obvious formula  $\beta = F(A \wedge PB) \wedge PB \wedge C$  will not do. We need extra expressive power. We can, for example, use an additional atom  $q$  and write

$$\gamma = q \wedge Hq \wedge Gq \wedge F(A \wedge P(B \wedge \sim q)) \wedge PB \wedge C.$$

This will do the job.

However, by far the simplest approach is to allow names of points in the language and write  $s : A$  to mean that we want  $s \models A$  to hold. Then we can write a theory  $\Delta$  as

$$\Delta = \{t < d < s, t : A, d : C, s : B, r : A, \sim (r < d \vee r = d \vee d < r)\}.$$

$\Delta$  is satisfied if we find a model  $(S, R, a, h)$  in which  $d$  can be identified with  $g(d) = a$  and  $t, s, r$  can be identified with some points  $g(t), g(s), g(r)$  such that the above ordering relations hold and the respective formulae are satisfied in the appropriate points.

The above language has turned temporal logic into a labelled deductive system (LDS). It has brought some of the semantics into the syntax.

But how about proof theory?

Consider  $t : FFPB$ . This formula does hold at  $t$  (because  $B$  holds at  $r$ ). Thus we must have rules that allow us to show that

$$\Delta \vdash FFPB.$$

It is convenient to write  $\Delta$  as:

Assumptions	Configuration
$t : B$	$t < d < s$
$d : C$	$\sim (r < d \vee d = r \vee d < r)$
$s : A$	$r < s$
$r : B$	

and give rules to manipulate the configuration until we get  $t : FFPB$ .

Thus formally a temporal database is a set of labelled formulae  $\{t_i : A_i\}$  together with a configuration on  $\{t_i\}$ , given in the form of an earlier-later relation  $<$ . A query is a labelled formula  $t : Q$ . The proof rules have the form

$$\frac{t_1 : A_1; \dots; t_n : A_n, \text{ configuration}}{s : B \text{ configuration}'}$$

where the configuration is a set of conditions on the ordering of  $\{t_i\}$ .

The use of labels is best illustrated via examples:

**EXAMPLE 50.** This example shows the LDS in the case of modal logic. Modal logic has to do with possible worlds. Thus we think of our basic database (or assumptions) as a finite set of information about possible

worlds. This consists of two parts. The configuration part, the finite configuration of possible worlds for the database, and the assumptions part which tells us what formulae hold in each world. The following is an example of a database:

<b>Assumptions</b>	<b>Configuration</b>
(1) $t : \Box\Box B$	$t < s$
(2) $s : \Diamond(B \rightarrow C)$	

The conclusion to show (or query) is

$$t : \Diamond\Diamond C.$$

The derivation is as follows:

- (3) From (2) create a new point  $r$  with  $s < r$  and get  $r : B \rightarrow C$ .

We thus have

<b>Assumptions</b>	<b>Configuration</b>
(1), (2), (3)	$t < s < r$

- (4) From (1), since  $t < s$  get  $s : \Box B$ .  
 (5) From (4) since  $s < r$  get  $r : B$ .  
 (6) From (5) and (3) we get  $r : C$ .  
 (7) From (6) since  $s < r$  get  $s : \Diamond C$ .  
 (8) From (7) using  $t < s$  we get  $t : \Diamond\Diamond C$ .

**Discussion** The object rules involved are:

$\Box E$  rule:

$$\frac{t < s; t : \Box A}{s : A}$$

$\Diamond I$  rule:

$$\frac{t < s, s : B}{t : \Diamond B}$$

$\Diamond E$  rule:

$$\frac{t : \Diamond A}{\text{create a new point } s \text{ with } t < s \text{ and deduce } s : A}$$

Note that the above rules are not complete. We do not have rules for deriving, for example,  $\Box A$ . Also, the rules are all for intuitionistic modal logic.

The metalevel considerations which determine which logic we are working in, may be properties of  $<$ , e.g.  $t < s \wedge s < r \rightarrow t < r$ , or linearity, e.g.  $t < s \vee t = s \vee s < t$  etc.

There are two serious problems in modal and temporal theorem proving. One is that Skolem functions for  $\exists x \Diamond A(x)$  and  $\Diamond \exists x A(x)$  are not logically the same. If we ‘Skolemize’ we get  $\Diamond A(c)$ . Unfortunately it is not clear where  $c$  exists, in the current world ( $\exists x = c \Diamond A(x)$ ) or the possible world ( $\Diamond \exists x = c A(x)$ ).

If we use labelled assumptions then  $t : \exists x \Diamond A(x)$  becomes  $t : \Diamond A(c)$  and it is clear that  $c$  is introduced at  $t$ .

On the other hand, the assumption  $t : \Diamond \exists x A(x)$  will be used by the  $\Diamond E$  rule to introduce a new point  $s, t < s$  and conclude  $s : \exists x A(x)$ . We can further ‘Skolemize’ at  $s$  and get  $s : A(c)$ , with  $c$  introduced at  $s$ . We thus need the mechanism of remembering or labelling constants as well, to indicate where they were first introduced.

EXAMPLE 51. Another example has to do with the Barcan formula

Assumption	Configuration
(1) $t : \forall x \Box A(x)$	$t < s$

We show

$$(2) \quad s : \forall x A(x).$$

We proceed intuitively

$$(3) \quad t : \Box A(x) \text{ (stripping } \forall x, \text{ remembering } x \text{ is arbitrary).}$$

$$(4) \quad \text{Since the configuration contains } s, t < s \text{ we get}$$

$$s : A(x).$$

$$(5) \quad \text{Since } x \text{ is arbitrary we get}$$

$$s : \forall x A(x).$$

The above intuitive proof can be restricted.

The rule

$$\frac{t : \Box A(x), t < s}{s : A(x)}$$

is allowed only if  $x$  is instantiated.

To allow the above rule for arbitrary  $x$  is equivalent to adopting the Barcan formula axiom

$$\forall x \Box A(x) \rightarrow \Box \forall x A(x).$$

EXAMPLE 52. To show  $\forall x \Box A(x) \rightarrow \Box \forall x A(x)$  in the modal logic where it is indeed true.

- (1) Assume  $t : \forall x \Box A(x)$ .

We show  $\Box \forall x A(x)$  by the use of the metabox:

create $\alpha$ ,	$t < \alpha$
(2) $t : \Box A(x)$	from (1)
(3) $\alpha : A(x)$	from (2) using a rule which allows this with $x$ a variable.
(4) $\alpha : \forall x A(x)$	universal generalization.

- (5) Exit:  $t : \Box \forall x A(x)$ .

This rule has the form

Create $\alpha$ ,	$t < \alpha$
Argue to get	$\alpha : B$
Exit with	$t : \Box B$

## 5.2 LDS semantics

We can now formally define a simplified version of LDS, sufficient for our temporal logics. The reader is referred to [Gabbay, 1996] for full details.

An algebraic LDS is built up from two components: an algebra  $\mathcal{A}$  and a logic  $\mathbf{L}$ . To make things specific, let us assume that we are dealing with a particular algebraic model  $\mathcal{A} = (S, <, f_1, \dots, f_k)$ , where  $S$  is the domain of the algebra,  $<$  is a strict order, i.e. irreflexive and transitive, relation on  $S$  and  $f_1, \dots, f_k$  are function symbols on  $S$  of arities  $r_1, \dots, r_k$  respectively. The sequence  $\Sigma = (<, f_1, \dots, f_k)$  is called the signature of  $\mathcal{A}$ . It is really the language of  $\mathcal{A}$  in logical terms but we use  $\Sigma$  to separate it from  $\mathbf{L}$ . We assume the functions are *isotonic*, i.e. they are either monotonic up or monotonic down in each variable, namely for each coordinate  $x$  in  $f$  we have that either

$$\forall x, y (x < y \rightarrow f(\dots, x, \dots) < f(\dots, y, \dots))$$

or

$$\forall x, y (x < y \rightarrow f(\dots, y, \dots) < f(\dots, x, \dots))$$

holds.

A typical algebra is a binary tree algebra where each point  $x$  in the tree has two immediate successor points  $r_1(x)$  and  $r_2(x)$  and one predecessor point  $p(x)$ .  $<$  is the (branching) earlier–later relation and we have  $p(x) < x < r_i(x)$ ,  $i = 1, 2$ .

The general theory of LDS (see [Gabbay, 1996, Section 3.2]) requires a source of labels and a source of formulae. These together are used to form the declarative units of the form  $t : A$ , where  $t$  is a label and  $A$  is a formula.

The labels can be syntactical terms in some algebraic theory. The algebraic theory itself can be characterized either syntactically by giving axioms which the terms must satisfy or semantically by giving a class of models (algebras) for the language.

The formulae part of an LDS declarative unit is defined in the traditional way in some language  $\mathbf{L}$ .

An LDS database (theory)  $\Delta$  is a set of terms and their formulae (i.e. a set of declarative units) with some relationships between the terms. In a general LDS, the terms themselves are syntactical and one always has to worry whether the required relations between the terms of  $\Delta$  are possible (i.e. are consistent).

If, however, we have a semantic theory for the labels characterized by one single model (algebra), then we can take the labels to be elements of this model and consistency and relationships among the labels (elements) of  $\Delta$  will always be clear—they are as dictated by the model. This represents a temporal logic with a concrete specific flow of time (e.g. integers, rationals, reals, etc.).

We therefore present for the purpose of this chapter, a concrete definition of LDS based on a single model as an algebra of labels.

**DEFINITION 53** (Concrete algebraic LDS).

1. A concrete algebraic LDS has the form  $(\mathcal{A}, \mathbf{L})$  where:
  - (a)  $\mathcal{A} = (S, <, f_1, \dots, f_k)$  is a concrete algebraic model. The elements of  $S$  are called labels.
  - (b)  $\mathbf{L}$  is a predicate language with connectives  $\sharp_1, \dots, \sharp_m$  with arities  $r_1, \dots, r_m$ , and quantifiers  $(Q_1x), \dots, (Q_mx)$ . The connectives can be some well-known modalities, binary conditional, etc., and the quantifiers can be some known generalized or traditional quantifiers. We assume that the traditional syntactical notions for  $\mathbf{L}$  are defined. We also assume that  $\mathbf{L}$  has only constants, no function symbols and the constants of  $\mathbf{L}$  are indexed by elements of the algebra  $\mathcal{A}$ , i.e. have the form  $c_i^t, t \in S, i = 1, 2, 3, \dots$
2. A declarative unit has the form  $t : A$ , where  $A$  is a wff and  $t \in S$ , or it has the form  $t : c_i^s$ . The unit  $t : A$  intuitively means ‘ $A$  holds at label  $t$ ’ and the unit  $t : c_i^s$  means ‘the element  $c_i$  which was created at label  $s$  does exist in the domain of label  $t$ ’.

3. A database has the form  $\Delta = (D, \mathbf{f}, d, U)$  where  $D \subseteq S$  is non-empty and  $\mathbf{f}$  is a function associating with each  $t \in D$  a set of wffs  $\mathbf{f}(t) = \Delta_t$ .  $U$  is a function associating with each  $t \in D$  a set of terms  $U_t$ .  $d \in D$  is a distinguished point in  $D$ . The theory  $\Delta$  can be displayed by writing  $\{t : A_1, t : A_2, s : B, r : c_3^s, \dots\}$ , where  $t : A$  indicates  $A \in \mathbf{f}(t)$  and  $r : c^s$  indicates  $c^s \in U_r$ .

DEFINITION 54 (Semantics for LDS). Let  $(\mathcal{A}, \mathbf{L})$  be a concrete algebraic LDS, with algebra  $\mathcal{A}$  and language  $\mathbf{L}$  with connectives  $\{\#_1, \dots, \#_m\}$  and quantifiers  $\{Q_1, \dots, Q_{m'}\}$  where  $\#_i$  is  $r_i$  place.

1. A semantical interpretation for the LDS has the form  $\mathcal{I} = (\Psi_0(x, X), \Psi_1, \dots, \Psi_i(x, X_1, \dots, X_{r_i}), \dots, \Psi_m, \Psi'_1(x, Z), \dots, \Psi'_{m'}(x, Z))$  where  $\Psi_i$  is a formula of the language of  $\mathcal{A}$ , possibly second order, with the single free element variable  $x$  and the free set variables  $X$  as indicated, and  $\Psi'_i$  have a single free element variable  $x$  and free binary relation variable  $Z$ . We need to assume that  $\Psi_i$  and  $\Psi'_j$  have the property that if we substitute in them for the set variables closed under  $\leq$  then the element variable coordinate is monotonic up under  $\leq$ . In symbols:

$$\bullet \bigwedge_j \forall x, y (x \in X_j \wedge x \leq y \rightarrow y \in X_j) \rightarrow [x \leq y \rightarrow \Psi(x, X_j) \rightarrow \Psi(y, X_j)].$$

2. A model for the LDS has the form  $\mathbf{m} = (V, h, g, d)$  where  $d \in S$  is the distinguished world,  $V$  is a function associating a domain  $V_t$  with each  $t \in S$ ,  $h$  is a function associating with each  $n$ -place atomic predicate  $P$  a subset  $h(t, P) \subseteq V_t^n$ .

$g$  is an assignment giving each variable  $x$  an element  $g(x) \in V_d$  and for each constant  $c_i^s$  an element  $g(c_i^s) \in V_s$ .

3. Satisfaction is defined by structural induction as follows:

- $t \models P(b_1, \dots, b_n)$  iff  $(b_1, \dots, b_n) \in h(t, P)$ ;
- $t \models \exists x A(x)$  iff for some  $b \in V_t$ ,  $t \models A(b)$ ;
- $t \models A \wedge B$  iff  $t \models A$  and  $t \models B$ ;
- $t \models \sim A$  iff  $t \not\models A$ ;
- $t \models \#_i(A_1, \dots, A_{r_i})$  iff  $\mathcal{A} \models \Psi_i(t, \hat{A}_1, \dots, \hat{A}_{r_i})$ , where  $\hat{A} = \{s \in S \mid s \models A\}$ ;
- $t \models (Q_i y)A(y)$  iff  $\mathcal{A} \models \Psi'_i(t, \widehat{\lambda y A(y)})$ , where  $\widehat{\lambda y A(y)} = \{(t, y) \mid t \models A(y)\}$ .

4. We say  $A$  holds in  $\mathbf{m}$  iff  $\mathcal{A} \models \Psi_0(d, \hat{A})$ .

5. The interpretation  $\mathcal{I}$  induces a translation  $*$  of  $\mathbf{L}$  into a two-sorted language  $\mathbf{L}^*$  based on the two domains  $S$  (of the algebra  $\mathcal{A}$ ) and  $U = \bigcup_t V_t$  (of the predicates of  $\mathbf{L}$ ) as follows:

- each atomic predicate  $P(x_1, \dots, x_n)$  (interpreted over the domain  $U$ ) is translated into

$$[P(x_1, \dots, x_n)]_t^* = P^*(t, x_1, \dots, x_n),$$

where  $P^*$  is a two-sorted predicate with one more variable  $t$  ranging over  $S$  and  $x_1, \dots, x_n$  ranging over  $U$ ;

- $[A \wedge B]_t^* = [A]_t^* \wedge [B]_t^*$ ;
- $[\sim A]_t^* = \sim [A]_t^*$ ;
- $[\sharp_i(A_1, \dots, A_{r_i})]_t^* = \Psi_i(t, \lambda_s[A_1]_s^*, \dots, \lambda_s[A_{r_i}]_s^*)$ ;
- $[(Q_i y)A(y)]_t^* = \Psi'_i(t, \lambda_y \lambda_s[A(y)]_s^*)$ ;
- Let  $\|A\|_t^* = \Psi_0(d, \lambda_t[A]_t^*)$ .

6. It is easy to show by induction that:

- $t \vDash A$  iff  $[A]_t^*$  holds in the naturally defined two-sorted model.

The reader should compare this definition with [Gabbay, 1996, Definition 3.2.6] and with Chapter 5 of [Gabbay *et al.*, 1994].

Gabbay's book on LDS contains plenty of examples of such systems.

In the particular case of temporal logic, the algebra has the form  $(D, <, d)$ , where  $D$  is the flow of time and  $<$  is the earlier-later relation.  $d \in D$  is the present moment (actual world).

### 5.3 Sample temporal completeness proof

The previous section presented the LDS semantics. This section will choose a sample temporal logical system and present *LDS* proof rules and a completeness theorem for it. We choose the modal logic  $\mathbf{K}_t$  (See Section 3.2 of [Gabbay *et al.*, 1994]). This is the propositional logic with  $H$  and  $G$  complete for all Kripke frames  $(S, R, a)$ ,  $a \in S$ , such that  $R$  is transitive and irreflexive. A wff  $A$  is a theorem of  $\mathbf{K}_t$  iff for all models  $(S, R, a, h)$  with assignment  $h$ , we have  $a \vDash A$ .

We want to turn  $\mathbf{K}_t$  into a quantified logic  $Q\mathbf{K}_t$ . We take as semantics the class of all models of the form  $(S, R, a, V, h)$  such that  $V_t$  for  $t \in S$  is the domain of world  $t$ . The following is assumed to hold:

- $tRs \wedge sRs'$  and  $a \in V_{s'}$  and  $a \in V_t$  imply  $a \in V_s$  (i.e. elements are born, exist for a while and then possibly die).

DEFINITION 55 (Traditional semantics for  $Q\mathbf{K}_t$ ).

1. A  $Q\mathbf{K}_t$  Kripke structure has the form  $(S, R, a, V, h)$ , where  $S$  is a non-empty set of possible worlds,  $R \subseteq S^2$  is the irreflexive and transitive accessibility relation and  $a \in S$  is the actual world.  $V$  is a function giving for each  $t \in S$  a non-empty domain  $V_t$ .

- Let  $V_S = \bigcup_{t \in S} V_t$ .

$h$  is the assignment function assigning for each  $t$  and each  $n$ -place atomic predicate  $P$  its extension  $h(t, P) \subseteq V_t^n$ , and for each constant  $c$  of the language its extension  $h(c) \in V_S$ .

Each  $n$ -place function symbol of the language of the form  $f(x_1, \dots, x_n)$  and  $t$  is assigned a function  $h(f) : V_S^n \mapsto V_S$ .

Note that function symbols are rigid, i.e. the assignment to a constant  $c$  is a fixed rigid element which may or may not exist at a world  $t$ .<sup>2</sup>

2. Satisfaction  $\models$  is defined in the traditional manner.

- (a)  $h$  can be extended to arbitrary terms by the inductive clause  $h(f(x_1, \dots, x_n)) = h(f)(h(x_1), \dots, h(x_n))$ .

- (b) We define for atomic  $P$  and terms  $x_1, \dots, x_n$

$$t \models_h P(x_1, \dots, x_n) \text{ iff } (h(x_1), \dots, h(x_n)) \in h(t, P).$$

- (c) The cases of the classical connectives are the traditional ones.

- (d)  $t \models_h \exists x A(x)$  iff for some  $a \in V_t, t \models_h A(a)$ .

- (e)  $t \models_h \forall x A(x)$  iff for all  $a \in V_t, t \models_h A(a)$ .

3.  $t \models_h GA(a_1, \dots, a_n)$  (resp.  $t \models HA$ ) iff for all  $s$ , such that  $tRs$  (resp.  $sRt$ ) and  $a_1, \dots, a_n \in V_s, s \models A$ .

4.  $t \models U(A(a_1, \dots, a_n), B(b_1, \dots, b_k))$  iff for some  $s, tRs$  and  $a_i \in V_s, i = 1, \dots, n$ , we have  $s \models A$  and for all  $s', tRs'$  and  $s'R_s$  and  $b_j \in V_{s'}, j = 1, \dots, k$ , imply  $s' \models B$ . (The mirror image holds for  $S(A, B)$ .)

5. Satisfaction in the model is defined as satisfaction in the actual world.

---

<sup>2</sup>Had we wanted non-rigid semantics we would have stipulated that the extension of a function symbol is  $h(t, f) : V_t^n \mapsto V_t$ . There is no technical reason for this restriction and our methods still apply. We are just choosing a simpler case to show how LDS works. Note that by taking  $h(t, P) \subseteq V_t^n$  as opposed to  $h(t, P) \subseteq V_S^n$ , we are introducing peculiarities in the semantic evaluation.  $t \models P(a_1, \dots, a_n)$  becomes false if not all  $a_i$  are in  $V_t$ . We can insist that we give values to  $t \models A(a_1, \dots, a_n)$  only if all elements are in  $V_t$ , but then what value do we give to  $t \models GA$ ? One option is to let  $t \models GA(a_1, \dots, a_n)$  iff for all  $s$  such that  $tRs$  and such that all  $a_i \in V_s$  we have  $s \models A$ .

Anyway, there are many options here and a suitable system can probably be chosen for any application area.



Note that the logic  $Q\mathbf{K}_t$  is not easy to axiomatize traditionally.

We now define an LDS corresponding to the system  $Q\mathbf{K}_t$ .

**DEFINITION 56** (The algebra of labels).

1. Consider the first-order theory of one binary relation  $<$  and a single constant  $d$ . Consider the axiom  $\partial = \forall x \sim (x < x) \wedge \forall xyz(x < y \wedge y < z \rightarrow x < z)$ . Any classical model of this theory has the form  $\mathbf{m} = (S, R, a, g)$ , where  $S$  is the domain,  $R$  is a binary relation on  $S$  giving the extension of the syntactical ' $<$ ' and  $g$  gives the extension of the variables and of  $d$ .  $g(d)$  equals  $a$  and is the interpretation of the constant ' $d$ '. Since  $(S, R, a, g) \models \partial$ , we have that  $R$  is irreflexive and transitive.
2. Let  $U = \{t_1, t_2, \dots\}$  be a set of additional constants in the predicate language of  $<$  and  $d$ . Let  $\mathcal{A}$  be the set of all terms of the language. By a diagram  $\Delta = (D, <, d)$ , with  $D = (D_1, D_2)$ , we mean a set  $D_1 \subseteq \mathcal{A}, d \in D_1$  of constants and variables and a set  $D_2$  of formulae  $\varphi(t, s)$  of the form  $t < s, \sim (t < s), t = s, t \neq s$ , with constants and variables from  $D_1$ .
3. A structure  $\mathbf{m} = (S, R, a, g)$  is a model of  $\Delta$  iff the following hold:
  - (a)  $g : D_1 \mapsto S$ , with  $g(d) = a$ ;
  - (b)  $R$  is irreflexive and transitive (i.e. it is a model of  $\partial$ );
  - (c) whenever  $\varphi(t, s) \in D_2$  then  $\varphi(g(t), g(s))$  holds in the model.
4. Note that  $g$  assigns elements of  $S$  also to the variables of  $D_1$ . Let  $x$  be a constant or a variable. Denote by  $g =_x g'$  iff for all variables and constants  $y \neq x$  we have  $g(y) = g'(y)$ .

**DEFINITION 57** (LDS language for  $Q\mathbf{K}_t$ ).

1. Let  $\mathbf{L}$  be the predicate modal language with the following:
  - (a) Its connectives and quantifiers are  $\wedge, \vee, \rightarrow, \perp, \top, \forall, \exists, G, F, H, P$ .
  - (b) Its variables are  $\{x_1, x_2, \dots\}$ .
  - (c) It has atomic predicates of different arities.
  - (d) It has function symbols  $e_1, e_2, \dots$ , of different arities.
  - (e) Let  $\mathcal{A}$  be the language of the algebra of labels of Definition 56. We assume that  $\mathcal{A}$  may share variables with  $\mathbf{L}$  but its constants are distinct from the constants of  $\mathbf{L}$ . For each constant  $t \in U$  of  $\mathcal{A}$  and each natural number  $n$ , we assume we have in  $\mathbf{L}$  a sequence of  $n$ -place function symbols

$$c_{n,1}^t(x_1, \dots, x_n), c_{n,2}^t(x_1, \dots, x_n) \dots$$

parameterized by  $t$ .

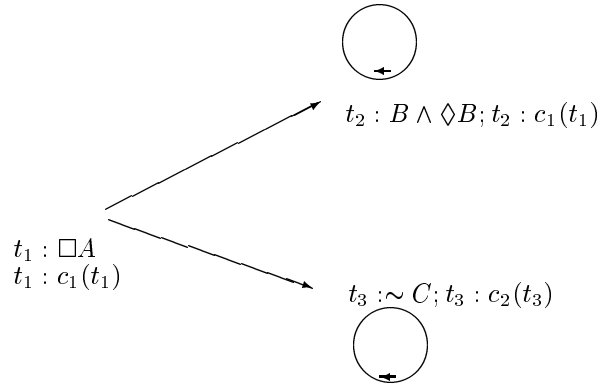


Figure 1.

Thus in essence we want an infinite number of Skolem functions of any arity parameterized by any  $t \in U$ . The elements of  $\mathcal{A}$  are our labels.

The LDS language is presented as  $(\mathcal{A}, \mathbf{L})$ .

2. A declarative unit is a pair  $t : A$ , where  $t$  is a constant from  $\mathcal{A}$  and  $A$  is a wff of  $\mathbf{L}$ .

Note that because some of the function symbols of  $\mathbf{L}$  are parameterized from  $U$ , we can get labels in  $\mathcal{A}$  as well. For example,

$$t : P(x, c_{1,1}^t(x))$$

is a declarative unit.

3. A configuration has the form  $(D, <, \mathbf{f}, d, U)$ , where  $(D, <, d)$  is a diagram as defined in Definition 56 and  $\mathbf{f}$  and  $U$  are functions associating with each  $t \in D_1$  a set  $\mathbf{f}(t)$  of wffs of  $\mathbf{L}$  and a set  $U_t$  of terms of  $\mathbf{L}$ .

We also write:

$t : A$  to indicate that  $A \in \mathbf{f}(t)$ ;

$t : c$  to indicate that  $c \in U_t$ .

Configurations can also be presented graphically. See for example Fig. 1.

DEFINITION 58 (LDS semantics for  $Q\mathbf{K}_t$ ).

1. A model for an LDS language  $(\mathcal{A}, \mathbf{L})$  has the form  $\mathbf{n} = (S, R, a, V, h, g)$  where  $(S, R, a, V, h)$  is a traditional  $Q\mathbf{K}_t$  model for the language  $\mathbf{L}$  and  $g$  is an assignment from the labelling language  $\mathcal{A}$  into  $S$ , giving values in  $S$  to each label and each variable. The following must hold:
  - (a) for a variable or a constant  $x$  common to both  $\mathcal{A}$  and  $\mathbf{L}$ ,  $g(x) = h(x)$ ;
  - (b) for any  $c = c_{n,k}^t(x_1, \dots, x_n)$  we have  $h(x) \in V_{g(t)}$ ;
  - (c) let  $t_1(x), t_2(y)$  be two terms of  $\mathbf{L}$  containing the subterms  $x$  and  $y$ . Assume that  $\partial$  (when augmented with the function symbols of  $\mathbf{L}$  and equality axioms) satisfies
 
$$\partial \vdash_{\mathcal{A}} (x = y) \rightarrow (t_1 = t_2);$$
 then if  $g(x) = g(y)$  then  $h(t_1) = h(t_2)$ .
  - (d)  $g(d) = a$ .
2. Let  $\Delta = (D, <, \mathbf{f}, d, U)$  be a modal configuration in a language  $\mathbf{L}$ . We define the notion of  $\mathbf{n} \models \Delta$  to mean that the following hold:
  - (a) for every  $t \in D$  and  $A \in \mathbf{f}(t)$  we have  $g(t) \models_h A$  in  $(S, R, a, V, h)$ , according to Definition 55;
  - (b)  $(S, R, a, g) \models (D, <, d)$  according to Definition 56;
  - (c) if  $x \in D_1$  is a variable then for all  $g' =_x g$  we have  $\mathbf{n}' = (S, R, a, V, h, g') \models \Delta$ , provided  $\mathbf{n}'$  is an acceptable model (satisfying the restrictions in 1). This means that the free variables occurring in  $D$  are interpreted universally.
3. Let  $\mathbf{n} = (S, R, a, V, h, g)$  be a model in a language  $(\mathcal{A}, \mathbf{L})$ . Let  $(\mathcal{A}', \mathbf{L}')$  be an extension of the language. Then  $\mathbf{n}'$  is said to be an extension of  $\mathbf{n}$  to  $(\mathcal{A}', \mathbf{L}')$  if the restriction of  $V', h', g'$  to  $(\mathcal{A}, \mathbf{L})$  equals  $V, h, g$  respectively.
4. Let  $\Delta$  be a temporal configuration in  $(\mathcal{A}, \mathbf{L})$  and  $\Delta'$  in  $(\mathcal{A}', \mathbf{L}')$ . We write  $\Delta \models \Delta'$  iff any model  $\mathbf{n}$  of  $\Delta$  can be extended to a model  $\mathbf{n}'$  of  $\Delta'$ .
5. We write  $\emptyset \models \Delta$  or equivalently  $\models \Delta$  iff for any model  $\mathbf{n}$ ,  $\mathbf{n} \models \Delta$ .

DEFINITION 59 (LDS proof rules for  $Q\mathbf{K}_t$ ). Let  $\Delta = (D, <, \mathbf{f}, d, U)$  and  $\Delta' = (D', <', \mathbf{f}', d, U')$  be two temporal configurations. We say that  $\Delta'$  is obtained from  $\Delta$  by the application of a single forward proof rule if one of the following cases hold (let  $\diamond \in \{P, F\}$ , and  $\square \in \{G, H\}$ ).

1.  $\diamond$  **introduction case**

For some  $t, s \in D_1, t < s \in D_2, A \in \mathbf{f}(s)$  and  $\Delta'$  is the same as  $\Delta$  except that  $\mathbf{f}'(t) = \mathbf{f}(t) \cup \{\diamond A\}$ .

Symbolically we can write  $\Delta' = \Delta_{[t < s; s:A]}$  for  $F$  and  $\Delta_{[s < t, s:A]}$  for  $P$ .

2.  $\diamond$  **elimination case**

For some  $t \in D_1$  and some  $A, \diamond A \in \mathbf{f}(t)$  and for some new atomic  $s \in U$ , that does not appear in  $\Delta$ , we have that  $\Delta'$  is the same as  $\Delta$  except that  $D'_1 = D_1 \cup \{s\}$ .  $D'_2 = D_2 \cup \{t < s\}$  for  $\diamond = fF$  (resp.  $s < t$  for  $\diamond = P$ ).  $\mathbf{f}'(s) = \{A\}$ .

Symbolically we can write  $\Delta' = \Delta_{[t:\diamond A; s]}$ .

3.  $\square$  **elimination case**

For some  $t, s \in D_1$  such that  $t < s \in D_2$  for  $\square = G$  (resp.  $s < t$  for  $\square = H$ ) we have  $\square A \in \mathbf{f}(t)$  and  $\Delta'$  is like  $\Delta$  except that  $\mathbf{f}'(s) = \mathbf{f}(s) \cup \{A\}$ .

Symbolically we can write  $\Delta' = \Delta_{[t < s; t:GA]}$ , and  $\Delta' = \Delta_{[s < t; t:HA]}$ .

4. **Local classical case**

$\Delta'$  is like  $\Delta$  except that for  $t \in D_1$  we have  $\mathbf{f}'(t) = \mathbf{f}(t) \cup \{A\}$ , where  $A$  follows from  $\mathbf{f}(t)$  using classical logic inference only.<sup>3</sup>

Symbolically we can write  $\Delta' = \Delta_{[t-A]}$ .

5. **Local  $\forall$  elimination**

For some  $t \in D_1$  we have  $\forall x A(x) \in \mathbf{f}(t)$  and  $c \in U_t$ ,  $\Delta'$  is like  $\Delta$  except  $\mathbf{f}'(t) = \mathbf{f}(t) \cup \{A(c)\}$ .

Symbolically we can write  $\Delta' = \Delta_{[t:\forall x A(x); x/c]}$ .

6. **Local  $\exists$  elimination**

For some  $t \in D_1$  and  $\forall x_1, \dots, x_n \exists y A(x_1, \dots, x_n, y) \in \mathbf{f}(t)$  and some *new* function symbol  $c^t(x_1, \dots, x_n)$  we have

$$\mathbf{f}'(t) = \mathbf{f}(t) \cup \{\forall x_1, \dots, x_n A(x_1, \dots, x_n, c^t(x_1, \dots, x_n))\}.$$

Otherwise  $\Delta'$  is like  $\Delta$ .

Symbolically we can write  $\Delta' = \Delta_{[t:\forall x_1, \dots, x_n \exists y A; y/c^t]}$ .

---

<sup>3</sup>Every formula of  $Q\mathbf{K}_t$  can be presented in the form  $B(Q_1/\square_1 A_1, \dots, Q_n/\square_n A_n)$  where  $\square_i \in \{G, H\}$  and  $B(Q_1, \dots, Q_n)$  is a modal free classical formula and  $A_i$  are general  $Q\mathbf{K}_t$  formulae.  $P/A$  means the substitution of  $A$  for  $P$ . A set of formulae of the form  $B_i(Q_j^i/\square_j A_j^i)$  proves using classical rules only a formula  $B(Q_j/\square_j A_j)$  iff  $\{B_i(Q_j^i)\}$  proves classically  $B(Q_j)$ .

**7. Visa rules**

- (a) For  $t, s \in D_1, t < s < t' \in D_2$  and  $c \in U_t$  and  $c \in U_{t'}$ ,  $\Delta'$  is like  $\Delta$  except that  $U'_s = U_s \cup \{c\}$ .
- (b)  $U'_t = U_t \cup \{c^t\}$ . In other words any  $c^t$  can be put in the domain of world  $t$ .

Symbolically we write  $\Delta' = \Delta_{[t:c, t':c \text{ to } s:c]}$  and  $\Delta' = \Delta_{[t:c^t]}$  respectively.

**8. Inconsistency rules**

- (a) If  $t, s \in D_1$  and  $\perp \in \mathbf{f}(t)$  then let  $\Delta'$  be like  $\Delta$  except that  $\mathbf{f}'(s) = \mathbf{f}(s) \cup \{\perp\}$  and  $\mathbf{f}'(x) = \mathbf{f}(x)$ , for  $x \neq s$ .  
Symbolically we write  $\Delta' = \Delta_{[t:\perp \text{ to } s:\perp]}$ .
- (b) If  $(D, <, d)$  is classically inconsistent and  $s \in D_1$ , we let  $\Delta'$  be as in (a) above.  
Symbolically we write  $\Delta' = \Delta_{[\perp \text{ to } s:\perp]}$ .

**9.  $\square$  introduction rule**

We say  $\Delta'$  is obtained from  $\Delta$  by a single-level  $n+1$  introduction rule if the following holds. For some  $t \in D_1$  we have  $\mathbf{f}'(t) = \mathbf{f}(t) \cup \{\square A\}$ ,  $\square \in \{G, H\}$  and  $\Delta_2$  follows from  $\Delta_1$  using a sequence of applications of single, forward rules or of single-level  $m \leq n$  introduction rules, and  $\Delta_1$  is like  $\Delta$  except that

$$D_1^1 = D_1 \cup \{t_1\}, D_2^1 = D_2 \cup \{t < t_1\} \text{ for } \square = G$$

$$\text{(and respectively } D_2^1 = D_2 \cup \{t_1 < t\} \text{ for } \square = H)$$

where  $t_1$  is a completely new constant label and  $\mathbf{f}^1(t_1) = \{\top\}$ .

$\Delta_2$  is like  $\Delta_1$  except that  $\mathbf{f}^2(t_1) = \mathbf{f}^1(t_1) \cup \{A\}$ .

Symbolically we write  $\Delta' = \Delta_{[t:\square A]}$ .

**10. Diagram rule**

$\Delta'$  is an extension of  $\Delta$  as a diagram, i.e.  $D \subseteq D', \mathbf{f}'(t) = \mathbf{f}(t), t \in D_1$  and  $\mathbf{f}'(t) = \{\top\}$ , for  $t \in D'_1 - D_1$ , and we have  $(D, <, d) \vdash_{\mathcal{A}} (D', <, d)$ .

Symbolically we write  $\Delta' = \Delta_{[D \vdash D']}$ .

Note that for our logic,  $Q\mathbf{K}_t$ , this rule just closes  $<$  under transitivity.

11. We write  $\Delta \vdash \Delta'$  iff there exists a sequence of steps of any level leading from  $\Delta$  to  $\Delta'$ . We write  $\vdash \Delta'$  if  $\Delta_0 \vdash \Delta'$ , for  $\Delta_0$  the theory with  $\{d\}$  only with  $\mathbf{f}(d) = \{\top\}$ .

Let  $t$  be a label of  $\Delta$ . We write  $\Delta \vdash t : A$  iff for any  $\Delta'$  such that  $\Delta \vdash \Delta'$  there exists a  $\Delta''$  such that  $\Delta' \vdash \Delta''$  and  $A \in \mathbf{f}''(t)$ . In other

words,  $\Delta$  can be proof-theoretically manipulated until  $A$  is proved at node  $t$ .

12. Notice that if  $\Delta \vdash \Delta'$  then the language  $\mathbf{L}'$  of  $\Delta'$  is slightly richer in labels and Skolem functions than the language  $\mathbf{L}$ .

Note also that if  $\Delta \vdash \Delta'$  then there is a sequence of symbols  $\pi_1, \dots, \pi_n$  such that  $\Delta' = \Delta_{\pi_1, \dots, \pi_n}$ .

In fact  $\Delta'$  is uniquely determined up to symbolic isomorphism by the sequence  $\pi_1, \dots, \pi_n$ .

13. **Local cut rule**

$\vdash$  of item 11 above is without the following *local cut rule*. Let  $\Delta_{t:B}$  denote the database obtained from  $\Delta$  by adding  $B$  at label  $t$ . Then  $\Delta_{t:B} \vdash \Delta'$  and  $\Delta_{t:\sim B} \vdash \Delta'$  imply  $\Delta \vdash \Delta'$ .

14. The consequence  $\Delta \vdash \Delta'$  can be implicitly formulated in a ‘sequent’-like form as follows:

- $\Delta \vdash \Delta_\pi$  (axiom)  
for  $\pi$  as in any of items 1–10 above.
- $$\frac{\Delta \vdash \Delta'; \Delta' \vdash \Delta''}{\Delta \vdash \Delta''}.$$
- Local cut rule.

DEFINITION 60 (Inconsistency).

1. A theory  $\Delta = (D, <, \mathbf{f}, d, U)$  is immediately inconsistent iff either  $(D, <, d)$  is inconsistent as a classical diagram or for some  $t \in D_1$ ,  $\perp \in \mathbf{f}(t)$ .
2.  $\Delta$  is inconsistent iff  $\Delta \vdash \Delta'$  and  $\Delta'$  is immediately inconsistent.
3. Note that if we do not provide the following inconsistency rules, namely

$$\frac{t : \perp}{s : \perp}$$

and

$$\frac{(D, <, d) \text{ is inconsistent}}{s : \perp},$$

then two inconsistent modal configurations cannot necessarily prove each other.

**THEOREM 61 (Soundness).** *Let  $\Delta$  be in  $\mathbf{L}$  and  $\Delta'$  in  $\mathbf{L}'$ . Then  $\Delta \vdash \Delta'$  implies  $\Delta \vDash \Delta'$ . In words, if  $\mathbf{n}$  is a model in  $\mathbf{L}$  such that  $\mathbf{n} \vDash \Delta$ , then  $\mathbf{n}$  can be extended to  $\mathbf{n}' \vDash \Delta'$ , where  $\mathbf{n}'$  is like  $\mathbf{n}$  except that the assignments are extended to  $\mathbf{L}'$ . In particular  $\vdash \Delta'$  implies  $\vDash \Delta'$ .*

**Proof.** By induction on the number of single steps proving  $\Delta'$  from  $\Delta$ . We can assume  $\Delta$  is consistent. It is easy enough to show that if  $\Delta'$  is obtained from  $\Delta$  by a single step then there is an  $\mathbf{n}' \vDash \Delta'$ .

Let  $\mathbf{n} = (S, R, a, V, h, g)$  be a model of  $\Delta = (D, <, \mathbf{f}, d, U)$ . Let  $\Delta'$  be obtained from  $\Delta$  by a single proof step and assume  $\mathbf{n} \vDash \Delta$ . We show how to modify  $\mathbf{n}$  to an  $\mathbf{n}'$  such that  $\mathbf{n}' \vDash \Delta'$ .

We follow the proof steps case by case:

1.  $\diamond$  **introduction case**

Here take  $\mathbf{n}' = \mathbf{n}$ .

2.  $\diamond$  **elimination case**

Here  $\Delta'$  contains a completely new constant  $s \in D'_1$ . We can assume  $g$  is not defined for this constant. Since  $\mathbf{n} \vDash \Delta$ , we have  $g(t) \vDash \diamond A$  and so for some  $b \in S, g(t)Rb$  and  $b \vDash A$ . Let  $g'$  be like  $g$  except  $g'(s) = b$ . Then  $\mathbf{n}' \vDash \Delta'$ .

3.  $\square$  **elimination case**

Here let  $\mathbf{n}' = \mathbf{n}$ .

4. **Local classical case**

Let  $\mathbf{n}' = \mathbf{n}$ .

5. **Local  $\forall$  elimination**

Let  $\mathbf{n}' = \mathbf{n}$ .

6. **Local  $\exists$  elimination**

We can assume the new function  $c^t(x_1, \dots, x_n)$  is new to the language of  $\Delta$  and that  $h$  is not defined for  $c^t$ . Since in  $\mathbf{n}$ ,  $t \vDash \forall x_1, \dots, x_n \exists y A(x_1, \dots, x_n, y)$ , for every  $(x_1, \dots, x_n) \in V_t^n$  a  $y \in V_t$  exists such that  $t \vDash A(x_i, y)$ . Thus an assignment  $h'(c^t)$  can be given to  $c^t$  by defining it to be this  $y$  for  $(x_1, \dots, x_n) \in V_t^n$  and to be a fixed element of  $V_t$  for tuples  $(x_1, \dots, x_n)$  not in  $V_t^n$ . Let  $\mathbf{n}'$  be like  $\mathbf{n}$  except we take  $h'$  instead of  $h$ .

7. **Visa rules**

(a) Take  $\mathbf{n}' = \mathbf{n}$ .

(b) Take  $\mathbf{n}' = \mathbf{n}$ .

(c) Take  $\mathbf{n}' = \mathbf{n}$ .

8. Inconsistency rules are not applicable as we assume  $\Delta$  is consistent.

9. Let  $\mathbf{n}_1$  be a model for the language of  $\Delta_1$  and assume  $\mathbf{n}_1 \models \Delta_1$ . Then by the induction hypothesis there exists  $\mathbf{n}_2 \models \Delta_2$ , where  $\mathbf{n}_2$  extends  $\mathbf{n}_1$ .

Assume now that  $\mathbf{n} \not\models \Delta'$ . Then  $t \not\models \Box A$ , and hence for some  $b$  such that  $tRb, b \not\models A$ .

Let  $\mathbf{n}_1$  be defined by extending  $g$  to  $g_1(t_1) = b$ . Clearly  $\mathbf{n}_1 \models \Delta_1$ . Since  $\Delta_1 \vdash \Delta_2$ ,  $g_1$  can be extended to  $g_2$  so that  $\mathbf{n}_2 \models \Delta_2$ , i.e.  $b \models A$ , but this contradicts our previous assumption. Hence  $\mathbf{n} \models \Delta'$ .

10. Let  $\mathbf{n}' = \mathbf{n}$ .

The above completes the proof of soundness.  $\blacksquare$

**THEOREM 62** (Completeness theorem for LDS proof rules). *Let  $\Delta$  be a consistent configuration in  $\mathbf{L}$ ; then  $\Delta$  has a model  $\mathbf{n} \models \Delta$ .*

**Proof.** We can assume that there are an infinite number of constants, labels and variables in  $\mathbf{L}$  which are not mentioned in  $\Delta$ . We can further assume that there are an infinite number of Skolem functions  $c_{n,i}^t(x_1, \dots, x_n)$  in  $\mathbf{L}$ , which are not in  $\Delta$ , for each  $t$  and  $n$ .

Let  $\delta(n)$  be a function such that

$$\delta(n) = (t_n, B_n, t'_n, \alpha_n, k_n),$$

where  $t_n, t'_n$  are labels or variables,  $\alpha_n$  a term of  $\mathbf{L}$  and  $B_n$  a formula of  $\mathbf{L}$ , and  $k_n$  a number between 0 and 7.

Assume that for each pair of labels  $t, t'$  of  $\mathbf{L}$  and each term  $\alpha$  and each formula  $B$  of  $\mathbf{L}$  and each  $0 \leq k \leq 7$  there exist an infinite number of numerals  $m$  such that  $\delta(m) = (t, B, t', \alpha, k)$ .

Let  $\Delta_0 = \Delta$ . Assume we have defined  $\Delta_n = (D^n, <, \mathbf{f}^n, d, U^n)$  and it is consistent.

We now define  $\Delta_{n+1}$ . Our assumption implies that  $(D^n, <, d)$  is classically consistent as well as each  $\mathbf{f}^n(t)$ . Consider  $(t_n, B_n, t'_n, \alpha_n, k_n)$ . It is possible that the formulae and labels of  $\delta(n)$  are not even in the language of  $\Delta_n$ , but if they are we can carry out a construction by case analysis.

**Case  $k_n = 0$**

This case deals with the attempt to add either the formula  $B_n$  or its negation  $\sim B_n$  to the label  $t_n$  in  $D^n$ . We need to do that in order to build eventually a Henkin model. We first try to add  $B_n$  and see if the resulting database  $(\Delta_n)_{t_n : B_n}$  is LDS-consistent. If not then we try to add  $\sim B_n$ . If both are LDS-inconsistent then  $\Delta_n$  itself is LDS-inconsistent, by the local cut rule (see item 13 of Definition 59).

The above is perfectly acceptable if we are prepared to adopt the local cut rule. However, if we want a system without cut then we must try to add  $B_n$  or  $\sim B_n$  using possibly other considerations, maybe a different notion of local consistency, which may be heavily dependent on our particular logic.



The kind of notion we use will most likely be correlated to the kind of traditional cut elimination available to us in that logic.

Let us motivate the particular notion we use for our logic, while at the same time paying attention to the principles involved and what possible variations are needed for slightly different logics.

Let us consider an LDS-consistent theory  $\Delta = (D, <, \mathbf{f}, d, U)$  and an arbitrary  $B$ . We want a process which will allow us to add either  $B$  or  $\sim B$  at node  $t$  (i.e. form  $\Delta_{t:B}$  or  $\Delta_{t:\sim B}$ ) and make sure that the resulting theory is LDS-consistent. We do not have the cut rule and so we must use a notion of local consistency particularly devised for our particular logic.

Let us try the simplest notion, that of classical logic consistency. We check whether  $\mathbf{f}(t) \cup \{B\}$  is classically consistent. If yes, take  $\Delta_{t:B}$ , otherwise take  $\Delta_{t:\sim B}$ . However, the fact that we have classical consistency at  $t$  does not necessarily imply that the database is LDS-consistent. Consider  $\{t : FA \rightarrow \sim B, s : A; t < s\}$ . Here  $\mathbf{f}(t) = FA \rightarrow \sim B$ .  $\mathbf{f}(s) = A$ . If we add  $B$  to  $t$  then we still have local consistency but as soon as we apply the  $F$  introduction rule at  $t$  we get inconsistency.

Suppose we try to be more sophisticated. Suppose we let  $\mathbf{f}_*(t) = \mathbf{f}(t) \cup \{F^k X \mid X \in \mathbf{f}(s), \text{ for some } k, t <^k s\} \cup \{Y \mid G^k Y \in \mathbf{f}(s), \text{ for some } k, s <^k t\}$  and try to add either  $B$  or  $\sim B$  to  $\mathbf{f}_*(t)$  and check for classical consistency. If neither is consistent then for some  $X_i, Y_k, X'_i, Y'_j$  we have

$$\begin{aligned} \mathbf{f}(t), X_i, Y_j &\vdash \sim B \\ \mathbf{f}(t), X'_i, Y'_j &\vdash B. \end{aligned}$$

Hence  $\mathbf{f}_*(t)$  is inconsistent. However, we do have LDS rules that can bring the  $X$  and  $Y$  into  $\mathbf{f}(t)$  which will make  $\Delta$  LDS-inconsistent.

So at least one of  $B$  and  $\sim B$  can be added. Suppose  $\Delta_{t:B}$  is locally consistent. Does that make it LDS-consistent?

Well, not necessarily. Consider

$$\{t < s; t : F^k A \rightarrow \sim B; s : A\}$$

and assume we have a condition  $\partial_1$  on the diagrams

$$\partial_1 = \forall xy(x < y \rightarrow \exists z(x < z < y)).$$

We would have to apply the diagram rule  $k$  times at the appropriate labels to get inconsistency.

It is obvious from the above discussion that to add  $B$  or  $\sim B$  at node  $t$  we have to make it consistent with all possible wffs that can be proved using LDS rules to have label  $t$  (i.e. to be at  $t$ ).

Let us define then, for  $t$  in  $\Delta$ ,

$$\Delta_{(t)} = \{A \mid \Delta \vdash t : A\}.$$

We say  $B$  is locally consistent with  $\Delta$  at  $t$ , or that  $\mathbf{f}(t) \cup \{B\}$  is locally consistent iff it is classically consistent with  $\Delta_{(t)}$ . Note that if  $\Delta_{(t)} \vdash A$  classically then  $\Delta \vdash t : A$ .

We can now proceed with the construction:

1. if  $t_n \notin D_1^n$  then  $\Delta_{n+1} = \Delta_n$ ;
2. if  $t_n \in D_1^n$  and  $\mathbf{f}^n(t_n) \cup \{B_n\}$  is locally consistent, then let  $\mathbf{f}^{n+1}(t_n) = \mathbf{f}^n(t_n) \cup \{B_n\}$ .  
Otherwise  $\mathbf{f}^n(t_n)$  is locally consistent with  $\sim B_n$  and we let  $\mathbf{f}^{n+1}(t_n) = \mathbf{f}^n(t_n) \cup \{\sim B_n\}$ . For other  $x \in D_1^n$ , let  $\mathbf{f}^{n+1}(x) = \mathbf{f}^n(x)$ . Let  $\Delta_{n+1} = (D^n, <, d, \mathbf{f}^{n+1}, U^n)$ .

We must show the following:

- if  $\Delta$  is LDS-consistent and  $\Delta_{t_i:B_i}$  is obtained from  $\Delta$  by simultaneously adding the wffs  $B_i$  to  $\mathbf{f}(t_i)$  of  $\Delta$  and if for all  $i$   $\mathbf{f}(t_i) \cup \{B_i\}$  is *locally*-consistent, then  $\Delta_{t_i:B_i}$  is LDS-consistent.

To show this, assume that  $\Delta_{t_i:B_i}$  is LDS-inconsistent. We show by induction on the complexity of the inconsistency proof that  $\Delta$  is also inconsistent.

**Case one step**

In this case  $\Delta_{t_i:B_i}$  is immediately inconsistent. It is clear that  $\Delta$  is also inconsistent, because the immediate inconsistency cannot be at any label  $t_i$ , and the diagram is consistent.

**Case  $(l + 1)$  steps**

Consider the proof of inconsistency of  $\Delta_{t_i:B_i}$ . Let  $\pi$  be the first proof step leading to this inconsistency. Let  $\Delta' = (\Delta_{t_i:B_i})_\pi$ .  $\pi$  can be one of several cases as listed in Definition 59. If  $\pi$  does not touch the labels  $t_i$ , then it can commute with the insertion of  $B_i$ , i.e.  $\Delta' = (\Delta_\pi)_{t_i:B_i}$  and by the induction hypothesis  $\Delta_\pi$  is LDS-inconsistent and hence so is  $\Delta$ .

If  $\pi$  does affect some label  $t_i$ , we have to make a case analysis:

- $\pi = [t_i < s; s : A]$ , i.e.  $\diamond A$  is put in  $t$ . In this case consider  $(\Delta_\pi)_{t_i:B_i}$ . If adding  $B_i$  is still locally consistent then by the induction hypothesis  $(\Delta_\pi)_{t_i:B_i}$  is consistent. But this is  $\Delta'$  and so  $\Delta'$  is consistent. If adding  $B_i$  is locally inconsistent, this means that  $\Delta_{(t_i)} \cup \{B_i\}$  classically proves  $\perp$ , contrary to assumption.
- $\pi = [s < t_i; s : GA]$  (resp.  $\pi = [t_i < s; s : HA]$ ) i.e.  $A$  is put in  $t_i$ . The reasoning is similar to the previous case.
- $\pi = [t_i : \square A]$ ,  $\square \in \{G, H\}$ , the reasoning is similar to previous cases.
- $\pi = [t_i \vdash A]$ , similar to the previous ones.
- $\pi =$  classical quantifier rules. This case is also similar to previous ones;

- $\pi = \text{visa rule}$ . This means some new constants are involved in the new inconsistency from  $\Delta_{(t_i)} \cup \{B_i\}$ . These will turn into universal quantifiers and contradict the assumptions.
- The inconsistency rules are not a problem.
- The diagram rule does not affect  $t_i$ .
- $\pi$  applies to  $B_i$  itself. Assume that  $B_i = \Box C_i$  where  $\Box$  is  $G$  (resp.  $H$ ) and that  $C_i$  is put in some  $s, t_i < s$  (resp.  $s < t_i$ ).

Let us first check whether  $C_i$  is locally consistent at  $s$ . This will not be the case if  $\Delta_{(s)} \vdash \sim C_i$ . This would imply  $\Delta_{(t)} \vdash \Diamond \sim C_i$  contradicting the fact that  $B_i$  is consistent with  $\Delta_{(t)}$ . Thus consider  $\Delta_{t_i; B_i; s; C_i}$ . This is the same as  $\Delta'$ . It is *LDS*-inconsistent, by a shorter proof; hence by the induction hypothesis  $\Delta$  is *LDS*-consistent;

- $B_i$  is  $\Diamond C_i$  and  $\pi$  eliminates  $\Diamond$ , i.e. a new point  $s$  is introduced with  $t_i < s$  for  $\Diamond = F$  (resp.  $s < t_i$  for  $\Diamond = P$ ) and  $s : C_i$  is added to the database.

We claim that  $C_i$  is locally consistent in  $s$ , in  $\Delta^+$ , where  $\Delta^+$  is the result of adding  $s$  to  $\Delta$  but not adding  $s : C_i$ . Otherwise  $\Delta_{(s)}^+ \vdash \sim C_i$ , and since  $s$  is a completely new constant and  $\Delta^+ \vdash s : \sim C_i$  this means that  $\Delta \vdash t_i : \Box \sim C_i$ , a contradiction. Hence  $C_i$  is locally consistent in  $\Delta_{(s)}^+$ . Hence by the induction hypothesis if  $\Delta_{s; C_i; t_i; B_i}^+$  is *LDS*-inconsistent so is  $\Delta^+$ . If  $\Delta^+$  is *LDS*-inconsistent then  $\Delta^+ \vdash s : \sim C_i$  and hence  $\Delta \vdash t : \Box \sim C_i$  contradicting the local consistency of  $B_i$  at  $t$ .

- $\pi$  applies to  $B_i$  and  $\pi$  adds  $\Diamond B_i$  to  $s < t_i$  for  $\Diamond = F$  (resp.  $t_i < s$  for  $\Diamond = P$ ).

Again we claim  $\Diamond B_i$  is locally consistent at  $s$ . Otherwise  $\Delta_{(s)} \vdash \Box \sim B_i$  and so  $B_i$  would not be locally consistent at  $t_i$ . We now consider  $\Delta_{t_i; B_i; s; \Diamond B_i}$  and get a contradiction as before.

- $\pi$  is a use of a local classical rule, i.e.  $\mathbf{f}(t) \cup \{B_i\} \vdash C_i$ , and  $\pi$  adds  $C_i$  at  $t_i$ .

We claim we can add  $B_i \wedge C_i$  at  $t$ , because it is locally consistent. Otherwise  $\Delta_{(t)} \vdash B_i \rightarrow \sim C_i$  contradicting consistency of  $\Delta_{(t)} \cup \{B_i\}$  since  $\mathbf{f}(t) \vdash B_i \rightarrow C_i$ .

- $\pi$  is a Skolemization on  $B_i$  or an instantiation from  $B_i$ . All these classical operations are treated as in the previous case.

**Case  $k_n = 1$** 

1. If  $t_n, t'_n \in D_1^n$  and  $t_n < t'_n \in D_2^n$  and  $B_n \in \mathbf{f}^n(t'_n)$  then let  $\mathbf{f}^{n+1}(t_n) = \mathbf{f}^n(t_n) \cup \{\diamond B_n\}$  and  $\mathbf{f}^{n+1}(x) = \mathbf{f}^n(x)$ , for  $x \neq t_n$ . Let  $\Delta_{n+1} = (D^n, <, \mathbf{f}^{n+1}, d, U^n)$ .
2. Otherwise let  $\Delta_{n+1} = \Delta_n$ .

**Case  $k_n = 2$** 

1. If  $t_n \in D_1^n$  and  $B_n = \diamond C \in \mathbf{f}^n(t_n)$ , then let  $s$  be a completely new constant and let  $D_1^{n+1} = D_1^n \cup \{s\}$ ,  $D_2^{n+1} = D_2^n \cup \{t_n < s\}$  for  $\diamond = F$  (resp.  $s < t_n$  for  $\diamond = P$ ). Let  $\mathbf{f}^{n+1}$  be like  $\mathbf{f}^n$  or  $D_1$  and let  $\mathbf{f}^{n+1}(s) = \{C\}$ . Let  $\Delta_{n+1} = (D^{n+1}, <, \mathbf{f}^{n+1}, d, U^n)$  and let the new domain at  $s, U_s^{n+1}$ , contain all free variables of  $C$ .
2. Otherwise let  $\Delta_{n+1} = \Delta_n$ .

**Case  $k_n = 3$** 

1. If  $t_n, t'_n \in D_1^n$  and  $t_n < t'_n \in D_2^n$  and  $B_n = \square C$  and  $\square = G$  (resp.  $t'_n < t_n$  and  $\square = H$ ) and  $B_n \in \mathbf{f}^n(t_n)$  and all free variables of  $C$  are in the domain  $U_{t'_n}^n$  then let  $\mathbf{f}^{n+1} = \mathbf{f}^n(x)$ , for  $x \neq t'_n$  and  $\mathbf{f}^{n+1}(t'_n) = \mathbf{f}^n(x) \cup \{C\}$ .  
Let  $\Delta_{n+1} = (D^n, <, \mathbf{f}^{n+1}, d, U^n)$ .
2. Otherwise let  $\Delta_{n+1} = \Delta_n$ .

**Case  $k_n = 4$** 

1. We have  $t_n \in D_1^n$  and  $\mathbf{f}^n(t_n) \vdash B_n$  classically. Let  $\mathbf{f}^{n+1}(x) = \mathbf{f}^n(x)$  for  $x \neq t_n$  and let  $\mathbf{f}^{n+1}(t_n) = \mathbf{f}^n(t_n) \cup \{B_n\}$ .  
Let  $\Delta_{n+1} = (D^n, <, \mathbf{f}^{n+1}, d, U^n)$ .
2. Otherwise let  $\Delta_{n+1} = \Delta_n$ .

**Case  $k_n = 5$** 

1.  $t_n \in D_1^n$  and  $B_n = \forall x C(x) \in \mathbf{f}^n(t_n)$  and  $\alpha_n \in U^n$ . Then let  $\mathbf{f}^{n+1}(t_n) = \mathbf{f}^n(t_n) \cup \{C(\alpha_n)\}$  and  $\mathbf{f}^{n+1}(x) = \mathbf{f}^n(x)$  for  $x \neq t_n$  and  $\Delta_{n+1} = (D^n, <, \mathbf{f}^{n+1}, d, U^n)$ .
2. Otherwise let  $\Delta_{n+1} = \Delta_n$ .

**Case  $k_n = 6$** 

1.  $t_n \in D_1^n$  and  $B_n \in \mathbf{f}^n(t_n)$  and  $B_n = \exists u C(u, y_1, \dots, y_k)$ .

Let  $c^{t_n}(y_1, \dots, y_k)$  be a completely new Skolem function of this arity not appearing in  $\Delta_n$  and let

$$\mathbf{f}^{n+1}(t_n) = \mathbf{f}^n(t_n) \cup \{C(c^{t_n}(y_1, \dots, y_k), y_1, \dots, y_k)\}.$$

This is consistent by classical logic. Let  $U_{t_n}^{n+1} = U_{t_n}^n \cup \{c^{t_n}(y_1, \dots, y_k)\}$ .

Let  $U^{n+1}$  and  $\mathbf{f}^{n+1}$  be the same as  $U^n$  and  $\mathbf{f}^n$  respectively, for  $x \neq t_n$ .

Take  $\Delta_{n+1} = (D^n, <, \mathbf{f}^{n+1}, d, U^{n+1})$ .

2. Otherwise let  $\Delta_{n+1} = \Delta_n$ .

**Case  $k_n = 7$** 

1. If  $t_n, t'_n < s_n \in D_1^n$  and  $t_n < t'_n < s_n \in D_2^n$  and  $\alpha_n \in U_{t_n}^n \cap U_{s_n}^n$  then let  $U_{t'_n}^{n+1} = U_{t'_n}^n \cup \{\alpha_n\}$  and  $U_x^{n+1} = U_x^n$  for  $x \neq t'_n$ . Let  $\Delta_{n+1} = (D^n, <, \mathbf{f}^n, d, U^{n+1})$ .

2. Otherwise let  $\Delta_{n+1} = \Delta_n$ .

Let  $\Delta_\infty$  be defined by  $D_i^\infty = \bigcup_n D_i^n$ ,  $\mathbf{f}^\infty = \bigcup_n \mathbf{f}^n$ ,  $U^\infty = \bigcup_n U^n$ .

$\Delta_\infty$  is our Henkin model. Let  $\mathbf{n} = (S, R, a, V, h, g)$ , where

$$S = D_1^\infty$$

$$R = \{(x, y) \mid x < y \in D_2^\infty\}$$

$$a = d$$

$$V_t = U_t$$

$$V = U^\infty$$

$$g = \text{identity};$$

then  $h(t, P) = \{(x_1, \dots, x_n) \mid P(x_1, \dots, x_n) \in \mathbf{f}^\infty(t)\}$  for  $t \in S$  and  $P$   $n$ -place atomic predicate, and  $x_1, \dots, x_n \in V_t$ .  $\blacksquare$

**LEMMA 63.**

1.  $\mathbf{n}$  is an acceptable structure of the semantics.
2. For any  $t$  and  $B$ ,  $t \vDash B$  iff  $B \in \mathbf{f}^\infty(t)$ .

**Proof.**

1. We need to show that  $R$  is irreflexive and transitive. This follows from the construction and the diagram rule.

2. We prove this by induction. Assume  $\diamond A \in \mathbf{f}^\infty(t)$ . Then for some  $n$  we have  $t \in D_1^n$  and  $\diamond A \in \mathbf{f}^n(t)$ . Thus at some  $n' \geq n$  we put  $s \in D_1^{n'}, t < s \in D_2^{n'}$  for  $\diamond = F$  (resp.  $s < t$  for  $\diamond = P$ ) and  $A \in \mathbf{f}^{n'}(s)$ .

Assume  $\diamond A \notin \mathbf{f}^\infty(t)$ . At some  $n$ ,  $B_n = \diamond A$  and had  $\mathbf{f}^n(t) \cup \{\diamond A\}$  been consistent,  $B_n$  would have been put in  $\mathbf{f}^{n+1}$ .

Hence  $\square \sim A \in \mathbf{f}^{n+1}$ .

Assume  $t < s$  (resp.  $s < t$ ),  $s \in D_1^\infty$ . Hence for some  $n'' \geq n'$  we have  $\delta(n'') = (t, \square \sim A, s, -, 3)$ . At this stage  $\sim A$  would be in  $\mathbf{f}^{n''+1}(s)$ . Thus for all  $s \in S$  such that  $tRs$  (resp.  $sRt$ ) we have  $s \vDash \sim A$ .

The classical cases follow the usual Henkin proof.

This completes the proof of the lemma and the proof of Theorem 62. ■

#### 5.4 Label-dependent connectives

We saw earlier that *since* and *until* cannot be defined from  $\{G, H, F, P\}$  but if we allow names for worlds we can write

$$\sim q \wedge Gq \wedge Hq \rightarrow [U(A, B) \leftrightarrow F(A \wedge H(P \sim q \rightarrow B))].$$

We can introduce the label-dependent connectives  $G^x A, H^x A$  meaning

$$t \vDash G^x A \text{ iff for all } y(t < y < x \rightarrow y \vDash A);$$

$$t \vDash H^x A \text{ iff for all } y(x < y < t \rightarrow y \vDash A).$$

We can then define

$$t : U(A, B) \text{ as } t : F(A \wedge H^t B).$$

Let  $F^x A$  be  $\sim G^x \sim A$  and  $P^x A$  be  $\sim H^x \sim A$ . Then

$$t \vDash F^x A \text{ iff for some } y, t < y < x \text{ and } y \vDash A \text{ hold.}$$

$$t \vDash P^x A \text{ iff for some } y, x < y < t \text{ and } y \vDash A \text{ hold.}$$

Consider  $t \vDash G^x \perp$  and  $t \vDash F^x \top$ . The first holds iff  $\sim \exists y(t < y < x)$  which holds if either  $\sim(t < x)$  or  $x$  is an immediate successor of  $t$ . The second holds if  $\exists y(t < y < x)$ .

Label-dependent connectives are very intuitive semantically since they just restrict the temporal range of the connectives. There are many applications where such connectives are used. In the context of LDS such connectives are also syntactically natural and offer no additional complexity costs.

We have the option of defining two logical systems. One is an ordinary predicate temporal logic (which is not an LDS) where the connectives  $G, H$  are labelled. We call it  $LQK_t$  (next definition). This is the logic analogous

to  $Q\mathbf{K}_t$ . The other system is an LDS formulation of  $LQ\mathbf{K}_t$ . This system will have (if we insist on being pedantic) two lots of labels: labels of the LDS and labels for the connectives. Thus we can write  $t : F^x A$ , where  $t$  is an LDS label from the labelling algebra  $\mathcal{A}$  and  $x$  is a label from  $L$ ; when we give semantics, both  $t$  and  $x$  will get assigned possible worlds. So to simplify the LDS version we can assume  $L = \mathcal{A}$ .

DEFINITION 64 (The logic  $LQ\mathbf{K}_t$ ).

1. Let  $L$  be a set of labels, and for each  $x \in L$ , let  $G^x$  and  $H^x$  be temporal connectives. The language of  $LQ\mathbf{K}_t$  has the classical connectives, the traditional connectives  $G, H$  and the labelled connectives  $G^x, H^x$ , for each  $x \in L$ .
2. An  $LQ\mathbf{K}_t$  model has the form  $(S, R, a, V, h, g)$ , where  $(S, R, A, V, h)$  is a  $Q\mathbf{K}_t$  model (see Definition 55) and  $g : L \mapsto S$ , assigning a world to each label. Satisfaction for  $G^x$  (resp.  $H^x$ ) is defined by

$$(3x) \quad t \vDash_{h,g} G^x(a_1, \dots, a_n) \text{ iff for all } s \text{ such that } tRs \wedge sRg(x) \text{ such that } a_1, \dots, a_n \in B_s, \text{ we have } s \vDash_{h,g} A.$$

The mirror image condition is required for  $H^x$ .

DEFINITION 65 (LDS version of  $LQ\mathbf{K}_t$ ). Our definition is in parallel to Definitions 56–58. We have the added feature that the language  $\mathbf{L}$  of the LDS allows for the additional connectives  $F^t, P^t, G^t, H^t$ , where  $t$  is from the labelling algebra. For this reason we must modify the LDS notion of an  $LQ\mathbf{K}_t$  theory and require that all the labels appearing in the connectives of the formulae of the theory are also members of  $D_1$ , the diagram of labels of the theory.

DEFINITION 66 (LDS proof rules for  $LQ\mathbf{K}_t$ ). We modify Definition 59 as follows:

1.  $F^x$  introduction case  
For some  $t, s \in D_1, t < s < x \in D_2$  and  $A \in \mathbf{f}(s)$ ,  $\Delta'$  is the same as  $\Delta$  except that  $\mathbf{f}'(t) = \mathbf{f}(t) \cup \{F^x A\}$ .

Symbolically we write

$$\Delta' = \Delta_{[t < s < x; s : A]}^x.$$

2.  $F^x$  elimination case  
For some  $t \in D_1$  and some  $A, F^x A \in \mathbf{f}(t)$ , and for some new atomic

$s \in U$  that does not appear in  $\Delta$ , we have that  $\Delta'$  is the same as  $\Delta$  except that

$$\begin{aligned} D'_1 &= \Delta_1 \cup \{s\}. \\ D'_2 &= D_2 \cup \{t < s < x\}. \\ \mathbf{f}'(s) &= \{A\}. \end{aligned}$$

Note that it may be that  $\Delta_2 \cup \{t < s < x\}$  is inconsistent in which case  $\Delta$  is inconsistent.

3.  $G^x$  elimination case

For some  $t \in D_1$  such that  $t < s < x \in D_2$  we have  $G^x A \in \mathbf{f}(t)$  and  $\Delta'$  is like  $\Delta$  except that  $\mathbf{f}'(s) = \mathbf{f}(s) \cup \{A\}$ .

The  $P^x, H^x$  rules are the mirror images of all the above and all the other rules remain the same.

4.  $G^x, H^x$  introduction case

This case is the same as in Definition 59 except that in the text we replace

$$D_2^1 = D_2 \cup \{t < t_1\}$$

by  $D_2^1 = D_2 \cup \{t < t_1 < x\}$  for the case of  $G^x$  and the mirror image for the case of  $H^x$ .

Similarly we write  $\Delta' = \Delta_{[t:\Box^x A]}$ .

**THEOREM 67** (Soundness and completeness). *The LDS version of  $LQ\mathbf{K}_t$  is sound and complete for the proposed semantics.*

**Proof.** The soundness and completeness are proved along similar lines to the  $Q\mathbf{K}_t$  case see Theorems 61 and 62. ■

## 6 TEMPORAL LOGIC PROGRAMMING

We can distinguish two views of logic, the declarative and the imperative. The declarative view is the traditional one, and it manifests itself both syntactically and semantically. Syntactically a logical system is taken as being characterized by its set of theorems. It is not important how these theorems are generated. Two different algorithmic systems generating the same set of theorems are considered as producing the same logic. Semantically a logic is considered as a set of formulae valid in all models. The model  $\mathcal{M}$  is a static



semantic object. We evaluate a formula  $\varphi$  in a model and, if the result of the evaluation is positive (notation  $\mathcal{M} \models \varphi$ ), the formula is valid. Thus the logic obtained is the set of all valid formulae in some class  $\mathcal{K}$  of models.

In contrast to the above, the imperative view regards a logic syntactically as a dynamically generated set of theorems. Different generating systems may be considered as different logics. The way the theorems are generated is an integral part of the logic. From the semantic viewpoint, a logical formula is not *evaluated* in a model but performs *actions* on a model to get a *new* model. Formulae are accepted as valid according to what they do to models. For example, we may take  $\varphi$  to be valid in  $\mathcal{M}$  if  $\varphi(\mathcal{M}) = \mathcal{M}$ . (i.e.  $\mathcal{M}$  is a fixed point of  $\varphi$ ).

Applications of logic in computer science have mainly concentrated on the exploitation of its declarative features. Logic is taken as a language for describing properties of models. The formula  $\varphi$  is evaluated in a model  $\mathcal{M}$ . If  $\varphi$  holds in  $\mathcal{M}$  (evaluation successful) then  $\mathcal{M}$  has property  $\varphi$ . This view of logic is, for example, most suitably and most successfully exploited in the areas of databases and in program specification and verification. One can present the database as a deductive logical theory and query it using logical formulae. The logical evaluation process corresponds to the computational querying process. In program verification, for example, one can describe in logic the properties of the programs to be studied. The description plays the role of a model  $\mathcal{M}$ . One can now describe one's specification as a logical formula  $\varphi$ , and the query whether  $\varphi$  holds in  $\mathcal{M}$  (denoted  $\mathcal{M} \vdash \varphi$ ) amounts to verifying that the program satisfies the specification. These methodologies rely solely on the declarative nature of logic.

Logic programming as a discipline is also declarative. In fact it advertises itself as such. It is most successful in areas where the declarative component is dominant, e.g. in deductive databases. Its procedural features are not imperative (in our sense) but computational. In the course of evaluating whether  $\mathcal{M} \vdash \varphi$ , a procedural reading of  $\mathcal{M}$  and  $\varphi$  is used.  $\varphi$  does not imperatively act on  $\mathcal{M}$ , the declarative logical features are used to guide a procedure—that of taking steps for finding whether  $\varphi$  is true. What does not happen is that  $\mathcal{M}$  and  $\varphi$  are read imperatively, resulting in some action. In logic programming such actions (e.g. assert) are obtained by side-effects and special non-logical imperative predicates and are considered undesirable. There is certainly no conceptual framework within logic programming for allowing only those actions which have logical meaning.

Some researchers have come close to touching upon the imperative reading of logic. Belnap and Green [1994] and the later so-called data semantics school regard a formula  $\varphi$  as generating an action on a model  $\mathcal{M}$ , and changing it. See [van Benthem, 1996]. In logic programming and deductive databases the handling of integrity constraints borders on the use of logic imperatively. Integrity constraints have to be maintained. Thus one can either reject an update or do some corrections. Maintaining integrity

constraints is a form of executing logic, but it is logically *ad hoc* and has to do with the local problem at hand. Truth maintenance is another form. In fact, under a suitable interpretation, one may view any resolution mechanism as model building which is a form of execution. In temporal logic, model construction can be interpreted as execution. Generating the model, i.e. finding the truth values of the atomic predicates in the various moments of time, can be taken as a sequence of execution.

As the need for the imperative executable features of logic is widespread in computer science, it is not surprising that various researchers have touched upon it in the course of their activity. However, there has been no conceptual methodological recognition of the imperative paradigm in the community, nor has there been a systematic attempt to develop and bring this paradigm forward as a new and powerful logical approach in computing.

The area where the need for the imperative approach is most obvious and pressing is temporal logic. In general terms, a temporal model can be viewed as a progression of ordinary models. The ordinary models are what is true at each moment of time. The imperative view of logic on the other hand also involves step-by-step progression in virtual ‘time’, involving both the syntactic generation of theorems and the semantic actions of a temporal formula on the temporal model. Can the two intuitive progressions, the semantic time and the action (transaction) time, be taken as the same? In the case of temporal logic the answer is ‘yes’. We can act upon the models in the same time order as their chronological time. This means acting on earlier models first. In fact intuitively a future logical statement can be read (as we shall see) both declaratively and imperatively. Declaratively it describes what should be true, and imperatively it describes the actions to be taken to ensure that it becomes true. Since the chronology of the action sequence and the model sequence are the same, we can virtually *create* the future model by our actions. The logic USF, presented in Chapter 10 of [Gabbay *et al.*, 1994], was the first attempt at promoting the imperative view as a methodology, with a proposal for its use as a language for controlling processes.

The purpose of this section is twofold:

1. to present a practical, sensible, logic programming machine for handling time and modality;
2. to present a general framework for extending logic programming to non-classical logics.

Point 1 is the main task of this section. It is done within the framework of 2.

Horn clause logic programming has been generalized in essentially two major ways:

1. using the metalevel features of ordinary Horn clause logic to handle time while keeping the syntactical language essentially the same;
2. enriching the syntax of the language with new symbols and introducing additional computation rules for the new symbols.

The first method is basically a simulation. We use the expressive power of ordinary Horn clause logic to talk about the new features. The *Demo* predicate, the *Hold* predicate and other metapredicates play a significant role.

The second method is more direct. The additional computational rules of the second method can be broadly divided into two:

### 2.1 Rewrites

### 2.2 Subcomputations

The rewrites have to do with simplifying the new syntax according to some rules (basically eliminating the new symbols and reducing goals and data to the old Horn clause language) and the subcomputations are the new computations which arise from the reductions.

Given a temporal set of data, this set has the intuitive form:

‘ $A(x)$  is true at time  $t$ ’.

This can be represented in essentially two ways (in parallel to the two methods discussed):

1. adding a new argument for time to the predicate  $A$ , writing  $A^*(t, x)$  and working within an ordinary Horn clause computational environment;
2. leaving time as an external indicator and writing ‘ $t : A(x)$ ’ to represent the above temporal statement.

To compare the two approaches, imagine that we want to say the following:

‘If  $A(x)$  is true at  $t$ , then it will continue to be true’.

The first approach will write it as

$$\forall s(A^*(t, x) \wedge t < s \rightarrow A^*(s, x)).$$

The second approach has to talk about  $t$ . It would use a special temporal connective ‘ $G$ ’ for ‘always in the future’. Thus the data item becomes

$$t : A(x) \rightarrow GA(x).$$

It is equivalent to the following in the first approach:

$$A^*(t, x) \rightarrow \forall s[t < s \rightarrow A^*(s, x)].$$

The statement ‘ $GA$  is true at  $t$ ’ is translated as

$$\forall s(t < s \rightarrow A^*(s)).$$

The second part of this section introduces temporal connectives and wants to discover what kind of temporal clauses for the new temporal language arise in ordinary Horn clause logic when we allow time variables in the atoms (e.g.  $A^*(t, x), B^*(s, y)$ ) and allow time relations like  $t < s, t = s$  for time points. This would give us a clue as to what kind of temporal Horn clauses to allow in the second approach. The computational tractability of the new language is assured, as it arises from Horn clause computation. Skolem functions have to be added to the Horn clause language, to eliminate the  $F$  and  $P$  connectives which are existential. All we have to do is change the computational rules to rely on the more intuitive syntactical structure of the second approach.

### 6.1 Temporal Horn clauses

Our problem for this section is to start with Horn clause logic with the ability to talk about time through time coordinates, and see what expressive power in term of connectives ( $P, F, G, H$ , etc.) is needed to do the same job. We then extend Prolog with the ability to compute directly with these connectives. The final step is to show that the new computation defined for  $P, F, G, H$  is really ordinary Prolog computation modified to avoid Skolemization.

Consider now a Horn clause written in predicate logic. Its general form is of course  $\bigwedge \text{atoms} \rightarrow \text{atom}$ . If our atomic sentences have the form  $A(x)$  or  $R(x, y)$  or  $Q(x, y)$  then these are the atoms one can use in constructing the Horn clause. Let us extend our language to talk about time by following the first approach; that is, we can add time points, and allow special variables  $t, s$  to range over a flow of time  $(T, <)$  ( $T$  can be the set of integers, for example) and write  $Q^*(t, x, y)$  instead of  $Q(x, y)$ , where  $Q^*(t, x, y)$  (also written as  $Q(t, x, y)$ , abusing notation) can be read as

$$\text{‘}Q(x, y) \text{ is true at time } t\text{.’}$$

We allow the use of  $t < s$  to mean ‘ $t$  is earlier than  $s$ ’. Recall that we do not allow mixed atomic sentences like  $x < t$  or  $x < y$  or  $A(t, s, x)$  because these would read as

‘John loves 1980 at 1979’ or

‘John < 1980’ or

‘John < Mary’.

Assume that we have organized our Horn clauses in such a manner: what kind of time expressive power do we have? Notice that our expressive power is potentially increased. We are committed, when we write a formula of the form  $A(t, x)$ , to  $t$  ranging over time and  $x$  over our domain of elements. Thus our model theory for classical logic (or Horn clause logic) does not accept any model for  $A(t, x)$ , but only models in which  $A(t, x)$  is interpreted in this very special way. Meanwhile let us examine the syntactical expressive power we get when we allow for this two-sorted system and see how it compares with ordinary temporal and modal logics, with the connectives  $P, F, G, H$ .

When we introduce time variables  $t, s$  and the earlier-later relation into the Horn clause language we are allowing ourselves to write more atoms. These can be of the form

$$A(t, x, y) \\ t < s$$

(as we mentioned earlier,  $A(t, s, y), x < y, t < y, y < t$  are excluded).

When we put these new atomic new sentences into a Horn clause we get the following possible structures for Horn clauses.  $A(t, x), B(s, y)$  may also be *truth*.

$$(a0) \quad A(t, x) \wedge B(s, y) \rightarrow R(u, z). \\ \text{Here } < \text{ is not used.}$$

$$(a1) \quad A(t, x) \wedge B(s, y) \wedge t < s \rightarrow R(u, z). \\ \text{Here } t < s \text{ is used in the body but the time variable } u \text{ is not the same} \\ \text{as } t, s \text{ in the body.}$$

$$(a2) \quad A(t, x) \wedge B(s, y) \wedge t < s \rightarrow R(t, z). \\ \text{Same as (a1) except the time variable } u \text{ appears in the body as } u = t.$$

$$(a3) \quad A(t, x) \wedge B(s, y) \wedge t < s \rightarrow R(s, z). \\ \text{Same as (a1) with } u = s.$$

$$(a4) \quad A(t, x) \wedge B(s, y) \rightarrow R(t, z). \\ \text{Same as (a0) with } u = t, \text{ i.e. the variable in the head appears in the} \\ \text{body.}$$

The other two forms (b) and (c) are obtained when the head is different: (b) for time independence and (c) for a pure  $<$  relation.

$$(b) \quad A(t, x) \wedge B(s, y) \rightarrow R1(z)$$

$$(b') \quad A(t, x) \wedge B(s, y) \wedge t < s \rightarrow R1(z).$$

$$(c) \quad A(t, x) \wedge B(s, y) \rightarrow t < s.$$

$$(d) \quad A((1970, x), \text{ where } 1970 \text{ is a } \textit{constant} \text{ date.}$$

Let us see how ordinary temporal logic with additional connectives can express directly, using the temporal connectives, the logical meaning of the above sentences. Note that if time is linear we can assume that one of  $t < s$  or  $t = s$  or  $s < t$  always occur in the body of clauses because for linear time

$$\vdash \forall t \forall s [t < s \vee t = s \vee s < t],$$

and hence  $A(t, x) \wedge B(s, y)$  is equivalent to

$$(A(t, x) \wedge B(t, y)) \vee (A(t, x) \wedge B(s, y) \wedge t < s) \vee (A(t, x) \wedge B(s, y) \wedge s < t).$$

Ordinary temporal logic over linear time allows the following connectives:

$Fq$ , read: ‘ $q$  will be true’

$Pq$ , read: ‘ $q$  was true’

$$\diamond q = q \vee Fq \vee Pq$$

$$\square q = \sim \diamond \sim q$$

$\square q$  is read: ‘ $q$  is always true’.

If  $[A](t)$  denotes, in symbols, the statement that  $A$  is true at time  $t$ , then we have

$$[Fq](t) \equiv \exists s > t ([q](s))$$

$$[Pq](t) \equiv \exists s < t ([q](s))$$

$$[\diamond q](t) \equiv \exists s ([q](s))$$

$$[\square q](t) \equiv \forall s ([q](s)).$$

Let us see now how to translate into temporal logic the Horn clause sentences mentioned above.

*Case (a0)*

Statement (a0) reads

$$\forall t \forall s \forall u [A(t, x) \wedge B(s, y) \rightarrow R(u, z)].$$

If we push the quantifiers inside we get

$$\exists t A(t, x) \wedge \exists s B(s, y) \rightarrow \forall u R(u, z),$$

which can be written in the temporal logic as

$$\diamond A(x) \wedge \diamond B(y) \rightarrow \square R(z).$$

If we do not push the  $\forall u$  quantifier inside we get  $\square(\diamond A(x) \wedge \diamond B(y) \rightarrow R(z))$ .

*Case (a1)*

The statement (a1) can be similarly seen to read (we do not push  $\forall t$  inside)

$$\forall t\{A(t, x) \wedge \exists s[B(s, y) \wedge t < s] \rightarrow \forall uR(u, z)\}$$

which can be translated as:  $\Box\{A(x) \wedge FB(y) \rightarrow \Box R(z)\}$ . Had we pushed  $\forall t$  to the antecedent we would have got

$$\exists t[A(t, x) \wedge \exists s(B(s, y) \wedge t < s)] \rightarrow \forall uR(u, z),$$

which translates into

$$\Diamond[A(x) \wedge FB(y)] \rightarrow \Box R(z).$$

*Case (a2)*

The statement (a2) can be rewritten as

$$\forall t[A(t, x) \wedge \exists s(B(s, y) \wedge t < s) \rightarrow R(t, z)],$$

and hence it translates to  $\Box(A(x) \wedge FB(y) \rightarrow R(z))$ .

*Case (a3)*

Statement (a3) is similar to (a2). In this case we push the external  $\forall t$  quantifier in and get

$$\forall s[\exists t[A(t, x) \wedge t < s] \wedge B(s, y) \rightarrow R(s, z)],$$

which translates to

$$\Box[PA(x) \wedge B(y) \rightarrow R(z)].$$

*Case (a4)*

Statement (a4) is equivalent to

$$\forall t[A(t, x) \wedge \exists sB(s, y) \rightarrow R(t, z)],$$

and it translates to

$$\Box(A(x) \wedge \Diamond B(y) \rightarrow R(z)).$$

*Case (b)*

The statement (b) is translated as

$$\Diamond(A(x) \wedge \Diamond B(y)) \rightarrow R1(z).$$

*Case (b')*

The statement (b') translates into

$$\diamond(A(x) \wedge FB(y)) \rightarrow R1(z).$$

*Case (c)*

Statement (c) is a problem. It reads  $\forall t \forall s [A(t, x) \wedge B(s, y) \rightarrow t < s]$ ; we do not have direct connectives (without negation) to express it. It says for any two moments of time  $t$  and  $s$  if  $A(x)$  is true at  $t$  and  $B(y)$  true at  $s$  then  $t < s$ . If time is linear then  $t < s \vee t = s \vee s < t$  is true and we can write the conjunction

$$\sim \diamond(A(x) \wedge PB(y)) \wedge \diamond(A(x) \wedge B(y)).$$

Without the linearity of time how do we express the fact that  $t$  *should be* '<-related to  $s$ '?

We certainly have to go beyond the connectives  $P, F, \diamond, \square$  that we have allowed here.

*Case (d)*

$A(1970, x)$  involves a constant, naming the date 1970. The temporal logic will also need a propositional constant  $1970$ , which is true *exactly* when the time is 1970, i.e.

$$\mathbf{M}t \models 1970 \text{ iff } t = 1970.$$

Thus (d) will be translated as  $\square(1970 \rightarrow A(x))$ .  $1970$  can be read as the proposition 'The time now is 1970'.

The above examples show what temporal expressions we can get by using Horn clauses with time variables as an *object* language. We are not discussing here the possibility of 'simulating' temporal logic in Horn clauses by using the Horn clause as a *metalanguage*. Horn clause logic can do that to any logic as can be seen from Hodges [Hodges, 1985].

**DEFINITION 68.** The language contains  $\wedge, \rightarrow, F$  (it will be the case)  $P$  (it was the case) and  $\square$  (it is always the case).

We define the notions of:

Ordinary clauses;

Always clauses;

Heads;

Bodies;

Goals.



1. A *clause* is either an always clause or an ordinary clause.
2. An *always clause* is  $\Box A$  where  $A$  is an ordinary clause.
3. An *ordinary clause* is a head or an  $A \rightarrow H$  where  $A$  is a body and  $H$  is a head.
4. A *head* is either an atomic formula or  $FA$  or  $PA$ , where  $A$  is a conjunction of ordinary clauses.
5. A *body* is an atomic formula, a conjunction of bodies, an  $FA$  or a  $PA$ , where  $A$  is a body.
6. A goal is any body.

EXAMPLE 69.

$$a \rightarrow F((b \rightarrow Pq) \wedge F(a \rightarrow Fb))$$

is an acceptable clause.

$$a \rightarrow \Box b$$

is not an acceptable clause.

The reason for not allowing  $\Box$  in the head is computational and not conceptual. The difference between a (temporal) logic programming machine and a (temporal) automated theorem prover is *tractability*. Allowing disjunctions in heads or  $\Box$  in heads crosses the boundary of tractability. We can give computational rules for richer languages and we will in fact do so in later sections, but we will lose tractability; what we will have then is a theorem prover for full temporal logic.

EXAMPLES 70.

$$P[F(FA(x) \wedge PB(y) \wedge A(y)) \wedge A(y) \wedge B(x)] \rightarrow \\ P[F(A(x) \rightarrow FP(Q(z) \rightarrow A(y)))]$$

is an ordinary clause. So is  $a \rightarrow F(b \rightarrow Pq) \wedge F(a \rightarrow Fb)$ , but not  $A \rightarrow \Box b$ .

First let us check the expressive power of this *temporal Prolog*. Consider

$$a \rightarrow F(b \rightarrow Pq).$$

This is an acceptable clause. Its predicate logic reading is

$$\forall t[a(t) \rightarrow \exists s > t[b(s) \rightarrow \exists u < sq(u)]].$$

Clearly it is more directly expressive than the Horn clause *Prolog* with time variables. Ordinary Prolog can rewrite the above as

$$\forall t(a(t) \rightarrow \exists s(t < s \wedge (b(s) \rightarrow \exists u(u < s \wedge q(u))))),$$

which is equivalent to

$$\forall t(\exists s\exists u(a(t) \rightarrow t < s \wedge (b(s) \rightarrow u < s \wedge q(u)))).$$

If we Skolemize with  $\mathbf{s}_0(t)$  and  $\mathbf{u}_0(t)$  we get the clauses

$$\forall t[(a(t) \rightarrow t < \mathbf{s}_0(t)) \wedge (a(t) \wedge b(\mathbf{s}_0(t)) \rightarrow \mathbf{u}_0(t) < \mathbf{s}_0(t)) \wedge \\ (a(t) \wedge b(\mathbf{s}_0(t)) \rightarrow q(\mathbf{u}_0(t)))].$$

The following are representations of some of the problematic examples mentioned in the previous section.

$$(a1) \quad \Box(A(x) \wedge FB(y) \rightarrow \Box R(z)).$$

This is not an acceptable always clause but it can be equivalently written as

$$\Box(\Diamond(A(x) \wedge FB(y)) \rightarrow R(z)).$$

$$(a2) \quad \Box(A(x) \wedge FB(y) \rightarrow R(z)).$$

$$(b') \quad \Diamond(A(x) \wedge FB(y)) \rightarrow R1(z).$$

(b) can be written as the conjunction below using the equation

$$\begin{aligned} \Diamond q &= Fq \vee Pq \vee q : \\ (A(x) \wedge FB(y) \rightarrow R1(z)) \wedge \\ (F(A(x) \wedge FB(y)) \rightarrow R1(z)). \end{aligned}$$

(a2) Can be similarly written.

(c) can be written as

$$\forall t, s(A(x)(t) \wedge B(y)(s) \rightarrow t < s).$$

This is more difficult to translate. We need negation as failure here and write

$$\begin{aligned} \Box(A(x) \wedge PB(y) \rightarrow \perp) \\ \Box(A(x) \wedge B(y) \rightarrow \perp) \end{aligned}$$

From now on we continue to develop the temporal logic programming machine.

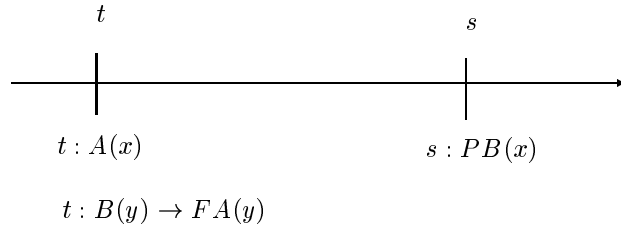


Figure 2. A temporal configuration.

### 6.2 LDS—Labelled Deductive System

This section will use the labelled deductive methodology of the previous section as a framework for developing the temporal Prolog machine. We begin by asking ourselves what is a temporal database? Intuitively, looking at existing real temporal problems, we can say that we have information about things happening at different times and some connections between them. Figure 2 is such an example.

The diagram shows a finite set of points of time and some labelled formulae which are supposed to hold at the times indicated by the labels. Notice that we have labelled not only assertions but also Horn clauses showing dependences across times. Thus at time  $t$  it may be true that  $B$  will be true. We represent that as  $t : FB$ . The language we are using has  $F$  and  $P$  as connectives. It is possible to have more connectives and still remain within the Horn clause framework. Most useful among them are ' $t : F^s A$ ' and ' $t : P^s A$ ', reading ' $t < s$  and  $s : A$ ' and ' $s < t$  and  $s : A$ '. In words: ' $A$  will be true at time  $s > t$ '.

The temporal configuration comprises two components.

1. A (finite) graph  $(\rho, <)$  of time points and the temporal relationships between them.
2. With each point of the graph we associate a (finite) set of clauses and assertions, representing what is true at that point.

In Horn clause computational logic, there is an agreement that if a formula of the form  $A(x) \rightarrow B(x)$  appears in the database with  $x$  free then it is understood that  $x$  is universally quantified. Thus we assume  $\forall x(A(x) \rightarrow B(x))$  is in the database. The variable  $x$  is then called universal (or type 1). In the case of modal and temporal logics, we need another type of variable, called type 2 or a Skolem variable. To explain the reason, consider the item of data

$$'t : FB(x)'$$

This reads, according to our agreement,

$$'\forall xFB(x) \text{ true at } t.'$$

For example, it might be the sentence:  $t$ : 'Everyone will leave'.

The time in the future in which  $B(x)$  is true depends on  $x$ . In our example, the time of leaving depends on the person  $x$ . Thus, for a given unknown (uninstantiated)  $u$ , i.e. for a given person  $u$  which we are not yet specifying, we know there must be a point  $t_1$  of time ( $t_1$  is dependent on  $u$ ) with  $t_1 : B(u)$ . This is the time in which  $u$  leaves.

This  $u$  is by agreement not a type 1 variable. It is a  $u$  to be chosen later. Really  $u$  is a Skolem constant and we do not want to and cannot read it as  $t_1 : \forall uB(u)$ . Thus we need two types of variables. The other alternative is to make the dependency of  $t_1$  on  $u$  explicit and to write

$$t_1(x) : B(x)$$

with  $x$  a universal type 1 variable, but then the object language variable  $x$  appears in the world indices as well. The world indices, i.e. the  $t$ , are external to the formal clausal temporal language, and it is simpler not to mix the  $t$  and the  $x$ . We chose the two types of variable approach. Notice that when we ask for a goal  $?G(u)$ ,  $u$  is a variable to be instantiated, i.e. a type 2 variable. So we have these variables anyway, and we prefer to develop a systematic way of dealing with them.

To explain the role of the two types of variables, consider the following classical Horn clause database and query:

$$\begin{aligned} A(x, y) \rightarrow B(x, y) \quad ?B(u, u) \\ A(a, a). \end{aligned}$$

This means 'Find an instantiation  $u_0$  of  $u$  such that  $\forall x, y[A(x, y) \rightarrow B(x, y)] \wedge A(a, a) \vdash B(u_0, u_0)$ '. There is no reason why we cannot allow for the following

$$\begin{aligned} A(u, y) \rightarrow B(x, u) \quad ?B(u, u) \\ A(a, a). \end{aligned}$$

In this case we want to find a  $u_0$  such that

$$\forall x, y[A(u_0, y) \rightarrow B(x, u_0)] \wedge A(a, a) \vdash B(u_0, u_0)$$

or to show

$$\vdash \exists u\{\forall x, y[A(u, y) \rightarrow B(x, u)] \wedge A(a, a) \rightarrow B(u, u)\}$$

$u$  is called a type 2 (Skolem) variable and  $x, y$  are universal type 1 variables. Given a database and a query of the form  $\Delta(x, y, u)?Q(u)$ , success means  $\vdash \exists u[\forall x, y \Delta(x, y, u) \rightarrow Q(u)]$ .

The next sequence of definitions will develop the syntax of the temporal Prolog machine. A lot depends on the flow of time. We will give a general definition (Definition 73 below), which includes the following connectives:

□ Always.

$F$  It will be the case.

$P$  It was the case.

$G$  It will always be the case (not including now).

$H$  It has always been the case (up to now and not including now);.

○ Next moment of time (in particular it implies that such a moment of time exists).

● Previous moment of time (in particular it implies that such a moment of time exists).

Later on we will also deal with  $S$  (Since) and  $U$  (Until).

The flows of time involved are mainly three:

- general partial orders  $(T, <)$ ;
- linear orders;
- the integers or the natural numbers.

The logic and theorem provers involved, even for the same connectives, are different for different partial orders. Thus the reader should be careful to note in which flow of time we are operating. Usually the connectives ○ and ● assume we are working in the flow of time of integers.

Having fixed a flow of time  $(T, <)$ , the temporal machine will generate finite configurations of points of time according to the information available to it. These are denoted by  $(\rho, <)$ . We are supposed to have  $\rho \subseteq T$  (more precisely,  $\rho$  will be homomorphic into  $T$ ), and the ordering on  $\rho$  will be the same as the ordering on  $T$ . The situation gets slightly complicated if we have a new point  $s$  and we do not know where it is supposed to be in relation to known points. We will need to consider all possibilities. Which possibilities do arise depend on  $(T, <)$ , the background flow of time we are working with. Again we should watch for variations in the sequel.

DEFINITION 71. Let  $(\rho, <)$  be a finite partial order. Let  $t \in \rho$  and let  $s$  be a new point. Let  $\rho' = \rho \cup \{s\}$ , and let  $<'$  be a partial order on  $\rho'$ . Then

$(\rho', <, t)$  is said to be a (one new point) *future* (resp. *past*) *configuration* of  $(\rho, <, t)$  iff  $t <' s$  (resp.  $s <' t$ ) and  $\forall xy \in \rho(x < y \leftrightarrow x <' y)$ .

EXAMPLE 72. Consider a general partial flow  $(T, <)$  and consider the subflow  $(\rho, <)$ .

The possible future configurations (relative to  $T, <$ ) of one additional point  $s$  are displayed in Fig. 3.

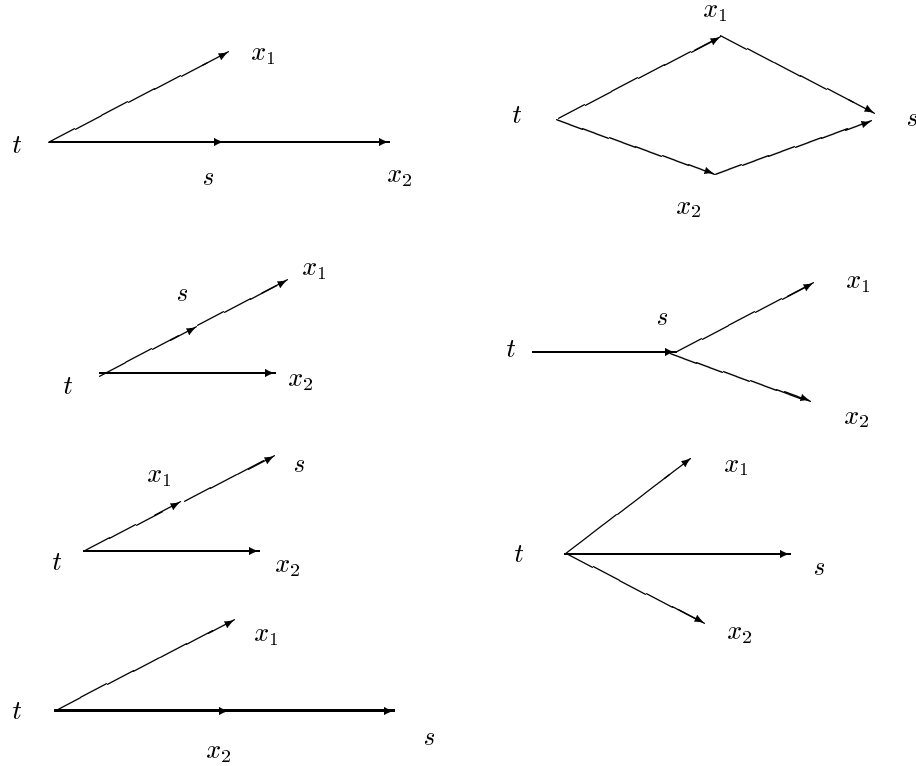


Figure 3.

For a finite  $(\rho, <)$  there is a finite number of future and past non-isomorphic configurations. This finite number is exponential in the size of  $\rho$ . So in the general case without simplifying assumptions we will have an intractable exponential computation. A configuration gives all possibilities of putting a point in the future or past.

In the case of an ordering in which a next element or a previous element exists (like  $t + 1$  and  $t - 1$  in the integers) the possibilities for configurations

are different. In this case we must assume that we know the *exact* distance between the elements of  $(\rho, <)$ .

For example, in the configuration  $\{t < x_1, t < x_2\}$  of Fig. 4 we may have the further following information as part of the configuration:

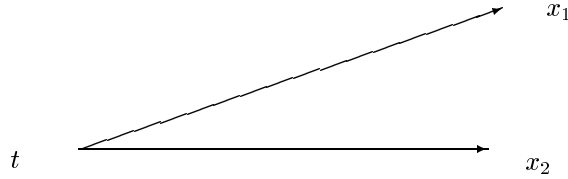


Figure 4.

$$t = \bullet^{18} x_1$$

$$t = \bullet^6 x_2$$

so that we have only a finite number of possibilities for putting  $s$  in.

Note that although  $\bullet$  operates on propositions, it can also be used to operate on points of time, denoting the predecessor function.

**DEFINITION 73.** Consider a temporal Prolog language with the following connectives and predicates:

1. atomic predicates;
2. function symbols and constants;
3. two types of variables:  
 universal variables (type 1)  $V = \{x_1, y_1, z_1, x_2, y_2, z_2, \dots\}$   
 and Skolem variables (type 2)  $U = \{u_1, v_1, u_2, v_2, \dots\}$ ;
4. the connectives  $\wedge, \rightarrow, \vee, F, P, \bigcirc, \bullet, \square$  and  $\neg$ .

$FA$  reads: it will be the case that  $A$ .

$PA$  reads: it was the case that  $A$ .

$\bigcirc A$  reads:  $A$  is true tomorrow (if a tomorrow exists; if tomorrow does not exist then it is false).

$\bullet A$  reads:  $A$  was true yesterday (if yesterday does not exist then it is false).

$\neg$ : represents negation by failure.

We define now the notions of an *ordinary clause*, an *always clause*, a *body*, a *head* and a *goal*.

1. A *clause* is either an always clause or an ordinary clause.
2. An *always clause* has the form  $\Box A$ , where  $A$  is an ordinary clause.
3. An *ordinary clause* is a head or an  $A \rightarrow H$ , where  $A$  is a body and  $H$  is a head.
4. A *head* is an atomic formula or an  $FA$  or a  $PA$  or an  $\bigcirc A$  or an  $\bullet A$ , where  $A$  is a finite conjunction of ordinary clauses.
5. A *body* is an atomic formula or an  $FA$  or a  $PA$  or an  $\bigcirc A$  or an  $\bullet A$  or  $\neg A$  or a conjunction of bodies where  $A$  is a body.
6. A *goal* is a body whose variables are *all* Skolem variables.
7. A disjunction of goals is also a goal.

REMARK 74. Definition 73 included all possible temporal connectives. In practice different systems may contain only some of these connectives. For example, a modal system may contain only  $\diamond$  (corresponding to  $F$ ) and  $\Box$ . A future discrete system may contain only  $\bigcirc$  and  $F$  etc.

Depending on the system and the flow of time, the dependences between the connectives change. For example, we have the equivalence

$$\Box(a \rightarrow \bigcirc b) \text{ and } \Box(\bullet a \rightarrow b)$$

whenever both  $\bullet a$  and  $\bigcirc b$  are meaningful.

DEFINITION 75. Let  $(T, <)$  be a flow of time. Let  $(\rho, <)$  be a finite partial order. A *labelled temporal database* is a set of labelled ordinary clauses of the form  $(t_i : A_i)$ ,  $t \in \rho$ , and always clauses of the form  $\Box A_i$ ,  $A_i$  a clause. A labelled goal has the form  $t : G$ , where  $G$  is a goal.

$\Delta$  is said to be a labelled temporal database over  $(T, <)$  if  $(\rho, <)$  is homomorphic into  $(T, <)$ .

DEFINITION 76. We now define the computation procedure for the temporal Prolog for the language of Definitions 73 and 75. We assume a flow of time  $(T, <)$ .  $\rho \subseteq T$  is the finite set of points of time involved so far in the computation. The exact computation steps depend on the flow of time. It is different for branching, discrete linear, etc. We will give the definition for linear time, though not necessarily discrete. Thus the meaning of  $\bigcirc A$  in this logic is that there exists a next moment and  $A$  is true at this next moment. Similarly for  $\bullet A$ .  $\bullet A$  reads: there exists a previous moment and  $A$  was true at that previous moment.

We define the success predicate  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0, \Theta)$  where  $t \in \rho$ ,  $(\rho, <)$  is a finite partial order and  $\Delta$  is a set of labelled clauses  $(t : A)$ ,  $t \in \rho$ .

$\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0, \Theta)$  reads: the labelled goal  $t : G$  succeeds from  $\Delta$  under the substitution  $\Theta$  to all the type 2 variables of  $G$  and  $\Delta$  in the computation with starting labelled goal  $t_0 : G_0$ .



When  $\Theta$  is known, we write  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)$  only.

We define the simultaneous success and failure of a set  $\mathbf{\Pi}$  of metapredicates of the form  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)$  under a substitution  $\Theta$  to type 2 variables. To explain the intuitive meaning of success or failure, assume first that  $\Theta$  is a substitution which grounds all the Skolem type 2 variables. In this case  $(\mathbf{\Pi}, \Theta)$  succeeds if by definition all  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0, \Theta) \in \mathbf{\Pi}$  succeed and  $(\mathbf{\Pi}, \Theta)$  fails if at least one of  $\mathbf{S} \in \mathbf{\Pi}$  fails. The success or failure of  $\mathbf{S}$  for a  $\Theta$  as above has to be defined recursively. For a general  $\Theta$ ,  $(\mathbf{\Pi}, \Theta)$  succeeds, if for some  $\Theta'$  such that  $\Theta\Theta'$  grounds all type 2 variables  $(\mathbf{\Pi}, \Theta\Theta')$  succeeds.  $(\mathbf{\Pi}, \Theta)$  fails if for all  $\Theta'$  such that  $\Theta\Theta'$  grounds all type 2 variables we have that  $(\mathbf{\Pi}, \Theta\Theta')$  fails. We need to give recursive procedures for the computation of the success and failure of  $(\mathbf{\Pi}, \Theta)$ . In the case of the recursion, a given  $(\mathbf{\Pi}, \Theta)$  will be changed to a  $(\mathbf{\Pi}', \Theta')$  by taking  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0) \in \mathbf{\Pi}$  and replacing it by  $\mathbf{S}(\rho', <', \Delta', G', t', G_0, t_0)$ . We will have several such changes and thus get several  $\mathbf{\Pi}'$  by replacing several  $\mathbf{S}$  in  $\mathbf{\Pi}$ . We write the several possibilities as  $(\mathbf{\Pi}'_i, \Theta'_i)$ . If we write  $(\mathbf{\Pi}, \Theta)$  to mean  $(\mathbf{\Pi}, \Theta)$  succeeds and  $\sim(\mathbf{\Pi}, \Theta)$  to read  $(\mathbf{\Pi}, \Theta)$  fails, then our recursive computation rules have the form:  $(\mathbf{\Pi}, \Theta)$  succeeds (or fails) if some Boolean combination of  $(\mathbf{\Pi}'_i, \Theta'_i)$  succeeds (or fails). The rules allow us to pick an element in  $\mathbf{\Pi}$ , e.g.  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)$ , and replace it with one or more elements to obtain the different  $(\mathbf{\Pi}'_i, \Theta'_i)$ , where  $\Theta'_i$  is obtained from  $\Theta$ . In case of failure we require that  $\Theta$  grounds all type 2 variables. We do not define failure for a non-grounding  $\Theta$ .

To summarize the general structure of the rules is:

$(\mathbf{\Pi}, \Theta)$  succeeds (or fails) if some Boolean combination of the successes and failures of some  $(\mathbf{\Pi}'_i, \Theta'_i)$  holds and  $(\mathbf{\Pi}, \Theta)$  and  $(\mathbf{\Pi}'_i, \Theta'_i)$  are related according to one of the following cases:

Case I If  $\mathbf{\Pi} = \emptyset$  then  $(\mathbf{\Pi}, \Theta)$  succeeds (i.e. the Boolean combination of  $(\mathbf{\Pi}'_i, \Theta'_i)$  is *truth*).

Case II  $(\mathbf{\Pi}, \Theta)$  fails if for some  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)$  in  $\mathbf{\Pi}$  we have  $G$  is atomic and for all  $\Box(A \rightarrow H) \in \Delta$  and for all  $(t : A \rightarrow H) \in \Delta, H\Theta$  does *not* unify with  $G\Theta$ . Further, for all  $\Omega$  and  $s$  such that  $t = \Omega s$  and for all  $s : A \rightarrow \Omega H$  and all  $\Box(A \rightarrow \Omega H)$  we have  $H\Theta$  does not unify with  $G\Theta$ , where  $\Omega$  is a sequence of  $\bigcirc$  and  $\bullet$ .

REMARK 77. We must qualify the conditions of the notion of failure. If we have a goal  $t : G$ , with  $G$  atomic, we know for sure that  $t : G$  finitely fails under a substitution  $\Theta$ , if  $G\Theta$  cannot unify with any head of a clause. This is what the condition above says. What are the candidates for unification? These are either clauses of the form  $t : A \rightarrow H$ , with  $H$  atomic, or  $\Box(A \rightarrow H)$ , with  $H$  atomic.

Do we have to consider the case where  $H$  is not atomic? The answer depends on the flow of time and on the configuration  $(\rho, <)$  we are dealing

with. If we have, say,  $t : A \rightarrow FG$  then if  $A \rightarrow FG$  is true at  $t$ ,  $G$  would be true (if at all) in some  $s$ ,  $t < s$ . This  $s$  is irrelevant to our query  $?t : G$ . Even if we have  $t' < t$  and  $t' : A \rightarrow FG$  and  $A$  true at  $t'$ , we can still ignore this clause because we are not assured that any  $s$  such that  $t' < s$  and  $G$  true at  $s$  would be the desired  $t$  (i.e.  $t = s$ ).

The only case we have to worry about is when the flow of time and the configuration are such that we have, for example,  $t' : A \rightarrow \bigcirc^5 G$  and  $t = \bigcirc^5 t'$ .

In this case we must add the following clause to the notion of failure: for every  $s$  such that  $t = \bigcirc^n s$  and every  $s : A \rightarrow \bigcirc^n H$ ,  $G\Theta$  and  $H\Theta$  do not unify.

We also have to check what happens in the case of always clauses.

Consider an integer flow of time and the clause  $\Box(A \rightarrow \bigcirc^5 \bullet^{27} H)$ . This is true at the point  $s = \bullet^5 \bigcirc^{27} t$  and hence for failure we need that  $G\Theta$  does not unify with  $H\Theta$ .

The above explains the additional condition on failure.

The following conditions 1–10, 12–13 relate to the success of  $(\Pi, \Theta)$  ( $\Pi'_i, \Theta'_i$ ) succeed. Condition (11) uses the notion of failure to give the success of negation by failure. Conditions 1–10, 12–13 give certain alternatives for success. They give failure if each one of these alternatives ends up in failure.

**1. Success rule for atomic query:**

$\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0) \in \Pi$  and  $G$  is atomic and for some head  $H$ ,  $(t : H) \in \Delta$  and for some substitutions  $\Theta_1$  to the universal variables of  $H$  and  $\Theta_2$  to the existential variables of  $H$  and  $G$  we have  $H\Theta_1\Theta_2 = G\Theta_1\Theta_2$  and  $\Pi' = \Pi - \{\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)\}$  and  $\Theta' = \Theta\Theta_2$ .

**2. Computation rule for atomic query:**

$\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0) \in \Pi$  and  $G$  is atomic and for some  $(t : A \rightarrow H) \in \Delta$  or for some  $\Box(A \rightarrow H) \in \Delta$  and for some  $\Theta_1, \Theta_2$ , we have  $H\Theta_1\Theta_2 = G\Theta_1\Theta_2$  and  $\Pi' = (\Pi - \{\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, A\Theta_1, t, G_0, t_0)\}$  and  $\Theta' = \Theta\Theta_2$ .

The above rules deal with the atomic case. Rules 3, 4 and 4\* deal with the case the goal is  $FG$ . The meaning of 3, 4 and 4\* is the following. We ask  $FG$  at  $t$ . How can we be sure that  $FG$  is true at  $t$ ? There are two possibilities, (a) and (b):

- (a) We have  $t < s$  and at  $s : G$  succeeds. This is rule 3;
- (b) Assume that we have the fact that  $A \rightarrow FB$  is true at  $t$ . We ask for  $A$  and succeed and hence  $FB$  is true at  $t$ . Thus there should exist a point  $s'$  in the future of  $t$  where  $B$  is true. Where can  $s'$  be? We don't know where  $s'$  is in the future of  $t$ . So we consider all future

configurations for  $s'$ . This gives us all future possibilities where  $s'$  can be. We assume for each of these possibilities that  $B$  is true at  $s'$  and check whether either  $G$  follows at  $s'$  or  $FG$  follows at  $s'$ . If we find that for all future constellations of where  $s'$  can be  $G \vee FG$  succeeds in  $s'$  from  $B$ , then  $FG$  holds at  $t$ . Here we use the transitivity of  $<$ . Rule 4a gives the possibilities where  $s'$  is an old point  $s$  in the future of  $t$ ; Rule 4b gives the possibilities where  $s'$  is a new point forming a new configuration. Success is needed from *all* possibilities.

**3. Immediate rule for  $F$ :**

$\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0) \in \mathbf{\Pi}$  and for some  $s \in \rho$  such that  $t < s$  we have  $\mathbf{\Pi}' = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, G, s, G_0, t_0)\}$  and  $\Theta' = \Theta$ .

**4. First configuration rule for  $F$ :**

$\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0) \in \mathbf{\Pi}$  and for some  $(t : A \rightarrow F \wedge_j B_j) \in \Delta$  and some  $\Theta_1, \Theta_2$  we have both (a) and (b) below are true.  $A$  may not appear in which case we pretend  $A = \text{truth}$ .

- (a) For *all*  $s \in \rho$  such that  $t < s$  we have that  
 $\mathbf{\Pi}'_s = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, E\Theta_1, t, G_0, t_0)\} \cup \{\mathbf{S}(\rho, <, \Delta \cup \{(s : B_j\Theta_1) \mid j = 1, 2, \dots\}, D, s, G_0, t_0)\}$   
 succeeds with  $\Theta'_s = \Theta\Theta_2$  and  $D = G \vee FG$  and  $E = A$ .
- (b) For all future configurations of  $(\rho, <, t)$  with a new letter  $s$ , denoted by the form  $(\rho_s, <_s)$ , we have that  
 $\mathbf{\Pi}'_s = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, E\Theta_1, t, G_0, t_0)\} \cup \{\mathbf{S}(\rho_s, <_s, \Delta \cup \{(s : B_j) \mid j = 1, 2, \dots\}, D, s, G_0, t_0)\}$   
 succeeds with  $\Theta'_s = \Theta\Theta_2$  and  $D = G \vee FG$  and  $E = A$ .

The reader should note that conditions 3, 4a and 4b are needed only when the flow of time has some special properties. To explain by example, assume we have the configuration of Fig. 5 and  $\Delta = \{t : A \rightarrow FB, t' : C\}$  as data, and our query is  $?t : FG$ .

Then according to rules 3, 4 we have to check and succeed in all the following cases:

1. from rule 3 we check  $\{t' : C, t : A \rightarrow FB\}?t' : G$ ;
2. from rule 4a we check  $\{t' : C, t : A \rightarrow FB, t' : B\}?t' : G$ ;
3. from rule 4b we check  $\{t' : C, t : A \rightarrow FB, s : B\}?s : G$ ;

for the three configurations of Fig. 6.

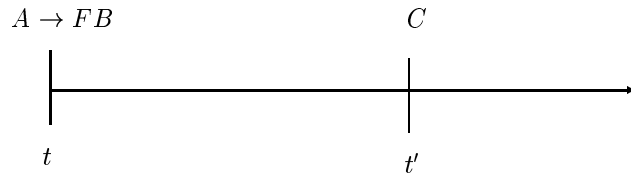


Figure 5.

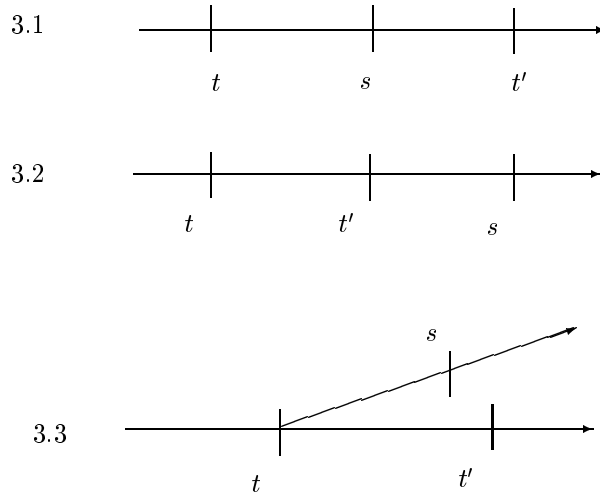


Figure 6.

If time is linear, configuration 3.3, shown in Fig. 6, does not arise and we are essentially checking 3.1, 3.2 of Fig. 6 and the case 4a corresponding to  $t' = s$ .

If we do not have any special properties of time, success in case 3.2 is required. Since we must succeed in all cases and 3.2 is the case with *least* assumptions, it is enough to check 3.2 alone.

Thus for the case of no special properties of the flow of time, case 4 can be replaced by case 4 general below:

4 **general**  $\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0) \in \mathbf{\Pi}$  and for the future configuration  $(\rho_1, <_1)$  defined as  $\rho_1 = \rho \cup \{s\}$  and  $<_1 = < \cup \{t < s\}$ ,  $s$  a new letter, we have that:  $\mathbf{\Pi}'s = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, E\Theta_1, t, G_0, t_0)\} \cup \{\mathbf{S}(\rho_1, <_1, \Delta \cup \{(s : B_j) \mid j = 1, 2, \dots\}, D, s, G_0, t_0)\}$  succeeds with  $\Theta'_s = \Theta\Theta_1$  and  $D = G \vee FG$  and  $E = A$ .

4\*. **Second configuration rule for  $F$ :**

For some  $\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0)$  and some  $\Box(A \rightarrow F \wedge_j B_j) \in \Delta$  and some  $\Theta_1\Theta_2$  we have both cases 4a and 4b above true with  $E = A \vee FA$  and  $D = G \vee FG$ .

4\* **general** Similar to (4 **general**) for the case of general flow.

5. This is the the mirror image of 3 with ' $PG$ ' replacing ' $FG$ ' and ' $s < t$ ' replacing ' $t < s$ '.

6; 6\* This is the mirror image of 4 and 4\* with ' $PG$ ' replacing ' $FG$ ', ' $s < t$ ' replacing ' $t < s$ ' and 'past configuration' replacing 'future configuration'.

6 **general** This is the image of 4 **general**.

We now give the computation rules 7-10 for  $\bigcirc$  and  $\bullet$  for orderings in which a next point and/or previous points exist. If  $t \in T$  has a next point we denote this point by  $s = \bigcirc t$ . If it has a previous point we denote it by  $s = \bullet t$ . For example, if  $(T, <)$  is the integers then  $\bigcirc t = t + 1$  and  $\bullet t = t - 1$ . If  $(T, <)$  is a tree then  $\bullet t$  always exists, except at the root, but  $\bigcirc t$  may or may not exist. For the sake of simplicity we must assume that if we have  $\bigcirc$  or  $\bullet$  in the language then  $\bigcirc t$  or  $\bullet t$  always exist. Otherwise we can sneak negation in by putting  $(t : \bigcirc A) \in \Delta$  when  $\bigcirc t$  does not exist!

7. **Immediate rule for  $\bigcirc$ :**

$\mathbf{S}(\rho, <, \Delta, \bigcirc G, t, G_0, t_0) \in \mathbf{\Pi}$  and  $\bigcirc t$  exists and  $\bigcirc t \in \rho$  and  $\Theta' = \Theta$  and  $\mathbf{\Pi}' = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, \bigcirc G, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, G, \bigcirc t, G_0, t_0)\}$ .

8. **Configuration rule for  $\bigcirc$ :**

$\mathbf{S}(\rho, <, \Delta, \bigcirc G, t, G_0, t_0) \in \mathbf{\Pi}$  and for some  $\Theta_1, \Theta_2$  some  $(t : A \rightarrow \bigcirc \wedge_j B_j) \in \Delta$  and  $\mathbf{\Pi}' = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, \bigcirc G, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, A\Theta_1, t, G_0, t_0)\} \cup \{\mathbf{S}(\rho \cup \{\bigcirc t\}, <', \Delta \cup \{(\bigcirc t : B_j)\}, G, \bigcirc t, G_0, t_0)\}$  succeeds with  $\Theta' = \Theta\Theta_2$ , and  $<'$  is the appropriate ordering closure of  $< \cup \{(t, \bigcirc t)\}$ .

Notice that case 8 is parallel to case 4. We do not need 8a and 8b because of  $\bigcirc t \in \rho$ ; then what would be case 8b becomes 7.

9. The mirror image of 7 with ' $\bullet$ ' replacing ' $\bigcirc$ '.

10. The mirror image of 8 with ' $\bullet$ ' replacing ' $\bigcirc$ '.

11. **Negation as failure rule:**

$\mathbf{S}(\rho, <, \Delta, \neg G, t, G_0, t_0) \in \mathbf{\Pi}$  and  $\Theta$  grounds every type 2 variable and the computation for success of  $\mathbf{S}(\rho, <, \Delta, G, t, \Theta)$  ends up in failure.

12. **Disjunction rule:**

$\mathbf{S}(\rho, <, \Delta, G_1 \vee G_2, t, G_0, t_0) \in \mathbf{\Pi}$  and  $\mathbf{\Pi}' = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, G_1 \vee G_2, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, G_i, t, G_0, t_0) \mid i \in \{1, 2\}\}$  and  $\Theta' = \Theta$  and  $i \in \{1, 2\}$ .

13. **Conjunction rule:**

$\mathbf{S}(\rho, <, \Delta, G_1 \wedge G_2, t, G_0, t_0) \in \mathbf{\Pi}$  and  $\mathbf{\Pi}' = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, G_1 \wedge G_2, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, G_i, t, G_0, t_0) \mid i \in \{1, 2\}\}$ .

14. **Restart rule:**

$\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0) \in \mathbf{\Pi}$  and  $\mathbf{\Pi}' = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, G_1, t_0, G_0, t_0)\}$  where  $G_1$  is obtained from  $G_0$  by substituting completely new type 2 variables  $u'_i$  for the type 2 variables  $u_i$  of  $G_0$ , and where  $\Theta'$  extends  $\Theta$  by giving  $\Theta'(u'_i) = u'_i$  for the new variables  $u'_i$ .

15. **To start the computation:**

Given  $\Delta$  and  $t_0 : G_0$  and a flow  $(T, <)$ , we start the computation with  $\mathbf{\Pi} = \{\mathbf{S}(\rho, <, \Delta, G_0, t_0, G_0, t_0)\}$ , where  $(\rho, <)$  is the configuration associated with  $\Delta$ , over  $(T, <)$  (Definition 75).

Let us check some examples.

EXAMPLE 78.

**Data:**

1.  $t : a \rightarrow Fb$

2.  $\Box(b \rightarrow Fc)$

3.  $t : a$ .

**Query:**  $?t : Fc$

**Configuration:**  $\{t\}$

Using rule 4\* we create a future  $s$  with  $t < s$  and ask the two queries (the notation  $A?B$  means that we add  $A$  to the data 1, 2, 3 and ask  $?B$ ).

4.  $?t : C \vee Fb$

and

5.  $s : c?s : c \vee Fc$

5 succeeds and 4 splits into two queries by rule 4.

6.  $?t : a$   
and  
7.  $s' : b?s' : b$ .

EXAMPLE 79.

**Data:**

1  $t : FA$

2  $t : FB$

**Query:**  $t : F\varphi$  where  $\varphi = (A \wedge B) \vee (A \wedge FB) \vee (B \wedge FA)$ .

The query will fail in any flow of time in which the future is not linear. The purpose of this example is to examine what happens when time is linear. Using 1 we introduce a point  $s$ , with  $s : A$ , and query from  $s$  the following:

$?s : \varphi \vee F\varphi$

If we do not use the restart rule, the query will fail. Now that we are at a point  $s$  there is no way to go back to  $t$ . We therefore cannot reason that we also have a point  $s' : B$  and  $t < s$  and  $t < s'$  and that because of linearity  $s = s'$  or  $s < s'$  or  $s' < s$ . However, if we are allowed to restart, we can continue and ask  $t : F\varphi$  and now use the clause  $t : FB$  to introduce  $s'$ . We now reason using linearity in rule 4 that the configurations are:

$$\begin{aligned} & t < s < s' \\ \text{or } & t < s' < s \\ \text{or } & t < s = s' \end{aligned}$$

and  $\varphi$  succeeds at  $t$  for each configuration.

The reader should note the reason for the need to use the restart rule. When time is just a partial order, the two assumptions  $t : FA$  and  $t : FB$  do not interact. Thus when asking  $t : FC$ , we know that there are two points  $s_1 : A$  and  $s_2 : B$ ; see Fig. 7.

$C$  can be true in either one of them.  $s_1 : A$  has no influence on  $s_2 : B$ . When conditions on time (such as linearity) are introduced,  $s_1$  does influence  $s_2$  and hence we must introduce both at the same time. When one does forward deduction one can introduce both  $s_1$  and  $s_2$  going forward. The backward rules do not allow for that. That is why we need the restart rule. When we restart, we keep all that has been done (with, for example,  $s_1$ ) and have the opportunity to restart with  $s_2$ . The restart rule can be used to solve the linearity problem for classical logic only. Its side-effect is that it turns intuitionistic logic into classical logic; see Gabbay's paper on *N-Prolog* [Gabbay, 1985] and [Gabbay and Olivetti, 2000]. In theorem

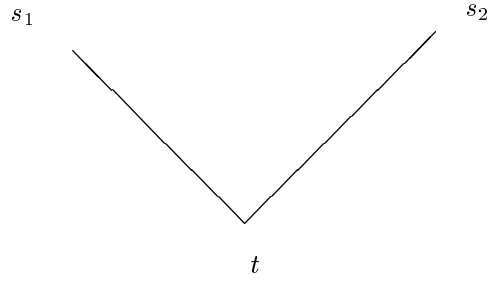


Figure 7.

proving based on intuitionistic logic where disjunctions are allowed, forward reasoning cannot be avoided. See the next example.

It is instructive to translate the above into Prolog and see what happens there.

EXAMPLE 80.

1.  $t : FA$  translates into  $(\exists s_1 > t)A^*(s_1)$ .

2.  $t : FB$  translates into  $(\exists s_2 > t)A^*(s_2)$ .

The query translates into the formula  $\psi(t)$ :

$$\psi = \exists s > t[A^*(s) \wedge B^*(s)] \vee \exists s_1 > t[A^*(s_1) \wedge \exists s_2 > s_1 B^*(s_2)] \vee \exists s_2 > t[B^*(s_2) \wedge \exists s_1 > s_2 A^*(s_2)]$$

which is equivalent to the disjunction of:

- (a)  $[t < s \wedge A^*(s) \wedge B^*(s)]$ ;
- (b)  $t < s_1 \wedge s_1 < s_2 \wedge A^*(s_1) \wedge B^*(s_2)$ ;
- (c)  $t < s_1 \wedge s_2 < s_1 \wedge A^*(s_1) \wedge B^*(s_2)$ .

All of (a), (b), (c) fail from the data, unless we add to the data the disjunction

$$\forall xy(x < y \vee x = y \vee y < x).$$

Since this is not a Horn clause, we are unable to express it in the database.

The logic programmer might add this as an integrity constraint. This is wrong as well. As an integrity constraint it would require the database to indicate which of the three possibilities it adopts, namely:



- $x < y$  is in the data;
- or  $x = y$  is in the data;
- or  $y < x$  is in the data.

This is stronger than allowing the disjunction in the data.

The handling of the integrity constraints corresponds to our metahandling of what configurations  $(\rho, <)$  are allowed depending on the ordering of time. By labelling data items we are allowing for the metalevel considerations to be done separately on the labels.

*This means that we can handle properties of time which are not necessarily expressible by an object language formula of the logic. In some cases (finiteness of time) this is because they are not first order; in other cases (irreflexivity) it is because there is no corresponding formula (axiom) and in still other cases because of syntactical restrictions (linearity).*

We can now make clear our classical versus intuitionistic distinction. If the underlying logic is classical then we are checking whether  $\Delta \vdash \psi$  in classical logic. If our underlying logic is intuitionistic, then we are checking whether  $\Delta \vdash \psi$  in intuitionistic logic where  $\Delta$  and  $\psi$  are defined below.

$\Delta$  is the translation of the data together with the axioms for linear ordering, i.e. the conjunction of:

1.  $\exists s_1 > tA^*(s_1)$ ;
2.  $\exists s_2 > tB^*(s_2)$ ;
3.  $\forall xy(x < y \vee x = y \vee y < x)$ ;
4.  $\forall x\exists y(x < y)$ ;
5.  $\forall x\exists y(y < x)$ ;
6.  $\forall xyz(x < y \wedge y < z \rightarrow x < z)$ ;
7.  $\forall x\neg(x < x)$ .

$\psi$  is the translation of the query as given above.

The computation of Example 79, using restart, answers the question  $\Delta \vdash ?\psi$  in classical logic. To answer the question  $\Delta \vdash ?\psi$  in intuitionistic logic we cannot use restart, but must use forward rules as well.

EXAMPLE 81. See Example 79 for the case that the underlying logic is intuitionistic. *Data* and *Query* as in Example 79.

Going forward, we get:

- 3  $s : A$  from 1;
- 4  $s' : B$  from 2.

By linearity,

$$t < s < s'$$

or  $t < s' < s$

or  $t < s = s'$ .

$\psi$  will succeed for each case.

Our language does not allow us to ask queries of the form  $\Box G(x)$ , where  $x$  are all universal variables (i.e.  $\forall x \Box G(x)$ ). However, such queries can be computed from a database  $\Delta$ . The *only* way to get *always* information out of  $\Delta$  for a general flow of time is via the always clauses in the database. Always clauses are true everywhere, so if we want to know what else is true everywhere, we ask it from the always clauses. Thus to ask

$?\Box G(x)$ ,  $x$  a universal variable

we first Skolemize and then ask

$\{X, \Box X \mid \Box X \in \Delta\} ? G(c)$

where  $c$  is a Skolem constant.

We can add a new rule to Definition 76:

16. **Always rule:**

$\mathbf{S}(\rho, <, \Delta, \Box G, t, G_0, t_0) \in \mathbf{\Pi}$  and

$\mathbf{\Pi}' = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, \Box G, t, G_0, t_0)\}) \cup \{\mathbf{S}(\{s\}, \emptyset, \Delta', G', s, g', s)\}$

where  $s$  is a completely new point and  $G'$  is obtained from  $G$  by substituting new Skolem constants for all the universal variables of  $G$  and

$$\Delta' = \{B, \Box B \mid \Box B \in \Delta\}.$$

We can use 16 to add another clause to the computation of Definition 76, namely:

17.  $\mathbf{S}(\rho, <, \Delta, F(A \wedge B), t, G_0, t_0) \in \mathbf{\Pi}$

and  $\mathbf{\Pi}' = (\mathbf{\Pi} - \{\mathbf{S}(\rho, <, \Delta, F(A \wedge B), t, G_0, t_0)\}) \cup \{\mathbf{S}(\rho, <, \Delta, FA, t, G_0, t_0), \mathbf{S}(\rho, <, \Delta, \Box B, t, G_0, t_0)\}$ .

EXAMPLE 82.

Data	Query	Configuration
$\Box a$	$t : F(a \wedge b)$	$\{t\}$
$t : Fb$		

**First computation**

Create  $s, t < s$  and get

Data	Query	Configuration
$\Box a$	$s : a \wedge b$	$t < s$
$t : Fb$		
$s : b$		

$s : b$  succeeds from the data.  $s : a$  succeeds by rule 2, Definition 76.

**Second computation**

Use rule 17. Since  $?\Box a$  succeeds ask for  $Fb$  and proceed as in the first computation.

6.3 *Different flows of time*

We now check the effect of different flows of time on our logical deduction (computation). We consider a typical example.

EXAMPLE 83.

Data	Query	Configuration
$t : FFA$	$?t : FA$	$\{t\}$

The possible world flow is a general binary relation.

We create by rule 4b of Definition 76 a future configuration  $t < s$  and add to the database  $s : FA$ . We get

Data	Query	Configuration
$t : FFA$	$?t : FA$	$t < s$
$s : FA$		

Again we apply rule 4a of Definition 76 and get the new configuration with  $s < s'$  and the new item of data  $s' : A$ . We get

Data	Query	Configuration
$t : FFA$	$?t : FA$	$t < s$
$s : FA$		$s < s'$
$s' : A$		

Whether or not we can proceed from here depends on the flow of time. If  $<$  is transitive, then  $t < s'$  holds and we can get  $t : FA$  in the data by rule 3.

Actually by rule 4\* we could have proceeded along the following sequence of deduction. Rule 4\* is especially geared for transitivity.

Data	Query	Configuration
$t : FFA$	$t : FA$	$t$

Using rule 4\* we get

Data	Query	Configuration
$t : FFA$	$s : FA \vee FFA$	$t < s$
$s : FA$		

The first disjunct of the query succeeds.

If  $<$  is not transitive, rule 3 does not apply, since  $t < s'$  does not hold.

Suppose our query were  $?t : FFA$ .

If  $<$  is reflexive then we can succeed with  $?t : FFA$  because  $t < t$ .

If  $<$  is dense (i.e.  $\forall xy(x < y \rightarrow \exists z(x < z \wedge z < y))$ ) we should also succeed because we can create a point  $z$  with  $t < z < s$ .

$z : FFA$  will succeed and hence  $t : FFA$  will also succeed.

Here we encounter a new rule (density rule), whereby points can always be ‘landed’ between existing points in a configuration.

We now address the flow of time of the type natural numbers,  $\{1, 2, 3, 4, \dots\}$ . This has the special property that it is generated by a function symbol  $\mathbf{s}$ :

$$\{1, \mathbf{s}(1), \mathbf{ss}(1), \dots\}.$$

EXAMPLE 84.

Data	Query	Configuration
$\Box(q \rightarrow \bigcirc q)$ $1 : \bigcirc q$ $1 : Fp$	$1 : F(p \wedge q)$	$\{1\}$

If time is the natural numbers, the query should succeed from the data. If time is not the natural numbers but, for example,  $\{1, 2, 3, \dots, w, w+1, w+2, \dots\}$  then the query should fail.

How do we represent the fact that time is the natural numbers in our computation rule? What is needed is the ability to do some induction. We can use rule 4b and introduce a point  $t$  with  $1 < t$  into the configuration and even say that  $t = n$ , for some  $n$ . We thus get

Data	Query	Configuration
$\Box(q \rightarrow \bigcirc q)$ $1 : \bigcirc q$ $1 : Fp$ $n : p$	$1 : F(p \wedge q)$	$1 < n$

Somehow we want to derive  $n : q$  from the first two assumptions. The key reason for the success of  $F(p \wedge q)$  is the success of  $\Box q$  from the first two assumptions. We need an induction axiom on the flow of time.

To get a clue as to what to do, let us see what Prolog would do with the translations of the data and goal.

#### Translated data

$$\begin{aligned} &\forall t[1 \leq t \wedge Q^*(t) \rightarrow Q^*(t+1)] \\ &Q^*(1) \\ &\exists tP^*(t). \end{aligned}$$

#### Translated query

$$\exists t(P^*(t) \wedge Q^*(t)).$$

After we Skolemize, the database becomes:

1.  $1 \leq t \wedge Q^*(t) \rightarrow Q^*(t+1)$
2.  $Q^*(1)$
3.  $P^*(c)$

and the query is

$$P^*(s) \wedge Q^*(s).$$

We proceed by letting  $s = c$ . We ask  $Q^*(c)$  and have to ask after a slightly generalized form of unification  $?1 \leq c \wedge Q^*(c-1)$ .

Obviously this will lead nowhere without an induction axiom. The induction axiom should be that for *any* predicate  $PRED$

$$PRED(1) \wedge \forall x[1 \leq x \wedge PRED(x) \rightarrow PRED(x+1)] \rightarrow \forall x PRED(x).$$

Written in Horn clause form this becomes

$$\exists x \forall y [PRED(1) \wedge [1 \leq x \wedge PRED(x) \rightarrow PRED(x+1)] \rightarrow PRED(y)].$$

Skolemizing gives us

$$4. PRED(1) \wedge (1 \leq d \wedge PRED(d) \rightarrow PRED(d+1)) \rightarrow PRED(y)$$

where  $d$  is a Skolem constant.

Let us now ask the query  $P^*(s) \wedge Q^*(s)$  from the database with 1–4. We unify with clause 3 and ask  $Q^*(c)$ . We unify with clause 4 and ask  $Q^*(1)$  which succeeds and ask for the implication

$$?1 \leq d \wedge Q^*(d) \rightarrow Q^*(d+1).$$

This should succeed since it is a special case of clause 1 for  $t = d$ .

The above shows that we need to add an induction axiom of the form

$$\bigcirc x \wedge \square(x \rightarrow \bigcirc x) \rightarrow \square x.$$

Imagine that we are at time  $t$ , and assume  $t' < t$ . If  $A$  is true at  $t'$  and  $\square(A \rightarrow \bigcirc A)$  is true, then  $A$  is true at  $t$ .

We thus need the following rule:

### 18. Induction rule:

$t : F(A \wedge B)$  succeeds from  $\Delta$  at a certain configuration if the following conditions all hold.

1.  $t : FB$  succeed.
2. For some  $s < t, s : A$  succeeds.

3.  $m : \bigcirc A$  succeeds from the database  $\Delta'$ , where  $\Delta' = \{X, \square X \mid \square X \in \Delta\} \cup \{A\}$  and  $m$  is a completely new time point and the new configuration is  $\{m\}$ .

The above shows how to compute when time is the natural number. This is not the best way of doing it. In fact, the characteristic feature involved here is that the ordering of the flow of time is a Herbrand universe generated by a finite set of function symbols.  $FA$  is read as ‘ $A$  is true at a point generated by the function symbols’. This property requires a special study. See Chapter 11 of [Gabbay *et al.*, 1994].

#### 6.4 A theorem prover for modal and temporal logics

This section will briefly indicate how our temporal Horn clause computation can be extended to an automated deduction system for full modal and temporal logic. We present computation rules for propositional temporal logic with  $F, P, \bigcirc, \bullet, \wedge \rightarrow$  and  $\perp$ . We intend to approach predicate logic in Volume 3 as it is relatively complex. The presentation will be intuitive.

DEFINITION 85. We define the notions of a *full clause*, a *body* and a *head*.

- (a) A *full clause* is an atom  $q$  or  $\perp$  or  $B \rightarrow H$ , or  $H$  where  $B$  is a body and  $H$  is a head.
- (b) A *body* is a conjunction of full clauses.
- (c) A *head* is an atom  $q$  or  $\perp$  or  $FH$  or  $PH$  or  $\bigcirc H$  or  $\bullet H$ , where  $H$  is a body.

Notice that negation by failure is not allowed. We used the connectives  $\wedge, \rightarrow, \perp$ . The other connectives,  $\vee$  and  $\sim$ , are definable in the usual way:  $\sim A = A \rightarrow \perp$  and  $A \vee B = (A \rightarrow \perp) \rightarrow B$ . The reader can show that every formula of the language with the connectives  $\{\sim, \wedge, \vee, F, G, P, H\}$  is equivalent to a conjunction of full clauses. We use the following equivalences:

$$A \rightarrow (B \wedge C) = (A \rightarrow B) \wedge (A \rightarrow C);$$

$$A \rightarrow (B \rightarrow C) = A \wedge B \rightarrow C;$$

$$GA = F(A \rightarrow \perp) \rightarrow \perp;$$

$$HA = P(A \rightarrow \perp) \rightarrow \perp.$$

DEFINITION 86. A database is a set of labelled full clauses of the form  $(\Delta, \rho, <)$ , where  $\rho = \{t \mid t : A \in \Delta, \text{ for some } A\}$ . A query is a labelled full clause.

DEFINITION 87. The following is a definition of the predicate  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)$ , which reads: the labelled goal  $t : G$  succeeds from  $(\Delta, \rho, <)$  with parameter (initial goal)  $t_0 : G_0$ .

- 1(a)  $\mathbf{S}(\rho, <, \Delta, q, t, G_0, t_0)$  for  $q$  atomic or  $\perp$  if for some  $t : A \rightarrow q$ ,  $\mathbf{S}(\rho, <, \Delta, A, t, G_0, t_0)$ .
- (b) If  $t : q \in \Delta$  or  $s : \perp \in \Delta$  then  $\mathbf{S}(\rho, <, \Delta, q, t, G_0, t_0)$ .
- (c)  $\mathbf{S}(\rho, <, \Delta, \perp, t, G_0, t_0)$  if  $\mathbf{S}(\rho, <, \Delta, \perp, s, G_0, t_0)$ .  
 This rule says that if we can get a contradiction from any label, it is considered a contradiction of the whole system.
2.  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)$  if for some  $s : A \rightarrow \perp$ ,  $\mathbf{S}(\rho, <, \Delta, A, s, G_0, t_0)$ .
3.  $\mathbf{S}(\rho, <, \Delta, t, FG, G_0, t_0)$  if for some  $s \in \rho, t < s$  and  $\mathbf{S}(\rho, <, \Delta, G, s, G_0, t_0)$ .
4.  $\mathbf{S}(\rho, <, \Delta, FG, t, G_0, t_0)$  if for some  $t : A \rightarrow FB \in \Delta$  we have that both (a) and (b) below hold true:
- (a) For all  $s \in \rho$  such that  $t < s$  we have  $\mathbf{S}(\rho, <, \Delta^*, s, D, G_0, t_0)$  and  $\mathbf{S}(\rho, <, \Delta, E, t, G_0, t_0)$  hold, where  $\Delta^* = \Delta \cup \{s : B\}$  and  $D \in \{G, FG\}$  and  $E \in \{A, FA\}$ .  
**Note:** The choice of  $D$  and  $E$  is made here for the case of transitive time. In modal logic, where  $<$  is not necessarily transitive, we take  $D = G, E = A$ . Other conditions on  $<$  correspond to different choices of  $D$  and  $E$ .
- (b) For all future configurations of  $(\rho, <, t)$  with a new letter  $s$ , denoted by  $(\rho_s, <_s)$ , we have  $\mathbf{S}(\rho_s, <_s, \Delta^*, s, D, G_0, t_0)$  and  $\mathbf{S}(\rho_s, <_s, \Delta, E, t, G_0, t_0)$  hold, where  $\Delta^*, E, D$  are as in (a).
5. This is the mirror image of 3.
6. This is the mirror image of 4.
- 7(a)  $\mathbf{S}(\rho, <, \Delta, A_1 \wedge A_2, t, G_0, t_0)$  if both  $\mathbf{S}(\rho, <, \Delta, A_i, t, G_0, t_0)$  hold for  $i = 1, 2$ .
- (b)  $\mathbf{S}(\rho, <, \Delta, A \rightarrow B, t, G_0, t_0)$  if  $\mathbf{S}(\rho, <, \Delta \cup \{t : A\}, B, t, G_0, t_0)$ .
8. **Restart rule:**  
 $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0)$  if  $\mathbf{S}(\rho, <, \Delta, G_0, t_0, G_0, t_0)$ .

If the language contains  $\circ$  and  $\bullet$  then the following are the relevant rules.

9.  $\mathbf{S}(\rho, <, \Delta, \circ G, t, G_0, t_0)$  if  $\circ t$  exists and  $\circ t \in \rho$  and  $\mathbf{S}(\rho, <, \Delta, G, \circ t, G_0, t_0)$ .
10.  $\mathbf{S}(\rho, <, \Delta, \circ G, t, G_0, t_0)$  if for some  $t : A \rightarrow \circ B \in \Delta$  both  $\mathbf{S}(\rho, <, \Delta, A, t, G_0, t_0)$  and  $\mathbf{S}(\rho \cup \{\circ t\}, <', \Delta \cup \{\circ t : B\}, G, \circ t, G_0, t_0)$  hold where  $<'$  is the appropriate ordering closure of  $< \cup \{t < \circ t\}$ .

11. This is the mirror image of 9 for  $\bullet$ .
12. This is the mirror image of 10 for  $\bullet$ .

EXAMPLE 88. (Here  $\square$  can be either  $G$  or  $H$ .)

	Data	Query	Configuration
1.	$t : \square a$	$?t : \square b$	$\{t\}$
2.	$t : \square(a \rightarrow b)$		$t$ is a constant

**Translation:**

	Data	Query	Configuration
1.	$t : F(a \rightarrow \perp) \rightarrow \perp$	$t : F(b \rightarrow \perp) \rightarrow \perp$	$\{t\}$
2.	$t : F((a \rightarrow b) \rightarrow \perp) \rightarrow \perp$		

**Computation**

The problem becomes

	Additional data	Current query	Configuration
3.	$t : F(b \rightarrow \perp)$	$?t : \perp$	$\{t\}$
from 2		$?t_0 : F((a \rightarrow b) \rightarrow \perp)$	

From 3 using \*\* create a new point  $s$ :

	Additional data	Current query	Configuration
4.	$s : b \rightarrow \perp$	$?s : (a \rightarrow b) \rightarrow \perp$	$t < s$
	Add $s : a \rightarrow b$ to the database and ask		
5.	$s : (a \rightarrow b)$	$?s : \perp$	

From 4 and 5 we ask:

$$?s : a.$$

From computation rule 2 and clause 1 of the data we ask

$$?t : F(a \rightarrow \perp).$$

From computation rule 2 we ask

$$?s : a \rightarrow \perp$$

We add  $s : a$  to the data and ask

	Additional data	Current query	Configuration
6.	$s : a$	$?s : \perp$	$t < s$

The query succeeds.



### 6.5 Modal and temporal Herbrand universes

This section deals with the soundness of our computation rules. In conjunction with soundness it is useful to clarify the notion of modal and temporal Herbrand models. For simplicity we deal with temporal logic with  $P, F$  only and transitive irreflexive time or with modal logic with one modality  $\Diamond$  and a general binary accessibility relation  $<$ . We get our clues from some examples:

EXAMPLE 89. Consider the database

1.  $t : a \rightarrow \Diamond b$
2.  $\Box(b \rightarrow c)$
3.  $t : a.$

The constellation is  $\{t\}$ .

If we translate the clauses into predicate logic we get:

1.  $a^*(t) \rightarrow \exists s > tb^*(s)$
2.  $\forall x[b^*(x) \rightarrow c^*(x)]$
3.  $a^*(t).$

Translated into Horn clauses we get after Skolemising:

- 1.1  $a^*(t) \rightarrow b^*(s)$
- 1.2  $a^*(t) \rightarrow t < s$
- 2  $b^*(x) \rightarrow c^*(x)$
- 3  $a^*(t).$

$t, s$  are Skolem constants.

From this program, the queries

$$a^*(t), \neg b^*(t), \neg c^*(t), \neg a(s), b(s), c^*(s)$$

all succeed.  $\neg$  is negation by failure.

It is easy to recognize that  $\neg a^*(s)$  succeeds because there is no head which unifies with  $a^*(s)$ . The meaning of the query  $\neg a^*(s)$  in terms of modalities is the query  $\Diamond \neg a$ .

The question is: how do we recognize syntactically what fails in the modal language? The heads of clauses can be whole databases and there is no immediate way of syntactically recognizing which atoms are not heads of clauses.

EXAMPLE 90. We consider a more complex example:

1.  $t : a \rightarrow \Diamond b$
2.  $\Box(b \rightarrow c)$
3.  $t : a$
4.  $t : a \rightarrow \Diamond d$ .

We have added clause 4 to the database in the previous example. The translation of the first three clauses will proceed as before. We will get

- 1.1  $a^*(t) \rightarrow c^*(s)$
- 1.2  $a^*(t) \rightarrow t < s$
- 2  $b^*(x) \rightarrow c^*(x)$
- 3  $a^*(t)$ .

We are now ready to translate clause 4. This should be translated like clause 1 into

- 4.1  $a^*(t) \rightarrow d^*(r)$
- 4.2  $a^*(t) \rightarrow t < r$ .

The above translation is correct if the set of possible worlds is just an ordering. Suppose we know further that in our modal logic the set of possible worlds is linearly ordered. Since  $t < s \wedge t < r \rightarrow s = r \vee s < r \vee r < s$ , this fact must be reflected in the Horn clause database. The only way to do it is to add it as an integrity constraint.

Thus our temporal program translates into a Horn clause program with integrity constraints.

This will be true in the general case. Whether we need integrity constraints or not will depend on the flow of time.

Let us begin by translating from the modal and temporal language into Horn clauses. The labelled wff  $t : A$  will be translated into a set of formulae of predicate logic denoted by  $Horn(t, A)$ .  $Horn(t, A)$  is supposed to be logically equivalent to  $A$ . The basic translation of a labelled atomic predicate formula  $t : A(x_1 \dots x_n)$  is  $A^*(t, x_1 \dots x_n)$ .  $A^*$  is a formula of a two-sorted predicate logic where the first sort ranges over labels and the second sort over domain elements (of the world  $t$ ).

DEFINITION 91. Consider a temporal predicate language with connectives  $P$  and  $F$ , and  $\neg$  for negation by failure.

Consider the notion of labelled temporal clauses, as defined in Definition 73.

Let  $Horn(t, A)$  be a translation function associating with each labelled clause or goal a set of Horn clauses in the two-sorted language described above. The letters  $t, s$  which appear in the translation are Skolem constants. They are assumed to be *all different*.

We assume that we are dealing with a general transitive flow of time. This is to simplify the translation. If time has extra conditions, i.e. linearity, additional integrity constraints may need to be added. If time is characterized by non-first-order conditions (e.g. finiteness) then an adequate translation into Horn clause logic may not be possible.

The following are the translation clauses:

1.  $Horn(t, A(x_1 \dots x_n)) = A^*(t, x_1 \dots x_n)$ , for  $A$  atomic;
2.  $Horn(t, FA) = \{t < s\} \cup Horn(s, A)$   
 $Horn(t, PA) = \{s < t\} \cup Horn(s, A)$ ;
3.  $Horn(t, A \wedge B) = Horn(t, A) \cup Horn(t, B)$ ;
4.  $Horn(t, \neg A) = \neg \bigwedge Horn(t, A)$ ;
5.  $Horn(t, A \rightarrow F \wedge B_j) = \{\bigwedge Horn(t, A) \rightarrow t < s\} \cup \bigcup_{B_j} \{\bigwedge Horn(s, A) \wedge C \rightarrow D \mid (C \rightarrow D) \in Horn(s, B_j)\}$ ;
6.  $Horn(t, A \rightarrow P \wedge B_j) = \{\bigwedge Horn(t, A) \rightarrow s < t\} \cup \bigcup_{B_j} \{\bigwedge Horn(s, A) \wedge C \rightarrow D \mid (C \rightarrow D) \in Horn(s, B_j)\}$ ;
7.  $Horn(t, \Box A) = Horn(x, A)$  where  $x$  is a universal variable.

EXAMPLE 92. To explain the translation of  $t : A \rightarrow F(B_1 \wedge (B_2 \rightarrow B_3))$ , let us write it in predicate logic.  $A \rightarrow F(B_1 \wedge (B_2 \rightarrow B_3))$  is true at  $t$  if  $A$  true at  $t$  implies  $F(B_1 \wedge (B_2 \rightarrow B_3))$  is true at  $t$ .  $F(B_1 \wedge (B_2 \rightarrow B_3))$  is true at  $t$  if for some  $s$ ,  $t < s$  and  $B_1 \wedge (B_2 \rightarrow B_3)$  are true at  $s$ .

Thus we have the translation

$$A^*(t) \rightarrow \exists s(t < s \wedge B_1^*(s) \wedge (B_2^*(s) \rightarrow B_3^*(s))).$$

Skolemizing on  $s$  and writing it in Horn clauses we get the conjunction

$$\begin{aligned} A^*(t) &\rightarrow t < s \\ A^*(t) &\rightarrow B_1^*(s) \\ A^*(t) \wedge B_2^*(s) &\rightarrow B_3^*(s). \end{aligned}$$

Let us see what the translation  $Horn$  does:

$$\begin{aligned} Horn(t, A \rightarrow F(B_1 \wedge (B_2 \rightarrow B_3))) &= \{\bigwedge Horn(t, A) \rightarrow t < s\} \cup \\ &\{\bigwedge Horn(t, A) \rightarrow Horn(s, B_2)\} \cup \{\bigwedge Horn(t, A) \wedge \bigwedge Horn(s, B_2) \rightarrow \\ &\bigwedge Horn(s, B_3)\} = \{A^*(t) \rightarrow t < s, A^*(t) \rightarrow B_2^*(s), A^*(t) \wedge B_2^*(s) \rightarrow B_3^*(s)\}. \end{aligned}$$

We prove soundness of the computation of Definition 76, relative to the Horn clause computation for the Horn database in classical logic. In other words, if the translation  $Horn(t, A)$  is accepted as sound, as is intuitively clear, then the computation of  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0, \Theta)$  can be translated isomorphically into a classical Horn clause computation of the form  $Horn(t, \Delta)?Horn(t, G)$ , and the soundness of the classical Horn clause computation would imply the soundness of our computation.

This method of translation will also relate our temporal computation to that of an ordinary Horn clause computation.

The basic unit of our temporal computation is  $\mathbf{S}(\rho, <, \Delta, G, t, G_0, t_0, \Theta)$ . The current labelled goal is  $t : G$  and  $t_0 : G_0$  is the original goal. The database is  $(\rho, <, \Delta)$  and  $\Theta$  is the current substitution.  $t_0 : G_0$  is used in the restart rule. For a temporal flow of time which is ordinary transitive  $<$ , we do not need the restart rule. Thus we have to translate  $(\rho, <, \Delta)$  to classical logic and translates  $t : G$  and  $\Theta$  to classical logic and see what each computation step of  $\mathbf{S}$  of the source translates into the classical logic target.

**DEFINITION 93.** Let  $(\rho, <)$  be a constellation and let  $\Delta$  be a labelled database such that

$$\rho = \{t \mid \text{for some } A, t : A \in \Delta\}.$$

Let  $Horn((\rho, <), \Delta) = \{t < s \mid t, s \in \rho \text{ and } t < s\} \cup \bigcup_{t:A \in \Delta} Horn(t, A)$ .

**THEOREM 94 (Soundness).**  $\mathbf{S}(\rho, <, \Delta, G, t, \Theta)$  *succeeds in temporal logic if and only if in the sorted classical logic  $Horn((\rho, <), \Delta)?Horn(t, G)$  succeeds with  $\Theta$ .*

**Proof.** The proof is by induction on the complexity of the computation tree of  $\mathbf{S}(\rho, <, \Delta, G, t, \Theta)$ .

We follow the inductive steps of Definition 75. The translation of  $(\mathbf{\Pi}, \Theta)$  is a conjunction of Horn clause queries, all required to succeed under the same substitution  $\Theta$ .

Case I The empty goal succeeds in both cases.

Case II  $(\mathbf{\Pi}, \Theta)$  fails if for some  $\mathbf{S}(\rho, <, \Delta, G, t)$ , we have  $G$  is atomic and for all  $\square(A \rightarrow H) \in \Delta$  and all  $t : A \rightarrow H \in \Delta$ ,  $G\Theta$  and  $H\Theta$  do not unify. The reason they do not unify is because of what  $\Theta$  substitutes to the variables  $u_i$ .

The corresponding Horn clause predicate programs are

$$\bigwedge Horn(x, A) \rightarrow H^*(x)$$

and

$$\bigwedge Horn(t, A) \rightarrow H^*(t)$$

and the goal is  $?G^*(t)$ .

Clearly, since  $x$  is a general universal variable, the success of the two-sorted unification depends on the other variables and  $\Theta$ . Thus unification does *not* succeed in the classical predicate case iff it does not succeed in the temporal case.

Rules 1 and 2 deal with the atomic case: the query is  $G^*(t)$  and in the database among the data are

$$\bigwedge \text{Horn}(t, A) \rightarrow H^*(t) \text{ and } \bigwedge \text{Horn}(x, A) \rightarrow H^*(x)$$

for the cases of  $t : A \rightarrow H$  and  $\Box(A \rightarrow H)$  respectively.

For the Horn clause program to succeed  $G^*(t)$  must unify with  $H^*(t)$ . This will hold if and only if the substitution for the domain variables allows unification, which is exactly the condition of Definition 75.

Rules 3, 4(general) and 4\*(general) deal with the case of a goal of the form  $?t : FG$ . The translation of the goal is  $t < u \wedge \bigwedge \text{Horn}(u, G)$  where  $u$  is an existential variable.

Rule 3 gives success when for some  $s, t < s \in \Delta$  and  $?s : G$  succeeds. In this case let  $u = s$ ; then  $t < u$  succeeds and  $\bigwedge \text{Horn}(s, G)$  succeeds by the induction hypothesis.

We now turn to the general rules 4(general) and 4\*(general). These rules yield success when for some clause of the form

$$t : A \rightarrow F \wedge B_j$$

or

$$\Box(A \rightarrow F \wedge B_j).$$

$\Delta?t : A$  succeeds and  $\Delta \cup \{(s : B_j)\} ?s : G \vee FG$  both succeed.  $s$  is a new point.

The translation  $\bigwedge \text{Horn}(t, A)$  succeeds by the induction hypothesis.

The translation of

$$t : A \rightarrow F \wedge B_j$$

or

$$\Box(A \rightarrow F \wedge B_j)$$

contains the following database:

1.  $\bigwedge \text{Horn}(t, A) \rightarrow t < s$ .
2. For every  $B_j$  and every  $C \rightarrow D$  in  $\text{Horn}(s, B_j)$  the clause  $\bigwedge \text{Horn}(s, A) \wedge C \rightarrow D$ .

Since  $\bigwedge \text{Horn}(t, A)$  succeeds we can assume we have in our database:

$$1^* t < s;$$

2\*  $C \rightarrow D$ , for  $C \rightarrow D \in \text{Horn}(s, B_j)$  for some  $j$ .

These were obtained by substituting *truth* in 1 and 2 for  $\bigwedge \text{Horn}(t, A)$ .

The goal is to show  $t < u \wedge \bigwedge \text{Horn}(u, G)$ .

Again for  $u = s$ ,  $t < u$  succeeds from (1\*) and by the induction hypothesis, since  $\Delta \cup \{s : B_j\} ?s : G \vee FG$  is successful, we get

$$\bigcup_j \text{Horn}(s, B_j) ? \bigwedge \text{Horn}(s, G) \vee (s < u' \wedge \bigwedge \text{Horn}(u', G))$$

should succeed, with  $u'$  an existential variable.

However, 2\* is exactly  $\bigcup_j \text{Horn}(s, \bigwedge B_j)$ . Therefore we have shown that rules 4(general) and 4\*(general) are sound.

Rules 6(general) and 6\*(general) are sound because they are the mirror images of 4(general) and 4\*(general).

The next relevant rules for our soundness cases are 11–13. These follow immediately since the rules for  $\wedge, \vee, \neg$  are the same in both computations.

Rule 14, the restart rule, is definitely sound. If we try to show in general that  $\Delta \vdash A$  then since in classical logic  $\sim A \rightarrow A$  is the same as  $A$  ( $\sim$  is classical negation) it is equivalent to show  $\Delta, \sim A \vdash A$ .

If  $\sim A$  is now in the data, we can *at any time* try to show  $A$  instead of the current goal  $G$ . This will give us  $A$  (shown) and  $\sim A$  (in Data) which is a contradiction, and this yields *any goal* including the current goal  $G$ .

We have thus completed the soundness proof.  $\blacksquare$

## 6.6 Tractability and persistence

We defined a temporal database  $\Delta$  essentially as a finite piece of information telling us which temporal formulae are true at what times. In the most general case, for a general flow of time  $(T, <)$ , all a database can do is to provide a set of the form  $\{t_i : A_i\}$ , meaning that  $A_i$  is true at time  $t_i$  and a configuration  $(\{t_i\}, <)$ , giving the temporal relationships among  $\{t_i\}$ . A query would be of the form  $?t : Q$ , where  $t$  is one of the  $t_i$ . The computation of the query from the data is in the general case exponential, as we found in Section 6.2, from the case analysis of clause 4 of Definition 76 and from Example 72. We must therefore analyse the reasons for the complexity and see whether there are simplifying natural assumptions, which will make the computational problem more tractable.

There are three main components which contribute to complexity:

1. The complexity of the temporal formulae allowed in the data and in the query. We allow  $t : A$  into the database, with  $A$  having temporal operators. So, for example,  $t : FA$  is allowed and also  $t : \bigcirc A$ .  $t : FA$  makes life more difficult because it has a hidden Skolem function in it. It really means  $\exists s[t < s \text{ and } (s : A)]$ . This gives rise to case

analysis, as we do not know in general where  $s$  is. See Example 80 and Examples 89 and 90. In this respect  $t : \bigcirc A$  is a relatively simple item. It says  $(t+1) : A$ . In fact any temporal operator which specifies the time is relatively less complex. In practice, we do need to allow data of the form  $t : FA$ . Sometimes we know an event will take place in the future but we do not know when. The mere fact that  $A$  is going to be true can affect our present actions. A concrete example where such a case may arise is when someone accepts a new appointment beginning next year, but has not yet resigned from their old position. We know they are going to resign but we do not know when;

2. The flow of time itself gives rise to complexity. The flow of time may be non-Horn clause (e.g. linear time which is defined by a disjunctive axiom

$$\forall xy[x < y \vee y < x \vee x = y].$$

This complicates the case analysis of 1 above.

3. Complexity arises from the behaviour. If atomic predicates get truth values at random moments of time, the database can be complex to describe. A very natural simplifying assumption in the case of temporal logic is *persistence*. If atomic statements and their negations remain true for a while then they give rise to less complexity. Such examples are abundant. For example, people usually stay at their residences and jobs for a while. So for example, any payroll or local tax system can benefit from persistence as a simplifying assumption. Thus in databases where there is a great deal of persistence, we can use this fact to simplify our representation and querying. In fact, we shall see that a completely different approach to temporal representation can be adopted when one can make use of persistence.

Another simplifying assumption is *recurrence*. Saturdays, for example, recur every week, so are paydays. This simplifies the representation and querying. Again, a payroll system would benefit from that.

We said at the beginning that a database  $\Delta$  is a finitely generated piece of temporal information stating what is true and when. If we do not have any simplifying assumptions, we have to represent  $\Delta$  in the form  $\Delta = \{t_i : A_i\}$  and end up needing the computation rules of Section 6.2 to answer queries.

Suppose now that we adopt all three simplifying assumptions for our database. We assume that the  $A_i$  are only atoms and their negations, we further assume that each  $A_i$  is either persistent or recurrent, and let us assume, to be realistic, that the flow of time is linear. Linearity does not make the computation more complicated in this particular case, because we are not allowing data of the form  $t : FA$ , and so complicated case analysis

does not arise. In fact, together with persistence and recurrence, linearity becomes an additional simplifying assumption!

Our aim is to check what form our temporal logic programming machine should take in view of our chosen simplifying assumptions.

First note that the most natural units of data are no longer of the form:

$$t : A$$

reading  $A$  is true at  $t$ , but either of the form

$$[t, s] : A, [t < s]$$

reading  $A$  is true in the closed interval  $[t, s]$ , or the form

$$t||d : A$$

reading  $A$  is true at  $t$  and recurrently at  $t + d, t + 2d, \dots$ , that is, every  $d$  moments of time.

$A$  is assumed to be a literal (atom or a negation of an atom) and  $[t, s]$  is supposed to be a maximal interval where  $A$  is true. In  $t||d$ ,  $d$  is supposed to be the minimal cycle for  $A$  to recur. The reasons for adopting the notation  $[t, s] : A$  and  $t||d : A$  are not mathematical but simply intuitive and practical. This is the way we think about temporal atomic data when persistence or recurrence is present. In the literature there has been a great debate on whether to evaluate temporal statements at points or intervals. Some researchers were so committed to intervals that they tended, unfortunately, to disregard any system which uses points. Our position here is clear and intuitive. First perform all the computations using intervals. Evaluation at points is possible and trivial. To evaluate  $t : A$ , i.e. to ask  $?t : A$  as a query from a database, compute the (maximal) intervals at which  $A$  is true and see whether  $t$  is there. To evaluate  $[t, s] : A$  do the same, and check whether  $[t, s]$  is a subset.

The query language is left in its full generality. i.e. we can ask queries of the form  $t : A$  where  $A$  is unrestricted (e.g.  $A = FB$  etc.). It makes sense also to allow queries of the form  $[t, s] : A$ , although exactly how we are going to find the answer remains to be seen. The reader should be aware that the data representation language and the query language are no longer the same. This is an important factor. There has been a lot of confusion, especially among the AI community, in connection with these matters. We shall see later that as far as computational tractability is concerned, the restriction to persistent data allows one to strengthen the query language to full predicate quantification over time points.

At this stage we might consider allowing recurrence within an interval, i.e. we allow something like

$$'A \text{ is true every } d \text{ days in the interval } [t, s].'$$



We can denote this by

$$[t||d, s] : A$$

meaning  $A$  is true at  $t, t + d, t + 2d$ , as long as  $t + nd \leq s, n = 1, 2, 3, \dots$

We may as well equally have recurrent intervals. An example of that would be taking a two-week holiday every year. This we denote by

$$[t, s]||d : A, \quad t < s, \quad (s - t) < d,$$

reading  $A$  is true at the intervals  $[t, s], [t + d, s + d], [t + 2d, s + 2d]$ , etc.

The reader should note that adopting this notation takes us outside the realm of first-order logic. Consider the integer flow of time. We can easily say that  $q$  is true at all even numbers by writing  $[0, 0]||1$  as a truth set for  $q$  and  $[1, 1]||1$  as a truth set for  $\sim q$  (i.e.  $q$  is true at 0 and recurs every 1 unit and  $\sim q$  is true at 1 and recurs every 1 unit).

The exact expressive power of this language is yet to be examined. It is connected with the language USF which we meet later.

The above seem to be the most natural options to consider. We can already see that it no longer makes sense to check how the computation rules of Definition 76 simplify for our case. Our case is so specialized that we may as well devise computation rules especially for it. This should not surprise us. It happens in mathematics all the time. The theory of Abelian groups, for example, is completely different from the theory of semigroups, although Abelian groups are a special case of semigroups. The case of Abelian groups is so special that it does not relate to the general case any more.

Let us go back to the question of how to answer a query from our newly defined simplified databases. We start with an even more simple case, assuming only persistence and assuming that the flow of time is the integers. This simple assumption will allow us to present our point of view of how to evaluate a formula at a point or at an interval. It will also ensure we are still within what is expressible in first-order logic. Compare this with Chapter 13 of [Gabbay *et al.*, 1994].

Assume that the atom  $q$  is true at the maximal intervals  $[Xx_n, y_n], x_n \leq y_n < x_{n+1}$ . Then  $\sim q$  is true at the intervals  $[y_n + 1, x_{n+1} - 1]$ , a sequence of the same form, i.e.  $y_n + 1 \leq x_{n+1} - 1$  and  $x_{n+1} - 1 < y_{n+1} + 1$ .

It is easy to compute the intervals corresponding to the truth values of conjunctions: we take the intersection:

If  $I_j = \bigcup_n [x_n^j, y_n^j]$  then  $I_1 \cap I_2 = \bigcup_n [x_n, y_n]$  and the points  $x_n, y_n$  can be effectively linearly computed. Also, if  $I_j$  is the interval set for  $A_j$ , the interval set for  $U(A_1, A_2)$  can be effectively computed.

In Fig. 8,  $U(A_1, A_2)$  is true at  $[u_k, y_n - 1], [u_k, y_{n+1} - 1]$  which simplifies to the maximal  $[u_k, y_{n+1} - 1]$ .

The importance of the above is that we can regard a query formula of the full language with Until and Since as an operator on the model (database)

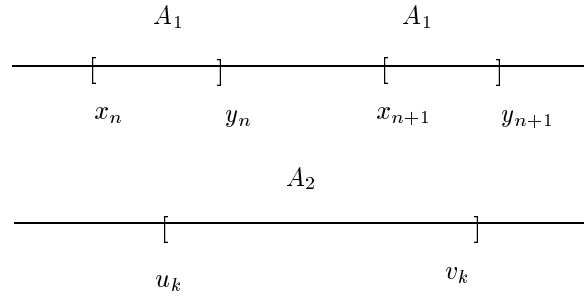


Figure 8.

to give a new database. If the database  $\Delta$  gives for each atom or its negation the set of intervals where it is true, then a formula  $A$  operates on  $\Delta$  to give the new set of intervals  $\Delta_A$ ; thus to answer  $\Delta?t : A$  the question we ask is  $t \in \Delta_A$ . The new notion is that the query operates on the model.

This approach was adopted by I. Torsun and K. Manning when implementing the query language USF. The complexity of computation is polynomial ( $n^2$ ). Note that although we have restricted the database formulae to atoms, we discovered that for no additional cost we can increase the query language to include the connectives *Since* and *Until*. As we have seen in Volume 1, in the case of integers the expressive power of *Since* and *Until* is equivalent to quantification over time points.

To give the reader another glimpse of what is to come, note that intuitively we have a couple of options:

1. We can assume persistence of atoms and negation of atoms. In this case we can express temporal models in first-order logic. The query language can be full *Since* and *Until* logic. This option does not allow for recurrence. In practical terms this means that we cannot generate or easily control recurrent events. Note that the database does not need to contain Horn clauses as data. Clauses of the form  $\Box(\text{present wff}_1 \rightarrow \text{present wff}_2)$  are redundant and can be eliminated (this has to be properly proved!). Clauses of the form  $\Box(\text{past wff}_1 \rightarrow \text{present wff}_2)$  are not allowed as they correspond to recurrence;
2. This option wants to have recurrence, and is not interested in first-order expressibility. How do we generate recurrence?

The language USF (which was introduced for completely different reasons) allows one to generate the database using rules of the form  $\Box(\text{past formula} \rightarrow \text{present or future formula})$ .

The above rules, together with some initial items of data of the form  $t : A$ ,  $A$  a literal, can generate persistent and recurrent models.

## 7 NATURAL NUMBERS TIME LOGICS

Certainly the most studied and widely used temporal logic is one based on a natural numbers model of time. The idea of discrete steps of time heading off into an unbounded future is perfect for many applications in computer science. The natural numbers also form quite a well-known structure and so a wide-range of alternative techniques can be brought to bear. Here we will have a brief look at this temporal logic, PTL, at another, stronger natural numbers time temporal logic *USF* and at the powerful automata technique which can be used to help reason in and about such logics.

### 7.1 PTL

PTL is a propositional temporal logic with semantics defined on the natural numbers time. It does not have past time temporal connectives because, as we will see, they are not strictly necessary, and, anyway, properties of programs (or systems or machines) are usually described in terms of what happens in the future of the start. So PTL is somewhat, but not exactly, like the propositional logic of the until connective over natural numbers time. By the way, PTL stands for Propositional Temporal Logic because computer scientists are not very interested in any of the other propositional temporal logics which we have met. PTL is also sometimes known as PLTL, for Propositional Linear Temporal Logic, because the only other propositional temporal logic of even vague interest to computer scientists is one based on branching time.

PTL does have a version of the until connective and also a next-time (or tomorrow) connective. The until connective, although commonly written as  $U$ , is not the same as the connective  $U$  which we have met before. Also, and much less importantly, it is usually written in an infix manner as in  $pUq$ . In PTL we write  $pUq$  iff either  $q$  is true now or  $q$  is true at some time in the future and we have  $p$  true at all points between now and then, including now but not including then. This is called a non-strict version of  $U$ : hence here we will use the notation  $U^{ns}$ . From the beginning in [Pnueli, 1977], temporal logic work with computer science applications in mind has used the non-strict version of until. In much of the work on temporal logic with a philosophical or linguistic bent, and in this chapter,  $U(q, p)$  means that  $q$  is true in the future and  $p$  holds at all points strictly in between. This is called the strict version of until.

To be more formal let us here define a structure as a triple  $(\mathbb{N}, <, h)$  where  $h : L \rightarrow \wp(\mathbb{N})$  is the valuation of the atoms from  $L$ . In effect, we have

an  $\omega$ -long sequence  $\sigma = \langle \sigma_0, \sigma_1, \dots \rangle$  of states (i.e. subsets of  $L$ ) with each  $\sigma_i = \{p \in L \mid i \in h(p)\}$ . Many presentations of PTL use this sequence of states notation. The truth of a PTL formula  $A$  at a point  $n$  of a structure  $\mathcal{T}$ , written  $\mathcal{T}, n \models A$  is defined by induction on the construction of  $A$  as usual. The clauses for atoms and Booleans are as usual. Define:

$$\begin{aligned} \mathcal{T}, n \models XA \text{ iff } & \quad \mathcal{T}, n+1 \models A, \text{ and} \\ \mathcal{T}, n \models AU^{ns}B \text{ iff } & \quad \text{there is some } m \geq n \text{ such that } \mathcal{T}, m \models B \text{ and} \\ & \quad \text{for all } j, \text{ if } n \leq j < m \text{ then } \mathcal{T}, j \models A. \end{aligned}$$

Note that in PTL, there are also non-strict versions of all the usual abbreviations. For example  $FA = \top U^{ns}A$  holds iff  $A$  holds now or at some time in the future, and  $GA = \neg F\neg A$  holds iff  $A$  holds now and at all time in the future. In many presentations of PTL, the symbols  $\diamond$  and  $\square$  are used instead of  $F$  and  $G$ .

In the case of natural numbers time (as in PTL) it is easy to show that the language with strict  $U$  is equally expressive as the language with both the next operator  $X$  and non-strict  $U^{ns}$ . We can define a meaning preserving translation  $\pi$  which preserves atoms and respects Boolean connectives via:

$$\pi U(A, B) = X(\pi(B)U^{ns}\pi(A)).$$

Similarly we can define a meaning preserving translation  $\rho$  which preserves atoms and respects Boolean connectives via:

$$\begin{aligned} \rho XA &= U(A, p_0 \wedge \neg p_0), \\ \rho(BU^{ns}A) &= \rho(A) \vee (\rho(B) \wedge U(\rho(A), \rho(B))). \end{aligned}$$

It is easy to show that these are indeed meaning preserving. Notice, though, that is computationally expensive to translate between the non-strict until and strict until (using  $\rho$ ) as there may be an exponential blow-up in formula length.

In some earlier presentations of PTL, the definitions of the concepts of satisfaction and satisfiability are different from ours. In typical PTL applications, it makes sense to concentrate almost exclusively on the truth of formulas in structures when evaluated at time 0. Thus we might say that structure  $(\mathbb{N}, <, h)$  satisfies  $A$  iff  $(\mathbb{N}, <, h), 0 \models A$ . Other presentations define satisfaction by saying  $(\mathbb{N}, <, h)$  satisfies  $A$  iff for all  $n$ ,  $(\mathbb{N}, <, h), n \models A$ . Recall that in this chapter we actually say that  $(\mathbb{N}, <, h)$  satisfies  $A$  iff there is some  $n$  such that  $(\mathbb{N}, <, h), n \models A$ . In general it is said that a formula  $A$  is satisfiable iff there is some structure which satisfies  $A$  (whatever notion of satisfaction is used). We will not explore the subtle details of what is sometimes known as the *anchored* versus *floating* version of temporal logics, except to say that no difficult or important issues of expressiveness or axiomatization etc, are raised by these differences. See [Manna and Pnueli, 1988] for more on this.

Given the equivalence of PTL and the temporal logic with  $U$  over natural numbers time and our separation result for the  $L(U, S)$  logic over the natural numbers, it is easy to see why  $S$  (and any other past connective) is not needed. Suppose that we want to check the truth of a formula  $A$  in the  $L(U, S)$  language at time 0 in a structure  $\mathcal{T} = (\mathbb{N}, <, h)$ . If we separate  $A$  into a Boolean combination of syntactically pure formulas we will see that, for each of the pure past formulas  $C$ , either  $C$  evaluates at time 0 to true in every structure or  $C$  evaluates at time 0 to false in every structure. Thus, we can effectively find some formula  $B$  in the language with  $U$  such that evaluating  $A$  at time 0 is equivalent to evaluating  $B$  at time 0. We know that  $B$  is equivalent to some formula of PTL which we can also find effectively. Thus adding  $S$  adds no expressiveness to PTL (in this specific sense). However, some of the steps in eliminating  $S$  may be time-consuming and it may be natural to express some useful properties with  $S$ . Thus there are motivations for introducing past-time connectives into PTL. This is the idea seen in [Lichtenstein *et al.*, 1985].

## 7.2 An axiomatization of PTL

We could axiomatize PTL using either the IRR rule or the techniques of [Reynolds, 1992]. However, the first axiomatization for PTL, given in [Gabbay *et al.*, 1980], uses a different approach with an interesting use of something like the computing concept of fairness. The axioms and proof in [Gabbay *et al.*, 1980] were actually given for a strict version of the logic but, as noted in [Gabbay *et al.*, 1980], it is easy to modify it for the non-strict version, what has now become the official PTL.

The inference rules are modus ponens and generalization,

$$\frac{A, A \rightarrow B}{B} \quad \frac{A}{G^{ns}A} \quad .$$

The axioms are all substitution instances of the following:

- (1) all classical tautologies,
- (2)  $G^{ns}(A \rightarrow B) \rightarrow (G^{ns}A \rightarrow G^{ns}B)$
- (3)  $X\neg A \rightarrow \neg XA$
- (4)  $X(A \rightarrow B) \rightarrow (XA \rightarrow XB)$
- (5)  $G^{ns}A \rightarrow A \wedge XG^{ns}A$
- (6)  $G^{ns}(A \rightarrow XA) \rightarrow (A \rightarrow G^{ns}A)$
- (7)  $(AU^{ns}B) \rightarrow F^{ns}A$
- (8)  $(AU^{ns}B) \leftrightarrow (B \vee (A \wedge X(AU^{ns}B)))$

The straightforward induction on the lengths of proof gives us the soundness result. The completeness result which is really a *weak completeness* result — the logic is not compact — follows.

**THEOREM 95.** *If  $A$  is valid in PTL then  $\vdash A$  ( $A$  is a theorem of the axiom system).*

**Proof.** We give a sketch. The details are left to the reader: or see [Gabbay *et al.*, 1980] (but note the use of strict versions of connectives in that proof). It is enough to show that if  $A$  is consistent then  $A$  is satisfiable.

We use the common Henkin technique of forming a model of a consistent formula in a modal logic out of the maximal consistent sets of formulae. These are the infinite sets of formulae which are each maximal in not containing some finite subsets whose conjunction is inconsistent. In our case this model will not have the natural numbers flow of time but will be a more general, not necessarily linear structure with a more general definition of truth for the temporal connectives.

Let  $\mathcal{C}$  contain all the maximally consistent sets of formulae. This is a non-linear model of  $A$  with truth for the connectives defined via the following (accessibility) relations: for each  $\Gamma, \Delta \in \mathcal{C}$ , say  $\Gamma R_+ \Delta$  iff  $\{B \mid XB \in \Gamma\} \subseteq \Delta$  and  $\Gamma R_< \Delta$  iff  $\{B \mid XG^{ns} B \in \Gamma\} \subseteq \Delta$ . For example, if we call this model  $M$  then for each  $\Gamma \in \mathcal{C}$ , we define  $M, \Gamma \models p$  iff  $p \in \Gamma$  for any atom  $p$  and  $M, \Gamma \models XB$  iff there is some  $\Delta \in \mathcal{C}$ , such that  $\Gamma R_+ \Delta$  and  $M, \Delta \models B$ . The truth of formulas of the form  $B_1 U^{ns} B_2$  is defined via paths through  $\mathcal{C}$  in a straightforward way.

The Lindenbaum technique shows us that there is some  $\Gamma_0 \in \mathcal{C}$  with  $A \in \Gamma_0$ . Using this and the fact that  $R_<$  is the transitive closure of  $R_+$ , we can indeed show that  $M, \Gamma_0 \models A$ .

There is also a common technique for taking this model and factoring out by an equivalence relation to form a finite but also non-linear model. This is the method of *filtration*. See [Gabbay *et al.*, 1994].

To do this in our logic, we first limit ourselves to a finite set of interesting formulae:

$$\text{cl}(A) = \{B, \neg B, XB, X\neg B, F^{ns} B, F^{ns} \neg B \mid B \text{ is a subformula of } A\}.$$

Now we define  $C = \{\Gamma \cap \text{cl}(A) \mid \Gamma \in \mathcal{C}\}$  and we impose a relation  $R_X$  on  $C$  via  $a R_X b$  iff there exist  $\Gamma, \Delta \in \mathcal{C}$  such that  $a = \Gamma \cap \text{cl}(A)$ ,  $b = \Delta \cap \text{cl}(A)$  and  $\Gamma R_+ \Delta$ .

To build natural numbers flowed model of  $A$  we next find an  $\omega$ -sequence  $\sigma$  of sets from  $C$  starting at  $\Gamma_0 \cap \text{cl}(A)$  and proceeding via the  $R_X$  relation in such a way that if the set  $\Gamma$  appears infinitely often in the sequence then each of its  $R_X$ -successors do too. This might be called a *fair* sequence. We can turn  $\sigma$  into a structure  $(\mathbb{N}, <, h)$  via  $i \in h(p)$  iff  $p \in \sigma_i$  (for all atoms  $p$ ). This is enough to give us a truth lemma by induction on all formulae  $B \in \text{cl}(A)$ : namely,  $B \in \sigma_i$  iff  $(\mathbb{N}, <, h), i \models B$ . Immediately we have  $(\mathbb{N}, <, h), 0 \models A$  as required. ■

### 7.3 *S1S and Fixed point languages*

PTL is expressively complete in respect of formulas evaluated at time 0. That is, for any first-order monadic formula  $\phi(P_1, \dots, P_n, x)$  in the language with  $<$  with one free variable, there is a PTL formula  $A$  such that for any structure  $(\mathbb{N}, <, h)$ ,

$$(\mathbb{N}, <, h), 0 \models A \text{ iff } (\mathbb{N}, <) \models \phi(h(p_1), \dots, h(p_n), 0).$$

To see this just use the expressive completeness of the  $L(U, S)$  logic over natural numbers time, separation, the falsity of since at zero, and the translation to PTL.

However, there are natural properties which can not be expressed. For example, there is no PTL formula which is true exactly at the even numbers (see [Wolper, 1983] for details) and we can not say that property  $p$  holds at each even-numbered time point. The reason for this lack of expressivity in an expressively complete language is simply that there are natural properties like evenness which can not be expressed in the first-order monadic theory of the natural numbers.

For these reasons there have been many and varied attempts to increase the expressiveness of temporal languages over the natural numbers. There has also been a need to raise a new standard of expressiveness. Instead of comparing languages to the first-order monadic theory of the natural numbers, languages are compared to another traditional second-order logic, the full second-order logic of one successor function, commonly known as *S1S*.

There are several slightly different ways of defining *S1S*. We can regard it as an ordinary first-order logic interpreted in a structure which actually consists of sets of natural numbers. The signature contains the 2-ary subset relation  $\subseteq$  and a 2-ary ordering relation symbol *succ*. Subset is interpreted in the natural way while *succ*( $A, B$ ) holds for sets  $A$  and  $B$  iff  $A = \{n\}$  and  $B = \{n + 1\}$  for some number  $n$ . To deal with a temporal structure using atoms from  $L$  we also allow the symbols in  $L$  as constant symbols in the language: given an  $\omega$ -structure  $\sigma$ , the interpretation of the atom  $p$  is just the set of times at which  $p$  holds.

Consider the example of the formula

$$J(a, b, z)(b \subseteq z) \wedge \forall uv(succ(u, v) \wedge (v \subseteq z) \wedge (u \subseteq a) \rightarrow (u \subseteq z))$$

with constants  $a, b$  and  $z$ . This will be true of a natural numbers flowed temporal structure with atoms  $a, b$  and  $z$  iff at every time at which  $aU^nsb$  holds, we also have  $z$  holding.

In fact, it is straightforward to show that the language of *S1S* is exactly as expressive as the full second-order monadic language of the natural number order. Thus it is more expressive than the first-order monadic language. A well-known and straightforward translation gives an *S1S* version of any

temporal formula. We can translate any temporal formula  $\alpha$  using atoms from  $L$  into an  $S1S$  formula  $(*\alpha)(x)$  with a free variable  $x$ :

$$\begin{aligned}
*p &= (x = p) \\
*(\neg\alpha) &= \neg(*\alpha) \\
*(\alpha \wedge \beta) &= *\alpha \wedge *\beta \\
*(X\alpha) &= \forall y((*\alpha)(y) \rightarrow \forall uv(succ(u, v) \wedge (u \subseteq x) \rightarrow (v \subseteq y))) \\
*(\alpha U^{ns} \beta) &= \forall ab((*\alpha)(a) \wedge (*\beta)(b)) \rightarrow \\
&\quad (J(a, b, x) \wedge (\forall y(J(a, b, y) \rightarrow (x \subseteq y)))) \\
\text{where } J(a, b, z) &= (b \subseteq z) \\
&\quad \wedge \forall uv(succ(u, v) \wedge (v \subseteq z) \wedge (u \subseteq a) \rightarrow (u \subseteq z))
\end{aligned}$$

An easy induction (on the construction of  $\alpha$ ) shows that  $\sigma \models (*\alpha)(S)$  iff  $S$  is the set of times at which  $\alpha$  holds.

In order to reach the expressiveness of  $S1S$ , temporal logics are often given an extra second-order capability involving some kind of quantification over propositions. See, for example, ETL [Wolper, 1983] and the quantified logics in chapter 8 of [Gabbay *et al.*, 1994]. One of the most computationally convenient ways of adding quantification is via the introduction of fixed-point operators into the language. This has been done in [Banniegbal and Barringer, 1986] and in [Gabbay, 1989]. We briefly look at the example of  $USF$  from [Gabbay, 1989].

In chapter 8 of [Gabbay *et al.*, 1994] it is established that  $S1S$  is exactly as expressive as  $USF$ . In order to use the automata-based decision procedures (which we meet in subsection 7.5 below) to give us decision procedures about temporal logics we need only know that the temporal logic of interest is less expressive than  $USF$  (or equivalently  $S1S$ ) — as they mostly are. So it is worth here briefly recalling the definition of  $USF$ .

In fact we start with the very similar language  $UYF$ . We often write  $\bar{x}, \bar{a}, \dots$ , for *tuples* — finite sequences of variables, atoms, elements of a structure, etc. If  $S \subseteq \mathbb{N}$ , we write  $S + 1$  (or  $1 + S$ ) for  $\{s + 1 \mid s \in S\}$ .

We start by developing the syntax and semantics of the fixed point operator. This is not entirely a trivial task. We will fix an infinite set of propositional atoms, with which our formulae will be written; we write  $p, q, r, s, \dots$  for atoms.

DEFINITION 96.

1. The set of formulae of  $UYF$  is the smallest class closed under the following:
  - (a) Any atom  $q$  is a formula of  $UYF$ , as is  $\top$ .
  - (b) If  $A$  is a formula so is  $\neg A$ .



- (c) If  $A$  is a formula so is  $YA$ . We read  $Y$  as ‘yesterday’.
- (d) If  $A$  and  $B$  are formulae, so are  $A \wedge B$  and  $U(A, B)$ . ( $A \vee B$  and  $A \rightarrow B$  are regarded as abbreviations.)
- (e) Suppose that  $A$  is a formula such that every occurrence of the atom  $q$  in  $A$  not within the scope of a  $\varphi q$  is within the scope of a  $Y$  but not within the scope of a  $U$ . Then  $\varphi qA$  is a formula. (The conditions ensure that  $\varphi qA$  has fixed point semantics.)
2. The *depth of nesting* of  $\varphi$ s in a formula  $A$  is defined by induction on its formation: formulae formed by clause (a) have depth 0, clause (e) adds 1 to the depth of nesting, clauses (b) and (c) leave it unchanged, and in clause (d), the depth of nesting of  $U(A, B)$  and  $A \wedge B$  is the maximum of the depths of nesting of  $A$  and  $B$ . So, for example,  $\neg\varphi r(\neg Yr \wedge \varphi qY(q \rightarrow r))$  has depth of nesting of 2.
3. A UYF-formula is said to be a *YF-formula* if it does not involve  $U$ .
4. Let  $A$  be a formula and  $q$  an atom. A *bound occurrence* of  $q$  in  $A$  is one in a subformula of  $A$  of the form  $\varphi qB$ . All other occurrences of  $q$  in  $A$  are said to be *free*. An occurrence of  $q$  in  $A$  is said to be *pure past* in  $A$  if it is in a subformula of  $A$  of the form  $YB$  but not in a subformula of the form  $U(B, C)$ . So  $\varphi qA$  is well-formed if and only if all free occurrences of  $q$  in  $A$  are pure past.

An *assignment* is a map  $h$  providing a subset  $h(q)$  of  $\mathbb{N}$  for each atom  $q$ . If  $h, h'$  are assignments, and  $\bar{q}$  a tuple of atoms, we write  $h =_{\bar{q}} h'$  if  $h(r) = h'(r)$  for all atoms  $r$  not occurring in  $\bar{q}$ . If  $S \subseteq \mathbb{N}$  and  $q$  is an atom, we write  $h_{q/S}$  for the unique assignment  $h'$  satisfying:  $h' =_q h$ ,  $h'(q) = S$ .

For each assignment  $h$  and formula  $A$  of UYF we will define a subset  $h(A)$  of  $\mathbb{N}$ , the interpretation of  $A$  in  $\mathbb{N}$ . Intuitively,  $h(A) = \{n \in \mathbb{N} \mid A \text{ is true at } n \text{ under } h\} = \{n \in \mathbb{N} \mid (\mathbb{N}, <, h), n \models A\}$ . We will ensure that, whenever  $\varphi qA$  is well-formed,

$$(*) \quad h(\varphi qA) \text{ is the unique } S \subseteq \mathbb{N} \text{ such that } S = h_{q/S}(A).$$

**DEFINITION 97.** We define the semantics of UYF by induction. Let  $h$  be an assignment. If  $A$  is atomic then  $h(A)$  is already defined. We set:

- $h(\top) = \mathbb{N}$ .
- $h(\neg A) = \mathbb{N} \setminus h(A)$ .
- $h(YA) = h(A) + 1$ .
- $h(A \wedge B) = h(A) \cap h(B)$ .
- $h(U(A, B)) = \{n \in \mathbb{N} \mid \exists m > n(m \in h(A) \wedge \forall m'(n < m' < m \rightarrow m' \in h(B))\}$ .

- Finally, if  $\varphi qA$  is well-formed we define  $h(\varphi qA)$  as follows. First define assignments  $h_n$  ( $n \in \mathbb{N}$ ) by induction:  $h_0 = h, h_{n+1} = (h_n)_{q/h_n(A)}$ . Then  $h(\varphi qA) \stackrel{\text{def}}{=} \{n \in \mathbb{N} \mid n \in h_n(A)\} = \{n \in \mathbb{N} \mid n \in h_{n+1}(q)\}$ .

To establish (\*) we need a theorem.

THEOREM 98 (fixed point theorem).

1. Suppose that  $A$  is any UYF-formula and  $\varphi qA$  is well formed. Then if  $h$  is any assignment, there is a unique subset  $S = h(\varphi qA)$  of  $\mathbb{N}$  such that  $S = h_{q/S}(A)$ . Thus, regarding  $S \mapsto h_{q/S}(A)$  as a map  $\alpha : \wp(\mathbb{N}) \rightarrow \wp(\mathbb{N})$  (depending on  $h, A$ ),  $\alpha$  has a unique fixed point  $S \subseteq \mathbb{N}$ , and we have  $S = h(\varphi qA)$ . For any  $h, h(A) = h(q) \iff h(\varphi qA) = h(q)$ .
2. If  $q$  has no free occurrence in a formula  $A$  and  $g =_q h$ , then  $g(A) = h(A)$ .
3. If  $\varphi qA$  is well-formed and  $r$  is an atom not occurring in  $A$ , then for all assignments  $h, h(\varphi qA) = h(\varphi rA(q/r))$ , where  $A(q/r)$  denotes substitution by  $r$  for all free occurrences of  $q$  in  $A$ .

We define USF using the first-order connectives Until and Since as well as the fixed point operator. The logic UYF is just as expressive as USF:  $Yq$  is definable in USF by the formula  $S(q, \perp)$ , while  $S(p, q)$  is definable in UYF by  $\varphi rY(p \vee (q \wedge r))$ . Using UYF allows easier proofs and stronger results.

As an example, consider the formula

$$A = \varphi q(\neg Yq).$$

It is easy to see that  $A$  holds in a structure iff  $q$  holds exactly at the even numbered times.

#### 7.4 Decision Procedures

There are many uses for PTL (and extensions such as USF) in describing and verifying systems. Once again, an important task required in many of these applications is determining the validity (or equivalently satisfiability) of formulas. Because it is a widely-used logic and because the natural numbers afford a wide-variety of techniques of analysis, there are several quite different ways of approaching decision procedures here. The main avenues are via finite model properties, tableaux, automata and resolution techniques.

The first proof of the decidability of PTL was based on automata. The pioneer in the development of automata for use with infinite linear structures is Büchi in [1962]. He was interested in proving the decidability of  $S1S$  as

a very restricted version of second-order arithmetic. We will look at his proof briefly in subsection 7.5 below. By the time that temporal logic was being introduced to computer scientists in [Pnueli, 1977], it was well known (via [Kamp, 1968b]) that temporal logic formulae can be expressed in the appropriate second-order logic and so via *S1S* we had the first decision procedure for PTL (and *USF*).

Unfortunately, deciding *S1S* is non-elementarily complex (see [Robertson, 1974]) and so this is not an efficient way to decide PTL. Tableaux [Wolper, 1983; Lichtenstein *et al.*, 1985; Emerson, 1990], resolution [Fisher, 1997] and automata approaches (which we meet in subsection 7.5 below) can be much more efficient.

The first PSPACE algorithm for deciding PTL was given in [Sistla and Clarke, 1985]. This result uses a finite model property. We show that if a PTL formula  $A$  of length  $n$  is satisfiable over the natural numbers then it is also satisfiable in a non-linear model of size bounded by a certain exponential in  $n$ . The model has a linear part followed by a loop. It is a straightforward matter to guess the truths of atoms at the states on this structure and then check the truth of the formula. The guessing and checking can be done “on the fly”, i.e. simultaneously as we move along the structure in such a way that we do not need to store the whole structure. So we have an NPSpace algorithm which by the well-known result in [Savitch, 1970] can give us a PSPACE one.

In [Sistla and Clarke, 1985] it was also shown that deciding PTL is PSPACE-hard. This is done by encoding the running of any polynomial space bounded Turing machine into the logic.

For complexities of deciding the logic with strict  $U$  and the *USF* version see chapter 15 of [Gabbay *et al.*, 1994]. They are both PSPACE-complete. The same finite model property ideas are used.

## 7.5 Automata

Automata are finite state machines which are very promising objects to help with deciding the validity of temporal formulae. In some senses they are like formulae: they are finite objects and they distinguish some temporal structures—the ones which they accept—from other temporal structures in much the same way that formulae are true (at some point) in some structures are not in others. In other senses automata are like structures: they contain states and relate each state with some successor states. Being thus mid-way between formulae and structures allows automata to be used to answer questions—such as validity—about the relation between formulae and structures.

An automaton is called *empty* iff it accepts no structures and it turns out to be relatively easy to decide whether a given automaton is empty or not. This is surprising because empty automata can look quite complicated

in much the same way as unsatisfiable formulae can. This fact immediately suggests a possible decision procedure for temporal formulae. Given a formula we might be able to find an automaton which accepts exactly the structures which are models of the formula. If we now test the automaton for emptiness then we are effectively testing the formula for unsatisfiability. Validity of a formula corresponds to emptiness of an automaton equivalent to the negation of the formula.

This is the essence of the incredibly productive automata approach to theorem proving. We only look in detail at the case of PTL on natural numbers time.

The idea of (finite state) automata developed from pioneering attempts by Turing to formalize computation and by Kleene [1956] to model human psychology. The early work (see, for example, [Rabin and Scott, 1959]) was on finite state machines which recognized finite words. Such automata have provided a formal basis for many applications from text processing and biology to the analysis of concurrency. There has also been much mathematical development of the field. See [Perrin, 1990] for a survey.

The pioneer in the development of automata for use with infinite linear structures is Büchi in [1962] in proving the decidability of  $S1S$ . This gives one albeit inefficient decision procedure for PTL. There are now several useful ways of using the automata stepping stone for deciding the validity of PTL formulae.

The general idea is to translate the temporal formula into an automaton which accepts exactly the models of the formula and then to check for emptiness of the automaton. Variations arise when we consider that there are several different types of automata which we could use and that the translation from the formula can be done in a variety of ways.

Let us look at the automata first. For historical reasons we will switch now to a language  $\Sigma$  of letters rather than keep using a language of propositional atoms. The nodes of trees will be labelled by a single letter from  $\Sigma$ . In order to apply the results in this section we will later have to take the alphabet  $\Sigma$  to be  $2^P$  where  $P$  is the set of atomic propositions.

A  $\Sigma$  (*linear*) *Büchi automaton* is a 4-tuple  $A = (S, T, S_0, F)$  where

- $S$  is a finite non-empty set called the set of *states*,
- $T \subseteq S \times \Sigma \times S$  is the *transition table*,
- $S_0 \subseteq S$  is the *initial state set* and
- $F \subseteq S$  is the *set of accepting states*.

A *run* of  $A$  on an  $\omega$ -structure  $\sigma$  is a sequence of states  $(s_0, s_1, s_2, \dots)$  from  $S$  such that  $s_0 \in S_0$  and for each  $i < \omega$ ,  $(s_i, \sigma_i, s_{i+1}) \in T$ . We assume that automata never grind to a halt: i.e. we assume that for all  $s \in S$ , for all  $a \in \Sigma$ , there is some  $s' \in S$  such that  $(s, a, s') \in T$ .

We say that the automaton accepts  $\sigma$  iff there is a run  $\langle s_0, s_1, \dots \rangle$  such that  $s_i \in F$  for infinitely many  $i$ .

One of the most useful results about Büchi automata, is that we can complement them. That is given a Büchi automata  $A$  reading from the language  $\Sigma$  we can always find another  $\Sigma$  Büchi automata  $\bar{A}$  which accepts exactly the  $\omega$ -sequences which  $A$  rejects. This was first shown by Büchi in [1962] and was an important step on the way to his proof of the decidability of  $S1S$ . The automaton  $\bar{A}$  produced by Büchi's method is double exponential in the size of  $A$  but more recent work in [Sistla *et al.*, 1987] shows that complementation of Büchi automata can always be singly exponential.

As we will see below, it is easy to complement an automaton if we can find a deterministic equivalent. This means an automaton with a unique initial state and a transition table  $T \subseteq S \times \Sigma \times S$  which satisfies the property that for all  $s \in S$ , for all  $a \in \Sigma$ , there is a unique  $s' \in S$  such that  $(s, a, s') \in T$ . A deterministic automaton will have a unique run on any given structure.

Two automata are equivalent iff they accept exactly the same structures.

The problem with Büchi automata is that it is not always possible to find a deterministic equivalent. A very short argument, see example 4.2 in [Thomas, 1990], shows that the non-deterministic  $\{a, b\}$  automaton which recognizes exactly the set

$$L = \{\sigma \mid a \text{ appears only a finite number of times in } \sigma\}$$

can have no deterministic equivalent.

One of our important tasks is to decide whether a given automaton is empty i.e. accepts no  $\omega$ -structures. For Büchi automata this can be done in linear time [Emerson and Lei, 1985] and co-NLOGSPACE [Vardi and Wolper, 1994].

The lack of a determinization result for Büchi automata led to a search for a class of automata which is as expressive as the class of Büchi automata but which is closed under finding deterministic equivalents. Muller automata were introduced by Muller in [Muller, 1963] and in [Rabin, 1972] variants, now called Rabin automata, were introduced.

The difference is that the accepting condition can require that certain states do not come up infinitely often. There are several equivalent ways of formalizing this. The Rabin method is, for a  $\Sigma$ -automata with state set  $S$ , to use a set  $\mathcal{F}$ , called the set of *accepting pairs*, of pairs of sets of states from  $S$ , i.e.  $\mathcal{F} \subseteq \wp(S) \times \wp(S)$ .

We say that the Rabin automaton  $A = (S, S_0, T, \mathcal{F})$  accepts  $\sigma$  iff there is some run  $\langle s_0, s_1, s_2, \dots \rangle$  (as defined for Büchi automata) and some pair  $(U, V) \in \mathcal{F}$  such that no state in  $V$  is visited infinitely often but there is some state in  $U$  visited infinitely often.

In fact, Rabin automata add no expressive power compared to Büchi automata, i.e. for every Rabin automaton there is an equivalent Büchi automaton. The translation [Choueka, 1974] is straightforward and, as it

essentially just involves two copies of the Rabin automata in series with a once-only non-deterministic transition from the first to the second, it can be done in polynomial time. The converse equivalence is obvious.

The most important property of the class of Rabin automata is that it is closed under determinization. In [1966], McNaughton, showed that any Büchi automaton has a deterministic Rabin equivalent. There are useful accounts of McNaughton's theorem in [Thomas, 1990] and [Hodkinson, 200]. McNaughton's construction is doubly exponential. It follows from McNaughton's result that we can find a deterministic equivalent of any Rabin automaton: simply first find a Büchi equivalent and then use the theorem.

The determinization result gives us an easy complementation result for Rabin automata: given a Rabin automata we can without loss of generality assume it is deterministic and complementing a deterministic automaton is just a straightforward negation of the acceptance criteria.

To decide whether Rabin automata are empty can be done with almost the same procedure we used for Büchi case. Alternatively, one can determinize the automaton  $A$ , and translate the deterministic equivalent into a deterministic Rabin automaton  $A'$  recognizing  $\omega$ -sequences from the one symbol alphabet  $\{a_0\}$  such that  $A'$  accepts some sequence iff  $A$  does. It is very easy to tell if  $A'$  is empty.

### *Translating formulae into Automata*

The first step in using automata to decide a temporal formula is to translate the temporal formula into an equivalent automata: i.e. one that accepts exactly the models of the formula. There are direct ways of making this translation, e.g., in [Sherman *et al.*, 1984] (via a nonelementarily complex procedure). However, it is easier to understand some of the methods which use a stepping stone in the translation:  $S1S$ .

We have seen that the translation from PTL into  $S1S$  is easy. The translation of  $S1S$  into an automaton is also easy, given McNaughton's result: it is via a simple induction. Suppose that the  $S1S$  sentence uses constants from the finite set  $P$ . We proceed by induction on the construction of the sentence. The automaton for  $p \subseteq q$  simply keeps checking that  $p \rightarrow q$  is true of the current state and falls into a fail state sink if not. The other base cases, of  $p = q$  and  $succ(p, q)$  are just as easy. Conjunction requires a standard construction of conjoining automata using the product of the state sets. Negation can be done using McNaughton's result to determinize the automaton for the negated subformula. It is easy to find the complement of a deterministic automaton. The case of an existential quantification, e.g.,  $\exists y \rho(y)$ , is done by simply using non-determinism to guess the truth of the quantified variable at each step.

Putting together the results above gives us several alternative approaches to deciding validity of PTL formulae. One route is to translate the formula

into  $S1S$ , translate the  $S1S$  formula into a Büchi automaton as above and then check whether that is empty.

The quickest route is via the alternating automata idea of [Brzozowski and Leiss, 1980] — a clever variation on the automata idea. By translating a formula into one of these automata, and then using a guess and check on the fly procedure, we need only check a polynomial number of states (in the size of  $\phi$ ) and then (nondeterministically) move on to another such small group of states. This gives us a PSPACE algorithm. From the results of [Sistla and Clarke, 1985], we know this is best possible as a decision procedure.

### *Other Uses of Automata*

The decision algorithm above using the translation into the language  $S1S$  can be readily extended to allow for past operators or fixed point operators or both to appear in the language. This is because formulae using these operators can be expressed in  $S1S$ .

Automata do not seem well suited to reasoning about dense time or general linear orders. However, the same strategy as we used for PTL also works for the decidability of branching time logics such as CTL\*. The only difference is that we must use tree automata. These were invented by Rabin in his powerful results showing the decidability of  $S2S$ , the second-order logic of two successors. See [Gurevich, 1985] for a nice introduction.

## 8 EXECUTABLE TEMPORAL LOGIC

Here we describe a useful paradigm in executable logic: that of the *declarative past and imperative future*. A future statement of temporal logic can be understood in two ways: the declarative way, that of describing the future as a temporal extension; and the imperative way, that of making sure that the future will happen the way we want. Since the future has not yet happened, we have a language which can be both declarative and imperative. We regard our theme as a natural meeting between the imperative and declarative paradigms.

More specifically, we describe a temporal logic with Since, Until and fixed point operators. The logic is based on the natural numbers as the flow of time and can be used for the specification and control of process behaviour in time. A specification formula of this logic can be automatically rewritten into an executable form. In an executable form it can be used as a program for controlling process behaviour. The executable form has the structure ‘If  $A$  holds in the past then do  $B$ ’. This structure shows that declarative and imperative programming can be integrated in a natural way.

Let  $\mathcal{E}$  be an environment in which a program  $\mathcal{P}$  is operating. The exact nature of the environment or the code and language of the program are not

immediately relevant for our purposes. Suppose we make periodic checks at time  $t_0, t_1, t_2, t_3, \dots$  on what is going on in the environment and what the program is doing. These checks could be made after every unit execution of the program or at some key times. The time is not important to our discussion. What is important is that we check at each time the truth values of some propositions describing features of the environment and the program. We shall denote these propositions by  $a_1, \dots, a_m, b_1, \dots, b_k$ . These propositions, which we regard as units taking truth values  $\top$  (true) or  $\perp$  (false) at every checkpoint, need not be expressible in the language of the program  $\mathcal{P}$ , nor in the language used to describe the environment. The program may, however, in its course of execution, change the truth values of some of the propositions. Other propositions may be controlled only by the environment. Thus we assume that  $a_1, \dots, a_m$  are capable of being influenced by the program while  $b_1, \dots, b_k$  are influenced by the environment. We also assume that when at checktime  $t_n$  we want the program to be executed in such a way as to make the proposition  $a_i$  true, then it is possible to do so. We express this command by writing **exec** ( $a_i$ ). For example,  $a_1$  can be ‘print the screen’ and  $b_1$  can be ‘there is a read request from outside’;  $a_1$  can be controlled by the program while  $b_1$  cannot. **exec** ( $a_1$ ) will make  $a_1$  true.

To illustrate our idea further, we take one temporal sentence of the form

$$(1) \quad G[\bullet a \Rightarrow Xb].$$

$\bullet$  is the ‘yesterday’ operator,  $X$  is the ‘tomorrow’ operator, and  $G$  is the ‘always in the future’ operator. One can view (1) as a wff of temporal logic which can be either true or false in a temporal model. One can use a temporal axiom system to check whether it is a temporal theorem etc. In other words, we treat it as a formula of logic.

There is another way of looking at it. Suppose we are at time  $n$ . In a real ticking forward temporal system, time  $n + 1$  (assume that time is measured in days) has not happened yet. We can find the truth value of  $\bullet a$  by checking the past. We do not know yet the value of  $Xb$  because tomorrow has not yet come. Checking the truth value of  $\bullet a$  is a declarative reading of  $\bullet a$ . However, we need not read  $Xb$  declaratively. We do not need to wait and see what happens to  $b$  tomorrow. Since tomorrow has not yet come, we can make  $b$  true tomorrow if we want to, and are able to. We are thus reading  $Xb$  imperatively: ‘make  $b$  true tomorrow’.

If we are committed to maintaining the truth of the specification (1) throughout time, then we can read (1) as: ‘at any time  $t$ , if  $\bullet a$  holds then execute  $Xb$ ’, or schematically, ‘if declarative past then imperative future’. This is no different from the Pascal statement

```
if x<5 then x:=x+1
```

In our case we involve whole formulae of logic.



The above is our basic theme. This section makes it more precise. The rest of this introduction sets the scene for it, and the conclusion will describe existing implementations. Let us now give several examples:

EXAMPLE 99 (Simplified Payroll). Mrs Smith is running a babysitter service. She has a list of reliable teenagers who can take on a babysitting job. A customer interested in a babysitter would call Mrs Smith and give the date on which the babysitter is needed. Mrs Smith calls a teenager employee of hers and arranges for the job. She may need to call several of her teenagers until she finds one who accepts. The customer pays Mrs Smith and Mrs Smith pays the teenager. The rate is £10 per night unless the job requires overtime (after midnight) in which case it jumps to £15.

Mrs Smith uses a program to handle her business. The predicates involved are the following:

$A(x)$	$x$ is asked to babysit
$B(x)$	$x$ does a babysitting job
$M(x)$	$x$ works after midnight
$P(x, y)$	$x$ is paid $y$ pounds .

In this set-up,  $B(x)$  and  $M(x)$  are controlled mainly by the environment and  $A(x)$  and  $P(x, y)$  are controlled by the program.

We get a temporal model by recording the history of what happens with the above predicates. Mrs Smith laid out the following (partial) specification:

1. Babysitters are not allowed to take jobs three nights in a row, or two nights in a row if the first night involved overtime.
2. Priority in calling is given to babysitters who were not called before as many times as others.
3. Payment should be made the next day after a job is done.

Figure 9 is an example of a partial model of what has happened to a babysitter called Janet. This model may or may not satisfy the specification.

We would like to be able to write down the specification in an intuitive temporal language (or even English) and have it automatically transformed into an executable program, telling us what to do day by day.

EXAMPLE 100 (J. Darlington, L. While [1987]). Consider a simple program  $\mathcal{P}$ , written in a rewrite language, to merge two queues. There are two merge rules:

**R1** Merge( $a.x, y$ ) =  $a.merge(x,y)$ ;

**R2** Merge( $x,a.y$ ) =  $a.merge(x,y)$ .



EXAMPLE 101 (Loop checking in Prolog). Imagine a Prolog program  $\mathcal{P}$  and imagine predicates  $A_1, \dots, A_m$  describing at each step of execution which rule is used and what is the current goal and other relevant data. Let  $B$  describe the history of the computation. This can be a list of states defined recursively. The loop checking can be done by ensuring that certain temporal properties hold throughout the computation. We can define in this set-up any loop-checking system we desire and change it during execution.

In the above examples, the propositions  $A_i, B_j$  change the truth value at each checktime  $t_k$ . We thus obtain a natural temporal model for these propositions (see Fig. 10).

$$\begin{array}{rcl}
 & & \vdots \\
 3 & a_1 = \perp & b_2 = \top \\
 2 & a_1 = \perp & b_2 = \top \\
 1 & a_1 = \top & b_2 = \top \\
 0 & a_1 = \top & b_2 = \perp
 \end{array}$$

Figure 10. An example temporal model.

In the above set-up the programmer is interested in influencing the execution of the program within the non-deterministic options available in the programming language. For example, in the merge case one may want to say that if the left queue is longer than the right queue then use the left merge next. In symbols

$$G[B \Rightarrow XA1].$$

In the Prolog case, we may want to specify what the program should do in case of loops, i.e.

$$G[C \wedge PC \Rightarrow D],$$

where  $C$  is a complex proposition describing the state of the environment of interest to us ( $P$  is the ‘in the past’ operator).  $C \wedge PC$  indicate a loop and  $D$  says what is to be done. The controls may be very complex and can be made dependent on the data and to change as we go along.

Of course in many cases our additional controls of the execution of  $\mathcal{P}$  may be synthesized and annotated in  $\mathcal{P}$  to form a new program  $\mathcal{P}^*$ . There are several reasons why the programmer may not want to do that:

1. The programming language may be such that it is impossible or not natural to synthesize the control in the program. We may lose in clarity and structure.

2. Changes in the control structure may be expensive once the program  $\mathcal{P}^*$  is defined and compiled.
3. It may be impossible to switch controlling features on and off during execution, i.e. have the control itself respond to the way the execution flows.
4. A properly defined temporal control module may be applicable as a package to many programming languages. It can give both practical and theoretical advantages.

In this section we follow option 4 above and develop an executable temporal logic for interactive systems. The reader will see that we are developing a logic here that on the one hand can be used for specification (of what we want the program to do) and on the other hand can be used for execution. (How to pass from the specification to the executable part requires some mathematical theorems.) Since logically the two formulations are equivalent, we will be able to use logic and proof theory to prove correctness.

This is what we have in mind:

1. We use a temporal language to specify the desirable behaviour of  $\{a_i, b_j\}$  over time. Let  $\mathcal{S}$  be the specification as expressed in the temporal language (e.g.  $G[b_2 \Rightarrow Xa_1]$ ).
2. We rewrite automatically  $\mathcal{S}$  into  $\mathcal{E}$ ,  $\mathcal{E}$  being an executable temporal module. The program  $\mathcal{P}$  can communicate with  $\mathcal{E}$  at each checktime  $t_i$  and get instructions on what to do.

We have to prove that:

- if  $\mathcal{P}$  follows the instructions of  $\mathcal{E}$  then any execution sequence satisfies  $\mathcal{S}$ , i.e. the resulting temporal model for  $\{a_i, b_j\}$  satisfies the temporal formula  $\mathcal{S}$ ;
- any execution sequence satisfying  $\mathcal{S}$  is non-deterministically realizable using  $\mathcal{P}$  and  $\mathcal{E}$ .

The proofs are tough!

Note that our discussion also applies to the case of shared resources. Given a resource to be shared by several processes, the temporal language can specify how to handle concurrent demands by more than one process. This point of view is dual to the previous one. Thus in the merge example, we can view the merge program as a black box which accepts items from two processes (queues), and the specification organizes how the box (program) is to handle that merge. We shall further observe that since the temporal language can serve as a metalanguage for the program  $\mathcal{P}$  (controlling its

execution),  $\mathcal{P}$  can be completely subsumed in  $\mathcal{E}$ . Thus the temporal language itself can be used as an imperative language ( $\mathcal{E}$ ) with an equivalent specification element  $\mathcal{S}$ . Ordinary Horn logic programming can be obtained as a special case of the above.

We can already see the importance of our logic from the following point of view. There are two competing approaches to programming; the declarative one as symbolized in logic programming and Prolog, and the imperative one as symbolized in many well-known languages. There are advantages to each approach and at first impression there seems to be a genuine conflict between the two. The executable temporal logic described in this section shows that these two approaches can truly complement each other in a natural way.

We start with a declarative specification  $\mathcal{S}$ , which is a formula of temporal logic.  $\mathcal{S}$  is transformed into an executable form which is a conjunction of expressions of the form

**hold**  $C$  **in the past**  $\Rightarrow$  **execute**  $B$  **now**.

At any moment of time, the past is given to us as a database; thus **hold**( $C$ ) can be evaluated as a goal from this database in a Prolog program. **execute**( $B$ ) can be performed imperatively. This creates more data for the database, as the present becomes past. Imperative languages have a little of this feature, e.g. **if**  $x < 5$  **let**  $x := x + 1$ . Here **hold**( $C$ ) equates to  $x < 5$  and **execute**( $B$ ) equates to  $x := x + 1$ . The  $x < 5$  is a very restricted form of a declarative test. On the other hand Prolog itself allows for imperative statements. Prolog clauses can have the form

**write**(Term)  $\Rightarrow$  b

and the goal **write**(Term) is satisfied by printing. In fact, one can accomplish a string of imperative commands just by stringing goals together in a clever way.

We thus see our temporal language as a pointer in the direction of unifying in a logical way the declarative and imperative approaches. The temporal language can be used for planning. If we want to achieve  $B$  we try to execute  $B$ . The temporal logic will give several ways of satisfying **execute**( $B$ ) while **hold**( $C$ ) remains true. Any such successful way of executing  $B$  is a plan. In logical terms we are looking for a model for our atoms but since we are dealing with a temporal model and the atoms can have an imperative meaning we get a plan. We will try and investigate these points further.

### 8.1 The logic USF

We describe a temporal system for specification and execution. The logic is USF which we met briefly in section 6 above. It contains the temporal

connectives *since*( $S$ ) and *until*( $U$ ) together with a fixed point operator  $\varphi$ . The formulae of USF are used for specifying temporal behaviour and these formulae will be syntactically transformed into an executable form. We begin with the definitions of the syntax of USF. There will be four types of well-formed formulae, pure future formulae (talking only about the strict future), pure past formulae (talking only about the strict past), pure present formulae (talking only about the present) and mixed formulae (talking about the entire flow of time).

**DEFINITION 102** (Syntax of USF for the propositional case). Let  $Q$  be a sufficiently large set of *atoms* (atomic propositions). Let  $\wedge, \vee, \neg, \Rightarrow, \top, \perp$  be the usual classical connectives and let  $U$  and  $S$  be the temporal connectives and  $\varphi$  be the fixed point operator. We define by induction the notions of

- $wff$  (well-formed formula);
- $wff^+$  (pure future wff);
- $wff^-$  (pure past wff);
- $wff^0$  (pure present wff).

1. An atomic  $q \in Q$  is a pure present wff and a wff. Its atoms are  $q$ .
2. Assume  $A$  and  $B$  are wffs with atoms  $\{q_1, \dots, q_n\}$  and  $\{r_1, \dots, r_m\}$  respectively. Then  $A \wedge B$ ,  $A \vee B$ ,  $A \Rightarrow B$ ,  $U(A, B)$  and  $S(A, B)$  are wffs with atoms  $\{q_1, \dots, q_n, r_1, \dots, r_m\}$ .
  - (a) If both  $A$  and  $B$  are in  $wff^0 \cup wff^+$ , then  $U(A, B)$  is in  $wff^+$ .
  - (b) If both  $A$  and  $B$  are in  $wff^0 \cup wff^-$ , then  $S(A, B)$  is in  $wff^-$ .
  - (c) If both  $A$  and  $B$  are in  $wff^*$ , then so are  $A \wedge B$ ,  $A \vee B$ ,  $A \Rightarrow B$ , where  $wff^*$  is one of  $wff^+$ ,  $wff^-$  or  $wff^0$ .
3.  $\neg A$  is also a wff and it is of the same type as  $A$  with the same atoms as  $A$ .
4.  $\top$  (truth) and  $\perp$  (falsity) are wffs in  $wff^0$  with no atoms.
5. If  $A$  is a wff in  $wff^-$  (pure past) with atoms  $\{q, q_1, \dots, q_n\}$  then  $(\varphi q)A$  is a pure past wff (i.e. in  $wff^-$ ) with the atoms  $\{q_1, \dots, q_n\}$ .

The intended model for the above propositional temporal language is the set of natural numbers  $\mathbb{N} = 0, 1, 2, 3, \dots$  with the ‘smaller than’ relation  $<$  and variables  $P, Q \subseteq \mathbb{N}$  ranging over subsets. We allow quantification  $\forall x \exists y$  over elements of  $\mathbb{N}$ . So really we are dealing with the monadic language of the model  $(\mathbb{N}, <, =, 0, P_i, Q_j \subseteq \mathbb{N})$ . We refer to this model also as the *non-negative integers (flow of) time*. A formula of the monadic language will in general have free set variables, and these correspond to the atoms of temporal formulae. See Volume 1 for more details.

DEFINITION 103 (Syntax of USF for the predicate case). Let  $Q^* = \{Q_{n_1}^1, Q_{n_2}^2, \dots\}$  be a set of predicate symbols.  $Q_{n_i}^i$  is a symbol for an  $n_i$ -place predicate. Let  $f^* = \{f_{n_1}^1, f_{n_2}^2, \dots\}$  be a set of function symbols.  $f_{n_i}^i$  is a function symbol for an  $n_i$ -place function. Let  $V^* = \{v_1, v_2, \dots\}$  be a set of variables. Let  $\wedge, \vee, \neg, \Rightarrow, \top, \perp, \forall, \exists$  be the usual classical connectives and quantifiers and let  $U$  and  $S$  be the temporal connectives and  $\varphi$  be the fixed point operator. We define by induction the notions of:

- $wff\{x_1, \dots, x_n\}$     wff with free variables  $\{x_1, \dots, x_n\}$ ;
- $wff^+\{x_1, \dots, x_n\}$     pure future wff with the indicated free variables;
- $wff^-\{x_1, \dots, x_n\}$     pure past wff with the indicated free variables;
- $wff^0\{x_1, \dots, x_n\}$     pure present wff with the indicated free variables;
- $term\{x_1, \dots, x_n\}$     term with the indicated free variables.

1.  $x$  is a term in  $term\{x\}$ , where  $x$  is a variable.
2. If  $f$  is an  $n$ -place function symbol and  $t_1, \dots, t_n$  are terms with variables  $V_1^*, \dots, V_n^* \subseteq V^*$  respectively, then  $f(t_1, \dots, t_n)$  is a term with variables  $\bigcup_{i=1}^n V_i^*$ .
3. If  $Q$  is an  $n$ -place atomic predicate symbol and we have  $t_1, \dots, t_n$  as terms with variables  $V_1^*, \dots, V_n^* \subseteq V^*$  respectively, then  $Q(t_1, \dots, t_n)$  is an atomic formula with free variables,  $\bigcup_{i=1}^n V_i^*$ . This formula is pure present as well as a wff.
4. Assume  $A, B$  are formulae with free variables  $\{x_1, \dots, x_n\}$  and  $\{y_1, \dots, y_m\}$  respectively. Then  $A \wedge B, A \vee B, A \Rightarrow B, U(A, B)$  and  $S(A, B)$  are wffs with the free variables  $\{x_1, \dots, x_n, y_1, \dots, y_m\}$ .
  - (a) If both  $A$  and  $B$  are in  $wff^0 \cup wff^+$ , then  $U(A, B)$  is in  $wff^+$ .
  - (b) If both  $A$  and  $B$  are in  $wff^0 \cup wff^-$ , then  $S(A, B)$  is in  $wff^-$ .
  - (c) If both  $A$  and  $B$  are in  $wff^*$ , then so are  $A \wedge B, A \vee B, A \Rightarrow B$ , where  $wff^*$  is one of  $wff^+, wff^-$  or  $wff^0$ .
5.  $\neg A$  is also a wff and it is of the same type and has the same free variables as  $A$ .
6.  $\top$  and  $\perp$  are in  $wff^0$  with no free variables.
7. If  $A$  is a formula in  $wff^*\{x, y_1, \dots, y_m\}$  then  $\forall x A$  and  $\exists x A$  are wffs in  $wff^*\{y_1, \dots, y_m\}$ .
8. If  $(\varphi q)A(q, q_1, \dots, q_n)$  is a pure past formula of propositional USF as defined in Definition 102, and if  $B_i \in wff V_i^*, i = 1, \dots, m$ , as defined in the present Definition 103, then  $A' = (\varphi q)A(q, B_1, \dots, B_m)$  is a wff in  $wff \bigcup_{i=1}^m V_i^*$ . If all of the  $B_i$  are pure past, then so is  $A'$ .

9. A wff  $A$  is said to be *essentially propositional* iff there exists a wff  $B(q_1, \dots, q_n)$  of propositional USF and wffs  $B_1, \dots, B_n$  of classical predicate logic such that  $A = B(B_1, \dots, B_n)$ .

REMARK 104. Notice that the fixed point operator  $(\varphi x)$  is used in propositional USF to define the new connectives, and after it is defined it is exported to the predicate USF. We can define a language HTL which will allow fixed-point operations on predicates as well; we will not discuss this here.

DEFINITION 105. We define the semantic interpretation of propositional USF in the monadic theory of  $(\mathbb{N}, <, =, 0)$ . An assignment  $h$  is a function associating with each atom  $q_i$  of USF a subset  $h(q_i)$  of  $\mathbb{N}$  (sometimes denoted by  $Q_i$ ).  $h$  can be extended to any wff of USF as follows:

$$\begin{aligned} h(A \wedge B) &= h(A) \cap h(B); \\ h(A \vee B) &= h(A) \cup h(B); \\ h(\neg A) &= \mathbb{N} - h(A); \\ h(A \Rightarrow B) &= (\mathbb{N} - h(A)) \cup h(B); \\ h(U(A, B)) &= \{t | \exists s > t (s \in h(A) \text{ and } \forall y (t < y < s \Rightarrow y \in h(B)))\}; \\ h(S(A, B)) &= \{t | \exists s < t (s \in h(A) \text{ and } \forall y (s < y < t \Rightarrow y \in h(B)))\}. \end{aligned}$$

The meaning of  $U$  and  $S$  just defined is the Until and Since of English, i.e. ‘ $B$  is true until  $A$  becomes true’ and ‘ $B$  is true since  $A$  was true’, as in Fig. 11 (notice the existential meaning in  $U$  and  $S$ ).

Finally we have the fixed point operator

$$h((\varphi q)A(q, q_i)) = \{n | n \in Q_n\}$$

where the sets  $Q_n \subseteq \mathbb{N}$  are defined inductively by

$$\begin{aligned} Q_0 &= h(A) \\ Q_{(n+1)} &= h_n(A(q, q_i)) \end{aligned}$$

where for  $n \geq 0$

$$\begin{aligned} h_n(r) &= h(r) \quad \text{for } r \neq q \\ &Q_n \quad \text{for } r = q. \end{aligned}$$

This is an inductive definition. If  $n$  is a natural number, we assume inductively that we know the truth values of the formula  $(\varphi q)A(q, q_i)$  at each  $m < n$ ; then we obtain its value at  $n$  by first changing the assignment so that  $q$  has the same values as  $(\varphi q)A$  for all  $m < n$ , and then taking the new value of  $A(q, q_i)$  at  $n$ . Since  $A$  is pure past, the values of  $q$  at  $m \geq n$  do not matter. Hence the definition is sound. So  $(\varphi q)A$  is defined in terms of its own previous values.



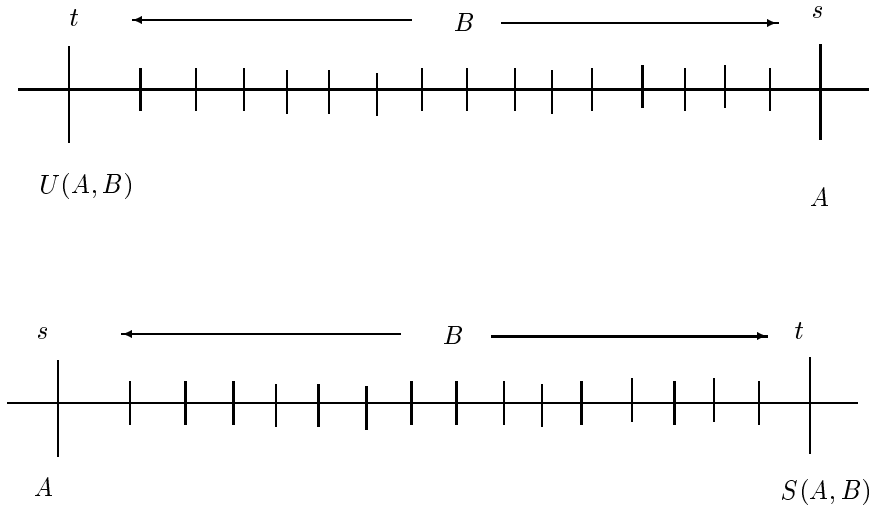


Figure 11.

This gives a fixed point semantics to formulae  $(\varphi q)A(q, q_i)$ , in the following sense. Suppose we have an assignment  $h$ . For any subset  $S$  of  $\mathbb{N}$ , let  $h_S$  be the assignment given by  $h_S(r) = h(r)$  if  $r \neq q$ , and  $S$  if  $r = q$ . Then given  $A$  as above, we obtain a function  $f : \wp\mathbb{N} \Rightarrow \wp\mathbb{N}$ , given by  $f(S) = h_S(A)$ .  $f$  depends on  $h$  and  $A$ .

It is intuitively clear from the above that if  $S = h((\varphi q)A)$  then  $f(S) = S$ , and that  $S$  is the unique solution of  $f(x) = x$ . So  $h((\varphi q)A)$  is the unique fixed point of  $f$ . This is what we mean when we say that  $\varphi$  has a fixed point semantics. There are some details to be checked, in particular that the value at  $n$  of any past formula (even a complicated one involving  $\varphi$ ) depends only on the values of its atoms at each  $m < n$ . For a full proof see [Hodkinson, 1989].

**DEFINITION 106** (Semantic definition of predicate USF). Let  $D$  be a non-empty set, called the domain, and  $g$  be a function assigning the following:

1. for each  $m$ -place function symbol  $f$  and each  $n \in \mathbb{N}$  a function  $g(n, f) : D^m \Rightarrow D$ ;
2. for each variable  $x$  and each  $n \in \mathbb{N}$ , an element  $g(n, x) \in D$ ;
3. for each  $m$ -place predicate symbol  $Q$  and each  $n \in \mathbb{N}$ , a function  $g(n, Q) : D^m \Rightarrow \{0, 1\}$ .

The function  $g$  can be extended to a function  $g(n, A)$ , giving a value in  $\{0, 1\}$  for each wff  $A(x_1, \dots, x_n)$  of the predicate USF as follows:

1.  $g(n, f(t_1, \dots, t_m)) = g(n, f)(g(n, t_1), \dots, g(n, t_m))$ ;
2.  $g(n, Q(t_1, \dots, t_m)) = g(n, Q)(g(n, t_1), \dots, g(n, t_m))$ ;
3.  $g(n, A \wedge B) = 1$  iff  $g(n, A) = 1$  and  $g(n, B) = 1$ ;
4.  $g(n, A \vee B) = 1$  iff either  $g(n, A) = 1$  or  $g(n, B) = 1$  or both;
5.  $g(n, A \Rightarrow B) = 1$  iff either  $g(n, A) = 0$  or  $g(n, B) = 1$  or both;
6.  $g(n, \neg A) = 1$  iff  $g(n, A) = 0$ ;
7.  $g(n, \top) = 1$  and  $g(n, \perp) = 0$  for all  $n$ ;
8.  $g(n, U(A, B)) = 1$  iff for some  $m > n$ ,  $g(m, A) = 1$  and for all  $n < k < m$ ,  $g(k, B) = 1$ ;
9.  $g(n, S(A, B)) = 1$  iff for some  $m < n$ ,  $g(m, A) = 1$  and for all  $m < k < n$ ,  $g(k, B) = 1$ ;
10.  $g(n, \forall x A(x)) = 1$  for a variable  $x$  iff for all  $g'$  such that  $g'$  gives the same values as  $g$  to all function symbols and all predicate symbols and all variables different from  $x$ , we have  $g'(n, A(x)) = 1$ ;
11.  $g(n, \exists x A(x)) = 1$  for a variable  $x$  iff for some  $g'$  such that  $g'$  gives the same values as  $g$  to all function symbols and all predicate symbols and all variables different from  $x$ , we have  $g'(n, A(x)) = 1$ ;
12. let  $(\varphi q)A(q, q_1, \dots, q_m)$  be a pure past formula of propositional USF, and let  $B_i \in wffV_i^*$  for  $i = 1, \dots, m$ . We want to define  $g(n, A')$ , where  $A' = (\varphi q)A(q, B_1, \dots, B_m)$ . First choose an assignment  $h$  such that  $h(q_i) = \{n \in \mathbb{N} | g(n, B_i) = 1\}$ . Then define  $g(n, A') = 1$  iff  $n \in h((\varphi q)A(q, q_1, \dots, q_m))$ .

REMARK 107. If we let  $h_g^*(A)$  be the set  $\{n | g(n, A) = 1\}$  we get a function  $h^*$  like that of Definition 105.

EXAMPLE 108. Let us evaluate  $(\varphi x)A(x)$  for  $A(x) = H\neg x$ , where  $Hx = \neg S(\neg x, \top)$ ; see Example 109.1. We work out the value of  $(\varphi x)A(x)$  at each  $n$ , by induction on  $n$ . If we know its values for all  $m < n$ , we assume that the atom  $x$  has the same value as  $(\varphi x)A(x)$  for  $m < n$ . We then calculate the value of  $A(x)$  at  $n$ . So, really,  $(\varphi x)A(x)$  is a definition by recursion.

Since  $H\neg x$  is a pure past formula, its value at 0 is known and does not depend on  $x$ . Thus  $A(x)$  is true at 0. Hence  $(\varphi x)A(x)$  is true at 0.

Let us compute  $A(x)$  at 1. Assume that  $x$  is true at 0. Since  $A(x)$  is pure past, its value at 1 depends on the value of  $x$  at 0, which we know. It

does not depend on the value of  $x$  at  $n \geq 1$ . Thus at 1,  $(\varphi x)A(x) = A(x) = H\neg\top = \perp$ .

Assume inductively that we know the values of  $(\varphi x)A(x)$  at  $0, 1, \dots, n$ , and suppose that  $x$  also has these values at  $m \leq n$ . We compute  $A(x)$  at  $n + 1$ . This depends only on the values of  $x$  at points  $m \leq n$ , which we know. Hence  $A(x)$  at  $n + 1$  can be computed; for our example we get  $\perp$ . So  $(\varphi x)A(x)$  is false at  $n + 1$ . Thus  $(\varphi x)H\neg x$  is (semantically) equivalent to  $H\perp$ , because  $H\perp$  is true at 0 and nowhere else.

Another way to get the answer is to use the fixed point semantics directly. Let  $f(S) = h(A)$ , where  $h(x) = S$ , as above. Then by definition of  $f$  and  $g$ ,

$$\begin{aligned} f(S) &= \{n \in \mathbb{N} \mid \neg \exists m < n (m \in S \wedge \forall k (m < k < n \Rightarrow k \in h(\top)))\} \\ &= \{n \in \mathbb{N} \mid \forall m < n (m \notin S)\}. \end{aligned}$$

So  $f(S) = S$  iff  $S = \{0\}$ . Hence the fixed point is  $\{0\}$ , as before.

Let us evaluate  $(\varphi x)B(x)$  where  $B(x) = S(S(x, a), \neg a)$ . At time 0 the value of  $B(x)$  is  $\perp$ . Let  $x$  be  $\perp$  at 0. At time 1 the value of  $B(x)$  is  $S(S(\perp, a), \neg a) = S(\perp, \neg a) = \perp$ . Let  $x$  be  $\perp$  at 1 etc. . . . It is easy to see that  $(\varphi x)B(x)$  is independent of  $a$  and is equal to  $\perp$ .

EXAMPLE 109. We give examples of connectives definable in this system.

1. The basic temporal connectives are defined as follows:

Connective	Meaning	Definition
$\bullet q$	$q$ was true 'yesterday'	$S(q, \perp)$
$Xq$	$q$ will be true 'tomorrow'	$U(q, \perp)$
$Gq$	$q$ 'will always' be true	$\neg U(\neg q, \top)$
$Fq$	$q$ 'will sometimes' be true	$U(q, \top)$
$Hq$	$q$ 'was always' true	$\neg S(\neg q, \top)$
$Pq$	$q$ 'was sometimes' true	$S(q, \top)$

Note that at 0, both  $\bullet q$  and  $Pq$  are false.

2. The first time point (i.e.  $n = 0$ ) can be identified as the point at which  $H\perp$  is true.
3. The fixed point operator allows us to define non-first-order definable subsets. For example,  $e = (\varphi x)(\bullet \bullet x \vee H\perp)$  is a constant true exactly at the even points  $\{0, 2, 4, 6, \dots\}$ .
4.  $S(A, B)$  can be defined from  $\bullet$  using the fixed point operator.:

$$S(A, B) = (\varphi x)(\bullet A \vee \bullet (x \wedge B))'$$

5. If we have

$$\text{block}(a, b) \triangleq (\varphi x)S(b \wedge S(a \wedge (x \vee H\perp \vee HH\perp), a), b)$$

then  $\text{block}(a, b)$  says that we have the sequence of the form

$$(\text{block of } bs) + (\text{block of } as) + \dots$$

recurring in the pure past, beginning yesterday with  $b$  and going into the past. In particular  $\text{block}(a, b)$  is false at time 0 and time 1 because the smallest recurring block is  $(b, a)$  which requires two points in the past.

**DEFINITION 110** (Expressive power of USF). Let  $\Psi(t, Q_1, \dots, Q_n)$  be a formula in the monadic language of  $(\mathbb{N}, <, =, 0, Q_1, \dots, Q_n \subseteq \mathbb{N})$ . Let  $Q = \{t \mid \Psi(t, Q_i) \text{ is true}\}$ . Then  $Q$  is said to be monadic first-order definable from  $Q_i$ .

**EXAMPLE 111.**  $\text{even} = \{0, 2, 4, \dots\}$  is not monadic first-order definable from any family of finite or cofinite subsets. It is easy to see that every quantificational wff  $\Psi(t, Q_i)$ , with  $Q_i$  finite or cofinite subsets of  $\mathbb{N}$ , defines another finite or co-infinite subset. But  $\text{even}$  is definable in USF (Example 109.3 above).  $\text{even}$  is also definable in monadic second-order logic. In fact, given any formula  $A(q_1, \dots, q_n)$  of USF, we can construct a formula  $A'(x, Y_1, \dots, Y_n)$  of monadic second-order logic in the language with relations  $\subseteq$  and  $<$ , such that for all  $h$ ,

$$h(A) = \{m \mid A'(m, h(q_1), \dots, h(q_n)) \text{ holds in } \mathbb{N}\}.$$

(See [Hodkinson, 1989].) Since this monadic logic is decidable, we get the following theorem.

**THEOREM 112.** *Propositional USF is decidable. In other words, the set of wffs  $\{A \mid h(A) = \mathbb{N} \text{ for all } h\}$  is recursive. See for example [Hodkinson, 1989].*

**THEOREM 113.** *Many nested applications of the fixed point operator are no stronger than a single one. In fact, any pure past wff of USF is semantically equivalent to a positive Boolean combination (i.e. using  $\wedge, \vee$  only) of wffs of the form  $(\varphi x)A$ , where  $A$  is built from atoms using only the Boolean connectives and  $\bullet$  (as in Example 109.1,4). See [Hodkinson, 1989].*

**THEOREM 114** (Full expressiveness of  $S$  and  $U$  [Kamp, 1968a]). *Let  $Q_1, \dots, Q_n \subseteq \mathbb{N}$  be  $n$  set variables and let  $\Psi(t, Q_1, \dots, Q_n)$  be a first-order monadic formula built up from  $Q_1, \dots, Q_n$  using  $<, =$  and the quantifiers over elements of  $\mathbb{N}$  and Boolean connectives. Then there exists a wff of USF,  $A_\Psi(q_1, \dots, q_n)$ , built up using  $S$  and  $U$  only (without the use of the fixed point operator  $\varphi$ ) such that for all  $h$  and all  $Q_i$  the following holds:*

$$\text{If } h(q_i) = Q_i \text{ then } h(A_\Psi) = \{t \mid \Psi(t, Q_i) \text{ holds in } \mathbb{N}\}.$$

**Proof.** H. Kamp proved this theorem directly by constructing  $A_\Psi$ . Another proof was given in [Gabbay *et al.*, 1994]. The significance of this theorem is that  $S$  and  $U$  alone are exactly as expressive (as a specification language) as first-order quantification  $\forall, \exists$  over temporal points. The use of  $\varphi$  takes USF beyond first-order quantification. ■

**THEOREM 115** (Well known). *Predicate USF without fixed point applications is not arithmetical ([Kamp, 1968a]).*

## 8.2 USF as a specification language

The logic USF can be used as a specification language as follows. Let  $A(a_1, \dots, a_m, b_1, \dots, b_k)$  be a wff of USF. Let  $h$  be an assignment to the atoms  $\{a_i, b_j\}$ . We say that  $h$  satisfies the specification  $A$  iff  $h(A) = \mathbb{N}$ .

In practice, the atoms  $b_j$  are controlled by the environment, and the atoms  $a_i$  are controlled by the program. Thus the truth values of  $b_j$  are determined as *events* and the truth values of  $a_i$  are determined by the program *execution*. As time moves forward and the program interacts with the environment, we get a function  $h$ , which may or may not satisfy the specification.

Let  $\mathbf{event}(q, n)$  mean that the value of  $q$  at time  $n$  is truth, as determined by the environment and let  $\mathbf{exec}(q, n)$  mean that  $q$  is executed at time  $n$  and therefore the truth value of  $q$  at time  $n$  is *true*. Thus out of  $\mathbf{event}$  and  $\mathbf{exec}$  we can get a full assignment  $h = \mathbf{event} + \mathbf{exec}$  by letting:

$$\begin{aligned} h(q) &= \{n \mid \mathbf{event}(q, n) \text{ holds}\} \text{ for } q \text{ controlled by the environment;} \\ h(q) &= \{n \mid \mathbf{exec}(q, n) \text{ holds}\} \text{ for } q \text{ controlled by the program.} \end{aligned}$$

Of course our aim is to execute in such a way that the  $h$  obtained satisfies the specification.

We now explain how to execute any wff of our temporal language. Recall that the truth values of the atoms  $a_i$  come from the program *via*  $\mathbf{exec}(a_i, m)$  and the truth values for the atoms  $b_j$  come from the environment via the function  $\mathbf{event}(b_j, m)$ . We define a predicate  $\mathbf{exec}^*(A, m)$  for any wff  $A$ , which actually defines the value of  $A$  at time  $m$ .

For atoms of the form  $a_i$ ,  $\mathbf{exec}^*(a_i, m)$  will be  $\mathbf{exec}(a_i, m)$  and for atoms of the form  $b_j$  (i.e. controlled by the environment)  $\mathbf{exec}^*(b_j, m)$  will be  $\mathbf{event}(b_j, m)$ . We  $\mathbf{exec}^*$  an atom controlled by the environment by ‘agreeing’ with the environment. For the case of  $A$  a pure past formula,  $\mathbf{exec}^*(A, m)$  is determined by past truth values. Thus  $\mathbf{exec}^*$  for pure past sentences is really a *hold* predicate, giving truth values determined already. For pure future sentences  $B$ ,  $\mathbf{exec}^*(B, m)$  will have an operational meaning. For example, we will have

$$\mathbf{exec}^*(G \text{ print}, m) = \mathbf{exec}^*(\text{print}, m) \wedge \mathbf{exec}^*(G \text{ print}, m + 1).$$

We can assume that the wffs to be executed are pure formulae (pure past, pure present or pure future) such that all negations are pushed next to atoms. This can be done because of the following semantic equivalences.

1.  $\neg U(a, b) = G\neg a \vee U(\neg b \wedge \neg a, \neg a)$ ;
2.  $\neg S(a, b) = H\neg a \vee S(\neg b \wedge \neg a, \neg a)$ ;
3.  $\neg Ga = F\neg a$ ;
4.  $\neg Ha = P\neg a$ ;
5.  $\neg Fa = G\neg a$ ;
6.  $\neg Pa = H\neg a$ ;
7.  $\neg[(\varphi x)A(x)] = (\varphi x)\neg A(\neg x)$ .

See [Gabbay *et al.*, 1994] for 1 to 6. For 7, let  $h$  be an assignment and suppose that  $h(\neg[(\varphi x)A(x)]) = S$ . Because  $\varphi$  has fixed point semantics, to prove 7 it is enough to show that if  $h'$  is the assignment that agrees with  $h$  on all atoms except  $x$ , and  $h'(x) = S$ , then  $h'(\neg(A\neg x)) = S$ . Clearly,  $h(\neg[(\varphi x)A(x)]) = \mathbb{N} \setminus S$ . We may assume that  $h(x) = \mathbb{N} \setminus S$ . Then  $h(A(x)) = \mathbb{N} \setminus S$ . So  $h(\neg A(x)) = S$  and  $h'(\neg(A\neg x)) = S$  as required; 7 is also easy to see using the recursive approach.

**DEFINITION 116.** We assume that  $\text{exec}^*(A, m)$  is defined for the system for any  $m$  and any  $A$  which is atomic or the negation of an atom. For atomic  $b$  which is controlled by the environment, we assume  $\text{event}(b, m)$  is defined and  $\text{exec}^*(b, m) = \text{event}(b, m)$ . For  $\text{exec}^*(a, m)$ , for  $a$  controlled by the program, execution may be done by another program. It may be a graphical or a mathematical program. It certainly makes a difference what  $m$  is relative to now. If we want to  $\text{exec}^*(a, m)$  for  $m$  in the past of now, then  $a$  has already been executed (or not) and so  $\text{exec}^*(a, m)$  is a hold predicate. It agrees with what has been done. Otherwise (if  $m \geq \text{now}$ ) we do execute\*.

1.  $\text{exec}^*(\top, m) = \top$ .
2.  $\text{exec}^*(\perp, m) = \perp$ .
3.  $\text{exec}^*(A \wedge B, m) = \text{exec}^*(A, m) \wedge \text{exec}^*(B, m)$ .
4.  $\text{exec}^*(A \vee B, m) = \text{exec}^*(A, m) \vee \text{exec}^*(B, m)$ .
5.  $\text{exec}^*(S(A, B), 0) = \perp$ .
6.  $\text{exec}^*(S(A, B), m+1) = \text{exec}^*(A, m) \vee [\text{exec}^*(B, m) \wedge \text{exec}^*(S(A, B), m)]$ .

7.  $\text{exec}^*(U(A, B), m) = \text{exec}^*(A, m+1) \vee [\text{exec}^*(B, m+1) \wedge \text{exec}^*(U(A, B), m+1)]$ .
8.  $\text{exec}^*((\varphi x)A(x), 0) = A_0$ , where  $A_0$  is obtained from  $A$  by substituting  $\top$  for any wff of the form  $HB$  and  $\perp$  for any wff of the form  $PB$  or  $S(B_1, B_2)$ .
9.  $\text{exec}^*((\varphi x)A(x), m+1) = \text{exec}^*(A(C), m+1)$ , where  $C$  is a new atom defined for  $n \leq m$  by  $\text{exec}^*(C, n) = \text{exec}^*((\varphi x)A(x), n)$ . In other words  $\text{exec}^*((\varphi x)A(x), m+1) = \text{exec}^*(A((\varphi x)A(x)), m+1)$  and since in the execution of  $A$  at time  $m+1$  we go down to executing  $A$  at time  $n \leq m$ , we will have to execute  $(\varphi x)A(x)$  at  $n \leq m$ , which we assume by induction that we already know.
10. In the predicate case we can let

$$\begin{aligned} \text{exec}^*(\forall y A(y)) &= \forall y \text{exec}^*(A(y)) \\ \text{exec}^*(\exists y A(y)) &= \exists y \text{exec}^*(A(y)). \end{aligned}$$

We are now in a position to discuss how the execution of a specification is going to be carried out in practice. Start with a specification  $S$ . For simplicity we assume that  $S$  is written in essentially propositional USF which means that  $S$  contains  $S$ ,  $U$  and  $\varphi$  operators applied to pure past formulae, and is built up from atomic units which are wffs of classical logic. If we regard any fixed point wff  $(\varphi x)D(x)$  as atomic, we can apply the separation theorem and rewrite  $S$  into an executable form  $\mathcal{E}$ , which is a conjunction of formulae such that

$$\Omega \equiv \bigwedge_k \left[ \bigwedge_i C_{i,k} \Rightarrow \bigvee_j B_{j,k} \right]$$

where  $C_{i,k}$  are pure past formulae (containing  $S$  only) and  $B_{j,k}$  are either atomic or pure future formulae (containing  $U$ ). However, since we regarded any  $(\varphi x)$  formula as an atom, the  $B_{j,k}$  can contain  $(\varphi x)D(x)$  formulae in them. Thus  $B_{j,k}$  can be for example  $U(a, (\varphi x)[\bullet \neg x])$ . We will assume that any such  $(\varphi x)D(x)$  contains only atoms controlled by the environment; this is a restriction on  $\mathcal{E}$ . Again, this is because we have no separation theorem as yet for full propositional USF, but only for the fragment US of formulae not involving  $\varphi$ . We conjecture that—possibly in a strengthened version of USF that allows more fixed point formulae—any formula can be separated. This again remains to be done.

However, even without such a result we can still make progress. Although  $(\varphi x)[\bullet \neg x]$  is a pure past formula within  $U$ , it is still an executable formula that only refers to environment atoms, and so we do not mind having it there. If program atoms were involved, we might have a formula equivalent

to  $X \bullet \text{print}$  (say), so that we would have to execute  $\bullet \text{print}$  tomorrow. This is not impossible: when tomorrow arrives we check whether we did in fact print yesterday, and return  $\top$  or  $\perp$  accordingly. But it is not a very intelligent way of executing the specification, since clearly we should have just printed in the first instance. This illustrates why we need to separate  $\mathcal{S}$ .

Recall the equation for executing  $U(A, B)$ :

$$\text{exec}^*(U(A, B)) \equiv X \text{exec}^*(A) \vee (X(\text{exec}^*(B) \wedge \text{exec}^*(U(A, B))).$$

If either  $A$  or  $B$  is of the form  $(\varphi x)D(x)$ , we know how to compute  $\text{exec}^*((\varphi x) D(x))$  by referring to past values. Thus  $(\varphi x)D(x)$  can be regarded as atomic because we know how to execute it, in the same way as we know how to execute `write`.

Imagine now that we are at time  $n$ . We want to make sure the specification  $\mathcal{E}$  remains true. To keep  $\mathcal{E}$  true we must keep true each conjunct of  $\mathcal{E}$ . To keep true a conjunct of the form  $C \Rightarrow B$  where  $C$  is past and  $B$  is future, we check whether  $C$  is true in the past. if it is true, then we have to make sure that  $B$  is true in the future. *Since the future has not happened yet, we can read  $B$  imperatively, and try to force the future to be true.* Thus the specification  $C \Rightarrow B$  is read by us as

$$\text{hold}(C) \Rightarrow \text{exec}^*(B).$$

Some future formulae cannot be executed immediately. We already saw that to execute  $U(A, B)$  now we either execute  $A$  tomorrow or execute  $B$  tomorrow together with  $U(A, B)$ . Thus we have to pass a list of formulae to execute from today to tomorrow. Therefore at time  $n + 1$ , we have a list of formulae to execute which we inherit from time  $n$ , in addition to the list of formulae to execute at time  $n + 1$ . We can thus summarize the situation at time  $n + 1$  as follows:

1. Let  $G_1, \dots, G_m$  be a list of wffs we have to execute at time  $n + 1$ . Each  $G_i$  is a disjunction of formulae of the form atomic or negation of atomic or  $FA$  or  $GA$  or  $U(A, B)$ .
2. In addition to the above, we are required to satisfy the specification  $\mathcal{E}$ , namely

$$\bigwedge_k \left[ \bigwedge_i C_{i,k} \Rightarrow \bigvee_j B_{j,k} \right]$$

for each  $k$  such that  $\bigwedge_i C_{i,k}$  holds (in the past). We must execute the future (and present) formula  $B_k = \bigvee_j B_{j,k}$  which is again a disjunction of the same form as in 1 above.

We know how to execute a formula; for example,



$$\text{exec}^*(FA) = X\text{exec}^*A \vee X \text{exec}^*(FA).$$

$FA$  means ‘ $A$  will be true’. To execute  $FA$  we can either make  $A$  true tomorrow or make  $FA$  true tomorrow. What we should be careful not to do is not to keep on executing  $FA$  day after day because this way  $A$  will never become true. Clearly then we should try to execute  $A$  tomorrow and if we cannot, only then do we execute  $FA$  by doing  $X \text{exec}^*(FA)$ . We can thus read the disjunction  $\text{exec}^*(A \vee B)$  as *first* try to  $\text{exec}^*A$  and *then only if we fail*  $\text{exec}^*B$ . This priority (left to right) is not a logical part of ‘ $\vee$ ’ but a procedural addition required for the correctness of the model. We can thus assume that the formulae given to execute at time  $n$  are written as disjunctions with the left disjuncts having priority in execution. Atomic sentences or their negations always have priority in execution (though this is not always the best practical policy).

Let  $D = \bigvee_j D_j$  be any wff which has to be executed at time  $n + 1$ , either because it is inherited from time  $n$  or because it has to be executed owing to the requirements of the specification at time  $n + 1$ . To execute  $D$ , either we execute an atom and discharge our duty to execute, or we pass possibly several disjunctions to time  $n + 2$  to execute then (at  $n + 2$ ), and the passing of the disjunctions will discharge our obligation to execute  $D$  at time  $n + 1$ . Formally we have

$$\text{exec}^*(D) = \bigvee_j \text{exec}^*(D_j).$$

Recall that we try to execute left to right. The atoms and their negations are supposed to be on the left. If we can execute any of them we are finished with  $D$ . If an atom is an environment atom, we check whether the environment gives it the right value. If the atom is under the program’s control, we can execute it. However, the negation of the atom may appear in another formula  $D'$  to be executed and there may be a clash. See Examples 117 and 118 below. At any rate, should we choose to execute an atom or negation of an atom and succeed in doing so, then we are finished. Otherwise we can execute another disjunct of  $D$  of the form  $D_j = U(A_j, B_j)$  or of the form  $GA_j$  or  $FA_j$ . We can pass the commitment to execute to the time  $n + 2$ . Thus we get

$$\text{exec}^*(D) = \bigvee \text{exec}^*(\text{atoms of } D) \vee \text{exec}^*(\text{future formulae of } D).$$

Thus if we cannot execute the atoms at time  $n + 1$ , we pass to time  $n + 2$  a conjunction of disjunctions to be executed, ensuring that atoms and subformulae should be executed before formulae. We can write the disjunctions to reflect these priorities. Notice further that although, on first impression, the formulae to be executed seem to multiply, they actually do not.

At time  $n = 0$  all there is to execute are heads of conditions in the specification. If we cannot execute a formula at time 0 then we pass execution to time 1. This means that at time 1 we inherit the execution of

$A \vee (B \wedge U(A, B))$ , where  $U(A, B)$  is a disjunct in a head of the specification. This same  $U(A, B)$  may be passed on to time 2, or some subformula of  $A$  or  $B$  may be passed. The number of such subformulae is limited and we will end up with a limited stock of formulae to be passed on. In practice this can be optimized. We have thus explained how to execute whatever is to be executed at time  $n$ . When we perform the execution sequence at times  $n, n+1, n+2, \dots$ , we see that there are now two possibilities:

- We cannot go on because we cannot execute all the demands at the same time. In this case we stop. The specification cannot be satisfied either because it is a contradiction or because of a wrong execution choice (e.g. we should not have printed at time 1, as the specification does not allow anything to be done after printing).
- Another possibility is that we see after a while that the same formulae are passed for execution from time  $n$  to time  $n+1$  to  $n+2$  etc. This is a loop. Since we have given priority in execution to atoms and to the  $A$  in  $U(A, B)$ , such a loop means that it is not possible to make a change in execution, and therefore either the specification cannot be satisfied because of a contradiction or wrong choice of execution, or the execution is already satisfied by this loop.

EXAMPLE 117. All atoms are controlled by the program. Let the specification be

$$Ga \wedge F\neg a.$$

Now the rules to execute the subformulae of this specification are

$$\begin{aligned} \text{exec}^*(Ga) &\equiv \text{exec}^*(a) \wedge \text{exec}^*(Ga) \\ \text{exec}^*(F\neg a) &\equiv \text{exec}^*(\neg a) \vee \text{exec}^*(F\neg a). \end{aligned}$$

To execute  $Ga$  we must execute  $a$ . Thus we are forced to discharge our execution duty of  $F\neg a$  by passing  $F\neg a$  to time  $n+1$ . Thus time  $n+1$  will inherit from time  $n$  the need to execute  $Ga \wedge F\neg a$ . This is a loop. The specification is unsatisfiable.

EXAMPLE 118. The specification is

$$b \vee Ga$$

$$Pb \Rightarrow F\neg a \wedge Ga.$$

According to our priorities we execute  $b$  first at time 0. Thus we will have to execute  $F\neg a \wedge Ga$  at time 1, which is impossible. Here we made the wrong execution choice. If we keep on executing  $\neg b \wedge Ga$  we will behave as specified.

In practice, since we may have several choices in execution we may want to simulate the future a little to see if we are making the correct choice.

Having defined  $\text{exec}^*$ , we need to add the concept of updating. Indeed, the viability of our notion of the declarative past and imperative future depends on adding information to our database. In this section we shall assume that every event that occurs in the environment, and every action  $\text{exec}$ -ed by our system, are recorded in the database. This is of course unnecessary, and in a future paper we shall present a more realistic method of updating.

### 8.3 The logic USF2

The fixed point operator that we have introduced in propositional USF has to do with the solution of the equation

$$x \leftrightarrow B(x, q_1, \dots, q_m)$$

where  $B$  is a pure past formula. Such a solution always exists and is unique. The above equation defines a connective  $A(q_1, \dots, q_m)$  such that

$$A(q_1, \dots, q_m) \leftrightarrow B(A(q_1, \dots, q_m), q_1, \dots, q_m).$$

Thus, for example,  $S(p, q)$  is the solution of the equation

$$x \leftrightarrow \bullet p \vee \bullet (q \wedge x)$$

as we have  $S(p, q) \leftrightarrow \bullet p \vee \bullet (q \wedge S(p, q))$ . Notice that the connective to be defined ( $x = S(p, q)$ ) appears as a unit in both sides of the equation.

To prove existence of a solution we proceed by induction. Suppose we know what  $x$  is at time  $\{0, \dots, n\}$ . To find what  $x$  is supposed to be at time  $n + 1$ , we use the equation  $x \leftrightarrow B(x, q_i)$ . Since  $B$  is pure past, to compute  $B$  at time  $n + 1$  we need to know  $\{x, q_i\}$  at times  $\leq n$ , which we do know. This is the reason why we get a unique solution.

Let us now look at the following equation for a connective  $Z(p, q)$ . We want  $Z$  to satisfy the equation

$$Z(p, q) \leftrightarrow \bullet p \vee \bullet (q \wedge Z(\bullet p, q)).$$

Here we did not take  $Z(p, q)$  as a unit in the equation, but substituted a value  $\bullet p$  in the right-hand side, namely  $Z(\bullet p, q)$ .  $\bullet p$  is a pure past formula. We can still get a unique solution because  $Z(p, q)$  at time  $n + 1$  still depends on the values of  $Z(p, q)$  at earlier times, and certainly we can compute the values of  $Z(\bullet p, q)$  at earlier times.

The general form of the new fixed point equation is as follows:

**DEFINITION 119** (Second-order fixed points). Let  $Z(q_1, \dots, q_m)$  be a candidate for a new connective to be defined. Let  $B(x, q_1, \dots, q_m)$  be a pure

past formula and let  $D_i(q_1, \dots, q_m)$  for  $i = 1 \dots m$  be arbitrary formulae. Then we can define  $Z$  as the solution of the following equation:

$$Z(q_1, \dots, q_m) \leftrightarrow B(Z[D_1(q_1, \dots, q_m), \dots, D_m(q_1, \dots, q_m)], q_1, \dots, q_m).$$

We call this definition of  $Z$  *second order*, because we can regard the equation as

$$Z \equiv \text{Application}(Z, D_i, q_j).$$

We define USF2 to be the logic obtained from USF by allowing nested applications of second-order fixed point equations. USF2 is more expressive than USF (Example 120).

Predicate USF2 is defined in a similar way to predicate USF.

EXAMPLE 120. Let us see what we get for the connective  $Z_1(p, q)$  defined by the equation

$$Z_1(p, q) \leftrightarrow \bullet p \vee \bullet (q \wedge Z_1(\bullet p, q)).$$

The connective  $Z_1(p, q)$  says what is shown in Fig.12:

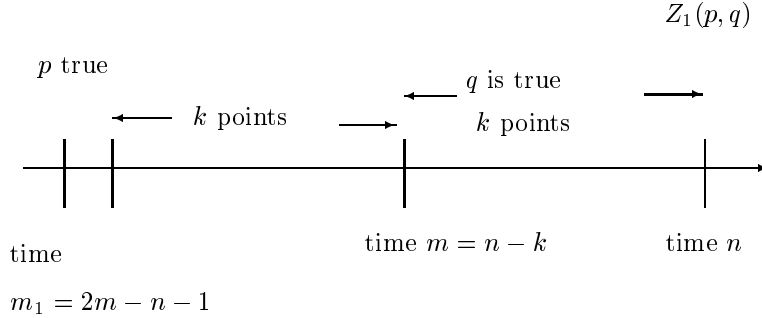


Figure 12.

$Z_1(p, q)$  is true at  $n$  iff for some  $m \leq n$ ,  $q$  is true at all points  $j$  with  $m \leq j < n$ , and  $p$  is true at the point  $m_1 = m - (n - m + 1) = 2m - n - 1$ . If we let  $k = n - m$ , then we are saying that  $q$  is true  $k$  times into the past and before that  $p$  is true at a point which is  $k + 1$  times further into the past. This is not expressible with any pure past formula of USF; see [Hodkinson, 1989].

Let us see whether this connective satisfies the fixed point equation

$$Z_1(p, q) \leftrightarrow \bullet p \vee \bullet (q \wedge Z_1(\bullet p, q)).$$

If  $\bullet p$  is true then  $k = 0$  and the definition of  $Z_1(p, q)$  is correct. If  $\bullet (q \wedge Z_1(\bullet p, q))$  is true, then we have for some  $k$  the situation in Fig. 13:

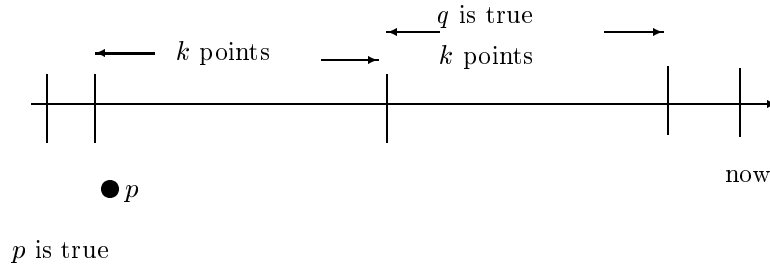


Figure 13.

The definition of  $Z_1(p, q)$  is satisfied for  $k + 1$ .

EXAMPLE 121 (Coding of dates). We can encode dates in the logic as follows:

1. The proposition  $\neg \bullet \top$  is true exactly at time 0, since it says that there is no yesterday. Thus if we let

$$\begin{aligned} \mathbf{n} &= \perp \text{ if } n \leq 0 \\ 0 &= \neg \bullet \top \\ \mathbf{n} &= \bullet (\mathbf{n} - \mathbf{1}). \end{aligned}$$

then we have that  $\mathbf{n}$  is true exactly at time  $n$ . This is a way of naming time  $n$ . In predicate temporal logic we can use elements to name time. Let  $\text{date}(x)$  be a predicate such that the following hold at all times  $n$ :

$$\begin{aligned} &\exists x \text{ date}(x) \\ &\forall x (\text{date}(x) \Rightarrow G\neg \text{date}(x) \wedge H\neg \text{date}(x)) \\ &\forall x (\text{date}(x) \vee P \text{date}(x) \vee F \text{date}(x)). \end{aligned}$$

These axioms simply say that each time  $n$  is identified by some element  $x$  in the domain that uniquely makes  $\text{date}(x)$  true, and every domain element corresponds to a time.

2. We can use this device to count in the model. Suppose we want to define a connective that counts how many times  $A$  was true in the past. We can represent the number  $m$  by the date formula  $\mathbf{m}$ , and define  $\text{count}(A, \mathbf{m})$  to be true at time  $n$  iff the number of times before  $n$  in which  $A$  was true is exactly  $\mathbf{m}$ . Thus in Fig. 14,  $\text{count}(A, \bullet \top \wedge \neg \bullet \bullet \top)$  is false at time 3, true at time 2, true at time 1 and false at time 0.

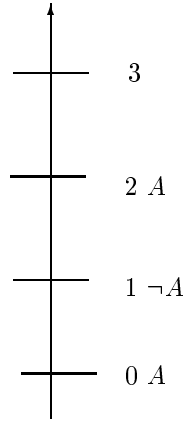


Figure 14.

The connective `count` can be defined by recursion as follows:

$$\begin{aligned} \text{count}(p, \mathbf{n}) \quad \leftrightarrow \quad & \bullet (\neg p \wedge \text{count}(p, \mathbf{n})) \\ & \vee \bullet (p \wedge \text{count}(p, X\mathbf{n})) \\ & \vee (\neg \bullet \top \wedge \mathbf{n}). \end{aligned}$$

Note that  $X\mathbf{n}$  is equivalent to  $\mathbf{n} - \mathbf{1}$ . We have cheated in this example. For the formula  $B(x, q_1, q_2)$  in the definition of second-order fixed points is here

$$\bullet (\neg q_1 \wedge x) \vee \bullet (q_1 \wedge x) \vee (\neg \bullet \top \wedge q_2).$$

This is not pure past, as  $q_2$  occurs in the present tense. To deal with this we could define the notion of a formula  $B(x, q_1, \dots, q_m)$  being pure past in  $x$ . See [Hodkinson, 1989]. We could then amend the definition to allow any  $B$  that is pure past in  $x$ . This would cover the  $B$  here, as all  $x$ s in  $B$  occur under a  $\bullet$ . So the value of the connective at  $n$  still depends only on its values at  $m \leq n$ , which is all we need for there to be a fixed point solution. We do not do this formally here, as we can express `count` in standard USF2; see the next example.

EXAMPLE 122. We can now define the connective `more`( $A, B$ ) reading ‘ $A$  was true more times than  $B$ ’.

$$\begin{aligned} \text{more}(A, B) \quad \leftrightarrow \quad & \bullet (A \wedge \text{more}(A, B)) \\ & \vee \bullet (\neg A \wedge \neg B \wedge \text{more}(A, B)) \\ & \vee \bullet (\neg A \wedge \text{more}((A \wedge PA), B)). \end{aligned}$$

(If  $k > 0$ , then at any  $n$ ,  $A \wedge PA$  has been true  $k$  times iff  $A$  has been true  $k + 1$  times.)

Note that for any  $k > 0$ , the formula  $E_k = \neg \bullet^k \top$  is true exactly  $k$  times, at  $0, 1, \dots, k - 1$ . If we define

$$\text{count}^*(p, k) = \text{more}(E_{k+1}, p) \wedge \neg \text{more}(E_k, p),$$

then at any  $n$ ,  $p$  has been true  $k$  times iff  $\text{count}^*(p, k)$  holds. So we can do the previous example in standard USF2.

**THEOREM 123** (For propositional USF2). *Nested applications of the second-order fixed point operator are equivalent to one application. Any wff  $A$  of USF2 is equivalent to a wff  $B$  of USF2 built up using no nested applications of the second-order fixed point operator.*

#### 8.4 Payroll example in detail

This section will consider in detail the execution procedures for the payroll example in Section 8.

First let us describe, in the temporal logic USF2, the specification required by Mrs Smith. We translate from the English in a natural way. This is important because we want our logical specification to be readable and have the same structure as in English.

Recall that the intended interpretation of the predicates to be used is

$A(x)$	$x$ is asked to babysit
$B(x)$	$x$ does a babysitting job
$M(x)$	$x$ works after midnight
‘ $P(x, y)$	$x$ is paid $y$ pounds.

‘Babysitters are not allowed to take jobs three nights in a row, or two nights in a row if the first night involves overtime’ is translated as

(a)  $\forall x \neg [B(x) \wedge \bullet B(x) \wedge \bullet \bullet B(x)]$

(b)  $\forall x \neg [B(x) \wedge \bullet (B(x) \wedge M(x))]$

(c)  $\forall x [M(x) \Rightarrow B(x)]$ .

Note that these wffs are not essentially propositional.

‘Priority in calling is given to those who were not called before as many times as others’ is translated as

(d)  $\neg \exists x \exists y [\text{more}(A(x), A(y)) \wedge A(x) \wedge \neg A(y) \wedge \neg \bullet M(y) \wedge \neg \bullet (B(y) \wedge \bullet B(y))]$ .

‘Payment should be made the next day after the job was done, with £15 for a job involving overtime, and £10 for a job not involving overtime’ is translated as

- (e)  $\forall x[M(x) \Rightarrow XP(x, 15)]$
- (f)  $\forall x[B(x) \wedge \neg M(x) \Rightarrow XP(x, 10)]$
- (g)  $\forall x[\neg B(x) \Rightarrow X\neg\exists yP(x, y)]$ .

Besides the above we also have

- (h)  $\forall x[B(x) \Rightarrow A(x)]$ .

Babysitters work only when they are called.

We have to rewrite the above into an executable form, namely

Past  $\Rightarrow$  Present  $\vee$  Future.

We transform the specification to the following:

- (a')  $\forall x[\bullet B(x) \wedge \bullet \bullet B(x) \Rightarrow \neg B(x)]$
- (b')  $\forall x[\bullet (B(x) \wedge M(x)) \Rightarrow \neg B(x)]$
- (c')  $\forall x[\neg M(x) \vee B(x)]$ .
- (d')  $\forall x\forall y[\text{more}(A(x), A(y)) \wedge \neg \bullet M(y) \wedge \neg \bullet (B(y) \wedge \bullet B(y)) \Rightarrow \neg A(x) \vee \neg A(y)]$
- (e')  $\forall x[\neg M(x) \vee XP(x, 15)]$
- (f')  $\forall x[\neg B(x) \vee M(x) \vee XP(x, 10)]$
- (g')  $\forall x[B(x) \vee X\forall y\neg P(x, y)]$
- (h')  $\forall x[\neg B(x) \vee A(x)]$ .

Note that (e'), (f') and (h') can be rewritten in the following form using the  $\bullet$  operator.

- (e'')  $\forall x[\bullet M(x) \Rightarrow P(x, 15)]$
- (f'')  $\forall x[\bullet (B(x) \wedge \neg M(x)) \Rightarrow P(x, 10)]$
- (g'')  $\forall x[\neg \bullet B(x) \Rightarrow \forall y\neg P(x, y)]$ .

Our executable sentences become

- (a\*)  $\text{hold}(\bullet B(x) \wedge \bullet \bullet B(x)) \Rightarrow \text{exec}(\neg B(x))$
- (b\*)  $\text{hold}(\bullet (B(x) \wedge M(x))) \Rightarrow \text{exec}(\neg B(x))$
- (c\*)  $\text{exec}(\neg M(x) \vee B(x))$
- (d\*)  $\text{hold}(\text{more}(A(x), A(y)) \wedge \neg \bullet M(y) \wedge \neg \bullet (B(y) \wedge \bullet B(y))) \Rightarrow \text{exec}(\neg A(x) \vee \neg A(y))$



(**e\***)  $\text{exec}(\neg M(x) \vee XP(x, 15))$

(**f\***)  $\text{exec}(\neg B(x) \vee M(x) \vee XP(x, 10))$

(**g\***)  $\text{exec}(B(x) \vee X\forall y\neg P(x, y))$

(**h\***)  $\text{exec}(\neg B(x) \vee A(x))$ .

If we use (**e''**), (**f''**), (**g''**) the executable form will be

(**e\*\***)  $\text{hold}(\bullet M(x)) \Rightarrow \text{exec}(P(x, 15))$

(**f\*\***)  $\text{hold}(\bullet (B(x) \wedge \neg M(x))) \Rightarrow \text{exec}(P(x, 10))$

(**g\*\***)  $\text{hold}(\neg \bullet B(x)) \Rightarrow \text{exec}(\forall y\neg P(x, y))$ .

In practice there is no difference whether we use (**e\*\***) or (**e\***). We execute  $XP$  by sending  $P$  to tomorrow for execution. If the specification is (**e\*\***), we send nothing to tomorrow but we will find out tomorrow that we have to execute  $P$ .

D. Gabbay

*Department of Computer Science, King's College, London.*

M. Finger

*Departamento de Ciência da Computação, University of Sao Paulo, Brazil.*

M. Reynolds

*School of Information Technology, Murdoch University, Australia.*

## BIBLIOGRAPHY

- [Amir, 1985] A. Amir. Separation in nonlinear time models. *Information and Control*, 66:177 – 203, 1985.
- [Bannieqbal and Barringer, 1986] B. Bannieqbal and H. Barringer. A study of an extended temporal language and a temporal fixed point calculus. Technical Report UMCS-86-10-2, Department of Computer Science, University of Manchester, 1986.
- [Belnap and Green, 1994] N. Belnap and M. Green. Indeterminism and the red thin line. In *Philosophical Perspectives, 8, Logic and Language*, pages 365–388. 1994.
- [Brzozowski and Leiss, 1980] J. Brzozowski and E. Leiss. Finite automata, and sequential networks. *TCS*, 10, 1980.
- [Büchi, 1962] J.R. Büchi. On a decision method in restricted second order arithmetic. In *Logic, Methodology, and Philosophy of Science: Proc. 1960 Intern. Congress*, pages 1–11. Stanford University Press, 1962.
- [Bull and Segerberg, in this handbook] R. Bull and K. Segerberg. Basic modal logic. In D.M. Gabbay and F. Guentner, editors, *Handbook of Philosophical Logic, second edition*, volume 2, page ? Kluwer, in this handbook.
- [Burgess, 2001] J. Burgess. Basic tense logic. In D.M. Gabbay and F. Guentner, editors, *Handbook of Philosophical Logic, second edition*, volume 7, pp. 1–42, Kluwer, 2001.
- [Burgess and Gurevich, 1985] J. P. Burgess and Y. Gurevich. The decision problem for linear temporal logic. *Notre Dame J. Formal Logic*, 26(2):115–128, 1985.
- [Burgess, 1982] J. P. Burgess. Axioms for tense logic I: 'since' and 'until'. *Notre Dame J. Formal Logic*, 23(2):367–374, 1982.

- [Choueka, 1974] Y. Choueka. Theories of automata on  $\omega$ -tapes: A simplified approach. *JCSS*, 8:117–141, 1974.
- [Darlington and While, 1987] J. Darlington and L. While. Controlling the behaviour of functional programs. In *Third Conference on Functional Programming Languages and Computer Architecture*, 1987.
- [Doets, 1989] K. Doets. Monadic  $\Pi_1^1$ -theories of  $\Pi_1^1$ -properties. *Notre Dame J. Formal Logic*, 30:224–240, 1989.
- [Ehrenfeucht, 1961] A. Ehrenfeucht. An application of games to the completeness problem for formalized theories. *Fund. Math.*, 49:128–141, 1961.
- [Emerson and Lei, 1985] E. Emerson and C. Lei. Modalities for model checking: branching time strikes back. In *Proc. 12th ACM Symp. Princ. Prog. Lang.*, pages 84–96, 1985.
- [Emerson, 1990] E.A. Emerson. Temporal and modal logic. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B. Elsevier, Amsterdam, 1990.
- [Fine and Schurz, 1991] K. Fine and G. Schurz. Transfer theorems for stratified multi-modal logics. 1991.
- [Finger and Gabbay, 1992] M. Finger and D. M. Gabbay. Adding a Temporal Dimension to a Logic System. *Journal of Logic Language and Information*, 1:203–233, 1992.
- [Finger and Gabbay, 1996] M. Finger and D. Gabbay. Combining Temporal Logic Systems. *Notre Dame Journal of Formal Logic*, 37(2):204–232, 1996. Special Issue on Combining Logics.
- [Finger, 1992] M. Finger. Handling Database Updates in Two-dimensional Temporal Logic. *J. of Applied Non-Classical Logic*, 2(2):201–224, 1992.
- [Finger, 1994] M. Finger. *Changing the Past: Database Applications of Two-dimensional Temporal Logics*. PhD thesis, Imperial College, Department of Computing, February 1994.
- [Fisher, 1997] M. Fisher. A normal form for temporal logic and its application in theorem-proving and execution. *Journal of Logic and Computation*, 7(4):?, 1997.
- [Gabbay and Hodkinson, 1990] D. M. Gabbay and I. M. Hodkinson. An axiomatisation of the temporal logic with until and since over the real numbers. *Journal of Logic and Computation*, 1(2):229 – 260, 1990.
- [Gabbay and Olivetti, 2000] D. M. Gabbay and N. Olivetti. *Goal Directed Algorithmic Proof*. APL Series, Kluwer, Dordrecht, 2000.
- [Gabbay and Shehtman, 1998] D. Gabbay and V. Shehtman. Products of modal logics, part 1. *Logic Journal of the IGPL*, 6(1):73–146, 1998.
- [Gabbay *et al.*, 1980] D. M. Gabbay, A. Pnueli, S. Shelah, and J. Stavi. On the temporal analysis of fairness. In *7th ACM Symposium on Principles of Programming Languages, Las Vegas*, pages 163–173, 1980.
- [Gabbay *et al.*, 1994] D. Gabbay, I. Hodkinson, and M. Reynolds. *Temporal Logic: Mathematical Foundations and Computational Aspects, Volume 1*. Oxford University Press, 1994.
- [Gabbay *et al.*, 2000] D. Gabbay, M. Reynolds, and M. Finger. *Temporal Logic: Mathematical Foundations and Computational Aspects, Vol. 2*. Oxford University Press, 2000.
- [Gabbay, 1981] D. M. Gabbay. An irreflexivity lemma with applications to axiomatizations of conditions on tense frames. In U. Monnich, editor, *Aspects of Philosophical Logic*, pages 67–89. Reidel, Dordrecht, 1981.
- [Gabbay, 1985] D. Gabbay. *N-Prolog*, part 2. *Journal of Logic Programming*, 5:251–283, 1985.
- [Gabbay, 1989] D. M. Gabbay. Declarative past and imperative future: Executable temporal logic for interactive systems. In B. Banieqbal, H. Barringer, and A. Pnueli, editors, *Proceedings of Colloquium on Temporal Logic in Specification, Altrincham, 1987*, pages 67–89. Springer-Verlag, 1989. Springer Lecture Notes in Computer Science 398.
- [Gabbay, 1996] D. M. Gabbay. *Labelled Deductive Systems*. Oxford University Press, 1996.
- [Gabbay, 1998] D. M. Gabbay. *Fibring Logics*. Oxford University Press, 1998.

- [Gabbay *et al.*, 2002] D. M. Gabbay, A. Kurucz, F. Wolter and M. Zakharyashev. *Many Dimensional Logics*, Elsevier, 2002. To appear.
- [Gurevich, 1964] Y. Gurevich. Elementary properties of ordered abelian groups. *Algebra and Logic*, 3:5–39, 1964. (Russian; an English version is in *Trans. Amer. Math. Soc.* 46 (1965), 165–192).
- [Gurevich, 1985] Y. Gurevich. Monadic second-order theories. In J. Barwise and S. Feferman, editors, *Model-Theoretic Logics*, pages 479–507. Springer-Verlag, New York, 1985.
- [Hodges, 1985] W. Hodges. Logical features of horn clauses. In D.M. Gabbay, C.J. Hogger, and J.A. Robinson, editors, *The Handbook of Logic in Artificial Intelligence and Logic Programming, vol. 1*, pages 449–504. Oxford University Press, 1985.
- [Hodkinson, 1989] I. Hodkinson. Decidability and elimination of fixed point operators in the temporal logic USF. Technical report, Imperial College, 1989.
- [Hodkinson, 200] I. Hodkinson. Automata and temporal logic, forthcoming. chapter 2, in [Gabbay *et al.*, 2000].
- [Hopcroft and Ullman, 1979] J. Hopcroft and J. Ullman. *Introduction to Automata Theory, Languages, and Computation*. Addison-Wesley, 1979.
- [Kamp, 1968a] H. Kamp. Seminar notes on tense logics. *J. Symbolic Logic*, 1968.
- [Kamp, 1968b] H. Kamp. *Tense logic and the theory of linear order*. PhD thesis, University of California, Los Angeles, 1968.
- [Kesten *et al.*, 1994] Y. Kesten, Z. Manna, and A. Pnueli. Temporal verification of simulation and refinement. In *A decade of concurrency: reflections and perspectives: REX school/symposium, Noordwijkerhout, the Netherlands, June 1–4, 1993*, pages 273–346. Springer-Verlag, 1994.
- [Kleene, 1956] S. Kleene. Representation of events in nerve nets and finite automata. In C. Shannon and J. McCarthy, editors, *Automata Studies*, pages 3–41. Princeton Univ. Press, 1956.
- [Konolige, 1986] K. Konolige. *A Deductive Model of Belief*. Research notes in Artificial Intelligence. Morgan Kaufmann, 1986.
- [Kracht and Wolter, 1991] M. Kracht and F. Wolter. Properties of independently axiomatizable bimodal logics. *Journal of Symbolic Logic*, 56(4):1469–1485, 1991.
- [Kuhn, 1989] S. Kuhn. The domino relation: flattening a two-dimensional logic. *J. of Philosophical Logic*, 18:173–195, 1989.
- [Läuchli and Leonard, 1966] H. Läuchli and J. Leonard. On the elementary theory of linear order. *Fundamenta Mathematicae*, 59:109–116, 1966.
- [Lichtenstein *et al.*, 1985] O. Lichtenstein, A. Pnueli, and L. Zuck. The glory of the past. In R. Parikh, editor, *Logics of Programs (Proc. Conf. Brooklyn USA 1985)*, volume 193 of *Lecture Notes in Computer Science*, pages 196–218. Springer-Verlag, Berlin, 1985.
- [Manna and Pnueli, 1988] Z. Manna and A. Pnueli. The anchored version of the temporal framework. In *REX Workshop, Noordwijkerh.*, 1988. LNCS 354.
- [Marx, 1999] M. Marx. Complexity of products of modal logics, *Journal of Logic and Computation*, 9:221–238, 1999.
- [Marx and Reynolds, 1999] M. Marx and M. Reynolds. Undecidability of compass logic. *Journal of Logic and Computation*, 9(6):897–914, 1999.
- [Venema and Marx, 1997] M. Marx and Y. Venema. *Multi Dimensional Modal Logic*. Applied Logic Series No.4 Kluwer Academic Publishers, 1997.
- [McNaughton, 1966] R. McNaughton. Testing and generating infinite sequences by finite automata. *Information and Control*, 9:521–530, 1966.
- [Muller, 1963] D. Muller. Infinite sequences and finite machines. In *Proceedings 4th Ann. IEEE Symp. on Switching Circuit Theory and Logical Design*, pages 3–16, 1963.
- [Németi, 1995] I. Németi. Decidable versions of first order logic and cylindric-relativized set algebras. In L. Csirmaz, D. Gabbay, and M. de Rijke, editors, *Logic Colloquium '92*, pages 171–241. CSLI Publications, 1995.
- [Ono and Nakamura, 1980] H. Ono and A. Nakamura. On the size of refutation Kripke models for some linear modal and tense logics. *Studia Logica*, 39:325–333, 1980.
- [Perrin, 1990] D. Perrin. Finite automata. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B. Elsevier, Amsterdam, 1990.

- [Pnueli, 1977] A. Pnueli. The temporal logic of programs. In *Proceedings of the Eighteenth Symposium on Foundations of Computer Science*, pages 46–57, 1977. Providence, RI.
- [Prior, 1957] A. Prior. *Time and Modality*. Oxford University Press, 1957.
- [Rabin and Scott, 1959] M. Rabin and D. Scott. Finite automata and their decision problem. *IBM J. of Res.*, 3:115–124, 1959.
- [Rabin, 1969] M. O. Rabin. Decidability of second order theories and automata on infinite trees. *American Mathematical Society Transactions*, 141:1–35, 1969.
- [Rabin, 1972] M. Rabin. *Automata on Infinite Objects and Church's Problem*. Amer. Math. Soc., 1972.
- [Rabinovich, 1998] A. Rabinovich. On the decidability of continuous time specification formalisms. *Journal of Logic and Computation*, 8:669–678, 1998.
- [Reynolds and Zakharyashev, 2001] M. Reynolds and M. Zakharyashev. On the products of linear modal logics. *Journal of Logic and Computation*, 6, 909–932, 2001.
- [Reynolds, 1992] M. Reynolds. An axiomatization for Until and Since over the reals without the IRR rule. *Studia Logica*, 51:165–193, May 1992.
- [Reynolds, 1994] M. Reynolds. Axiomatizing  $U$  and  $S$  over integer time. In D. Gabbay and H.-J. Ohlbach, editors, *Temporal Logic, First International Conference, ICTL '94, Bonn, Germany, July 11-14, 1994, Proceedings*, volume 827 of *Lecture Notes in A.I.*, pages 117–132. Springer-Verlag, 1994.
- [Reynolds, 1998] M. Reynolds. A decidable logic of parallelism. *Notre Dame Journal of Formal Logic*, 38, 419–436, 1997.
- [Reynolds, 1999] M. Reynolds. The complexity of the temporal logic with until over general linear time, submitted 1999. Draft version of manuscript available at <http://www.it.murdoch.edu.au/~mark/research/online/cult.html>
- [Robertson, 1974] E.L. Robertson. Structure of complexity in weak monadic second order theories of the natural numbers. In *Proc. 6th Symp. on Theory of Computing*, pages 161–171, 1974.
- [Savitch, 1970] W. J. Savitch. Relationships between non-deterministic and deterministic tape complexities. *J. Comput. Syst. Sci.*, 4:177–192, 1970.
- [Sherman *et al.*, 1984] R. Sherman, A. Pnueli, and D. Harel. Is the interesting part of process logic uninteresting: a translation from PL to PDL. *SIAM J. on Computing*, 13:825–839, 1984.
- [Sistla and Clarke, 1985] A. Sistla and E. Clarke. Complexity of propositional linear temporal logics. *J. ACM*, 32:733–749, 1985.
- [Sistla *et al.*, 1987] A. Sistla, M. Vardi, and P. Wolper. The complementation problem for Buchi automata with applications to temporal logic. *Theoretical Computer Science*, 49:217–237, 1987.
- [Spaan, 1993] E. Spaan. *Complexity of Modal Logics*. PhD thesis, Free University of Amsterdam, Faculteit Wiskunde en Informatica, Universiteit van Amsterdam, 1993.
- [Thomas, 1990] W. Thomas. Automata on infinite objects. In J. van Leeuwen, editor, *Handbook of Theoretical Computer Science*, volume B. Elsevier, Amsterdam, 1990.
- [Thomason, 1984] R. H. Thomason. Combinations of Tense and Modality. In D. Gabbay and F. Guenther, editors, *Handbook of Philosophical Logic*, volume II, pages 135–165. D. Reidel Publishing Company, 1984. Reproduced in this volume.
- [van Benthem, 1991] J. F. A. K. van Benthem. *The logic of time*. 2nd edition. Kluwer Academic Publishers, Dordrecht, 1991.
- [van Benthem, 1996] J. van Benthem. *Exploring Logical Dynamics*. Cambridge University Press, 1996.
- [Vardi and Wolper, 1994] M. Vardi and P. Wolper. Reasoning about infinite computations. *Information and Computation*, 115:1–37, 1994.
- [Venema, 1990] Y. Venema. Expressiveness and Completeness of an Interval Tense Logic. *Notre Dame Journal of Formal Logic*, 31(4), Fall 1990.
- [Venema, 1991] Y. Venema. Completeness via completeness. In M. de Rijke, editor, *Colloquium on Modal Logic, 1991*. ITLI-Network Publication, Instit. for Lang., Logic and Information, University of Amsterdam, 1991.
- [Venema, 1993] Y. Venema. Derivation rules as anti-axioms in modal logic. *Journal of Symbolic Logic*, 58:1003–1034, 1993.

- [Wolper, 1983] P. Wolper. Temporal logic can be more expressive. *Information and computation*, 56(1-2):72-99, 1983.
- [Xu, 1988] Ming Xu. On some  $U, S$ -tense logics. *J. of Philosophical Logic*, 17:181-202, 1988.
- [Zanardo, 1991] A. Zanardo. A complete deductive system for since-until branching time logic. *J. Philosophical Logic*, 1991.



## COMBINATIONS OF TENSE AND MODALITY

### 1 INTERACTIONS WITH TIME

Physics should have helped us to realise that a temporal theory of a phenomenon  $X$  is, in general, more than a simple combination of two components: the statics of  $X$  and the ordered set of temporal instants. The case in which all functions from times to world-states are allowed is uninteresting; there are too many such functions, and the theory has not begun until we have begun to restrict them. And often the principles that emerge from the interaction of time with the phenomena seem new and surprising. The most dramatic example of this, perhaps, is the interaction of space with time in relativistic space-time.

The general moral, then, is that we shouldn't expect the theory of time + $X$  to be obtained by mechanically combining the theory of time and the theory of  $X$ .<sup>1</sup>

Probability is a case that is closer to our topic. Much ink has been spilled over the evolution of probabilities: take, for instance, the mathematical theory of Markov processes (Howard [1971a; 1971b] make a good text), or the more philosophical question of rational belief change (see, for example, Chapter 11 of Jeffrey [1990] and Harper [1975].) Again, there is more to these combinations than can be obtained by separate reflection on probability measure and the time axis.

probability shares many features with modalities and, despite the fact that (classical) probabilities are numbers, perhaps in some sense probability *is* a modality. It is certainly the classic case of the use of possible worlds in interpreting a calculus. (Sample points in a state space are merely possible worlds under another name.) But the literature on probability is enormous, and almost none of it is presented from the logician's perspective. So, aside from the references I have given, I will exclude it from this survey. However, it seems that the techniques we will be using can also help to illuminate problems having to do with probability; this is illustrated by papers such as D. Lewis [1981] and Van Fraassen [1971]. For lack of space, these are not discussed in the present essay.

---

<sup>1</sup>For a treatment that follows this procedure, see [Woolhouse, 1973]; [Werner, 1974] may also fit into this category, but I have not been able to obtain a copy of it. The tense logic of Woolhouse's paper is fairly crude: e.g. moments of time appear both in models and in the object language. The paper seems mainly to be of historical interest.

## 2 INTRODUCTION TO HISTORICAL NECESSITY

Modern modal logic began with necessity (or with things definable with respect to necessity), and the earliest literature, like C. I. Lewis [1918], confuses this with validity. Even in later work that is formally scrupulous about distinguishing these things, it is sometimes difficult to tell what concepts are really metalinguistic. Carnap, for instance [1956, p. 10], begins his account of necessity by directing our attention to *l*-truth; a sentence of a semantical system (or language) is *L*-true when its truth follows from the semantical rules of the language, without auxiliary assumptions. This, of course, is a metalinguistic notion. But later, when he introduces necessity into the object language [Carnap, 1956, p. 174], he stipulates that  $\Box\varphi$  is true if and only if  $\varphi$  is *L*-true.

Carnap thinks of the languages with which he is working as fully determinate; in particular, their semantical rules are fixed. This has the consequence that whatever is *L*-true in a language is eternally *L*-true in that language. (See [Schlipp, 1963, p. 921], for one passage in which Carnap is explicit on the point: he says ‘analytic sentences cannot change their truth-value’.) Combining this consequence with Carnap’s explication of necessity, we see that<sup>2</sup>

$$(1) \quad \Box\varphi \rightarrow HG\Box\varphi$$

will be valid in languages containing both necessity and tense operators: necessary truths will be eternally true. The combination of necessity with tense would then be trivialised.

But there are difficulties with Carnap’s picture of necessity; indeed, it seems to be drastically misconceived.<sup>3</sup> For one thing, many things appear to be necessary, even though the sentences that express them can’t be derived from semantical rules. In Kripke [1982], for instance, published 26 years after *Meaning and Necessity*, Saul Kripke argues that it is necessary that Hesperus is Phosphorous, though ‘Hesperus’ and ‘Phosphorous’ are by no means synonymous. Also at work in Kripke’s conception of necessity, and that of many other contemporaries, is the distinction between  $\varphi$  expressing a necessary truth, and  $\varphi$  necessarily expressing a truth. In a well-known defence of the analytic-synthetic distinction, Grice and Strawson [1956] write as follows:

---

<sup>2</sup>I use the tense logical notation of the first Chapter in this volume.

<sup>3</sup>For an early appreciation of the philosophical importance of making necessity time-dependent (the point I myself am leading up to), see [Lehrer and Taylor, 1965]. The puzzles they raise in this paper are genuine and well presented. But the solution they suggest is very implausible, and the considerations that motivate it seem to confuse semantic and pragmatic phenomena. This is a good example of a case in which philosophical reflections could have been aided by an appeal to the technical apparatus of model theory (in this case, to the model theory of tense logic).



Any form of words at one time held to express something true may, no doubt, at another time come to be held to express something false. but it is not only philosophers who would distinguish between the case where this happens as the result of a change of opinion solely as to matters of fact, and the case where this happens at least partly as a result of shift in the sense of the words (p. 157).

This distinction, at least in theory, makes it possible that a sentence  $\varphi$  should necessarily (perhaps, because of semantical rules) express a truth, even though the truth that it expresses is contingent. This idea is developed most clearly in [Kaplan, 1978].

On this view of necessity, it attaches not primarily to sentences, but to propositions. A sentence will express a proposition, which may or may not be necessary. This can be explicated using possible worlds: propositions take on truth values in these worlds, and a proposition is necessary if and only if it is true in all possible worlds.<sup>4</sup>

This conception can be made temporal without trivialising the results. Probably the simplest way of managing this is to begin with nonempty sets  $T$  of times and  $W$  of worlds;<sup>5</sup>  $T$  is linearly ordered by a relation  $<$ . I will call this the  $T \times W$  approach.

Recall that a tensed formula, say  $F\varphi$ , is true at  $\langle w, t \rangle$ , where  $w \in W$  and  $t \in T$ , if and only if  $\varphi$  is true at  $\langle w, t' \rangle$ , for some  $t'$  such that  $t < t'$ .<sup>6</sup> We now want to ask under what conditions  $\Box\varphi$  is true at  $\langle w, t \rangle$ . (In putting it this way we are suppressing propositions; this is legitimate, as long as we treat propositional attitudes as unanalysed, and assume that sentences express the same proposition everywhere.)

If we appeal to intuitions about languages like English, it seems that we should treat formulas like  $\Box\varphi$  as nontrivially tensed. This is shown most clearly by sentences involving the adjective 'possible', such as 'In 1932 it was possible for Great Britain to avoid war with Germany; but in 1937 it was impossible'. This suggests that when  $\Box\varphi$  is evaluated at  $\langle w, t \rangle$  we

<sup>4</sup>To simplify matters, I confine the discussion to the absolute necessity of S5. But perhaps I should mention in passing that in explicating the relative breeds of necessity, such as that of S4, it is easy to confuse modal relations with temporal relations, relative necessity with evanescent necessity. And, of course, tense logic was inspired in part by work on relative necessity. But the two notions are separate; an S5 breed of necessity, for instance, can be evanescent. And when tense and modality are combined, it is very important to attend to the distinction.

<sup>5</sup>I dislike this way of arranging things for philosophical reasons. it doesn't strike me as a logical truth that all worlds have the same temporal orderings: some may have an earliest time, for instance, and others not. Also, the notion of different worlds sharing the same time is philosophically problematic; it is hard to reconcile with a plausible theory of time, when the possible worlds differ widely. Finally, I like to think of possible worlds as overlapping, so that at the same moment may have alternative futures. This requires a more complicated representation. However, the  $T \times W$  arrangement will do for now.

<sup>6</sup>See the discussion of the interpretation of tense in Chapter 1 of this volume.

are considering what is *then necessary*; what is true in all worlds at that particular time,  $t$ .

The rule then is that  $\Box\varphi$  is true at  $\langle w, t \rangle$  if and only if  $\varphi$  is true at  $\langle w', t \rangle$  for all  $w' \in W$ . If we like, we can make this relational. Let  $\{\approx_t: t \in T\}$  be a family of equivalence relations on  $W$ , and let  $\Box\varphi$  be true at  $\langle t, w \rangle$  if and only if  $\varphi$  is true at  $\langle t, w' \rangle$  for all  $w' \in W$  such that  $w \approx_t w'$ .

The resulting theory generates some validities arising from the assumption that the worlds share a common temporal ordering. Formulas (2) and (3) are two such validities, corresponding to the principle that one world has a first moment if and only if all worlds do.

$$(2) \quad P[\varphi \vee \neg\varphi] \leftrightarrow \Box P[\varphi \vee \neg\varphi]$$

$$(3) \quad H[\varphi \wedge \neg\varphi] \leftrightarrow \Box H[\varphi \wedge \neg\varphi]$$

In case  $\approx_t$  is the universal relation for every  $t$  (or the relations  $\approx_t$  are simply omitted from the satisfaction conditions) there are other validities, such as (4) and (5).

$$(4) \quad P\Box\varphi \rightarrow \Box P\varphi$$

$$(5) \quad F\Box\varphi \rightarrow \Box F\varphi$$

As far as I know, the general problem of axiomatising these logics has not been solved. But I'm not sure that it is worth doing, except as an exercise. The completeness proofs should not be difficult, using Gabbay's techniques (described in Section 4, below). and these logics do not seem particularly interesting from a philosophical point of view.

But a more interesting case is near to hand. The tendency we have noted to bring Carnap's metalinguistic notion of necessity down to earth has made room for the reintroduction of one of the most important notions of necessity: practical necessity, or historical necessity.<sup>7</sup> This is the sort of necessity that figures in Aristotle's discussion of the Sea Battle (*De Int.* 18<sup>b</sup>25–19<sup>b</sup>4), and that arises when free will is debated. It also seems to be an important background notion in practical reasoning. Jonathan Edwards, in his usual lucid way, gives a very clear statement of the matter.

Philosophical necessity is really nothing else than the full and fixed connection between the things signified by the subject and predicate of a proposition, which affirms something to be true . . . . [This connection] may be fixed and made certain, because the existence of that thing is already come to pass; and either

---

<sup>7</sup>I am not sure if there are personal, or relational varieties of inevitability; it seems a bit peculiar to my ear to speak of an accident John caused as inevitable for Mary, but not inevitable for John. If there are such sorts of inevitability I mean to exclude them, and to speak only of impersonal inevitability. Thus 'inevitable' does not belong to the same modal family as 'able', since the latter is personal.

now is, or has been; and so has as it were made sure of existence. And therefore, the proposition which affirms present and past existence of it, may be this means be made certain, and necessarily and unalterably true; the past event has fixed and decided that matter, as to its existence; and has made it impossible but that existence should be truly predicted of it. Thus the existence of whatever is already come to pass, is not become necessity; 'tis become impossible it should be otherwise than true, that such a thing has been. [Edwards, 1957, pp. 152–3]

Historical necessity can be fitted into the  $T \times W$  framework; it is merely a matter of adjusting the relations  $\approx_t$  so that if  $w \approx_t w'$ , then  $w$  and  $w'$  share the same past up to and including  $t$ . So for  $t' < t$ , atomic formulas must be treated the same way in  $w$  and  $w'$ . Furthermore, we have to stipulate that historical possibilities diminish monotonically with the passage of time: if  $t < t'$ , then  $\{w' : w \approx_{t'} w'\} \subseteq \{w; : w \approx_t w'\}$ . This interaction between time and relative necessity creates distinctive validities, such as (6) and (7).

(6)  $\varphi \leftrightarrow \Box\varphi$ , if  $\varphi$  contains no occurrences of  $F$ .

(7)  $P\Box\varphi \rightarrow \Box P\varphi$

Formula (8), on the other hand, is clearly invalid.

(8)  $\Box Pp \rightarrow P\Box p$

These correspond to rather natural intuitions relating the flow of time to the loss of possibilities.

There is another way of representing historical necessity, which perhaps will seem less straightforward to logicians steeped in possible worlds. Time can be treated as non-linear (branching only towards the future), and worlds represented as branches on the resulting ordered structure. This corresponds very closely to the  $T \times W$  account: (6) and (7) remain valid, and (8) invalid. But the validities are not the same. This matter will be taken up below, in Section 4.

So much for necessity; I will deal more briefly with 'ought' and conditionals.

As Aristotle points out, we don't deliberate about just anything; in particular, we deliberate only about what is in our power to determine. [Ne 1112<sup>a</sup> 19f.] But the past, and the instantaneous present, are not in our power: deliberation is confined to future alternatives.

This suggests that deontic logic, insofar as it investigates practical oughts, should identify its possibilities with the ones of historical necessity. Unfortunately, this conception played little or no role in the early interpretation of deontic logic; those who developed the deontic applications of possible worlds semantics seemed to think of deontic possibilities ahistorically, as

'perfect worlds' in which all norms are fulfilled.<sup>8</sup> Historical possibilities, on the other hand, are typically imperfect; life is full of occasions on which we have to make the best of a bad situation. In my opinion, this is one reason why deontic logic has seemed to most philosophers to consist largely of a sterile assortment of paradoxes, and why its influence on moral philosophy has been so fruitless.

Conditionals have been intensively studied by philosophical logicians over the last fifteen years, and this has created an extensive literature. Relatively little of this effort has been devoted to the interaction of conditionals with tense. But there is reason to think that important insights may be lost if conditionals are studied ahistorically. One very common sort of conditional (the philosopher's novel example of a 'subjunctive conditional') is exemplified by (9).

(9) If Oswald hadn't shot Kennedy, then Kennedy would be alive today.

These conditionals seem to be closely related to historical possibilities; they envisage courses of events that diverge at some point in the past from the actual one. And this in turn suggests that there may be close connections between historical necessity and some conditionals.

Examples like the following four provide evidence of a different sort.

(10) He would go if she would go.

(11) He will go if she will go.

(12) he would have gone if she were to have gone.

(13) he went if she went.

Sentences (10) and (11) seem hardly to differ in meaning, if (10) has to do with the future. On the other hand, (12) and (13) are very different. If he didn't go, but would have gone if she had, (12) is true and (13) false.

This suggests that there may be systematic connections between tense, mood and the truth conditions of conditionals. According to one extreme proposal, the difference in 'mood' between (9) and the past-present conditionals form 'If Oswald didn't shoot Kennedy, then Kennedy is alive today' can be accounted for solely in terms of the interaction of tense operators and the conditional.<sup>9</sup> This has in favour of it the grammatical fact that 'would' is the past tense of 'will'. But the matter is complex, and it is difficult to see how much merit there is in the suggestion.

There have been recent signs of interest in the interaction of tense and conditionals; the most systematic of these is [Thomason and Gupta, 1981]

<sup>8</sup>See [Von Wright, 1968] and Hintikka [1969; 1971].

<sup>9</sup>See [Thomason and Gupta, 1981, pp. 304-305].

If this study is any indication, the topic is surprisingly complicated. But the complications may prove to be of philosophical interest.

The relation between historical necessity and quantum mechanics is a topic that I will not discuss at any great length. The indeterminacy that is associated with microphenomena seems at first glance to invite a treatment using alternative futures; and one of the approaches to the measurement problem in quantum theory, the ‘many-worlds interpretation’, does appear to do just this. (See [DeWitt and Graham, 1973] for more information about the approach.)

But alternative futures don’t provide in themselves an adequate representation of the physical situation, because the quantum mechanical probabilities can’t be treated as distributions over a set of fully determinate worlds.<sup>10</sup> Some further apparatus would have to be introduced to secure the right system of nonboolean probabilities, and as far as I can see, what is required would have to go beyond the resources of possible worlds semantics: there is no escaping an analysis of measurement interactions, or of interactions in general.

Possible world semantics may help to make the ‘many worlds’ approach to quantum indeterminacy seem less frothy; the prose of philosophical modal realists, such as D. Lewis [1970], is much more judicious than that in which the physicists sometimes indulge. (See, for instance, [DeWitt, 1973, p. 161].) So, modal logic may be of some help in sorting out the philosophical issues; but this leaves the fundamental problems untouched. Possible worlds are not in themselves a key to the problem of measurement in quantum mechanics.

The following sections will aim at fleshing out this general introduction with further historical information, more detailed descriptions of the relevant logical theories, and more extensive references to the literature.

### 3 HISTORICAL NECESSITY

The first sustained discussion of this topic, from the standpoint of modern tense logic is (as far as I know) Chapter 7 of [Prior, 1967] entitled ‘Time and determinism’.<sup>11</sup> Prior’s judgement and philosophical depth, as well as his

---

<sup>10</sup>See, for instance [Wigner, 1971] and [Fine, 1982].

<sup>11</sup>The mention of historical necessity and its combination with deontic operators in the *tour de force* at the end of [Montague, 1968] probably takes precedence if you go by date of composition. But Montague’s discussion is very compressed, and neglects philosophical motivation. And some interesting things are said about indeterminism in Prior’s earlier book [Prior, 1957]. But the connection does not seem to be made there between the philosophical issues and the problem of interpreting future tense in treelike frames. The ingredients of a model theoretic treatment of historical necessity also occurred at an early date to Dana Scott. Like much of his work in modal and tense logic, it remained unpublished, but there is a mimeographed paper [Scott, 1967].

readable style, make this required reading for anyone seriously interested in historical necessity.

Prior's exposition is informal, and sprinkled with historical references to the philosophical debate over determinism. In this debate he unearths a logical determinist argument, that probably goes back to ancient times. According to this argument, if  $\varphi$  is true then, at any previous time,  $F\varphi$  must have been true. But choose such a time, and suppose that at this earlier time  $\varphi$  could have failed to come about; then  $F\varphi$  could not have been true at this time. It seems to follow that the determinist principle

$$(14) \quad \varphi \rightarrow H\Box F\varphi$$

holds good.

In discussing the argument and some ways of escaping from it, Prior is fairly flexible about this object language; in particular, he allows metric tense operators. Since these complicate matters from a semantic point of view, I will ignore them, and consider languages whose only modal operators are  $\Box$  and the nonmetric tenses.

Also (and this is more unfortunate), Prior [1967] speaks loosely in describing models. At the place where his indeterminist models are introduced, for instance, he writes as follows.

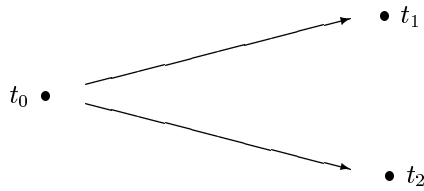
... we may define an Ockhamist *model* as a line without beginning or end which may break up into branches as it moves from left to right (i.e. from past to future), though not the other way ...  
(p. 126)

From this description it is clear that Prior is representing historical necessity by means of non-linear time, rather than according to the  $T \times W$  format described in Section 2, above. But it is a little difficult to tell exactly what mathematical structures have been characterised; probably, Prior had in mind trees whose branches all have the order type of the (negative and non-negative) integers.

To bring this into accord with the usual treatment of linear nonmetric tense logic, we will liberalise Prior's account.

**DEFINITION 1.** A treelike frame  $\mathfrak{A}$  for tense logic is a pair  $\langle T, < \rangle$ , where  $T$  is a non-empty set and  $<$  is a transitive ordering on  $T$  such that if  $t_1 < t_2$  and  $t_2 < t_3$  then either  $t_1 = t_3$  or  $t_1 < t_3$  or  $t_3 < t_1$ .

As in ordinary tense logic, we imagine an assignment of truth values to atomic formulas at each  $t \in T$ , and truth-functional connectives are treated in the usual way. But things become perplexing when you try to interpret future tense in these structures. Take a very simple branching case, with just three moments, and imagine that  $p$  is true at  $t_0$  and  $t_1$ , and false at  $t_2$ .



Is  $Fp$  true at  $t_0$ ? It is hard to say.

Moreover, as you reflect on the problem, it becomes clear that Prior's juxtaposition of this technical problem with bits from figures like Diodorus Cronus, Peter de Rivo, and Jonathan Edwards is not merely an antiquarian quirk. There is a genuine connection. These treelike frames represent ways in which things can evolve indeterministically. A definition of satisfaction for a language with tense operators that is suited to such structures would automatically provide a way of making tense compatible with indeterministic cases. And it is just this that the logical argument for determinism claims can't be done. The technical problem can't be solved without getting to the bottom of this argument.

If the argument is correct, any definition of satisfaction for these structures will be incorrect—will generate validities that are at variance with the intended interpretation. Łukasiewicz's [1967] earlier three-valued solution is like this, I believe. Not because it makes some formulas neither true nor false, but because the formulas it endorses as valid are so far off the mark. It is bad enough that  $Fp \vee \neg Fp$  is invalid, but also the approach would make  $[[\Diamond Fp \wedge \Diamond \neg Fp] \wedge [\Diamond Fq \wedge \Diamond \neg Fq]] \rightarrow [Fp \leftrightarrow Fq]$  valid, if  $\varphi$  is true if and only if  $\varphi$  takes the intermediate truth value.<sup>12</sup>

Nor does the logic that Prior calls 'Peircian' strike me as more satisfactory, from a philosophical standpoint, though it does lead to some interesting technical problems relating to axiomatisability. Here,  $F_v arphi$  is treated as true at  $t$  in case the moments at which  $\varphi$  is true bar the future paths through  $t$ ; i.e., every branch through  $t$  contains a moment subsequent to  $t$  at which  $\varphi$  is true. On the Peircian approach,  $F\varphi \vee \neg F\varphi$  is valid, but  $Fp \vee F\neg p$  is not; nor is  $p \rightarrow PFP$ .<sup>13</sup> As Prior says, sense can be made of this by reading  $F$  as 'will inevitably'. Though this helps us to see what is going on, it is not the intended interpretation.

The most promising of Prior's suggestions for dealing with indeterminist future tense is the one he calls 'Ockhamist'. The theory will be easier to present if we first work out the satisfaction conditions for  $\Box\varphi$  in treelike frames. Intuitively,  $\Box\varphi$  is true at  $t$  if  $\varphi$  is true at  $t$  no matter what the

<sup>12</sup>Prior briefly criticises Łukasiewicz's treatment of future tense [Prior, 1967, p. 135]; for a more extended criticism, see [Seeskin, 1971].

<sup>13</sup>Notice that this corresponds to one of the informal principles used in the logical argument for determinism.

future is like. And a way the future can be like will be represented by a fully determinate—i.e. linear—path beyond  $t$ . Since the frames are treelike, these correspond to the branches, or maximal chains, through  $t$ .

DEFINITION 2. Where  $\langle T, < \rangle$  is a treelike model structure and  $t \in T$ , a branch through  $t$  is a maximal linearly ordered subset of  $T$  containing  $t$ ;  $B_t$ .

To make sense of  $\varphi$  being true at  $t$  no matter what the future is like, we will have to think of formulas being satisfied not just at moments  $t$ , but at pairs  $\langle t, b \rangle$ , where  $b \in B_t$ . Prior explains it this way. On the Ockhamist approach formulas like  $Fp$  are given ‘*prima facie* assignments’ at  $t$ ; such an assignment is made by choosing a particular  $b$  in  $B_t$ . His idea seems to be that there is something more provisional about the selection of  $b$  than about that of  $t$ ; but this he does not articulate or defend very fully. In the technical formulation of the theory there is no asymmetry between moments and branches; it is just that two parameters need to be fixed in evaluating formulas.

Once satisfaction is made relative to pairs  $\langle t, b \rangle$  for some formulas, it must be relativised in the same way for all formulas; otherwise the recursive definition of satisfaction will become snarled. So, except for future tense, the definition will go like this.

DEFINITION 3. A function  $h$  assigning each atomic formula a subset of  $T$  is called an (Ockhamist) *assignment*, as in Chapter 1 of this volume.

DEFINITION 4. The  $h$  truth value  $\|\varphi\|_{\langle t, b \rangle}^h$  of  $\varphi$  at the pair  $\langle t, b \rangle$  is defined as follows. We assume here that  $\|\varphi\|_{\langle t, b \rangle}^h = 0$  iff  $\|\varphi\|_{\langle t, b \rangle}^h \neq 1$ .

$$\begin{aligned} \|\varphi\|_{\langle t, b \rangle}^h &= 1 \text{ iff } t \in h(\varphi), \text{ if } \varphi \text{ is atomic,} \\ \|\neg\varphi\|_{\langle t, b \rangle}^h &= 1 \text{ iff } \|\varphi\|_{\langle t, b \rangle}^h = 0, \\ \|\varphi \wedge \psi\|_{\langle t, b \rangle}^h &= 1 \text{ iff } \|\varphi\|_{\langle t, b \rangle}^h = 1 \text{ and } \|\psi\|_{\langle t, b \rangle}^h = 1, \\ \|\varphi \vee \psi\|_{\langle t, b \rangle}^h &= 1 \text{ iff } \|\varphi\|_{\langle t, b \rangle}^h = 1 \text{ or } \|\psi\|_{\langle t, b \rangle}^h = 1, \\ \|\varphi \rightarrow \psi\|_{\langle t, b \rangle}^h &= 1 \text{ iff } \|\varphi\|_{\langle t, b \rangle}^h = 0 \text{ or } \|\psi\|_{\langle t, b \rangle}^h = 1, \\ \|\mathcal{P}\varphi\|_{\langle t, b \rangle}^h &= 1 \text{ iff for some } t' < t, \|\varphi\|_{\langle t', b \rangle}^h = 1, \\ \|\Box\varphi\|_{\langle t, b \rangle}^h &= 1 \text{ iff for all } b \in B_t, \|\varphi\|_{\langle t, b \rangle}^h = 1. \end{aligned}$$

This definition renders  $p \rightarrow \Box p$  valid, though not every substitution instance of it is valid. This can be easily changed by letting assignments take atomic formulas into subsets of  $\{\langle t, b \rangle : t \in T \text{ and } b \in B_t\}$ ; [Prior, 1967, pp. 123–123] discusses the matter.

At this point, the way to handle  $F\varphi$  is forced on us. We use the branch that is provided by the index.

$$\|F\varphi\|_{\langle t, b \rangle}^h = 1 \text{ iff for some } t' \in b \text{ such that } t < t', \|\varphi\|_{\langle t', b \rangle}^h = 1.$$



The Ockhamist logic is conservative; it's easy to show that if  $\varphi$  contains no occurrences of  $\Box$  then  $\varphi$  is valid for treelike frames if and only if  $\varphi$  is valid in ordinary tense logic. So indeterminist frames can be accommodated without sacrificing any orthodox validities. This is good for those who (like me) are not determinists, but feel that these validities are intuitively plausible. Finally, the Ockhamist solution thwarts the logical argument for determinism by denying that if  $\varphi$  is true at  $t$  (i.e. at  $\langle b, t \rangle$ , for some selected  $b$  in  $B_t$ ) then  $\Box\varphi$  is.

This way out of the argument bears down on its weakest joint; but the argument is so powerful that even this link resists the pressure; it is hard for an indeterminist to deny that  $\Box\varphi$  must be true if  $\varphi$  is. To a thoroughgoing indeterminist, the choice of a branch  $b$  through  $t$  has to be entirely *prima facie*; there is no special branch that deserves to be called the 'actual' future through  $t$ .<sup>14</sup> Consider two different branches  $b_1$  and  $b_2$ , through  $t$ , with  $t < t_1 \in b_1$  and  $t < t_2 \in b_2$ . From the standpoint of  $t_1$ ,  $b_1$  is actual (at least up to  $t_1$ ). From the standpoint of  $t_2$ ,  $b_2$  is actual (at least up to  $t_2$ ). And neither standpoint is correct in any absolute sense. In exactly the same way, no particular moment of linear time is 'present'.

But then it seems that the Ockhamist theory gives no account of truth relative to a moment  $t$ , and it also suggests very strongly that if  $\varphi$  is true at  $t$  then  $\Box\varphi$  is also true at  $t$ . The only way that a thing can be true at a moment is for it to be settled at that moment.

In Thomason [1970], it is suggested that such an absolute notion of truth can be introduced by superimposing Van Fraassen's treatment of truth-value gaps onto Prior's Ockhamist theory.<sup>15</sup> The resulting definition is very simple.

DEFINITION 5.

$$\begin{aligned} \|\varphi\|_t^h = 1 & \text{ iff } \|\varphi\|_{\langle t, b \rangle}^h = 1 \text{ for all } b \in B_t, \\ \|\varphi\|_t^h = 0 & \text{ iff } \|\varphi\|_{\langle t, b \rangle}^h = 0 \text{ for all } b \in B_t. \end{aligned}$$

This logic preserves the validities of linear tense logic; indeed,  $\varphi$  is Ockhamist valid if and only if it is valid here. Also, the rule holds good that if  $\|\varphi\|_t^h = 1$  then  $\|\Box\varphi\|_t^h = 1$ .

Thus, this theory endorses the principle (rejected by the Ockhamist theory) that if a thing is true at  $t$  then its truth at  $t$  is settled. It may seem at first that it validates all the principles needed for the logical determinist argument, but of course (since the logic allows branching frames) it must sever the argument somewhere. The way in which this is done is subtle. The scheme  $\varphi \rightarrow HF\varphi$  is valid, but this does not mean that if  $\varphi$  is true at  $t$  then  $F\varphi$  is true at any  $t' < t$ . That is, it is *not* the case that if  $\|\varphi\|_t^h = 1$  then

<sup>14</sup>See [Lewis, 1970], and substitute 'the actual future' for 'the actual world' in what he says. *That* is the view of the thoroughgoing indeterminist.

<sup>15</sup>See, for instance Van Fraassen [1966; 1971].

$\|F\varphi\|_{t'}^h = 1$  for all  $t' < t$ . The validity of this scheme only means that for all  $b \in B_t$ , if  $\|\varphi\|_{\langle t, b \rangle}^h = 1$  then  $\|F\varphi\|_{\langle t', b \rangle}^h = 1$  for all  $t' < t$ . But there may be  $b' \in B_{t'}$  which are not in  $B_t$ .

To put it another way, the fact that  $F\varphi$  is true at  $t$  from the perspective of a later  $t'$  does not make  $F\varphi$  *absolutely* true at  $t$ , and so need not imply that  $\Box F\varphi$  is true at  $t$ .

This manoeuvre makes use of the availability of truth-value gaps. To make this clearer, take a future-oriented version of the logical determinist argument:  $F\varphi \vee F\neg\varphi$  is true at any  $t$ ; so  $F\varphi$  is true at  $t$  or  $\neg F\varphi$  is true at  $t$ ; so  $\Box F\varphi$  is true at  $t$  or  $\Box\neg F\varphi$  is true at  $t$ . The supervaluational theory blocks the second step of this argument: in any such theory, the truth of  $\psi \vee \neg\psi$  does not imply that  $\psi$  is true or  $\neg\psi$  is true.

Thomason suggests that this logic represents the position endorsed by Aristotle in *De Int.* 18<sup>b</sup>25 – 19<sup>b</sup>4, but his suggestion is made without any analysis of the very controversial text, or discussion of the exegetical literature. For a close examination of the texts, with illuminating philosophical discussion, see [Frede, 1970]; see also [Sorabji, 1980]. for a broadly-based examination of Aristotle's views that nicely illustrates the value of treelike frames as an interpretive device, see [Code, 1976]. The suggestion is also made in [Jeffrey, 1979]. For information about the medieval debate on this topic, see [Normore, 1982].

#### 4 THE TECHNICAL SIDE OF HISTORICAL NECESSITY

The mathematical dimension of the picture painted in the above section is still relatively undeveloped. At present, most of the results known to me deal with axiomatisability in the propositional case.

We have already characterised one important variety of propositional validity:  $\varphi$  is *Ockhamist valid* if it is satisfied at all pairs  $\langle t, b \rangle$ , relative to all Ockhamist assignments on all treelike frames. (See Definitions 2– 4, above.) The time has come to give an official definition of  $T \times W$  validity.

**DEFINITION 6.** A  $T \times W$  frame is a quadruple  $\langle W, T, <, \approx \rangle$ , where  $W$  and  $T$  are non-empty sets,  $<$  is a transitive relation on  $T$  which is also irreflexive and linear (i.e.  $t \not< t$  for all  $t \in T$ , and either  $t < t'$  or  $t' < t$  or  $t = t'$  for all,  $t, t' \in T$ ), and  $\approx$  is a 3-place relation on  $T \times W \times W$ , such that (1) for all  $t, \approx_t$  is an equivalence relation (i.e.  $w \approx_t w$  for all  $t \in T$  and  $2 \in W$ , etc.), and (2) for all  $w_1, w_2 \in W$  and  $t, t' \in T$ , if  $w_1 \approx_t w_2$  and  $t' < t$  then  $w_1 \approx_{t'} w_2$ . The intention is that  $w \approx_t w'$  if  $w$  and  $w'$  are historical alternatives through  $t$ , and so differ only in what is future to  $t$ .

**DEFINITION 7.** A function  $h$  assigning each atomic formula a subset of  $T \times W$  is an *assignment*, provided that if  $w \approx_t w'$  and  $t_1 \leq t$  then  $\langle t_1, w \rangle \in h(\varphi)$  iff  $\langle t_1, w' \rangle \in h(\varphi)$ .

DEFINITION 8. The  $h$ -truth value  $\|\varphi\|_{\langle t, w \rangle}^h$  of  $\varphi$  at the pair  $\langle t, w \rangle$  is defined by a recursion that treats truth-functional connectives in the usual way. The clauses for tense and necessity run as follows.

$$\begin{aligned} \|P\varphi\|_{\langle t, w \rangle}^h &= 1 \quad \text{iff for some } t' < t, \|\varphi\|_{\langle t', w \rangle}^h = 1, \\ \|F\varphi\|_{\langle t, w \rangle}^h &= 1 \quad \text{iff for some } t' \text{ such that } t < t', \|\varphi\|_{\langle t', w \rangle}^h = 1, \\ \|\Box\varphi\|_{\langle t, w \rangle}^h &= 1 \quad \text{iff for all } w' \text{ such that } w \approx_t w', \|\varphi\|_{\langle t, w' \rangle}^h = 1. \end{aligned}$$

A formula is  $T \times W$  valid if it is satisfied at every pair  $\langle t, w \rangle$  by every assignment on every  $T \times W$  frame.<sup>16</sup>

There are some validities that are peculiar to these  $T \times W$  frames, and that arise from the fact that only a single temporal ordering is involved in these frames: (15) and (16) are examples,

$$(15) \quad FG[p \wedge \neg p] \rightarrow \Box FG[p \wedge \neg p]$$

$$(16) \quad GF[p \vee \neg p] \rightarrow \Box GF[p \vee \neg p]$$

Example (15) is valid because its antecedent is true at  $\langle t, w \rangle$  if and only if there is a  $t'$  that is  $<$ -maximal with respect to  $w$ ; but this holds if and only if there is a  $t'$  that is  $<$ -maximal absolutely. Example (16) is similar, except that this time what is at stake is the non-existence of a maximal time.

Burgess remarked (in correspondence) that  $T \times W$  validity is recursively axiomatisable, since it is essentially first-order. But as far as I know the problem of finding a reasonable axiomatisation for  $T \times W$  validity is open. I would expect the techniques discussed below, in connection with Kamp validity, to yield such an axiomatisation.

Although (15) and (16) may be reasonable given certain physical assumptions, they do not seem so plausible from a logical perspective. After all, if  $w \approx_t w'$ , all that is required is that  $w$  and  $w'$  should share a certain segment of the past, and this implies that the structure of time should be the same in  $w$  and  $w'$  on this segment. But it is not so clear that  $w$  and  $w'$  should participate in the same temporal structure after  $t$ . This suggests a more liberal sort of  $T \times W$  frame, first characterised by Kamp [1979].<sup>17</sup>

<sup>16</sup>I will adhere to this terminology here, but I am not confident that it is the best terminology, over the long run. Varieties of  $T \times W$  validity tend to proliferate, and this is only one of them, and probably not the most interesting. Perhaps it would be better to speak of (Fixed  $T$ )  $\times W$  validity— but this is awkward.

<sup>17</sup>This paper of Kamp's deserves summary because it became widely known and is historically important, though it was never published. Kamp had evidently been thinking about these matters for several years; the latest draft of the paper that I have seen was finished early in 1979. The paper contains much valuable philosophical discussion of historical necessity, and defines the type of validity that I here call 'Kamp validity'. An axiomatisation is proposed in the paper, which was never published because of difficulties that came to light in the completeness argument that Kamp had sketched for validity in dense Kamp frames. In 1979, Kamp discovered a formula that was Kamp valid but

DEFINITION 9. A Kamp frame is a triple  $\langle \mathcal{T}, W, \approx \rangle$  where  $W$  is a non-empty set,  $\mathcal{T}$  is a function from  $W$  to transitive, irreflexive linear orderings (i.e. if  $w \in W$  then  $\mathcal{T}(w) = \langle T_w, <_w \rangle$ , where  $<_w$  is an ordering on  $T_w$  as in  $T \times W$  frames), and  $\approx$  is a relation on  $\{ \langle t, w, w' \rangle : w, w' \in W \text{ and } t \in \mathcal{T}(w) \cap \mathcal{T}(w') \}$  such that for all  $t, \approx_t$  is an equivalence relation, and if  $w \approx_t w'$  then  $\{ t_1 : t_1 \in T_w \text{ and } t_1 <_w t \} = \{ t_1 : t_1 \in T_{w'} \text{ and } t + 1 <_{w'} t_1 \}$ . Also, if  $w \approx_t w'$  and  $t' <_w t$  then  $w \approx_{t'} w'$ .

The definitions of an assignment, and of the  $h$ -truth value  $\|\varphi\|_{\langle t, w \rangle}^h$  of  $\varphi$  at the pair  $\langle t, w \rangle$  (where  $t \in \mathcal{T}_w$  and  $h$  is an assignment) are readily adapted from Definitions 7 and 8.

Besides (AK0) all classical tautologies, Kamp takes as axioms all instances of the following schemes.<sup>18</sup>

$$(AK1) \quad H[\varphi \rightarrow \psi] \rightarrow [P\varphi \rightarrow P\psi]$$

$$(AK2) \quad G[\varphi \rightarrow \psi] \rightarrow [F\varphi \rightarrow F\psi]$$

$$(AK3) \quad PP\varphi \rightarrow P\varphi$$

$$(AK4) \quad FF\varphi \rightarrow F\varphi$$

$$(AK5) \quad PG\varphi \rightarrow \varphi$$

$$(AK6) \quad FH\varphi \rightarrow \varphi$$

$$(AK7) \quad PF\varphi \rightarrow [P\varphi \vee \varphi \vee F\varphi]$$

$$(AK8) \quad FP\varphi \rightarrow [P\varphi \vee \varphi \vee F\varphi]$$

$$(AK9) \quad \Box[\varphi \rightarrow \psi] \rightarrow [\Box\varphi \rightarrow \Box\psi]$$

$$(AK10) \quad \Box\varphi \rightarrow \varphi$$

$$(AK11) \quad \Diamond\varphi \rightarrow \Box\Diamond\varphi$$

$$(AK12) \quad \Diamond P\varphi \rightarrow P\Diamond\varphi$$

$$(AK13) \quad \Box\varphi \vee \Box\neg\varphi, \text{ if } \varphi \text{ is atomic.}$$

not provable from the axioms of the paper. Later, Thomason discovered other sorts of counterexamples. As far as I know, the axiomatisation problem for Kamp validity was open until Dov Gabbay encountered the problem at a workshop for [the first edition of] this *Handbook*, in the fall of 1981. Due to the wide circulation of [Kamp, 1979], a number of erroneous references have crept into the literature concerning the existence of an axiomatisation of Kamp validity. Gabbay has not yet published his result [Editors' note: see Section 7.7 of D. M. Gabbay, I. Hodkinson and M. Reynolds. *Temporal Logic*, volume 1, Oxford University Press, 1994.], and any such reference in a work published before [1981, the first edition of] this *Handbook* is likely to be mistaken.

<sup>18</sup>I omit the axiom scheme for density, and omit a redundant axiom for  $\Box$ . And the system I describe is only one of several discussed by Kamp.

As well as (RK0) *modus ponens*, Kamp posits the rules given by the following three schemes.

$$(RK1) \quad \varphi/H\varphi$$

$$(RK2) \quad \varphi/G\varphi$$

$$(RK3) \quad \varphi/\Box\varphi$$

Readers familiar with axioms for modal and tense logic will see that this list falls into three natural parts. Classical tautologies, *modus ponens*, (RK1), (RK2) and (AK1)–(AK8) are familiar principles of ordinary tense logic without modality. Classical tautologies, *modus ponens*, (RK3), and (AK9)–(AK11) are principles of the modal logic **S5**. Axiom (AK12) is a principle combining tense and modality; this principle was explained informally in Section 2. The validity of (AK13) reflects the treatment of atomic formulas as noncontingent; see the provision in Definition 7. If tense operators were not present, (AK13) would of course trivialise  $\Box$ , rendering every formula noncontingent.

To establish the incompleteness of (AK0)–(AK13) + (RK0)–(RK3), consider the formula (17), discovered by Kamp, where  $E_1(\varphi)$  is  $F\varphi \wedge G[\varphi \vee P\varphi]$ .

$$(17) \quad [PE_1(p) \wedge \Diamond PE_1(q)] \rightarrow [P[E_1(p) \wedge P\Diamond E_1(q)] \vee \vee P[\Diamond E_1(q) \wedge \Diamond E_1(p)] \vee P[E_1(p) \wedge \Diamond E_1(q)]]$$

The validity of (17) in Kamp frames follows from the fact that these frames are closed under the sort of diagram completion given in Figure 1. Given  $t_1, t_2, t_3, t'_1, t'_3$  and the relations of the diagram, it must be possible to interpolate at  $t'_2$  in  $w'$  alternative to  $t_2$ , with  $t'_3 < t'_2 < t'_1$ .<sup>19</sup> Formula (17) is complicated, but I think I can safely leave the task to checking its Kamp validity to the reader.

One proof that (17) is independent of (AK0)–(AK13) + (RK0)–(RK3) makes use of still another sort of frame, which is closer to a Henkin construction than the  $T \times W$  frames. If our task is to build models out of maximal consistent sets of formulas, the ‘times’ of a  $T \times W$  frame are rather artificial; they would have to be equivalence sets of maximal consistent sets. In *neutral frames*, the basic elements are moments, or instantaneous slices of evolving worlds, which are organised by intra-world temporal relations, and interworld alternativeness relations. Neutral frames tend to proliferate, because of the many conditions that can be imposed on the relations; hence my use of subscripts in describing them.

**DEFINITION 10.** A neutral frame<sub>1</sub> is a triple  $\langle W, \mathcal{U}, \approx \rangle$ , where (1)  $W$  is a nonempty set, (2)  $\mathcal{U}$  is a function whose arguments are members of  $w$  of  $W$

<sup>19</sup>In fact, since we are working with Kamp frames,  $t'_2 = t_2$ . But I put it in this more general way in anticipation of what I will call *neutral frames*, so that Figure 1 will be similar in format to Figure 2.



Figure 1. Interpolation

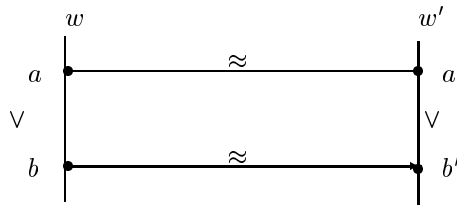


Figure 2. One-way completion.

and whose values are orderings  $\langle U_w, <_w \rangle$  such that  $U - w$  is a nonempty set and  $<_w$  is a transitive<sup>20</sup> ordering on  $U_w$  such that for all  $a, b \in U_w$  either  $a <_w b$  or  $b <_w a$  or  $a = b$ , (3) if  $w, w' \in W$  and  $w \neq w'$  then  $U_w$  and  $U_{w'}$  are disjoint, and (4)  $\approx$  is an equivalence relation on  $\cup\{U_w : w \in W\}$ , and (5) if  $a \approx a'$  and  $b <_w a$  then there is some  $b' \in U_{w'}$  (where  $a' \in U_{w'}$ ) such that  $b \approx b'$  and  $b' <_{w'} a'$ .

Here, each  $U_w$  corresponds to the set of instantaneous slices of the world  $w$ ;  $\approx$  is the alternativeness relation between these slices.

The diagram-completion property expressed in (5) looks as shown in this picture. It's important to realise that nothing prevents  $a \neq b$  while at the same time  $a' = b'$  in Figure 2. Of course, in this case we will also have  $a \approx b$ , which would be something like history repeating itself, at least in all respects that are settled by the past.

Everything generated by (AK0)–(AK13) + (RK0)–(RK3) is valid in neutral frames<sub>1</sub>. (And the fact that the converse holds, so that we have a completeness result; see [Thomason, 1981c].) But it is easy to show that (17) is invalid in neutral frames<sub>1</sub>.

<sup>20</sup>But not necessarily irreflexive!

This process can be continued; for instance, stronger conditions of diagram completion can be imposed on neutral frames, which extend the propositional validities, and completeness conditions obtained for the resulting sorts of frames. Details can be found in [Thomason, 1981c]; but this effort did not produce an axiomatisation of Kamp validity.

Using his method of constructing irreflexive models (see [Gabbay, 1981]), Gabbay has shown that all the Kamp validities will be obtained if (AG1) and (RG1) are added to (AK0)–(AK13) + (RK0)–(RK3). Some terminology is needed to formulate (RG1); we need a way of talking, to put it intuitively, about formulas which record a finite number of steps forwards, backwards and sideways in Kamp frames.

**DEFINITION 11.** Let  $\alpha_i \in \{P, F, \Diamond\}$  for  $1 \leq i \leq n, n \geq 0$ . Then  $f(\varphi) = \varphi$ , and

$$f_{\alpha_1, \dots, \alpha_n}(\varphi_0, \dots, \varphi_{n-1}, \varphi_n) = \varphi_0 \wedge \alpha_1[\varphi_1 \wedge \dots \wedge \alpha_{n-1}[\varphi_{n-1} \wedge \alpha_n \varphi_n] \dots].$$

So, for instance,  $f_{F, \Diamond, P}(p_0, p_1, p_2, p_3)$  is  $p_0 \wedge F[p_1 \wedge \Diamond[p_2 \wedge Pp_3]]$ .

The axiom and rule are as follows:

$$(AG1) \quad [\Box \neg \varphi \wedge H\Box \varphi \wedge \Box \psi] \rightarrow G\Box H[[\neg \varphi \wedge H\varphi] \rightarrow \psi]$$

$$(RG2) \quad \frac{f_{\alpha_1, \dots, \alpha_n}(\varphi_0, \dots, \varphi_{n-1}, [\varphi_n \wedge \neg p \wedge Hp]) \rightarrow \psi}{f_{\alpha_1, \dots, \alpha_n}(\varphi_0, \dots, \varphi_{n-1}, \varphi_n) \rightarrow \psi}$$

In RG2,  $p$  must be foreign to  $\psi$  and  $\varphi_0, \dots, \varphi_n$ .

I leave it to the reader to verify that the axiom and rules are Kamp valid.

It looks as if the ordering properties of linear frames that can be axiomatised in ordinary tense logic (endlessness towards the past, endlessness towards the future, density, etc.)<sup>21</sup> can be axiomatised against the background of Kamp frames, using Gabbay's techniques. But even so, there are still many simple questions that need to be settled; to take just two examples, it would be nice to know whether Kamp satisfaction is compact, and whether (RG2) is independent.

The situation with respect to treelike frames (i.e. with respect to Ockhamist validity) was even less well explored until recently. To begin with, a number of people have noticed that, although every propositional formula that is Kamp valid is Ockhamist valid,<sup>22</sup> the converse fails. Nishimura [1979b] points this out, giving (18) as a counterexample.<sup>23</sup>

$$(18) \quad \frac{GH\Box FP[H\neg p \wedge \neg p \wedge Gp] \rightarrow FP\Diamond FP[\neg p \wedge \Box Gp]}{}$$

<sup>21</sup>See Chapter 1 of this Volume.

<sup>22</sup>This follows from the fact that a Kamp model can be made from a treelike model, by making worlds out of its branches.

<sup>23</sup>Kamp [1979] ascribes a similar example to Burgess.

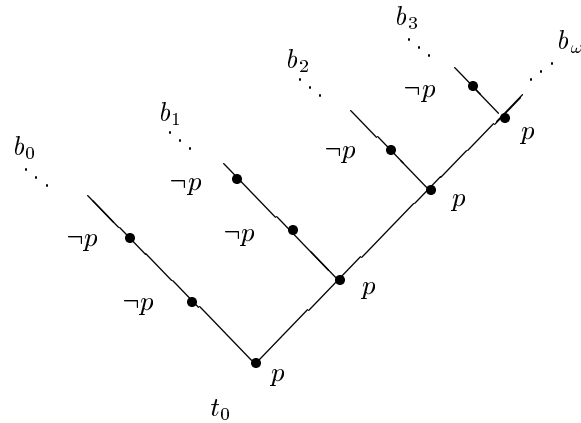


Figure 3.

In 1977, Burgess discovered (19), the simplest counterexample known to me.

$$(19) \quad \Box G \Diamond F p \rightarrow \Diamond G F p.$$

And in 1978, Thomason independently constructed the following counterexample.

$$(20) \quad [p \wedge \Box G H [p \rightarrow F p]] \rightarrow G F p$$

Both examples trade on the fact that any linearly-ordered subset of a tree can be extended to a branch (a consequence of the Axiom of Choice). If the antecedent of (20) is true at  $\langle t, b \rangle$  in a frame  $\langle T, < \rangle$  then there is a linear set  $X$  of moments such that  $p$  is true at every member of  $X$ , and such that there is no upper bound to  $T$  to  $X$ . This set  $X$  can be extended to a branch  $b^*$ , and  $G F p$  will be true at  $\langle t, b^* \rangle$ .

But the situation shown in Figure 3 can arise in Kamp models, allowing (19) to be falsified. If we make a Kamp model out of  $\{b_i : i \in \omega\}$ , omitting  $b_\omega$ , (19) is false at  $\langle t_0, b_0 \rangle$ .

Nishimura [1979a; 1979b] seems to feel that these examples show treelike frames to be inadequate. But the technical results only establish a difference between the treelike and the  $T \times W$  approaches. Adequacy has to do with intuitions about what should be valid. Intuitions may differ, but to me the natural notion is that of a possible future—not that of a possible course of events. Thus, (20) strikes me as clearly valid. The metric examples discussed in Nishimura [1979a] affect me similarly. The person who holds both (21) and (22) seems to me to have contradicted himself.



- (21) Inevitably, life on earth will come to an end at some date in the future.
- (22) For every date in the future, it is not inevitable that life on earth will have come to an end by that date.

For this reason, it seems to me that the  $T \times W$  frames do not have the philosophical interest of the treelike ones, though they are certainly interesting for technical reasons. This makes Ockhamist validity appear worth investigating; until recently, however, very little was known about it.<sup>24</sup> In [Burgess, 1979], it is claimed that Ockhamist validity is recursively axiomatisable, and a proof is sketched. Later (in conversation), Kripke challenged the proof and Burgess has been unable to substantiate all the details. Very recently, Gurevich and Shelah have proved a result implying that Ockhamist validity is decidable. (See [Gurevich and Shelah, 1985].) At present (October 1982) their paper is not yet written, and I have not had an opportunity to see their proof. The main result is that the theory of trees with second-order quantification over maximal chains is decidable.

Of course a proof of decidability would allow axioms to be recovered for Ockhamist validity; but this would be done in Craigian fashion. And unfortunately, Gabbay's completeness techniques do not seem (at first glance, anyway) to extend to the treelike case. Burgess' example [1979, p. 577], of an Ockhamist invalid formula valid in countable treelike frames, helps to bring home the complexity of the case.<sup>25</sup>

There are some interesting technical results regarding logics other than the treelike and  $T \times W$  ones that I have stressed here; the most important of these is Burgess' [1980] proof that the Peircian validities are decidable. Burgess has pointed out to me that the method of Section 5 of Burgess [1980] can be used to prove Kamp validity decidable, as well as  $T \times W$  validity with dense time. He also remarks that the most interesting technical questions about the case in which atomic formulas are treated like complex ones (so that  $p \rightarrow \Box p$  is not valid, and substitution is an admissible rule) are unresolved, and may prove more difficult.

## 5 DEONTIC LOGIC COMBINED WITH HISTORICAL NECESSITY

For a general discussion of deontic logic, with historical background, see Føllesdal and Hilpinen [1971] and Aqvist's chapter on Deontic Logic in Volume 8 of this *Handbook*. This presentation will concentrate on combinations

<sup>24</sup>In note 15 of [Thomason, 1970], Thomason says that he 'means to present an axiomatisation' of Ockhamist validity in a forthcoming paper. Since the paper has never appeared, this intention was evidently premature.

<sup>25</sup>I owe much of the information in this paragraph to Burgess and Gurevich.

of deontic modalities with temporal ones, and, in particular, with historical necessity.

Deontic logic seems to have suffered from a lack of communication. Even now, papers are written in which the relevant literature is not mentioned and the authors appear to be reinventing the wheel. In the hope that a survey of the literature will help to correct this situation, I have tried to make the bibliographical coverage of the present discussion thorough.

One facet of this lack of communication can be seen at work in the late 1960s. On the one hand, quite sophisticated model theoretic studies were developed during this time, treating deontic possibilities historically, as future alternatives. Montague [1968, pp. 116–117] and Scott [1967] represent the earliest such studies. (Unfortunately, Montague's presentation is tucked away in a rather forbidding technical paper that discusses many other topics, and the publication of the paper was delayed. And Scott's paper was never published.) But in 1969 Chellas [1969] appeared, giving an extended and very readable presentation of the California Theory.<sup>26</sup> (Montague's, Scott's and Chellas' theories are quite similar variations of the  $T \times W$  approach; the treatment of historical necessity is similar, and indeed identical in all important respects to Kamp's.)

But although Chellas' monograph contains an extensive and valuable bibliography of deontic logic, including many references to the literature in moral philosophy and practical reasoning, and though Chellas is evidently familiar with this literature, there is no attempt in the work to relate the theory to the more general philosophical issues, or even to discuss its application to the 'paradoxes' of deontic logic, which by then were well known. Chellas concentrates on the mathematical portion of the task. Thus although the presentation is less compressed than Montague's it remains relatively impenetrable to most moral philosophers, and there is no advertisement of the genuine help that the theory can give in dealing with these puzzles.

On the other hand, in the philosophical literature, it is easy to find studies that would have benefited from vigorous contact with logical theories such as Chellas'. To consider an example almost at random, take [Chisholm, 1974]. This paper has the word 'logic' in its title and deals with a topic that is thoroughly entangled with Chellas' investigation, but Chisholm's paper has no references to such logical work and seems entirely ignorant of it. Moreover, the paper is written in an axiomatic style that makes no use of semantical techniques that had been current in the logical literature for many years. And Chisholm commits errors that could have been avoided by awareness of these things. ([Chisholm, 1974, D9 on p. 13] is an example;

---

<sup>26</sup>Chellas' study is concerned with imperatives; but he starts with the assumption that these express obligations, so that the work belongs to deontic logic. Chellas [1969] is difficult to obtain; the theory is also presented in [Chellas, 1971], which is more accessible. McKinney [1975], a dissertation written under Chellas, is a later work belonging to this genre.

compare this with [Thomason, 1981a, pp. 183–184].)

To take another example, Wiggins [1973] provides an informal discussion, within the context of the determinist-libertarian debate in moral philosophy, of issues very similar to those treated by Chellas [1971] (and, so far as historical necessity goes in [Prior, 1967, Chapter 7]). Again, the paper contains no references to the logical literature. Though Wiggins' firm intuitive grasp of the issues prevents his argument from being affected,<sup>27</sup> it would have been nice to see him connect his account to the very relevant modal theoretic work.

There are a number of general discussions of the 'paradoxes' of deontic logic: see, for example, Åqvist [1967, pp. 364–373], Føllesdal and Hilpinen [1971, pp. 21–26], Hansson [1971, pp. 130–133], Al-Hibri [1978, pp. 22–29] and Van Eck [1981, pp. 28–35]. It should be apparent from the shudder quotes that I prefer not to dignify these puzzles with the same term that is applied to profoundly deep (perhaps unanswerable) questions like the Liar Paradox. The puzzles are a disparate assortment, and require a spectrum of solutions, some of which have little to do with tense. The Good Samaritan problem, for instance (the problem of reparational obligations), seems from one point of view to simply be a rediscovery of the frailties of the material conditional as a formalisation of natural language conditionals.

But part of the solution<sup>28</sup> of this problem seems to lie in the development of an *ought kinematics*,<sup>29</sup> in analogy to the probability kinematics that is the topic of [Jeffrey, 1990, Chapter 11]. As we would expect from probability, where there are interactions (surprisingly complex ones) between the rules of probability kinematics and locutions that combine conditionality and probability, we should expect there to be close relationships between ought kinematics and the semantics of conditional oughts. Nevertheless, we can formulate the kinematics of ought without having to work with conditional oughts in the object language.

---

<sup>27</sup>The one exception is Wiggins' definition of 'deterministic theory' and his assumption that macroscopic physical theories are deterministic. Here his argument could have been genuinely improved by familiarity with Montague [1962]. I haven't discussed this much-neglected paper here, because it belongs more to philosophy of science than to modal logic.

<sup>28</sup>Another part consists in developing an account of conditional oughts. Since this is a combination of the modalities rather than of tense and modality, I do not discuss it here. See De Cew [1981] for a recent survey of the topic.

<sup>29</sup>I have coined this term, since there seems to be a terminological niche for it. The sources are Jeffrey's 'probability kinematics' (I believe the term is his) and Greenspan's use of 'ought' as a substantive. I think there are good reasons for the terminology she recommends in [Greenspan, 1975] where she speaks of 'having an ought'. A theory like Chellas' cannot, of course, be deployed without containing an ought kinematics; but, as I said, the California writers did not advertise the theory as a solution of some of the deontic puzzles. Powers [1967] juxtaposes the two, though his pay-off machines present the model theory informally. Thomason [1981b] states the modal theory in terms of treelike frames, and discusses it in relation to some of the deontic puzzles.

The technical resolution of this problem is very simple, and this is precisely what the California theory that I referred to above provides: e.g. the theory of Chellas [1969]. for the sake of variety, I will formulate it with respect to treelike frames; this is something that, to my knowledge, was first done in Thomason [1981b].<sup>30</sup>

DEFINITION 12. A treelike frame  $\mathfrak{A}$  for deliberative deontic tense logic is a pair  $\langle T, <, O \rangle$ , where  $\langle T, < \rangle$  is a treelike frame for tense logic and  $O$  is a function on  $T$  such that for all  $t \in T$ , (1)  $O_t$  is nonempty, (2)  $O_t \subseteq B_t$ , and (3) if  $t < t'$  and  $t' \in b$  for some  $b \in O_t$ , then  $O_{t'} = O_t \cap B_{t'}$ .

The satisfaction clause for oughts is as follows.

$$\|O_\varphi\|_{\langle t, b \rangle}^h = 1 \text{ iff for all } b' \in O_t, \quad \|\varphi\|_{\langle t, b' \rangle}^h = 1.$$

We are dealing with the result of extending a propositional language of the sort we discussed in previous sections (truth- functional connectives, past and future tense, and historical  $\square$ ) by adding a one-place connective  $O$  for ought.<sup>31</sup>

Something needs to be said about Condition (3) on frames, which is not to be found in Thomason [1981b], and, as far as I know, has also been overlooked by other authors who (like Chellas) formulate the model theory of ought kinematics. This condition yields validities such as the following two.

$$(23) \quad OG[F\varphi \rightarrow \neg OG\neg\varphi]$$

$$(24) \quad OG\varphi \rightarrow OGO\varphi$$

These do strike me as valid in this context, and at any rate are interesting tense-deontic principles having to do with the coherence of plans. Principle (23), for instance, disallows an alternative future in  $O_t$  along which some outcome will happen, but is forbidden from ever happening. Such an alternative can't correspond to a coherent plan.

This set-up can readily deal with reparational obligations. Suppose that, because of a promise to my aunt, at 4:00 I ought to catch an airplane at 5:00, but that at 5:00 I have broken my promise because of the attractions of the airport bar. Then at 5:00 I should call my aunt to tell her I won't be on the plane. Though this is the sort of situation that is sometimes represented as paradoxical in the literature, it is easily modelled in ought kinematics, with no apparent conceptual strain. At one time we have  $O\neg Fp$  true, where  $p$  stands for 'I tell my aunt I won't be on the plane'. At a later time (one that involves the occurrence of something that shouldn't have happened)  $OFp$  is true.

<sup>30</sup>The theory of Thomason [1981b] is more general, but yields an account of deliberative ought as a special case.

<sup>31</sup>Since 'P' is already used for past tense, I will not use any special symbol for the dual of  $O$ .

If we press the account a bit harder, we can change the example; it also is true at 5:00 that I ought to inform my aunt I won't be on the plane, and this can be taken to entail that I ought not to be on the plane, since I can only inform someone of what is true, and because  $\Box[\varphi \rightarrow \psi] \rightarrow [O\varphi \rightarrow O\psi]$  is a validity of our logic.

The response to this pressure is, of course, a Gricean manoeuvre;<sup>32</sup> it is true at 5:00 (on the understanding of 'ought' in question) that I ought not to be on the plane. But it is not worth saying at 5:00, and if I were to say it then, I would be taken to have said something else, something false. And all of this can be made plausible in terms of general principles of reasonable conversation. I will not give the details here, since they can easily be reconstructed from [Lewis, 1979b].<sup>33</sup> On balance, the approach seems to be well braced against pressure from this direction. It is hard to imagine a reasonable theory of truth conditions that will not have to deploy Gricean tactics at some point. And this can be made to look like a very reasonable place to deploy them.

These linguistic reflections are nicely supported by quite independent philosophical considerations. Greenspan [1975] is a sustained study of ought kinematics from a philosophical standpoint, in which it is argued that a time-bound treatment of oughts is essential to an understanding of their logic, and that the proper view of deontic detachment is that a conditional ought licenses a 'consequent ought' when the antecedent is unalterably true. The paper contains many useful references to the philosophical literature, and provides a good example of the results that can be obtained by combining this philosophical material with the logical apparatus.

The fact that 'I ought to be on the plane' would ordinarily be taken to be false at 5:00 in the example we discussed above shows that 'ought' has employments that are not practical or deliberative: ones that perhaps have to do with wishful thinking. Also, there is its common use to express a kind of necessity: 'The butter is warm enough now; it ought to melt'.

Philosophers are inclined to speak of ambiguity in cases like this, but this is either a failure to appreciate the facts or an abuse of the word 'ambiguity'. The word 'ought' is indexical or context-sensitive, not ambiguous. The matter is argued, and some of its consequences are explored, in Kratzer [1977]; see [Lewis, 1979b] for a study of the consequences in a more general setting.

In Thomason [1981b] this is taken into account by a more general interpretation of  $O$ , according to which the relevant alternatives at  $t$  need not be possible futures for  $t$ . Probably the most general account would make the interpretation of  $O$  in a context relative to a set of alternatives which are regarded as possible in some sense relative to that context.

---

<sup>32</sup>With an added wrinkle, due to the contextual adjustability of oughts.

<sup>33</sup>See especially [Lewis, 1979b, pp. 354–355].

This is of course related to the philosophical debate over whether ‘ought’ implies ‘can’.<sup>34</sup> The most sophisticated linguistic account makes the issue appear rather boring: if we attend to a reasonable distinction between ambiguity and context-sensitivity, ‘ought’ doesn’t imply ‘can’, since there are contexts that provide counterexamples. But in practical contexts, when the one is true the other will be, even if for technical reasons we can’t relate this to an implication among linguistic forms.<sup>35</sup> This result is rather disappointing. But maybe there are ways of extracting interesting consequences for moral philosophy from a pragmatic account of oughts and other practical phenomena. An idea that I find intriguing is that manipulation of the context is the typical—perhaps the only—mechanism of moral weakness. The idea is suggested in [Thomason, 1981a], but is not much developed.

There seems to be no point in discussing the technical side of temporal deontic logic here. Not much work has been done in the area,<sup>36</sup> and the best strategy seems to be to let the matter wait until more is known about the interpretation of historical necessity.

Strictly speaking, conditional oughts are more closely related to the combination of conditionals with oughts than that of tense with modalities. Because the topic is complex and a thorough discussion of it would take up much space I have decided to neglect it in this article, even though (as Greenspan [1975] makes clear) tense enters into the matter. For an illuminating discussion of some of the problems, see DeCew [1981].<sup>37</sup>

## 6 CONDITIONAL LOGIC COMBINED WITH HISTORICAL NECESSITY

In the present essay, ‘modality’ has been confined to what can be interpreted using possible worlds semantics. So here, ‘conditional logic’ has to do with the modern theories that were introduced by Stalnaker and then by D. Lewis; see [Lewis, 1973].<sup>38</sup> For surveys of this work, see the chapter on

<sup>34</sup>See, for instance [Hare, 1963, Chapter 4].

<sup>35</sup>The fate of the validity of ‘I exist’ seems to be much the same. Maybe linguistics is a graveyard for some philosophical slogans.

<sup>36</sup>Some results are presented in [Åqvist and Hoepelman, 1981].

<sup>37</sup>I have not discussed Castañeda’s work here, though it may offer an alternative to the kinematic approach to some of the deontic puzzles; see for instance [Castañeda, 1981]. For one thing, the work falls outside the topic of this article; for another, Castañeda’s writings on deontic logic strike me as too confused and poorly presented to repay close study.

<sup>38</sup>Thus, for instance, I will not discuss [Slote, 1978], for although it deals with conditionals and the time it seeks to replace the possible worlds semantics with an analysis in the philosophical tradition. Another recent paper, [Kvart, 1980] uses techniques that are more model theoretic; but philosophical preconceptions have crept into the semantic theory at every point and uglified it. The usefulness of logical techniques in philosophy is largely dependent on the independence of the intuitions that guide the two disciplines; this enables them to reinforce one another in certain instances.

Conditional Logic by Nute and Cross in Volume 4 of this *Handbook*, and [Harper, 1981] and the volume in which it appears: Harper *et al.* [1981].

The interaction of conditionals and historical necessity is a topic that is only beginning to receive attention.<sup>39</sup> As in the case of deontic logic there is a fairly venerable philosophical tradition, involving issues that are still debated in the philosophical literature, and a certain amount of technical model theoretic work that may be relevant to these issues. But this time, the philosophical topic is causality.<sup>40</sup>

A conditional like D. Lewis', which does not satisfy the principle of conditional excluded middle,  $[\varphi > \psi] \vee [\varphi > \neg\psi]$ , is much more easy to relate to causal notions than Stalnaker's, since it is possible to say (with only a little hedging) that such a conditional is true in case there is a connection of determination of some sort between the antecedent and the consequent. But it has seemed less easy to reconcile the theory of such a conditional with simple tensed examples, like the case of Jim and Jack, invented much earlier by Downing [1959].

Jim and Jack quarrelled yesterday; Jack is unforgiving, and Jim is proud. The example is this.

- (25) If Jim were to ask Jack for help today,  
       Jack would help him.

Most authors feel that (25) could be taken in two ways. It could be taken to be false, because they quarrelled and Jack is so unforgiving. It could be taken to be true, because Jim is so proud that if he were to ask for help they would not have quarrelled yesterday. But the preferred understanding of (25) seems to be the first of these; and this is only one way in which a systematic preference for alternatives that involve only small changes in the past<sup>41</sup> seems to affect our habits of evaluating such conditionals. An examination of such preferences and their influence on the sort of similarity that is involved in interpreting conditions can be found in [Lewis, 1979a]. Further information can be found in [Bennett, 1982; Thomason, 1982].

This, of course, relates conditionals to time in a philosophical way. But Lewis' informal way of attacking the problem assumes that the logic has

---

<sup>39</sup>Judging from some unpublished manuscripts that I have recently received, it is likely to receive more before long. But these are working drafts, which I should not discuss here.

<sup>40</sup>See [Sosa, 1975] for a collection juxtaposing some recent papers on causality with ones on conditionals. (Many of the papers seek to establish links between the two.) Also see [Downing, 1959; Jackson, 1977].

<sup>41</sup>Though I put it this way D. Lewis (who assumes determinism for the sake of argument in [Lewis, 1979a], so that changes in the future must be accompanied by changes in the past) has to resort to a hierarchy of maxims to achieve a like effect. (In particular, Maxim 1, [Lewis, 1979a, p. 472], enjoins us to avoid wholesale violations of law, and Maxim 2 tells us to seek widespread perfect match of particular fact; together, these have much the same effect as the principle I stated in the text.)

to come to an end, and in particular that there are no new validities to be discovered by placing conditionals in a temporal setting.<sup>42</sup> In view of the lessons we have learned about the combination of tense with other modalities, this may be a methodological oversight; it runs the risk of not getting the most out of the possible worlds semantics that is to be put to philosophical use.<sup>43</sup>

Thomason and Gupta [1981] and Van Fraassen [1981] are two recent studies that pursue this model theoretic route: the former uses treelike frames and the latter a version of the  $T \times W$  approach. The two treatments are very similar in essentials, though a decision about how to secure the validity of (27) leads to much technical complexity in [Thomason and Gupta, 1981]. Since Van Fraassen's exposition is so clear, and I fear I would be repeating myself if I attempted a detailed account of [Van Fraassen, 1981], I will be very brief.

Both papers endorse certain validities involving a mixture of tense, historical necessity, and the conditional. The following examples are representative.

$$(26) \quad [\Diamond\varphi \wedge [\varphi > \psi]] \rightarrow \Diamond\psi$$

$$(27) \quad [\Box\neg\varphi \wedge \Box[\varphi > \psi]] \rightarrow [\varphi \rightarrow \Box[\varphi \rightarrow \psi]]$$

The first of these represents one way in which selection principles for  $>$  can be formulated in terms of alternative histories; the validity of (26) corresponds to a preference at  $\langle t, w \rangle$  for alternatives that are possible futures for  $\langle t, w \rangle$ .

Example (27) called the Edelberg inference after Walter Edelberg, who first noticed it, represents a principle of 'conditional transmission of settledness' that can be made quite plausible; see the discussion in Thomason and Gupta [1981, pp. 306–307]. Formally, its validity depends on there being no 'unattached' counterfactual futures—ones that are not picked out as counterfactual alternatives on condition  $\varphi$  with respect to actual futures. At least, this is the way that Van Fraassen secures the validity of (28); Thomason and Gupta do it in a circuitous way, at the cost of making their theory much more difficult to explain.

If there advantages to offset this cost, they have to do with causality. The key notion introduced in [Thomason and Gupta, 1981], which is not

<sup>42</sup>I mean that in [Lewis, 1979a], Lewis phrases the discussion in terms of 'similarity'. This is the intuitive notion used to explicate the technical gadgets (assignments of sets of possible worlds to each world) that yield Lewis' theory in [Lewis, 1973] of the satisfaction conditions for the conditional. In [Lewis, 1979a], he leaves things at this informal level, and doesn't try to build temporal conditional frames which can be used to define satisfaction for an extended language.

<sup>43</sup>If Lewis' adoption of a determinist position in [Lewis, 1979a] is not for the sake of argument, there may be a philosophical issue at work here. A philosophical determinist would be much more likely to follow an approach like Lewis' than to base conditional logic on the logic of historical necessity.



needed by Van Fraassen,<sup>44</sup> is that of a *future choice function*: a function  $\mathcal{F}$  that for each moment  $t$  in a treelike frame chooses a branch in  $B_t$ . future choice functions must choose coherently, so that if  $t' \in \mathcal{F}_t$  then  $\mathcal{F}_t = \mathcal{F}_{t'}$ . By considering restricted sets of choice functions, Thomason and Gupta are able to introduce a modal notion that, they claim, may help to explicate causal independence. If this claim could be made good it would be worth the added complexity, since so far the techniques of possible worlds semantics have not been of much direct help in clarifying the philosophical debate about causality. But in [Thomason and Gupta, 1981], the idea is not developed enough to see very clearly what the prospects of success are.

*University of Michigan, USA.*

#### EDITORIAL NOTE

The present chapter is reproduced from the first edition of the Handbook. A continuation chapter will appear in a later volume of the present, second edition. The logic of historical necessity is technically a special combination of modality and temporal operators. Combinations with temporal logic, (or *temporalising*) are discussed in chapter 2 of this volume. Branching temporal logics with modalities in the spirit of the logics of this chapter have been very successfully introduced in theoretical computer science. These are the CTL (computation tree logic) family of logics. For a survey, see the chapter by Colin Stirling in Volume 2 of the *Handbook of Logic in Computer Science*, S. Abramsky, D. Gabbay and T. Maibaum, editors, pp. 478–551. Oxford University Press, 1992.

The following two sources are also of interest:

1. E. Clarke, Jr., O. Grumberg and D. Peled. *Model Checking*, MIT Press, 2000.
2. E. Clarke and B.-H. Schlingloff. Model checking. In *Handbook of Automated Reasoning*, A. Robinson and a. Voronkov, eds. pp. 1635 and 1790. Elsevier and MIT Press, 2001.

#### BIBLIOGRAPHY

- [Al-Hibri, 1978] A. Al-Hibri. *Deontic Logic*. Washington, DC, 1978.  
 [Åqvist and Hoepelman, 1981] L. Åqvist and J. Hoepelman. Some theorems about a 'tree' system of deontic tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 187–221. Reidel, Dordrecht, 1981.

---

<sup>44</sup>And which also would not be required on a strict theory of the conditional, such as D. Lewis'. It is important to note in the present context that Thomason and Gupta, as well as Van Fraassen, are working with Stalnaker's theory.

- [Åqvist, 1967] L. Åqvist. Good samaritans, contrary-to-duty imperatives, and epistemic obligations. *Nôus*, 1:361–379, 1967.
- [Bennett, 1982] J. Bennett. Counterfactuals and temporal direction. Technical report, Xerox, Syracuse University, 1982.
- [Burgess, 1979] J. Burgess. Logic and time. *Journal of Symbolic Logic*, 44:566–582, 1979.
- [Burgess, 1980] J. Burgess. Decidability and branching time. *Studia Logica*, 39:203–218, 1980.
- [Carnap, 1956] R. Carnap. *Meaning and Necessity*. Chicago University Press, 2nd edition, 1956.
- [Castañeda, 1981] H.-N. Castañeda. The paradoxes of deontic logic: the simplest solution to all of them in one fell swoop. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 37–85. Reidel, Dordrecht, 1981.
- [Chellas, 1969] B. Chellas. *The Logical Form of Imperatives*. Perry Lane Press, Stanford, 1969.
- [Chellas, 1971] B. Chellas. Imperatives. *Theoria*, 37:114–129, 1971.
- [Chisholm, 1974] R. Chisholm. Practical reason and the logic of requirement. In S. Körner, editor, *Practical Reason*, pages 1–16. New Haven, 1974.
- [Code, 1976] A. Code. Aristotle's response to quine's objections to modal logic. *Journal of Philosophical Logic*, 5:159–186, 1976.
- [DeCew, 1981] J. DeCew. Conditional obligation and counterfactuals. *Journal of Philosophical Logic*, 10:55–72, 1981.
- [DeWitt and Graham, 1973] B. DeWitt and N. Graham. *The Many-Worlds Interpretation of Quantum Mechanics*. Princeton, 1973.
- [DeWitt, 1973] B. DeWitt. Quantum mechanics and reality. In B. DeWitt and N. Graham, editors, *The Many-Worlds Interpretation of Quantum Mechanics*, pages 155–165. Princeton, 1973.
- [Downing, 1959] P. Downing. Subjunctive conditionals, time order, and causation. *Proc Aristotelian Society*, 59:125–140, 1959.
- [Edwards, 1957] J. Edwards. *Freedom of the Will*. Yale University Press, New Haven, 1957. First published in Boston, 1754.
- [Fine, 1982] A. Fine. Joint distributions, quantum correlations, and commuting observables. *Journal of Mathematical Physics*, 23:1306–1310, 1982.
- [Føllesdal and Hilpinen, 1971] D. Føllesdal and R. Hilpinen. Deontic logic: an introduction. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 1–35. Reidel, Dordrecht, 1971.
- [Frede, 1970] D. Frede. *Aristoteles und die 'Seeschlacht'*. Göttingen, 1970.
- [Gabbay, 1981] D. M. Gabbay. An irreflexivity lemma with applications to axiomatisations of conditions on tense frames. In U. Mönnich, editor, *Aspects of Philosophical Logic*, pages 67–89. Reidel, Dordrecht, 1981.
- [Greenspan, 1975] P. Greenspan. Conditional oughts and hypothetical imperatives. *Journal of Philosophy*, 72:259–276, 1975.
- [Grice and Strawson, 1956] P. Grice and P. Strawson. In defense of a dogma. *Philosophical Review*, 65:141–158, 1956.
- [Gurevich and Shelah, 1985] Y. Gurevich and S. Shelah. The decision problem for branching time logic. *Journal of Symbolic Logic*, 50:668–681, 1985.
- [Hansson, 1971] B. Hansson. An analysis of some deontic logics. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 121–147. Reidel, Dordrecht, 1971.
- [Hare, 1963] R. Hare. *Freedom and Reason*. Oxford University Press, Oxford, 1963.
- [Harper, 1975] W. Harper. Rational belief change, Popper functions and counterfactuals. *Synthese*, 30:221–262, 1975.
- [Harper, 1981] W. Harper. A sketch of some recent developments in the theory of conditionals. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance, and Time*, pages 3–38. Reidel, Dordrecht, 1981.
- [Hintikka, 1969] J. Hintikka. Deontic logic and its philosophical morals. In J. Hintikka, editor, *Models for Modalities*, pages 184–214. Reidel, Dordrecht, 1969.

- [Hintikka, 1971] J. Hintikka. Some main problems of deontic logic. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, pages 59–104. Reidel, Dordrecht, 1971.
- [Howard, 1971a] R. Howard. *Dynamic Probabilistic Systems, Volume I: Markov Models*. New York, 1971.
- [Howard, 1971b] R. Howard. *Dynamic Probabilistic systems, Volume II: Semi-Markov and Decision Processes*. New York, 1971.
- [Jackson, 1977] F. Jackson. A causal theory of counterfactuals. *Australasian Journal of Philosophy*, 55:3–21, 1977.
- [Jeffrey, 1990] R. Jeffrey. *The Logic of Decision*, 2nd edition. University of Chicago Press, 1990.
- [Jeffrey, 1979] R. Jeffrey. Coming true. In C. Diamond and J. Teichman, editors, *Intention and Intentionality*, pages 251–260. Ithaca, NY, 1979.
- [Kamp, 1979] H. Kamp. The logic of historical necessity, part i. Unpublished typescript, 1979.
- [Kaplan, 1978] D. Kaplan. On the logic of demonstratives. *Journal of Philosophical Logic*, 8:81–98, 1978.
- [Kratzer, 1977] A. Kratzer. What ‘must’ and ‘can’ must and can mean. *Linguistics and Philosophy*, 1:337–355, 1977.
- [Kripke, 1982] S. Kripke. Naming and necessity. Harvard University Press, 1982.
- [Kvart, 1980] I. Kvart. Formal semantics for temporal logic and counterfactuals. *Logique et analyse*, 23:35–62, 1980.
- [Lehrer and Taylor, 1965] K. Lehrer and R. Taylor. time, truth and modalities. *Mind*, 74:390–398, 1965.
- [Lemmon and Scott, 1977] E. J. Lemmon and D. S. Scott. *An Introduction to Modal Logic: the Lemmon Notes*. Blackwell, 1977.
- [Lewis, 1918] C. I. Lewis. *A Survey of Symbolic Logic*. Univeristy of California Press, Berkeley, 1918.
- [Lewis, 1970] D. Lewis. Anselm and acutality. *Noûs*, 4:175–188, 1970.
- [Lewis, 1973] D. Lewis. *Counterfactuals*. Oxford University Press, Oxford, 1973.
- [Lewis, 1979a] D. Lewis. Counterfactual dependence and tim’es arrow. *Noûs*, 13:455–476, 1979.
- [Lewis, 1979b] D. Lewis. Scorekeeping in a language game. *Journal of Philosophical Logic*, 8:339–359, 1979.
- [Lewis, 1981] D. Lewis. A subjectivist’s guide to objective chance. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance and Time*, pages 259–265. Reidel, Dordrecht, 1981.
- [Lukasiewicz, 1967] J. Lukasiewicz. On determinism. In S. McCall, editor, *Polish Loigc*, pages 19–39. Oxford University Press, Oxford, 1967.
- [McKinney, 1975] A. McKinney. *Conditional obligation and temporally dependent necessity: a study in conditional deontic logic*. PhD thesis, University of Pennsylvania, 1975.
- [Montague, 1962] R. Montague. Deterministic theories. In *Decisions, Values and Groups 2*, pages 325–370. Pergamon Press, Oxford, 1962.
- [Montague, 1968] R. Montague. Pragmatics. In R. Klibansky, editor, *Contemporary Philosophy: A Survey*, pages 101–122. Florence, 1968.
- [Nishimura, 1979a] H. Nishimura. Is the semantics of branching structures adequate for chronological modal logics? *Journal of Philosophical Logic*, 8:469–475, 1979.
- [Nishimura, 1979b] H. Nishimura. Is the semantics of branching structures adequate for non-metric ochamist tense logics? *Journal of Philosophical Logic*, 8:477–478, 1979.
- [Normore, 1982] C. Normore. Future contingents. In N. Kretzman *et al.*, editor, *Cambridge History of Later Medieval Philosophy*, pages 358–381. Cambridge University Press, Cambridge, 1982.
- [Powers, 1967] L. Powers. Some deontic logicians. *Noûs*, 1:381–400, 1967.
- [Prior, 1957] A. Prior. *Time and Modality*. Oxford University Press, Oxford, 1957.
- [Prior, 1967] A. Prior. *Past, Present and Future*. Oxford University Press, Oxford, 1967.
- [Schlipp, 1963] P. Schlipp, editor. *The Philosophy of Rudolf Carnap*. Open Court, LaSalle, 1963.

- [Scott, 1967] D. Scott. A logic of commands. mimeograph, Stanford University, 1967.
- [Seeskin, 1971] K. Seeskin. Many-valued logic and future contingencies. *Logique et analyse*, 14:759–773, 1971.
- [Slote, 1978] M. Slote. Time and counterfactuals. *Philosophical Review*, 87:3–27, 1978.
- [Sorabji, 1980] R. Sorabji. *Necessity, Cause and Blame: Perspectives on Aristotle's Theory*. Cornell University Press, Ithaca, NY, 1980.
- [Sosa, 1975] E. Sosa, editor. *Causation and Conditionals*. Oxford University Press, Oxford, 1975.
- [Thomason and Gupta, 1981] R. Thomason and A. Gupta. A theory of conditionals in the context of branching time. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance and Time*, pages 299–322. Reidel, Dordrecht, 1981.
- [Thomason, 1970] R. Thomason. Indeterministic time and truth-value gaps. *Theoria*, 36:264–281, 1970.
- [Thomason, 1981a] R. Thomason. Deontic logic and the role of freedom in moral deliberation. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 177–186. Reidel, Dordrecht, 1981.
- [Thomason, 1981b] R. Thomason. Deontic logic as founded on tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, pages 177–186. Reidel, Dordrecht, 1981.
- [Thomason, 1981c] R. Thomason. Notes on completeness problems with historical necessity. Xerox, 1981.
- [Thomason, 1982] R. Thomason. Counterfactuals and temporal direction. Xerox, University of Pittsburgh, 1982.
- [Van Eck, 1981] J. Van Eck. *A system of temporally relative modal and deontic predicate logic and its philosophical applications*. PhD thesis, Rujksuniversiteit de Groningen, 1981.
- [Van Fraassen, 1966] B. Van Fraassen. Singular terms, truth-value gaps and free logic. *Journal of Philosophy*, 63:481–495, 1966.
- [Van Fraassen, 1971] B. Van Fraassen. *Formal Semantics and Logic*. Macmillan, New York, 1971.
- [Van Fraassen, 1981] B. Van Fraassen. A temporal framework for conditionals and chance. In W. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs: Conditionals, Belief, Decision, Chance and Time*, pages 323–340. Reidel, Dordrecht, 1981.
- [Von Wright, 1968] G. Von Wright. An essay in deontic logic. *Acta Philosophica Fennica*, 21:1–110, 1968.
- [Werner, 1974] B. Werner. *Foundations of temporal modal logic*. PhD thesis, University of Wisconsin at Madison, 1974.
- [Wiggins, 1973] D. Wiggins. Towards a reasonable libertarianism. In T. Honderich, editor, *Essays on Freedom of Action*, pages 33–61. London, 1973.
- [Wigner, 1971] E. Wigner. Quantum-mechanical distribution functions revisited. In W. Yourgraw and A. van der Merwe, editors, *Perspectives in Quantum Theory*, pages 25–36. MIT, Cambridge, MA, 1971.
- [Woolhouse, 1973] R. Woolhouse. Tensed modalities. *Journal of Philosophical Logic*, 2:393–415, 1973.

NINO B. COCCHIARELLA

# PHILOSOPHICAL PERSPECTIVES ON QUANTIFICATION IN TENSE AND MODAL LOGIC

## INTRODUCTION

The trouble with modal logic, according to its critics, is quantification into modal contexts—i.e. *de re* modality. For on the basis of such quantification, it is claimed, essentialism ensues, and perhaps a bloated universe of *possibilia* as well. The essentialism is avoidable, these critics will agree, but only by turning to a Platonic realm of individual concepts whose existence is no less dubious or problematic than mere *possibilia*. Moreover, basing one's semantics on individual concepts, it is claimed, would in effect render all identity statements containing only proper names either necessarily true or necessarily false— i.e. there would then be no contingent identity statements containing only proper names.

None of these claims is true quite as it stands, however; and in what follows we shall attempt to separate the chaff from the grain by examining the semantics of (first-order) quantified modal logic in the context of different philosophical theories. Beginning with the primary semantics of logical necessity and the philosophical context of logical atomism, for example, we will see that essentialism not only does not ensue but is actually rejected in that context by the validation of the modal thesis of anti-essentialism, and that in consequence all *de re* modalities are reducible to *de dicto* modalities.

Opposed to logical atomism, but on a par with it in its referential interpretation of quantifiers and proper names, is Kripke's semantics for what he properly calls metaphysical necessity. Unlike the primary semantics of logical necessity, in other words, Kripke's semantics for metaphysical necessity is in direct conflict with some of the basic assumptions of logical atomism; and in the form which that conflict takes, which we shall refer to here as the form of a *secondary* semantics for necessity, Kripke's semantics amounts to the initial step toward a proper formulation of Aristotelian essentialism. (A secondary semantics for necessity stands to the primary semantics in essentially the same way that non-standard models for second-order logic stand to standard models.) The problem with this initial step toward Aristotelian essentialism, however, is the problem of all secondary semantics; viz. that of its objective, as opposed to its merely formal, significance—a problem which applies all the more so to Kripke's deepening of his formal semantics by the introduction of an accessibility relation between possible worlds. This, in fact, is the real problem of essentialism.

There are no individual concepts, it will be noted in what follows, in either logical atomism or Kripke's implicit philosophical semantics, and yet in both contexts proper names are rigid designators; that is, in both there can be no contingent identity statements containing only proper names. One need not, accordingly, turn to a Platonic realm of individual concepts in order to achieve this result. Indeed, quite the opposite is the case. That is, it has in fact been for the defence of contingent identity, and not its rejection, that philosophical logicians have turned to a Platonic realm of individual concepts, since, on this view, it is only through the mere coincidence of the denotations of the individual concepts expressed by proper names that an identity statement containing those names can be contingent. Moreover, unless such a Platonic realm is taken as the intensional counterpart of logical atomism (a marriage of dubious coherence), it will not validate the modal thesis of anti-essentialism. That is, one can in fact base a Platonic or logical essentialism—which is not the same thing at all as Aristotelian essentialism—upon such a realm. However, under suitable assumptions, essentialism can also be avoided in such a realm; or rather it can in the weaker sense in which, given these assumptions, all *de re* modalities are reducible to *de dicto* modalities.

Besides the Platonic view of intensionality, on the other hand, there is also a socio-biologically based conceptualist view according to which concepts are not independently existing Platonic forms but cognitive capacities or related structures of the human mind whose realisation in thought is what informs a mental act with a predicable or referential nature. This view, it will be seen, provides an account in which there can be contingent identity statements, but not such as to depend on the coincidence of individual concepts in the platonic sense. Such a conceptualist view will also provide a philosophical foundation for quantified tense logic and paradigmatic analyses thereby of metaphysical modalities in terms of time and causation. The problem of the objective significance of the secondary semantics for the analysed modalities, in other words, is completely resolved on the basis of the nature of time, local or cosmic. The related problem of a possible ontological commitment to *possibilia*, moreover, is in that case only the problem of how conceptualism can account for direct references to past or future objects.

## 1 THE PRIMARY SEMANTICS OF LOGICAL NECESSITY

We begin by describing what we take to be the primary semantics of logical necessity. Our terminology will proceed as a natural extension of the syntax and semantics of standard first-order logic with identity. Initially, we shall assume that the only singular terms are individual variables. As primitive *logical constants* we take  $\rightarrow$ ,  $\neg$ ,  $\forall$ ,  $=$ , and  $\square$  for the material conditional sign, the negation sign, the universal quantifier, the identity sign and the

necessity sign, respectively. (The conjunction, disjunction, biconditional, existential quantifier and possibility signs— $\wedge, \vee, \leftrightarrow, \exists$  and  $\diamond$ , respectively—are understood to be defined in the usual way as metalinguistic abbreviatory devices.) The only non-logical or *descriptive constants* at this point are predicates of arbitrary (finite) degree. We call a set of such predicates a *language* and understand the well-formed formulas (*wffs*) of a language to be defined in the usual way.

A *model*  $\mathfrak{A}$  indexed by a language  $L$ , or for brevity, an *L-model*, is a structure of the form  $\langle D, R \rangle$ , where  $D$ , the *universe* of the model, is a non-empty set and  $R$  is a function with  $L$  as domain and such that for each positive integer  $n$  and each  $n$ -place predicate  $F^n$  in  $L$ ,  $R(F^n) \subseteq D^n$ , i.e.  $R(F^n)$  is a set of  $n$ -tuples of members of  $D$ . An *assignment in  $D$*  is a function  $A$  with the set of individual variables as domain and such that  $A(x) \in D$ , for each variable  $x$ . Where  $d \in D$ , we understand  $A(d/x)$  to be that assignment in  $D$  which is exactly like  $A$  except for its assigning  $d$  to  $x$ . The *satisfaction* of a wff  $\varphi$  of  $L$  in  $\mathfrak{A}$  by an assignment  $A$  in  $D$ , in symbols  $\mathfrak{A}, A \vDash \varphi$ , is recursively defined as follows:

1.  $\mathfrak{A}, A \vDash (x = y)$  iff  $A(x) = A(y)$ ;
2.  $\mathfrak{A}, A \vDash P^n(x_1, \dots, x_n)$  iff  $\langle A(x_1), \dots, A(x_n) \rangle \in R(P^n)$ ;
3.  $\mathfrak{A}, A \vDash \neg\varphi$  iff  $\mathfrak{A}, A \not\vDash \varphi$ ;
4.  $\mathfrak{A}, A \vDash (\varphi \rightarrow \psi)$  iff either  $\mathfrak{A}, A \not\vDash \varphi$  or  $\mathfrak{A}, A \vDash \psi$ ;
5.  $\mathfrak{A}, A \vDash \forall x\varphi$  iff for all  $d \in D$ ,  $\mathfrak{A}, A(d/x) \vDash \varphi$ ; and
6.  $\mathfrak{A}, A \vDash \Box\varphi$  iff for all  $R'$ , if  $\langle D, R' \rangle$  is an  $L$ -model, then  $\langle D, R' \rangle, A \vDash \varphi$ .

The *truth* of a wff in a model (indexed by a language suitable to that wff) is as usual the satisfaction of the wff by every assignment in the universe of the model. *Logical truth* is then truth in every model (indexed by any appropriate language). One or another version of this primary semantics for logical necessity, it should be noted, occurs in [Carnap, 1946]; [Kanger, 1957]; [Beth, 1960] and [Montague, 1960].

## 2 LOGICAL ATOMISM AND QUANTIFIED MODAL LOGIC

These definitions, as already indicated, are extensions of essentially the same semantical concepts as defined for the modal free wffs of standard first-order predicate logic with identity. The clause for the necessity operator has a particularly natural motivation within the framework of logical atomism. In such a framework, a model  $\langle D, R \rangle$  for a language  $L$  represents a *possible world* of a *logical space* based upon (1)  $D$  as the universe of objects of that space and (2)  $L$  as the predicates characterising the *atomic states of affairs*

of that space. So based, in other words, a logical space consists of the totality of atomic states of affairs all the constituents of which are in  $D$  and the characterising predicates of which are in  $L$ . A possible world of such a logical space then amounts in effect to a partitioning of the atomic states of affairs of that space into two cells: those that obtain in the world in question and those that do not.

Every model, it is clear, determines both a unique logical space (since it specifies both a domain and a language) and a possible world of that space. In this regard, the clause for the necessity operator in the above definition of satisfaction is the natural extension of the standard definition and interprets that operator as ranging over *all* the possible worlds (models) of the logical space to which the given one belongs.

Now it may be objected that logical atomism is an inappropriate framework upon which to base a system of quantified modal logic; for if any framework is a paradigm of anti-essentialism, it is logical atomism. The objection is void, however, since in fact the above semantics provides the clearest validation of the modal thesis of anti-essentialism. Quantified modal logic, in other words, does not in itself commit one to any non-trivial form of essentialism (cf. [Parsons, 1969]).

The general idea of the modal thesis of anti-essentialism is that if a predicate expression or open wff  $\varphi$  *can be true of* some individuals in a given universe (satisfying a given identity-difference condition with respect to the variables free in  $\varphi$ ), then  $\varphi$  *can be true of* any individuals in that universe (satisfying the same identity-difference conditions). In other words, no conditions are essential to some individuals that are not essential to all, which is as it should be if necessity means logical necessity.

The restriction to identity-difference conditions mentioned (parenthetically) above can be dropped, it should be noted, if nested quantifiers are interpreted exclusively and not (as we have done) inclusively where, e.g. it is allowed that the value of  $y$  in  $\forall x\exists y\varphi(x, y)$  can be the same as the value of  $x$ . (Cf. [Hintikka, 1956] for a development of the exclusive interpretation.) Indeed, as Hintikka has shown, when nested quantifiers are interpreted exclusively, identity and difference wffs are superfluous—which is especially apropos of logical atomism where an identity wff does not represent an atomic state of affairs. (Cf. Wittgenstein's *Tractatus Logico-Philosophicus* 5.532–5.53 and [Cocchiarella, 1975a, Section V].)

Retaining the inclusive interpretation and identity as primitive, however, an *identity-difference condition* for distinct individual variables  $x_1, \dots, x_n$  is a conjunction of one each but not both of the wffs  $(x_i = x_j)$  or  $(x_i \neq x_j)$ , for all  $i, j$  such that  $1 \leq i < j \leq n$ . It is clear of course that such a conjunction specifies a complete identity-difference condition for the variables  $x_1, \dots, x_n$ . Since there are only a finite number of non-equivalent such conditions for  $x_1, \dots, x_n$ , moreover, we understand  $ID_j(x_1, \dots, x_n)$ , relative to an assumed ordering of such non-equivalent conjunctions, to be the  $j$ th



conjunction in the ordering. *The modal thesis of anti-essentialism* may now be stated as the thesis that every wff of the form

$$\begin{aligned} & \exists x_1 \dots \exists x_n (\text{ID}_j(x_1, \dots, x_n) \wedge \diamond\varphi) \\ & \rightarrow \forall x_1 \dots \forall x_n (\text{ID}_j(x_1, \dots, x_n) \rightarrow \diamond\varphi) \end{aligned}$$

is to be logically true, where  $x_1, \dots, x_n$  are all the distinct individual variables occurring free in  $\varphi$ . (Where  $n = 0$ , the above wff is understood to be just  $(\diamond\varphi \rightarrow \diamond\varphi)$ ; and where  $n = 1$ , it is understood to be just  $\exists x \diamond\varphi \rightarrow \forall x \diamond\varphi$ .) The validation of the thesis in our present semantics is easily seen to be a consequence of the following lemma (whose proof is by a simple induction on the wffs of L).

LEMMA If L is a language,  $\mathfrak{A}, \mathfrak{B}$  are L-models, and  $h$  is an isomorphism of  $\mathfrak{A}$  with  $\mathfrak{B}$ , then for all wffs  $\varphi$  of L and all assignments  $A$  in the universe of  $\mathfrak{A}$ ,  $\mathfrak{A}, A \models \varphi$  iff  $\mathfrak{B}, A/h \models \varphi$ .

One of the nice consequences of the modal thesis of anti-essentialism in the present semantics, it should be noted, is the reduction of all *de re* wffs to *de dicto* wffs. (A *de re* wff is one in which some individual variable has a free occurrence in a subwff of the form  $\Box\psi$ . A *de dicto* wff is a wff that is not *de re*.) Naturally, such a consequence is a further sign that all is well with our association of the present semantics with logical atomism.

THEOREM (De Re Elimination Theorem) For each *de re* wff  $\varphi$ , there is a *de dicto* wff  $\psi$  such that  $(\varphi \leftrightarrow \psi)$  is logically true.<sup>1</sup>

These niceties aside, however, another result of the present semantics is its essential incompleteness with respect to any language containing at least one relational predicate. (It is not only complete but even decidable when restricted to monadic wffs—of which more anon.) The incompleteness is easily seen to follow from the following lemma and the well-known fact that the modal free non-logical truths of a language containing at least one relational predicate is not recursively enumerable (cf. [Cocchiarella, 1975b]). (It is also for the statement of the infinity condition of this lemma that a relational predicate is needed.)

LEMMA If  $\psi$  is a sentence which is satisfiable, but only in an infinite model, and  $\varphi$  is a modal and identity-free sentence, then  $(\psi \rightarrow \neg\Box\varphi)$  is logically true iff  $\varphi$  is not logically true.

<sup>1</sup>A proof of this theorem can be found in [McKay, 1975]. Briefly, where  $x_1, \dots, x_n$  are all the distinct individual variables occurring free in  $\varphi$  and  $ID_1(x_1, \dots, x_n), \dots, ID_k(x_1, \dots, x_n)$  are all the non-equivalent identity-difference conditions for  $x_1, \dots, x_n$ , then the equivalence in question can be shown if  $\psi$  is obtained from  $\varphi$  by replacing each subwff  $\Box\chi$  of  $\varphi$  by:

$$\begin{aligned} & [ID_1(x_1, \dots, x_n) \wedge \Box\forall x_1 \dots \forall x_n (ID_1(x_1, \dots, x_n) \rightarrow \chi)] \vee \dots \\ & \vee [ID_k(x_1, \dots, x_n) \wedge \Box\forall x_1 \dots \forall x_n (ID_k(x_1, \dots, x_n) \rightarrow \chi)]. \end{aligned}$$

**THEOREM** If  $L$  is a language containing at least one relational predicate, then the set of wffs of  $L$  that are logically true is not recursively enumerable.

This last result does not affect the association we have made of the primary semantics with logical atomism. Indeed, given the Löwenheim–Skolem theorem, what this lemma shows is that there is a complete concurrence between logical necessity as an internal condition of modal free propositions (or of their corresponding states of affairs) and logical truth as a semantical condition of the modal free sentences expressing those propositions (or representing their corresponding states of affairs). And that of course is as it should be if the operator for logical necessity is to have only formal and no material content.

Finally, it should be noted that the above incompleteness theorem explains why Carnap was not able to prove the completeness of the system of quantified modal logic formulated in [Carnap, 1946]. For on the assumption that the number of objects in the universe is denumerably infinite, Carnap's state description semantics is essentially that of the primary semantics restricted to denumerably infinite models; and, of course, precisely because the models are denumerably infinite, the above incompleteness theorem applies to Carnap's formulation as well. Thus, the reason why Carnap was unable to carry through his proof of completeness is finally answered.

### 3 THE SECONDARY SEMANTICS OF METAPHYSICAL NECESSITY

Like the situation in standard second-order logic, the incompleteness of the primary semantics can be avoided by allowing the quantificational interpretation of necessity in the metalanguage to refer not to *all* the possible worlds (models) of a given logical space but only to those in a given non-empty set of such worlds. Of course, since a model may belong to many such sets, the relativisation to the one in question must be included as part of the definition of satisfaction.

Accordingly, where  $L$  is a language and  $D$  is a non-empty set, we understand a *model structure* based on  $D$  and  $L$  to be a pair  $\langle \mathfrak{A}, K \rangle$ , where  $K$  is a set of  $L$ -models all having  $D$  as their universe and  $\mathfrak{A} \in K$ . The *satisfaction* of a wff  $\varphi$  of  $L$  in such a model structure by an assignment  $A$  in  $D$ , in symbols  $\langle \mathfrak{A}, K \rangle, A \models \varphi$ , is recursively defined exactly as in Section 1, except for clause (6) which is defined as follows:

6.  $\langle \mathfrak{A}, K \rangle, A \models \Box\varphi$  iff for all  $\mathfrak{B} \in K$ ,  $\langle \mathfrak{B}, K \rangle, A \models \varphi$ .

Instead of logical truth, a wff is understood to be *universally valid* if it is satisfied by every assignment in every model structure based on a language to which the wff belongs. Where **QS5** is standard first-order logic with

identity supplemented with the axioms of **S5** propositional modal logic, a completeness theorem for the secondary semantics of logical necessity was proved by Kripke in [1959].

**THEOREM (Completeness Theorem).** A set  $\Gamma$  of wffs is consistent in **QS5** iff all the members of  $\Gamma$  are simultaneously satisfiable in a model structure; and (therefore) a wff  $\varphi$  is a theorem of **QS5** iff  $\varphi$  is universally valid.

The secondary semantics, despite the above completeness theorem, has too high a price to pay as far as logical atomism is concerned. In particular, unlike the situation in the primary semantics, the secondary semantics does not validate the modal thesis of anti-essentialism—i.e. it is false that every instance of the thesis is universally valid. This is so of course because necessity no longer represents an invariance through all the possible worlds of a given logical space but only through those in arbitrary non-empty sets of such worlds; that is, necessity is now allowed to represent an internal condition of propositions (or of their corresponding states of affairs) which has material and not merely formal content—for what is invariant through all the members of such a non-empty set need not be invariant through all the possible worlds (models) of the logical space to which those in the set belong.

One example of how such material content affects the implicit metaphysical background can be found in monadic modal predicate logic. It is well-known, for example, that modal free monadic predicate logic is decidable and that no modal free monadic wff can be true in an infinite model unless it is true in a finite model as well. Consequently, any substitution instance of a modal free monadic wff for a relational predicate in an infinity axiom is not only false but logically false. It follows, accordingly, that there can be no modal free analysis or reduction otherwise of all relational predicates or open wffs in terms only of monadic predicates, i.e. in terms only of modal free monadic wffs.

Now it turns out that the same result also holds in the primary semantics for quantified modal logic. That is, in the primary semantics, modal monadic predicate logic is also decidable and no monadic wff, modal free or otherwise, can be true in an infinite model unless it is also true in a finite model (cf. [Cocchiarella, 1975b]). Consequently, there can also be no modal analysis or reduction otherwise of all relational predicates or open wffs in terms only of monadic wffs, modal free or otherwise.

With respect to the secondary semantics, however, the situation is quite different. In particular, as Kripke has shown in [1962], modal monadic predicate logic, as interpreted in the secondary semantics, is not decidable. Moreover, on the basis of that semantics a modal analysis of relational predicates in terms of monadic predicates can in general be given. E.g.,

substituting  $\diamond(Fx \wedge Gy)$  for the binary predicate  $R$  in the infinity axiom

$$\begin{aligned} &\forall x \neg R(x, x) \wedge \forall x \exists y R(x, y) \wedge \forall x \forall y \forall z [R(x, y) \\ &\quad \wedge R(y, z) \rightarrow R(x, z)] \end{aligned}$$

results in a modal monadic sentence which is true in some model structure based on an infinite universe and false in all model structures based on a finite domain. Somehow, in other words, relational content has been incorporated in the semantics for necessity, and thereby of possibility as well. In this respect, the secondary semantics is not the semantics of a merely formal or logical necessity but of a necessity having additional content as well.

Kripke himself, it should be noted, speaks of the necessity of his semantics not as a formal or logical necessity but as a metaphysical necessity (cf. [Kripke, 1971, p. 150]). Indeed, it is precisely because he is concerned with a metaphysical or material necessity and not a logical necessity that not every necessary proposition needs to be *a priori*, nor every *a posteriori* proposition contingent (*ibid.*). Needless to say, however, but that the latter result should obtain does not of itself amount to a refutation, as it is often taken, of the claim of logical atomism that every logically necessary proposition is *a priori* and that every *a posteriori* proposition is logically contingent. We are simply in two different metaphysical frameworks, each with its own notion of necessity and thereby of contingency as well.

#### 4 PROPER NAMES AS RIGID DESIGNATORS

Ordinary proper names in the framework of logical atomism are not what Bertrand Russell called ‘logically proper names’, because the things they name, if they name anything at all, are not the simple objects that are the constituents of atomic states of affairs. However, whereas the names of ordinary language have a sense (*Sinn*) insofar as they are introduced into discourse with identity criteria (usually provided by a sortal common noun with which they are associated—cf. [Geach, 1962, p. 43 f]), the logically proper names of logical atomism have no sense other than what they designate. In other words, in logical atomism, ‘a name means (*bedeutet*) an object. The object is its meaning (*Bedeutung*)’ (*Tractatus* 3.203). Different identity criteria have no bearing on the simple objects of logical atomism, and (pseudo) identity propositions, strictly speaking, have no sense (*Sinn*)—i.e. they do not represent an atomic state of affairs.

Semantically, what this comes to is that logically proper names, or *individual constants*, are rigid designators; that is, their introduction into formal languages requires that the wff

$$\exists x \Box(a = x)$$

be logically true in the primary semantics for each individual constant  $a$ . Carnap, in his formulation of the primary semantics, also required that  $(a \neq b)$  be logically true for distinct individual constants; but that was because his semantics was given in terms of state descriptions where redundant proper names have a complicating effect. Carnap's additional assumption that there is an individual constant for every object in the universe is, of course, also an assumption demanded by his use of state descriptions and is not required by our present model-theoretic approach. (It is noteworthy, however, that the assumption amounted in effect to perhaps the first substitution interpretation of quantifiers, and that in fact it was Carnap who first observed that a strong completeness theorem even for modal free wffs could not be established for an infinite domain on the basis of such an interpretation. Cf. [Carnap, 1938, p. 165].)

Kripke also claims that proper names are rigid designators, but his proper names are those of ordinary language and, as already noted, his necessity is a metaphysical and not a logical necessity. Nevertheless, in agreement with logical atomism the function of a proper name, according to Kripke, is simply to refer, and not to describe the object named [Kripke, 1971, p. 140]; and this applies even when we fix the reference of a proper name by means of a definite description—for the relation between a proper name and a description used to fix the reference of the name is not that of synonymy [Kripke, 1971, p. 156f]. To the objection that we need a criterion of identity across possible worlds before we can determine whether a name is rigid or not, Kripke notes that we should distinguish *how we would speak* in a counterfactual situation from *how we do speak* of a counterfactual situation [Kripke, 1971, p. 159]. That is, the problem of cross-world identity, according to Kripke, arises only through confusing the one way of speaking with the other and that it is otherwise only a pseudo-problem.

## 5 NON-CONTINGENT IDENTITY AND THE CARNAP–BARCAN FORMULA

As rigid designators, proper names cannot be the only singular terms occurring in contingent identity statements. That is, a contingent identity statement must contain at least one definite description whose descriptive content is what accounts for the possibility of different designata and thereby of the contingency of the statement in question. However, in general, as noted by [Smullyan, 1948], there is no problem about contingent identity in quantified modal logic if one of the singular terms involved is a definite description; or rather there is no problem so long as one is careful to observe the proper scope distinctions. On the other hand, where scope distinctions are not assumed to have a bearing on the occurrence of a proper name, the problem of contingent identity statements involving only proper names is trivially resolved by their construal as rigid designators. That is, where  $a$

and  $b$  are proper names or individual constants, the sentences

$$(a = b) \rightarrow \Box(a = b), \quad (a \neq b) \rightarrow \Box(a \neq b)$$

are to be logically true in the primary semantics and universally valid in the secondary. In other words, whether in the context of logical atomism or Kripke's metaphysical necessity, there are only non-contingent identity statements involving only proper names.

Now it is noteworthy that the incorporation of identity-difference conditions in the modal thesis of anti-essentialism disassociates these conditions from the question of essentialism. This is certainly as it should be in logical atomism, since in that framework, as F. P. Ramsey was the first to note, 'numerical identity and differences are necessary relations' [Ramsey, 1960, p. 155]. In other words, even aside from the use of logically proper names, the fact that there can be no contingent identities or non-identities in logical atomism is reflected in the logical truth of both of the wffs

$$\forall x \forall y (x = y \rightarrow \Box x = y), \quad \forall x \forall y (x \neq y \rightarrow \Box x \neq y)$$

in the primary semantics. But then even in the framework of Kripke's metaphysical necessity (where quantifiers also refer *directly* to objects), an object cannot but be the object that it is, nor can one object be identical with another—a metaphysical fact which is reflected in the above wffs being universally valid as well.

Another observation made by Ramsey in his adoption of the framework of logical atomism was that the number of objects in the world is part of its logical scaffolding [Ramsey, 1960]. That is, for each positive integer  $n$ , it is either necessary or impossible that there are exactly  $n$  individuals in the world; and if the number of objects is infinite, then, for each positive integer  $n$ , it is necessary that there are at least  $n$  objects in the world (cf. [Cocchiarella, 1975a, Section 5]. This is so in logical atomism because every possible world consists of the same totality of objects that are the constituents of the atomic states of affairs constituting the actual world. In logical atomism, in other words, an object's existence is not itself an atomic state of affairs but consists in that object's being a constituent of atomic states of affairs.

One important consequence of the fact that every possible world (of a given logical space) consists of the same totality of objects is the logical truth in the primary semantics of the well-known Barcan formula (and its converse):

$$\forall x \Box \varphi \leftrightarrow \Box \forall x \varphi.$$

Carnap, it should be noted, was the first to argue for the logical truth of this principle (in [Carnap, 1946, Section 10] and [1947, Section 40]) which he validated in terms of the substitution interpretation of quantifiers in

his state description semantics. The validation does not depend, of course, on the number of objects being denumerably infinite, though, as already noted, Carnap did impose that condition on his state descriptions. But then—even though Carnap himself did not give this argument— given the non-contingency of identity, the logical truth of the Carnap–Barcan formula, and the assumption for each positive integer  $n$  that it is not necessary that there are just  $n$  objects in the world, it follows that the number of objects in the world must be infinite. (For if everything is one of a finite number  $n$  of objects, then, by the non-contingency of identity, everything is necessarily one of  $n$  objects, and therefore by the Carnap–Barcan formula, necessarily everything is one of  $n$  objects; i.e. contrary to the assumption, it is necessary after all that there are just  $n$  objects in the world.)

As the above remarks indicate, the validation of the Carnap–Barcan formula in the framework of logical atomism is unproblematic; and therefore its logical truth in the primary semantics is as it should be. However, besides being logically true in the primary semantics the principle is also universally valid in the secondary semantics; and it is not clear that this is as it should be for Kripke’s metaphysical necessity. Indeed, Kripke’s later modified semantics for quantified modal logic in [Kripke, 1963] suggests he thinks otherwise, since there the Carnap–Barcan formula is no longer validated. Nevertheless, as indicated above, even with the rejection of the Carnap–Barcan formula, it is clear that Kripke intends his metaphysical context to be such as to support the validation of the non-contingency of identity.

## 6 EXISTENCE IN THE PRIMARY AND SECONDARY SEMANTICS

In rejecting the Carnap–Barcan formula, one need not completely reject the assumption upon which it is based, *viz.* that every possible world (of a given logical space) consists of the same totality of objects. All one need do is take this totality not as the set of objects existing in each world but as the sum of objects that exist in some world or other (of the same logical space), i.e. as the totality of possible objects (of that logical space). Quantification with respect to a world, however, is always to be restricted to the objects existing in that world—though free variables may, as it were, range over the possible objects, thereby allowing a single interpretation of both *de re* and *de dicto* wffs. The resulting quantificational logic is of course free of the presupposition that singular terms (individual variables and constants) always designate an existing object and is for this reason called *free logic* (cf. [Hintikka, 1969]).

Thus, where  $L$  is a language and  $D$  is a non-empty set, then  $\langle\langle\mathfrak{A}, X\rangle, K\rangle$  is a *free model structure* based on  $D$  and  $L$  iff (1)  $\langle\mathfrak{A}, X\rangle \in K$ , (2)  $K$  is a set every member of which is a pair  $\langle\mathfrak{B}, Y\rangle$  where  $\mathfrak{B}$  is an  $L$ -model having

$D$  as its universe and  $Y \subseteq D$  and (3)  $D = \cup\{Y : \text{for some L-model } \mathfrak{B}, \langle \mathfrak{B}, Y \rangle \in K\}$ . Possible worlds are now represented by the pairs  $\langle \mathfrak{B}, Y \rangle$ , where the (possibly empty) set  $Y$  consists of just the objects existing in the world in question; and of course the pair  $\langle \mathfrak{A}, X \rangle$  is understood to represent the actual world. Where  $A$  is an assignment in  $D$ , the *satisfaction by*  $A$  of a wff  $\varphi$  of  $L$  in  $\langle \langle \mathfrak{A}, X \rangle, K \rangle$  is defined as in the secondary semantics, except for clause (5) which is now as follows:

5.  $\langle \langle \mathfrak{A}, X \rangle, K \rangle, A \models \forall x\varphi$  iff for all  $d \in X$ ,  $\langle \langle \mathfrak{A}, X \rangle, K \rangle, A(d/x) \models \varphi$ .

Now if  $K$  is the set of *all* pairs  $\langle \mathfrak{B}, Y \rangle$ , where  $\mathfrak{B}$  is an L-model having  $D$  as its universe and  $Y \subseteq D$ , then  $\langle \langle \mathfrak{A}, X \rangle, K \rangle$  is a *full* free model structure. Of course, whereas validity with respect to all free model structures (based on an appropriate language) is the free logic counterpart of the secondary semantics, validity with respect to all full free model structures is the free logic version of the primary semantics. Moreover, because of the restricted interpretation quantifiers are now given, neither the Carnap–Barcan formula nor its converse is valid in either sense, i.e. neither is valid in either the primary or secondary semantics based on free logic.

If a formal language  $L$  contains proper names or individual constants, then their construal as rigid designators requires that a free model structure  $\langle \langle \mathfrak{A}, X \rangle, K \rangle$  based on  $L$  be such that for all  $\langle \mathfrak{B}, Y \rangle \in K$  and all individual constants  $a$  in  $L$ , the designation of  $a$  in  $\mathfrak{B}$  is the same as the designation of  $a$  in  $\mathfrak{A}$ , i.e. in the actual world. Note that while it is assumed that every individual constant designates a possible object, i.e. possibly designates an existing object, it need not be assumed that it designates an existing object, i.e. an object existing in the actual world. In that case, the rigidity of such a designator is not given by the validity of  $\exists x\Box(a = x)$  but by the validity of  $\Diamond\exists x\Box(a = x)$  instead. Existence of course is analysable as follows:

$$E!(a) = df \exists x(a = x).$$

Note that since possible worlds are now differentiated from one another by the objects existing in them, the concept of existence, despite its analysis in logical terms, must be construed here as having material and not merely formal content. In logical atomism, however, that would mean that the existence or non-existence of an object is itself an atomic state of affairs after all, since now even merely possible objects are constituents of atomic states of affairs. To exclude the later situation, i.e. to restrict the constituents of atomic states of affairs to those that exist in the world in question, would mean that merely possible worlds are not after all merely alternative combinations of the same atomic states of affairs that constitute the actual world; that is, it would involve rejecting one of the basic features of logical atomism, and indeed one upon which the coherence of the framework depends. In this regard, it should be noted, while it is one thing to reject logical



atomism (as probably most of us do) as other than a paradigm of logical analysis, it is quite another to accept some of its basic features (such as the interpretation of necessity as referring to *all* the possible worlds of a given logical space) while rejecting others (such as the constitutive nature of a possible world); for in that case, even if it is set-theoretically consistent, it is no longer clear that one is dealing with a philosophically coherent framework.

That existence should have material content in the secondary semantics, on the other hand is no doubt as it should be, since as already noted, necessity is itself supposed to have such content in that semantics. The difficulty here, however, is that necessity can have such content in the secondary semantics only in a free model structure that is *not* full; for with respect to the full free model structure, the modal thesis of anti-essentialism (with quantifiers now interpreted as respecting existing objects only) can again be validated, just as it was in the original primary semantics. (A full free model structure, incidentally, is essentially what Parsons in [1969] calls a maximal model structure.) The key lemma that led to its validation before continues to hold, in other words, only now for free model structures  $\langle\langle\mathfrak{A}, X\rangle, K\rangle$  and  $\langle\langle\mathfrak{B}, Y\rangle, K\rangle$  instead of the models  $\mathfrak{A}$  and  $\mathfrak{B}$ , and for an isomorphism  $h$  between  $\mathfrak{A}$  and  $\mathfrak{B}$  such that  $Y = h''(X)$ . Needless to say, moreover, but the incompleteness theorem of the primary semantics for logical truth also carries over to universal validity with respect to all full free model structures.

No doubt one can attempt to avoid this difficulty by simply excluding full free model structures; but that in itself would hardly constitute a satisfactory account of the metaphysical content of necessity (and now of existence as well). For there remains the problem of explaining how arbitrary non-empty subsets of the set of possible worlds in a free model structure can themselves be the referential basis for necessity in other free model structures. Indeed, in general, the problem with the secondary semantics is that it provides no explanation of why *arbitrary* non-empty sets of possible worlds can be the referential basis of necessity. In this regard, the secondary semantics of necessity is quite unlike the secondary semantics of second-order logic where, e.g. general models are subject to the constraints of the compositional laws of a comprehension principle.

## 7 METAPHYSICAL NECESSITY AND RELATIONAL MODEL STRUCTURES

It is noteworthy that in his later rejection of the Carnap–Barcan formula, Kripke also introduced a further restriction into the quantificational semantics of necessity, viz. that it was to refer not to all the possible worlds in a given model structure but only to those that are possible alternatives to

the world in question. In other words, not only need not all the worlds in a given logical space be in the model structure (the first restriction), but now even the worlds in the model structure need not all be possible alternatives to one another (the second restriction). Clearly, such a restriction within the first restriction only deepens the sense in which the necessity in question is no longer a logical but a material or metaphysical modality.

The virtue of a relational interpretation, as is now well-known, is that it allows for a general semantical approach to a whole variety of modal logics by simply imposing in each case certain structural conditions on the relation of accessibility (or alternative possibility) between possible worlds. Of course, in each such case, the question remains as to the real nature and content of the structural conditions imposed, especially if our concern is with giving necessity a metaphysical or material interpretation as opposed to a merely formal or set-theoretical one. How this content is explained and filled in, needless to say, will no doubt affect how we are to understand modality *de re* and the question of essentialism.

Retaining the semantical approach of the previous section where the restrictions on possible worlds (models with a restricted existence set) are rendered explicit, we shall understand a *relational model structure* based on a universe  $D$  and a language  $L$  to be a triple  $\langle\langle\mathfrak{A}, X\rangle, K, R\rangle$  where (1)  $\langle\langle\mathfrak{A}, X\rangle, K\rangle$  is a free model structure and (2)  $R \subseteq K \times K$ . If  $A$  is an assignment in  $D$ , then *satisfaction* by  $A$  is defined as in Section 6, except for clause (6) which now is as follows:

6.  $\langle\langle\mathfrak{A}, X\rangle, K, R\rangle, A \models \Box\varphi$  iff for all  $\langle\mathfrak{B}, Y\rangle \in K$ , if  $\langle\mathfrak{A}, X\rangle R \langle\mathfrak{B}, Y\rangle$ , then  $\langle\langle\mathfrak{B}, Y\rangle, K, R\rangle, A \models \varphi$ .

Needless to say, but if  $R = K \times K$  and  $K$  is full, then once again we are back to the free logic version of the primary semantics; and, as before, even excluding relational model structures that are full in this extended sense still leaves us with the problem of explaining how otherwise arbitrary non-empty sets of possible worlds of a given logical space, together now with a relation of accessibility between such worlds, can be the basis of a metaphysical modality.

No doubt it can be assumed regarding the implicit metaphysical framework of such a modality that in addition to the objects that exist in a given world there are properties and relations which these objects either do or do not have and which account for the truths that obtain in that world. They do so, of course, by being what predicate expressions stand for as opposed to the objects that are the designata of singular terms. (Nominalism, it might be noted, will not result in a coherent theory of predication in a framework which contains a metaphysical modality—a point nominalists themselves insist on.)

On the other hand, being only what predicates stand for, properties and relations do not themselves exist in a world the way objects do. That

is, unlike the objects that exist in a world and which might not exist in another possible world, the properties and relations that are predicable of objects in one possible world are the same properties and relations that are predicable of objects in any other possible world. In this regard, what is semantically peculiar to a world about a property or relation is not the property or relation itself but only its extension, i.e. the objects that are in fact conditioned by that property or relation in the world in question. Understood in this way, a property or relation may be said to have in itself only a transworld or non-substantial mode of being.

Following Carnap [1955], who was the first to make this sort of proposal, we can represent a property or relation in the sense indicated by a function from possible worlds to extensions of the relevant sort. With respect to the present semantics, however, it should be noted that the extension which a predicate expression has in a given world need not be drawn exclusively from the objects that exist in that world. That is, the properties and relations that are part of the implicit metaphysical framework of the present semantics may apply not only to existing objects but to possible objects as well—even though quantification is only with respect to existing objects. Syntactically, this is reflected in the fact that the rule of substitution:

$$\text{if } \models \varphi, \text{ then } \models \check{S}_{\psi}^{F(x_1, \dots, x_n)} \varphi \mid$$

is validated in the present semantics; and this in turn indicates that any open wff  $\psi$ , whether *de re* or *de dicto*, may serve as the definiens of a possible definition for a predicate. That is, it can be shown by means of this rule that such a definiens will satisfy both the criterion of eliminability and the criterion of non-creativity for explicit definitions of a new predicate constant. (Beth's Definability Theorem fails for the logic of this semantics, however, and therefore so does Craig's Interpolation Lemma which implies the Definability Theorem, cf. [Fine, 1979].)

We can, of course, modify the present semantics so that the extension of a predicate is always drawn exclusively from existing objects. That perhaps would make the metaphysics implicit in the semantics more palatable—especially if metaphysical necessity in the end amounts to a physical or natural necessity and the properties and relations implicit in the framework are physical or natural properties and relations rather than properties and relations in the logical or intensional sense. However, in that case we must then also give up the above rule of substitution and restrict the conditions on what constitutes a possible explicit definition of a predicate; e.g. not only would modal wffs in general be excluded as possible definiens but so would the negation of any modal free wff which itself was acceptable. In consequence, not only would many of the predicates of natural language not be representable by predicates of a formal modal language—their associated 'properties' being dispositional or modal—but even their possible analyses in terms of predicates that are acceptable would also be excluded.

## 8 QUANTIFICATION WITH RESPECT TO INDIVIDUAL CONCEPTS

One way out of the apparent impasse of the preceding semantics is the turn to intensionality, i.e. the turn to an independently existing Platonic realm of intensional existence (and inexistence) and away from the metaphysics of either essentialism (natural kinds, physical properties, etc.) or anti-essentialism (logical atomism). In particular, it is claimed, problems about states of affairs, possible objects and properties and relations (in the material sense) between such objects are all avoidable if we would only turn instead to propositions, individual concepts and properties and relations in the logical sense, i.e. as the intensions of predicate expressions and open wffs in general. Thus, unlike the problem of whether there can be a state of affairs having merely possible objects among its constituents, there is nothing problematic, it is claimed, about the intensional existence of a proposition having among its components intensionally inexistent individual concepts, i.e. individual concepts that fail to denote (*bedeuten*) an existing object.

Intensional entities (*Sinne*), on this approach, do not exist in space and time and are not among the individuals that differentiate one possible world from another. They are rather non-substantial transworld entities which, like properties and relations, may have different extensions (*Bedeutungen*) in different possible worlds. For example, the extension of a proposition in a given world is its truth-value, i.e. truth or falsity (both of which we shall represent here by 1 and 0, respectively), and the extension of an individual concept in that world is the object which it denotes or determines. We may, accordingly, follow Carnap once again and represent different types of intensions in general as functions from possible worlds to extensions of the relevant type. In doing so, however, we shall no longer identify possible worlds with the extensional models of the preceding semantics; that is, except for the objects that exist in a given world, the nature and content of that world will otherwise be left unspecified. Indeed, because intensionality is assumed to be conceptually prior to the functions on possible worlds in terms of which it will herein be represented, the question of whether a merely possible world, or of whether a merely possible object existing in such a world, has an ontological status independent of the realm of intensional existence (or inexistence) is to be left open on this approach (if not closed in favour of an analysis or reduction of such worlds and objects in terms of propositions and the intensional inexistence of individual concepts).

Accordingly, a triple  $\langle W, R, E \rangle$  will be said to be a *relational world system* if  $W$  is a non-empty set of possible worlds,  $R$  is an accessibility relation between the worlds in  $W$ , i.e.  $R \subseteq W \times W$ , and  $E$  is a function on  $W$  into (possibly empty) sets, though for some  $w \in W$  (representing the actual world),  $E(w)$  is non-empty. (The sets  $E(w)$ , for  $w \in W$ , consist of the objects existing in each of the worlds in  $W$ .) Where  $D = \cup_{w \in W} E(w)$ , an

*individual concept* in  $\langle W, R, E \rangle$ , is a function in  $D^W$ ; and for each natural number  $n$ ,  $P$  is an  $n$ -place *predicate intension* in  $\langle W, R, E \rangle$  iff  $P \in \{X : X \subseteq D^n\}^W$ . (Note: for  $n = 0$ , we take an  $n$ -place predicate intension in  $\langle W, R, E \rangle$  to be a *proposition* in  $\langle W, R, E \rangle$ , and therefore since  $D^0 = \{0\}$  and  $2 = \{0, 1\}$ ,  $P$  is a proposition in  $\langle W, R, E \rangle$  iff  $P \in 2^W$ . For  $n \geq 2$ , an  $n$ -place predicate intension is also called an  $n$ -ary *relation-in-intension*, and for  $n = 1$ , it is taken as a property in the logical sense.)

Where  $L$  is a language,  $I$  is said to be an *interpretation* for  $L$  based on a relational world system  $\langle W, R, E \rangle$  if  $I$  is a function on  $L$  such that (1) for each individual constant  $a \in L$ ,  $I(a)$  is an individual concept in  $\langle W, R, E \rangle$ ; and (2) if  $F^n$  is an  $n$ -place predicate in  $L$ , then  $I(F^n)$  is an  $n$ -place predicate intension in  $\langle W, R, E \rangle$ . An *assignment*  $A$  in  $\langle W, R, E \rangle$  is now a function on the individual variables such that  $A(x)$  is an individual concept in  $\langle W, R, E \rangle$ , for each such variable. The *intension* with respect to  $I$  and  $A$  of an individual variable or constant  $b$  in  $L$ , in symbols  $\text{Int}(b, I, A)$  is defined to be  $(I \cup A)(b)$ .

Finally, the *intension with respect to  $I$  and  $A$  of an arbitrary wff  $\varphi$  of  $L$*  is defined recursively as follows:

1. where  $a, b$  are individual variables or constants in  $L$ ,  $\text{Int}(a = b, I, A) =$  the  $p \in 2^W$  such that for  $w \in W$ ,  $P(w) = 1$  iff  $\text{Int}(a, I, A)(w) = \text{Int}(b, I, A)(w)$ ;
2. where  $a_1, \dots, a_n$  are individual variables or constants in  $L$  and  $F^n \in L$ ,  $\text{Int}(F(a_1, \dots, a_n), I, A) =$  the  $P \in 2^W$  such that for  $w \in W$ ,  $P(w) = 1$  iff  $\langle \text{Int}(a_1, I, A)(w), \dots, \text{Int}(a_n, I, A)(w) \rangle \in I(F^n)(w)$ ;
3.  $\text{Int}(\neg\varphi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W$ ,  $P(w) = 1$  iff  $\text{Int}(\varphi, I, A)(w) = 0$ ;
4.  $\text{Int}((\varphi \rightarrow \psi), I, A) =$  the  $P \in 2^W$  such that for  $w \in W$ ,  $P(w) = 1$  iff either  $\text{Int}(\varphi, I, A)(w) = 0$  or  $\text{Int}(\psi, I, A)(w) = 1$ ;
5.  $\text{Int}(\forall x\varphi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W$ ,  $P(w) = 1$  iff for *all* individual concepts  $f$  in  $\langle W, R, E \rangle$ ,  $\text{Int}(\varphi, I, A(f/x))(w) = 1$ ; and
6.  $\text{Int}(\Box\varphi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W$ ,  $P(w) = 1$  iff for all  $v \in W$ , if  $wRv$ , then  $\text{Int}(\varphi, I, A)(v) = 1$ .

Different notions of *intensional validity*, needless to say, can now be defined as in the earlier semantics depending on the different structural properties that the relation of accessibility might be assumed to have. It can be shown, however, from results announced by Kripke in [1976] that if the relation is only assumed to be reflexive, or reflexive and symmetric but not also transitive, or reflexive and transitive but not also symmetric, then the wffs that are intensionally valid with respect to the relational structures in

question are not recursively enumerable; that is, the resulting semantics is then essentially incomplete. Whether the semantics is also incomplete for intensional validity with respect to the class of relational world systems in which the relation of accessibility is an equivalence relation, or, equivalently, in which it is universal between all the worlds in the system, has apparently not yet been determined (or at any rate not yet announced or published in the literature). However, because of its close similarity to Thomason's system **Q2** in [Thomason, 1969], the **S5** version of which Kripke in [1976] has claimed to be complete, we conjecture that it too is complete, i.e. that the set of wffs (of a given language) that are intensionally valid with respect to all relational world systems in which the relation of accessibility is universal is recursively enumerable. For convenience, we shall speak of the members of this set hereafter as being *intensionally valid simpliciter*; that is, we shall take the members of this set as being intensionally valid in the primary sense (while those that are valid otherwise are understood to be so in a secondary sense).

It is possible of course to give a secondary semantics in the sense in which quantification need not be with respect to *all* of the individual concepts in a given relational world system but only with respect to some non-empty set of such. (Cf. [Parks, 1976] where this gambit is employed—but in a semantics in which predicates have their extensions drawn in a given world from the restricted set of individual concepts and not from the possible objects.) As might be expected, completeness theorems are then forthcoming in the usual way even for classes of relational world systems in which the relation of accessibility is other than universal. Of course the question then arises as to the rationale for allowing *arbitrary* non-empty sets of individual concepts to be the basis for quantifying over such in any given relational world system. This question, moreover, is not really on a par with that regarding allowing arbitrary non-empty subsets of the set of possible worlds to be the referential basis of necessity (even where the relation of accessibility is universal); for in a framework in which the realm of intensionality is conceptually prior to its representation in terms of functions on possible worlds, the variability of the sets of possible world may in the end be analogous to the similar variability of the universes of discourses in standard (modal free) first-order logic. Such variability within the intensional realm itself, on the other hand, would seem to call for a different kind of explanation.

Thomason's semantical system **Q2** in [Thomason, 1969], it should be noted, differs from the above semantics for intensional validity *simpliciter* in requiring first that the set of existing objects of each possible world be non-empty, and, secondly, that although free individual variables range over the entire set of individual concepts, quantification in a given world is to be restricted to those individual concepts which denote objects that exist in that world. (Thomason also gives an 'outer domain' interpretation for improper definite descriptions which we can ignore here since definite de-

scriptions are not singular terms of the formal languages being considered.) Thus, clause (5) for assigning an intension to a quantified wff is replaced in **Q2** by:

(5')  $\text{Int}(\forall x\varphi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W, P(w) = 1$  if for all individual concepts  $f$  in  $\langle W, R, E \rangle$ , if  $f(w) \in E(w)$ , then  $\text{Int}(\varphi, I, A(f/x))(w) = 1$ .

Intensional **Q2**-validity can now be defined as intensional validity with respect to all relational world systems in which (1) the set of objects existing in each world is non-empty, (2) the relation of accessibility is universal, and (3) quantification is interpreted as in clause(5'). According to Kripke [1976], the set of wffs (of a given language) that are intensionally **Q2**-valid is recursively enumerable. (Cf. [Bacon, 1980] for some of the history of these results and of an earlier erroneous claim by David Kaplan.) The completeness proof given in [Kamp, 1977], it should be noted, is not for **Q2**-validity (as might be thought from Kamp's remark that he is reconstructing Kripke's proof), but for a semantics in which individual concepts always denote only existing objects and in which the extension of a predicate's intension in a given world is drawn exclusively from the objects that exist in that world. Both conditions are too severe, however—at least from the point of view of the realm of intensional existence (an inexistence). In particular, whereas the rule of substitution:

$$\text{if } \models \varphi, \text{ then } \models \tilde{S}_{\psi}^{F(x_1, \dots, x_n)} \varphi \mid$$

is validated in both the semantics of intensional validity *simpliciter* and in the **Q2**-semantics, it is not validated in Kamp's more restricted semantics. Not all open wffs, in other words, represent predicate intensions, i.e. properties or relations in the logical sense, in Kamp's semantics—a result contrary to one of the basic motivations for the turn to intensionality.

## 9 INDIVIDUAL CONCEPTS AND THE ELIMINATION OF *DE RE* MODALITIES

One of the nice things about Thomason's **Q2**-semantics is that existence remains essentially a quantifier concept; that is, the definition of  $E!$  given earlier remains in effect in the **Q2** semantics. This is not so of course in the semantics for intensional validity *simpliciter* where  $E!$  would have to be introduced as a new intensional primitive (cf. [Bacon, 1980]). There would seem to be nothing really objectionable about doing so, however; or at least not from the point of view of the realm of intensional existence (and inexistence). Quantifying over all individual concepts, whether existent or in-existent—i.e. whether they denote objects that exist in the world in

question or not—can hardly be compared from this point of view with the different situation of quantifying over all possible objects in the semantics of a metaphysical necessity (as opposed to quantifying only over the objects that exist in the world in question in that metaphysical context).

One of the undesirable features of the **Q2**-semantics, however, is its validation of the wff

$$\exists x \Box E!(x)$$

which follows from the **Q2**-validity of

$$\Box \exists x \varphi \rightarrow \exists x \Box \varphi \quad \text{and} \quad \Box \exists x E!(x).$$

This situation can be easily rectified, of course, by simply rejecting the semantics of the latter wff, i.e. by not requiring the set of objects existing in each world of the **Q2**-semantics to be non-empty. The converse of the above conditional is not intensionally **Q2**-valid, incidentally, though both are intensionally valid in the primary sense; that is

$$(\Box \exists / \exists \Box) \exists x \Box \varphi \leftrightarrow \Box \exists x \varphi$$

is intensionally valid *simpliciter*. So of course is the Carnap–Barcan formula (and its converse), which also fails (in both directions) in the **Q2**-semantics; for this formula and its converse, it is well-known, is a consequence of the **S5** modal principles together with those of standard first-order predicate logic without identity (LPC). Of course, whereas every wff which is an instance of a theorem of LPC is intensionally valid in the primary sense, it is only their modal free-logic counterparts that are valid in the **Q2**-semantics. For example, whereas

$$\forall x \varphi \rightarrow \varphi(a/x)$$

is intensionally valid *simpliciter*, only its modal free-logic counterpart

$$\exists x \Box (a = x) \rightarrow [\forall x \varphi \rightarrow \varphi(a/x)]$$

is intensionally **Q2**-valid.

One rather important consequence of these results of the semantics for intensional validity in the primary sense, it should be noted, is the validation in this sense, of Von Wright's principle of predication (cf. [von Wright, 1951]) i.e. the principle (as restated here in terms of individual concepts) that if a property or relation in the logical sense is contingently predicabile of the denotata of some individual concepts, then it is contingently predicabile of the denotata of all individual concepts:

$$(Pr) \quad \exists x_1 \dots \exists x_n (\Diamond \varphi \wedge \Diamond \neg \varphi) \rightarrow \forall x_1 \dots \forall x_n (\Diamond \varphi \wedge \Diamond \neg \varphi).$$



The fact that (Pr) is intensionally valid *simpliciter* can be seen from the syntactic proof in [Broido, 1976] that

$$\text{LPC} + \mathbf{S5} + (\Box\exists/\exists\Box) \vdash (\text{Pr}).$$

That is, since every wff which is an axiom of  $\text{LPC} + \mathbf{S5} + (\Box\exists/\exists\Box)$  is intensionally valid *simpliciter*, and *modus ponens* and universal modal generalisation preserve validity in this sense, then (Pr) is also intensionally valid *simpliciter*.

Now although not every wff which is an axiom of  $\text{LPC} + (\Box\exists/\exists\Box)$  is valid in the **Q2**- semantics, nevertheless, it follows from the intensional validity of (Pr) in the primary sense that (Pr) is intensionally **Q2**-valid as well. To see this, assume that  $E!$  is added as a new intensional primitive with the following clause added to the semantics of the preceding section:

$$\begin{aligned} \text{Int}(E!(a), I, A) &= \text{the } P \in 2^W \text{ such that for } w \in W, P(w) = 1 \\ &\text{iff } \text{Int}(a, I, A)(w) \in E(w). \end{aligned}$$

Then, where  $t$  translates each wff into its  $E!$  restricted counterpart, i.e. where  $t(\varphi) = \varphi$ , for atomic wffs,  $t(\neg\varphi) = \neg t(\varphi)$ ,  $t(\varphi \rightarrow \psi) = (t(\varphi) \rightarrow t(\psi))$ ,  $t(\Box\varphi) = \Box t(\varphi)$  and  $t(\forall x\varphi) = \forall x(E!(x) \rightarrow t(\varphi))$ , it can be readily seen that a wff  $\varphi$  is intensionally **Q2**-valid iff  $[\exists x\Box E!(x) \rightarrow t(\varphi)]$  is intensionally valid *simpliciter*; and therefore if  $t(\varphi)$  is intensionally valid *simpliciter*, then  $\varphi$  is intensionally **Q2**-valid. Now since

$$\exists x_1 \dots \exists x_n [\Diamond t(\varphi) \wedge \Diamond \neg t(\varphi)] \rightarrow \forall x_1 \dots \forall x_n [\Diamond t(\varphi) \wedge \Diamond \neg t(\varphi)]$$

is an instance of (Pr), it is intensionally valid *simpliciter*, and therefore so is

$$\begin{aligned} \exists x_1 \dots \exists x_n [E!(x_1) \wedge \dots \wedge E!(x_n) \wedge \Diamond t(\varphi) \wedge \Diamond \neg t(\varphi)] \rightarrow \\ \rightarrow \forall x_1 \dots \forall x_n [E!(x_1) \wedge \dots \wedge E!(x_n) \rightarrow \Diamond t(\varphi) \wedge \Diamond \neg t(\varphi)]. \end{aligned}$$

This last wff, however, is trivially equivalent to  $t(\text{Pr})$ . That is,  $t(\text{Pr})$  is intensionally valid *simpliciter*, and therefore (Pr) is intensionally **Q2**-valid.

It is noteworthy, finally, that on the basis of  $\text{LPC} + \mathbf{S5} + (\text{Pr}) + (\Box\exists/\exists\Box)$ , [Broido, 1976] has shown that every *de re* wff is provably equivalent to a *de dicto* wff. Accordingly, since all of the assumptions or wffs essential to Broido's proof are intensionally valid *simpliciter*, it follows that every *de re* wff is eliminable in favour of only *de dicto* wffs (of modal degree  $\leq 1$ ) in the semantics of intensional validity in the primary sense.

**THEOREM (De Re Elimination Theorem)** For each *de re* wff  $\varphi$ , there is a *de dicto* wff  $\psi$  such that  $(\varphi \leftrightarrow \psi)$  is intensionally valid *simpliciter*.

Kamp [1977] has shown, incidentally, that a *de re* elimination theorem also holds for the more restricted semantics in which individual concepts always denote existing objects and the extensions of predicate intensions

are always drawn exclusively from the objects existing in the world in question. (This theorem is in fact the basis of Kamp's completeness theorem for his semantics—and therefore perhaps also the basis for a similar proof of completeness for the semantics of intensional validity *simpliciter*.) Accordingly, since the logic of the **Q2**-semantics is intermediate between Kamp's semantics and the semantics of intensional validity in the primary sense, it is natural to conjecture that a similar *de re* elimination theorem also holds for intensional **Q2**-validity—though of course not one which depends on the consequences of  $\text{LPC} + (\Box\exists/\exists\Box)$ .

## 10 CONTINGENT IDENTITY

Identity in logical atomism, as we have already noted, does not stand for an atomic state of affairs; that is, despite its being represented by an atomic wff, an object's self-identity is part of the logical scaffolding of the world (which the world shares with every other possible world) and not part of the world itself. This is why identity is a non-contingent relation in logical atomism.

Identity in the realm of intensionality, on the other hand, is really not an identity of individual concepts but a world-bound relation of coincidence between these concepts. That is, as a relation in which individual concepts need not themselves be the same but only have the same denotation in a given world, 'identity' need not hold between the same individual concepts from world to world. This is why 'identity' can be a contingent relation from the intensional point of view.

In other words, whereas

$$\exists x\exists y(x = y \wedge \Diamond x \neq y), \quad \exists x\exists y(x \neq y \wedge \Diamond x = y)$$

are logically false from the point of view of the primary semantics for logical atomism, both can be true (and in fact must be true if there are at least two objects in the world) from the point of view of the realm of intensionality.

One argument in favour of contingent identity as a relation of coincidence between individual concepts is given in [Gibbard, 1975]. In Gibbard's example a clay statue named Goliath (hereafter represented by  $a$ ) is said to be contingently identical with the piece of clay of which it is made and which is named Lump1 (hereafter represented by  $b$ ). For convenience, we may suppose that  $a$  and  $b$  begin to exist at the same time; e.g. the statue is made first in two separate pieces which are then struck together 'thereby bringing into existence simultaneously a new piece of clay and a new statue' [Gibbard, 1975, p. 191]. Now although Goliath is Lump1—i.e. ( $a = b$ ) is true in the world in question—it is nevertheless possible that the clay is squeezed into a ball before it dries; and if that is done, then 'at that point . . . the statue Goliath would have ceased to exist, but the piece of clay Lump1 would still

exist in a new shape. Hence Lumpl would not be Goliath, even though both existed' [Gibbard, 1975]. That is, according to Gibbard, the wff

$$a = b \wedge \diamond[a \neq b \wedge E!(a) \wedge E!(b)]$$

would be true in the world in question.

Contrary to Gibbard's claim, however, the above wff is not really a correct representation of the situation he describes. In particular, it is not true that Goliath exists in the world in which Lumpl has been squeezed into a ball. The correct description, in other words, is given by

$$a = b \wedge \diamond[\neg E!(a) \wedge E!(b)],$$

which, since

$$a = b \rightarrow [E!(a) \leftrightarrow E!(b)]$$

is both intensionally valid *simpliciter* and intensionally **Q2**-valid, implies:

$$a = b \wedge \diamond[a \neq b \wedge \neg E!(a) \wedge E!(b)]$$

and this wff in turn implies

$$\neg[a = b \rightarrow \Box a = b],$$

which is the conclusion Gibbard was seeking in any case. That is, the identity of Goliath with Lumpl, though true, is only contingently true.

Now it should be noted in this context that the thesis that proper names are rigid designators can be represented neither by

$$\exists x \Box(a = x) \quad \text{nor} \quad \diamond \exists x \Box(a = x)$$

in the present semantics. For both wffs are in fact intensionally valid *simpliciter*, and the latter would be **Q2**-valid if we assumed that any individual concept expressed by a proper name always denotes an object which exists in at least one possible world—and yet, it is not required in either of these semantics that the individual concept expressed by a proper name is to denote the same object in every possible world. The question arises, accordingly, whether and in what sense Gibbard's example shows that the names 'Goliath' and 'Lumpl' are not rigid designators. For surely there is nothing in the way each name is introduced into discourse to indicate that its designation can change even when the object originally designated has continued to exist; and yet if it is granted that 'Goliath' and 'Lumpl' are rigid designators in the sense of designating the same object in every world in which it exists, then how is it that 'Goliath' can designate Lumpl in the one world where Goliath and Lumpl exist but not in the other where Lumpl but not Goliath exists? Doesn't the same object which 'Goliath' designates in the one world exist in the other? An answer to this problem is forthcoming, as we shall see, but from an entirely different perspective of the realm of intensionality; and indeed one in which identity is not a contingent relation either between objects or individual concepts.

## 11 QUANTIFIERS AS REFERENTIAL CONCEPTS

Besides the Platonic view of intensionality there is also the conceptualist view according to which concepts are not independently existing Platonic forms but cognitive capacities or related structures whose realisation in thought is what informs our mental acts with a predicable or referential nature. However, as cognitive capacities which may or may not be exercised on a given occasion, concepts, though they are not Platonic forms, are also neither mental images nor ideas in the sense of particular mental occurrences. That is, concepts are not objects or individuals but are rather unsaturated cognitive structures or dispositional abilities whose realisation in thought is what accounts for the referential and predicable aspects of particular mental acts.

Now the conceptual structures that account for the referential aspect of a mental act on this view are not the same as those that inform such acts with a predicable nature. A categorical judgement, for example, is a mental act which consists in the joint application of both types of concepts; that is, it is a mental event which is the result of the combination and mutual saturation of a referential concept with a predicable concept. Referential concepts, in other words, have a type of structure which is complementary to that of predicable concepts in that each can combine with the other in a kind of mental chemistry which results in a mental act having both a referential aspect and a predicable nature.

Referential concepts, it should be noted, are not developed initially as a form of reference to individuals *simpliciter* but are rather first developed as a form of reference to individuals of a given sort of kind. By a *sort* (or *sortal concept*) we mean in this context a type of common noun concept whose use in thought and communication is associated with certain identity criteria, i.e. criteria by which we are able to distinguish and count individuals of the kind in question. Typically, perceptual criteria such as those for shape, size and texture (hard, soft, liquid, etc.) are commonly involved in the application of such a concept; but then so are functional criteria (especially edibility) as well as criteria for the identification of natural kinds of things (animals, birds, fish, trees, plants, etc.) (cf. [Lyons, 1977, Vol. 2, Section 11.4]).

Though sortal concepts are expressed by common (count) nouns, not every common (count) noun, on the other hand, stands for a sort or kind in the sense intended here. Thus, e.g., whereas 'thing' and 'individual' are common (count) nouns, the concept of a thing or individual *simpliciter* is not associated in its use with any particular identity criteria, and therefore it is not a sortal concept in the sense intended here. Indeed, according to conceptualism, the concept of a thing or individual *simpliciter* has come to be constructed on the basis of the concept of a thing or individual of a certain sort (cf. [Sellars, 1963]). (It might be noted in this context, inci-

dentally, that while there are no explicit grammatical constructions which distinguish sortal common nouns from non-sortal common (count) nouns in the Indo-European language family, nevertheless there are ‘classifier-languages’—e.g. Tzeltal, a Mayan language spoken in Mexico, Mandarin Chinese, Vietnamese, etc.—which do contain explicit and obligatory constructions involving sortal classifiers (cf. [Lyons, 1977]).

Reference to individuals of a given sort, accordingly, is not a form of restricted reference to individuals *simpliciter*; that is, referential concepts regarding these individuals are not initially developed as derived concepts based on a quantificational reference to individuals in general, but are themselves basic or underived sortal quantifier concepts. Thus, where  $S$  and  $T$  stand for sortal concepts,  $(\forall xS)$ ,  $(\forall yT)$ ,  $(\exists zS)$ ,  $(\exists xT)$ , etc. can be taken on the view in question as basic forms of referential concepts whose application in thought enable us to refer to all  $S$ , all  $T$ , some  $S$ , some  $T$ , etc. respectively. For example, where  $S$  stands for the sort *man* and  $F$  stands for the predicable concept of *being mortal*, a categorical judgement that every man is mortal, or that some man is not mortal, can be represented by  $(\forall xS)F(x)$  and  $(\exists xS)\neg F(x)$ , respectively. These formulas, it will be noted, are especially perspicuous in the way they represent the judgements in question as being the result of a combination and mutual saturation of a referential and predicable concept.

Though they are themselves basic or underived forms of referential concepts, sortal quantifiers are nevertheless a special type of common (count) noun quantifier—including, of course, the ultimate common (count) noun quantifiers  $\forall x$  and  $\exists x$  (as applied with respect to a given individual variable  $x$ ). Indeed, the latter, in regard to the referential concepts they represent, would be more perspicuous if written out more fully as  $(\forall x \textit{Individual})$  and  $(\exists x \textit{Individual})$ , respectively. The symbols  $\forall$  and  $\exists$ , in other words, do not stand in conceptualism for separate cognitive elements but are rather ‘incomplete symbols’ occurring as parts of common (count) noun quantifiers. For convenience, however, we shall continue to use the standard notation  $\forall x$  and  $\exists x$  as abbreviations of these ultimate common (count) noun quantifiers.

## 12 SINGULAR REFERENCE

As represented by common (count) noun quantifiers, referential concepts are indeed complementary to predicable concepts in exactly the way described by conceptualism; that is, they are complementary in the sense that when both are applied together it is their combination and mutual saturation in a kind of mental chemistry which accounts for the referential and predicable aspects of a mental act. It is natural, accordingly, that a parallel interpretation should be given for the referential concepts underlying the use of singular terms.

Such an interpretation, it will be observed, is certainly a natural comitant of Russell's theory of definite descriptions—or rather of Russell's theory somewhat modified. Where  $S$ , for example, is a common (count) noun, including the ultimate common (count) noun 'individual', the truth-conditions for a judgement of the form

1. the  $S$  wh. is  $F$  is  $G$

will be semantically equivalent in conceptualism to those for the wff

2.  $(\exists xS)[(\forall yS)(F(y) \leftrightarrow y = x) \wedge G(x)]$

if, in fact, the definite description is being used in that judgement with an existential presupposition. If it is not being so used, however, then the truth-conditions for the judgement are semantically equivalent to

3.  $(\forall xS)[(\forall yS)[F(y) \leftrightarrow y = x] \rightarrow G(x)]$

instead. Note however that despite the semantical equivalence of one of these wffs with the judgement in question, neither of them can be taken as a direct representation of the cognitive structure of that judgement. Rather, where ' $S$  wh  $F$ '; abbreviates ' $S$  wh. is  $F$ ', a more perspicuous representation of the judgement can be given either by

4.  $(\exists_1 xS \text{ wh } F)G(x)$

or

5.  $(\forall_1 xS \text{ wh } F)G(x)$

respectively, depending on whether the description is being used with or without an existential presupposition. (The 'incomplete' quantifier symbols  $\exists_1$  and  $\forall_1$  are, of course, understood here in such a way as to render (4) and (5) semantically equivalent to (2) and (3), respectively.) The referential concept which underlies using the definite description with an existential presupposition, in other words, is the concept represented by  $(\exists_1 xS \text{ wh } F)$ ; and of course the referential concept which underlies using the description without an existential presupposition is similarly represented by  $(\forall_1 xS \text{ wh } F)$ .

Now while definite descriptions are naturally assimilated to quantifiers, proper names are in turn naturally assimilated to sortal common nouns. For just as the use of a sortal is associated in thought with certain identity criteria, so too is the introduction and use of a proper name (whose identity criteria are provided in part by the most specific sortal associated with the introduction of that name and to which the name is thereafter subordinate). In this regard, the referential concept underlying the use of a proper name is determined by the identity criteria associated with that name's introduction into discourse.

On this interpretation, accordingly, the referential concept underlying the use of a proper name corresponds to the referential concept underlying the use of a sortal common noun; that is, both are to be represented by sortal quantifiers (where 'sortal' is now taken to encompass proper names as well). The only difference between the two is that when such a quantifier contains a proper name, it is always taken to refer to at most a single individual. Thus, if in someone's statement that Socrates is wise, 'Socrates' is being used with an existential presupposition, then the statement can be represented by

$$(\exists x \text{ Socrates})(x \text{ is wise}).$$

If 'Socrates' is being used without an existential presupposition, on the other hand, then the statement can be represented by

$$(\forall x \text{ Socrates})(x \text{ is wise})$$

instead. That is, the referential concepts underlying using 'Socrates' with and without an existential presupposition can be represented by  $(\exists x \text{ Socrates})$  and  $(\forall x \text{ Socrates})$ , respectively. (Such a quantifier interpretation of the use of proper names will also explain, incidentally, why the issue of scope is relevant to the use of a proper name in contexts involving the expression of a propositional attitude.)

Now without committing ourselves at this point as to the sense in which conceptualism can allow for the development of alethic modal concepts, i.e. modal concepts other than those based upon a propositional attitude, it seems clear that the identity criteria associated with the use of a proper name do not change when that name is used in such a modal context. That is, the demand that we need a criterion of identity across the possible worlds associated with such a modality in order to determine whether a proper name is a rigid designator or not is without force in conceptualism since, in fact, such a criterion is already implicit in the use of a proper name. In other words, where  $S$  is a proper name, we can take it as a conceptual truth that the identity criteria associated with the use of  $S$  (1) always picks out at most one object and (2) that it is the same object which is so picked out whenever it exists:

$$(PN) \quad (\forall xS)[\Box(\forall yS)(y = x) \wedge \Box(E!(x) \rightarrow (\exists yS)(x = y))].$$

Nothing in this account of proper names conflicts, it should be noted, with Gibbard's example of the statue Goliath which is identical with Lump1, the piece of clay of which it consists, during the time of its existence, but which ceases to be identical with Lump1 because it ceases to exist when Lump1 is squeezed into a ball. Both names, in other words, can be taken as rigid designators in the above sense without resulting in a contradiction in the situation described by Gibbard. Where  $S$  and  $T$ , for example, are proper

name sortals for ‘Goliath’ and ‘Lumpl’, respectively, the situation described by Gibbard is consistently represented by the following wff:

$$(\exists xS)(\exists yT)(x = y) \wedge \diamond(\exists yT)(\forall xS)(x \neq y).$$

That is, whereas the identity criteria associated with ‘Goliath’ and ‘Lumpl’ enable us to pick out the same object in the original world or time in question, it is possible that the criteria associated with ‘Lumpl’ enable us to pick out an object identifiable as Lumpl in a world or time in which there is no object identifiable as Goliath. In this sense, conceptualism is compatible with the claim that there can be contingent identities containing only proper names—even though proper names are rigid designators in the sense of satisfying (PN).

It does not follow, of course, that identity is a contingent relation in conceptualism; and, indeed, quite the opposite is the case. That is, since *reference in conceptualism is directly to objects*, albeit mediated by referential concepts, it is a conceptual truth to say that an object cannot but be the object that it is or that one object cannot be identical with another. In other words, the following wffs:

$$\forall x\forall y(x = y \rightarrow \Box x = y), \quad \forall x\forall y(x \neq y \rightarrow \Box x \neq y)$$

are to be taken as valid theses of conceptualism. This result is the complete opposite, needless to say, from that obtained *on the Platonic view* where *reference is directly to individual concepts* (as independently existing platonic forms) and only indirectly to the objects denoted by these concepts in a given possible world.

### 13 CONCEPTUALISM AND TENSE LOGIC

As forms of conceptual activity, thought and communication are inextricably temporal phenomena, and to ignore this fact in the semantics of a formal representation of such activity is to court possible confusion of the Platonic with the conceptual view of intensionality. Propositions, for example, on the conceptual view, are not abstract entities existing in a platonic realm independently of all conceptual activity. Rather, according to conceptualism, they are really conceptual constructs corresponding to the truth-conditions of our temporally located assertions; and on the present level of analysis where propositional attitudes are not being considered, their status as constructs can be left completely in the metalanguage.

What is also a construction, but which should not be left to the metalanguage, are certain cognitive schemata characterising our conceptual orientation in time and implicit in the form and content of our assertions as mental acts. These schemata, whether explicitly recognised as such or not, are usually represented or modelled in terms of a tenseless idiom (such as



our set-theoretic metalanguage) in which reference can be made to moments or intervals of time (as individuals of a special type); and for most purposes such a representation is quite in order. But to represent them only in this way in a context where our concern is with a perspicuous representation of the form of our assertions as mental acts might well mislead us into thinking that the schemata in question are not essential in conceptualism to the form and content of an assertion after all—the way they are not essential to the form and content of a proposition on the Platonic view. Indeed, even though the cognitive schemata in question can be modelled in terms of a tenseless idiom of moments or intervals of time (as in fact they will be in our set-theoretic metalanguage), they are themselves the conceptually prior conditions that lead to the construction of our referential concepts for moments or intervals of time, and therefore of the very tenseless idiom in which they are subsequently modelled. In this regard, no assumption need be made in conceptualism about the ultimate nature of moments or intervals of time, i.e. whether such entities are really independently existing individuals or only constructions out of the different events that actually occur.

Now since what the temporal schemata implicit in our assertions fundamentally do is enable us to orientate ourselves in time in terms of the distinction between the past, the present, and the future, a more appropriate or perspicuous representation of these schemata is one based upon a system of quantified tense logic containing at least the operators  $\mathcal{P}$ ,  $\mathcal{N}$ ,  $\mathcal{F}$  for ‘it was the case that’, ‘it is now the case that’, and ‘it will be the case that’, respectively. As applied in thought and communication, what these operators correspond to is our ability to refer to what was the case, what is now the case, and what will be the case—and to do so, moreover, without having first to construct referential concepts for moments or intervals of time.

Keeping our analysis as simple as possible, accordingly, let us now understand a language to consist of symbols for common (count) nouns, including always one for ‘individual’, as well as proper names and predicates. Where  $L$  is such a language, the atomic wffs of  $L$  are expressions of the form  $(x = y)$  and  $F(x_1, \dots, x_n)$ , where  $x, y, x_1, \dots, x_n$  are variables and  $F$  is an  $n$ -place predicate in  $L$ . The wffs of  $L$  are then the expressions in every set  $K$  containing the atomic wffs of  $L$  and such that  $\neg\varphi, \mathcal{P}\varphi, \mathcal{N}\varphi, \mathcal{F}\varphi, (\varphi \rightarrow \psi), (\forall xS)\varphi$  are all in  $K$  whenever  $\varphi, \psi \in K, x$  is an individual variable and  $S$  is either a proper name or a common (count) noun symbol in  $L$ . As already noted, where  $S$  is the symbol for ‘individual’, we take  $\forall x\varphi$  to abbreviate  $(\forall xS)\varphi$ , and similarly  $\exists x\varphi$  abbreviates  $\neg(\forall xS)\neg\varphi$ .

In regard to a set-theoretic semantics for these wffs, let us retain the notion of a relational world system  $\langle W, R, E \rangle$  already defined, but with the understanding that the members of  $W$  are now to be the moments of a *local time* (*Eigenzeit*) rather than complete possible worlds, and that the

relation  $R$  of accessibility is the earlier-than relation between the moments of that local time. The only constraint imposed by conceptualism on the structure of  $R$  is that it be a linear ordering of  $W$ , i.e. that  $R$  be asymmetric, transitive and connected in  $W$ . This constraint is based upon the implicit assumption that a local time is always the local time of a continuant.

There is nothing in set theory itself, it should be noted, which directly corresponds to the unsaturated nature of concepts as cognitive capacities; and for this reason we shall once again follow the Carnapian approach and represent concepts as functions from the moments of a local time to the classes of objects falling under the concepts at those times. Naturally, on this approach one and the same type of function will be used to represent the concepts underlying the use of common (count) nouns, proper names, and one-place predicates—despite the conceptual distinctions between them and in the way they account for different aspects of a mental act.

Accordingly, where  $L$  is a language and  $\langle W, R, E \rangle$  is a relational world system, we shall now understand an *interpretation* for  $L$  based upon  $\langle W, R, E \rangle$  to be a function  $I$  on  $L$  such that (1) for each  $n$ -place predicate  $F^n$  in  $L$ ,  $I(F^n)$  is an  $n$ -place predicate intension in  $\langle W, R, E \rangle$ ; (2) for each common (count) noun symbol  $S$  in  $L$ ,  $I(S)$  is a one-place predicate intension in  $\langle W, R, E \rangle$ ; and for the symbol  $S$  for 'individual' in particular,  $I(S)(w) = E(w)$ , for all  $w \in W$ ; and (3) for each proper name  $S$  in  $L$ ,  $I(S)$  is a one-place predicate intension in  $\langle W, R, E \rangle$  such that for some  $d \in \cup_{w \in W} E(w)$ ,  $I(S)(w) \subseteq \{d\}$ , for all  $w \in W$ . (Note that at any given time  $w \in W$ , nothing need, in fact, be identifiable by means of the identity criteria associated with the use of a proper name—though if anything is so identifiable, then it is always the same individual. In this way we trivially validate the thesis (PN) of the preceding section that proper names are rigid designators with respect to the modalities analysable in terms of time.)

By a *referential assignment* in a relational world system  $\langle W, R, E \rangle$ , we now understand a function  $A$  which assigns to each variable  $x$  a member of  $\cup_{w \in W} E(w)$ , hereafter called the *realia* of  $\langle W, R, E \rangle$ . *Realia*, of course, are the objects that exist at some time or other of the local time in question. Referential concepts, at least in the semantics formulated below, do not refer directly to *realia*, but only indirectly (of which more anon); and in this regard *realia* are in tense logic what *possibilia* are in modal logic.

Finally, where  $t$  is construed as *the present moment* of a local time  $\langle W, R, E \rangle$ , i.e.  $t \in W$ ,  $A$  is a referential assignment in  $\langle W, R, E \rangle$  and  $I$  is an interpretation for a language  $L$  based on  $\langle W, R, E \rangle$ , we recursively define with respect to  $I$  and  $A$  *the proposition* (or intension in the sense of the truth-conditions) *expressed by a wff  $\varphi$  of  $L$  when part of an assertion made at  $t$*  as follows:

1.  $\text{Int}_t(x = y, I, A) =$  the  $P \in 2^W$  such that for  $w \in W, P(w) = 1$  iff  $A(x) = A(y)$ ;
2.  $\text{Int}_t(F^n(x_1, \dots, x_n), I, A) =$  the  $P \in 2^W$  such that for  $w \in W, P(w) = 1$  iff  $\langle A(x_1), \dots, A(x_n) \rangle \in I(F^n)(w)$ ;
3.  $\text{Int}_t(\neg\varphi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W, P(w) = 1$  iff  $\text{Int}_t(\varphi, I, A)(w) = 0$ ;
4.  $\text{Int}_t(\varphi \rightarrow \psi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W, P(w) = 1$  iff  $\text{Int}_t(\varphi, I, A)(w) = 0$  or  $\text{Int}_t(\psi, I, A)(w) = 1$ ;
5.  $\text{Int}_t((\forall xS)\varphi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W, P(w) = 1$  iff for all  $d \in E(w)$ , if  $d \in I(S)(w)$ , then  $\text{Int}_t(\varphi, I, A(d/x))(w) = 1$ ;
6.  $\text{Int}_t(\mathcal{P}\varphi, I, A) =$  the  $P \in 2^W$  such that for all  $w \in W, P(w) = 1$  iff  $\text{Int}_t(\varphi, I, A)(u) = 1$ , for some  $u$  such that  $uRw$ ;
7.  $\text{Int}_t(\mathcal{N}\varphi, I, A) =$  the  $P \in 2^W$  such that for all  $w \in W, P(w) = 1$  iff  $\text{Int}_t(\varphi, I, A)(t) = 1$ ; and
8.  $\text{Int}_t(\mathcal{F}\varphi, I, A) =$  the  $P \in 2^W$  such that for all  $w \in W, P(w) = 1$  iff  $\text{Int}_t(\varphi, I, A)(u) = 1$ , for some  $u$  such that  $wRu$ .

The double-indexing involved in this semantics and critically used in clause (7) is to account for the role of the now-operator. It was first given in [Kamp, 1971] and, of course, is particularly appropriate for conceptualism's concern with the semantics of assertions as particular mental acts. That is, as constructed in terms of the truth-conditions for assertions, propositions on the conceptualist's view of intensionality differ from those of the Platonist in being bound to the time at which the assertion in question occurs. For the Platonist, propositions exist independently of time, and therefore of the truth-conditions for assertions as well.

In regard to truth and validity, we shall say, relative to an interpretation  $I$  and referential assignment  $A$  in a local time  $\langle W, R, E \rangle$ , that a wff  $\varphi$  of the language in question is *true* if  $\text{Int}_t(\varphi, I, A)(t) = 1$ , where  $t$  is the present moment of the local time  $\langle W, R, E \rangle$ . The wff  $\varphi$  is said to be *valid* or *tense-logically true*, on the other hand, if for all local time systems  $\langle W, R, E \rangle$ , all  $t \in W$ , all referential assignments  $A$  in  $\langle W, R, E \rangle$ , and all interpretations  $I$  for a language of which  $\varphi$  is a wff,  $\text{Int}_t(\varphi, I, A)(t) = 1$ .

A completeness theorem is forthcoming for this semantics, but we shall not concern ourselves with establishing one in the present essay—especially since the overall logic is rather weak or minimal in the way it accounts for our conceptual orientation in time. Instead, let us briefly examine the problem of referring to *realia* in general, and in particular to past or future objects—i.e. the problem of how the conceptual structure of such a minimal system can either account for such reference or lead to a conceptual development where such an account can be given.

#### 14 THE PROBLEM OF REFERENCE TO PAST AND FUTURE OBJECTS

Our comparison of the status of *realia* in tense logic with *possibilia* in modal logic is especially appropriate, it might be noted, insofar as quantificational reference to either is said to be feasible only *indirectly*—i.e. through the occurrence of a quantifier within the scope of a modal or tense operator (cf. [Prior, 1967, Chapter 8]). The reference to a past individual in ‘Someone did exist who was a King of France’, for example, can be accounted for by the semantics of  $\mathcal{P}(\exists xS)(\exists yT)(x = y)$ , where  $S$  and  $T$  are sortal common noun symbols for ‘person’ and ‘King of France’, respectively. What is apparently not feasible about a *direct* quantificational reference to such objects, on this account, is our present inability to actually confront and apply the relevant identity criteria to objects which do not now exist.

A present ability to identify past or future objects of a given sort, however, is not the same as the ability to actually confront and identify those objects in the present; that is, our existential inability to do the latter is not the same as, and should not be confused with, what is only presumed to be our inability to directly refer to past or future objects. Indeed, the fact is that we can and do make *direct reference* to *realia*, and to past and future objects in particular, and that we do so not only in ordinary discourse but also, and especially, in most if not all of our scientific theories. The real problem is not that we cannot directly refer to past and future objects, but rather how it is that conceptually we come to do so.

One explanation of how this comes to be can be seen in the analysis of the following English sentences:

1. There did exist someone who is an ancestor of everyone now existing.
2. There will exist someone who will have everyone now existing as an ancestor.

Where  $S$  is a sortal common noun symbol for ‘person’ and  $R(x, y)$  is read as ‘ $x$  is an ancestor of  $y$ ’, it is clear that (1) and (2) cannot be represented by:

3.  $\mathcal{P}(\exists xS)(\forall yS)R(x, y)$
4.  $\mathcal{F}(\exists xS)(\forall yS)R(y, x)$ .

For what (3) and (4) represent are the different sentences:

5. There did exist someone who was an ancestor of everyone then existing.
6. There will exist someone who will have everyone then existing as an ancestor.

Of course, if referential concepts that enabled us to refer directly to past and future objects were already available, then the obvious representation of (1) and (2) would be:

$$7. (\exists x \textit{ Past-S})(\forall yS)R(x, y)$$

$$8. (\exists x \textit{ Future-S})(\forall yS)\mathcal{F}R(y, x)$$

where ‘Past-’ and ‘Future-’ are construed as common noun modifiers. (We assume here that the relational ancestor concept is such that  $x$  is an ancestor of  $y$  only at those times when either  $y$  exists and  $x$  did exist, though  $x$  need not still exist at the time in question, or when  $x$  has continued to exist even though  $y$  has ceased to exist. When  $y$  no longer exists as well as  $x$ , we say that  $x$  *was* an ancestor of  $y$ ; and where  $y$  has yet to exist, we say that  $x$  *will be* an ancestor of  $y$ .)

Now although these last analyses are not available in the system of tense logic formulated in the preceding section, nevertheless semantical equivalences for them are. In this regard, note that although the indirect references to past and future objects in (3) and (4) fail to provide adequate representations of (1) and (2), the same indirect references followed by the now-operator succeed in capturing the direct references given in (7) and (8):

$$9. \mathcal{P}(\exists xS)\mathcal{N}(\forall yS)R(x, y)$$

$$10. \mathcal{F}(\exists xS)\mathcal{N}(\forall yS)\mathcal{F}R(y, x).$$

In other words, at least relative to any present tense context, we can in general account for direct reference to past and future objects as follows:

$$(\forall x \textit{ Past-S})\varphi \leftrightarrow \neg\mathcal{P}\neg(\forall xS)\mathcal{N}\varphi$$

$$(\forall x \textit{ Future-S})\varphi \leftrightarrow \neg\mathcal{F}\neg(\forall xS)\mathcal{N}\varphi.$$

These equivalences, it should be noted, cannot be used other than in a present tense context; that is, the above use of the now-operator would be inappropriate when the equivalences are stated within the scope of a past- or future-tense operator, since in that case the direct reference to past or future objects would be from a point of time other than the present. Formally, what is needed in such a case is the introduction of so-called ‘backwards-looking’ operators, such as ‘then’, which can be correlated with occurrences of past or future tense operators within whose scope they lie and which semantically evaluate the wffs to which they are themselves applied in terms of the past or future times already referred to by the tense operators with which they are correlated (cf. [Vlach, 1973] and [Saarinen, 1976]). Backwards-looking operators, in other words, enable us to conceptually return to a past or future time already referred to in a given context in the same way that the now-operator enables us to return to the present. In that regard, their role in the cognitive schemata characterising our conceptual orientation in time

and implicit in each of our assertions is essentially a projection of the role of the now-operator.

We shall not formulate the semantics of these backwards-looking operators here, however; but we note that with their formulation equivalences of the above sort can be established for all contexts of tense logic, past and future as well as present. In any case, it is clear that the fact that conceptualism can account for the development of referential concepts that enable us to refer directly to past or future objects is already implicit in the fact that such references can be made with respect to the present alone. For this already shows that whereas the reference is direct at least in effect, nevertheless the application of any identity criteria associated with such reference will itself be indirect, and in particular, not such as to require a present confrontation, even if only in principle, with a past or future object.

## 15 TIME AND MODALITY

One important feature of the cognitive schemata characterising our conceptual orientation in time and represented in part by quantified tense logic, according to conceptualism, is the capacity they engender in us to form modal concepts having material content. Indeed, some of the first such modal concepts every to be formulated in the history of thought are based precisely upon the very distinction between the past, the present, and the future which is contained in these schemata. For example, the Megaric logician Diodorus is reported as having argued that the possible is that which either is or will be the case, and that therefore the necessary is that which is and always will be the case (cf. [Prior, 1967, Chapter 2]):

$$\Diamond^f \varphi = df \varphi \vee \mathcal{F}\varphi, \quad \Box^f \varphi = df \varphi \wedge \neg \mathcal{F}\neg \varphi.$$

Aristotle, on the other hand, included the past as part of what is possible; that is, for Aristotle the possible is that which either was, is, or will be the case (in what he assumed to be the infinity of time), and therefore the necessary is what is always the case (cf. [Hintikka, 1973]):

$$\Diamond^t \varphi = df \mathcal{P}\varphi \vee \varphi \vee \mathcal{F}\varphi, \quad \Box^t \varphi = df \neg \mathcal{P}\neg \varphi \wedge \varphi \wedge \neg \mathcal{F}\neg \varphi.$$

Both Aristotle and Diodorus, it should be noted, assumed that time is real and not ideal—as also does the socio-biologically based conceptualism being considered here. The temporal modalities indicated above, accordingly, are in this regard intended to be taken as material or metaphysical modalities (of a conceptual realism); and, indeed, they serve this purpose rather well, since in fact they provide a paradigm by which we might better understand what is meant by a material or metaphysical modality. In particular, not only do these modalities contain an explanatory, concrete

interpretation of the accessibility relation between possible worlds (now reconstrued as momentary states of the universe), but they also provide a rationale for the secondary semantics of a metaphysical necessity—since clearly not every possible world (of a given logical space) need ever actually be realised in time (as a momentary state of the universe). Moreover, the fact that the semantics (as considered here) is concerned with concepts and not with independently real material properties and relations (which may or may not correspond to some of these concepts but which can in any case also be considered in a supplementary semantics of conceptual realism) also explains why predicates can be true of objects at a time when those objects do not exist. For concepts, such as that of being an ancestor of everyone now existing, are constructions of the mind and can in that regard be applied to past or future objects no less so than to presently existing objects. In addition, because the intellect is subject to the closure conditions of the laws of compositionality for systematic concept-formation, there is no problem in conceptualism regarding the fact that a concept can be constructed corresponding to every open wff—thereby validating the rule of substitution of wffs for predicate letters.

As a paradigm of a metaphysical modality, on the other hand, one of the defects of Aristotle's notion of necessity is its exclusion of certain situations that are possible in special relativity. For example, relative to the present of a given local time, a state of affairs can come to have been the case, according to special relativity, without its ever actually being the case (cf. [Putnam, 1967]). That is, where  $\mathcal{FP}\varphi$  represents  $\varphi$ 's coming (future) to have been (past) the case, and  $\neg\Diamond^t\varphi$  represents  $\varphi$ 's never actually being the case, the situation envisaged in special relativity might be thought to be represented by:

$$\mathcal{FP}\varphi \wedge \neg\Diamond^t\varphi.$$

This conjunction, however, is incompatible with the linearity assumption of the local time in question; for on the basis of that assumption

$$\mathcal{FP}\varphi \rightarrow \mathcal{P}\varphi \vee \varphi \vee \mathcal{F}\varphi$$

is tense-logically true, and therefore  $\mathcal{FP}\varphi$ , the first conjunct, implies  $\Diamond^t\varphi$ , which contradicts the second conjunct,  $\neg\Diamond^t\varphi$ . The linearity assumption, moreover, cannot be given up without violating the notion of a local time or that of a continuant upon which it is based; and the notion of a continuant, as already indicated, is a fundamental construct of conceptualism. In particular, the notion of a continuant is more fundamental even than that of an event, which (at least initially) in conceptualism is always an occurrence in which one or more continuants are involved. Indeed, the notion of a continuant is even more fundamental in a socio-biologically based conceptualism than the notion of the self as a centre of conceptual activity, and it

is in fact one of the bases upon which the tense-logical cognitive schemata characterising our conceptual orientation in time are constructed.

This is not to say, on the other hand, that in the development of the concept of a self as a centre of conceptual activity we do not ever come to conceive of the ordering of events from perspectives other than our own. Indeed, by a process which Jean Piaget calls decentering, children at the stage of concrete operational thought (7–11 years) develop the ability to conceive of projections from their own positions to that of others in their environment; and subsequently, by means of this ability, they are able to form operational concepts of space and time whose systematic co-ordination results essentially in the structure of projective geometry. Spatial considerations aside, however, and with respect to time alone, the cognitive schemata implicit in the ability to conceive of such projections can be represented in part by means of tense operators corresponding to those already representing the past and the future as viewed from one's own local time. That is, since the projections in question are to be based on actual causal connections between continuants, we can represent the cognitive schemata implicit in such projections by what we shall here call *causal tense operators*, viz.  $\mathcal{P}_c$  for 'it causally was the case that' and  $\mathcal{F}_c$  for 'it causally will be the case that'. Of course, the possibility in special relativity of a state of affairs coming to have been the case without its ever actually being the case is a possibility that should be represented in terms of these operators and not in terms of those characterising the ordering of events within a single local time.

Semantically, in other words, the causal tense operators go beyond the standard tenses by requiring us to consider not just a single local time but a causally connected system of such local times. In this regard, the causal connections between the different continuants upon which such local times are based can simply be represented by a signal relation between the momentary states of those continuants—or rather, and more simply yet, by a signal relation between the moments of the local times themselves, so long as we assume that the sets of moments of different local times are disjoint. (This assumption is harmless if we think of a moment of a local time as an ordered pair one constituent of which is the continuant upon which that local time is based.) The only constraint that should be imposed on such a signal relation is that it be a strict partial ordering, i.e. transitive and asymmetric. Of course, since we assume that there is a causal connection from the earlier to the later momentary states of the same continuant, we shall also assume that the signal relation contains the linear temporal ordering of the moments of each local time in such a causally connected system. (Cf. [Carnap, 1958, Sections 49–50], for one approach to the notion of a causally connected system of local times.) Needless to say, but such a signal relation provides yet another concrete interpretation of the accessibility relation between possible worlds (reconstructed as momentary



states of the universe); and it will be in terms of this relation that the semantics of the causal tense operators will be given.

Accordingly, by a *system of local times* we shall understand a pair  $\langle K, S \rangle$  such that (1)  $K$  is a non-empty set of relational world systems  $\langle W, R, E \rangle$  for which (a)  $R$  is a linear ordering of  $W$  and (b) for all  $\langle W', R', E' \rangle \in K$ , if  $\langle W, R, E \rangle \neq \langle W', R', E' \rangle$ , then  $W$  and  $W'$  are disjoint; and (2)  $S$  is a strict partial ordering of  $\{w : \text{for some } \langle W, R, E \rangle \in K, w \in W\}$  and such that for all  $\langle W, R, E \rangle \in K, R \subseteq S$ . Furthermore, if  $\langle W, R, E \rangle, \langle W', R', E' \rangle \in K, t \in W$ , and  $t' \in W'$ , then  $t$  is said to be *simultaneous* with  $t'$  in  $\langle K, S \rangle$  iff neither  $tSt'$  nor  $t'St$ ; and  $t$  is said to *coincide with*  $t'$  iff for all  $\langle W'', R'', E'' \rangle \in K$  and all  $w \in W''$ , (1)  $w$  is simultaneous with  $t$  in  $\langle K, S \rangle$  iff  $w$  is simultaneous with  $t'$  in  $\langle K, S \rangle$ , and (2)  $tSw$  if  $t'Sw$ .

Now a system  $\langle K, S \rangle$  of local times is said to be *causally connected* iff for all  $\langle W, R, E \rangle, \langle W', R', E' \rangle \in K$ , (1) for all  $t \in W, t' \in W'$ , if  $t$  coincides with  $t'$  in  $\langle K, S \rangle$ , then  $E(t) = E'(t')$ , i.e. the same objects exist at coinciding moments of different local times; and (2) for all  $t, w \in W$ , all  $t', w' \in W'$ , if  $t$  is simultaneous with  $t'$  in  $\langle K, S \rangle$ ,  $w$  is simultaneous with  $w'$  in  $\langle K, S \rangle$ ,  $tRw$  and  $t'R'w'$ , then  $\{(t, u) : tRu \wedge uRw\} \cong \{(t', u) : t'R'u \wedge uR'w'\}$ ; i.e. the structure of time is the same in any two local intervals whose end-points are simultaneous in  $\langle K, S \rangle$ .

Note that although the relation of coincidence in a causally connected system is clearly an equivalence relation, the relation of simultaneity, at least in special relativity, need not even be transitive. This will, in fact, be a consequence of the principal assumption of special relativity, viz. that the signal relation  $S$  of a causally connected system  $\langle K, S \rangle$  has a *finite limiting velocity*; i.e. for all  $\langle W, R, E \rangle, \langle W', R', E' \rangle \in K$  and all  $w \in W$ , if  $w$  does not coincide in  $\langle K, S \rangle$  with any moment of  $W'$ , then there are moments  $u, v$  of  $W'$  such that  $uR'v$  and yet  $w$  is simultaneous with both  $u$  and  $v$  in  $\langle K, S \rangle$  (cf. [Carnap, 1958]). It is, of course, because of this assumption that a state of affairs can come (causal future) to have been (causal past) the case without its ever actually being the case (in the local time in question).

Finally, where

$$[t]_{\langle K, S \rangle} = \{t' : t' \text{ coincides with } t \text{ in } \langle K, S \rangle\},$$

$$W_{\langle K, S \rangle} = \{[t]_{\langle K, S \rangle} : \text{for some } \langle W, R, E \rangle \in K, t \in W\},$$

$$R_{\langle K, S \rangle} = \{([t]_{\langle K, S \rangle}, [w]_{\langle K, S \rangle}) : tSw\},$$

$$E_{\langle K, S \rangle} = \{([t]_{\langle K, S \rangle}, E(t)) : \text{for some } W, R, \langle W, R, E \rangle \in K \text{ and } t \in W\},$$

then  $\langle W_{\langle K, S \rangle}, R_{\langle K, S \rangle}, E_{\langle K, S \rangle} \rangle$  is a relational world system (in which every theorem of **S4** is validated). Accordingly, if  $I$  is an interpretation for a language  $L$  based on  $\langle W_{\langle K, S \rangle}, R_{\langle K, S \rangle}, E_{\langle K, S \rangle} \rangle$ ,  $A$  is a referential assignment in  $\langle W_{\langle K, S \rangle}, R_{\langle K, S \rangle}, E_{\langle K, S \rangle} \rangle$ ,  $\langle W, R, E \rangle \in K$ , and  $t \in W$ , then we recursively define with respect to  $I$  and  $A$  the *proposition expressed by a wff  $\varphi$  of  $L$  when part of an assertion made at  $t$*  (as the present of the local time

$\langle W, R, E \rangle$  which is causally connected in the system  $\langle K, S \rangle$ , in symbols  $\text{Int}_t(\varphi, I, A)$ , exactly as before (in Section 13), except for the addition of the following two clauses:

9.  $\text{Int}_t(\mathcal{P}_c\varphi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W, P(w) = 1$  iff there are a local time  $\langle W', R', E' \rangle \in K$  and moments  $t', u \in W'$  such that  $t$  is simultaneous with  $t'$  in  $\langle K, S \rangle, uSw$ , and  $\text{Int}_{t'}(\varphi, I, A)(u) = 1$ ; and
10.  $\text{Int}_t(\mathcal{F}_c\varphi, I, A) =$  the  $P \in 2^W$  such that for  $w \in W, P(w) = 1$  iff there are a local time  $\langle W', R', E' \rangle \in K$  and moments  $t', u \in W'$  such that  $t$  is simultaneous with  $t'$  in  $\langle K, S \rangle, wSu$  and  $\text{Int}_{t'}(\varphi, I, A)(u) = 1$ .

Except for an invariance with respect to the added parameter  $\langle K, S \rangle$ , validity or tense-logical truth is understood to be defined exactly as before. It is clear of course that although

$$\mathcal{P}\varphi \rightarrow \mathcal{P}_c\varphi, \quad \mathcal{F}\varphi \rightarrow \mathcal{F}_c\varphi$$

are valid, their converses can be invalidated in a causally connected system which has the finite limiting velocity. On the other hand, were we to exclude such systems (as was done in classical physics) and validate the converse of the above wffs as well (as perhaps is still implicit in our common sense framework), then, of course, the causal tense operators would be completely redundant (which perhaps explains why they have no counterparts in natural language). It should perhaps be noted here that unlike the cognitive schemata of the standard tense operators whose semantics is based on a single local time, those represented by the causal tense operators are not such as must be present in one form or another in every act of thought. That is, they are derived schemata, constructed on the basis of those de-centering abilities whereby we are able to conceive of the ordering of events from a perspective other than our own. Needless to say, but the importance and real significance of these derived schemata was unappreciated until the advent of special relativity.

One important consequence of the divergence of the causal from the standard tense operators is the invalidity of

$$\mathcal{F}_c\mathcal{P}_c\varphi \rightarrow \mathcal{P}_c\varphi \vee \varphi \vee \mathcal{F}_c\varphi$$

and therefore the consistency of

$$\mathcal{F}_c\mathcal{P}_c\varphi \wedge \neg\Diamond^t\varphi.$$

Unlike its earlier counterpart in terms of the standard tenses, this last wff of course is the appropriate representation of the possibility in special relativity of a state of affairs coming (in the causal future) to have been the case (in the causal past) without its ever actually being the case (in a given local

time). Indeed, not only can this wff be true at some moment of a local time of a causally connected system, but so can the following wff:

$$[\mathcal{P}_c \diamond^t \varphi \vee \mathcal{F}_c \diamond^t \varphi] \wedge \neg \diamond^t \varphi.$$

Quantification over *realia*, incidentally, finds further justification in special relativity. For just as some states of affairs can come to have been the case (in the causal past of the causal future) without their actually ever being the case, so too there can be things that exist only in the past or future of our own local time, but which nevertheless might exist in a causally connected local time at a moment which is simultaneous with our present. In this regard, reference to such objects as real even if not presently existing would seem hardly controversial—or at least not at that stage of conceptual development where our decentering abilities enable us to construct referential concepts that respect other points of view causally connected with our own.

Finally, it should be noted that whereas the original Diodorean notion of possibility results in the modal logic **S4.3**, i.e. the system **S4** plus the additional thesis

$$\diamond^f \varphi \wedge \diamond^f \psi \rightarrow \diamond^f (\varphi \wedge \psi) \vee \diamond^f (\varphi \wedge \diamond^f \psi) \vee \diamond^f (\psi \wedge \diamond^f \varphi),$$

the same Diodorean notion of possibility, but redefined in terms of  $\mathcal{F}_c$  instead, results in the modal logic **S4**. If we also assume, as is usual in special relativity, that the causal futures of any two moments  $t, t'$  of two local times of a causally connected system  $\langle K, S \rangle$  *eventually* intersect, i.e. that there is a local time  $\langle W, R, E \rangle \in K$  and a moment  $w \in W$  such that  $tSw$  and  $t'Sw$ , then the thesis

$$\mathcal{F}_c \neg \mathcal{F}_c \neg \varphi \rightarrow \neg \mathcal{F}_c \neg \mathcal{F}_c \varphi$$

will be validated, and the Diodorean modality defined in terms of  $\mathcal{F}_c$  will result in the modal system **S4.2** (cf. [Prior, 1967, p. 203]), i.e. the system **S4** plus the thesis

$$\diamond^{fc} \square^{fc} \varphi \rightarrow \square^{fc} \diamond^{fc} \varphi.$$

Many other modal concepts, it is clear, can also be characterised in terms of the semantics of a causally connected system of local times, including, e.g. the notion of something being necessary because of the way the past has been. What is distinctive about them all, moreover, is the unproblematic sense in which they can be taken as material or metaphysical modalities. This may indeed not be all there is to such a modality, but taking account of more will confront us once again with the problem of providing a philosophically coherent interpretation of the secondary semantics for such.

*Indiana University, USA.*

## BIBLIOGRAPHY

- [Bacon, 1980] J. Bacon. Substance and first-order quantification over individual concepts. *J. Symbolic Logic*, **45**, 193–203, 1980.
- [Beth, 1960] E. W. Beth. Extension and Intension. *Synthese*, **12**, 375–379, 1960.
- [Broido, 1976] On the eliminability of *de re* modalities in some systems. *Notre Dame J. Formal Logic*, **17**, 79–88, 1976.
- [Carnap, 1938] R. Carnap. Foundations of logic and mathematics. In *International Encyclopedia of Unified Science*, Vol. 1, Univ. Chicago Press, 1938.
- [Carnap, 1946] R. Carnap. Modalities and quantification. *J. Symbolic Logic*, **11**, 33–64, 1946.
- [Carnap, 1947] R. Carnap. *Meaning and Necessity*. Univ Chicago Press, 1947.
- [Carnap, 1955] R. Carnap. Notes on Semantics. Published posthumously in *Philosophia (Phil. Quant. Israel)*, **2**, 1–54 (1972).
- [Carnap, 1958] R. Carnap. *Introduction to Symbolic Logic and its Applications*. Dover Press, 1958.
- [Cocchiarella, 1975a] N. B. Cocchiarella. Logical atomism, nominalism, and modal logic. *Synthese*, **3**, 23–62, 1975.
- [Cocchiarella, 1975b] N. B. Cocchiarella. On the primary and secondary semantics of logical necessity. *Journal of Philosophical Logic*, **4**, 13–27, 1975.
- [Fine, 1979] K. Fine. Failures of the interpolation lemma in quantified modal logic. *J. Symbolic Logic*, **44**, 201–206, 1979.
- [Geach, 1962] P. Geach. *Reference and Generality*, Cornell Univ. Press, 1962.
- [Gibbard, 1975] A. Gibbard. Contingent identity. *J. Philosophical Logic*, **4**, 187–221, 1975.
- [Hintikka, 1956] J. Hintikka. Identity, variables and inpredicative definitions. *J. Symbolic Logic*, **21**, 225–245, 1956.
- [Hintikka, 1969] J. Hintikka. *Models for Modalities*, Reidel, Dordrecht, 1969.
- [Hintikka, 1973] J. Hintikka. *Time and Necessity*, Oxford University Press, 1973.
- [Hintikka, 1982] J. Hintikka. Is alethic modal logic possible? *Acta Phil. Fennica*, **35**, 227–273, 1982.
- [Kamp, 1971] J. A. W. Kamp. Formal properties of ‘Now’. *Theoria*, **37**, 227–273, 1971.
- [Kamp, 1977] J. A. W. Kamp. Two related theorems by D. Scott and S. Kripke. Xeroxed, London, 1977.
- [Kanger, 1957] S. Kanger. *Provability in Logic*, Univ. of Stockholm, 1957.
- [Kripke, 1959] S. Kripke. A completeness theorem in modal logic. *J. Symbolic Logic*, **24**, 1–14, 1959.
- [Kripke, 1962] S. Kripke. The undecidability of monadic modal quantification theory. *Zeitsch f. Math. Logic und Grundlagen d. Math*, **8**, 113–116, 1962.
- [Kripke, 1963] S. Kripke. Semantical considerations on modal logic, *Acta Philosophica Fennica*, **16**, 83–94, 1963.
- [Kripke, 1971] S. Kripke. Identity and necessity. In M. Munitz, ed. *Identity and Individuation*, New York University Press, 1971.
- [Kripke, 1976] S. Kripke. Letter to David Kaplan and Richmond Thomason. March 12, 1976.
- [Lyons, 1977] J. Lyons. *Semantics*, Cambridge Univ. Press, 1977.
- [McKay, 1975] T. McKay. Essentialism in quantified modal logic. *J. Philosophical Logic*, zbf 4, 423–438, 1975.
- [Montague, 1960] R. M. Montague. Logical necessity, physical necessity, ethics and quantifiers. *Inquiry*, **4**, 259–269, 1960. Reprinted in R. Thomason, ed. *Formal Philosophy*, Yale Univ. Press, 1974.
- [Parks, 1976] Z. Parks. Investigations into quantified modal logic - I. *Studia Lgoica*, **35**, 109–125, 1976.
- [Parsons, 1969] T. Parsons. Essentialism and quantified modal logic. *Philosophical Review*, **78**, 35–52, 1969.
- [Prior, 1967] A. N. Prior. *Past, Present and Future*, Oxford Univ. Press, 1967.

- [Putnam, 1967] H. Putnam. Time and physical geometry. *J. Philosophy*, **64**, 240–247, 1967. Reprinted in *Mathematics, Matter and Method, Phil. Papers*, Vol. 1, Cambridge Univ. Press, 1975.
- [Ramsey, 1960] F. P. Ramsey. In R. B. Braithwaite, ed. *The Foundation of Mathematics*, Littlefield, Adams, Paterson, 1960.
- [Saarinen, 1976] E. Saarinen. Backwards-looking operators in tense logic and in natural language. In J. Hintikka *et al.*, eds. *Essays on Mathematical Logic*, D. Reidel, Dordrecht, 1976.
- [Sellars, 1963] W. Sellars. Grammar and existence: a preface to ontology. In *Science, Perception and Reality*, Routledge and Kegan Paul, London, 1963.
- [Smullyan, 1948] A. Smullyan. Modality and description. *J. Symbolic Logic*, **13**, 31–37, 1948.
- [Thomason, 1969] R. Thomason. Modal logic and metaphysics. In K. Lambert, ed. *The Logical Way of Doing Things*, Yale Univ. Press, 1969.
- [Vlach, 1973] F. Vlach. ‘Now’ and ‘then’: a formal study in the logic of tense anaphora. PhD Dissertation, UCLA, 1973.
- [von Wright, 1951] G. H. von Wright. *An Essay in Modal Logic*, North-Holland, Amsterdam, 1951.



## TENSE AND TIME

### 1 INTRODUCTION

The semantics of tense has received a great deal of attention in the contemporary linguistics, philosophy, and logic literatures. This is probably due partly to a renewed appreciation for the fact that issues involving tense touch on certain issues of philosophical importance (viz., determinism, causality, and the nature of events, of time and of change). It may also be due partly to neglect. Tense was noticeably omitted from the theories of meaning advanced in previous generations. In the writings of both Russell and Frege there is the suggestion that tense would be absent altogether from an ideal or scientifically adequate language. Finally, in recent years there has been a greater recognition of the important role that all of the so-called indexical expressions must play in an explanation of mental states and human behavior. Tense is no exception. Knowing that one's friend *died* is cause for mourning, knowing that he *dies* is just another confirmation of a familiar syllogism.

This article will survey some attempts to make explicit the truth conditions of English tenses, with occasional discussion of other languages. We begin in Section 2 by discussing the most influential early scholarship on the semantics of tense, that of Jespersen, Reichenbach, and Montague. In Section 3 we outline the issues that have been central to the more linguistically-oriented work since Montague's time. Finally, in Section 4 we discuss recent developments in the area of tense logic, attempting to clarify their significance for the study of the truth-conditional semantics of tense in natural language.

### 2 EARLY WORK

#### 2.1 *Jespersen*

The earliest comprehensive treatment of tense and aspect with direct influence on contemporary writings is that of Otto Jespersen. Jespersen's *A Modern English Grammar on Historical Principles* was published in seven volumes from 1909 to 1949. Jespersen's grammar includes much of what we would call semantics and (since he seems to accept some kind of identification between meaning and use) a good deal of pragmatics as well. The

aims and methods of Jespersen's semantic investigations, however, are not quite the same as ours.<sup>1</sup>

First, Jespersen is more interested than we are in cataloging and systematizing the various uses of particular English constructions and less interested in trying to characterize their meanings in a precise way. This leads him to discuss seriously uses we would consider too obscure or idiomatic to bother with. For example, Jespersen notes in the *Grammar* that the expressions of the form *I have got A* and *I had got A* are different than other present perfect and past perfect sentences. *I have got a body*, for example, is true even though there was no past time at which an already existent me received a body. Jespersen suggests *I have in my possession* and *I had in my possession* as readings for *I have got* and *I had got*. And this discussion is considered important enough to be included in his *Essentials of English Grammar*, a one volume summary of the *Grammar*.

Jespersen however does *not* see his task as being merely to collect and classify rare flora. He criticizes Henry Sweet, for example, for a survey of English verb forms that includes such paradigms as *I have been being seen* and *I shall be being seen* on the grounds that they are so extremely rare that it is better to leave them out of account altogether. Nevertheless there is an *emphasis* on cataloging, and this emphasis is probably what leads Jespersen to adhere to a methodological principle that we would ignore; viz., that example sentences should be drawn from published literature wherever possible rather than manufactured by the grammarian. Contemporary linguists and philosophers of language see themselves as investigating fundamental intuitions shared by all members of a linguistic community. For this reason it is quite legitimate for them to produce a sentence and assert without evidence that it is well-formed or ill-formed, ambiguous or univocal, meaningful or unmeaningful. This practice has obvious dangers. Jespersen's methodological scruples, however, provide no real safety. On the one hand, if one limits one's examples to a small group of masters of the language one will leave out a great deal of commonly accepted usage. On the other hand, one can't accept *anything* as a legitimate part of the language just because it has appeared in print. Jespersen himself criticizes a contemporary by saying of his examples that 'these three passages are the only ones adduced from the entire English literature during nearly one thousand years'.

A final respect in which Jespersen differs from the other authors discussed here is his concern with the recent history of the language. Although the *Grammar* aims to be a compendium of contemporary idiom, the history of a construction is recited whenever Jespersen feels that such a discussion might be illuminating about present usage. A good proportion of the discussion of the progressive form, for example, is devoted to Jespersen's

---

<sup>1</sup>By 'ours' we mean those of the authors discussed in the remainder of the article. Some recent work, like that of F. Palmer and R. Huddleston, is more in the tradition of Jespersen than this.



thesis that *I am reading* is a relatively recent corruption of *I am a-reading* or *I am on reading*, a construction that survives today in expressions like *I am asleep* and *I am ashore*. This observation, Jespersen feels, has enabled him to understand the meaning of the progressive better than his contemporaries.<sup>2</sup> In discussing Jespersen's treatment of tense and aspect, no attempt will be made to separate what is original with Jespersen from what is borrowed from other authors. Jespersen's grammar obviously extends a long tradition. See Binnick for a recent survey.<sup>3</sup> Furthermore there is a long list of grammarians contemporaneous with Jespersen who independently produced analyses of tenses. See, for example, Curme, Kruisinga and Poutsma. Jespersen, however, is particularly thorough and insightful and, unlike his predecessors and contemporaries, he continues to be widely read (or at least cited) by linguists and philosophers. Jespersen's treatment of tense and aspect in English can be summarized as follows:

### 2.1.1 *Time*

It is important to distinguish *time* from *tense*. Tense is the linguistic device which is used (among other things) for expressing time relations. For example, *I start tomorrow* is a present tense statement about a future time. To avoid time-tense confusion it is better to reserve the term *past* for time and to use *preterit* and *pluperfect* for the linguistic forms that are more commonly called past tense and past perfect. Time must be thought of as something that can be represented by a straight line, divided by the present moment into two parts: the past and the future. Within each of the two divisions we may refer to some point as lying either before or after the main point of which we are speaking. For each of the seven resulting divisions of time there are *retrospective* and *prospective* versions. These two notions are not really a part of time itself, but have rather to do with the perspective from which an event on the time line is viewed. The prospective present time, for example, is a variety of present that looks forward into the future. In summary, time can be pictured as in Figure 2.1.1. The three divisions marked with *A*'s are past; those marked with *C*'s are future. The short pointed lines at each division indicate retrospective and prospective times.

### 2.1.2 *Tense morphology*

The English verb has only two tenses proper, the present tense and the preterit. There are also two tense phrases, the perfect (e.g., *I have written*) and the pluperfect or anteperfect (e.g., *I had written*). (Modal verbs,

<sup>2</sup>A similar claim is made in Vlach [1981]. For the most part, however, the history of English is ignored in contemporary semantics.

<sup>3</sup>Many of the older grammars have been reprinted in the series *English Linguistics: 1500-1800 (A Collection of Facsimile Reprints)* edited by R.C. Alston and published by Scholar Press Limited, Menston, England in 1967.

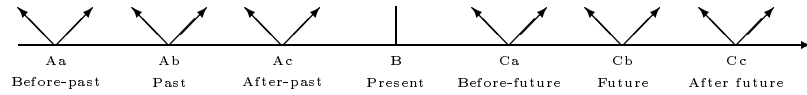


Figure 1.

including *can*, *may*, *must*, *ought*, *shall*, and *will*, cannot form perfects and pluperfects.) Corresponding to each of the four tenses and tense phrases there is an *expanded* (what is more commonly called today the *progressive*) form. For example, *had been writing* is the expanded pluperfect of *write*. It is customary to admit also future and future perfect tenses, as in *I will write* and *I shall have written*. But these constructions lack the fixity of the others. On the one hand, they are often used to express nontemporal ideas (e.g., volition, obstinacy) and on the other hand future time can be indicated in many other ways.

The present tense is primarily used about the present time, by which we mean an interval containing the present moment whose length varies according to circumstances. Thus the time we are talking about in *He is hungry* is shorter than in *None but the brave deserve the fair*. Tense tells us nothing about the duration of that time. The same use of present is found in expressions of intermittent occurrences (*I get up every morning at seven* and *Whenever he calls, he sits close to the fire*). Different uses of the present occur in statements of what might be found at all times by all readers (*Milton defends the liberty of the press in his Areopagitica*) and in expressions of feeling about what is just happening or has just happened (*That's capital!*). The present can also be used to refer to past times. For example, the *dramatic* or *historical* present can alternate with the preterit: *He perceived the surprise, and immediately pulls a bottle out of his pocket, and gave me a dram of cordial*. And the present can play the same role as the perfect in subordinate clauses beginning with *after*: *What happens to the sheep after they take its kidney out?* Present tense can be used to refer to future time when the action described is considered part of a plan already fixed: *I start for Italy on Monday*. The present tense can also refer to future events when it follows *I hope*, *as soon as*, *before*, or *until*.

The perfect is actually a kind of present tense that seems to connect the present time with the past. It is both a retrospective present, which looks upon the present as a result of what happened in the past and an inclusive present, which speaks of a state that is continued from the past into the present time (or at least one that has results or consequences bearing on the present time).

The preterit differs from the perfect in that it refers to some time in the past without telling anything about its connection with the present moment. Thus *Did you finish?* refers to a past time while *Have you finished?* is a question about present status. It follows that the preterit is appropriate with words like *yesterday* and *last year* while the perfect is better with *today*, *until now* and *already*. *This morning* requires a perfect tense when uttered in the morning and a preterit in the afternoon. Often the correct form is determined by context. For example, in discussing a schoolmate's Milton course, *Did you read Samson Agonistes?* is appropriate,

whereas in a more general discussion *Have you read Samson Agonistes?* would be better. In comparing past conditions with present the preterit may be used (*English is not what it was*), but otherwise vague times are not expressed with the preterit but rather by means of the phrase *used to* (*I used to live at Chelsea*). The perfect often seems to imply repetition where the preterit would not. (Compare *When I have been in London*, with *When I was in London*).

The pluperfect serves primarily to denote before-past time or retrospective past, two things which cannot easily be kept apart. (An example of the latter use is *He had read the whole book before noon*.) After *after*, *when*, or *as soon as*, the pluperfect is interchangeable with the preterit.

The expanded tenses indicate that the action or state denoted provides a temporal frame encompassing something else described in the sentence or understood from context. For example, if we say *He was writing when I entered*, we mean that his writing (which may or may not be completed now) had begun, but was not completed, at the moment I entered. In the expanded present the shorter time framed by the expanded time is generally considered to be *very recently*. The expanded tenses also serve some other purposes. In narration simple tenses serve to carry a story forward while expanded tenses have a retarding effect. In other cases expanded tense forms may be used in place of the corresponding simple forms to indicate that a fact is already known rather than new, that an action is incomplete rather than complete or that an act is habitual rather than momentary. Finally, the expanded form is used in two clauses of a sentence to mark the simultaneity of the actions described. (In that case neither really frames the other.)

In addition to the uses already discussed, all the tenses can have somewhat different functions in passive sentences and in indirect speech. They also have uses apparently unrelated to temporal reference. For example, forms which are primarily used to indicate past time are often used to denote unreality, impossibility, improbability or non-fulfillment, as in *If John had arrived on time, he would have won the prize*.<sup>4</sup>

---

<sup>4</sup>From the contemporary perspective we would probably prefer to say here that *had*

### 2.1.3 Tense syntax

In the preceding discussion we started with the English tense forms and inquired about their meanings. Alternatively we can start with various temporal notions and ask how they can be expressed in English. If we do so, several additional facts emerge:

1. The future time can be denoted by present tense (*He leaves on Monday*), expanded present tense (*I am dining with him on Monday*), *is sure to*, *will*, *shall*, *come to* or *get to*.
2. The after-past can be expressed by *would*, *should*, *was to*, *was destined to*, expanded preterit (*They were going out that evening* and *When he came back from the club she was dressing*) or *came to* (*In a few years he came to control all the activity of the great firm*).
3. The before-future can be expressed by *shall have*, *will have* or present (*I shall let you know as soon as I hear from them* or **Wait until the rain stops**).
4. The after-future is expressed by the same means as the future (*If you come at seven, dinner will soon be ready*).
5. Retrospective pasts and futures are not distinguished in English from before-pasts and before-futures. (But retrospective presents, as we have seen, are distinct from pasts. The former are expressed by the perfect, the latter by the preterit.)
6. Prospectives of the various times can be indicated by inserting expressions like *on the point of*, *about to* or *going to*. For example, *She is about to cry* is a prospective present.

## 2.2 Reichenbach

In his general outlook Reichenbach makes a sharp and deliberate break with the tradition of grammarians like Jespersen. Jespersen saw himself as studying the English language by any means that might prove useful (including historical and comparative investigations). Reichenbach saw himself as applying the methods of contemporary logic in a new arena. Thus, while Jespersen's writings about English comprise a half dozen scholarly treatises, Reichenbach's are contained in a chapter of an introductory logic text. (His treatment of tense occupies twelve pages.) Where Jespersen catalogs dozens of uses for an English construction, Reichenbach is content to try to characterize carefully a single use and then to point out that this paradigm does not cover all the cases. While Jespersen uses, and occasionally praises, the *arrived* is a subjunctive preterit which happens to have the same form as a pluperfect.

efforts of antecedent and contemporary grammarians, Reichenbach declares that the state of traditional grammar is hopelessly muddled by its two-millennial ties to a logic that cannot account even for the simplest linguistic forms.

Despite this difference in general outlook, however, the treatment of tenses in Reichenbach is quite similar to that in Jespersen. Reichenbach's chief contribution was probably to recognize the importance of the distinction between what he calls the *point of the event* and the *point of reference* (and the relative unimportance and obscurity of Jespersen's notions of prospective and retrospective time.) In the sentence *Peter had gone*, according to Reichenbach, the point of the event is the time when Peter went. The point of reference is a time between this point and the point of speech, whose exact location must be determined by context. Thus Reichenbach's account of the past perfect is very similar to Jespersen's explanation that the past perfect indicates a 'before past' time. Reichenbach goes beyond Jespersen, however, in two ways.

First, Reichenbach is a little more explicit about his notion of reference times than is Jespersen about the time of which we are speaking. He identifies the reference time in a series of examples and mentions several rules that might be useful in determining the reference time in other examples. Temporally specific adverbials like *yesterday*, *now* or *November 7, 1944*, for example, are said to refer to the reference point. Similarly, words like *when*, *after*, and *before* relate the reference time of an adjunct clause to that of the main clause. And if a sentence does not say anything about the relations among the reference times of its clauses, then every clause has the same point of reference.

Second, Reichenbach argues that the notion of reference time plays an important role in *all* the tenses. The present perfect, for example, is distinguished by the fact that the event point is before the point of reference and the point of reference coincides with the point of speech. (So *I have seen Sharon* has the same meaning as *Now I have seen Sharon*.) In general, each tense is determined by the relative order of the point of event (*E*), the point of speech (*S*), and the point of reference (*R*). If *R* precedes *S* we have a kind of past tense, if *S* precedes *R* we have a kind of future tense and if *R* coincides with *S* we have a kind of present. This explains Jespersen's feeling that the simple perfect is a variety of the present. Similarly the labels 'anterior', 'posterior' and 'simple' indicate that *E* precedes, succeeds or coincides with *R*. The account is summarized in the following table.

Each of the tenses on this table also has an expanded form which indicates, according to Reichenbach, that the event covers a certain stretch of time.

Notice that the list of possible tenses is beginning to resemble more closely the list of tenses realized in English. According to Jespersen there are seven divisions of time, each with simple, retrospective and prospective versions. This makes twenty-one possible tenses. According to Reichenbach's scheme

Structure	New Name	Traditional Name
$E\_R\_S$	Anterior past	Past perfect
$E, R\_S$	Simple past	Simple past
$R\_E\_S$		
$R\_S, E$	Posterior past	_____
$R\_S\_E$		
$E\_S, R$	Anterior present	Present perfect
$S, R, E$	Simple present	Present
$S, R\_E$	Posterior present	Simple future
$S\_E\_R$		
$S, E\_R$	Anterior future	Future perfect
$E\_S\_R$		
$S\_R, E$	Simple future	Simple future
$S\_R\_E$	Posterior future	_____

there should be thirteen possible tenses, corresponding to the thirteen orderings of  $E$ ,  $S$ , and  $R$ . Looking more closely at Reichenbach, however, we see that the *tense* of a sentence is determined only by the relative order of  $S$  and  $R$ , and the *aspect* by the relative order of  $R$  and  $E$ . Since there are three possible orderings of  $S$  and  $R$ , and independently three possible orderings of  $R$  and  $E$ , there are really only nine possible complex tenses (seven of which are actually realized in English).<sup>5</sup>

Finally, Reichenbach acknowledges that actual language does not always keep to the scheme set forth. The expanded forms, for example, sometimes indicate repetition rather than duration: *Women are wearing larger hats this year*. And the present perfect is used to indicate that the event has a certain duration which reaches up to the point of speech: *I have lived here for ten years*.

### 2.3 Montague

Despite Reichenbach's rhetoric, it is probably Montague, rather than Reichenbach, who should be credited with showing that modern logic can be fruitfully applied to the study of natural language. Montague actually had very little to say about tense, but his writings on language have been very influential among those who do have something to say. Two general principles underlie Montague's approach.

<sup>5</sup>There are actually only six English tense constructions on Reichenbach's count, because two tenses are realized by one construction. The simple future is ambiguous between  $S, R\_E$ , as in *Now I shall go* or  $S\_R, E$ , as in *I shall go tomorrow*. Reichenbach suggests that, in French the two tenses may be expressed by different constructions: *je vais voir* and *je verrai*.

- (1a) Compositionality. The meaning of an expression is determined by the meaning of its parts.
- (1b) Truth conditions. The meaning of a declarative sentence is something that determines the conditions under which that sentence is true.

Neither of these principles, of course, is original with Montague, but it is Montague who shows how these principles can be used to motivate an explicit account of the semantics of particular English expressions.

Initially, logic served only as a kind of paradigm for how this can be done. One starts with precisely delineated sets of *basic expressions* of various categories. *Syntactic rules* show how complex expressions can be generated from the basic ones. A class of permissible *models* is specified, each of which assigns interpretations to the basic expressions. *Rules of interpretation* show how the interpretation of complex expressions can be calculated from the interpretations of the expressions from which they are built.

The language of classical predicate logic, for example, contains predicates, individual variables, quantifiers, sentential connectives, and perhaps function symbols. Generalizations of this logic are obtained by adding additional expressions of these categories (as is done in modal and tense logic) or by adding additional categories (as is done in higher order logics). It was Montague's contention that if one generalized enough, one could eventually get English itself. Moreover, clues to the *direction* this generalization should take are provided by modal and tense logic. Here sentences are interpreted by functions from possible worlds (or times or *indices* representing aspects of context) to truth values. English, for Montague, is merely an exceedingly baroque intensional logic. To make this hypothesis plausible, Montague constructed, in [1970; 1970a] and [1973], three 'fragments' of English of increasing complexity. In his final fragment, commonly referred to as PTQ, Montague finds it convenient to show how the expressions can be translated into an already-interpreted intensional logic rather than to specify an interpretation directly. The goal is now to find a translation procedure by which every expression of English can be translated into a (comparatively simple) intensional logic.

We will not attempt here to present a general summary of PTQ. (Readable introductions to Montague's ideas can be found in Montague [1974] and Dowty [1981].) We will, however, try to describe its treatment of tense. To do so requires a little notation. Montague's intensional logic contains tense operators **W** and **H** meaning roughly *it will be the case that* and *it was the case that*. It also contains an operator  $\wedge$  that makes it possible to refer to the intension of an expression. For example, if *a* is an expression referring to the object **a**, then  $\wedge a$  denotes the function that assigns **a** to every pair of a possible world *w* and a time *t*.

Among the expressions of English are *terms* and *intransitive verb phrases*. An intransitive verb phrase  $B$  is translated by an expression  $\mathbf{B}'$  which denotes a function from entities to truth values. (That is,  $\mathbf{B}'$  is of type  $\langle e, t \rangle$ .) A term  $A$  is translated by an expression  $\mathbf{A}'$  which denotes a function whose domain is intensions of functions from entities to truth values and whose range is truth values. (That is,  $\mathbf{A}'$  is of type  $\langle \langle s, \langle e, t \rangle, t \rangle \rangle$ .) Tense and negation in PTQ are treated together. There are six ways in which a term may be combined with an intransitive verb phrase to form a sentence. These generate sentences in the present, future, present perfect, negated present, negated future and negated present perfect forms. The rules of translation corresponding to these six constructions are quite simple. If  $B$  is an intransitive verb phrase with translation  $\mathbf{B}'$  and  $A$  is a term with translation  $\mathbf{A}'$  then the translations of the six kinds of sentences that can be formed by combining  $A$  and  $B$  are just  $\mathbf{A}'(\wedge \mathbf{B}')$ ,  $\mathbf{WA}'(\wedge \mathbf{B}')$ ,  $\mathbf{HA}'(\wedge \mathbf{B}')$ ,  $\neg \mathbf{A}'(\wedge \mathbf{B}')$ ,  $\neg \mathbf{WA}'(\wedge \mathbf{B}')$  and  $\neg \mathbf{HA}'(\wedge \mathbf{B}')$ .

A simple example will illustrate. Suppose that  $A$  is *Mary* and that  $B$  is *sleeps*. The future tense sentence *Mary will sleep* is assigned translation  $\mathbf{WMary}(\wedge \mathbf{sleeps})$ .  $\mathbf{Mary}$  denotes that function which assigns 'true' to a property  $P$  in world  $w$  at time  $t$  if and only if Mary has  $P$  in  $w$  at  $t$ . The expression  $\wedge \mathbf{sleeps}$  denotes the property of sleeping, i.e. the function  $f$  from indices to functions from individuals to truth values such that  $f(\langle w, t \rangle)(a) = \text{'true'}$  if and only if  $a$  is an individual who is asleep in world  $w$  at time  $t$  (for any world  $w$ , time  $t$ , and individual  $a$ ). Thus  $\mathbf{Mary}(\wedge \mathbf{sleeps})$  will be true at  $\langle w, t \rangle$  if and only if Mary is asleep in  $w$  at  $t$ . Finally, the sentence  $\mathbf{WMary}(\wedge \mathbf{sleeps})$  is true in a world  $w$  at a time  $t$  if and only if  $\mathbf{Mary}(\wedge \mathbf{sleeps})$  is true at some  $\langle w, t' \rangle$ , where  $t'$  is a later time than  $t$ .

This treatment is obviously crude and incomplete. It was probably intended merely as an illustration of how tense *might* be handled within Montague's framework. Nevertheless, it contains the interesting observation that the past tense operator found in the usual tense logics corresponds more closely to the present perfect tense than it does to the past. In saying *John has kissed Mary* we seem to be saying that there was *some* time in the past when *John kisses Mary* was true. In saying *John kissed Mary*, we seem to be saying that *John kisses Mary* was true at *the* time we happen to be talking about. This distinction between *definite* and *indefinite* past times was pointed out by Jespersen, but Jespersen does not seem to have thought it relevant to the distinction between present perfect and past.

Reichenbach's use of both event time and reference time, leading to a three-dimensional logic, may suggest that it will not be easy to add the past tenses to a PTQ-like framework. However, one of the differences between Reichenbach's reference time and event time seems to be that the former is often fixed by an adverbial clause or by contextual information whereas the latter is less often so fixed. So it is approximately correct to say that the reference time is determinate whereas the event time is indetermi-



nate. This may help explain the frequent remarks that only two times are needed to specify the truth conditions of all the tenses. In one sense these remarks are wrong.  $S$ ,  $R$  and  $E$  all play essential roles in Reichenbach's explanation of the tenses. But only  $S$  and  $R$  ever need to be extracted from the context. All that we need to know about  $E$  is its position relative to  $R$  and this information is contained in the sentence itself. Thus a tense logic following Reichenbach's analysis could be two-dimensional, rather than three-dimensional. If  $s$  and  $r$  are the points of speech and reference, for example, we would have  $(s, r) \models \text{PASTPERFECT}(A)$  if and only if  $r < s$  and, for some  $t < r$ ,  $t \models A$ . (See Section 4 below.)

Still, it seems clear that the past tenses cannot be added to PTQ without adding something like Reichenbach's point of reference to the models. Moreover, adherence the idea that there should be a separate way of combining tenses and intransitive verb phrases for every negated and unnegated tense would be cumbersome and would miss important generalizations. Montague's most important legacies to the study of tense were probably his identification of meaning with truth conditions, and his high standards of rigor and precision. It is striking that Jespersen, Reichenbach and Montague say successively less about tense with correspondingly greater precision. A great deal of the contemporary work on the subject can be seen as an attempt to recapture the insights of Jespersen without sacrificing Montague's precision.

### 3 CONTEMPORARY VIEWS

In Sections 3.1 and 3.2 below we outline what seem to us to be two key issues underlying contemporary research into the semantics of tense. The first has to do with whether tense should be analyzed as an operator or as something that refers to particular time or times; this is essentially a type-theoretic issue. The second pertains to a pair of truth-conditional questions which apparently are often confused with the type-theoretic ones: (i) *does the semantics of tense involve quantification over times, and if so how does this quantification arise?*, and (ii) *to what extent is the set of times relevant to a particular tensed sentence restricted or made determinate by linguistic or contextual factors?* Section 3.3 then outlines how contemporary analytical frameworks have answered these questions. Finally, Section 3.4 examines in more detail some of the proposals which have been made within these frameworks about the interpretation of particular tenses and aspects.

#### 3.1 *Types for Tense*

The analyses of Reichenbach and Montague have served as inspiration for two groups of theorists. Montague's approach is the one more familiar

from traditional tense logics developed by Prior and others. The simplest non-syncategorematic treatment of tense which could be seen as essentially that of Montague would make tenses propositional operators, expressions of type  $\langle\langle s, t \rangle, t\rangle$  or  $\langle\langle s, t \rangle, \langle s, t \rangle\rangle$ , that is, either as functions from propositions to truth values or as functions from propositions to propositions (where propositions are taken to be sets of world-time pairs). For example, the present perfect might have the following interpretation:

- (2) **PrP** denotes that function  $f$  from propositions to propositions such that, for any proposition  $p$ ,  $f(p)$  = the proposition  $q$ , where for any world  $w$  and time  $t$ ,  $q(\langle w, t \rangle)$  = ‘true’ iff for some time  $t'$  preceding  $t$ ,  $p(\langle w, t' \rangle)$  = ‘true’.

Two alternative, but closely related, views would take tense to have the type of a verb phrase modifier  $\langle\langle s, \langle e, t \rangle \rangle, \langle e, t \rangle\rangle$  ([Bäuerle, 1979; Kuhn, 1983]) or as a ‘mode of combination’ in  $\langle\text{type}(\text{TERM}), \langle\langle s, \langle e, t \rangle \rangle, t \rangle\rangle$  or  $\langle\langle s, \langle e, t \rangle \rangle, \langle\text{type}(\text{TERM}), t \rangle\rangle$ . We will refer to these approaches as representative of the *operator* view of tense.

The alternative approach is more directly inspired by Reichenbach’s views. It takes the semantics of tense to involve reference to particular times. This approach is most thoroughly worked out within the framework of Discourse Representation Theory (DRT; [Kamp, 1983; Kamp and Rohrer, 1983; Hinrichs, 1986; Partee, 1984]), but for clarity we will consider the type-theoretic commitments of the neo-Reichenbachian point of view through the use of a Predicate Calculus-like notation. We may take a tense morpheme to introduce a free variable to which a time can be assigned. Depending on which tense morpheme is involved, the permissible values of the variable should be constrained to fall within an appropriate interval. For example, the sentence *Mary slept* might have a logical form as in (3).

- (3) **PAST**( $t$ ) & **AT**( $t$ , **sleeps**(**Mary**)).

With respect to an assignment  $g$  of values to variables, (3) should be true if and only if  $g(t)$  is a time that precedes the utterance time and one at which Mary sleeps. On this approach the semantics of tense is analogous to that of pronouns, a contention defended most persuasively by Partee.

A more obviously Reichenbachian version of this kind of analysis would introduce more free variables than simply  $t$  in (3). For example, the pluperfect *Mary had slept* might be rendered as in (4):

- (4) **PAST**( $r$ ) &  $t < r$  & **AT**( $t$ , **sleeps**(**Mary**)).

This general point of view could be spelled out in a wide variety of ways. For example, times might be taken as arguments of predicates, or events and states might replace times. We refer to this family of views as *referential*.

### 3.2 Quantification and determinacy

#### 3.2.1 Quantification

In general, the operator theory has taken tense to involve quantification over times. Quantification is not an inherent part of the approach, however; one might propose a semantics for the past tense of the following sort:

- (5)  $(r, u) \models \mathbf{PAST}(S)$  iff  $r < u$  and  $(r, r) \models S$ .

Such an analysis of a non-quantificational past tense might be seen as especially attractive if there are other tense forms that are essentially quantificational. An operator-based semantics would be a natural way to introduce this quantification, and in the interest of consistency one might then prefer to treat all tenses as operators—just as PTQ argues that all NP's are quantifiers because some are inherently quantificational. On the other hand, if no tenses are actually quantificational it might be preferable to utilize a less powerful overall framework.

The issue of quantification for the referential theory of tense is not entirely clear either. If there are sentences whose truth conditions must be described in terms of quantification over times, the referential theory cannot attribute such quantification to the tense morpheme. But this does not mean that such facts are necessarily incompatible with the referential view. Quantification over times may arise through a variety of other, more general, means. Within DRT and related frameworks, several possibilities have been discussed. The first is that some other element in the sentence may bind the temporal variable introduced by tense. An adverb of quantification like *always*, *usually*, or *never* would be the classical candidate for this role.

- (6) When it rained, it always poured.

- (7)  $\forall t[(\mathbf{PAST}(t) \ \& \ \mathbf{AT}(t, \mathbf{it-rains})) \rightarrow (\mathbf{PAST}(t) \ \& \ \mathbf{AT}(t, \mathbf{it-pours}))]$ .

DRT follows Lewis [1975] in proposing that *always* is an unselective universal quantifier which may bind any variables present in the sentence. Hinrichs and Partee point out that in some cases it may turn out that a variable introduced by tense is thus bound; their proposals amount to assigning (6) a semantic analysis along the lines of (7).

The other way in which quantification over times may arise in referential analyses of tense is through some form of default process. The most straightforward view along these lines proposes that, in the absence of explicit quantificational adverbs, the free variable present in a translation like (3), repeated here, is subject to a special rule that turns it into a quantified formula like (8):

(3) **PAST**( $t$ ) & AT( $t$ , **sleeps**(Mary)).

(8)  $\exists t$  [ **PAST**( $t$ ) & AT( $t$ , **sleeps**(Mary))].

This operation is referred to as *existential closure* by Heim; something similar is proposed by Parsons [1995]. It is also possible to get the effect of existential quantification over times through the way in which the truth of a formula is defined. This approach is taken by DRT as well as Heim [1982, Ch. III]. For example, a formula like (3) would be true with respect to a model  $M$  if and only if there is *some* function  $g$  from free variables in (3) to appropriate referents in  $M$  such that  $g(t)$  precedes the utterance time in  $M$  and  $g(t)$  is a time at which Mary is asleep in  $M$ .

To summarize, we may say that one motivation for the operator theory of tense comes from the view that some tense morphemes are inherently quantificational. The referential analysis, in contrast, argues that all examples of temporal quantification are to be attributed not to tense but to independently needed processes.

### 3.2.2 *Determinacy*

An issue which is often not clearly distinguished from questions of the type and quantificational status of tense is that of the determinacy or definiteness of tense. Classical operator-based tense logics treat tense as all but completely indeterminate: a past tense sentence is true if and only if the untensed version is true at *any* past time. On the other hand, Reichenbach's referential theory seemingly considers tense to be completely determinate: a sentence is true or false with respect to the *particular* utterance time, reference time, and event time appropriate for it. However, we have already seen that a referential theory might allow that a time variable can be bound by some quantificational element, thus rendering the temporal reference less determinate. Likewise, we have seen that an operator-based theory may be compatible with completely determinate temporal reference, as in (5). In this section, we would like to point out how varying degrees of determinacy can be captured within the two systems.

If temporal reference is fully indeterminate, it is natural to adopt an operator view: **PAST**( $B$ ) is true at  $t$  if and only if  $B$  is true at some  $t' < t$ . A referential theory must propose that in every case the time variable introduced by tense is bound by some quantificational operator (or effectively quantified over by default, perhaps merely through the effects of the truth definition). In such cases it seems inappropriate to view the temporal parameters as 'referring' to times.

If temporal reference is fully determinate, the referential theory need make no appeal to any ancillary quantification devices. The operator theory may use a semantics along the lines of (3). Alternatively, tense might be seen as an ordinary quantificational operator whose domain of quantification has

been severely restricted. We might implement this idea as follows: Suppose that each tense morpheme bears an index, as *Mary PAST<sub>3</sub>sleeps*. Sentences are interpreted with respect to a function  $R$  from indices to intervals. (The precedence order is extended from instants to intervals and instants in the appropriate way, with  $<$  indicating ‘completely precedes’.) The formula in (9a) would then have the truth conditions of (9b).

(9a) **PAST<sub>3</sub> (sleeps(Mary))**.

(9b)  $(R, u) \models \mathbf{PAST}_3(\mathbf{sleeps}(\mathbf{Mary}))$  iff for some time  $t \in R(3)$ ,  $t < u$  and  $(R, t) \models \mathbf{sleeps}(\mathbf{Mary})$ .

Plainly,  $R$  in (9b) is providing something very similar to that of the reference time in Reichenbach’s system. This can be seen by the fact that the identity of  $R(3)$  should be fixed by temporal adverbs like *yesterday*, as in *Yesterday, Mary slept*.

Finally, we should examine what could be said about instances of tense which are partially determinate. The immediately preceding discussion makes it clear what the status of such examples would be within an operator account; they would simply exemplify restricted quantification ([Bennett and Partee, 1972; Kuhn, 1979]). Instead of the analysis in (9), we would propose that  $R$  is a function from indices to sets of intervals, and give the truth conditions as in (10).

(10)  $(R, u) \models \mathbf{PAST}_3(\mathbf{sleeps}(\mathbf{Mary}))$  iff for some time  $t \in R(3)$ ,  $t < u$  and  $(R, t) \models \mathbf{sleeps}(\mathbf{Mary})$ .

According to (10), (9a) is true if and only if Mary was asleep at some past time which is within the set of contextually relevant past times. Temporal quantification would thus be seen as no different from ordinary nominal quantification, as when *Everyone came to the party* is taken to assert that everyone *relevant* came to the party.

Referential analyses of tense would have to propose that partial determinacy arises when temporal variables are bound by restricted quantifiers. Let us consider a Reichenbach-style account of *Mary slept* along the lines of (11).

(11)  $\exists t [\mathbf{PAST}(r) \ \& \ t \in r \ \& \ \mathbf{AT}(t, \mathbf{sleeps}(\mathbf{Mary}))]$ .

The remaining free variable in (11), namely  $r$ , will have to get its value (the reference *set*) from the assignment function  $g$ . The formula in (11) has  $t \in r$  where Reichenbach would have  $t = r$ ; the latter would result in completely determinate semantics for tense, while (11) results in restricted quantification. The sentence is true if and only if Mary slept during some past interval contained in  $g(r)$ .

The only difference between (10) and (11) is whether the quantificational restriction is represented in the translation language as a variable, the  $r$  in (11), or as a special index on the operator, the subscripted  $\beta$  in (10). In each case, one parameter of interpretation must be some function which identifies the set of relevant times for the quantification. In (11), it is the assignment function,  $g$ , while in (10) it is  $R$ . Clearly at this point the differences between the two theories are minor. To summarize, we need to distinguish three closely related ways in which theories of tense may differ: (i) They may take tense to be an operator or to introduce elements which refer to times; (ii) they may involve quantification over times through a considerable variety of means—the inherent semantics of tense itself, the presence of some other quantificational element within the sentence, or a default rule; and (iii) they may postulate that the temporal reference of sentences is fully determinate, fully indeterminate, or only partially determinate.

### 3.3 *Major contemporary frameworks*

Most contemporary formal work on the semantics of tense takes place within two frameworks: Interval Semantics and Discourse Representation Theory. In this section we describe the basic commitments of each of these, noting in particular how they settle the issues discussed in 3.1 and 3.2 above. We will then consider in a similar vein a couple of other influential viewpoints, those of Situation Semantics [Cooper, 1986] and the work of Enç [1986; 1987].

By *Interval Semantics* we refer to the framework which has developed out of the Intensional Logic of Montague's PTQ. There are a number of implementations of a central set of ideas; for the most part these differ in fairly minor ways, such as whether quantification over times is to be accomplished via operators or explicit quantifiers. The key aspects of Interval Semantics are: (i) the temporal part of the model consists of set  $I$  of *intervals*, the set of open and closed intervals of the reals, with precedence and temporal overlap relations defined straightforwardly; (ii) the interpretation of sentences depends on an *evaluation interval* or *event time*, an *utterance time*, and perhaps a *reference interval* or set of reference intervals; (iii) interpretation proceeds by translating natural language sentences into some appropriate higher-order logic, typically an intensional  $\lambda$ -calculus; and (iv) tenses are translated by quantificational operators or formulas involving first-order quantification to the same effect. The motivation for (i) comes initially from the semantics for the progressive, a point which we will see in Section 3.4 below. We have already examined the motivation for (ii), though in what follows we will see more clearly what issues arise in trying to understand the relationship between the reference interval and the evaluation interval. Points (iii) and (iv) are implementation details with which we will not much concern ourselves.

From the preceding, it can be seen what claims Interval Semantics makes

concerning the issues in 3.1 and 3.2. Tense has the type of an operator. It is uniformly quantificational, but shows variable determinacy, depending on the nature of the reference interval or intervals.

*Discourse Representation Theory* is one of a number of theories of *dynamic interpretation* to be put forth since the early 1980's; others include File Change Semantics [Heim, 1982] and Dynamic Montague Grammar [Groenendijk and Stokhof, 1990]. What the dynamic theories share is a concern with the interpretation of multi-sentence texts, concentrating on establishing means by which information can be passed from one sentence to another. The original problems for which these theories were designed had to do with nominal anaphora, in particular the relationships between antecedents and pronouns in independent sentences like (12) and donkey sentences like (13).

(12) A man walked in. He sat down.

(13) When a man walks in, he always sits down.

Of the dynamic theories, by far the most work on tense has taken place within DRT. It will be important over time to determine whether the strengths and weaknesses of DRT analyses of tense carry over to the other dynamic approaches.

As noted above, work on tense within DRT has attempted to analogize the treatment of tense to that of nominal anaphora. This has resulted in an analytical framework with the following general features: (i) the temporal part of the model consists of a set of *eventualities* (events, processes, states, etc.), and possibly of a set of intervals as well; (ii) the semantic representation of a discourse (or sub-part thereof) contains explicit variables ranging over to reference times, events, and the utterance time; (iii) interpretation proceeds by building up a *Discourse Representation Structure* (DRS), a partial model consisting of a set of objects (*discourse markers*) and a set of conditions specifying properties of and relations among them; the discourse is true with respect to a model  $M$  if and only if the partial model (DRS) can be embedded in the full model  $M$ ; (iv) tenses are translated as conditions on discourse markers representing events and/or times. For example, consider the discourse in (14).

(14) Pedro entered the kitchen. He took off his coat.

We might end up with discourse markers representing Pedro ( $x$ ), the kitchen ( $y$ ), the coat ( $z$ ), the event of entering the kitchen ( $e_1$ ), the event of taking off the coat ( $e_2$ ), the utterance time ( $u$ ), the reference time for the first sentence ( $r_1$ ) and the reference time for the second sentence ( $r_2$ ). The DRS would contain at least the following conditions: **Pedro= $x$ , kitchen( $y$ ), coat( $z$ ), entering( $e_1, x, y$ ), taking-off( $e_2, x, z$ ),  $r_1 < u$ ,  $r_2 < u$ ,  $r_1 < r_2$ ,  $e_1 \circ$**

$r_1$ , and  $e_2 \circ r_2$  (where  $\circ$  represents temporal overlap). The algorithms for introducing conditions may be rather complex, and typically are sensitive to the aspectual class of the eventualities represented (that is, whether they are events, processes, states, etc.).

DRT holds a referential theory of tense, treating it via discourse markers plus appropriate conditions. It therefore maintains that tense is not inherently quantificational, and that any quantificational force which is observed must come from either an independent operator, as with (6), or default rule. Given the definition of truth mentioned above, tense will be given a default existential quantificational force—the DRS for (14) will be true if there is *some* mapping from discourse markers to entities in the model satisfying the conditions. The DRT analysis of tense also implies that temporal reference is highly determinate, since the events described by a discourse typically must overlap temporally with a contextually determined reference time.

Closely related to the DRT view of tense are a pair of *indexical* theories of tense. The first is developed by Cooper within the framework of *Situation Semantics* (Barwise and Perry [1983]). Situation Semantics constructs objects known as *situations* or *states of affairs* set-theoretically out of properties, relations, and individuals (including space-time locations). Let us say that the situation of John loving Mary is represented as  $\langle l, \langle \langle \mathbf{love}, \mathbf{John}, \mathbf{Mary} \rangle, 1 \rangle \rangle$ ,  $l$  being a spatiotemporal location and 1 representing ‘truth’. A set of states of affairs is referred to as a *history*, and it is the function of a sentence to describe a history. A simple example is given in (15).

- (15) *John loved Mary* describes a history  $h$  with respect to a spatiotemporal location  $l$  iff  $\langle l, \langle \langle \mathbf{love}, \mathbf{John}, \mathbf{Mary} \rangle, 1 \rangle \rangle \in h$ .

Unless some theory is given to explain how the location  $l$  is arrived at, a semantics like (15) will of course not enlighten us much as to the nature of tense. Cooper proposes that the location is provided by a *connections function*; for our purposes a connections function can be identified with a function from words to individuals. When the word is a verb, a connections function  $c$  will assign it a spatiotemporal location. Thus,

- (16) *John loved Mary* describes a history  $h$  with respect to a connections function  $c$  iff  $\langle c(\mathit{loved}), \langle \langle \mathbf{love}, \mathbf{John}, \mathbf{Mary} \rangle, 1 \rangle \rangle \in h$ .

Cooper’s theory is properly described as an ‘indexical’ approach to tense, since a tensed verb directly picks out the location which the sentence is taken to describe.<sup>6</sup>

---

<sup>6</sup>Unlike ordinary indexicals, verbs do not refer to the locations which they pick out. The verb *loved* still denotes the relation **love**.



Enç's analysis of tense is somewhat similar to Cooper's. She proposes that tense morphemes refer to intervals. For example, the past tense morpheme *-ed* might refer, at an utterance time  $u$ , to the set of moments preceding  $u$ . For Enç, a verb is a semi-indexical expression, denoting a contextually relevant subrelation of the relation which it is normally taken to express—e.g., any occurrence of *kiss* will denote a subset of  $\{\langle x, y \rangle: x \text{ kisses } y \text{ (at some time)}\}$ . Tense serves as one way of determining which subrelation is denoted. The referent of a verb's tense morpheme serves to constrain the denotation of the verb, so that, for instance, the verb *kissed* must denote a set of pairs of individuals where the first kissed the second during the past, i.e. during the interval denoted by the tense.

- (17) *kissed* denotes a (contextually relevant) subset of  $\{\langle x, y \rangle: \text{for some } t \in \text{-ed}, x \text{ kissed } y \text{ at } t\}$ .

In (17), *-ed* is the set of times denoted by *-ed*, i.e. that set of times preceding the utterance time.

Both Enç's theory and the Situation Semantics approach outlined above seem to make the same commitments on the issues raised in Sections 3.1 and 3.2 as DRT. Both consider tense to be non-quantificational and highly determinate. They are clearly referential theories of tense, taking its function to be to pick out a particular time with respect to which the eventualities described by the sentence are temporally located.

### 3.4 *The compositional semantics of individual tenses and aspects*

Now that we have gone through a general outline of several frameworks which have been used to semantically analyze tense in natural language, we turn to seeing what specific claims have been made about the major tenses (present, past, and future) and aspects (progressive and perfect) in English.

#### 3.4.1 *Tense*

**Present Tense.** In many contemporary accounts the semantic analysis of the present underlies that of all the other tenses.<sup>7</sup> But despite this allegedly fundamental role, the only use of the present that seems to have been treated formally is the 'reportive' use, in which the sentence describes an event that is occurring or a state that obtains at the moment of utterance.<sup>8</sup> The preoccupation with reportive sentences is unfortunate for two reasons. First, the reportive uses are often the less natural ones—consider the sentence

<sup>7</sup>This is true, for example, of Bennett and Partee. But there is no consensus here. Kuhn [1983], for example, argues that past, present, and future should be taken as (equally fundamental) modes of combination of noun phrases and verb phrases.

<sup>8</sup>Many authors restrict the use of the term 'reportive' to event sentences.

*Jill walks to work* (though many languages do not share this feature with English). Second, if the present tense is taken as fundamental, the omission of a reading in the present tense can be transferred to the other tenses. (*John walked to work* can mean that John habitually walked to work.) The neglect is understandable, however, in view of the variety of uses the present can have and the difficulty of analyzing them. One encounters immediately, for example, the issue discussed below.

**Statives and non-statives.** There is discussion in the philosophical literature beginning with Aristotle about the kinds of verb phrases there are and the kinds of things verb phrases can describe. Details of the classification and terminology vary widely. One reads about events, processes, accomplishments, achievements, states, activities and performances. The labels are sometimes applied to verb phrases, sometimes to sentences and sometimes to eventualities. There seems to be general agreement, however, that some kind of classification of this kind will be needed in a full account of the semantics of tense. In connection with the present tense there is a distinction between verb phrases for which the reportive sense is easy (e.g., *John knows Mary*, *The cat is on the mat*, *Sally is writing a book*) and those for which the reportive sense is difficult (e.g., *John swims in the channel*, *Mary writes a book*). This division almost coincides with a division between verb phrases that have a progressive form and those that do not. (Exceptions—noted by Bennett and Partee—include *John lives in Rome* and *John resides in Rome*, both of which have easy reportive uses but common progressive forms.) It also corresponds closely to a division of sentences according to the kind of *when* clauses they form. The sentence *John went to bed when the cat came in* indicates that John went to bed after the cat came in, while *John went to bed when the cat was on the mat* suggests that the cat remained on the mat for some time after John went to bed. In general, if the result of prefixing a sentence by *when* can be paraphrased using *just after* it will have difficult reportive uses and common progressive forms. If it can be paraphrased using *still at the time* it will have easy reportive uses and no common progressive forms. (Possible exceptions are ‘inceptive readings’ like *She smiled when she knew the answer*; see the discussion in Section 3.4.4 below.)

The correspondence among these three tests suggests that they reflect some fundamental ways in which language users divide the world. The usual suggestion is that sentences in the second class (easy reportive readings, no progressives and *when = still at the time*) describe *states*. States are distinguished by the fact that they seem to have no temporal parts. The way Emmon Bach puts it is that it is possible to imagine various states obtaining even in a world with only one time, whereas it is impossible to imagine events or processes in such a world. (Other properties that have been regarded as

characteristic of states are described in Section 4.2 below.) Sentences that describe states are *statives*; those that do not are *non-statives*.

There is some disagreement about whether sentences in the progressive are statives. The fact that Harry is building a house, for example, can go on at discontinuous intervals and the fact that Mary is swimming in the Channel is composed of a sequence of motions, none of which is itself swimming, lead Gabbay and Moravcsik to the conclusion that present progressives do not denote states. But according to the linguistic tests discussed above progressives clearly do belong with the state sentences. For this reason, Vlach, Bach, and Bennett all take the other side. The exact importance of this question depends on what status one assigns to the property of being a stative sentence. If it means that the sentence implies that a certain kind of eventuality known as a state obtains, then it seems that language users assume or pretend that there is some state that obtains steadily while Mary makes the swimming motions and another while Harry is involved in those house-building activities. On the other hand, if 'stative' is merely a label for a sentence with certain temporal properties, for example passing the tests mentioned above, then the challenge is just to assign a semantics to the progressive which gives progressive sentences the same properties as primitive statives; this alternative does not commit us to the actual existence of states (cf. Dowty's work). Thus, the implications of deciding whether to treat progressives as statives depends on one's overall analytical framework, in particular on the basic eventuality/time ontology one assumes.

A recent analysis of the present tense which relates to these issues has been put forth by Cooper. As mentioned above, Cooper works within the Situation Semantics framework, and is thereby committed to an analysis of tense as an element which describes a spatiotemporal region. A region of this kind is somewhat more like an eventuality, e.g. a state, than a mere interval of time; however, his analysis does not entail a full-blown eventuality theory in that it doesn't (necessarily) propose primitive classes of states, events, processes, etc. Indeed, Cooper proposes to define states, activities, and accomplishments in terms very similar to those usual in interval semantics. For instance, stative and process sentences share the property of describing some temporally included sublocation of any spatiotemporal location which they describe (*temporal ill-foundedness*); this is a feature similar to the *subinterval property*, which arises in purely temporal analyses of the progressive (see 3.4.2 below).

Cooper argues that this kind of framework allows an explanation for the differing effects of using the simple present with stative, activity, and accomplishment sentences. The basic proposal about the present tense is that it describes a present spatiotemporal location—i.e. the location of discourse. Stative sentences have both temporal ill-foundedness and the property of *independence of space*, which states that, if they describe a location  $l$ , they also describe the location  $l+$  which is  $l$  expanded to include

all of space. This means that if, for example, John loves Mary anywhere for a length of time including the utterance time, *John loves Mary* will describe all of space for the utterance time. This, according to Cooper, allows the easy use of the present tense here. It seems, though, that to get the result we need at least one more premise: either a stative must describe any spatial sublocation of any location it describes (so that it will precisely describe the utterance location) or we must count the location of utterance for a stative to include all of space.

Activity sentences do not have independence of space. This means that, if they are to be true in the present tense, the utterance location will have to correspond spatially to the event's location. This accounts for the immediacy of sentences like *Mary walks away*. On the other hand, they do have temporal ill-foundedness, which means that the sentence can be said even while the event is still going on. Finally, accomplishment sentence lack the two above properties but have *temporal well-foundedness*, a property requiring them not to describe of any temporal subpart of any location they describe. This means that the discourse location of a present tense accomplishment sentence will have to correspond exactly to the location of the event being described. Hence such sentences have the sense of narrating something in the vicinity just as it happens (*He shoots the ball!*)

Cooper goes on to discuss how locations other than the one where a sentence is actually uttered may become honorary utterance locations. This happens, for example, in the historical present or when someone narrates events they see on TV (following Ejerhed). Cooper seems correct in his claim that the variety of ways in which this occurs should not be a topic for formal semantic analysis; rather it seems to be understandable only in pragmatic or more general discourse analytic terms.

**Past Tense.** Every account of the past tense except those of Dowty and Parsons accommodates in some way the notion that past tense sentences are more definite than the usual tense logic operators. Even Dowty and Parsons, while claiming to treat the more fundamental use of the past tense, acknowledge the strength of the arguments that the past can refer to a definite time. Both cite Partee's example:

When uttered, for instance, half way down the turnpike such a sentence [as *I didn't turn off the stove*] clearly does not mean that there exists some time in the past at which I did not turn off the stove or that there exists no time in the past at which I turned off the stove.

There are, however, some sentences in which the past does seem completely indefinite. We can say, for example, *Columbus discovered America* or *Oswald killed Kennedy* without implying or presupposing anything about the date those events occurred beyond the fact that it was in the past. It

would be desirable to have an account of the past that could accommodate both the definite and indefinite examples. One solution, as discussed in Section 3.2, is that we interpret the past as a quantifier over a set of possible reference times.<sup>9</sup> *I left the oven on* is true now only if the oven was left on at one of the past times I might be referring to. The context serves to limit the set of possible reference times. In the absence of contextual clues to the contrary the set comprises *all* the past times and the past is completely indefinite. In any case, the suggestion that the context determines a set of possible reference times seems more realistic than the suggestion that it determines a unique such time.

There is still something a little suspicious, however, about the notion that context determines a reference interval or a range of reference times for past tense sentences to refer to. One would normally take the ‘context of utterance’ to include information like the time and place the utterance is produced, the identity of the speaker and the audience, and perhaps certain other facts that the speaker and the audience have become aware of before the time of the utterance. But in this case it is clear that *Baltimore won the Pennant* and *Columbus discovered America* uttered in *identical* contexts would have *different* reference times.

A way out of the dilemma might be to allow the sentence itself to help identify the relevant components of a rich utterance context. Klein [1994] emphasizes the connection between the topic or background part of a sentence and its reference time (for him *topic time*). A full explanation of the mechanism will require taking into account the presupposition-focus structure of a sentence—that is, what new information is being communicated by the sentence. For example, when a teacher tells her class *Columbus discovered America*, the sentence would most naturally be pronounced with focal intonation on *Columbus*:

(18) COLUMBUS discovered America.

(19) ??Columbus discovered AMERICA.

??Columbus DISCOVERED America.

---

<sup>9</sup>The proposal is made in these terms in Kuhn [1979]. In Bennett–Partee the idea is rather that the reference time is an interval over whose subintervals the past tense quantifies. Thus the main difference between these accounts has to do with whether the reference time (or range of reference times) can be discontinuous. One argument for allowing it to be is the apparent reference to such times in sentences like *John came on a Saturday*. Another such argument might be based on the contention of Kuhn [1979] that the possible reference times are merely the times that happen to be maximally *salient* for speaker and audience. Vlach [1980] goes Partee–Bennett one further by allowing the past to indicate what obtains *in*, *at*, or *for* the reference interval.

The teacher is presupposing that someone discovered America, and communicating the fact that the discovery was made by Columbus. Similarly, when the teacher says *Bobby discovered the solution to problem number seven*, teacher and students probably know that Bobby was trying to solve problem number seven. The new information is that he succeeded. In those cases it is plausible to suppose that possible reference times would be the times at which the sentence's presupposition is true—the time of America's discovery and the times after which Bobby was believed to have started working on the problem. (As support for the latter claim consider the following scenario: Teacher assigns the problems at the beginning of class period. At the end she announces *Bobby discovered the solution to problem seven*. Susy objects *No he didn't. He had already done it at home.*)

A variety of theories have been proposed in recent years to explain how the intonational and structural properties of a sentence serve to help identify the presuppositions and 'new information' in a sentence.<sup>10</sup> We will not go into the details of these here, but in general we can view a declarative sentence as having two functions. First, it identifies the relevant part of our mutual knowledge. Second, it supplies a new piece of information to be added to that part. It is the first function that helps delimit possible reference times. Previous discourse and non-linguistic information, of course, also play a role. When I say *Baltimore won the Pennant* it matters whether we have just been talking about the highlights of 1963 or silently watching this week's Monday Night Baseball.

**Frequency.** Bäuerle and von Stechow point out that interpreting the past tense as a quantifier ranging over possible reference times (or over parts of the reference time) makes it difficult to explain the semantics of frequency adverbs. Consider, for example, the sentence *Angelika sneezed exactly three times*, uttered with reference to the interval from two o'clock to three o'clock yesterday morning. We might take the sentence to mean that there are exactly three intervals between two and three with reference to which *Angelika sneezed* is true. But if *Angelika sneezed* means that she sneezed at least once within the time interval referred to, then whenever there is one such interval there will be an infinite number of them. So *Angelika sneezed exactly three times* could never be true. Alternatively we might take the sentence to mean that there was at least one time interval within which Angelika sneezed-three-times. But the intervals when Angelika sneezed three times will contain subintervals in which she sneezed twice. So in this case *Angelika sneezed exactly three times* would imply *Angelika sneezed exactly twice*.

This problem leads Bäuerle and von Stechow to insist that the past tense itself indicates simply that the eventuality described occupies that part of

---

<sup>10</sup>On the theory of focus, see for example Jackendoff, Rooth [1985; 1992], and Cresswell and von Stechow. On the nature of presupposition and factivity more generally, Levinson provides a good overview.

the reference time that lies in the past. On this interpretation, it does make sense to say that *Angelika sneezed three times* means that there were three times with reference to which *Angelika sneezed* is true. Tichý, using a different framework, arrives at a similar analysis. Unfortunately, this position also has the consequence that the simple sentence *Angelika sneezed*, taken literally, would mean that Angelika's sneeze lasted for the full hour between two and three. Bäuerle–von Stechow and Tichý both suggest that past tense sentences without explicit frequency operators often contain an *implicit* 'at least once' adverb. In a full treatment the conditions under which the past gets the added implicit adverb would have to be spelled out, so it is not clear how much we gain by this move. The alternative would seem to be to insist that the 'at least once' qualification is a normal part of the meaning of the tense which is dropped in the presence of frequency adverbs. This seems little better.

Vlach handles the frequency problem by allowing sentences to be true either 'in' or 'at' a time interval. *Angelika sneezed exactly three times* is true *at* the reference interval if it contains exactly three subintervals *at* which Angelika sneezes. On the other hand *Angelika sneezed* would normally be taken to assert that Angelika sneezed *in* the reference interval, i.e., that there is at least one time in the interval at which she sneezed. Again, a complete treatment would seem to require a way of deciding, for a given context and a given sentence, whether the sentence should be evaluated in or at the reference time.

We might argue that *all* the readings allowed by Vlach (or Bäuerle–von Stechow) are always present, but that language users tend to ignore the implausible ones—like those that talk about sneezes lasting two hours. But the idea that ordinary past tense sentences are riddled with ambiguities is not appealing.

The DRT analysis, on which frequency adverbs are examples of adverbs of quantification, can provide a somewhat more attractive version of the Bäuerle–von Stechow analysis. According to this view, *three times* binds the free time (or eventuality) variable present in the translation, as *always* did in (6)–(7) above. The situation is more straightforward when an additional temporal expression is present:

(20) On Tuesday, the bell rang three times.

(21) **three-times**<sub>*t*</sub>(**past**(*t*) & **Tuesday**(*t*))(**rang**(**the-bell**,*t*)).

Here *Tuesday* helps to identify the set of times *three-times* quantifies over. **Tuesday**(*t*) indicates that *t* is a subinterval of Tuesday. A representation of this kind would indicate that there were three assignments of times during Tuesday to *t* at which the bell rang, where we say that the bell rang at *t* iff *t* is precisely the full interval of bell-ringing. The issue is more difficult when there is no restrictive argument for the adverb, as with *Angelika sneezed*

*three times*. One possibility is that it ranges over all past times. More likely, context would again provide a set of reference times to quantify over. In still other cases, as argued by Klein [1994], it ranges over times which are identified by the ‘background’ or presuppositions of the sentence. Thus, *Columbus sailed to AMERICA four times* means that, of the times when Columbus sailed somewhere, four were ones at which he sailed to America.

In terms of a DRT analysis, when there is no adverbial, as with *Angelika sneezed*, the temporal variable would be bound by whatever default process normally takes care of free variables (‘existential closure’ or another, as discussed above). This parallels the suggestion in terms of Bäuerle–von Stechow’s analysis, that ‘at least once’ is a component of meaning which is ‘dropped’ in the presence of an overt adverbial. Thus, in the DRT account there wouldn’t need to be a special stipulation for this.

There is still a problem with adverbials of duration, such as in *On Tuesday, the bell rang for five minutes*. This should be true, according to the above, if for some subinterval  $t$  of Tuesday,  $t$  is precisely the full time of the bell’s ringing and  $t$  lasts five minutes. Whether the sentence would be true if the bell in fact rang for ten minutes depends on whether *for five minutes* means ‘for at least five’ or ‘for exactly five’. If the former, the sentence would be true but inappropriate (in most circumstances), since it would generate an implicature that the bell didn’t ring for more than five minutes. If the latter, it would be false. It seems better to treat the example via implicature, since it is not as bad as *The bell rang for exactly five minutes* in the same situation, and the implication seems defeasible (*The bell rang for five minutes, if not more.*)

**Future Tense.** The architects of fragments of English with tense seem to have comparatively little to say about the future. Vlach omits it from his very comprehensive fragment, suggesting he may share Jespersen’s view that the future is not a genuine tense. Otherwise the consensus seems to be that the future is a kind of mirror image of the past with the exception, noted by Bennett and Partee, that the times to which the future can refer include the present. (Compare *He will now begin to eat* with *He now began to eat.*)

There appears to be some disagreement over whether the future is definite or indefinite. Tichý adopts the position that it is ambiguous between the two readings. This claim is difficult to evaluate. The sentence *Baltimore will win* can indicate either that Baltimore will win next week or that Baltimore will win eventually. But this difference can be attributed to a difference in the set of possible reference times as easily as to an ambiguity in the word *will*. It is of course preferable on methodological grounds to adopt a uniform treatment if possible.



### 3.4.2 Aspect

**The Progressive.** Those who wrote about the truth conditions of English tenses in the 1960's assumed that sentences were to be evaluated at instants of time. Dana Scott suggested (and Mongague [1970] seconds) a treatment of the present progressive according to which *Mary is swimming in the Channel* is true at an instant  $t$  if *Mary swims in the Channel* is true at every instant in an open interval that includes  $t$ . This account has the unfortunate consequence of making the present progressive form of a sentence imply its (indefinite) past. For a large class of sentences this consequence is desirable. If John is swimming in the Channel he did, at some very recent time, swim in the Channel. On the other hand there are many sentences for which this property does not hold. *John is drawing a circle* does not imply that John drew a circle. *Mary is climbing the Zugspitze* does not imply that Mary climbed the Zugspitze.

In Bennett–Partee, Vlach [1980] and Kuhn [1979] this difficulty avoided by allowing some present tense sentences to be evaluated at extended intervals of time as well as instants. *John is drawing a circle* means that the present instant is in the interior of an interval at which *John draws a circle* is true. The present instant can clearly be in such an interval even though *John drew a circle* is false at that instant. Sentences like *John swims in the Channel*, on the other hand, are said to have what Bennett and Partee label the *subinterval* property: their truth at an interval entails their truth at all subintervals of that interval. This stipulation guarantees that *Mary is swimming in the Channel* does imply *Mary swam in the Channel*.

**Instantaneous events and gappy processes.** Objections have been made to the Bennett–Partee analysis having to do with its application to two special classes of sentences. The first class comprises sentences that cannot plausibly be said to be true at extended intervals, but that do have progressive forms. Vlach, following Gilbert Ryle, calls these achievement sentences. We will follow Gabbay–Moravcsik and Bach in calling them instantaneous event sentences. They include *Baltimore wins*, *Columbus reaches North America*, *Columbus leaves Portugal* and *Mary starts to sweat*. It seems clear that instantaneous event sentences fail all the tests for statives. But if they are really true only instantaneously then the interval analysis would predict that they would never form true progressives.

The second class contains just the sentences whose present progressive implies their indefinite past. These are the *process* sentences. The Bennett–Partee analysis (and its modalized variation discussed below) have the consequence that process sentences can't have 'gappy' progressives. If *I sat in the front row of the Jupiter theater* was true at the interval from two o'clock to four o'clock last Saturday afternoon, then *I was sitting in the front row of the Jupiter theater* was true at all instants between those times including, perhaps, some instants at which I was really buying popcorn. This accord-

ing to Vlach, Bennett, and Gabbay–Moravcsik, is a conclusion that must be avoided.<sup>11</sup>

Vlach’s solution to the problems of instantaneous events and gappy processes is to give up the idea that a uniform treatment of the progressive is possible. For every non-stative sentence  $A$ , according to Vlach, we understand a notion Vlach calls the *process of  $A$*  or, simply  $\text{proc}(A)$ . The present progressive form of  $A$  simply says that our world is now in the state of  $\text{proc}(A)$ ’s going on.

The nature of  $\text{proc}(A)$ , however, depends on the kind of sentence  $A$  is. If  $A$  is a process sentence then  $\text{proc}(A)$  is ‘the process that goes on when  $A$  is true.’ For the other non-stative sentences,  $\text{proc}(A)$  is a process that ‘leads to’ the truth of  $A$ , i.e., a process whose ‘continuation... would eventually cause  $A$  to become true.’ In fact, Vlach argues, to really make this idea precise we must divide the non-process, non-stative sentences into at least four subclasses.

The first subclass contains what we might (following Bach) call extended event sentences. Paradigm examples are *John builds a house* and *Mary swims across the Channel*. If an extended event sequence is true at an interval  $I$  then  $\text{proc}(A)$  starts at the beginning of  $I$  and ends at the end of  $I$ . For the second subclass (*John realizes his mistake*, *Mary hits on an idea*)  $\text{proc}$  is not defined at all. For the third class (*Mary finishes building the house*, *Columbus reaches North America*) the progressive indicates that the corresponding process is in its final stages. For the fourth class (*Max dies*, *The plane takes off*)  $\text{proc}$  must give a process that culminates in a certain state.

Vlach’s account is intended only as a rough sketch. As Vlach himself acknowledges, there remain questions of clarification concerning the boundaries of the classes of sentences and the formulation of the truth conditions. Furthermore, Vlach’s account introduces a new theoretical term. If the account is to be really enlightening we would like to be sure that we have an understanding of  $\text{proc}$  that is independent of, but consistent with, the truth conditions of the progressive. Even if all the questions of clarification were resolved, Vlach’s theory might not be regarded as particularly attractive because it abandons the idea of a uniform account of the progressive. Not even the sources of irregularity are regular. The peculiarity of the truth conditions for the progressive form of a sentence  $A$  are explained sometimes by the peculiarity of  $A$ ’s truth conditions, sometimes by the way  $\text{proc}$  operates on  $A$  and sometimes by what the progressive says about  $\text{proc}(A)$ . In

---

<sup>11</sup>This argument is not completely decisive. It would seem quite natural to tell a friend one meets at the popcorn counter *I am sitting in the front row*. On the other hand, if one is prepared to accept *I am not sitting in the front row* at popcorn buying time, then perhaps one should be prepared to accept *I sat in the front row before I bought the popcorn and again after*. This would suggest the process went on *twice* during the long interval rather than at one time with a gap.

this sense, Vlach's account is *pessimistic*. Other attempts have been made to give a more uniform account of the progressive. These *optimistic* theories may be divided into two groups depending on whether they propose that the progressive has a modal semantics.

**Non-Modal Accounts.** The analysis of Bennett-Partee discussed above was the first optimistic account presented developed in the formal semantic tradition. Since that time, two other influential non-modal proposals have been put forth. One is by Michael Bennett [1981] and one by Terence Parsons [1985; 1990]. The accounts of Vlach, Bennett and Parsons (and presumably anyone else) must distinguish between statives and non-statives because of the differences in their ability to form progressives. Non-statives must be further divided between processes and events if the inference from present progressive to past is to be selectively blocked. But in the treatments of Bennett and Parsons, as opposed to that of Vlach, all the differences among these three kinds of sentences are reflected in the untensed sentences themselves. Tenses and aspects apply uniformly.

Bennett's proposal is extremely simple.<sup>12</sup> The truth conditions for the present perfect form of *A* (and presumably all the other forms not involving progressives) require that *A* be true at a *closed* interval with the appropriate location. The truth conditions for the progressive of *A* require that *A* be true in an open interval with the appropriate location. Untensed process sentences have two special properties. First, if a process sentence is true at an interval, it is true at all closed subintervals of that interval. Second, if a process sentence is true at every instant in an interval (open or closed) then it is true at that interval. Neither of these conditions need hold for event sentences. Thus, if *John is building a house* is true, there must be an open interval at which *John builds a house* is true. But if there is no *closed* interval of that kind, then *John has built a house* will be false. On the other hand, *Susan is swimming* does imply *Susan has (at some time) swum* because the existence of an open interval at which *Susan swims* is true guarantees the existence of the appropriate closed intervals.

If this proposal has the merit of simplicity, it has the drawback of seeming very *ad hoc*—'a logician's trick' as Bennett puts it. Bennett's explanatory remarks are helpful. Events have a beginning and an end. They therefore occupy closed intervals. Processes, on the other hand, need not. But a process is composed, at least in part, of a sequence of parts. If Willy walks then there are many subintervals such that the eventualities described by *Willy walks* are also going on at these intervals. Events, however, need not be decomposable in this way.

The account offered by Parsons turns out to be similar to Bennett's. Parson's exposition seems more natural, however, because the metaphysical underpinnings discussed above are exposed. Parsons starts with the

---

<sup>12</sup>Bennett attributes the idea behind his proposal to Glen Helman.

assumption that there are three kinds of eventualities: *states*, *processes*, and *events*. Eventualities usually have *agents* and sometimes *objects*. An agent may or may not be *in* a state at a time. Processes may or may not be *going on* at a time. Events may or may not be *in development* at a time. In general, if  $e$  is an eventuality, we say that  $e$  *holds* at time  $t$  if the agent of  $e$  is in  $e$  at  $t$  or  $e$  is in development or going on at  $t$ . In addition, events can have the property of *culminating* at a time. The set of times at which an event holds is assumed to be an open interval and the time, if any, at which it culminates is assumed to be the least upper bound of the times at which it holds.

The structure of language mirrors this metaphysical picture. There are three kinds of untensed sentences: stative, process sentences and event sentences. Tensed sentences describe properties of eventualities. Stative and process sentences say that an eventuality *holds* at a time. Event sentences say that an eventuality *culminates* at a time. So, for example, *John sleeps* can be represented as (22) and *Jill bought a cat* as (23):

$$(22) \quad \exists e \exists t [\text{pres}(t) \wedge \text{sleeping}(e) \wedge \text{holds}(e, t) \wedge \text{agent}(e, \text{john})]$$

$$(23) \quad \exists e \exists t \exists x [\text{past}(t) \wedge \text{buying}(e) \wedge \text{culm}(e, t) \wedge \text{agent}(e, \text{jill}) \wedge \text{cat}(x) \wedge \text{obj}(e, x)].$$

The treatment of progressives is remarkably simple. Putting a sentence into the progressive has no effect whatsoever, other than changing the sentence from a non-stative into a stative. This means that, for process sentences, the present and progressive are equivalent. *John swims* is true if and only if *John is swimming* is true. Similarly, *John swam* is true if and only if *John was swimming* is true. For event sentences, the change in classification does affect truth conditions. *John swam across the Channel* is true if the event described *culminated* at some past time. *John was swimming across the Channel*, on the other hand, is true if the state of John's swimming across the Channel *held* at a past time. But this happens if and only if the event described by *John swims across Channel* was *in development* at that time. So it can happen that *John was swimming across the Channel* is true even though John never got to the other side.

Landman [1992] points out a significant problem for Parsons' theory. Because it is a purely extensional approach, it predicts that *John was building a house* is true if and only if there is a house  $x$  and a past event  $e$  such that  $e$  is an event of John building  $x$  and  $e$  holds. This seems acceptable. But Landman brings up examples like *God was creating a unicorn (when he changed his mind)*. This should be true iff there is a unicorn  $x$  and a past event  $e$  such that  $e$  is an event of God creating  $x$  and  $e$  holds. But it may be that the process of creating a unicorn involves some mental planning or magic words but doesn't cause anything to appear until the last moment,

when all of a sudden there is a fully formed unicorn. Thus no unicorn need ever exist for the sentence to be true. Landman's problem arises because of Parsons' assumption that eventualities are described primarily by the verb alone, as a swimming, drawing, etc., and by thematic relations connecting them to individuals, as **agent**(*e*, **jill**) or **obj**(*e*, *x*). There is no provision for more complex descriptions denoting a property like 'house-building'. The question is how intrinsic this feature is to Parsons' analysis of tense and aspect. One could adjust his semantics of verbs to make them multi-place intensional relations, so that *John builds a house* could be analyzed as:

$$(24) \quad \exists e \exists t [\text{pres}(t) \wedge \text{building}(e, \text{john}, \text{a house}) \wedge \text{culm}(e, t)].$$

But then we must worry about how the truth conditions of **building**(*e*, **john**, **a house**) are determined on a compositional basis and how one knows what it is for an eventuality of this type to hold or culminate. However, while the challenge is real, it is not completely clear that it is impossible to avoid Landman's conclusion that the progressive cannot be treated in extensional terms.

It seems likely that, with the proper understanding of theoretical terms, Parsons, Vlach, and Bennett could be seen as saying very similar things about the progressive. Parsons' exposition seems simpler than Vlach's, however, and more natural than Bennett's. These advantages may have been won partly by reversing the usual order of analysis from ordinary to progressive forms. Vlach's account proceeds from *A* to *proc*(*A*) to the state of *proc*(*A*)'s holding. In Bennett's, the truth conditions for the progressive of *A* are explained in terms of those for *A*. If one compares the corresponding progressive and non-progressive forms on Parson's account, however, one sees that in the progressive of an event sentence, something is *subtracted* from the corresponding non-progressive form. The relations between the progressive and non-progressive forms seem better accommodated by viewing events as processes plus culminations rather than by viewing processes as eventualities 'leading to' events.

On the other hand the economy of Parsons' account is achieved partly by ignoring some of the problems that exercise Vlach. The complexity of Vlach's theory increases considerably in the face of examples like *Max is dying*. To accommodate this kind of case Parsons has two options. He can say that they are ordinary event sentences that are in development for a time and then culminate, or he can say that they belong to a new category—achievement—of sentences that culminate but never hold. The first alternative doesn't take account of the fact that such eventualities can occur at an instant (compare *Max was dying and then died at 5:01* with *Jane was swimming across the Channel and then swam across the Channel at 5:01*). The second requires us to say that the progressive of these sentences, if it can be formed at all, involves a 'change in meaning' (cf. Parsons [1990,

p. 24, 36]). But the progressive *can* be formed and spelling out the details of the meaning changes involved will certainly spoil some of Parsons' elegance.

**Unfinished progressives and Modal Accounts.** According to the Bennett–Partee account of progressives, *John was building a house* does not imply that John built a house. It does, however, imply that John will eventually have built a house. Yet it seems perfectly reasonable to say:

(25) John was building a house when he died.

One attempt to modify the account to handle this difficulty is given by Dowty [1979]. Dowty's proposal is that we make the progressive a *modal* notion.<sup>13</sup> The progressive form of a sentence *A* is true at a time *t* in world *w* just in case *A* is true at an interval containing *t* in all worlds *w'* such that *w'* and *w* are exactly alike up to *t* and the course of events after *t* develops in the way most compatible with past events. The *w'*-worlds mentioned are referred to as 'inertia worlds'. (25) means that *John builds a house* is eventually true in all the worlds that are inertia worlds relative to ours at the interval just before John's death.

If an account like this is to be useful, of course, we must have some understanding of the notion of inertia world independent of its role in making progressive sentences true. The idea of a development maximally compatible with past events may not be adequate here. John's death and consequent inability to finish his house may have been natural, even inevitable, at the time he was building it. In Kuhn [1979] the suggestion is that it is the expectations of the language users that are at issue. But this seems equally suspect. It is quite possible that because of a bad calculation, we all mistakenly expect a falling meteor to reach earth. We would not want to say in this case that the meteor *is* falling to earth.

Landman attempts to identify in more precise terms the alternate possible worlds which must be considered in a modal semantics of the progressive. We may label his the *counterfactual analysis*, since it attempts to formalize the following intuition: Suppose we are in a situation in which John fell off the roof and died, and so didn't complete the house, though he would have finished it if he hadn't died. Then (25) is true *because* he would have finished if he hadn't died. Working this idea out requires a bit more complexity, however. Suppose not only that John fell off the roof and died, but also that if he hadn't fallen, he would have gotten ill and not finished the house anyway. The sentence is still true, however, and this is because he would have finished the house if he hadn't fallen and died and hadn't gotten ill. We can imagine still more convoluted scenarios, where other dangers lurk for John. In the end, Landman proposes that (25) is true iff *John builds a*

<sup>13</sup>Dowty attributes this idea to David Lewis.

*house* would be true if nothing were to interrupt some activity that John was engaged in.

Landman formalizes his theory in terms of the notion of the *continuation branch* of an event  $e$  in a world  $w$ . He assumes an ontology wherein events have *stages* (cf. Carlson [1977]); the notion of ‘stage of an eventuality’ is not defined in a completely clear way. Within a single world, all of the temporally limited subeventualities of  $e$  are stages of  $e$ . An eventuality  $e'$  may also be a stage of an eventuality  $e$  in another world. It seems that this can occur when  $e'$  is duplicated in the world of  $e$  by an eventuality which is a stage of  $e$ . The continuation branch of  $e$  in  $w$ ,  $C(e, w)$ , is a set of event-world pairs;  $C(e, w)$  contains all of the pairs  $\langle a, w \rangle$  where  $a$  is a stage of  $e$  in  $w$ . If  $e$  is a stage of a larger event in some other possible world, we say that it *stops* in  $w$  (otherwise it simply ends in  $w$ ). If  $e$  stops in  $w$  at time  $t$ , the continuation branch moves to the world  $w_1$  most similar to  $w$  in which  $e$  does not stop at  $t$ . Suppose that  $e_1$  is the event in  $w_1$  of which  $e$  is a stage; then all pairs  $\langle a, w_1 \rangle$ , where  $a$  is a stage of  $e_1$  in  $w_1$ , are also in  $C(e, w)$ . If  $e_1$  stops in  $w_1$ , the continuation branch moves to the world most similar to  $w_1$  in which  $e_1$  does not stop, etc. Eventually, the continuation branch may contain a pair  $\langle e_n, w_n \rangle$  where a house gets built in  $e_n$ . Then the continuation branch ends. We may consider the continuation branch to be the maximal extension of  $e$ . *John was building a house* is true in  $w$  iff there is some event in  $w$  whose continuation branch contains an event of John building a house.

Landman brings up one significant problem for his theory. Suppose Mary picks up her sword and begins to attack the whole Roman army. She kills a few soldiers and then is cut down. Consider (26):

(26) Mary was wiping out the Roman army.

According to the semantics described above, (26) ought to be true. Whichever soldier actually killed Mary might not have, and so the continuation branch should move to a world in which he didn't. There some soldier kills Mary but might not have, so . . . Through a series of counterfactual shifts, the continuation branch of Mary's attack will eventually reach a world in which she wipes out the whole army. Landman assumes that (26) ought not be true in the situation envisioned. The problem, he suggests, is that the worlds in which Mary kills a large proportion of the Roman army, while possible, are outlandishly unreasonable. He therefore declares that only ‘reasonable worlds’ may enter the continuation branch.

Landman's analysis of the progressive is the most empirically successful optimistic theory. Its major weaknesses are its reliance on two undefined terms: *stage* and *reasonable*. The former takes part in the definition of when an event stops, and so moves the continuation branch to another world. How do we know with (as) the event John was engaged in didn't end when he

died? Lots of eventualities did end there; we wouldn't want to have *John was living to be 65* to be true simply because he would probably have lived that long if it weren't for the accident. We know that the construction event didn't end because we know it was supposed to be a house-building. Thus, Landman's theory requires a primitive understanding of when an event is complete, ending in a given world, and when it is not complete and so may continue on in another world. In this way, it seems to recast in an intensional theory Parsons' distinction between holding and culminating. The need for a primitive concept of reasonableness of worlds is perhaps less troubling, since it could perhaps be assimilated to possible worlds analyses of epistemic modality; still, it must count as a theoretical liability.

Finally, we note that Landman's theory gives the progressive a kind of interpretation quite different from any other modal or temporal operator. In particular, since it is nothing like the semantics of the perfect, the other aspect we will consider, one wonders why the two should be considered members of a common category. (The same might be said for Dowty's theory, though his at least resembles the semantics for modalities.)

**The Perfect.** Nearly every contemporary writer has abandoned Montague's position that the present perfect is a completely indefinite past. The current view (e.g. [McCoard, 1978; Richards, 1982; Mittwoch, 1988]) seems to be that the time to which it refers (or the range of times to which it might refer) must be an *Extended Now*, an interval of time that begins in the past and includes the moment of utterance. The event described must fall somewhere within this interval. This is plausible. When we say *Pete has bought a pair of shoes* we normally do not mean just that a purchase was made at some time in the past. Rather we understand that the purchase was made recently. The view also is strongly supported by the observation that the present perfect can always take temporal modifiers that pick out intervals overlapping the present and never take those that pick out intervals entirely preceding the present: *Mary has bought a dress since Saturday*, but not \**Mary has bought a dress last week*. These facts can be explained if the adverbials are constrained to have scope over the perfect, so that they would have to describe an extended now.

There is debate, however, about whether the extended now theory should incorporate two or even three readings for the perfect. The uncontroversial analysis, that suggested above, locates an event somewhere within the extended now. This has been called the *existential use*. Others have argued that there is a separate *universal* or *continuative use*. Consider the following, based on some examples of Mittwoch:

- (27) Sam has lived in Boston since 1980.

This sentence is compatible with Sam's still living in Boston, or with his having come, stayed for a while, and then left. Both situations are com-



patible with the following analysis: the extended now begins in 1980, and somewhere within this interval Sam lives in Boston. However, supporters of the universal use (e.g. [McCawley, 1971; Mittwoch, 1988; Michaelis, 1994]) argue that there is a separate reading which requires that Sam's residence in Boston continue at the speech time: (27) is true iff Sam lives in Boston throughout the whole extended now which begins in 1980.

Michaelis argues that the perfect has a third reading, the *resultative use*. A resultative present perfect implies that there is a currently existing result state of the event alluded to in the sentence. For example, *John has eaten poison* could be used to explain the fact that John is sick. Others (McCawley [1971], Klein [1994]) argue that such cases should be considered examples of the existential use, with the feeling that the result is especially important being a pragmatic effect. At the least one may doubt analyses in terms of result state on the grounds that precisely which result is to be focused on is never adequately defined. Any event will bring about some new state, if only the state of the event having occurred, and most will bring about many. So it is not clear how this use would differ in its truth conditions from the existential one.

Stump argues against the Extended Now theory on the basis of the occurrence of perfects in nonfinite contexts like the following (his Chapter IV, (11); cf. McCoard, Klein, Richards who note similar data):

- (28) Having been on the train yesterday, John knows exactly why it derailed.

Stump provides an analysis of the perfect which simply requires that no part of the event described be located after the evaluation time. In a present perfect sentence, this means that the event can be past or present, but not future. Stump then explains the ungrammaticality of *\*Mary has bought a dress last week* in pragmatic terms. This sentence, according to Stump, is truth conditionally equivalent to *Mary bought a dress last week*. Since the latter is simpler and less marked in linguistic terms, the use of the perfect should implicate that the simple past is inappropriate. But since the two are synonymous, it cannot be inappropriate. Therefore, the present perfect with a definite past adverbial has an implicature which can never be true. This is why it cannot be used (cf. Klein [1992] for a similar explanation).

Klein [1992; 1994] develops a somewhat different analysis of perfect aspect from those based on interval semantics. He concentrates on the relevance of the aspectual classification of sentences for understanding different 'uses' of the perfect. He distinguishes *0-state*, *1-state*, and *2-state* clauses: A 0-state clause describes an unchanging state of affairs (*The Nile is in Africa*); a 1-state sentence describes a state which obtains at some interval while not obtaining at adjoining intervals (*Peter was asleep*); and a 2-state clause denotes a change from one lexically determined state to another (*John opened*

*the window*). Here, the first state (the window's being closed) is called the *source state*, and the second (the window's being open) the *target state*. He calls the maximal intervals which precede and follow the interval at which a state holds its *pretime* and *posttime* respectively.

Given this framework, Klein claims that all uses of the perfect can be analyzed as the reference time falling into the posttime of the most salient situation described by the clause. Since the states described by 0-state sentences have no posttime, the perfect is impossible (*\*The Nile has been in Africa*). With 1-state sentences, the reference time will simply follow the state in question, so that *Peter has been asleep* will simply indicate that Peter has at some point slept ('experiential perfect'). With 2-state sentences, Klein stipulates that the salient state is the source state, so that *John has opened the window* literally only indicates that the reference time (which in this case corresponds to the utterance time) follows a state of the window being closed which itself precedes a state of the window being open. It may happen that the reference time falls into the target state, in which case the window must still be open ('perfect of result'); alternatively, the reference time may follow the target state as well—i.e. it may be a time after which the window has closed again—giving rise to another kind of experiential perfect.

One type of case which is difficult for Klein is what he describes as the 'perfect of persistent situation', as in *We've lived here for ten years*. This is the type of sentence which motivated the universal/continuative semantics within the Extended Now theory. In Klein's terms, here it seems that the reference time, the present, falls into the state described by a 1-state sentence, and not its posttime. Klein's solution is to suggest that the sentence describes a state which is a substate of the whole living-here state, one which comprises just the first ten years of our residency, a 'living-here-for-ten-years' state. The example indicates that we are in the posttime of this state, a fact which does not rule out that we're now into our eleventh year of living here. On the other hand, such an explanation does not seem applicable to other examples, such as *We've lived here since 1966*.

**Existence presuppositions.** Jespersen's observation that the present perfect seems to presuppose the present existence of the subject in cases where the past tense does not has been repeated and 'explained' many times. We are now faced with the embarrassment of a puzzle with too many solutions. The contemporary discussion begins with Chomsky, who argues that *Princeton has been visited by Einstein* is all right, but *Einstein has visited Princeton* is odd. James McCawley points out that the alleged oddity of the latter sentence actually depends on context and intonation. Where the existence presupposition does occur, McCawley attributes it to the fact that the present perfect is generally used when the present moment is included in an interval during which events of the kind being described *can* be true.

Thus, *Have you seen the Monet exhibition?* is inappropriate if the addressee is known to be unable to see it. (*Did you* is appropriate in this case.) *Frege has contributed a lot to my thinking* is appropriate to use even though Frege is dead because Frege *can* now contribute to my thinking. *My mother has changed my diapers many times* is appropriate for a talking two year old, but not for a normal thirty year old. *Einstein has visited Princeton* is odd because Einsteinian visits are no longer possible. *Princeton has been visited by Einstein* is acceptable because Princeton's being visited *is* still possible.

In Kuhn [1983] it is suggested that the explanation may be partly syntactic. Existence presuppositions can be canceled when a term occurs in the scope of certain operators. Thus *Santa is fat* presupposes that Santa exists, but *According to Virginia, Santa is fat* does not. There are good reasons to believe that past and future apply to sentences, whereas perfect applies only to intransitive verb phrases. But in that case it is natural that presuppositions concerning the subject that do hold in present perfect sentences fail in past and future sentences.

Guenther requires that at least one of the objects referred to in a present perfect sentence (viz., the topic of the sentence) must exist at utterance time. Often, of course, the subject will be the topic.

The explanation given by Tichý is that, in the absence of an explicit indication of reference time, a present perfect generally refers to the lifetime of its subject. If this does not include the present, then the perfect is inappropriate.

Overall, the question of whether these explanations are compatible, and whether they are equally explanatory, remains open.

### 3.4.3 Tense in Subordinate Clauses

The focus in all of the preceding discussion has been on occurrences of tense in simple sentences. A variety of complexities arise when one tries to accommodate tense in subordinate clauses. Of particular concern is the phenomenon known as *Sequence of Tense*. Consider the following:

(29) John believed that Mary left.

(30) John believed that Mary was pregnant.

Example (29) says that at some past time  $t$  John had a belief that at some time  $t' < t$ , Mary left. This reading is easily accounted for by a classic Priorean analysis: the time of evaluation is shifted into the past by the first tense operator, and then shifted further back by the second. (30), which differs from (29) in having a stative subordinate clause, has a similar reading, but has another as well, the so-called 'simultaneous reading', on which the time of Mary's alleged pregnancy overlaps with the time of John's belief. It would seem that the tense on *was* is not semantically active. A

traditional way of looking at things is to think of the tense form of *was* as triggered by the past tense of *believed* by a morphosyntactic sequence of tense (SOT) rule. Following Oghihara [1989; 1995], we could formalize this idea by saying that a past tense in a subordinate clause governed by another past tense verb is deleted prior to the sentence's being interpreted. For semantic purposes, (30) would then be *John believed that Mary be (tenseless) pregnant*. Not every language has the SOT rule. In Japanese, for example, the simultaneous reading of (30) would be expressed with present tense in the subordinate clause.

The SOT theory does not explain why simultaneous readings are possible with some clauses and not with others. The key distinction seems to be between states and non-states. One would hope to be able to relate the existence of simultaneous readings to the other characteristic properties of statives discussed in Section 3.4.1 above.

Sentences like (31) pose special problems. One might expect for it to be equivalent to either (30), on the simultaneous reading, or (32).

(31) John believed that Mary is pregnant.

(32) John believed that Mary would now be pregnant.

A simultaneous interpretation would be predicted by a Priorian account, while synonymy with (32) would be expected by a theory which said that present tense means 'at the speech time'. However, as pointed out by Enç [1987], (31) has a different, problematical interpretation; it seemingly requires that the time of Mary's alleged pregnancy extend from the belief time up until the speech time. She labels this the *Double Access Reading* (DAR). Recent theories of SOT, in particular those of Oghihara [1989; 1995] and Abusch [1991; 1995], have been especially concerned with getting a correct account of such 'present under past' sentences.

Enç's analysis of tense in intensional contexts begins with the proposal that tense is a referential expression. She suggests that the simultaneous interpretation of (30) should be obtained through a 'binding' relationship between the two tenses, indicated by coindexing as in (33). The connection is similar to that holding with nominal anaphora, as in (34).

(33) John PAST<sub>1</sub> believed that Mary PAST<sub>1</sub> was pregnant.

(34) John<sub>1</sub> thinks that he<sub>1</sub> is smart.

This point of view lets Enç say that both tense morphemes have a usual interpretation. Her mechanisms entail that all members of a sequence of coindexed tense morphemes denote the same time, and that each establishes the same temporal relationship as the highest ('first') occurrence. Oghihara elucidates the intended interpretation of structures like (33) by translating them into Intensional Logic.

$$(35) \quad t_1 < s^* \& \text{believe}'(t_1, j, ^\wedge [t_1 < s^* \wedge \text{be-pregnant}(t_1, m)]).$$

Here  $s^*$  denotes the speech time. If the two tenses were not coindexed, as in (36), the second would introduce  $t_2 < t_1$  to the translation:

$$(36) \quad \text{John PAST}_1 \text{ believed that Mary PAST}_2 \text{ was pregnant.}$$

$$(37) \quad t_1 < s^* \& \text{believe}'(t_1, j, ^\wedge [t_2 < t_1 \wedge \text{be-pregnant}(t_2, m)]).$$

This represents the non-simultaneous ('shifted') reading.

Accounting for the DAR is more complex. Enç proposes that there need to be two ways that temporal expressions may be linked. Expressions receive pairs of indices, so that with a configuration  $A\langle i, j \rangle \dots B\langle k, l \rangle$ , if  $i = k$ , then  $A$  and  $B$  refer to the same time, while if  $j = l$ , then the time if  $B$  is included in that of  $A$ . The complement clause *that Mary is pregnant* is then interpreted outside the scope of the past tense. The present tense is linked to the speech time. As usual, however, the two tenses may be coindexed, but only via their second indices. This gives us something like (38).

$$(38) \quad \exists x(x = [\text{Mary PRES}_{\langle 0,1 \rangle} \text{ be pregnant}] \text{ John PAST}_{\langle 2,1 \rangle} \text{ believes } x).$$

This representation says that Mary is pregnant at the speech time and that the time of John's belief is a subinterval of Mary's pregnancy. Thus it encodes the DAR.

The mechanisms involved in deriving and interpreting (38) are quite complicated. In addition, examples discussed by Abusch [1988], Baker [1989] and Ogihara [1995] pose a serious difficulty for Enç's view.

$$(39) \quad \text{John decided a week ago that in ten days at breakfast he would say to his mother that they were having their last meal together.}$$

Here, on the natural interpretation of the sentence, the past tense of *were* does not denote a time which is past with respect to either the speech time or any other time mentioned in the sentence. Thus it seems that the tense component of this expression cannot be semantically active.

As mentioned above, Ogihara proposes that a past tense in the right relation with another past tense may be deleted from a sentence prior to semantic interpretation. (Abusch has a more complex view involving feature passing, but it gets similar effects.) This would transform (39) into (40).

$$(40) \quad \text{John PAST decided a week ago that in ten days at breakfast he } \emptyset \text{ woll say to his mother that they } \emptyset \text{ be having their last meal together.}$$

Notice that we have two deleted tenses (marked ‘ $\emptyset$ ’) here. *Would* has become tenseless *woll*, a future operator evaluated with respect to the time of the deciding. Then breakfast time ten days after the decision serves as the time of evaluation for *he say to his mother that they be having their last meal together*. Since there are no temporal operators in this constituent, the time of the saying and that of the last meal are simultaneous.

The double access sentence (31) is more difficult story. Both Ogihara and Abusch propose that the DAR is actually a case of *de re* interpretation, similar to the famous Ortcutt examples of Quine [1956]. Consider example (31), repeated here:

(31) John believed that Mary is pregnant.

Suppose John has glimpsed Mary two months ago, noticing that she is quite large. At that time he thought ‘Mary is pregnant’. Now you and I are considering why Mary is so large, and I report John’s opinion to you with (31). The sentence could be paraphrased by *John believed of the state of Mary’s being large that it is a state of her being pregnant*. (Abusch would frame this analysis in terms of a *de re* belief about an interval, rather than a state, but the difference between these two formulations appears slight.) Both Ogihara and Abusch give their account in terms of the analysis of *de re* belief put forward by Lewis [1979] and extended by Cresswell and von Stechow [1982]. These amount to saying that (31) is true iff the following conditions are met: (i) John stands in a suitable *acquaintance relation* *R* to a state of Mary’s (such as her being large), in this case the relation of having glimpsed it on a certain occasion, and (ii) in all of John’s belief-worlds, the state to which he stands in relation *R* is a state of Mary being pregnant.

A *de re* analysis of present under past sentences may hope to give an account of the DAR. Suppose we have an analysis of tense whereby the present tense in (31) entails that the state in question holds at the speech time. Add to this the fact that the acquaintance relation, that John had glimpsed this state at the time he formed his belief, entails that the state existed already at that time. Together these two points require that the state stretch from the time of John’s belief up until the speech time. This is the DAR.

The preceding account relies on the acquaintance relation to entail that the state have existed already at the past time. The idea that it would do so is natural in light of Lewis’ suggestion that the relation must be a causal one: in this case that John’s belief has been caused, directly or indirectly, by the state. However, as Abusch [1995] points out, there is a problem with this assumption: it sometimes seems possible to have a future-oriented acquaintance relation. Consider Abusch’s example (41) (originally due to Andrea Bonomi).

- (41) Leo will go to Rome on the day of Lea's dissertation. Lia believes that she will go to Rome with him then.

Here, according to Abusch, we seem to have a *de re* attitude by Lia towards the future day of Lea's dissertation. Since the acquaintance relation cannot be counted on to require in (31) that the time of Mary's being large overlaps the time when John formed his belief, both Abusch and Ogihara have had to introduce extra stipulations to serve this end. But at this point the explanatory force of appealing to a *de re* attitude is less clear.

There are further reasons to doubt the *de re* account, at least in the form presented. Suppose that we're wondering whether the explanation for Mary's appearance is that she's pregnant. John has not seen Mary at all, but some months ago her mother told John that she is, he believed her, and he reported on this belief to me. It seems that I could say (31) as evidence that Mary is indeed pregnant. In such a case it seems that the sentence is about the state *we're* concerned with, not one which provided John's evidence.

#### 3.4.4 *Tense and discourse*

One of the major contributions of DRT to the study of tense is its focus on 'discourse' as the unit of analysis rather than the sentence. Sentential analyses treat reference times as either completely indeterminate or given by context. In fact the 'context' that determines the time a sentence refers to may just be the sentences that were uttered previously. Theorists working within DRT have sought to provide a detailed understanding of how the reference time of a sentence may depend on the tenses of the sentence and its predecessors.

As mentioned above, DRS's will include events, states, and times as objects in the universe of discourse and will specify relations of precedence and overlap among them. Precisely which relations hold depends on the nature of the eventualities being described. The key distinction here is between 'atelic' eventualities (which include both states and processes) and 'telic' ones. Various similar algorithms for constructing DRS's are given by Kamp, Kamp and Rohrer, Hinrichs, and Partee, among others. Let us consider the following pair of examples:

- (42) Mary was eating a sandwich. Pedro entered the kitchen.

- (43) Pedro went into the hall. He took off his coat.

In (42), the first sentence describes an atelic eventuality, a process, whereas the second describes a telic event. The process is naturally taken to temporally contain the event. In contrast, in (43) both sentences describe telic events, and the resulting discourse indicates that the two happened in sequence.

A DRS construction procedure for these two could work as follows: With both the context provides an initial past reference time  $r_0$ . Whenever a past tense sentence is uttered, it is taken to temporally coincide with the past reference time. A telic sentence introduces a new reference time that follows the one used by the sentence, while an atelic one leaves the reference time unchanged. So, in (42), the same reference time is used for both sentences, implying temporal overlap, while in (43) each sentence has its own reference time, with that for the second sentence following that for the first.

Dowty [1986a] presents a serious critique of the DRT analysis of these phenomena. He points out that whether a sentence describes a telic or atelic eventuality is determined by compositional semantics, and cannot be read off of the surface form in any direct way. He illustrates with the pair (44)–(45).

(44) John walked. (activity)

(45) John walked to the station. (accomplishment)

Other pairs are even more syntactically similar (*John baked a cake* vs. *John baked cakes*.) This consideration is problematical for DRT because that theory takes the unit of interpretation to be the entire DRS. A complete DRS cannot be constructed until individual sentences are interpreted, since it must be determined whether sentences describe telic or atelic eventualities before relations of precedence and overlap are specified. But the sentences cannot be interpreted until the DRS is complete.

Dowty proposes that the temporal sequencing facts studied by DRT can be accommodated more adequately within interval semantics augmented by healthy amounts of Gricean implicature and common-sense reasoning. First of all, individual sentences are compositionally interpreted within a Montague Grammar-type framework. Dowty [1979] has shown how differences among states, processes, and telic events can be defined in terms of their temporal properties within interval semantics. (For example, as mentioned above,  $A$  is a stative sentence iff, if  $A$  is true at interval  $I$ , then  $A$  is true at all moments within  $I$ .) The temporal relations among sentence are specified by a single, homogeneous principle, the *Temporal Discourse Interpretation Principle* (TDIP), which states:

- (46) **TDIP** Given a sequence of sentences  $S_1, S_2, \dots, S_n$  to be interpreted as a narrative discourse, the reference time of each sentence  $S_i$  (for  $i$  such that  $1 < i \leq n$ ) is interpreted to be:
- (a) a time consistent with the definite time adverbials in  $S_i$ , if there are any;
  - (b) otherwise, a time which immediately follows the reference time of the previous sentence  $S_{i-1}$ .



Part (b) is the novel part of this proposal. It gives the same results as DRT in all-telic discourses like (43), but seems to run into trouble with atelic sentences like the one in (42). Dowty proposes that (42) really does describe a sequence of a process or state of Mary eating a sandwich followed by an event of Pedro entering the kitchen; this is the literal contribution of the example (Nerbonne [1986] makes a similar proposal.) However, common sense reasoning allows one to realize that a process of eating a sandwich generally takes some time, and so the time at which Mary was actually eating a sandwich might have started some time before the reference time and might continue for some time afterwards. Thus (42) is perfectly consistent with Mary continuing to eat the sandwich while Pedro entered the kitchen. In fact, Dowty would suggest, in normal situations this is just what someone hearing (42) would be likely to conclude.

Dowty's analysis has an advantage in being able to explain examples of inceptive readings of atelic sentences like *John went over the day's perplexing events once more in his mind. Suddenly, he was fast asleep. Suddenly* tells us that the state of being asleep is new. World knowledge tells us that he could not have gone over the days events in his mind if he were asleep. Thus the state must begin after the event of going over the perplexing events in his mind. DRT would have a more difficult time with this example; it would have to propose that *be asleep* is ambiguous between an atelic (state) reading and a telic (achievement) reading, or that the word *suddenly* cancels the usual rule for atelics.

As Dowty then goes on to discuss, there are a great many examples of discourses in which the temporal relations among sentences do not follow the neat pattern described by the DRT algorithms and the TDIP. Consider:

- (47) Mary did the dishes carefully. She filled the sink with hot water. She added a half cup of soap. Then she gently dipped each glass into the sudsy liquid.

Here all of the sentences after the first one describe events which comprise the dish-washing. To explain such examples, an adherent of DRT must propose additional DRS construction procedures. Furthermore, there exists the problem of knowing which procedures to apply; one would need rules to determine which construction procedures apply before the sentences within the discourse are interpreted, and it is not clear whether such rules can be formulated in a way that doesn't require prior interpretation of the sentences involved. Dowty's interval semantics framework, on the other hand, would say that the relations among the sentences here are determined pragmatically, overriding the TDIP. The weakness of this approach is its reliance on an undeveloped pragmatic theory.

## 4 TENSE LOGICS FOR NATURAL LANGUAGE

4.1 *Motivations*

General surveys of tense logic are contained elsewhere in this Handbook (Burgess, Finger, Gabbay and Reynolds, and Thomason, all in this Volume). In this section we consider relations between tense logic and tense and aspect in natural language. Work on tense logic, even among authors concerned with linguistic matters, has been motivated by a variety of considerations that have not always been clearly delineated. Initially, tense logic seems to have been conceived as a generalization of classical logic that could better represent logical forms of arguments and sentences in which tense plays an important semantic role. To treat such items within classical logic requires extensive ‘paraphrase’. Consider the following example from Quine [1982]:

- (48) George *V* married Queen Mary, Queen Mary is a widow, therefore George *V* married a widow.

An attempt to represent this directly in classical predicate logic might yield

$$(48a) \quad Mgm, Wm \vDash \exists x(Mgx \wedge Wx),$$

which fallaciously represents it as valid. When appropriately paraphrased, however, the argument becomes something like:

- (49) Some time before the present is a time when George *V* married Queen Mary, Queen Mary is a widow at the present time, therefore some time before the present is a time at which George *V* married a widow,

which, in classical logic, is represented by the nonvalid:

$$(49a) \quad \exists t(Tt \wedge Btn \wedge Mgmt), Wmn \vDash \exists t(Tt \wedge Btn \wedge \exists x(Wxn \wedge Mgmt)).$$

If we want a logic that can easily be *applied* to ordinary discourse, however, such extensive and unsystematized paraphrase may be unsatisfying. Arthur Prior formulated several logical systems in which arguments like (48) could be represented more directly and, in a series of papers and books in the fifties and sixties, championed, chronicled and contributed to their development. (See especially [Prior, 1957; Prior, 1967] and [Prior, 1968].) A sentence like *Queen Mary is a widow* is not to be represented by a formula that explicitly displays the name of a particular time and that is interpreted simply as *true* or *false*. Instead it is represented as *Wm*, just as in (48), where such formulas are now understood to be true or false only relative to a time. Past

and future sentences are represented with the help of *tense logical operators* like those mentioned in previous sections. In particular, most of Prior's systems contained the past and future operators with truth conditions:

$$(50) \quad t \models \mathcal{P}A \text{ if and only if } \exists s(s < t \ \& \ s \models A)$$

$$t \models \mathcal{F}A \text{ if and only if } \exists s(t < s \ \& \ s \models A)$$

(where  $t \models A$  means  $A$  is true at time  $t$  and  $s < t$  means time  $s$  is before time  $t$ ). This allows (48) to be represented:

$$(48b) \quad \mathcal{P}Mgm, Wm \models \mathcal{P}\exists x(Mgx \wedge Wx).$$

Quine himself thought that a logic to help prevent us misrepresenting (48) as (48a) would be 'needlessly elaborate'. 'We do better,' he says, 'to make do with a simpler logical machine, and then, when we want to apply it, to paraphrase our sentences to fit it.' In this instance, Quine's attitude seems too rigid. The advantages of the simpler machine must be balanced against a more complicated paraphrase and representation. While (49a) may represent the form of (49), it does not seem to represent the form of (48) as well as (48b) does. But if our motivation for constructing new tense logics is to still better represent the logical forms of arguments and sentences of natural language, we should be mindful of Quine's worries about their being needlessly elaborate. We would not expect a logical representation to capture all the nuances of a particular tense construction in a particular language. We would expect a certain economy in logical vocabulary and rules of inference.

Motivations for many new systems of tense logic may be seen as more semantical than logical. A semantics should determine, for any declarative sentence  $S$ , context  $C$ , and possible world  $w$ , whether the thought expressed when  $S$  is uttered in  $C$  is true of  $w$ . As noted in previous sections, the truth conditions associated with Prior's  $\mathcal{P}$  and  $\mathcal{F}$  do not correspond very closely to those of English tenses. New systems of tense logic attempt to forge a closer correspondence. This might be done with the view that the tense logic would become a convenient *intermediary* between sentences of natural language and their truth conditions. That role was played by tensed intensional logic in Montague's semantics. An algorithm translates English sentences into formulas of that system and an inductive definition specifies truth conditions for the formulas. As noted above, Montague's appropriation of the Priorean connectives into his intensional logic make for a crude treatment of tense, but refined systems might serve better. Specifications of truth conditions for the tensed intensional logic (and, more blatantly for the refined tense logics), often seem to use a first order theory of temporal precedence (or containment, overlap, etc.) as yet another intermediary. (Consider clauses (50) above, for example.) One may wonder, then, whether it wouldn't be

better to skip the first intermediary and translate English sentences directly into such a first order theory. Certainly the most perspicuous way to give the meaning of a particular English sentence is often to ‘translate’ it by a formula in the language of the first order theory of temporal precedence, and this consideration may play a role in some of the complaints against tense logics found, for example, in [van Benthem, 1977] and [Massey, 1969]. Presumably, however, a *general* translation procedure could be simplified by taking an appropriate tense logic as the target language.

There is also another way to understand the attempt to forge a closer correspondence between tense logical connectives and the tense constructions of natural language. We may view tense logics as ‘toy’ languages, which, by isolating and idealizing certain features of natural language, help us to understand them. On this view, the tense logician builds models or simulations of features of language, rather than parts of linguistic theories. This view is plausible for, say Kamp’s logic for ‘now’ and Galton’s logic of aspect (see below), but it is difficult to maintain for more elaborate tense logics containing many operators to which no natural language expressions correspond.

Systems of tense logic are sometimes defended against classical first order alternatives on the grounds that they don’t commit language users to an ontology of temporal moments, since they don’t explicitly quantify over times. This defense seems misguided on several counts. First, English speakers do seem to believe in such an ontology of moments, as can be seen from their use of locutions like ‘at three o’clock sharp’. Second, it’s not clear what kind of ‘commitment’ is entailed by the observation that the language one uses quantifies over objects of a certain kind. Quine’s famous dictum, ‘to be is to be the value of a bound variable,’ was not intended to express the view that we are committed to what we quantify over in ordinary language, but rather that we are committed to what our best scientific theories quantify over, when these are cast in first order logic. There may be some weaker sense in which, by speaking English, we may be committing ourselves to the existence of entities like chances, sakes, average men and arbitrary numbers, even though we may not believe in these objects in any ultimate metaphysical sense. Perhaps we should say that *the language* is committed to such objects. (See Bach [1981].) But surely the proper test for this notion is simply whether the best interpretation of our language requires these objects: ‘to be is to be an element of a model.’ And, whether we employ tense logics or first order theories, our best models do contain (point-like and/or extended) times. Finally, even if one were sympathetic to the idea that the weaker notion of commitment was revealed by the range of first order quantifiers, there is reason to be suspicious of claims that a logic that properly models any substantial set of the temporal features of English would have fewer ontological commitments than a first order theory of temporal precedence. For, as Cresswell has argued in detail [1990; 1996],

the languages of such logics turn out to be equivalent in expressive power to the language of the first order theories. One might reasonably suppose in this case that the ontological commitments of the modal language should be determined by the range of the quantifiers of its first order equivalent. As discussed in Section 3.1, then, the proper defense of tense logic's replacement of quantifiers by operators is linguistic rather than metaphysical.

#### 4.2 Interval based logics

One of the most salient differences between the traditional tense logical systems and natural language is that all the formulas of the former are evaluated at instants of time, whereas at least some of the sentences of the latter seem to describe what happens at extended temporal periods. We are accustomed to thinking of such periods as comprising continuous stretches of instants, but it has been suggested, at least since Russell, that extended periods are the real objects of experience, and instants are abstractions from them. Various recipes for constructing instants from periods are contained in Russell [1914], van Benthem [1991], Thomason [1984; 1989] and Burgess [1984]. Temporal relations among intervals are more diverse than those among instants, and it is not clear which of these relations should be taken as primitive for an interval based tense logic. Figure 4.2 shows 13 possible relations that an interval  $A$  can bear to the fixed interval  $B$ .

We can think of  $<$  and  $>$  as precedence and succession,  $\ll$  and  $\gg$  as immediate precedence and succession and  $\subset$ ,  $\supset$ , and  $\circ$  as inclusion, containment and overlap. The subscripts  $l$  and  $r$  are for 'left' and 'right'. Under reasonable understandings of these notions and reasonable assumptions about the structure of time, these can all be defined in elementary logic from precedence and inclusion. For example,  $A \ll B$  can be defined by  $A < B \wedge \neg \exists x(A < x \wedge x < B)$ , and  $A \circ_l B$  by  $\exists x(x \subset A \wedge x < B) \wedge (\exists x)(x \subset A \wedge x \subset B) \wedge \exists x(x \subset A \wedge x > B)$ . It does not follow, however, that a tense operator based on any of these relations can be defined from operators based on  $<$  and  $\subset$ . Just as instant based tense logics include both  $\mathcal{P}$  and  $\mathcal{F}$  despite the fact that  $>$  is elementarily definable from  $<$ , we may wish to include operators based on a variety of the relations above in an interval based tense logic. For each of the relations  $R$  listed in the above chart, let  $[R]$  and  $\langle R \rangle$  be the box and diamond operators defined with  $R$  as the accessibility relation. (We are presupposing some acquaintance with the Kripke semantics for modal logics here. See Bull and Segerberg in this *Handbook* for background.) Then  $\langle < \rangle$  and  $\langle > \rangle$  are interval analogs of Prior's  $\mathcal{P}$  and  $\mathcal{F}$ , and  $\langle \subset \rangle$  is a connective that Dana Scott suggested as a rough analog of the progressive. Halpern and Shoham [1986] (and Shoham [1988]) point out that if we take the three converse pairs  $[\ll]$  and  $[\gg]$ ,  $[\subset_l]$  and  $[\supset_l]$  and  $[\subset_r]$  and  $[\supset_r]$  as primitive we can give simple definitions of the connectives associated with the remaining relations:

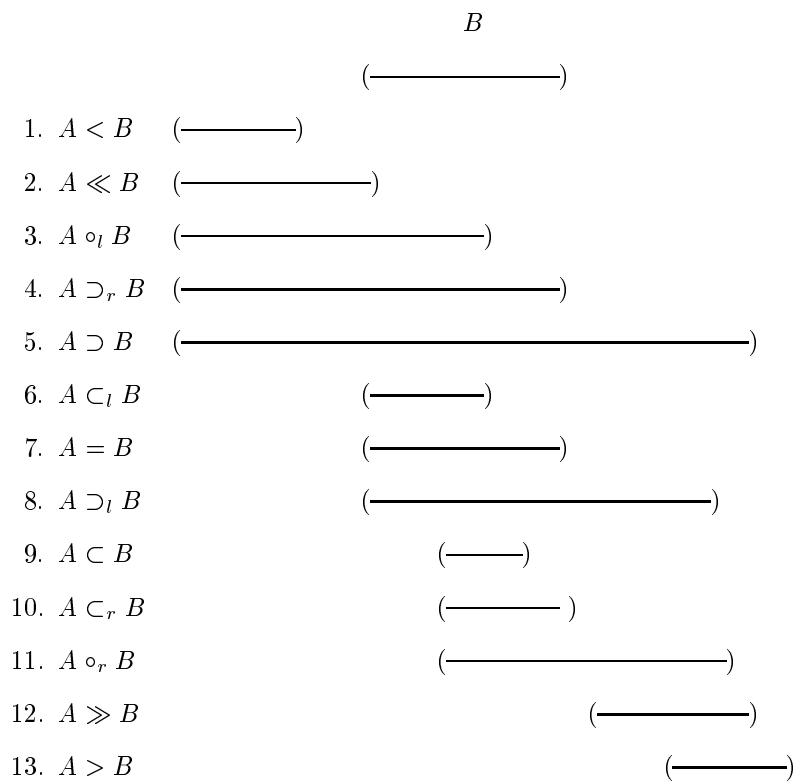


Figure 2.

$$[<] = [⟨⟨][⟨⟨] \quad [>] = [⟩⟩][⟩⟩]$$

$$[▷] = [▷_l][▷_r] \quad [◁] = [◁_l][◁_r]$$

$$[◦_l] = [▷_l][◁_r] \quad [◦_r] = [▷_r][◁_l]$$

If it is assumed that intervals always contain durationless atoms, i.e., subintervals  $s$  such that  $\neg\exists t(t \subset s)$ , then Venema shows that we can do better. For then  $[▷_l] \perp$  and  $[▷_r] \perp$  will be true only at atoms, and there are formulas  $[l]A = (A \wedge [▷_l] \perp) \vee \langle▷_l\rangle(A \wedge [▷_l] \perp)$  and  $[r]A = (A \wedge [▷_l] \perp) \vee \langle▷_r\rangle(A \wedge [▷_l] \perp)$  saying that  $A$  is true at the left and right ‘endpoints’ of an interval.  $[⟨⟨]$  and  $[⟩⟩]$  can now be defined by  $[r][◁_l]$  and  $[l][◁_r]$ . (The assumption that there are durationless ‘intervals’ undercuts the idea that instants are mere abstractions, but it seems appropriate for linguistic applications of tense logic, since language users do, at some level, presume the existence of both intervals and instants.)

Call the tense-logical language with operators  $[◁_l], [▷_l], [◁_r]$  and  $[▷_r]$ , HSV in honor of its inventors. Since HSV can so easily express all the relations on the table above, one might expect it to be sufficient to express any temporal relations that common constructions in natural language do. As Venema shows, however, there are limitations to its expressive power. Consider the binary connective  $\wedge^*$  such that  $(s, t) \models (A \wedge^* B)$  iff, for some  $r, s < r < t$ ,  $(s, r) \models A$  and  $(r, t) \models B$ . Lloyd Humberstone argues that  $\wedge^*$  is the tense logical connective that properly expresses temporal conjunction, i.e., *and* in the sense of *and next*. But no formula in HSV can express  $\wedge^*$ . Further, as Venema shows, there is a sense in which this expressive poverty is unavoidable in interval logics. Call a model  $M = (I, \subset_l, \subset_r, \supset_l, \supset_r, V)$  for HSV ‘instant generated’ if there is some nonempty set  $T$  ordered by  $<$  such that  $I$  is the set of all  $(x, y) \in (T \times T)$  for which  $x \leq y$ , and  $\subset_l, \subset_r, \supset_l$  and  $\supset_r$  are the appropriate relations on  $I$ . (For example  $(r, s) \supset_l (u, v)$  iff  $u = r$  and  $s > v$ .) Instant generated HSV-models, then, are models in which formulas are evaluated at pairs of indices, i.e., they are two-dimensional models. The truth conditions for the connectives determine a translation that maps formulas of HSV to ‘equivalent’ formulas in predicate logic with free variables  $r$  and  $s$ . Similar translations could be obtained for any language in which the truth conditions of the connectives can be expressed in elementary logic. Venema shows, however, that for no finite set of connectives will this translation include in its range every formula with variable  $r$  and  $s$ . This result holds even when the equivalent formulas are required to agree only on models for which the instants form a dense linear order. This contrasts with a fundamental result in instant-based tense logics, that for dense linear orders, the two connectives ‘since’ and ‘until’ are sufficient to express everything that can be said in elementary logic with one free variable. (See Burgess [2001]).

Several authors have suggested that in tense logics appropriate for natural language there should be constraints on the set of intervals at which a formula can be true. The set  $\|A\|_M$  of indices at which formula  $A$  is true in model  $M$  is often called the truth set of  $A$ . Humberstone requires that valuations be restricted so that truth sets of sentence letters be closed under containment. That ‘downward closure’ property seems natural for stative sentences (see Section 3.4.2). The truth of *The cat is on the mat* at the interval from two to two to two thirty apparently entails its truth at the interval from two ten to two twenty. But downward closure is not preserved under ordinary logical negation. If *The cat is on the mat* is true at (2:00,2:30) and all its subintervals, but not at (1:30, 3:00) then  $\neg(\textit{the cat is on the mat})$  is true at (1:30,3:00) but not all of its subintervals. Humberstone suggests a stronger form of negation, which we might call  $[\neg]$ .  $[\neg]A$  is true at interval  $i$  if  $A$  is false at all subintervals of  $i$ . Such a negation may occur in one reading of *The cat isn't on the mat*. It can also be used to express a more purely tense logical connective:  $[\supset]$  can be defined as  $[\neg][\neg]$ . We obtain a reasonable tense logic by adding the standard past and future connectives  $\langle \rangle$  and  $\langle \rangle$ .

Statives also seem to obey an upward closure constraint. If  $A$  is true in each of some sequence of adjoining or overlapping intervals, it is also true in the ‘sum’ of those intervals. Peter Röper observes that, in the presence of downward closure, upwards closure is equivalent to the condition that  $A$  is true in  $i$  if it is true ‘almost everywhere’ in  $i$ , i.e., if every subinterval of  $i$  contains a subinterval at which  $A$  is true. (See Burgess [1982a] for an interesting list of other equivalents of this and related notions.) Following Röper, we may call a truth set homogeneous if it satisfies both upwards and downwards closure. Humberstone’s strong negation preserves homogeneity, but the tense connectives  $\langle \rangle$  and  $\langle \rangle$  do not. For suppose the temporal intervals are the open intervals of some densely ordered set of instants, and  $A$  is true only at  $(s, t)$  and its subintervals. Then the truth set of  $A$  is homogeneous. But every proper subinterval of  $(s, t)$  verifies  $\langle \rangle A$ , and so every subinterval of  $(s, t)$  contains a subinterval that verifies the formula, whereas  $(s, t)$  itself does not verify the formula, and so the truth set of  $\langle \rangle A$  is not homogeneous. To ensure that homogeneity is preserved, Röper replaces the standard truth conditions for the future operator by a condition stating that  $\langle \rangle A$  is true at  $i$  if every subinterval of  $i$  contains a subinterval  $i'$  such that  $A$  is true at some  $w > i'$ . The past operator is similarly altered. This ensures that all formulas have homogeneous truth sets and the resulting system admits a simple axiomatization. One may wonder whether the future and past tenses of statives really are themselves statives in natural language, and thus whether homogeneity really ought to be preserved. But if one is thoroughgoing (as Humberstone and Röper seem to be, but Venema does not), about the attitude that (extended) intervals are the genuine temporal objects, then it does seem reasonable to suppose that for stative  $A$ ,  $A \rightarrow$



$\langle \rangle A$  and  $A \rightarrow \langle \rangle A$  are logical truths. If the cat is on the mat, then, if one looks sufficiently close to the present, it will be on the mat, and if one looks sufficiently close in the other direction, it was on the mat. For otherwise we would have to believe that the present was the instant at which it came or left. Indeed, the ‘present implies past’ property was cited by Aristotle (in *Metaphysics IX*) as a distinguishing feature of ‘energaie,’ a category that surely includes the statives. The formulas  $A \rightarrow \langle \rangle A$  and  $A \rightarrow \langle \rangle A$  are not theorems of HSV or standard tense logics unless  $\langle$  is reflexive, but they are theorems of Röper’s homogeneous interval tense logic.

### 4.3 ‘Now’, ‘then’, and keeping track of times

Another way in which natural language differs from Priorean tense logics is its facility in conveying that the eventualities described in various scattered clauses of a sentence obtain simultaneously. Consider first an example in which exterior and interior clauses describe what obtains at the moment of utterance.

(51) This is 1996 and one day everyone now alive will be dead.

If we represent this as  $P \wedge \forall x(Lx \rightarrow \mathcal{F}Dx)$ , we fail to imply that those alive today will all be dead at a common future moment. If we pull the future operator outside the quantifier, we get  $P \wedge \mathcal{F}\forall x(Lx \rightarrow Dx)$ , which wrongly implies that there will be a time when live people are (simultaneously) dead. A solution (following Kamp [1971] and Prior [1968]) is to evaluate formulas at pairs of times, the first of which ‘keeps track’ of the moment of utterance and the second of which is used to evaluate expressions inside tense operators.  $(s, t) \vDash A$  can be understood as asserting that  $A$  is true at  $t$  when part of an expression uttered at  $s$ . The truth conditions for the Priorean operators use the second coordinate:  $(s, t) \vDash \mathcal{P}A$  iff  $\exists t' < t(s, t') \vDash A$  and  $(s, t) \vDash \mathcal{F}A$  iff  $\exists t' > t(s, t') \vDash A$ . A new connective  $\mathcal{N}$  corresponding to the adverb *now* is added satisfying  $(s, t) \vDash \mathcal{N}A$  iff  $(s, s) \vDash A$ . Validity in a model is to be understood as truth whenever uttered, i.e.,  $M \vDash A$  iff for every time  $t$  in  $M$ ,  $(t, t) \vDash A$ . On this understanding  $A \leftrightarrow \mathcal{N}A$  is valid, so it may appear that  $\mathcal{N}$  is vacuous. Its effect becomes apparent when it appears within the scope of the other tense operators.  $\mathcal{P}(A \leftrightarrow \mathcal{N}A)$ , for example, is false when  $A$  assumes a truth value at utterance time that differs from the value it had until then. This condition can still be expressed without the new connective by  $(A \wedge \neg \mathcal{P}A) \vee (\neg A \wedge \neg \mathcal{P}\neg A)$ , and in general, as Kamp shows,  $\mathcal{N}$  is eliminable in propositional Priorean tense logics. If the underlying language has quantifiers, however,  $\mathcal{N}$  does increase its expressive power. For example, the troublesome example above can be represented as

(51a)  $P \wedge \mathcal{F}\forall x(\mathcal{N}Lx \rightarrow Dx)$ .

The new connective can be used to ensure that embedded clauses get evaluated after the utterance moment as well as simultaneously with it. Consider Kamp's

(52) A child was born who will be king.

To represent this as  $\mathcal{P}(A \wedge \mathcal{F}B)$  would imply only that the child is king after its birth. To capture the sense of the English *will*, that the child is king after the utterance moment, we need  $\mathcal{P}(A \wedge \mathcal{N}\mathcal{F}B)$ .

Vlach [1973] shows that in a somewhat more general setting  $\mathcal{N}$  can be used to cause evaluation of embedded clauses at still other times. Take the sentence *It is three o'clock and soon Jones will cite all those who are now speeding*, which has a structure like (51), and put it into the past:

(53) It was three o'clock and Jones would soon cite those who were then speeding.

We cannot represent this by simply applying a past operator to (51a) because the resulting formula would imply that Jones was going to ticket those who were speeding at the time of utterance. Vlach suggests we add an 'index' operator to the language with truth conditions very similar to  $\mathcal{N}$ 's

$$\langle s, t \rangle \models \mathcal{I}A \text{ iff } \langle t, t \rangle \models A.$$

If an  $\mathcal{N}$  occurs within the scope of an  $\mathcal{I}$  it can be read as *then*. This allows, for example, the sentence (44) to be represented as

$$\mathcal{P}\mathcal{I}(P \wedge \forall x(\mathcal{N}Sx \rightarrow Cx)).$$

In general, if  $A$  contains no occurrence of  $\mathcal{I}$ , the utterance time is 'fixed' in the sense that the truth value of  $A$  at  $\langle u, t \rangle$  depends on the truth values of its subformulas at pairs  $\langle u, t' \rangle$ . The occurrence of an  $\mathcal{I}$  'shifts' the utterance time so that evaluating  $A$  at  $\langle u, t \rangle$  may require evaluating the subformulas that are within the scope of the  $\mathcal{I}$  at pairs  $\langle u', t' \rangle$  for  $u'$  different than  $u$ .

With Kamp's *now*, we can keep track of the utterance time and one other time. With Vlach's *then*, we still track two times, although neither need coincide with utterance. Several authors have suggested that a tense-logical system adequate to represent natural language must allow us to keep track of more than two times. The evidence is not entirely convincing, but it has motivated some interesting revisions in the Priorean framework. Gabbay [1974; 1976] points to examples like the following:

(54) John said he would come.

(55) Ann will go to a school her mother attended and it will become better than Harvard,

which, he maintains, have interpretations suggested by the formulas

(54a)  $\exists t_1 < t_0$ (John says at  $t_1$  that  $\exists t_2(t_1 < t_2 < t_0 \wedge$  John comes at  $t_2$ ))

(55a)  $\exists t_1 > t_0 \exists s$ ( $s$  is a school Ann goes to  $s$  at  $t_1 \wedge \exists t_2 < t_0$ (Ann's mother goes to  $s$  at  $t_2 \wedge \exists t_3 > t_1$ ( $s$  is better than Harvard at  $t_3$ ))).

Saarinen's exhibits include

(56) Every man who ever supported the Vietnam War believes now that one day he will have to admit that he was an idiot then,

interpreted as

(56a)  $\forall x(x$  is a man  $\rightarrow \forall t_1 < t_0(x$  supports the Vietnam War at  $t_1)(x$  believes at  $t_0$  that  $\exists t_2 > t_0(x$  has to admit at  $t_2$  that  $x$  is an idiot at  $t_1))$ ,

and (57) Joe said that a child had been born who would become ruler of the world,

which, Saarinen argues, has at least the two readings

(57a)  $\exists t < t_0$ (Joe says at  $t$  that  $\exists s < t \exists x$ (Child  $x \wedge$  Born  $xs \wedge \exists u > s$  Ruler  $xu$ ))

(57b)  $\exists t < t_0$ (Joe says at  $t$  that  $\exists s < t \exists x$ (Child  $x \wedge$  Born  $xs \wedge \exists u > t$  Ruler  $xu$ ))

according to whether the sentence reported is *A child was born who would become ruler*, or *A child was born who will become ruler*. (Note that the sequence of tense theories discussed in Section 3.4.3 above conflict with the readings proposed here for (54) and (57).)<sup>14</sup>

Cresswell [1990] points to examples of a more explicitly quantificational form:

<sup>14</sup>They hold that requirement in (54a) that  $t_2$  precede  $t_0$  is not part of the truth conditions for (54) (though it may be implicated). Similarly, they hold that (57a) is the sole reading of (57).

(58) There will be times such that all persons now alive will be  $A_1$  at the first or  $A_2$  at the second or...  $A_n$  at the  $n$ th.

(58a)  $\exists t_1 \dots \exists t_n (t_0 < t_1 \wedge \dots \wedge t_0 < t_n \wedge \forall x (x \text{ is alive at } t_0 \rightarrow (x \text{ is } A_1 \text{ at } t_1 \vee \dots \vee x \text{ is } A_n \text{ at } t_n)))$ .

Some of the troublesome examples could be expressed in a Priorean language. For example, for (55) we might propose:

(55b)  $\exists s (\text{SCHOOL}(s) \wedge \mathcal{P} \text{ATTEND}(\text{ann's mother}, s) \wedge \mathcal{F}(\text{ATTEND}(\text{ann}, s) \wedge \mathcal{F} \text{BETTER}(s, \text{harvard})))$

But as a toy version of (55) or the result of applying a uniform English-to-tense-logic translation procedure, this may seem implausible. It requires a reordering of the clauses in (55), which removes *that her mother attended* from inside the scope of the main tense operator. Other troublesome examples can be represented with the help of novel two-dimensional operators. For example, Gabbay suggests that the appropriate reading of (54) might be represented  $\mathcal{P} \text{Johnsaythat} \mathcal{F}_2 A$ , where  $\langle u, t \rangle \models \mathcal{F}_2 A$  iff either  $t < u$  and  $\exists s (t < s < u \wedge \langle u, s \rangle \models A)$  or  $u < t$  and  $\exists s (u < s < t \wedge \langle u, s \rangle \models A)$ . (A variety of other two dimensional tense operators are investigated in Åqvist and Guenther ([1977; 1978]). This approach, however, seems somewhat *ad hoc*. In the general case, Gabbay argues, “we must keep record of the entire sequence of points that figure in the evaluation of a formula] and not only that, but also keep track of the kind of operators used.”

We sketch below five more general solutions to the problem of tracking times. Each of these introduces an interesting formal system in which the times that appear at one stage in the evaluation of a formula can be remembered at later stages, but none of these seems to provide a fully accurate model of the time-tracking mechanisms of natural language.

#### 4.3.1 Backwards-looking operators (Saarinen)

Add to the language of tense logic a special ‘operator functor’  $\mathbf{D}$ . For any operator  $\square$ ,  $\mathbf{D}(\square)$  is a connective that ‘looks back’ to the time at which the preceding  $\square$  was evaluated. For example, (47) can be represented

(56b)  $\forall x (x \text{ is a man} \rightarrow \neg \mathcal{P} \neg (x \text{ supported the Vietnam war} \rightarrow \mathbf{D}(\mathcal{P})(x \text{ believesthat } \mathcal{F}(x \text{ hastoadmitthat } \mathbf{D}(\mathbf{D}(\mathcal{P}))(x \text{ is an idiot}))))))$

if we have the appropriate **believesthat** and **hastoadmitthat** operators. Within a more standard language,

(59)  $A \wedge \mathcal{F}(B \wedge \mathcal{P}(C \wedge \mathcal{F}(\mathbf{D} \wedge \mathbf{D}(\mathcal{P})E) \wedge \mathbf{D}(\mathcal{F})F)$

is true at  $w$  iff  $\exists x\exists y\exists z(w < x, y < x, y < z, w \vDash A, x \vDash B, y \vDash C, z \vDash D, x \vDash E$  and  $y \vDash F)$ . In this example  $\mathbf{D}(\mathcal{P})$  and  $\mathbf{D}(\mathcal{F})$  ‘look back’ to the times at which the preceding  $\mathcal{P}$  and  $\mathcal{F}$  were evaluated, namely,  $x$  and  $y$ . This condition can be expressed without the backwards operators by

$$(59a) \quad A \wedge \mathcal{F}(B \wedge E \wedge \mathcal{P}(C \wedge F \wedge \mathcal{F}D)),$$

but (as with (55b)) this requires a reordering of the clauses, and (as with (56b)) the reordering may be impossible in a richer formal language. It is a little hard to see how the semantics for  $\mathbf{D}$  might be made precise in Tarski-style truth definition. Saarinen suggests a game-theoretic interpretation, in which each move is made with full knowledge of previous moves. Iterated  $\mathbf{D}(\Box)$ ’s look back to more distant  $\Box$ ’s so that, for example,

$$A \wedge \mathcal{P}(B \wedge \mathcal{F}(C \wedge \mathcal{F}(D \wedge \mathbf{D}(\mathcal{F})\mathbf{D}(\mathcal{F})E) \wedge \mathbf{D}(\mathcal{P})F))$$

is true at  $w$  iff  $\exists x\exists y\exists z(x < w, x < y < z, w \vDash A, x \vDash B, y \vDash C, z \vDash D, x \vDash E$  and  $w \vDash A)$ . Logics based on this language would differ markedly from traditional ones. For example, if time is dense  $\mathcal{F}A \rightarrow \mathcal{F}\mathcal{F}A$  is valid when  $A$  does not contain  $\mathbf{D}$ ’s, but not when  $A$  is of the form  $\mathbf{D}(\mathcal{F})B$ .

#### 4.3.2 Dating sentences (Blackburn [1992; 1994])

Add a special sort of sentence letters, each of which is true at exactly one moment of time. Blackburn thinks of these as naming instants and calls his systems ‘nominal tense logics,’ but they are more accurately viewed as ‘dating sentences’, asserting, for example *It is now three pm on July 1, 1995*. Tense logical systems in this language can be characterized by adding to the usual tense logical axioms the schema

$$n \wedge \mathcal{E}(n \wedge A) \rightarrow A$$

where  $n$  is a dating sentence and  $\mathcal{E}$  is any string of  $\mathcal{P}$ ’s and  $\mathcal{F}$ ’s. In place of (59), we can now write:

$$(59b) \quad A \wedge \mathcal{F}(B \wedge i \mathcal{P}(C \wedge j \wedge \mathcal{F}D)) \wedge \mathcal{P}\mathcal{F}(i \wedge E) \wedge \mathcal{P}\mathcal{F}(j \wedge F).$$

Here  $i$  and  $j$  ‘date’ the relevant times at which  $B$  and  $C$  are true, so that the truth of  $i \wedge E$  and  $j \wedge F$  requires the truth of  $E$  and  $F$  at those same times.

#### 4.3.3 Generalization of $\mathcal{N}\text{-}\mathcal{I}$ (Vlach [1973, appendix])

To the language of Priorean tense logic, add connectives  $\mathcal{N}_i$  and  $\mathcal{I}_i$  for all non-negative integers  $i$ . Let formulas be evaluated at pairs  $(s, i)$  where  $s = (s_0, s_1, \dots)$  is an infinite sequence of times and  $i$  is a non-negative integer, specifying the coordinate of  $s$  relevant to the evaluation.  $\mathcal{N}_i A$  indicates that

$A$  is to be evaluated at the time referred to when  $\mathcal{I}_i$  was encountered. More precisely,

$$(\mathbf{s}, i) \vDash \mathcal{P}A \text{ iff } \exists t < s_i((s_0, \dots, s_{i-1}, t, s_{i+1}, \dots), i) \vDash A$$

$$(\mathbf{s}, i) \vDash \mathcal{F}A \text{ iff } \exists t < s_i((s_0, \dots, s_{i-1}, t, s_{i+1}, \dots), i) \vDash A$$

$$(\mathbf{s}, i) \vDash \mathcal{I}_j A \text{ iff } ((s_0, \dots, s_{j-1}, s_i, s_{j+1}, \dots), i) \vDash A$$

$$(\mathbf{s}, i) \vDash \mathcal{N}_j A \text{ iff } (\mathbf{s}, j) \vDash A$$

The truth of sentence letters at  $(\mathbf{s}, i)$  depend only on  $s_i$  and formulas are to be considered valid in a model if they are true at all pairs  $((t, t, \dots), 0)$ . In this language (59) can be expressed

$$(59c) \quad A \wedge \mathcal{F}\mathcal{I}_1(B \wedge \mathcal{P}\mathcal{I}_2(C \wedge \mathcal{F}(D \wedge \mathcal{N}_2 E \wedge \mathcal{N}_1 F))).$$

Here  $\mathcal{I}_1$  and  $\mathcal{I}_2$  ‘store’ in  $s_1$  and  $s_2$  the times at which  $B$  and  $C$  are evaluated and  $\mathcal{N}_2$  and  $\mathcal{N}_1$  shift the evaluation to  $s_2$  and  $s_1$ , causing  $F$  and  $E$  to be evaluated at times there stored.

#### 4.3.4 The backspace operator (Vlach [1973, appendix])

Add to the language of Priorean tense logic a single unary connective  $\mathbf{B}$ . Let formulas be evaluated at finite (nonempty) sequences of times according to the conditions:

$$(t_1, \dots, t_n) \vDash \mathcal{P}A \text{ iff } \exists t_{n+1} < t_n((t_1, \dots, t_{n+1}) \vDash A)$$

$$(t_1, \dots, t_n) \vDash \mathcal{F}A \text{ iff } \exists t_{n+1} > t_n((t_1, \dots, t_{n+1}) \vDash A)$$

$$(t_1, \dots, t_{n+1}) \vDash \mathbf{B}A \text{ iff } (t_1 \dots, t_n) \vDash A$$

$$(\text{and, if } n = 0, (t_1) \vDash \mathbf{B}A \text{ iff } (t_1) \vDash A)$$

The truth value of sentence letters depends only on the last time in the sequence, and formulas are considered valid in a model when they are true at all length-one sequences. (59) is now represented

$$(59d) \quad A \wedge \mathcal{F}(B \wedge \mathcal{P}(C \wedge \mathcal{F}(D \wedge \mathbf{B} E \wedge \mathbf{B} B F))).$$

The indices of evaluation here form a stack. In the course of evaluating a formula a new time is pushed onto the stack whenever a Priorean tense connective is encountered and it is popped off whenever a  $\mathbf{B}$  is encountered. Thus,  $\mathbf{B}$  is a ‘backspace’ operator, which causes its argument to be evaluated at the time that had been considered in the immediately preceding stage of evaluation. In terms of this metaphor, Kamp’s original ‘now’ connective

was, in contrast, a ‘return’ operator, causing its argument to be evaluated at the time that was given at the initial moment of evaluation.

#### 4.3.5 Generalization of $\mathcal{N}\text{-}\mathcal{I}$ (Cresswell [1990])

Generalize the language of Vlach’s  $\mathcal{N}\text{-}\mathcal{I}$  system just as in solution 3. Let formulas be evaluated at infinite sequences of times and let the truth definition contain the following clauses:

$$(s_0, s_1, s_2, \dots) \models \mathcal{P}A \text{ iff } \exists s < s_0 ((s, s_1, s_2, \dots) \models A)$$

$$(s_0, s_1, s_2, \dots) \models \mathcal{F}A \text{ iff } \exists s > s_0 ((s, s_1, s_2, \dots) \models A)$$

$$(s_0, s_1, \dots, s_i, \dots) \models \mathcal{I}_i A \text{ iff } (s_0, s_1, \dots, s_{i-1}, s_0, s_{i+1}, \dots) \models A$$

$$(s_0, s_1, \dots, s_i, \dots) \models \mathcal{N}_i A \text{ iff } (s_i, s_1, s_2, \dots) \models A$$

A formula is considered valid if it is true at all constant sequences  $(s, s, \dots)$ . Then we can express (59) above as:

$$(59e) \quad A \wedge \mathcal{F}\mathcal{I}_1(B \wedge \mathcal{P}\mathcal{I}_2(C \wedge \mathcal{F}(D \wedge \mathcal{N}_2 E \wedge \mathcal{N}_1 F))).$$

As in solution 3,  $\mathcal{I}_1$  and  $\mathcal{I}_2$  store in  $s_1$  and  $s_2$  the times at which  $B$  and  $C$  are evaluated. Subsequent occurrences of  $\mathcal{N}_2$  and  $\mathcal{N}_1$  restore those times to  $s_0$  so that  $E$  and  $F$  can be evaluated—with respect to them.

Each of the systems described in 4.3.1–4.3.5 has a certain appeal, and we believe that none of them has been investigated as thoroughly as it deserves. We confine ourselves here to a few remarks about their expressive powers and their suitability to represent tense constructions of natural language. Of the five systems, only Cresswell’s  $\mathcal{N}\text{-}\mathcal{I}$  generalization permits atomic formulas to depend on more than one time. This makes it possible, for example, to represent *Johnson ran faster than Lewis*, meaning that Johnson ran faster in the 1996 Olympics than Lewis did in the 1992 Olympics, by Rmn. We understand  $R$  to be a predicate (*runs faster than*) which, at every pair of times, is true or false of pairs of individuals. Since the issues involved in these representations are somewhat removed from the ones discussed here, and since the other systems could be generalized in this way if desired, this difference is not significant. If we stipulate that the truth value of a sentence letter at  $s$  in Cresswell’s system depends only on  $s_0$  then, for each of the systems, there is a translation of formulas into the classical first order language with identity and a countable collection of temporally monadic predicates and a single temporally dyadic predicate  $<$  (and, in the case of nominal tense logic, a countable collection of temporal constants). We say ‘temporally’ monadic and dyadic because, if the base language of these systems is the language of predicate logic, it will already

contain polyadic predicates that apply to tuples of *individuals*. The translation maps these to predicates with an additional temporal argument, and it maps tense formulas with free individual variables into classical formulas with those same free variables and additional free temporal variables. The sentential version of Cresswell's  $\mathcal{N}$ - $\mathcal{I}$  provides an example. Associate with each sentence letter  $p$  a unary predicate letter  $p^\tau$  and fix two (disjoint) sequences of variables  $x_0, x_1, \dots$  and  $y_0, y_1, \dots$ . A translation  $\tau$  from Cresswell-formulas into classical formulas is defined by the following clauses (where  $A^x/y$  is the result of replacing all free occurrences of  $y$  in  $A$  by  $x$ ):

$$\text{i) } \tau p = p^\tau x_0$$

$$\text{ii) } \tau \mathcal{P}A = \exists y < x_0 (\tau A)^y /_{x_0}, \text{ where } y \text{ is the first } y_i \text{ that does not occur in } \tau A$$

$$\text{iii) } \tau \mathcal{F}A = \exists y > x_0 (\tau A)^y /_{x_0}, \text{ where } y \text{ is as above}$$

$$\text{iv) } \tau \mathcal{I}_j A = (\tau A)^{x^0} /_{x_j}$$

$$\text{v) } \tau \mathcal{N}_j A = (\tau A)^{x^j} /_{x_0}$$

To every model  $M$  for Cresswell's language there corresponds a classical model  $M'$  with the same domain which assigns to each predicate letter  $p^\tau$  the set of times at which  $p$  is true in  $M$ .  $\tau A$  expresses  $A$  in the sense that  $(s_0, s_1, \dots) \models_M A$  iff  $\tau A$  is true in  $M'$  under the assignment that assigns  $s_i$  to  $x_i$  for  $i = 0, 1, \dots$ . Viewing  $M$  and  $M'$  as the same model, we can say that a tense-logical formula expresses a classical one when the two formulas are true in the same models. (Of course in defining a tense-logical system, we may restrict the class of appropriate models. By 'true in the same models' we mean true in the same models appropriate for the tense logic.) A formula with one free variable in the first order language with unary predicates and  $<$  might be called a 'classical tense'. From the translation above we may observe that every Cresswell formula in which each occurrence of a connective  $\mathcal{N}_j$  lies within the scope of an occurrence of  $\mathcal{I}_j$  expresses a classical tense. If every classical tense is expressible in tense-logical system, the system is said to be temporally complete.

An argument in Chapter IV of Cresswell establishes that, as long as  $<$  is assumed to be connected (so that quantification over times can be expressed in the tense language), every classical tense without  $<$  can be expressed in his generalization of the  $\mathcal{N}$ - $\mathcal{I}$  language. It is not difficult to see that this holds as well for Vlach's generalization. For consider the following translation  $\tau$  mapping Cresswell's system into Vlach's:



$\tau A = \mathcal{N}_0 A$  if  $A$  is a sentence letter,

$\tau \mathcal{P} A = \mathcal{P} \tau A$ ,

$\tau \mathcal{F} A = \mathcal{F} \tau A$ ,

$\tau \mathcal{I}_i A = \mathcal{I}_i \tau A$ ,

$\tau \mathcal{N}_i A = \mathcal{I}_x \mathcal{I}_{x+1} \dots \mathcal{I}_{2x} \mathcal{N}_i \mathcal{I}_{x-i} \mathcal{N}_{x-i} \tau A$  where  $x$  is the successor of the least integer greater than every subscript that occurs in  $\mathcal{N}_i A$ .

Then, using the subscripts  $C$  and  $V$  for Vlach's system and Cresswell's,  $\mathfrak{s} \models_C A$  iff  $(\mathfrak{s}, 0) \models_V \tau A$ . So, if  $A$  is a classical tense without  $<$ , there is a formula  $A_C$  that expresses  $A$  in Cresswell's system, and  $\tau A_C$  will express  $A$  in Vlach's system.

The question of whether every classical tense is expressible is more difficult. As we saw with Kamp's  $\mathcal{N}$ , questions about expressive power are sensitive to the underlying language.  $\mathcal{N}$  adds nothing to the expressive power of sentential tense logic, but it does add to the expressive power of predicate tense logic. The examples suggest that the same is true of the backwards-looking and backspace operators. A well known result of Kamp (see Burgess [2001]) states that, if time is like the reals, every tense can be expressed with the connectives  $U$  (*until*) and  $S$  (*since*) with truth conditions  $U(A, B)$  iff  $\exists t > t_0 (t \models A \wedge \forall s (t_0 < s < t \rightarrow s \models B))$  and  $S(A, B)$  iff  $\exists t < t_0 (t \models A \wedge \forall s (t < s < t_0 \rightarrow s \models B))$ . By constructing a pair of models that can be discriminated by formulas with  $U$  and  $S$  but not by any Priorean formulas, one can show that Priorean tense logic is not temporally complete. A reduction of the sentential backwards-looking and backspace systems to the ordinary ones, therefore, would imply their temporal incompleteness. From the pairs of ordinary models that are indistinguishable by Priorean formulas, we can easily construct pairs that are indistinguishable in the language of Blackburn's dating sentences. (Pick corresponding times  $t$  and  $t'$  in the two models and require that every dating sentence be true exactly at  $t$  in the first model and exactly at  $t'$  in the second.) So that system also fails to be temporally complete.<sup>15</sup>

For a number of reasons, the suitability of a system of tense logic for natural language should not be identified with its expressive power, and the observation that the formulas in the five systems described here are all expressible as classical tenses does not imply that the language of classical tenses is itself a suitable tense logical system. Although we can express all

<sup>15</sup>There is a weaker sense in which  $U$  and  $S$  can be expressed with dating sentences. Let  $U(i, A, B)$  be  $i \wedge \mathcal{F}(A \wedge \neg \mathcal{P} \neg (\mathcal{F}_i \vee i \vee B))$  and  $S(i, A, B)$  be  $i \wedge \mathcal{P}(A \wedge \neg \mathcal{F}(\mathcal{P}_i \vee i \vee B))$ . Then  $U(i, A, B)$  is satisfiable in Blackburn's system iff  $U(A, B)$  is satisfiable in the since-until system and  $S(i, A, B)$  is satisfiable iff  $S(A, B)$  is.

the classical tenses in English, it is not the tense mechanism that allows us to do so. English sentences like *For every instant  $t$ , if  $t$  succeeds  $t_0$  there is an instant  $t'$ , such that  $t'$  succeeds  $t$  and  $t$  succeeds  $t_0$  and John is asleep at  $t'$* , however useful in explaining the meaning of first order formulas, are not the sort of sentences for which one would expect to find a phrase-by-phrase representatives in an idealized language isolating the tense-and-aspect features of English. One can object to Saarinen's  $\mathbf{D}$ , Blackburn's dating sentences, Vlach's  $\mathbf{B}$ , and Vlach and Cresswell's  $\mathcal{L}_j$ 's and  $\mathcal{N}_j$ 's on similar grounds. It is possible, of course, that some of these systems make particularly good intermediaries between tense constructions of natural language and truth conditions, or that there is some other sense in which they are especially suitable as tense logics for natural language, but such claims need arguments beyond demonstrations of expressive capacity. Indeed the fact that we can express very simply in these languages ideas that in English require complex constructions (perhaps involving quantifier phrases variable expressions) suggests that they are unsuitable on some conceptions of tense logic. On the other hand, if there are ideas we can express simply and uniformly in English, the mere observation that a tense-logical system has sufficient expressive power to somehow express them, may not be evidence in favor of the system. For example, the fact that prefixing a sufficiently long string of backspace operators to an embedded formula causes it to be evaluated at the moment of utterance does not mean that the backspace system is a good model of the English *now*.

Part of the difficulty in judging the adequacy of tense logical systems for natural language is discerning the linguistic data itself. It is not clear, for example, whether *John said he would come* does have the reading indicated in (54a) implying that he said he would come *by now*, or whether that inference, when legitimate, is based on (extralinguistic) contextual cues. Similarly, the observation that *Joe said that a child had been born who would become ruler of the world* is consistent with two possible utterances by Joe does not establish that (57) is ambiguous between (57a) and (57b). Saarinen maintains that a sentence of the form *A reported that B believed that C said John would go* has at least four readings, according to whether John's alleged departure occurs in the future of C's saying, B's believing, A's reporting, or the utterance time. Since the first of these readings is true if any of the others is, one can't expect to find a case which *requires* readings other than the first. The plausibility of there being such readings is undermined by observation that a similar ambiguity does not occur when the *would* is in the scope of future operators. *A will report (next week) that B said (the previous day) that C would go* is not made true by A's reporting next week that B said 'C went yesterday,' as it would if 'C would go' could refer to a time future to the utterance moment. While an adequate logic for the tenses of natural language may require greater 'time-tracking' capabilities than Priorian tense logic, there is not strong evidence for the

thesis that it be able to ‘remember’ at each stage in the evaluation times at which previous clauses were evaluated.

#### 4.4 Galton’s logic of aspects and events

English discourse presumes a universe of events and states with internal structure as well as temporal location. The language of Priorean tense logic is built solely from formulas, boolean connectives, and operators of temporal location. It is reasonable to try to enrich the language so that more of the internal structure of events can be described. In recent years there has been a proliferation of work in this area motivated by concerns in deontic logic and action theory. (See, for example, Jones and Sergot [1996] and the references therein.) For the most part, however, that work has not focussed on temporal or natural language considerations. There is a large and growing semantics literature on events and aspect, but much of it is too detailed to be considered part of a ‘logic’ of tense. In this section we sketch some ideas in the spirit of Galton [1984; 1987a], which seem to strike a good balance between simplicity and fidelity to ‘surface’ phenomena of English.

The idea that sentences in the future and present perfect can be represented by attaching  $\mathcal{F}$  and  $\mathcal{P}$  to some more basic sentence is plausible for sentences describing states but not those describing events. *The cat has been on the mat* is true now if *the cat is on the mat* was true before, but to say that *John has built a house* is true now if *John builds a house* was true before is confusing, since we don’t normally use the present tense to indicate that an event is true at the present time. (Indeed, since events like house building occur over extended intervals, it is not clear what the ‘present’ time would be in this case.) Let us instead add a class of *event letters*  $E_1, E_2, \dots$  to the language along with two event-to-formula *e-f aspect operators*  $\mathcal{P}erf$  and  $\mathcal{P}ros$ , which attach to event letters to produce formulas.<sup>16</sup> (The tense operators  $\mathcal{P}$  and  $\mathcal{F}$  and the boolean operators, as usual, apply to formulas to form formulas.) Let us provisionally say that an interpretation assigns to each event letter a set  $I(E)$  of *occurrence intervals*.  $\mathcal{P}rosE$  is true at  $t$  if  $t$  precedes (all of) some interval in  $I(E)$ ;  $\mathcal{P}erfE$ , if  $t$  succeeds (all of) some interval in  $I(E)$ . One may wonder what hinges on the distinction between an event’s occurring at a time and a formula’s being true at a time. Granting that we don’t normally say that *John builds a house* is true, say, in the spring of 1995, we might find it convenient to stipulate that it be true then if one of John’s house buildings occurs at that time. One advantage of not doing so is that the event/formula distinction

<sup>16</sup>Galton uses the label ‘imperfective’ in place of ‘e-f’, and the label ‘perfective’ in place of our *f-e*.

provides a sorting that blocks inappropriate iterations of aspect operators. Another is that the distinction makes it possible to retain the Priorean notion that all formulas are to be evaluated at instants even when the events they describe occupy extended intervals. The tense logical systems that result from this language so interpreted will contain the usual tense logical principles, like  $\mathcal{F}A \rightarrow \mathcal{F}\mathcal{P}A$  as well as event analogs of some of these, like  $\mathcal{P}rosE \rightarrow \mathcal{F}\mathcal{P}erfE$ . Some tense theorems lack event analogs. For example,  $\mathcal{F}\mathcal{P}A \rightarrow (\mathcal{F}A \vee \mathcal{P}A \vee A)$  is valid when time does not branch towards the past, but  $\mathcal{F}\mathcal{P}erfE \rightarrow \mathcal{P}erfE \vee \mathcal{P}rosE$  is not (because  $E$  may occur only at intervals containing the present moment).

We may add to this logic another *e-f* operator  $\mathcal{P}rog$  such that  $\mathcal{P}rogE$  is true at  $t$  iff  $t$  belongs to an interval at which  $E$  occurs. Thus  $\mathcal{P}rogE$  asserts that event  $E$  is in progress. In view of the discussion in Section 3.4 above, it should be obvious that  $\mathcal{P}rog$  is a poor representation of the English progressive. It can perhaps be viewed as an idealization of that construction which comes as close to its meaning as is possible with a purely temporal truth condition. (Analogous justifications are sometimes given for claims that the material conditional represents the English ‘if...then’ construction.) The new connective allows us to express the principle that eluded us above:  $\mathcal{F}\mathcal{P}erfE \rightarrow \mathcal{P}erfE \vee \mathcal{P}rosE \vee \mathcal{P}rogE$ .

Since Zeno of Elea posed his famous paradoxes in the fifth century *BC*, accounts of events and time have been tested by a number of puzzles. One Zeno-like puzzle, discussed in Hamblin [1971; 1971a], Humberstone, and Galton [Humberstone, 1979; Galton, 1984; Galton, 1987a], is expressed by the following question. ‘At the instant a car starts to move, is it moving or at rest?’ To choose one alternative would seem to distort the meaning of *starting to move*, to choose both or neither would seem to violate the laws of non-contradiction or excluded middle. Such considerations lead Galton to a slightly more complicated interpretation for event logic. Events are not assigned sets of occurrence intervals, but rather sets of interval pairs  $(B, A)$ , where  $B$  and  $A$  represent the times before the event and the times after the event (so that, if time is linear,  $B$  and  $A$  are disjoint initial and final segments of the set of times.) The clauses in the truth definition are modified appropriately. For example,  $\mathcal{P}rogE$  is true at  $t$  if, for some  $(B, A) \in I(E)$ ,  $t \in -B \cap -A$ , where  $-A$  and  $-B$  are those times of the model that do not belong to  $A$  and  $B$ .  $\mathcal{P}erfE$  is true at  $t$  if, for some  $(B, A) \in I(E)$ ,  $t \in B$ . For an event like the car’s starting to move, any  $(B, A)$  in the occurrence set will be exhaustive, i.e.,  $B \cup A$  will contain all times. Such events are said to be *punctual* (although we must distinguish these from events that occupy a ‘point’ in the sense that  $B \cup A$  always omits a single time). A punctual event does not really occur ‘at’ a time, nor is it ever in the process of occurring. Instead, it marks a boundary between two states, like the states of rest and motion. When  $E$  is punctual,  $\mathcal{P}rogE$  is always false, and so the principle  $\mathcal{F}\mathcal{P}erfE \rightarrow \mathcal{P}erfE \vee \mathcal{P}rosE \vee \mathcal{P}rogE$

reduces to  $\mathcal{F}\mathcal{P}erfE \rightarrow \mathcal{P}erfE \vee \mathcal{P}rosE$ , which we have observed not to be valid without the stipulation that  $E$  is punctual.

We may also wish to add  $f$ - $e$  aspect operators that apply to formulas to form event-expressions. Galton suggests the ‘ingressive’ operator  $\mathcal{I}ngr$  and the ‘pofective’ operator  $\mathcal{P}o$ , where, for any formula  $A$ ,  $\mathcal{I}ngrA$  is the event of  $A$ ’s *beginning* to be true, and  $\mathcal{P}oA$  is the event (or state) of  $A$ ’s being *true for a time*. In the ‘before-after’ semantics, these operators can be interpreted by the clauses below:

$$I(\mathcal{I}ngrA) = \{(B, -B) : A \text{ is true throughout a non-empty initial segment of } -B, \text{ and false throughout a nonempty final segment of } B\}$$

$$I(\mathcal{P}oA) = \{(B, C) : -B \cap -C \text{ is not empty, } A \text{ is true throughout } -B \cap -C \text{ and } A \text{ is false at some point in every interval that properly contains } -B \cap -C\}$$

Thus,  $\mathcal{I}ngrA$  is always punctual, and  $\mathcal{P}oA$  is never punctual. Notice that  $-B \cap -C$  can be a singleton, so that being true for ‘a time,’ on this interpretation, includes being true for an instant. We get principles like  $\mathcal{P}ros \mathcal{I}ngrA \rightarrow (\neg A \wedge \mathcal{F}A) \wedge \mathcal{F}(\neg A \wedge \mathcal{F}A)$ ,  $\mathcal{P}erf \mathcal{I}ngr \mathcal{P}erfE \rightarrow \mathcal{P}erfE$ , and  $\mathcal{P}rog \mathcal{P}oA \rightarrow \mathcal{P}\neg A \wedge A \wedge \mathcal{F}\neg A$ . It is instructive to consider the converses of these principles. If  $A$  is true and false at everywhere dense subsets of the times (for example if time is the reals and  $A$  is false at all rationals and true at all irrationals), then at the times  $A$  is false  $\neg A \wedge \mathcal{F}A$  is true, but  $\mathcal{I}ngrA$  has no occurrence pairs, and so  $\mathcal{P}ros \mathcal{I}ngrA$  is false. Thus the converse of the first principle fails. Likewise, if  $E$  occurs repeatedly throughout the past (for example, if time is the reals and  $I(E) = \{(-\infty, n][n + 1, \infty)\}$ ) then  $\mathcal{P}erfE$  is true at all times, which implies that  $\mathcal{I}ngr \mathcal{P}erfE$  has an empty occurrence set,  $\mathcal{P}erf \mathcal{I}ngr \mathcal{P}erfE$  is everywhere false, and the converse to the second principle fails. The converse to the third principle is valid, for if  $\mathcal{P}\neg A \wedge A \wedge \mathcal{F}\neg A$  is true at  $t$ , then, letting  $\cap S$  be the intersection of all intervals  $S$  such that  $t \in S$  and  $A$  is true throughout  $S$ , the occurrence set of  $\mathcal{P}oA$  includes the pair  $(\{x : \forall y \in \cap S, x < y\}, \{x : \forall y \in \cap S, x > y\})$  and  $\mathcal{P}rog \mathcal{P}oA$  is true at  $t$ . (The principle would fail, however, if we took  $\mathcal{P}oA$  to require that  $A$  be true throughout an *extended* period.) As a final exercise in Galtonian event logic, we observe that it provides a relatively straightforward expression of Dedekind continuity (see Burgess [2001]). The formula  $\mathcal{P}erf \mathcal{I}ngr \mathcal{P}erfE \rightarrow \mathcal{P}(\mathcal{P}erfE \wedge \neg \mathcal{P} \mathcal{P}erfE) \vee \mathcal{P}(\neg \mathcal{P}erfE \wedge \neg \mathcal{F} \neg \mathcal{P}erfE)$  states that, if there was a cut between times at which  $\mathcal{P}erfE$  was false and times at which it was true, then either there was a first time when it was true or a last time when it was false. It corresponds to Dedekind continuity in the sense that a dense frame verifies the formula if and only if the frame is Dedekind continuous.

The view represented by the ‘before-after’ semantics suggests that events of the form *IngrA* and other punctual events are never in the process of occurring, but somehow occur ‘between’ times. However plausible as a metaphysical theory, this idea seems not to be reflected in ordinary language. We sometimes accept as true sentences like *the car is starting to move*, which would seem to be of the form *ProgIngrA*. To accommodate these ordinary-language intuitions, we might wish to revert to the simpler occurrence-set semantics. *IngrA* can be assigned short intervals, each consisting of an initial segment during which *A* is false and a final segment at which *A* is true. On this view, *IngrA* exhibits *vagueness*. In a particular context, the length of the interval (or a range of permissible lengths) is understood. When the driver engages the gear as the car starts to move he invokes one standard, when the engineer starts the timer as the car starts to move she invokes a stricter one. As in Galton’s account, the *Zeno*-like puzzle is dissolved by denying that there is an instant at which the car starts to move. The modified account concedes, however, that there are instants at which the car is starting to move while moving and other instants at which it is starting to move while not moving.

Leaving aside particular issues like the semantics of punctual events and the distinction between event-letters and sentence-letters, Galton’s framework suggests general tense-logical questions. The *f–e* aspect operators, like *Ingr* and *Po* can be viewed as operations transforming instant-evaluated expressions into interval-evaluated (or interval-occupying?) expressions, and the *e–f* aspect operators, like *Perf* and *Prog*, as operations of the opposite kind. We might say that traditional tense logic has investigated general questions about instant/instant operations and that interval tense logic has investigated general questions about operations taking intervals (or pairs of intervals) to intervals. A general logic of aspect would investigate questions about operations between instants and intervals. Which such operations can be defined with particular metalinguistic resources? Is there anything logically special about those (or the set of all those) that approximate aspects of natural language? The logic of events and aspect would seem to be a fertile ground for further investigation.

## ACKNOWLEDGEMENTS

A portion of this paper was written while Portner was supported by a Georgetown University Graduate School Academic Research Grant. Helpful comments on an earlier draft were provided by Antony Galton. Some material is taken from Kuhn [1986] (in the earlier edition of this *Handbook*), which benefitted from the help of Rainer Bäuerle, Franz Guentner, and Frank Vlach, and the financial assistance of the Alexander von Humboldt foundation.

Steven Kuhn

*Department of Philosophy, Georgetown University*

Paul Portner

*Department of Linguistics, Georgetown University*

## BIBLIOGRAPHY

- [Abusch, 1988] D. Abusch. Sequence of tense, intensionality, and scope. In *Proceedings of the Seventh West Coast Conference on Formal Linguistics*, Stanford: CSLI, pp. 1–14, 1988.
- [Abusch, 1991] D. Abusch. The present under past as de re interpretation. In D. Bates, editor, *The Proceedings of the Tenth West Coast Conference on Formal Linguistics*, pp. 1–12, 1991.
- [Abusch, 1995] D. Abusch. Sequence of tense and temporal de re. To appear in *Linguistics and Philosophy*, 1995.
- [Åqvist, 1976] L. Åqvist. Formal semantics for verb tenses as analyzed by Reichenbach. In van Dijk, editor, *Pragmatics of Language and Literature*, North Holland, Amsterdam, pp. 229–236, 1976.
- [Åqvist and Guentner, 1977] L. Åqvist and F. Guentner. In L. Åqvist and F. Guentner, editors, *Tense Logic*, Nauwelaerts, Louvain, 1977.
- [Åqvist and Guentner, 1978] L. Åqvist and F. Guentner. Fundamentals of a theory of verb aspect and events within the setting of an improved tense logic. In F. Guentner and C. Rohrer, editors, *Studies in Formal Semantics*, North Holland, pp. 167–199, 1978.
- [Åqvist, 1979] L. Åqvist. A conjectured axiomatization of two-dimensional Reichenbachian tense logic. *Journal of Philosophical Logic*, 8:1–45, 1979.
- [Baker, 1989] C. L. Baker. *English Syntax*. Cambridge, MA: MIT Press, 1989.
- [Bach, 1981] E. Bach. On time, tense and events: an essay in English metaphysics. In Cole, editor, *Radical Pragmatics*, Academic Press, New York, pp. 63–81, 1981.
- [Bach, 1983] E. Bach. A chapter of English metaphysics, manuscript. University of Massachusetts at Amherst, 1983.
- [Bach, 1986] E. Bach. The algebra of events. In Dowty [1986, pp. 5–16], 1986.
- [Bäuerle, 1979] R. Bäuerle. Tense logics and natural language. *Synthese*, 40:226–230, 1979.
- [Bäuerle, 1979a] R. Bäuerle. *Temporale Deixis, Temporale Frage*. Gunter Narr Verlag, Tübingen, 1979.
- [Bäuerle and von Stechow, 1980] R. Bäuerle and A. von Stechow. Finite and non-finite temporal constructions in German. In Rohrer [1980, pp.375–421], 1980.
- [Barwise and Perry, 1983] J. Barwise and J. Perry. *Situations and Attitudes*. Cambridge MA: MIT Press, 1983.
- [Bennett, 1977] M. Bennett. A guide to the logic of tense and aspect in English. *Logique et Analyse* 20:137–163, 1977.

- [Bennett, 1981] M. Bennett. Of Tense and aspect: One analysis. In Tedeschi and Zaenen, editors, pp. 13–30, 1981.
- [Bennett and Partee, 1972] M. Bennett and B. Partee. Toward the logic of tense and aspect in English. Systems Development Corporation Santa Monica, California, reprinted by Indiana University Linguistics Club, Bloomington, 1972.
- [Binnick, 1991] R.I. Binnick. *Time and the Verb*. New York and Oxford: Oxford University Press, 1991.
- [Blackburn, 1992] P. Blackburn. Nominal Tense Logic. *Notre Dame Journal of Formal Logic* 34:56–83, 1992.
- [Blackburn, 1994] P. Blackburn. Tense, Temporal Reference, and Tense Logic. *Journal of Semantics* 11:83–101, 1994.
- [Brown, 1965] G. Brown. *The Grammar of English Grammars*. William Wood & Co., New York, 1965.
- [Bull and Segerberg, 2001] R. Bull and K. Segerberg. Basic modal logic. In *Handbook of Philosophical Logic*, Volume 3, pp. 1–82. Kluwer Academic Publishers, Dordrecht, 2001.
- [Bull, 1968] W. Bull. *Time, Tense, and The Verb*. University of California-Berkeley, 1968.
- [Burgess, 1982] J. Burgess. Axioms for tense logic I. "Since" and "until". *Notre Dame Journal of Formal Logic*, pp. 367–374, 1982.
- [Burgess, 1982a] J. Burgess. Axioms for tense logic II. Time periods. *Notre Dame Journal of Formal Logic* 23:375–383, 1982.
- [Burgess, 1984] J. Burgess. Beyond tense logic. *Journal of Philosophical Logic* 13:235–248, 1984.
- [Burgess, 2001] J. Burgess. Basic tense logic. In *Handbook of Philosophical Logic*, this volume, 2001.
- [Cooper, 1986] R. Cooper. Tense and discourse location in situation semantics. In Dowty [Dowty, 1986, pp. 5–16], 1986.
- [Carlson, 1977] G.N. Carlson. A unified analysis of the English bare plural. *Linguistics and Philosophy* 1:413–58, 1977.
- [Cresswell, 1990] M.J. Cresswell. *Entities and Indicatives*. Kluwer, Dordrecht, 1990.
- [Cresswell, 1996] M.J. Cresswell. *Semantic Indexicality*. Kluwer, Dordrecht, 1996.
- [Cresswell and von Stechow, 1982] M.J. Cresswell and A. von Stechow. De re belief generalized. *Linguistics and Philosophy* 5:503–35, 1982.
- [Curme, 1935] G. Curme. *A Grammar of the English Language* (vols III and II). D.C. Heath and Company, Boston, 1935.
- [Dowty, 1977] D. Dowty. Toward a semantic analysis of verb aspect and the English imperfective progressive. *Linguistics and Philosophy* 1:45–77, 1977.
- [Dowty, 1979] D. Dowty. *Word Meaning and Montague Grammar: The Semantics of Verbs and Times in Generative Grammar and Montague's PTQ*. D. Reidel, Dordrecht, 1979.
- [Dowty et al., 1981] D. Dowty et al. *Introduction to Montague Semantics*. Kluwer, Boston, 1981.
- [Dowty, 1986] D. Dowty, editor. Tense and Aspect in Discourse. *Linguistics and Philosophy* 9:1–116, 1986.
- [Dowty, 1986a] D. Dowty. The Effects of Aspectual Class on the Temporal Structure of Discourse: Semantics or Pragmatics? In Dowty [Dowty, 1986, pp. 37–61], 1986.
- [Ejerhed, ] B.E. Ejerhed. *The Syntax and Semantics of English Tense Markers*. Monographs from the Institute of Linguistics, University of Stockholm, No. 1.
- [Enç, 1986] M. Enç. Towards a referential analysis of temporal expressions. *Linguistics and Philosophy* 9:405–426, 1986.
- [Enç, 1987] M. Enç. Anchoring conditions for tense. *Linguistic Inquiry* 18:633–657, 1987.
- [Gabbay, 1974] D. Gabbay. Tense Logics and the Tenses of English. In Moravcsik, editor, *Logic and Philosophy for Linguists: A Book of Readings*. Mouton, The Hague, reprinted in Gabbay [Gabbay, 1976a, Chapter 12], 1974.
- [Gabbay, 1976] D. Gabbay. Two dimensional propositional tense logics. In Kasher, editor, *Language in Focus: Foundation, Method and Systems—Essays in Memory of Yehoshua Bar-Hillel*. D. Reidel, Dordrecht, pp. 569–583, 1976.



- [Gabbay, 1976a] D. Gabbay. *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. D. Reidel, Dordrecht, 1976.
- [Gabbay and Moravcsik, 1980] D. Gabbay and J. Moravcsik. Verbs, events and the flow of time. In Rohrer [1980], 1980.
- [Finger et al., 2001] M. Finger, D. Gabbay and M. Reynolds. Advanced tense logic. In *Handbook of Philosophical Logic*, this volume, 2001.
- [Galton, 1984] A.P. Galton. *The Logic of Aspect: An Axiomatic Approach*. Clarendon Press, Oxford, 1984.
- [Galton, 1987] A.P. Galton. Temporal Logics and Their Applications, A.P. Galton, editor, Academic Press, London, 1987.
- [Galton, 1987a] A.P. Galton, The logic of occurrence. In Galton [Galton, 1987], 1987.
- [Groenendijk and Stokhof, 1990] J. Groenendijk and M. Stokhof. Dynamic Montague Grammar, In L. Kalman and L. Polos, editors, *Papers from the Second Symposium on Logic and Language*, Budapest, Akademiai Kiado, 1990.
- [Guenther, 1980] F. Guenther, *Remarks on the present perfect in English*. In Rohrer [1980].
- [Halpern, 1986] J. Halpern and V. Shoham. A propositional modal logic of time intervals. In *Proceedings of IEEE Symposium on Logic in Computer Science*. Computer Society Press, Washington, D.C., 1986.
- [Hamblin, 1971] C.L. Hamblin. Instant and intervals. *Studia Generale* 24:127–134, 1971.
- [Hamblin, 1971a] C.L. Hamblin. Starting and stopping. In Freeman and Sellars, editors, *Basic Issues in the Philosophy of Time*, Open Court, LaSalle, Illinois, 1971.
- [Heim, 1982] I. Heim. The Semantics of Definite and Indefinite Noun Phrases. University of Massachusetts Ph.D. dissertation, 1982.
- [Hinrichs, 1981] E. Hinrichs. *Temporale Anaphora im Englischen, Zulassungarbeit*. University of Tbingen, 1981.
- [Hinrichs, 1986] E. Hinrichs. Temporal anaaphora in discourses of English. In Dowty [1986, pp. 63–82], 1986.
- [Humberstone, 1979] I.L. Humberstone. Interval Semantics for Tense Logic, Some Remarks. *Journal of Philosophical Logic* 8:171–196, 1979.
- [Jackendoff, 1972] R. Jackendoff. *Semantic Interpretation in Generative Grammar*. Cambridge, Ma.: MIT, 1972.
- [Jespersen, 1924] O. Jespersen. *The Philosophy of Grammar*. Allen & Unwin, London, 1924.
- [Jespersen, 1933] O. Jespersen. *Essentials of English Grammar*. Allen & Unwin, London, 1933.
- [Jespersen, 1949] O. Jespersen. *A Modern English Grammar Based on Historical Principles*, 7 vols. Allen & Unwin, London, 1949.
- [Jones and Sergot, 1996] A.J.I. Jones and J. Sergot (eds): *Papers in Deontic Logic, Studia Logica* 56 number 1, 1996.
- [Kamp, 1971] H. Kamp. Formal properties of ‘now’. *Theoria* 37:237–273, 1971.
- [Kamp, 1980] H. Kamp. A theory of truth and semantic representation. In Groenendijk et al., editor, *Formal Methods in the Study of Language Part I*, Mathematisch Centrum, Amsterdam, pp. 277–321, 1980.
- [Kamp, 1983] H. Kamp. Discourse representation and temporal reference, manuscript, 1983.
- [Kamp and Rohrer, 1983] H. Kamp and C. Rohrer. Tense in Texts. In R. Buerle et al., editors, *Meaning, Use, and Interpretation of Language, de Gruyter* Berlin, pp. 250–269, 1983.
- [Klein, 1992] W. Klein. The present-perfect puzzle. *Language* 68:525–52, 1992.
- [Klein, 1994] W. Klein. *Time in Language*. Routledge, London, 1994.
- [Kruisinga, 1932] E. Kruisinga. *A Handbook of Present Day English*. 4.; P. Noordhoff, Groningen, 1931.
- [Kuhn, 1979] S. Kuhn. The pragmatics of tense. *Synthese* 40:237–263, 1979.
- [Kuhn, 1983] S. Kuhn. Where does tense belong?’, manuscript, Georgetown University, Washington D.C., 1983.

- [Kuhn, 1986] S. Kuhn. Tense and time. In Gabbay and Guenther, editors, *Handbook of Philosophical Logic*, first edition, volume 4, pp. 513–551, D. Reidel, 1986.
- [Landman, 1992] F. Landman. The progressive. *Natural Language Semantics* 1:1–32, 1992.
- [Levinson, 1983] S.R. Levinson. *Pragmatics*. Cambridge: Cambridge University Press, 1983.
- [Lewis, 1975] D.K. Lewis. Adverbs of quantification. In E. Keenan, editor, *Formal Semantics of Natural Language*, Cambridge: Cambridge University Press, pp. 3–15, 1975.
- [Lewis, 1979] D.K. Lewis. Attitudes de dicto and de se. *The Philosophical Review* 88:513–543, 1979.
- [Lewis, 1979a] D.K. Lewis. Scorekeeping in a language game. *J. Philosophical Logic* 8:339–359, 1979.
- [McCawley, 1971] J. McCawley. Tense and time reference in English. In Fillmore and Langendoen, editors, *Studies in Linguistic Semantics*, Holt Rinehart and Winston, New York, pp. 96–113, 1971.
- [McCoard, 1978] R.W. McCoard. *The English Perfect: Tense-Choice and Pragmatic Inferences*, Amsterdam: North-Holland, 1978.
- [Massey, 1969] G. Massey. Tense logic! Why bother? *Noûs* 3:17–32, 1969.
- [Michaelis, 1994] L.A. Michaelis. The ambiguity of the English present perfect. *Journal of Linguistics* 30:111–157, 1994.
- [Mittwoch, 1988] A. Mittwoch. Aspects of English aspect: on the interaction of perfect, progressive, and durational phrases. *Linguistics and Philosophy* 11:203–254, 1988.
- [Montague, 1970] R. Montague. English as a formal language. In Visentini et al., editors, *Linguaggi nella Società e nella Tecnica*, Milan, 1970. Reprinted in Montague [Montague, 1974].
- [Montague, 1970a] R. Montague. Universal grammar. *Theoria* 36:373–398, 1970. Reprinted in Montague [Montague, 1974].
- [Montague, 1973] R. Montague. The proper treatment of quantification in ordinary English. In Hintikka et al., editors, *Approaches to Natural Language*, D. Reidel, Dordrecht, 1973. Reprinted in Montague [Montague, 1974].
- [Montague, 1974] R. Montague. *Formal Philosophy, Selected Papers of Richard Montague*, R. H. Thomason, editor, Yale University Press, New Haven, 1974.
- [Needham, 1975] P. Needham. *Temporal Perspective: A Logical Analysis of Temporal Reference in English*, *Philosophical Studies* 25, University of Uppsala, 1975.
- [Nerbonne, 1986] J. Nerbonne. Reference time and time in narration. In Dowty [Dowty, 1986, pp. 83–96], 1986.
- [Nishimura, 1980] H. Nishimura. Interval logics with applications to study of tense and aspect in English. *Publications of the Research Institute for Mathematical Sciences*, Kyoto University, 16:417–459, 1980.
- [Ogihara, 1989] T. Ogihara. Temporal Reference In English and Japanese, PhD dissertation, University of Texas at Austin, 1989.
- [Ogihara, 1995] T. Ogihara. Double-access sentences and reference to states. *Natural Language Semantics* 3:177–210, 1995.
- [Parsons, 1980] T. Parsons. Modifiers and quantifiers in natural language. *Canadian Journal of Philosophy* 6:29–60, 1980.
- [Parsons, 1985] T. Parsons. Underlying events in the logical analysis of English. In E. LePore and B.P. McLaughlin, editors, *Actions and Events, Perspectives on the Philosophy of Donald Davidson*, Blackwell, Oxford, pp. 235–267, 1985.
- [Parsons, 1990] T. Parsons. *Events in the Semantics of English*. Cambridge, MA: MIT Press, 1990.
- [Parsons, 1995] T. Parsons. Thematic relations and arguments. *Linguistic Inquiry* 26:(4)635–662, 1995.
- [Partee, 1973] B. Partee. Some structural analogies between tenses and pronouns in English. *The Journal of Philosophy* 18:601–610, 1973.
- [Partee, 1984] B. Partee. Nominal and Temporal Anaphora. *Linguistics and Philosophy* 7:243–286, 1984.

- [Poutsma, 1916] H. Poutsma. *A Grammar of Late Modern English*. 5 vols, P. Noordhoff, Groningen, 1916.
- [Prior, 1957] A. Prior. *Time and Modality*. Oxford University Press, Oxford, 1957.
- [Prior, 1967] A. Prior. *Past, Present, and Future*. Oxford University Press, Oxford, 1967.
- [Prior, 1968] A. Prior. *Papers on Time and Tense*. Oxford University Press, Oxford, 1968.
- [Prior, 1968a] A. Prior. 'Now'. *Nous* 2:101–119, 1968.
- [Prior, 1968b] A. Prior. 'Now' corrected and condensed. *Nous* 2:411–412, 1968.
- [Quine, 1956] W.V.O. Quine. Quantifiers and propositional attitudes. *Journal of Philosophy* 53:177–187, 1956.
- [Quine, 1982] W.V.O. Quine. *Methods of Logic*. 4th edition, Harvard University Press, Cambridge, Mass, 1982.
- [Reichenbach, 1947] H. Reichenbach. *Elements of Symbolic Logic*. MacMillan, New York, 1947.
- [Richards, 1982] B. Richards. Tense, aspect, and time adverbials. *Linguistics and Philosophy* 5:59–107, 1982.
- [Rohrer, 1980] C. Rohrer, editor. Time, Tense, and Quantifiers. *Proceedings of the Stuttgart Conference on the logic of Tense and Quantification*, Max Niemeyer Verlag, Tuebingen, 1980.
- [Rooth, 1985] M. Rooth. Association with Focus, University of Massachusetts PhD dissertation, 1985.
- [Rooth, 1992] M. Rooth. A theory of focus interpretation. *Natural Language Semantics* 1:75–116, 1992.
- [Röper, 1980] P. Röper. Intervals and Tenses. *Journal of Philosophical Logic* 9:451–469, 1980.
- [Russell, 1914] B. Russell. *Our Knowledge of the External World as a Field for Scientific Method in Philosophy*. Chicago: Open Court, 1914.
- [Saarinen, 1978] E. Saarinen. Backwards-looking operators in tense logic and in natural language. In Hintikka, Niiniluoto, and Saarinen, editors, *Essays in Mathematical and Philosophical Logic*, Dordrecht: Kluwer, 1978.
- [Scott, 1970] D. Scott. Advice on modal logic. In K. Lambert, editor, *Philosophical Problems in Logic*, D. Reidel, Dordrecht, pp. 143–174, 1970.
- [Shoham, 1988] Y. Shoham. *Reasoning about Change: Time and Causation from the Standpoint of Artificial Intelligence*, MIT, Cambridge, Mass, 1988.
- [Smith, 1986] C. Smith. A speaker-based approach to aspect. In Dowty [1986, pp. 97–115], 1986.
- [Stump, 1985] G. Stump. *The Semantic Variability of Absolute Constructions*. Dordrecht: Kluwer, 1985.
- [Sweet, 1898] H. Sweet. A New English Grammar. *Logical and Historical* 2:, Oxford University Press, Oxford, 1898.
- [Tedeschi and Zaenen, 1981] P. Tedeschi and A. Zaenen, editors. *Tense and Aspect (Syntax and Semantics 14)*. Academic Press, New York, 1981.
- [Thomason, 1997] R. H. Thomason. Combinations of tense and modality. In this Handbook, Vol. 3, 1997.
- [Thomason, 1984] S.K. Thomason. On constructing instants from events. *Journal of Philosophical Logic* 13:85–86, 1984.
- [Thomason, 1989] S.K. Thomason. Free Construction of Time from Events. *Journal of Philosophical Logic* 18:43–67, 1989.
- [Tichý, 1980] P. Tichý. The logic of temporal discourse. *Linguistics and Philosophy* 3:343–369, 1980.
- [van Benthem, 1977] J. van Benthem. Tense logic and standard logic. In Aqvist and Guentner [Åqvist and Guentner, 1977], 1977.
- [van Benthem, 1991] J. van Benthem. *The Logic of Time: A Model-Theoretic Investigation into the Varieties of Temporal Ontology and Temporal Discourse*. 2nd edition, Kluwer, Dordrecht, 1991.
- [Venema, 1988] Y. Venema. Expressiveness and completeness of an interval tense logic. *Notre Dame Journal of Formal Logic* 31:529–547, 1988.

- [Vlach, 1973] F. Vlach. *Now and Then: A Formal Study in the Logic of Tense Anaphora*, Ph.D. dissertation, UCLA, 1973.
- [Vlach, 1979] F. Vlach. *The semantics of tense and aspect*. Ms., University of New South Wales, 1979.
- [Vlach, 1980] F. Vlach. *The semantics of tense and aspect in English*. Ms., University of New South Wales, 1980.
- [Vlach, 1981] F. Vlach. *The semantics of the progressive*. In Tedeschi and Zaenen, editors, pp. 271–292, 1981.
- [Vlach, 1993] F. Vlach. *Temporal adverbs, tenses and the perfect*. *Linguistics and Philosophy* 16:231–283, 1993.
- [Ward, 1667] W. Ward. *An Account of the Principles of Grammar as Applied to the English Language*, 1767.

# INDEX

- Åqvist, L., 330
- a posteriori proposition, 242
- a priori, 242
- abbreviations, 45
- Abusch, D., 314
- antepreperfect, 279
- antisymmetry, 2
- Aristotelian essentialism, 235, 236
- Aristotle, 6, 208
- aspect, 284
- atomic states of affairs, 237
- augmented frame, 39
- automata, 169
- axiomatization, 163
- Bäuerle, R., 300
- Bach, E., 297
- backspace operator, 332
- Baker, C. L., 315
- Barwise, J., 294
- Bennett, M., 291
- Bentham, J. F. A. K. van, 2, 6, 34
- Beth's Definability Theorem, 249
- Blackburn, P., 331
- Bull, R. A., 13, 26
- Burgess, J., 323
- Büchi, , 26
- C. S. Peirce, 37
- canonical notation, 1
- Carlson, G. N., 309
- Carnap–Barcan formula, 245
- causal tense operators, 270
- chronicle, 10
- Cocchiarella, N., 13
- cognitive capacities, 258
- combining temporal logics, 82
- comparability, 2
- complete, 52
- completeness, 2, 54, 110, 163
- complexity, 57
- concepts, 258
- conceptualism, 262, 268
- conditional logic, 228
- consequence, 45
- consistent, 52
- contingent identity, 256
- continuity, 19
- continuous, 20
- Craig's Interpolation Lemma, 249
- Creswell, M., 316
- dating sentences, 331
- De Re Elimination Theorem, 239, 255
- decidability, 56, 88, 168
- Dedekind completeness axioms, 60
- de dicto* modalities, 235, 236, 239, 245, 255
- dense time, 48
- density, 2, 16
- deontic tense logic, 226
- de re* modalities, 235, 239, 245, 255
- determinacy (of tenses), 289
- Diodorean and Aristotelian modal fragments of a tense logic, 37
- Diodoros Kronos, 6
- discourse representation theory, DRT, 288
- discrete, 19
- discrete orders, 38
- discreteness, 17
- Dummett, M., 38

- dynamic logic, 6
- dynamic Montague grammar, 293
  
- Edelberg inference, 230
- Eng, M., 292, 314
- essentialism, 235, 239, 250
- event point, 283
- events, 303
- existence, 245, 246
- expanded tense, 280
- expressive completeness, 75, 165
- expressive power, 65
  
- file change semantics, 293
- filtrations, 23
- finite model property, 23, 56, 169
- first-order monadic formula , 165
- first-order monadic logic of order, 45, 57
- fixed point languages, 165
- frame, 4
- free logic, 245, 254
- full second-order logic of one successor function  $S1S$ , 165
- full second-order monadic logic of linear order, 46
- future choice function, 231
- future contingents, 6
  
- Gabbay, D. M., 8, 26, 30, 31, 218
- Galton, A., 337
- Goldblatt, R., 38
- greatest lower bound, 23
- Guenther, F., 313
- Gurevich, Y., 30
  
- H-dimension, 78
- Halpern, J. Y., 323
- Hamblin, C. L., 17, 338
- Heim, I., 290
- Henkin, L., 8
- hilbert system, 50
- Hinrichs, E., 288
- historical necessity, 206
  
- homogeneous, 19
- Humberstone, L., 325
  
- imperative view, 119
- independent combination, 83, 88
- individual concepts, 235, 250, 253, 256
- instants, 1
- intensional entitites, 250
- intensional logic, 285
- intensional validity, 251
- intensionality, 250
- interval semantics, 292
- IRR rule, 53
- IRR theories, 55
- irreflexive models, 221
  
- Jespersen, O., 277
  
- Kamp frame, 218
- Kamp validity, 217
- Kamp, H., 27, 29, 30, 33, 43, 187, 288
- Kessler, G., 38
- killing lemma, 12
- Klein, W., 299
- Kripke, S., 13
- Kuhn, S., 34, 291
  
- labelled deductive systems, 83, 94
- Landman, F., 306
- lattices, 22
- LDS, 129
- least upperbound, 23
- Lemmon, E. J., 8
- Lewis, C. I., 37
- Lewis, D. K., 289
- Lindenbaum's lemma, 9
- Lindenbaum, A., 9
- linear frames, 44
- linearity axiom, 50
- logical atomism, 235–241
- logical necessity, 235–238, 240, 241
- logical space, 237, 238

- Lukasiewicz, J., 37  
 maximal consistent, 7  
 maximal consistent set (MCS), 52  
 McCawley, J., 311  
 McCoard, R. W., 310  
 metaphysical necessity, 240  
 metric tense logic, 36  
 Michaelis, L. A., 311  
 minimality of the independent combination, 92  
 minimal tense logic, 7  
 Minkowski frame, 38  
 mirror image, 4, 45  
 Mittwoch, A., 310, 311  
 modal logic, 6, 285  
 modal thesis of anti-essentialism, 235, 236, 238, 239  
 monadic, 25  
 Montague, R., 277  
 mosaics, 64  
  
 natural numbers, 43, 62, 161  
 neutral frames, 219  
 Nietzsche, 39  
 nominalism, 248  
 now, 30  
  
 Ockhamist  
   assignment, 214  
   logic, 215  
   model, 212  
   valid, 215, 223  
 Ogihara, T., 314  
 one-dimensional connectives, 78  
 ought kinematics, 225  
  
 Parsons, T., 298  
 Partee, B., 288, 291  
 partial orders, 13  
 past, 279  
 past tense, 298  
 Perry, J., 294  
 persistence, 156  
  
 Peter Auriolo, 6  
 Platonic or logical essentialism, 236  
 pluperfect, 279  
 Poincaré, 39  
 possibilia, 235, 266  
 possible world, 237, 244, 245, 250  
 Pratt, V. R., 6  
 predecessors, 2  
 present tense, 295  
 preterit, 279  
 Prior, A. N., 3, 6, 37, 43, 320  
 processes, 303  
 program verification, 6  
 progressive, 280, 303  
 proposition, 251  
 PTL, 161  
 punctual, 338  
 pure past, 70  
  
 quantifiers, 40  
 Quine, W. V. O., 1  
  
 Röper, P., 326  
 Rabin, M. O., 23, 25, 26, 30, 57  
 rationals, 43, 59  
 reals, 43, 59  
 reference point, 283  
 refinement, 49  
 regimentation, 1  
 regimenting, 2  
 Reichenbach, H., 277  
 return operator, 333  
 Richards, B., 310  
 rigid designators, 242, 243, 257, 261  
 Rohrer, C., 288  
 Russell, B., 323  
  
 Saarinen, E., 330  
 satisfiable, 45  
 Scott, D., 8, 323  
 Sea Battle, 208  
 Second Law of Thermodynamics, 39

- second-order logic of two successors *S2S*, 57
- Segerberg, K., 13
- separability, 60
- separable, 70
- separation, 69, 73
- separation property, 28, 29, 71
- sequence of tense, 313
- Shelah, S., 26
- Shoham, Y., 323
- since, 26, 43, 44
- situation semantics, 292
- soundness, 51
- special theory of relativity, 38, 269, 271, 272
- specification, 49
- statives, 296
- Stavi connectives, 64
- Stavi, J., 29
- structure, 44
- substitution rule, 51
- successors, 2
- syntactically separated, 76
- system of local times, 271
  
- table, 47
- temporal generalisation, 4
- temporal Horn clauses, 122
- temporalising, 83
- temporalized logic, 83
- temporally complete, 27
- tense, 1, 3, 277
- tense logic, 285
- tense-logically true, 265
- then, 31
- thesis, 50
- Thomason, R., 323
- Thomason, S. K., 39
- Tichý, P., 301
- time, 277
- time periods, 33
- total orders, 14
- tractability, 156
- transitivity, 2
  
- treelike frames, 212, 223
- trees, 15
- truth table, 67
  
- universal, 25
- universally valid, 240
- unsaturated cognitive structures, 258, 264
- until, 26, 43, 44
- US/LT*, 43
- USF, 179
  
- valid, 45
- valuation, 4, 35
- van Benthem, J. F. A. K., 323
- variable assignment, 46
- Venema, Y., 325
- verb, 1, 3, 6
- Vlach, F., 297
- Vlach, P., 31
- von Stechow, A., 300
- Von Wright's principle of predication, 254
  
- weak completeness, 52
- well-orders, 21
- wellfoundedness, 2
- William of Ockham, 6



# Handbook of Philosophical Logic

2nd Edition

Volume 8

edited by Dov M. Gabbay and F. Guentner



## CONTENTS

Editorial Preface	vii
<b>Dov M. Gabbay</b>	
The Logic of Questions	1
<b>David Harrah</b>	
Sequent Systems for Modal Logics	61
<b>Heinrich Wansing</b>	
Deontic Logic	147
<b>Lennart Åqvist</b>	
Deontic Logic and Contrary-to-duties	265
<b>José Carmo and Andrew J. I. Jones</b>	
Index	345



## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London

Logic	IT			
	Natural language processing	Program control specification, verification, concurrency	Artificial intelligence	Logic programming
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs



<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical fragments</b>	Basic back-ground language	Program synthesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely useful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calculus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syntax	Modules. Combining languages	Logics of space and time	Combining features
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game semantics gaining ground		
<b>Object level/metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action-revision models</b>			ditto	

	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model



DAVID HARRAH

## THE LOGIC OF QUESTIONS

### 1 INTRODUCTION

#### 1.1 *Basic Notions*

Most theorists use ‘interrogative’ to refer to a type of sentence. Some theorists posit questions as distinct entities that may be *asked*, or *put*, or *expressed* by interrogatives, just as propositions may be expressed by declaratives and commands may be expressed by imperatives. Intuitively it seems that some questions may be expressed by sentences other than interrogatives, and some interrogatives can be used to do other things besides ask questions. Thus it is reasonable to say that there are two overlapping subject matters: the logic of interrogatives, and the logic of questions.

Most theorists use ‘reply’ to refer to any verbal response that can be given to a question, and use ‘answer’ to refer to a distinguished kind of reply. Many kinds of reply may be appropriate from the respondent’s point of view, but the replies that are appropriate from the questioner’s point of view, the replies that the question calls for, are the answers.

Most theorists define various types of answer. The most important distinction is between direct answers, each of which gives exactly what the question calls for, and partial answers, each of which may give some (but perhaps not all) of what the question calls for. The label ‘direct’ was introduced in Harrah [1961] because it connotes both logical sufficiency and immediacy, as in the request:

‘Please give me a direct answer?’

Just as statements and commands can be ‘good’ or ‘bad’ in various ways (valid or not, true or not, possible or not, and the like), so too with questions. The details, however, vary from theory to theory.

Most theorists agree on labels for question-types approximately as follows:

Label	Example
<i>whether</i>	‘Is two even or odd?’
<i>yes-no</i>	‘Is two a prime number?’
<i>which</i>	‘Which even numbers are prime?’
<i>what</i>	‘What is Church’s Thesis?’
<i>who</i>	‘Who is Bourbaki?’
<i>why</i>	‘Why does two divide zero?’
<i>deliberative</i>	‘What shall I do now?’

<i>disjunctive</i>	‘How long is your new proof, or do you have a shorter one?’
<i>hypothetical</i>	‘If you had a proof, how long would it be?’
<i>conditional</i>	‘If you now have a proof, how long is it?’
<i>given-that</i>	‘Given that Turing’s Conjecture is provable, is Church’s Thesis provable?’

No theorist holds that this is a complete list of question-types, and most would divide each of the types listed here into several sub-types. For discussions, see Hamblin [1967], Prior and Prior [1955], Belnap and Steel [1976], and Wiśniewski [1995].

To see how problematic these basic notions can be, and how theorists’ intuitions may differ, consider the English sentence:

‘Where is Jane or Ann?’

Harrah [1975] formalizes this as a disjunction of two interrogatives (one about Jane, one about Ann); each of these interrogatives expresses one question that has its own set of answers. Belnap and Steel [1976] formalizes this as one interrogative that expresses one disjunctive question whose answers are the answers about Jane plus the answers about Ann. Groenendijk and Stokhof [1984] construes it as one interrogative that expresses two questions; the semantics of the interrogative yields one set of answers that contains the Jane answers and the Ann answers.

There is another subject matter that is a generalization on the first two: erotetic logic. In the narrow sense, ‘erotetic’ can be paraphrased as ‘pertaining to questioning’. In this sense erotetic logic is the theory of questions, interrogatives, and the use of interrogatives. Generalizing on this notion, we may use ‘erotetic’ with the sense of ‘pertaining to calling-for-reply’. Under such an interpretation, erotetic logic is the theory of all the sentences (interrogative, imperative, declarative, or whatever) that call for reply, or that are vulnerable to replies of certain sorts, and the theory of all the entities that can thus be called for. Erotetic logic in this general sense has not yet been developed to any significant depth, and we discuss it only briefly in Sections 7.7 – 7.8. For this reason, and because most of the theories discussed in this chapter are logics of questions, this chapter is most appropriately titled ‘The Logic of Questions’.

(Helpful comments and advice concerning this chapter were supplied by Andrzej Wiśniewski.)

## 1.2 Motivations

The theories that have been developed up to now differ not only in superstructure and points of detail but also in foundation and basic conception. Many of these differences are due to differences in motivation. For this

reason we take note of motivation in our survey below. The following descriptive labels are used in our exposition:

*Empirical.* This is the motivation of linguists and psychologists, for example, who wish to describe the sentences of a natural language and describe how those sentences are understood and used by the speakers of the language.

*Platonic.* This is the motivation of some philosophers and logicians who wish to describe linguistic or semantic entities considered as mathematical objects, objects that exist in their own right, so to speak.

*Normative.* This is the motivation of some philosophers and logicians who wish to describe how one ought to ask and answer question, or how a rational person asks and answers.

*Engineering.* This is the motivation of those whose purpose is simply to construct a system that will be usable for certain practical purposes (e.g. computer-assisted information-retrieval) and satisfy certain criteria (e.g. efficiency, effectiveness).

*Metalogical.* This is the motivation of one who wishes to study how far, within a given logical system, a system of question-and-answer can be developed, or, more generally, to study what sorts of question-and answer system can be developed within a given logical system, where the given system might or might not have been intended by its creators to provide a question-and-answer system.

*Technical* and *aesthetic.* These are additional motivations felt by the theorist in the course of constructing the theory. The most important of course are the desire to facilitate construction of theory and the desire to achieve simplicity or elegance of other kinds.

Two cautions: First, the foregoing characterizations are rough and usually require some qualification when applied to particular cases. Second, in any particular case there may be more than one motivation present; we note some examples below.

### 1.3 *History and Bibliography*

Discussion of questions is at least as old as Aristotle; Kubiński [1980], pp. 118 – 119, says that Cohen [1929] “is the earliest known to me in which the logic of questions is treated by means of formal logic”. Cohen suggested identifying questions with sentential functions having free variables. Closely related ideas were considered by Carnap and Reichenbach. Gornstein [1967], p. v – 1, says that Carnap seems to be the first author who wrote a question in a formalized manner.

Current activity in the field began in the 1950s, stimulated to a large extent by Prior and Prior [1955], Stahl [1956], Hamblin [1958], and Kubiński [1960]. Since then many approaches have been suggested, and several have been developed in detail.

For a general discussion of the history of the field to about 1965, see Hamblin [1967]. Gornstein [1967] presents a detailed summary and discussion of many theories and logics from Aristotle to the 1960s. Kubiński [1980] offers many historical remarks and useful bibliography.

Probably the most comprehensive and complete bibliography published to 1976 is the one by Egli and Schleichert that is included in Belnap and Steel [1976]. It contains sections on Logic and Philosophy of Language; Linguistics; Automatic Question-Answering; and Psychology and Pedagogy. Abstracts are included for many of the items listed. Ficht [1978] presents an updated version of this bibliography with an additional section on dialogue. Other bibliographies, more limited in scope but quite useful, are found in Hiz (ed.) [1978], Lehnert [1978], Belnap [1981] and [1983], and Higginbotham [1993]. The bibliography presented in Wiśniewski [1995] is valuable in many ways and has good coverage of the important work done up to 1994. Note: The list of References given at the end of this chapter is very selective. In general it concentrates on logic and ignores psychology, pedagogy, and heuristics. In particular it concentrates on the topics and authors discussed in this chapter.

#### 1.4 *Scope of this Chapter*

The main aims of this chapter are to indicate the variety of motivations, basic conceptions, and approaches to theorizing that are now evident in the field, and to outline some topics and aspects that invite further study. This chapter does not aim at a complete historical account, or a complete catalogue of possible theories. We concentrate on some work, by a few logicians, that is especially significant and fruitful for further developments in various directions.

#### 1.5 *Abbreviations and Notation*

In most cases, when summarizing the work of another author, we use that author's terminology and notation. In a few cases we depart for the sake of clarity or simplicity. Occasionally, where there is no danger of ambiguity, we use symbols as names of themselves. Except where noted otherwise, we use

wff	for	well-formed formula
iff		if and only if
$d(I)$		direct answer to $I$ (or, in Sections 7.3 – 7.4, the set of these answers)
$D(I)$		the set of direct answers to $I$
$\neg X$		the negation of $X$
$\Lambda$		the empty set



## 2 SET-OF-ANSWERS METHODOLOGY

### 2.1 Hamblin's Postulates

In an informal paper Hamblin [1958] proposed three postulates:

1. An answer to a question is a statement.
2. Knowing what counts as an answer is equivalent to knowing the question.
3. The possible answers to a question are an exhaustive set of mutually exclusive possibilities.

Hamblin also suggested a calculus of questions to formalize such ideas as containment and equivalence. For example, one question *contains* another when from every answer to the first we can deduce an answer to the second; and the two questions are *equivalent* if they contain each other. This paper stimulated much formal work by others (see, e.g. Belnap and Steel [1976], p. 35).

Some linguists and logicians have argued against adopting Postulate (1) (Hintikka [1976], Tichý [1978]). For those who adopt it, however, it effects an important simplification. Replies that are not statements (e.g. noun phrases, nods, grunts) can be treated as *coded answers* that are abbreviations of statements. Thus the logic of answers is concerned only with statements.

Some logicians have argued against adopting Postulate (2) (Åqvist [1965], Hintikka [1976]). For those who adopt it, however, it represents another giant step toward formalization. The techniques inspired by it are perhaps best thought of under the label 'set-of-answers methodology' (or 'SA methodology' for short). Hamblin's own technique for defining containment and equivalence is one example; others appear below.

Postulate (2), and SA methodology in general, are compatible with several different theories about the logical nature of questions, and about the connection between a question and its answers. In fact, most of the approaches to theorizing surveyed in this chapter exemplify SA methodology at various points.

(Note: Postulate (3) is controversial and there is much unfinished business connected with it, but we do not discuss it in this chapter beyond the brief mentions in 6.6 and 7.2.)

### 2.2 The SA Reduction of Questions

One idea, not required by SA methodology but obviously compatible with it, is to identify a question with its set of answers. Let us call this the *SA reduction of questions*. In radical versions of SA reduction one allows

arbitrary sets of sentences to count as questions, regardless of whether these sets have defining characteristics that can be expressed in a given language. In conservative versions one considers only certain kinds of sets — e.g. sets that are definable in terms of the syntactical form of the sentences that are their members.

In several papers in the 1960s Stahl developed an SA reduction. The following is a summary of Stahl [1962].

We assume a higher-order function calculus, and then distinguish three types of questions: (1) *individual* questions (e.g.  $[Hx?]$ , read ‘Which things satisfy  $H?$ ’), (2) *function* questions (e.g.  $[F?a]$ , read ‘Which functions are satisfied by  $a?$ ’), and (3) *truth* questions (e.g.  $[Af?B]$ , read ‘Which truth-functions hold between  $A$  and  $B?$ ’). To (1), *simple* answers are  $Ha$ ,  $Hb$ , etc.; *direct* answers are simple answers that are not negations of theorems. We can form finite conjunctions  $[Ha \wedge Hb \wedge Hd]$ ,  $(\neg Ha \wedge \neg Hc)$ , etc., and also infinite conjunctions  $(x)\neg Hx$ ,  $(x)(H'x \rightarrow Hx)$ , etc. A *perfect* answer is such a conjunction which is not the negation of a theorem. A *sufficient* answer is a wff  $F$  such that  $F$  is not the negation of a theorem, and either  $F$  implies a perfect answer which is not a theorem or else  $F$  is a theorem and some perfect answer is a theorem. We now define the question  $[Hx?]$  as the class of its sufficient answers.

To (2) the simple answers are  $Ha$ ,  $H'a$ ,  $H''a$ , etc.; as before we then form direct, perfect, and sufficient answers, and the question is defined as the class of its sufficient answers. For (3) we proceed likewise, except that there are no infinite conjunctions; e.g.  $((A \vee B) \wedge (A \rightarrow B))$  is in  $[Af?B]$ .

The initial definitions can be generalized and made relative to a system  $X$  and a set of premises  $S$ ; we write  $[P]_X S$  to mean that  $P$  is a question in  $X$  relative to  $S$ . Here the sufficient answers are required to be consistent with consequences of  $S$ . For discussion and criticism of Stahl [1962], see Harrah [1963b]. See also Section 7.5 below.

### 2.3 Motivation

There are several motivations for adopting an SA reduction. Besides the metalogical one (to see what can be done within set theory) there are technical and aesthetic motivations. All the operations on sets and relations between sets are directly available as operations on questions and relations between questions, and theorems about sets become theorems about questions.

For some researchers there are also Platonic, normative, and engineering motivations. It has been argued that for a rational decision-maker in a choice situation the only essential thing about a question is its set of answers (e.g. Szaniawski [1973] and Dacey [1981]).

Another motivation (a technical one?) appears in connection with Fregean principles of language construction. If every declarative sentence is to be

assigned a sense and a denotation, uniformity would suggest a similar treatment for interrogatives. The natural technique is to let each interrogative denote the set of its direct answers. Some logicians and linguists have adopted this technique, though others have argued that an interrogative denotes just the set of its true direct answers. For an entry into the literature of this topic, see Karttunen [1978], Belnap [1981], Kiefer [1983], Groenendijk and Stokhof [1984], and Higginbotham [1993].

#### 2.4 *Some Problems*

Some logicians have argued that the SA reduction is not intuitively or empirically plausible (e.g. Tichý [1978], pp. 279 – 280). Even for those who favor it, however, it presents the problem of deciding what kind of answers are to be in the sets under consideration. It might seem that all the direct answers must be included; but, as noted above, some theorists in the tradition of Montague Grammar have equated questions with the sets of just their true direct answers.

Regardless of whether a question includes all of its direct answers, should it be allowed to include partial answers? and replies like ‘I don’t know’? Some linguists have argued for a view that in effect obliterates the distinction between direct and partial, and that would (on the SA reduction) treat a question as the set of its direct and partial answers together (e.g. Bolinger [1978], p. 104). For logicians who wish to preserve a sharp distinction between the direct and partial the simplest course (with the SA reduction) is to identify a question with the direct (or, true direct) answers, and then to define the partial answers separately. As noted above, this is the technique used in Stahl [1962].

On some motivations one who adopts SA reduction cares about questions but not about interrogatives, and specifically does not care whether there are enough interrogatives to express all questions. On other motivations one is not so indifferent, and decisions about what sets are to count as questions depend on what sentences are available as interrogatives for expressing questions.

Among many possible policies of question-definition the following are the natural ones:

1. Choose any arbitrary collection  $C$  of sets  $S$  of sentences.
2. Choose any  $C$  such that every  $S$  in  $C$  is describable in the assumed metalanguage.
3. Choose any  $C$  such that there is a set  $S'$  of sentences of  $L$  (the ‘interrogatives’) with this property: There is a many-one mapping from  $S'$  onto  $C$  (so that every question  $S$  in  $C$  is expressible by at least one interrogative in  $S'$ ).

4. As in (3), with the additional requirement that there is an effective procedure for recognizing interrogatives, and an effective procedure whereby, given an interrogative, we can recognize the question that it expresses.
5. Choose a  $C$  only if it, in effect, represents the set of questions belonging to some given natural language.

In fact, most logicians in the field thus far have followed (5) or (4). An empirical motivation of course leads to (5). A Platonic motivation might lead to any of (1)–(5), and would lead also to metatheoretic questions about completeness. The topics of effectiveness and completeness are of such importance that we discuss them in a separate section below.

### 3 EFFECTIVENESS AND COMPLETENESS

#### 3.1 Introduction

It is usually of interest to investigate whether certain aspects of a question-and-answer system are effective, and whether the system is complete in certain respects. On some motivations it is required that certain kinds of completeness and effectiveness do indeed obtain. Probably the first theorist to note the importance of these properties and to study them in a systematic way was Belnap [1963]; we outline his ideas and results in Section 4. In this section we give a more general discussion, because the topic is important for many theories of questions.

#### 3.2 The Problem of Effectiveness

Should the notion of *interrogative*, or *question-expresser*, be effective? Suppose it is not. Then, when questioner  $Q$  utters  $X$ , respondent  $R$  might have to ask ‘Are you asking a question?’ and then  $Q$  might have to ask ‘Is the latter a question?’ and so on back and forth indefinitely. This argument does not prove that all interrogatives must be recognizable as such, but it does suggest that some must be. Either all must be, or else there must be in the language some interrogatives with the force of ‘Does the expression  $X$  express a question?’, where these are recognizable as such.

A similar argument applies to answers. Suppose that  $Q$  utters  $X$ , and  $R$  recognizes that  $X$  expresses a question, but that *answer-to- $X$*  is not effective. Then, when  $R$  gives a reply,  $Q$  must ask ‘Was that an answer?’ As before, this argument does not prove that every question must have an effective set of answers, but it does suggest that some must. Either all must, or there must be some question with the force of ‘Is  $W$  an answer to  $X$ ?’ and with an effective set of answers.

On the other hand one might want to allow noneffectiveness for certain types of question and answer. There seems to be no effective method for determining whether English sentences of the form ‘I wonder whether . . .’ express questions as distinct from statements. Also it seems that many who-, what-, and why- questions do not have effective answer sets.

### 3.3 *Concepts of Expressive Completeness*

A question system might be expressively complete (or fail to be complete) in any of several senses. First, it might be empirically complete, in that it provides for all the questions that natural languages do. Second, it might be complete in a Platonic sense, in that it provides for all the questions that ‘really exist’.

The Platonic conception may be made precise via model theory without reference to a particular language. One theory for which this would be possible is that of Tichý [1978]; see below in Section 6.6. Alternatively, we may speak of ‘all the questions that really exist, relative to a given language’. This is Belnap’s conception; see Section 4. With this conception the problem of completeness is to determine whether everything that counts as a real question is expressible by an interrogative in the given language.

Further senses of completeness are generated by semantic and pragmatic considerations. For example, a system that fails to provide for all the real questions may nevertheless provide for all the questions that have a true direct answer, or all that are truly answerable by a human being.

With respect to answers, a system may be complete if, for each of its questions, all the ‘real answers’ (again, see Belnap) are expressible. A more reasonable alternative is to require merely that, for each question that has true real answers, at least one true real answer is expressible. For further discussion of concepts of completeness, see Belnap and Steel [1976] and Harrah [1969].

### 3.4 *Diagonalization and Expressive Incompleteness*

In the logic of questions generally, and in SA methodology especially, we deal with sets of linguistic expressions, or sets of sets of them. Sometimes there is enough structure in the situation to permit Cantorian diagonalization. We give one example here, from Harrah [1969].

We assume that we have a language  $L$  with denumerably many expressions, and that an effective alphabetical ordering of them has been established. We assume that there is a set  $S$  of questions, and that  $S$  is recursively enumerable. Suppose that each question has denumerably many direct answers, or can be assigned denumerably many in a harmless way (e.g. by adding instances of  $(P \wedge \neg P)$ ). Suppose next that, given a question  $q$ , the

set of direct answers to  $q$  is recursively enumerable. Finally suppose that direct answers are sentences, and that *sentence* is effective.

Now we can diagonalize to construct new sets of sentences each of which is not the set of direct answers to any question in the assumed enumeration of questions. Viz: Choose any positive integer  $j$ . Then for the alphabetically first member of the new set  $D$  we choose the alphabetically  $j$ th sentence after the first direct answer of the first question, and for the alphabetically  $(i + 1)$ th member of  $D$  we choose the alphabetically  $j$ th sentence after the first direct answer  $d$  of the  $(i + 1)$ th question such that  $d$  is alphabetically later than the  $i$ th member of  $D$ . To make  $D$  more interesting we can specify in advance some recursive property  $P$  and stipulate that the construction is to move at least  $j$  sentences and keep going until it finds a sentence with the property  $P$ .

One consequence that is relevant to many theories but to the SA reduction in particular is this: If *question* and *direct answer* are effective (or merely enumerable, as assumed above), then the system is incomplete in the sense that not every set of sentences can be the set of direct answers to a question. For further discussion, see Harrah [1969].

### 3.5 Deductive Completeness

There is another family of completeness concepts that has been suggested by work of Kubiński and Wiśniewski (see 8.5 in Wiśniewski [1995]). The basic idea is that relations like implication can hold between (1) wffs and questions and (2) questions, and the particular relationships that do hold can be described or expressed via statements of a certain form  $F$  in a metalanguage ML of the given language  $L$ . Then, in analogy with the development of systems of logic for declarative sentences, we may choose some set  $S$  of statements of ML and ask whether  $S$  is a complete axiom set for all of the true statements of the form  $F$ .

This area invites and awaits exploration.

## 4 BELNAP'S ANALYSIS

In this section we summarize a part of Belnap and Steel [1976]. That book presents (1) a formal system for question-and-answer, (2) a rich metatheory for the system, (3) discussion of application to English, (4) discussion of application to data processing and information retrieval, and (5) an extensive bibliography by U. Egli and H. Schleichert. Here we summarize just (1) and (2). Because they were the work of Belnap (completed in 1968), we shall call them Belnap's system and Belnap's theory. We present Belnap's theory in detail because many of his concepts apply to systems other than his own, and many of his concepts deserve to become standard.

#### 4.1 Motivation

The main motivation is a normative one: to construct a rational system applicable in situations of a certain kind — namely, where questioner  $Q$  and respondent  $R$  are motivated to help each other, and  $R$  has access to a well structured information source. In particular  $R$  may be a machine, and the information source may be a data bank.

For Belnap, a system is adequate for these situations only if it meets certain conditions of effectiveness. First, the interrogatives must be effectively recognizable as such. Second, given any question, its direct answers must be effectively recognizable as such. The latter is the fundamental criterion, emphasized in both Belnap [1963] and Belnap and Steel [1976].

#### 4.2 The Assertoric Basis

The language  $L$  is an applied first-order functional calculus with identity. There are denumerably many individual variables  $w, x, y, z, \dots$ , and countably many individual constants,  $n$ -ary function constants  $f, g, \dots$ , and  $n$ -ary predicate constants  $F, G, \dots$ . There are signs  $=$  for identity,  $\wedge$  for conjunction,  $\vee$  for disjunction,  $\rightarrow$  for the material conditional,  $\leftrightarrow$  for the material biconditional,  $\exists$  and  $\forall$  for the existential and universal quantifiers. *Term* and *wff* are defined as usual, except that  $n$ -ary conjunctions  $(A_1 \wedge \dots \wedge A_n)$  and disjunctions  $(A_1 \vee \dots \vee A_n)$  are permitted for each  $n$ . We use  $a, b, c, \dots$  for terms and  $A, B, \dots$  for wffs. An  *$n$ -place condition* is a wff with exactly  $n$  free variables. A *statement* is a wff with no free variables. A *name* is a term with no free variables. Given  $Ax_1 \dots x_n$ , we understand that  $Ab_1 \dots b_n$  comes from  $Ax_1 \dots x_n$  by proper substitution of  $b_i$  for  $x_i$ .

Some one-place conditions are designated as *elementary category conditions* (including  $(x = a)$  for each name  $a$ ). The set of *category conditions* is defined recursively by:

1. Every elementary category condition is a category condition.
2. If  $Ax$  and  $Bx$  are category conditions with  $x$  as the only free variable, then  $(Ax \wedge Bx)$  and  $(Ax \vee Bx)$  are category conditions, and so too is any result of changing variables (free or bound) in  $Ax$ .

With each elementary category condition  $Ax$  there is associated a decidable set of names, called the *nominal category determined by  $Ax$* . If  $Ax$  is  $(x = a)$ , the nominal category must be  $\{a\}$ . If  $Ax$  and  $Bx$  differ only in their variables (free or bound), they determine the same nominal category. If  $Ax$  is  $(Bx \wedge Cx)$  or  $(Bx \vee Cx)$ , then its nominal category is the intersection or union of the nominal categories determined by  $Bx$  and  $Cx$ .

For the semantics of  $L$ , a *candidate interpretation* consists of a nonempty domain of individuals  $D$  and an interpretation function of the usual extensional kind. *Denotation* and *truth* are defined as usual. The *range* of a

one-place condition  $Ax$  in a candidate interpretation  $M$  is the set of individuals  $i$  in the domain of  $M$  such that  $Ax$  is true in  $M'$ , where  $M'$  is like  $M$  except in assigning  $i$  to the free variable  $x$  in  $Ax$ . The range in  $M$  of a category condition is also called the *real range*, or the *real category determined by* that condition in  $M$ . Category conditions differing only in their free or bound variables are *equivalent*; we write  $\mathbf{C}x$  for the set of conditions that are equivalent to  $Cx$ .

An *interpretation* is a candidate interpretation  $M$  in which, for every category condition  $Ax$ , every name in the nominal category determined by  $Ax$  denotes in  $M$  some individual in the real category determined by  $Ax$  in  $M$ . *Consistency, validity, logical implication* and *logical equivalence* are defined as usual. Where there is no explicit reference to an interpretation, it is understood that there is implicit reference to some *principal interpretation*.

### 4.3 Elementary Questions

The elementary questions are whether-questions and which-questions. An elementary question is expressed by an *elementary interrogative*. These have the form  $?\rho\sigma$ , where  $\rho$  denotes a request,  $\sigma$  denotes a subject and  $?$  denotes the function which takes a request and a subject as arguments and produces a question as value.

An *abstract whether-subject* is a finite set of wffs. The *range determined by* this subject is the set itself; likewise, the set of *alternatives presented by* this subject is the set itself. A *lexical whether-subject* is a finite list of wffs enclosed in parentheses, as:  $(A_1, \dots, A_n)$ . To simplify matters it is required of both abstract whether-subjects  $\{A_1, \dots, A_n\}$  and lexical whether-subjects  $(A_1, \dots, A_n)$  that there be no repetitions among  $A_1, \dots, A_n$ , and that no  $A_i$  be a conjunction of other statements in the list  $A_1, \dots, A_n$ . The lexical whether-subject  $(A_1, \dots, A_n)$  *signifies* the abstract whether-subject  $\{A_1, \dots, A_n\}$ .

An *abstract which-subject* is a triple  $\langle X, g, A \rangle$  such that  $X$  is a nonempty set of variables (the *queriables*),  $g$  is a *category mapping* in  $X$  (i.e. a mapping from a subset of  $X$  into the set of equivalence classes of category conditions), and  $A$  is a matrix (a wff whose free variables include the queriables).

Let  $\langle X, g, Ax_1 \dots x_n \rangle$  be an abstract which-subject, where  $X = \{x_1, \dots, x_n\}$ , and  $g$  is a category mapping in  $X$ . Then the *nominal alternatives* making up the *nominal range determined by* this subject and *presented by* any question with this subject are the results  $Aa_1 \dots a_n$  of substituting a name  $a_i$  for a queriable  $x_i$  (for each  $i$ ) in the matrix  $Ax_1 \dots x_n$ , under the restriction that, if  $g(x_i)$  is defined and is  $\mathbf{C}x$ , then  $a_i$  must be in the nominal category determined by  $Cx$ .

For any interpretation  $M$ , the *real  $M$ -alternatives*, or *alternatives in  $M$* , which make up the *real  $M$ -range* or *range in  $M$* , *determined by*  $\langle X, g, Ax_1 \dots x_n \rangle$ , and *presented by* any question with this subject are all the



pairs  $\langle f, Ax_1 \dots x_n \rangle$ , where  $f$  is a function from  $X$  into the domain in  $M$ , under the restriction that, if  $g(x_i)$  is defined and is  $\mathbf{C}x$ , then  $f(x_i)$  is in the real range in  $M$  of  $Cx$ . A real  $M$ -alternative  $\langle f, Ax_1 \dots x_n \rangle$  is *true in  $M$*  just in case  $Ax_1 \dots x_n$  is true in that  $M'$  which is like  $M$  except in assigning  $f(x_i)$  to  $x_i$  for each  $i$ . Roughly, a nominal alternative  $Aa_1 \dots a_n$  *signifies in  $M$*  the real alternative  $\langle f, Ax_1 \dots x_n \rangle$ , provided that  $a_i$  denotes in  $M$  the individual  $f(x_i)$ .

A *lexical which-subject* is an expression of the form

$$(C_1x_1, \dots, C_rx_r, x_{r+1}, \dots, x_n \parallel Ax_1 \dots x_n),$$

where  $x_1, \dots, x_n$  is a nonempty nonrepeating sequence of variables, and  $C_1x_1, \dots, C_rx_r$  is a possibly empty sequence of category conditions, each  $C_ix_i$  being a category condition with  $x_i$  as its one free variable. Here each  $x_i$  is *governed by* the category condition  $C_ix_i$ , while  $x_{r+1}, \dots, x_n$  are *category-free*.

Given such a lexical which-subject, we can recover the abstract which-subject that it *signifies*, in the obvious way. The queriables are all of  $x_1, \dots, x_n$ , but  $g$  is defined only for  $x_1, \dots, x_r$ .

In a footnote added late (p. 26), Belnap and Steel say that it was a mistake to define an abstract which-subject as the triple  $\langle X, g, A \rangle$ , because then  $(x \parallel Fx)$  and  $(y \parallel Fy)$  signify distinct abstract which-subjects. They suggest defining an abstract which-subject not as  $\langle X, g, A \rangle$  but rather as “something amounting to the equivalence-class generated from this by means of uniform substitution for queriables”. To avoid complicating our exposition here we shall not make this change but will continue as in Belnap’s original development.

A *which-interrogative* has the form

$$?\rho(C_1x_1, \dots, C_rx_r, x_{r+1}, \dots, x_n \parallel Ax_1 \dots x_n).$$

The variables  $x_1, \dots, x_n$  occur free in  $Ax_1 \dots x_n$ , and are said to be free in the list  $C_1x_1, \dots, C_rx_r, x_{r+1}, \dots, x_n$ , but are considered to be *bound* in the interrogative as a whole.

Roughly: Every direct answer to an elementary question is a conjunction  $(S \wedge C \wedge D)$ , where  $S$  is a selection drawn from among the presented alternatives,  $C$  is a completeness-claim, and  $D$  is a distinctness-claim. The selection  $S$  is itself a conjunction  $(S_1 \wedge \dots \wedge S_p)$  without repetitions.  $(S_1 \wedge \dots \wedge S_p)$  is a *lexical selection*, and the corresponding set  $\{S_1, \dots, S_p\}$  is an *abstract selection*. In the case of which-questions, the *nominal selection signified by*  $(S_1 \wedge \dots \wedge S_p)$  is  $\{S_1, \dots, S_p\}$ ; if each of the  $S_i$  is in the range of a which-subject  $\sigma$ , then the *real selection signified by*  $(S_1 \wedge \dots \wedge S_p)$  *in  $M$  relative to  $\sigma$*  is the set of real alternatives signified by the  $S_i$  in  $M$  relative to  $\sigma$ .

Because direct answers have  $(S \wedge C \wedge D)$  structure, the request  $\rho$  in  $?\rho\sigma$  has a structure of the form  $(\mathbf{s} \mathbf{c} \mathbf{d})$ . Here  $\mathbf{s}$  is a selection-size specification,

which is a pair of numerals  $\overset{\mu}{\nu}$ , where  $\nu$  is a positive numeral representing a lower bound on the selection size, and  $\mu$  is either a positive numeral ( $\geq \nu$ ) representing an upper bound, or a dash signifying the absence of an upper bound. The  $\overset{\mu}{\nu}$  notation is a *lexical selection-size specification*, and it *signifies* the corresponding *abstract selection-size specification*, which is the corresponding ordered pair of cardinals (or the dash).

A subject  $\sigma$  *sanctions* a selection  $(S_1 \wedge \dots \wedge S_p)$  if each  $S_i$  is in the range determined by  $\sigma$ . A request  $\rho$  *sanctions* a selection if the length of the selection falls within the limits specified by the selection-size specification of  $\rho$ . An interrogative  $?\rho\sigma$  *sanctions* a selection if both  $\rho$  and  $\sigma$  do.

Roughly: The completeness-claim made by a direct answer is a claim as to how complete the selection is when measured against the totality of true alternatives presented by the question. Completeness-claims may be analyzed in terms of quantifiers, where a *quantifier*  $Q$  is defined as a binary relation between classes  $T$  and  $S$  (here  $T$  would be the set of true alternatives, and  $S$  would be the selection) such that whether or not  $Q(T, S)$  holds depends on the cardinalities of  $T \cap S$  and  $T - S$  (e.g. the universal quantifier is the quantifier in which  $T - S = 0$ ). Belnap mentions various examples of quantifiers that might be used, and the possibility of letting the specification indicate a range of completeness-claims. To develop these possibilities would require an enrichment of the assertoric basis, so Belnap confines attention to just the universal quantifier and the claim it represents, the maximal completeness-claim. Notation:  $\text{Max}(\sigma, S)$ .

For whether-questions: Given a lexical whether-subject and a selection  $S(S_1 \wedge \dots \wedge S_p)$  sanctioned by that subject, we define  $\text{Max}(\sigma, S)$ , the *maximum completeness-claim in  $\sigma$  and  $S$* , as  $(\neg B_1 \wedge \dots \wedge \neg B_r)$ , where  $B_1, \dots, B_r$  are (in order) all the members of the subject that are not in the selection  $S$ .

For which-questions: Assume as given a lexical which-subject

$$\sigma = (C_1 x_1, \dots, C_r x_r, x_{r+1}, \dots, x_n \parallel A x_1 \dots x_n)$$

and a selection

$$S = (A a_{1_1} \dots a_{1_n} \wedge \dots \wedge A a_{p_1} \dots a_{p_n})$$

sanctioned by  $\sigma$ . Then we define  $\text{Max}(\sigma, S)$  as

$$\forall x_1 \dots \forall x_n [C_1 x_1 \wedge \dots \wedge C_r x_r \rightarrow [A x_1 \dots x_n \rightarrow [(x_{1,n} = a_{1_1,n}) \vee \dots \vee (x_{1,n} = a_{p_1,n})]]],$$

where  $(x_{1,n} = a_{k_1,n}) = [(x_1 = a_{k_1}) \wedge \dots \wedge (x_n = a_{k_n})]$ .

A dash is used for the *lexical empty completeness-claim* specification. Then  $?(s - \mathbf{d})\sigma$  *specifies no completeness-claim*, and  $?(s \forall \mathbf{d})\sigma$  *specifies the maximum completeness-claim*. Given an interrogative  $I$  and a selection  $S$

sanctioned by  $I$ , the *completeness-claim sanctioned by  $I$  relative to  $S$* , or  $Comp(I, S)$ , is not defined if  $I$  specifies no completeness-claim, and is defined as  $Max(\sigma, S)$  if  $I$  specifies the maximum completeness-claim.

Single-example questions have the form  $?(1 - \mathbf{d})\sigma$ .

Some-examples questions have the form  $?(1 - \mathbf{d})\sigma$ .

Unique-alternative questions have the form  $?(1 \forall \mathbf{d})\sigma$ .

Complete-list questions have the form  $?(1 \forall \mathbf{d})\sigma$ .

For the following theorem we say that  $I_1$  is *erotetically equivalent to  $I_2$*  iff, for every  $d(I_1)$  there is an equivalent  $d(I_2)$ , and for every  $d(I_2)$  there is an equivalent  $d(I_1)$ .

**THEOREM 1.** *The completeness-claim specification is dispensable in some cases but not in all. Viz: (A) For each whether-question interrogative, there is an erotetically equivalent single-example whether-interrogative. (B) Unique-alternative which-interrogatives are erotetically equivalent to certain single-example which-interrogatives. (C) There are certain complete-list which-interrogatives that are not erotetically equivalent to any some-examples which-interrogative.*

Result (B) can be generalized from exactly-one to exactly- $n$ . The trick is to add to the subject the appropriate completeness-clause  $\forall y(Ay \rightarrow (y = x_1 \vee \dots \vee y = x_n))$ . Compare Åqvist [1965], pp. 123ff. Further results on equivalence of this sort are noted in Åqvist [1965] and Kubiński [1980], pp. 61–68.

Concerning distinctness-claims, Belnap says that only two kinds have a systematic use in question logic: the empty and the nonempty. Let  $\sigma$  and  $S$  be as in the definition of  $Max$ . Then  $Dist(\sigma, S)$  is defined as a conjunction of disjunctions of the form  $((a_{i_1} \neq a_{j_1}) \vee \dots \vee (a_{i_n} \neq a_{j_n}))$  which says that the conjuncts of  $S$  signify, relative to  $\sigma$ , distinct real alternatives.

*Notation.*  $-$  for the lexical empty distinctness-claim specification, and  $\neq$  for the lexical nonempty distinctness-claim specification. Given an interrogative  $I$  and a selection sanctioned by  $I$ , the *distinctness-claim sanctioned by  $I$  relative to  $S$* , or  $Dist(I, S)$ , is not defined if  $I$  specifies no distinctness-claim, and is defined as  $Dist(\sigma, S)$  if  $I$  does specify a distinctness-claim.

#### 4.4 Answers

Let  $I$  be an elementary interrogative and  $S$  be a selection sanctioned by  $I$ . Then  $A$  is a *direct answer to  $I$*  iff  $I$  and  $A$  have the corresponding forms indicated:

$I$	$A$
$?(s - -)\sigma$	$S$
$?(s\forall -)\sigma$	$S \wedge \text{Comp}(I, S)$
$?(s - \neq)\sigma$	$S \wedge \text{Dist}(I, S)$
$?(s\forall \neq)\sigma$	$S \wedge \text{Comp}(I, S) \wedge \text{Dist}(I, S)$

#### 4.5 Effectivity, Univocity, Completeness

Belnap says that his elementary interrogatives and their answers as defined above satisfy criteria of *effectivity* and *univocity*. One can effectively tell whether an expression is an elementary interrogative (effectivity) and, given that it is, what question it puts (univocity). Given a wff  $A$  and interrogative  $I$ , one can effectively tell whether  $A$  is a  $d(I)$  (effectivity) and, if it is, how it answers  $I$  (univocity). It is for the sake of effectivity that we require the nominal ranges of category conditions to be effective, allow  $n$ -ary conjunction, and bar conjunctions of alternatives from counting as alternatives.

Roughly, a *real answer* to a which-question is a sequence of real alternatives presented by that question, and the system is *complete* only if every real answer is expressible by some nominal answer. Completeness can fail if some entities do not have names, or do not have names in the proper nominal categories, and also if some true real answers are infinite. Also, as Belnap points out, no category system can be complete; for there are only denumerably many one-place wffs available to serve as category conditions, while there are nondenumerably many sets of names that might be wanted as categories. Roughly speaking, the positive result is that, for any given category system, Belnap's answer system is complete up to these limitations: the real answers must be finite, the entities involved must have names, and the names must be in the right nominal categories. (For the precise account, see Belnap and Steel [1976], Section 1.34.)

#### 4.6 Useful Abbreviations

Belnap suggests the following rules for abbreviating requests: (1) Drop parentheses around the request. (2) Omit dashes. (3) Omit '1' as a lower limit. Then, where  $\sigma$  is a subject, these rules allow

$?^1\sigma$	for	single-example questions,
$?\sigma$		some-examples questions,
$?\neq\sigma$		some-distinct-examples questions,
$?^1\forall\sigma$		unique alternative questions,
$?\forall\sigma$		complete-list questions,
$?\forall\neq\sigma$		complete-and-distinct-list questions.

For abbreviating subjects, where  $A$  has no free variables,  $A$  abbreviates  $(A, \neg A)$ ; and, where  $A$  contains exactly  $x_1, \dots, x_n$  free (given in order), then  $A$  abbreviates  $(x_1, \dots, x_n \parallel A)$ . Then ‘Is it the case that  $A$ ?’ can be expressed by either  $?^1 A$  or  $?A$ .

#### 4.7 Elementary-like Questions

In first-order functional calculi there are six parts of speech: open and closed wffs, open and closed terms, connectives, and quantifiers. These generate 36 types of *elementary-like* questions, each type *positing* an entity of some part of speech and having as *desiderata* entities of some parts of speech; e.g. a whether-question can be analyzed as positing a sequence of statements and asking for a truth-functional connection between them. Belnap credits this idea to Stahl [1962].

As noted above, Stahl assumed a higher-order function calculus and analyzed just individual-questions, function-questions, and truth-questions. Kubiński has considered enriched languages in which one can ask these and related questions (Kubiński [1980], Section I.12). It does not appear that anyone has yet made a full study of all 36 of the possibilities noted above, or of the analogous set for higher-order languages. The possibilities that Belnap discusses are six:

1. *Whether-questions.* Each of these posits a sequence (conjunction?) of statements and has truth-functional connectives as desiderata. The presented alternatives are formed by constructing truth-functional compounds of the posited wffs.
2. *Which-questions.* Each of these posits an open wff and has closed terms as desiderata. The presented alternatives are formed by substituting one of the closed terms for one of the free variables in the posited wff.
3. *Description-questions.* Each of these posits a closed term and has *descriptors* (i.e. open wffs) as desiderata. One way to formalize is this: Let  $L$  have a list of *determinables*, which are open wffs, and with each determinable let there be associated a list of *descriptors*. (E.g. with ‘ $x$  is a color’ we associate ‘ $x$  is red’, ‘ $x$  is green’, etc.) For semantic coherence each candidate interpretation must be such that an individual satisfies a determinable only if also satisfying some descriptor associated with it. Where  $Hx$  is a determinable with  $H_1x, \dots, H_ix, \dots$  with it, to posit a term  $b$  and call for  $H_1x, \dots, H_ix, \dots$  as desiderata we would use a new sort of subject  $\text{Des}(Hx \parallel b)$  whose range is the set of presented alternatives  $H_1b, \dots, H_ib, \dots$ . For this sort of question distinctness-claims can be expressed via wffs of the form

$\neg\forall x(H_i x \leftrightarrow H_j x)$ . Completeness-claims can be expressed in a first-order way only in special cases, however — e.g. in the case where the set of the associated descriptors is finite.

4. *Identity-questions*. Each of these posits a closed term and has closed terms as desiderata. We define the subject as  $\text{Ident}(Cx||b)$ , where  $b$  is the posit and  $Cx$  is a category condition with associated names  $a_1, \dots, a_i, \dots$ . The presented alternatives are  $(b = a_1)$ ,  $(b = a_2), \dots$
5. *What-questions*. Belnap distinguishes four sub-types: *equivalence questions*, with subject  $\text{Equiv}(Hx||Ax)$  and alternatives  $\forall x(Ax \leftrightarrow H_i x)$ ; *necessity-questions*, with subject  $\text{Nec}(Hx||Ax)$  and alternatives  $\forall x(Ax \rightarrow H_i x)$ ; *sufficiency-questions*, with subject  $\text{Suf}(Hx||Ax)$  and alternatives  $\forall x(H_i x \rightarrow Ax)$ ; and *intersection-questions*, with subject  $\text{Inter}(Hx||Ax)$  and alternatives  $\exists x(H_i x \wedge Ax)$ . In all cases here the  $H_i$  are descriptors associated with  $Hx$ . As with description-questions, completeness-claims can be expressed in a first-order language only in the finite case.
6. *How-many Questions*. Each of these posits an open wff and has quantifiers as desiderata. The alternatives are formed by prefixing the quantifier to the wff. Full study of this awaits further work on the logic of quantifiers.

#### 4.8 Compounding

Given interrogatives  $I_1, \dots, I_n$ , we write  $(I_1 \cup \dots \cup I_n)$  for the *unionized interrogative* of  $I_1, \dots, I_n$ . For this interrogative the concepts of subject and request are not defined. We say that  $A$  is a direct answer to it iff  $A$  is a direct answer to at least one of the  $I_i$ . In a similar way we can form intersection, complement, and set-difference questions. Belnap says that intersections and complements of questions do not appear to be useful, but set-difference might be (cf. ‘Tell me about  $\dots$ , without telling me about  $\dots$ ’).

The above-mentioned operations are boolean. In contrast there are other operations best thought of as logical or syntactical. Viz:

We define  $(I_1 \wedge \dots \wedge I_n)$  as the *conjunction* of  $I_1, \dots, I_n$ . Its direct answers are conjunctions  $(A_1 \wedge \dots \wedge A_n)$ , where each  $A_i$  is a direct answer to  $I_i$ . Belnap says that negation, disjunction, implication, and equivalence, conceived as logical operations on questions, do not seem to be of much interest.

Incidentally, the connective ‘or’ in English interrogatives is ambiguous.

Compare:

- ‘Who or what killed the dog?’ [exclusive *or*]
- ‘Have you been to Sweden, or have you been to Germany?’  
[inclusive *or*]
- ‘What day have you chosen, or what week?’ [nonsymmetric *or*]
- ‘Is it a bird or is it a plane?’ [simple whether-question]

For discussion see Stahl [1962], and Belnap and Steel [1976], p. 91.

Given a list of whether- and which-subjects  $\sigma_1, \dots, \sigma_n$ , we can form the *unionized subject*  $(\sigma_1 \cup \dots \cup \sigma_n)$ . The set of *presented alternatives* is the union of the sets presented by  $\sigma_1, \dots, \sigma_n$ , but the selections sanctioned by the unionized subject are conjunctions of alternatives presented by it. Expression of the appropriate completeness- and distinctness-claims is tedious.

Given elementary requests  $\rho_1, \dots, \rho_n$ , we let  $(\rho_1 \cup \dots \cup \rho_n)$  be a *unionized request*. The interrogative  $?( \rho_1 \cup \dots \cup \rho_n ) \sigma$  is to be treated like the union of the interrogatives  $?\rho_1 \sigma, \dots, ?\rho_n \sigma$ .

To formalize hypothetical questions, Belnap introduces *hypothetical interrogatives*  $(P \mid \rightarrow \mid I)$ , where  $P$  is a wff and  $I$  is an interrogative. The direct answers are wffs  $(P \rightarrow A)$ , where  $A$  is a  $d(I)$ .

To formalize *given-that questions*, Belnap introduces interrogatives  $(P \mid \wedge \mid I)$ , with direct answers  $(P \wedge A)$ , where  $A$  is a  $d(I)$ .

Belnap says that  $(P \mid \rightarrow \mid I)$  could be called an *added-condition* question, and  $(P \mid \wedge \mid I)$  an *added-conjunct* question. “We leave it to the reader to determine whether there is any point in introducing ‘added disjunction’ or ‘added equivalence’ questions” (pp. 98 – 99).

Let  $I$  be an interrogative with  $x$  free (hence, not a querable). Then we can allow  $\forall x I x$  to be an interrogative, and define the direct answers as conjunctions

$$(A_1 a_1 \wedge \dots \wedge A_n a_n \wedge \forall x (x \neq a_1 \wedge \dots \wedge x \neq a_n \rightarrow Bx)),$$

where  $A_1 x, \dots, A_n x, Bx$  are all direct answers to  $I$ .

To allow for qualification in terms of a category condition  $Cx$ , we use  $\forall_{[Cx]} I$  and define its direct answers as conjunctions

$$(A_1 a_1 \wedge \dots \wedge A_n a_n \wedge \forall x (Cx \rightarrow (x \neq a_1 \wedge \dots \wedge x \neq a_n \rightarrow Bx)))$$

as above, but requiring also that each  $a_i$  is in the nominal category determined by  $Cx$ .

To formalize ‘Answer  $I$  for some  $x$  in the domain’ (where  $x$  is not free in  $I$ ), Belnap uses not  $\exists x I$  but  $\cup x I$ , whose direct answers are the wffs  $Aa$  such that  $Ax$  is a direct answer to  $Ix$ . Thus  $\cup x ?^1(Px)$  has the same answers as  $?^1(x \parallel Px)$ . Belnap says that  $\exists x I$  might have a use if the operation of disjunction on questions can be shown to have a use.

As emphasized by many writers, conditional questions call for an answer only if a given condition obtains. A general theory may use these basic notions:  $I$  calls for an answer in an interpretation  $M$ , and  $A$  is a *direct answer to  $I$  in  $M$* . The latter is undefined unless  $I$  is *operative* (i.e. calls for an answer) in  $M$ . *Absolute* interrogatives are those for which direct answer is defined without relativization to  $M$ . *Relativized* interrogatives are those for which direct answerhood is relativized to interpretations. *Categorical* interrogatives call for an answer in every  $M$  and have the same direct answers in every  $M$ .

Belnap writes  $(P/I)$  for a *conditional interrogative* with *condition  $P$*  and *conditioned interrogative  $I$* . Then  $(P/I)$  calls for an answer in an interpretation  $M$  iff  $P$  is true in  $M$  and  $I$  calls for an answer in  $M$ . If  $(P/I)$  calls for an answer in  $M$ , then  $A$  is a *direct answer to  $(P/I)$*  iff  $A$  is a direct answer to  $I$  in  $M$ .

Belnap also discusses conjunction and union of relativized interrogatives, and points out the natural generalization via universal quantification.

#### 4.9 *Presupposition and Truth*

Belnap's intention is, roughly, that every question presupposes precisely that at least one of its direct answers is true. (For relativized questions we should prefix 'if the question is operative, then ...'.) Presuppositions can be attached to interrogatives in a way that parallels their attachment to questions. For simplicity we formulate the discussion here in terms of interrogatives. We assume throughout that  $A$  is any wff, and  $M$  is any interpretation. As before,  $d(I)$  abbreviates 'direct answer to  $I$ ', and  $D(I)$  denotes the set of direct answers to  $I$ .

Let  $I$  be a whether-interrogative. Then:  $I$  is *true in  $M$*  iff some  $d(I)$  is true in  $M$ , and  $I$  is *false in  $M$*  otherwise.  $I$  *presupposes  $A$*  if  $A$  is true in every  $M$  in which  $I$  is true.  $A$  *expresses-the-presupposition-of  $I$*  if  $A$  is true in exactly those  $M$  in which  $I$  is true.

Let  $I$  be a which-interrogative. Then:  $I$  is *really true* [*really false*] in  $M$  iff some real answer to  $I$  is true in  $M$  [every real answer to  $I$  is false in  $M$ ].  $I$  is *nominally true* [*nominally false*] in  $M$  iff some (nominal)  $d(I)$  is true in  $M$  [every  $d(I)$  is false in  $M$ ].  $I$  *really* [*nominally*] *presupposes  $A$*  iff  $A$  is true in every  $M$  in which  $I$  is really [nominally] true.  $A$  *expresses-the-real-* [*nominal-*] *presupposition of  $I$*  iff  $A$  is true in exactly those  $M$  in which  $I$  is really [nominally] true. For uniformity we use unqualified ' $I$  is true in  $M$ ', ' $I$  presupposes  $A$ ', and ' $A$  expresses-the-presupposition-of  $I$ ' for both which- and whether-interrogatives, meaning for which-interrogatives the 'real' variety.

If there is an  $A$  that expresses-the-presupposition-of  $I$ , then there are indefinitely many such  $A$ , so it is convenient to pick one such and call it *the* presupposition of  $I$ ; notation:  $\text{Pres}(I)$ . For whether-interrogatives we



choose some disjunction of the direct answers. For which-interrogatives the construction is straightforward but tedious. Belnap's example: Let  $I$  be  $? \rho(Cx \parallel Fx)$ . Then  $\text{Pres}(I)$  will be a conjunction  $(P_1 \wedge P_2 \wedge P_3)$  of at most three conjuncts.  $P_1$  is always present, and says that at least one  $C$  is an  $F$ .  $P_2$  is present just in case  $\rho$  is  $(\overset{\nu}{\mu} \forall \mathbf{d})$ , where  $\nu$  is an integer;  $P_2$  says that at most  $\nu$   $C$ 's are  $F$ 's.  $P_3$  is present just in case  $\rho$  has the form  $(\overset{\nu}{\mu} \mathbf{c} \neq)$ ;  $P_3$  says that at least  $\mu C$ 's are  $F$ 's.

It is possible to define  $\text{Pres}(I)$  in such a way that, for each elementary interrogative  $I$ ,  $\text{Pres}(I)$  is an effectively specified wff that expresses-the-presupposition-of  $I$ . In the case of nominal presuppositions it is not in general possible to find for each which-interrogative  $I$  a wff that expresses-the-presupposition-of  $I$ . See Belnap [1963], Section 7.5.

Let us say that  $X$  is a *quasi-wff* iff  $X$  is a wff or an interrogative. Then, using  $X$  for quasi-wffs and  $H$  for sets of quasi-wffs, we define:  $M$  is an *H-interpretation* iff every member of  $H$  is true in  $M$ .  $X$  is *logically H-true*, *H-consistent*, or *H-inconsistent* according as  $X$  is true in all, some, or no  $H$ -interpretations.  $H'$  *propositionally H-implies*  $X$  iff  $X$  is true in every  $H$ -interpretation in which every member of  $H'$  is true. Similarly with other semantic concepts. For brevity the 'propositionally' may be omitted before 'H-implies' and 'H-equivalent'.

#### 4.10 Types of Answer

Let us say that a wff, considered as a reply to  $I$ , is *H-uninformative* if it is  $H$ -implied by  $I$ , and otherwise *H-informative*. Such a wff is *H-foolish* if it is  $H$ -inconsistent, and otherwise *H-possible*. It is *relatively H-foolish* if it is  $H$ -inconsistent with  $I$ , and otherwise *relatively H-possible*.

Such a wff is an *H-complete answer to I* iff it  $H$ -implies some  $d(I)$ , and an *H-just-complete answer* if it is  $H$ -equivalent to some  $d(I)$ . It is an *H-partial answer* iff it is  $H$ -implied by some  $d(I)$ . It is an *H-eliminative answer* iff it  $H$ -implies the negation of some  $d(I)$ , and *H-quasi-eliminative answer* iff it is  $H$ -implied by the negation of some  $d(I)$ .

Belnap says that  $A$  is a *Harrah-H-complete answer to I* iff  $A$  is a  $(H \cup \{I\})$ -complete answer to  $I$ . The essential idea, which was suggested by Harrah [1961], is that in normal cases the questioner believes that  $I$  is true and thus includes  $\text{Pres}(I)$  in his background knowledge. Example: Let  $I$  be  $?^1((A \vee B), (C \wedge D))$ . Then  $\neg C$  is not a complete answer, but it is a Harrah-complete answer. The same idea can be used to generate 'Harrah' variants on many other concepts, some of which are noted below. Suggestion: Find a substitute for the label 'Harrah', preferably a short word meaning 'presupposition-aided'.

A *proper H-partial answer* is a partial answer that is implied by some  $H$ -consistent  $d(I)$ .  $A$  is a *highly proper H-partial answer* iff  $A$  is both  $H$ -

informative and a proper  $H$ -partial answer. (In contrast, the *safe* answers are the uninformative partial answers.)

$A$  is a *nominal  $H$ -corrective answer to  $I$*  iff  $A$  implies the negation of every  $d(I)$ .  $A$  is an  *$H$ -corrective answer to  $I$*  iff  $I$  is (really) false in every  $M$  in which  $A$  is true. For a standard corrective answer we may choose  $\neg\text{Pres}(I)$  and abbreviate it (following Åqvist [1965]) by  $\text{Corr}(I)$ .

Any wff that counts as an answer to  $I$  relative to  $H$ , in any of the senses of answer defined above, may be said to be *erotetically  $H$ -relevant to  $I$* .

#### 4.11 Properties of Interrogatives

An interrogative  $I$  is  *$H$ -safe* iff  $I$  is logically  $H$ -true (i.e.,  $H$  implies  $I$ ); otherwise  $I$  is  *$H$ -risky*.  $I$  is  *$H$ -foolish* iff  $I$  is  $H$ -inconsistent; otherwise  $I$  is  *$H$ -possible*. (Thus  $I$  is  $H$ -safe if  $H$  implies  $\text{Pres}(I)$ , and  $H$ -foolish if  $H$  implies  $\text{Corr}(I)$ .) Parallel to these (real) concepts for interrogatives there are nominal variants, and there are concepts for questions.

**THEOREM 2.** [*Belnap's Hauptsatz, or, the Theorem of the Fifth Gymnosophist (see Plutarch's Alexander)*] *Ask a foolish question and you get a foolish answer.*

Belnap gives a proof of the corresponding result for interrogatives. The proof is straightforward. (See Belnap and Steel [1976], pp. 131 – 133.)

Continuing in terms of interrogatives:  $I$  is *dumb* iff  $I$  has no direct answers.  $I$  is  *$H$ -exclusive* if in each  $H$ -interpretation there is at most one true real answer (for which-interrogatives) or abstract answer (for whether-interrogatives). (Thus  $?^1\forall(\dots)$  and  $?^1\forall(A_1, \dots, A_n)$  turn out to be exclusive.)  $I$  is a *Hobson's Choice* if  $I$  has exactly one direct answer (e.g.  $?^1(A)$ ).

$I$  is *answerable by  $H$*  iff  $H$  implies some  $d(I)$ ; otherwise  $I$  is  *$H$ -unanswerable*.  $I$  is *Harrah-answerable by  $H$*  if  $(H \cup \{I\})$  implies some  $d(I)$ . Here, if  $H$  is the questioner's beliefs, we can say that  $I$  is  *$H$ -rhetorical*. Similarly,  $I$  may be *Harrah-unanswerable by  $H$* , and, if  $H$  is the questioner's beliefs, we say that  $I$  is *moot*, or *open*.  $I$  is *hyper- $H$ -moot*, or  *$H$ -wide-open*, if  $H$  provides neither a Harrah-complete answer nor an eliminative answer.

Following the suggestion of Åqvist (who used the label 'normal'),  $I$  is  *$H$ -independent* if no  $d(I)$   $H$ -implies any other  $d(I)$ .  $I$  is  *$H$ -minimal* if, for every  $A$  in  $D(I)$ , there is an  $H$ -interpretation  $M$  in which  $A$  is the one and only true  $d(I)$ . Minimality implies independence, but not conversely.

$I_1$   *$H$ -contains  $I_2$*  just in case every  $H$ -consistent  $d(I_1)$  is an  $H$ -complete answer to  $I_2$ .  $I_1$  *Harrah- $H$ -contains  $I_2$*  iff every  $d(I_1)$  is a Harrah- $H$ -complete answer to  $I_2$ .

*Example:*  $?^1(A, B)$  Harrah-contains  $?^1\forall(A, B)$ .

$I_1$  is *erotetically  $H$ -equivalent* to  $I_2$  iff for every  $d(I_1)$  there is an  $H$ -equivalent  $d(I_2)$ , and, for every  $d(I_2)$  there is an  $H$ -equivalent  $d(I_1)$ .  $I_1$  is

*Harrah-erotetically H-equivalent* to  $I_2$  iff  $I_1$  is erotetically  $(\{I_1, I_2\} \cup H)$ -equivalent to  $I_2$ . Example:  $?^1(A, B)$  and  $?^1\forall(A, B)$ .

$I_1$  is *erotetically H-relevant* to  $I_2$  iff some  $d(I_1)$  is erotetically  $H$ -relevant to  $I_2$ .

*Example:*  $?^1(A, \neg A)$  is (propositionally) equivalent to  $?^1(B, \neg B)$  but is not erotetically relevant to it.

$I_1$  *H-obviates*  $I_2$  iff every  $d(I_1)$  is either an  $H$ -corrective answer or an  $H$ -complete answer to  $I_2$ .

#### 4.12 Extending Belnap's Analysis

There are many possibilities for further development of Belnap's system. Some of these are suggested in an obvious way by the system itself.

First: Belnap formalizes six types of elementary-like question (4.7 above), but, as Belnap and Steel point out, thirty types remain. How should they be formalized? Do they have a natural semantics? Do they have any interesting use? (For a relevant discussion, see Hiž [1962].)

Second: Belnap formalizes the maximum completeness-claim. As he points out, indefinitely many other types of completeness-claim remain to be studied.

Third: Belnap and Steel say that only one type of distinctness-claim seems to have a systematic use in question logic. This may be doubted, however, because of examples like the police chief who says

‘Give me a (nominal) list of at least ten suspects, of whom at least seven are (really) distinct.’

Surely the matter warrants further study.

Fourth: Several possibilities are suggested by ideas of Kubiński. (For a summary in English, see Kubiński [1980].) Kubiński assumes an indefinitely large stock of interrogative operators. Simple interrogatives are formed by prefixing an interrogative operator to a sentential function that contains free variables (which are then bound by the queriables in the operator). One interesting difference from Belnap concerns answer-size specifications. Belnap's interrogatives contain a selection-size specification, but each of Kubiński's interrogative operators is in effect a string of numerical quantifiers, with each quantifier binding its corresponding queriable, so the quantity specifications are attached to the queriables individually. This makes possible a straightforward formalization of interrogatives like:

‘Which two ministers voted against which three projects on which five committees?’

‘Which at most three kings have ruled which at least two countries?’

(Possible objection: English does not provide interrogatives like these. Reply: Polish shows that it ought to.) For more on interrogatives like these, and efficient ways of formalizing them, see Wiśniewski [1995], §2.2.6.2. In connection with Belnap’s system, what remains to be studied are (1) which of these are warranted in the system, and (2) how they can be accommodated in a smooth way.

Fifth: As noted by many authors, the more one enriches the underlying assertoric logic, the more questions one can construct. Some possible enrichments mentioned by Belnap and Steel are: adding variables of higher type, adding modal operators, and allowing infinite conjunctions.

Sixth: Although Belnap has argued for the importance of effectiveness at certain points in a logic of questions, and indeed one might take this as essential to his approach, we might study how the system could be extended if various of the effectivity requirements were relaxed. One example: If we drop the requirement that nominal categories be decidable, we can allow questions like

‘Which theorems of set theory should a number-theorist know?’

## 5 EPISTEMIC ANALYSIS OF QUESTIONS

### 5.1 *Motivation*

In this section we outline what is called the *imperative-epistemic* approach to questions. The object of study is said to be the ‘standard’ situation, in which (1) the questioner does not know any direct answer, and (2) the interrogative is taken to be synonymous with an imperative of the form

‘Let it be the case that I know . . .’ or

‘Make it the case that I know . . .’.

In our discussion below we may refer to this as the MMK (or ‘Make Me Know’) approach.

Kubiński mentions Bolzano and Loeser as precursors of this approach (Kubiński [1980], p. 131), but the first to give a substantial formal analysis was Åqvist. Accordingly we begin here by summarizing, in 5.2 – 5.6, the work presented in Åqvist [1965].

In the next sections (5.7 – 6.3) we outline some refinements and further developments, and some other theories that analyze questions in terms of imperatives of various kinds. Many of these theories have an empirical (perhaps phenomenological) motivation, with an implied claim that most actual question situations are “standard” (see, e.g., Wachowicz [1978], p. 157). Complications may arise if one formalizes via a logical framework (e.g., epistemic logic) that has a normative motivation.

On all variants of this general approach one may use SA methodology, but it is not necessary. It is possible to develop a substantial part of the logic of questions within the logic of imperatives, without considering answers at all. To show this in detail is a contribution of Åqvist.

## 5.2 Foundations

Åqvist assumes an applied first-order predicate calculus with the quantifiers  $(Ux)$  and  $(Ex)$  and identity, supplemented with the operators

$!$	(‘Let it turn out to be the case that’)
$i$	(‘It is permissible that’)
$K$	(‘I know that’)
$P$	(‘It is compatible with everything that I know that’)

To refer to free individual symbols (individual constants and free variables) Åqvist uses:  $a, b, c$ . To refer to bound individual variables:  $x, y, z$ . For one-place predicate constants:  $J_i$ . For  $n$ -place predicate constants:  $F_i^n$ . For the result of putting  $x$  for all free occurrences of the variable  $a$  in  $p$  we write:  $p(x/a)$ .

For QIE (quantified imperative-epistemic) logic:

1. Every propositional constant, predication (predicate followed by its arguments), and identity ( $a = b$ ) is a QIE-wff.
2. If  $p$  and  $q$  are QIE-wffs, then so are  $\neg p$ ,  $(p \wedge q)$ ,  $(p \vee q)$ .
3. If  $p$  is a QIE-wff containing free occurrences of  $a$ , but not containing  $!$  or  $i$ , then  $(Ux)p(x/a)$  and  $(Ex)p(x/a)$  are QIE-wffs.
4. If  $p$  is a QIE-wff not containing  $!$  or  $i$ , then  $Kp$  and  $Pp$  are QIE-wffs.
5. If  $p$  is a QIE-wff not containing  $!$  or  $i$  or any free variables, then  $!p$  and  $ip$  are QIE-wffs.

A *sentence* is a QIE-wff with no free variables. A *statement* is a sentence that does not contain  $!$  or  $i$ . An *ordinary statement* is a sentence that does not contain  $!$ ,  $i$ ,  $K$ , or  $P$ .

For the semantics Åqvist adopts (with some refinements) Hintikka’s notions of *model set* and *model system* for the logic of  $K$  and  $P$ , and (following Kanger) adds a relation of imperative alternativeness and imposes appropriate conditions, e.g. if  $\mu$  is a model set in a model system  $\Omega$  and  $\mu^+$  is an imperative alternative to  $\mu$  in  $\Omega$ , then, if  $!p$  is in  $\mu$  then  $p$  is in  $\mu^+$ . On the other hand,  $!$  and  $i$  are to act as genuine imperative operators only when their arguments are epistemic statements; if  $p$  is an ordinary statement in  $\mu$  and  $\mu^+$  is an imperative alternative to  $\mu$ , then  $p$  is in  $\mu^+$ . To have

a complete logic of imperatives one would have to add further imperative operators (with distinct properties) that apply to ordinary statements.

If one wants to satisfy both normative and empirical motivations, one must impose a relatively complex set of conditions on model systems. We omit details. The culminating idea is the usual one: A set of sentences is *consistent* if it can be embedded in a model set that is a member of a model system. A sentence  $p$  is *valid* if  $\{\neg p\}$  is not consistent, and  $p$  *entails*  $q$  if  $\{q\}$  is embeddable wherever  $\{p\}$  is.

### 5.3 Definition of Questions

Åqvist defines interrogatives by introducing various interrogative operators in abbreviations of sentences of the form  $!p$ . The following are examples.

*Whether-questions.*

$$(1) \quad ?_n(p_1, \dots, p_n) =_{df}!(Kp_1 \vee \dots \vee Kp_n),$$

where  $n \geq 2$  and each  $p_i$  is an ordinary statement.

$$(2) \quad ?_{n/m}(p_1, \dots, p_n \mid r_1, \dots, r_m) =_{df}!(r_1 \wedge \dots \wedge r_m \rightarrow (Kp_1 \vee \dots \vee Kp_n))$$

where  $n \geq 2$ ,  $m \geq 1$ , and each  $p_i$  and  $r_i$  is an ordinary statement.

Monadic *complete-list* which-questions. Let  $p$  be a QIE-wff not containing  $!$ ,  $i$ ,  $K$ , or  $P$  and containing just one free variable  $a$ ; let  $x$ ,  $y$ ,  $z$  be the alphabetically earliest variables not bound in  $p$ . Then we define:

$$(3) \quad (?_Aa)p =_{df}!(Ux)(p(x/a) \rightarrow (Ey)(y = x \wedge (Ez)K(z = y)))$$

$$(4) \quad (?_Ba)p =_{df}!(Ux)(p(x/a) \rightarrow (Ey)(y = x \wedge Kp(y/a)))$$

$$(5) \quad (?_KBa)p =_{df}!K(Ux)(p(x/a) \rightarrow (Ey)(y = x \wedge Kp(y/a)))$$

$$(6) \quad (?_EBa)p =_{df}!((Ex)Kp(x/a) \wedge (Ux)(p(x/a) \rightarrow (Ey)(y = x \wedge Kp(y/a))))$$

$$(7) \quad (?_EKBa)p =_{df}!((Ex)Kp(x/a) \wedge K(Ux)(p(x/a) \rightarrow (Ey)(y = x \wedge Kp(y/a))))$$

The latter two are equivalent to

$$(8) \quad (Ex)p(x/a) \wedge (?_Ba)p$$

$$(9) \quad (Ex)p(x/a) \wedge (?_KBa)p$$

Thus  $(?_{EKBa})$  and  $(?_{EBa})$  carry a nonemptiness claim;  $(?_{EKBa})$  and  $(?_{KBa})$  carry a completeness request. Åqvist says that  $(?_Aa)$  is too weak to be of interest, and that  $(?_{EKBa})$  best reflects ordinary use.

Monadic *at-least-which* and *exactly-which*.

$$(10) \quad (?_{C_n} a)p =_{df}!(Ex_1) \dots (Ex_n)K(p(x_1/a) \wedge \dots \wedge p(x_n/a) \wedge x_i \neq x_j)$$

$$(11) \quad (?_{D_n} a)p =_{df}!(Ex_1) \dots (Ex_n)K(p(x_1/a) \wedge \dots \wedge p(x_n/a) \wedge x_i \neq x_j \wedge \\ \wedge (Uy)(p(y/a) \rightarrow y = x_1 \vee \dots \vee y = x_n))$$

where  $x_1, \dots, x_n, y$  are the earliest distinct variables not bound in  $p$ , and where ' $x_i \neq x_j$ ' abbreviates the obvious distinctness-clause.

The interrogative  $(?_{D_n} a)p$  is equivalent to  $(?_{C_n} a)p'$ , where  $p'$  is

$$(p \wedge (Uy)(p(y/a) \rightarrow y = x_1 \vee \dots \vee y = x_n)).$$

Cf. the comments on the Theorem in Section 4.3 above.

Pure polyadic *relational* which-questions. The generalization to the case where monadic  $p$  is replaced by polyadic  $p$  is straightforward. Here, the defined operators have the form  $(?_{*}^m a_1, \dots, a_m)$ , where  $*$  is  $B, KB, EB, EKB, C_n$ , or  $D_n$ .

Mixed polyadic which-questions. Let  $R$  be a binary relation. Åqvist shows how to define operators for: 'Which [At least which  $n$ ] [Exactly which  $n$ ] objects bear  $R$  to which (or, at least which  $k$ , or, exactly which  $k$ ) objects?' e.g., the interrogative  $(?_{D_n-B}^2 a, b)$  is introduced to abbreviate

$$\begin{aligned} &!(Ex)(Ey)K((Ez)Rzx \wedge (Ez)Ryz \wedge x \neq y \wedge \\ &\wedge (Ut)((Ez)Rtz \rightarrow t = x \vee t = y) \wedge (Uu)(Rxu \rightarrow \\ &\quad \rightarrow (Ew)(w = u \wedge KRxw) \wedge \\ &\quad \wedge (Uu)(Ryu \rightarrow (Ew)(w = u \wedge KRyw))), \end{aligned}$$

which expresses

'Exactly which two objects bear  $R$  to which things?'

Generalization from binary  $R$  to  $n$ -ary  $R$  is straightforward but tedious.

*Categorically qualified* which-questions. One way to provide for categorial qualification is to use many sorts of variables  $x^\alpha$  and then define interrogatives of the form

$$(?_{*}^m x_1^{\alpha_1} \dots x_m^{\alpha_m})p.$$

Åqvist prefers to retain one-sorted theory and define interrogatives of the form

$$(?_{*a_1, \dots, a_m}^{J_1, \dots, J_m})p.$$

Roughly, the latter are defined by inserting  $[J_i x_i \rightarrow]$  or  $[J_i x_i \wedge]$  in the appropriate places.

*Mixed whether-which* and *conditional-which* questions. Åqvist defines *CoreQ* as the epistemic statement that is the scope of the ! operator in *Q*. Now let  $r_1, \dots, r_k$  be ordinary QIE-statements. Then:

$$(12) \quad \begin{aligned} & (?_{EKB}^m a_1, \dots, a_m; + k)(p; r_1, \dots, r_k) \\ & =_{df}!(\text{Core}(?_{EKB}^m a_1, \dots, a_m)p \vee Kr_1 \vee \dots \vee Kr_k) \end{aligned}$$

$$(13) \quad \begin{aligned} & (?_{EKB}^m a_1, \dots, a_m | k)(p | r_1, \dots, r_k) \\ & =_{df}!(r_1 \wedge \dots \wedge r_k \rightarrow \text{Core}(?_{EKB}^m a_1, \dots, a_m)p). \end{aligned}$$

Similar operators can be defined for *EB*,  $C_n$ ,  $D_n$ . Also, categorial qualification can be added in the obvious way.

#### 5.4 Presupposition, Riskiness, and Guarding

Let us use *Q* as a variable over questions. Then: *Q presupposes p* iff *p* is an ordinary statement, and *CoreQ* entails *p*. [For *Core*, see above.] A statement *p* is a *correction* to *Q* iff *p* is the negation of some presupposition of *Q*. A *safe* question is one whose presuppositions are valid. A *risky* question is one that is not safe.

*PresQ* is the result of dropping all occurrences of imperative and epistemic operators from *Q*. Claim: If *Q* entails an ordinary statement *p*, then *PresQ* entails *p* (Åqvist [1965], p. 133). We interpret *PresQ* as being *the* presupposition of *Q*.

The *correction of Q* = *CorrecQ* =  $\neg$ *PresQ*.

Roughly, guarding a risky question consists of transforming it into a safe one. Three methods for guarding a risky whether-question of the form  $?_n(p_1, \dots, p_n)$  are:

I. [Whether-Whether Method] Use

$$?_{n+1}(\text{Correc}(?_n(p_1, \dots, p_n)), p_1, \dots, p_n)$$

II. [Whately-Prior Method] Use

$$(?_1 \text{Pres}(?_n(p_1, \dots, p_n)) \wedge ?_{n/1}(p_1, \dots, p_n / \text{Pres}(?_n(p_1, \dots, p_n))))$$

III. [Whether-If Method] Use

$$?_{n/1}(p_1, \dots, p_n / \text{Pres}(?_n(p_1, \dots, p_n)))$$

Here methods I and II are equivalent.

Four methods for guarding a risky which-question of the form  $(?_{EKB}^m a_1, \dots, a_m)p$  are:



IV. [Mixed Whether-Which Method] Use

$$({}^m_{EKB}a_1, \dots, a_m; +1)(p; \text{Correc}({}^m_{EKB}a_1, \dots, a_m)p))$$

V. [Whately-Prior Method] Use

$$({}_1\text{Pres}({}^m_{EKB}a_1, \dots, a_m)p) \wedge \\ ({}^m_{EKB}a_1, \dots, a_m/1)(p/\text{Pres}({}^m_{EKB}a_1, \dots, a_m)p))$$

VI. [Which-If Method] Use

$$({}^m_{EKB}a_1, \dots, a_m/1)(p/\text{Pres}({}^m_{EKB}a_1, \dots, a_m)p))$$

VII. [Weak Mixed Whether-Which Method] Use

$$({}^m_{KB}a_1, \dots, a_m)p.$$

Here methods IV and V are equivalent. Similar techniques can be used for the cases of  $EB$ ,  $C_n$ , and  $D_n$ .

### 5.5 Direct Answers

Roughly, a direct answer  $d$  to  $Q$  should be such that, if the questioner comes to know that  $d$  is true, then the epistemic request expressed by  $Q$  is satisfied. Specific criteria that a theory should meet include the following. (These criteria will become clear in terms of the examples below.) First, if  $p$  is either a direct answer to  $Q$  [relative to  $r$ ] or a direct pseudo-answer to  $Q$ , then  $p[p \wedge r]$  should entail  $\text{Pres}Q$ . Second,  $Q_1$  is to entail  $Q_2$  iff every direct proper answer to  $Q_1$ , as well as every direct pseudo-answer (if any) to  $Q_1$ , entails some direct proper answer or direct pseudo-answer to  $Q_2$ . Åqvist does not yet have a complete theory of direct answers, but makes at least the following specific proposals. The numbers refer to the questions defined above in section 5.3.

To (1) the direct answers are  $p_1, \dots, p_n$ .

To (2) the direct answers are  $p_1, \dots, p_n$ ; and  $\neg(r_1 \wedge \dots \wedge r_m)$  is a direct pseudo-answer.

To (7) the direct answers are

$$(p(c_1/a) \wedge \dots \wedge p(c_k/a) \wedge (Ux)(p(x/a) \rightarrow \\ \rightarrow (x = c_1 \vee \dots \vee x = c_k)) \wedge [\dots (Ey)(y = c_k)]).$$

Here, and below,  $x$  and  $y$  are to be the earliest variables not bound in  $p$ , and  $[\dots (Ey)(y = c_k)]$  abbreviates

$$((Ey)(y = c_1) \wedge \dots \wedge (Ey)(y = c_k)).$$

To (6) the statements

$$(p(c_1/a) \wedge \dots \wedge p(c_k/a) \wedge [\dots (Ey)(y = c_k)])$$

are direct answers *relative to* the ordinary statement

$$(Ux)(p(x/a) \rightarrow (x = c_1 \vee \dots \vee x = c_k)).$$

To (10) the direct answers are

$$(p(c_1/a) \wedge \dots \wedge p(c_n/a) \wedge [(c_i \neq c_j)] \wedge [\dots (Ey)(y = c_n)]),$$

where there are no repetitions among the  $c$ 's and  $[(c_i \neq c_j)]$  is the obvious conjunction of nonidentities.

To (11) the direct answers are

$$\begin{aligned} &(p(c_1/a) \wedge \dots \wedge p(c_n/a) \wedge [(c_i \neq c_j)] \wedge (Ux)(p(x/a) \rightarrow \\ &\rightarrow (x = c_1 \vee \dots \vee x = c_n)) \wedge [\dots (Ey)(y = c_n)]). \end{aligned}$$

To (12) the direct answers are the direct answers to (7), plus  $r_1, \dots, r_k$ .

To  $(?_{EBa; +k})(p; r_1, \dots, r_k)$  the direct answers are the relativized direct answers to (6), plus  $r_1, \dots, r_k$ .

Each direct answer to  $(?_{C_n a} p)$  [i.e. (10)] is a direct answer to  $(?_{C_n}^J p)$  relative to the statement  $(Jc_1 \wedge \dots \wedge Jc_n)$ .

Each direct answer to  $(?_{D_n a} p)$  [i.e. (11)] is a direct answer to  $(?_{D_n}^J p)$  relative to  $(Jc_1 \wedge \dots \wedge Jc_n)$ , except that in the direct answer we insert  $[Jx \rightarrow]$  after  $(Ux)$ .

$(?_{EKB a}^J)$  is handled in the same way  $(?_{D_n}^J)$  is. For  $(?_{EB a}^J)$  we relativize to a complete-list statement.

For all the cases given above the generalization from monadic  $p$  to polyadic  $p$  is straightforward. For details see Åqvist [1965], pp. 156 – 158.

Åqvist indicates, but does not prove, that the foregoing proposals satisfy the criteria mentioned at the beginning of Section 5.5 (Åqvist [1965], p. 156). On the other hand, he says that the theory is not yet complete. There is, e.g., an open problem concerning how to define direct answer for  $(?_{KB})$  questions (*Ibid.*, p. 158).

Åqvist cites the following as an incoherence. Let  $Q$  be  $(?_{EKB}^1 x)Nx$ , expressing ‘Which things are the natural numbers?’. Then  $\text{Pres}Q$  is true, but no  $d(Q)$  is true. He conjectures that this is the only type of question that falsifies the *Coherence Principle*: For every  $Q$  and  $M$ ,  $\text{Pres}Q$  is true in  $M$  iff some  $d(Q)$  is true in  $M$ . (*Ibid.*, p. 160) (Cf. Section 4.9 above.)

## 5.6 Extending the System

Åqvist conjectures: We can accommodate all of Belnap’s questions if we add set theory to QIE, construe questions as presenting sets of alternatives, construe Belnap’s request-indicators as quantifiers over these sets, and interpret these quantifiers in terms of imperative, epistemic, and ordinary quantifiers (*Ibid.*, p. 82). This program remains to be carried out.

### 5.7 Revising the Foundations

Several motivations led to a revised proposal in Åqvist [1971]. Roughly: The 1965 analysis assumed a single context of use, a single knower, a nonindexical  $K$ , and monadic imperative operators  $!$  and  $i$ . The revision assumes a set of possible contexts (each context being a tuple  $C = \langle s, r, t, \dots \rangle$ , where  $s$  is a sender,  $r$  a receiver,  $t$  a time,  $\dots$ ). It also assumes a set of possible knowers, an indexical  $K$ , and imperative operators  $!$  and  $i$  that are dyadic (to express conditional obligation and permission) and may be indexical as well.

The new  $K$  carries a subscript for the knower and may carry other subscripts for other parameters such as time. Thereby we have not only standard questions ('Make me know') but also test questions ('Make me know that you know').

Because  $!$  is conditional, the basic form of interrogative is now the conditional. However, unconditional forms can be defined straightforwardly, e.g.

$$?_n(p_1, \dots, p_n) =_{df} (Kp_1 \vee \dots \vee Kp_n / p_1 \vee \dots \vee p_n) \wedge \wedge (p_1 \vee \dots \vee p_n).$$

To make the epistemic logic work more smoothly, Åqvist introduces several new kinds of existential quantifier. These might generate new types of interrogative, but this matter has not been studied in detail.

### 5.8 Hintikka's Development

In general, on the MMK approach the point of a question is to express an epistemic request, and the point of an answer is to satisfy this request. Loosely speaking, in the 1960s Åqvist developed a theory of epistemic requests, and, since the mid 1970s, Hintikka has been articulating a theory of epistemic request-satisfaction.

Consider

'Bring it about that I know that  $A$  or I know that  $B$ .'

In Hintikka's exposition this has a presupposition — namely,

' $A$  or  $B$ '

and a *desideratum* — namely,

'I know that  $A$  or I know that  $B$ .'

A reply that satisfies the epistemic request of the questioner completely is a *conclusive* or *full* answer. A reply that is not conclusive but does contribute some information toward satisfying the request is a *partial* answer. One of Hintikka's aims is to develop the logic of conclusiveness.

The conditions for conclusiveness may differ from one type of question to another. For example, the question

‘Who killed Julius Caesar?’

has

‘I know who killed Julius Caesar.’

as desideratum, and the reply

‘Cassius killed Julius Caesar.’

is a conclusive answer only if the questioner knows who Cassius is. Hintikka says that, for all simple questions (without intensional operators), if the desideratum has the form

$$(\exists x)KS(x),$$

then  $b$  is a conclusive answer to the question if and only if

$$(\exists x)K(b = x)$$

is established. Loosely speaking, conclusiveness requires that bound variables range over the questioner’s domain of acquaintance, and singular terms denote things in that domain. One consequence is that *conclusive answer* is not effectively recognizable from the syntactical form of the question.

For an introduction to Hintikka’s theory, see Hintikka [1976], [1983], and [1992], and Hintikka (ed.) [1988]. For further discussion and criticism of Hintikka, see Harrah [1979] and [1987]. For further studies of the epistemics and pragmatics of questions, see Kiefer (ed.) [1983] and Groenendijk and Stokhof [1984].

Hintikka and others have developed what has come to be called an *interrogative model of inquiry* that is intended to correspond to rational inquiry in science and other fields. In general, using the basic concepts of a given theory of questions one can formulate rules for rational question-and-answer procedures, including question-and-answer dialogues. In the case of Hintikka’s approach the idea is to choose rules whose rationale derives from his epistemic logic and game theoretic semantics.

One promising way of developing the semantics for this approach uses the concept of interrogative tableau. Such tableaux are formed from Beth-style deductive tableaux by adding interrogative rules — namely, rules that concern the formula that is being queried and the responses that might be given. Different sets of rules are possible, and correspondingly different systems of tableaux may be developed. For an introduction to this approach and the techniques involved, see Hintikka (ed.) [1988], Hintikka [1992], and Harris [1994].

## 6 OTHER APPROACHES

6.1 *MMB* ('*Make Me Believe*')

The MMK analysis of questions, and especially Hintikka's development of it, is really a family of different analyses. Family members share the reference to knowledge; family members may differ in their conception of knowledge. Consider the following (where  $B$  means 'I believe that'):

1.  $KP \rightarrow P$
2.  $KP \rightarrow KKP$
3.  $KP \rightarrow BP$
4.  $(KP \wedge [P \text{ implies } P']) \rightarrow KP'$

This set of assertions articulates a conception of knowledge that is appropriate for some normative models of knowledge and belief.

One might for various reasons hold that such a conception is too strong. One might decide to drop (4) and perhaps also (2). One might decide to drop knowledge altogether. Instead of using epistemic imperatives, one would use doxastic imperatives like

'Bring it about that I believe that  $A$  or I believe that  $B$ .'

Conclusive answers would bring sufficient evidence to produce stable firm belief. Obviously there is a spectrum of possible systems, corresponding to possible conceptions of belief and evidence. This spectrum awaits detailed study.

6.2 *TMT* ('*Tell Me Truly*')

D. and S. Lewis criticize Åqvist's MMK approach and argue that a more adequate theory results if one takes interrogatives to be synonymous with imperatives of the form 'Tell me truly ...' (Lewis [1975]). Adopting some of the Lewises' ideas, Åqvist [1983] outlines a way of developing this approach.

Åqvist assumes a three-place relation of *presentation* (sender  $X$  presents sentence  $S$  to receiver  $Y$ ). The language includes an 'empty' sentence representing silence, so that each  $X$  always presents some  $S$  to each  $Y$ . Then the question 'Is it the case that  $P$ ?' is to be analyzed as

Let it be the case in the immediate future that *either* there is a sentence  $S$  such that (i) you present  $S$  to me, (ii)  $S$  is true iff  $P$ , and (iii)  $S$  is true, *or* there is a sentence  $S$  such that (i) you present  $S$  to me, (ii)  $S$  is true iff not- $P$ , and (iii)  $S$  is true.

The question ‘For which time  $x$  is it the case that  $Fx$ ?’ is to be analyzed as:

Let it be the case in the immediate future that, for some time  $x$ , there is a sentence  $S$  such that (i) you present  $S$  to me, (ii)  $S$  is true iff  $Fx$ , and (iii)  $S$  is true.

The locution ‘ $S$  is true’ is to be analyzed as suggested in Kripke [1975]. A detailed working out of this approach remains to be undertaken.

### 6.3 *GMA (‘Give Me an Answer’)*

In some situations the questioner doesn’t need MMK or TMT. What will suffice is simply ‘Give me an answer.’ For arguments and examples, see Harrah [1987].

One way to develop this approach is to assume set theory and a predicate like Åqvist’s presentation predicate. Suppose that  $b$  is a term denoting a set  $D$  of sentences. Then, to ask the question whose direct answers are the members of  $D$ , we use ‘Let it be the case that, for some  $x$  in  $b$ , you present me with  $x$ ’.

Suppose we want GMACT (Give me an answer and claim that it is true’). One method is to add syntax (including the theory of the concatenation operator  $\hat{\ }^{\wedge}$ ) and a sentential operator  $\vdash$ , so that we have “... you present me with  $\vdash^{\wedge}x$ .”

The GMA and GMACT approaches, like TMT, await detailed study. The general problem is to determine which systems model the GMA or GMACT idea at some level of abstraction (see Belnap [1969], p. 122, concerning his own system). The specific problem is to develop systems in which the GMA or GMACT idea is directly expressed by interrogative-imperatives.

### 6.4 *Questions as Context Descriptions*

Hamblin discusses analyses like that of Jeffreys, in which an interrogative is taken to be synonymous with

‘I do not know ...; I want to know ...; and I think you know ...’

He says that even if such analyses can be made precise in a noncircular way (avoiding the phrase ‘know whether’), they nevertheless confuse two things that should always be kept distinct: (1) the description of the situation, and (2) the content of the question (Hamblin [1958]).

### 6.5 Questions as their Presuppositions

Another kind of analysis takes an interrogative to be synonymous with the declarative sentence that is (roughly) the presupposition in the sense of Belnap or Åqvist (see Section 4.9 and 5.4 above). In Harrah [1961, 1963a] a whether-question is an exclusive disjunction; its direct answers are the disjuncts. A which-question is an existential generalization (the existentially quantified variables being the queriables); the direct answers are substitution-instances of the quantified matrix. In Harrah [1961] a question is required to be true. In Harrah [1963a] a question may be true or false; if false, it is said to commit the fallacy of many questions.

The motivation for this sort of analysis is metalogical (to see what can be done within first-order languages) and technical, rather than empirical. The analysis becomes plausible for application if the question-and-answer situation is interpreted as an information-matching game. The questioner begins by making an assertion (e.g.  $(A_1 \vee A_2)$ ), and the respondent then replies by making another assertion (e.g.  $A_1$ ) that gives more information about the given subject matter.

There is of course no provision for ‘tagging’ sentences to indicate when they are being used to ask questions and when they are being used simply to make assertions. Thus the analysis is plausible where the communication situation can be interpreted as a question-and-answer situation and hence as an information-matching game, but is less plausible in wider contexts where this interpretation is not possible.

### 6.6 Questions as Intensional Entities

According to Tichý [1978], an interrogative expresses a question, and a question is an *office* — i.e. a function defined on possible worlds. The commonest types of question are propositions, individual concepts, and properties. These are functions whose values for a given world are a truth value, an individual, and a set of individuals, respectively.

To answer a question is to cite an entity of the right type (depending on the type of question); the answer is *right* if the entity is a value of the function at the actual world. A *complete* answer cites a single entity of the right type. An *incomplete* answer cites a class of entities of the right type. An incomplete answer is *correct* if the right complete answer is one of its members.

The motivation here is partly empirical (to account for intuitions about the informativeness of answers to questions), partly metalogical (to see what can be done within the logic of propositional entities), and partly philosophical. It is based on the assumption that logic is the study of logical objects or topics, rather than speakers’ concerns and attitudes. It assumes that the declarative-interrogative distinction is not one of logic, and that the logi-

cian does not have to provide distinct syntactic forms corresponding to the various distinct speech acts such as asserting, asking, and the like.

Higginbotham [1993] presents an intensional analysis that is motivated in large part by empirical-linguistic considerations. The aim is to discover the semantics of English interrogatives. The strategy is to develop a theory of questions as intensional entities, and then show how questions may be expressed by interrogatives.

For Higginbotham, an *elementary abstract question* is a nonempty partition  $\Pi$  of the possible states of nature into cells  $P$  such that no more than one cell corresponds to the true state of nature. A partition is *proper* if at least one cell must correspond to the true state of nature. (The elements of a cell may be thought of as statements; the cell corresponds to the true state of nature if all the statements in the cell are true.) An *answer* to a question  $\Pi$  is a set  $S$  of sentences that is inconsistent with one or more cells in  $\Pi$ . An answer is *proper* if it is consistent with at least one cell. A *partial answer* is one that is inconsistent with some (but not with all but one) of the cells.  $X$  is a *presupposition* of  $\Pi$  if every cell in  $\Pi$  implies  $X$ .

*Complex* abstract questions are constructed from elementary ones by quantification, conjunction, or disjunction; these form a hierarchy of orders, with abstract questions of order  $n$  being sets of abstract questions of order  $n - 1$ . Elementary abstract questions may be expressed by simple interrogatives and referred to by indirect-question phrases. Complex abstract questions may be expressed by various syntactical means. Most abstract questions are not expressed by any interrogative (there are too many of them).

Each interrogative may indicate partition of a limited universe, or partition of a limited part of a given universe.

‘Who did John see?’

has the form

$[WH\alpha : \text{person}(\alpha)]? \text{ John saw } \alpha$

where the quantification is restricted to persons. This interrogative expresses a partition whose cells describe the possibilities of John’s seeing (or not seeing) persons.

To account for multiple questions (see below), Higginbotham generalizes as follows. Where  $Q$  is a restricted quantifier,  $\varphi$  is the restriction on  $Q$ , and  $\Theta$  is another interrogative, an interrogative of the form

$[Qv : \varphi]\Theta$

expresses a question that is composed of sets of questions, one set for each way in which the quantifier, construed as a function from pairs of extensions to truth values, gives the value *true*. Consider



‘Where can I find two screwdrivers?’

This has the form

[Two  $x$  : screwdriver( $x$ )] [What  $\alpha$  : place( $\alpha$ )]  $x$  at  $\alpha$

and the question it expresses is the class of all classes of partitions each of which, for at least two screwdrivers  $a$  and  $b$  as values of  $x$  (and for no objects other than screwdrivers as values of  $x$ ), contains the partition for the interrogatives

[What  $\alpha$  : place( $\alpha$ )]  $a$  at  $\alpha$

[What  $\alpha$  : place( $\alpha$ )]  $b$  at  $\alpha$

Classes of partitions are *blocs*, and classes of them are *questions of order 1*. To *answer* a question of order 1 is to answer every question in one of its blocs.

On this approach the interrogative-question-answer relationship is not effective; one reason is that some English interrogatives are ambiguous. For example, where an interrogative seems to involve quantifiers, the semantics might not involve quantifiers and sets of questions but might instead involve a functional interpretation. E.g., to

‘Which of his poems does every poet like least?’

the answers that are wanted are not lists of poet-poem pairs but are replies like

‘His earliest poems.’

Concerning the various motivations for, and the rich potential of, the intensional approach, see Tichý [1978], Materna [1981], Belnap [1981], Higginbotham [1993], and Groenendijk and Stokhof [1984]. The latter work is rich in discussion and examples, and it argues that many theories can be articulated to provide a semantic equivalent of some Hintikka-type pragmatic dimensions.

(Incidental note: On many approaches, Hamblin’s Postulate (3) [recall 2.1 above] is false. On the intensional approach it might be true; see 7.2 below.)

### 6.7 Questions as Incomplete Entities

Several linguists have proposed theories in which the semantically meaningful unit (or at least the unit of truth) is not the question but the question-answer pair. One example is Keenan and Hull [1973]. Here the motivation is to account for our intuitions about the presuppositions of questions in

natural language — in particular, the intuitions that (1) questions have no truth value, but (2) questions have presuppositions, and (3) presuppositions are to be analyzed in terms of truth-value.

Keenan and Hull in effect identify questions with interrogatives. They define the class of *L-sentences* so that every declarative is an L-sentence, and, if  $Q$  is a question and  $A$  is a definite noun phrase, then  $\langle Q, A \rangle$  is an L-sentence.

For the case of wh-L-questions  $Q$ , which have the form (which,  $NP$ ,  $S$ ), we say:  $Q$  is *valid* in a state of affairs iff  $NP$  specifies a nonempty set (the *domain* of the question) and  $S$ , which is an L-sentence expressing the question property, is true of some members of the domain. The *answer set* determined by  $\langle Q, A \rangle$  is the set of objects denoted by  $A$  that are also in the domain. The L-sentence  $\langle Q, A \rangle$  is *true* in  $i$  iff (1)  $Q$  is valid in  $i$ , (2) the answer set determined by  $\langle Q, A \rangle$  is nonempty in  $i$ , and (3)  $S$  is true in  $i$  of every member of the answer set. Similarly,  $\langle Q, A \rangle$  is *false* in  $i$  iff (1) and (2) hold but  $S$  is false of some member of the answer set. In addition,  $\langle Q, A \rangle$  is *zero* in  $i$  iff it is neither true nor false in  $i$ .

On the basis of these definitions we can in an obvious way define consequence and presupposition. This definitional chain rests on (begins with) the definition of *valid*, and the concept of validity adopted here is similar to Belnap's concept of real truth. Thus it remains to be seen whether there is any formal advantage in using this framework rather than Belnap's or Åqvist's.

Hiz [1978] presents a development of the question-answer pair approach that is like Keenan and Hull's in some of its basic conceptions but differs considerably in details of superstructure. It remains to be seen whether there are any advantages from a formal point of view.

### 6.8 Questions as Hyper-complete Entities

According to an approach suggested by Finn [1974] (generalizing on Harrah [1963a]), a question may be treated as an ordered pair  $\langle A, B \rangle$  such that (roughly)  $B$  is an interrogative (or 'inquiry') term, and  $A$  is a statement that expresses the presuppositions of the question. In general this approach is motivated by technical and engineering considerations; and, if the context of application is narrowly conceived, then  $A$  and  $B$  can be narrowly conceived. This idea leads, however, to the following generalization. If the context of application is a relatively general type of problem-solving or inquiry situation, then  $A$  and  $B$  can be conceived in a relatively general way. Among the interrogative terms  $B$  there may be noun phrases like 'an  $x$  such that ...'; this phrase indicates what sort of entity is to be found. Among the presupposition statements  $A$  there may be any statements giving background information or imposing constraints on the search.

6.9 IQW (*‘It Is the Question Whether’*)

Hoepelman [1983] proposes to read the propositional operator ? as

‘It is the question whether’

Cf. the German

‘Es ist die Frage ob’

and in English

‘[For me] there is the question whether’

‘[For me] it is an open question whether’

The basic assumption is that

‘Is it the case that  $p$ ?’

is a question for me not when  $p$  has a truth value for me but when its truth value is still undetermined for me. Hoepelman develops a truth value analysis of interrogatives to serve an empirical motivation — in general, to account for interrogatives in natural language, and in particular to account for distinctions that reflect differences in the questioner’s certainty, as in the following pairs:

‘Is John ill?’

‘Isn’t John ill?’

Inter alia, Hoepelman adopts the following truth tables:

	$\neg p$	$(p \rightarrow q)$	$(p \leftrightarrow q)$	$?p$
q	...	11 10 01 00	11 10 01 00	...
p				
11	00	11 10 01 00	11 10 01 00	00
10	01	11 11 01 01	10 11 00 01	00
01	10	11 10 11 10	01 00 11 10	10
00	11	11 11 11 11	00 01 10 11	00

The idea here is to articulate the questioner’s certainty in terms of a comparison of two worlds — the world known by the questioner and the world known by the authority to whom the question is to be put. In the truth tables each pair of numbers represents a comparison; think of the first number as the questioner’s certainty about the given statement, and the second number as what the questioner believes is the authority’s certainty. According to these truth tables, the following are valid:

- $\neg?(p \rightarrow p)$
- $? \neg p \rightarrow \neg?p$
- $?p \rightarrow ? \neg?p$
- $?(p \rightarrow q) \rightarrow (?p \rightarrow?q)$
- $?p \rightarrow (p \rightarrow q)$

The following are not valid:

$$\begin{aligned} ?p &\rightarrow ?\neg p \\ (p \leftrightarrow q) &\rightarrow (?p \leftrightarrow ?q) \end{aligned}$$

To accommodate predicate logic and wh-questions Hoepelman extends the propositional apparatus by assuming a pair of models (each with its domain of individuals and assignment of denotations), incorporating the propositional truth value conditions, and then adding truth value conditions for the two quantifiers and wffs with free variables. Loosely speaking:

1.  $? \varphi x = 10$  if  $\varphi x = 01$ , and  $? \varphi x = 00$  otherwise.
2. The value of  $\forall x \varphi x$  is the minimum of the values for  $\varphi x$ .
3. The value of  $\exists x \varphi x$  is the maximum of the values for  $\varphi x$ .

If the two domains of individuals are identical, then the following are valid:

$$\begin{aligned} \forall x ? \varphi &\rightarrow ? \forall x \varphi \\ ? \exists x \varphi &\rightarrow \exists x ? \varphi \end{aligned}$$

If we add  $=$  as a predicate, then

$$\forall x \forall y \forall z (x = y \wedge ?(y = z) \rightarrow ?(x = z))$$

is valid, but

$$\forall x \forall y (? (x = y) \wedge P x \rightarrow ? P y)$$

is not.

For some readers it is an open question whether all of the validity results noted above accord well enough with intuition; but the assertion that this sort of approach is interesting and worth exploring further is not in question.

## 7 OTHER TOPICS AND FURTHER WORK

### *7.1 Other Types of Question*

Some logicians have been concerned to theorize in a global way about questions in general, setting aside or de-emphasizing the problem of distinguishing and analyzing particular types of question (e.g. Tichý [1978], and radical SA reduction). Most logicians, however, have concentrated on distinguishing and analyzing question types, one at a time; and most have concentrated on whether- and which-questions.

Little progress was made on who, what, and why questions until the 1970s. See Belnap and Steel [1976], pp. 78 – 87. Hintikka, beginning with Hintikka [1976], has made an aggressive attack on who questions. See also,

e.g., Grewendorf [1983]. Since the 1970s many logicians and linguists have studied the various types of wh-question found in natural languages, but everyone would agree that much more work needs to be done. In particular, why questions pose a special challenge.

Bromberger [1992] contains illuminating and suggestive discussion of why questions, and proposals concerning one important type. This type is exemplified by:

‘Why is it the case that  $X$  has property  $Y$  (instead of property  $Z$ )?’

Bromberger’s conception is that normally, for such questions, the questioner has had in mind a general rule  $R$  that represents some expectation  $E$  about  $X$  but the questioner fails to observe  $E$  and hence asks the question. The respondent answers the question by citing (1) an abnormic law  $L$  that specifies exceptions to the rule  $R$ , and (2) one or more exceptions that are specified by  $L$ . E.g., expecting the milk to taste good, the child asks

‘Why does this milk taste sour?’

and is told

‘All milk tastes good unless it is spoiled or adulterated, and this milk is spoiled.’

To formulate this proposal in a way that is general, precise, and not subject to counterexamples is a task that is not yet complete; see Bromberger [1992], pp. 88 – 97. On the difficulties of accommodating a Bromberger-style analysis within an extensional framework see Belnap and Steel [1976], pp. 84 – 87.

Koura [1988] outlines a Hintikka-style analysis of one family of why questions. In these the explanandum is the occurrence of event  $e$ , and the question in effect is

‘Which event caused the event  $e$ ?’

Different subtypes correspond to different types of causation. For example, let

$$E(x) =_{df} \exists y(y = x)$$

and let  $N$  be a primitive signifying nomical necessity. Then

$$N[E(x) \rightarrow E(y)]$$

expresses that  $x$  is a possible (sufficient) cause of  $y$ , and to say that  $x$  causes  $y$  we use

$$E(x) \wedge N[E(x) \rightarrow E(y)].$$

Following Hintikka’s approach the desideratum of

‘What caused  $e$ ?’

is

$$\exists xK[E(x) \wedge N[E(x) \rightarrow E(e)]].$$

Koura shows that a reply  $f$  is an adequate answer — i.e., brings about the desideratum — if and only if the conclusiveness conditions

$$\begin{aligned} &KN[E(f) \rightarrow E(e)] \text{ (relevance)} \\ &\exists xKN(x = f) \text{ (uniqueness)} \end{aligned}$$

both hold. Koura discusses other concepts of cause and suggests that some types of why question might involve pragmatic parameters.

Hintikka and Halonen [1995] (hereafter ‘H&H’) rejects this approach and says in particular that no modal element is needed for why questions per se. The gist of the H&H proposal is as follows. Consider

‘Why does  $b$  have property  $P$ ?’

Suppose we have a sentence  $T$  (e.g., a general theory) and a sentence  $A$  (e.g., some additional ad hoc information supplied by an oracle) such that

1.  $(T \wedge A) \vdash P(b)$
2. not  $T \vdash P(b)$
3. not  $A \vdash P(b)$
4.  $b$  does not occur in  $T$
5.  $P$  does not occur in  $A$

where ‘ $\vdash$ ’ indicates derivability in a Hintikka-style interrogation game based on first-order logic. Then by Craig’s Interpolation Theorem it follows that there is a formula  $H[b]$  such that

1.  $T \vdash (\forall x)(H[x] \rightarrow P(x))$
2. All the constants in  $H$ , except for  $b$ , occur in both  $T$  and  $A$ .
3.  $A \vdash H[b]$

Call  $H[b]$  the *initial condition* and  $(\forall x)(H[x] \rightarrow P(x))$  the *covering law*. The proposal is that we answer the given why question by citing the initial condition or the covering law or both. The claim is that these answers are conclusive for anyone who believes that  $(T \wedge A)$  is true and sees that  $P(b)$  is derivable from  $(T \wedge A)$ .

(Comment: Those who are not committed to Hintikka-style interrogation games or criteria of conclusiveness may wish to say simply that to ‘Why

$P(b)$ ?' the direct answers are all sentences of the form  $(H[b] \wedge (\forall x)(H[x] \rightarrow P(x)))$ , where  $H[x]$  is restricted in certain ways to exclude cases that are trivial or unacceptable for other reasons. The restrictions imposed by H&H are probably minimal; in practice, users of why interrogatives usually want to put further restrictions on the size or content of  $H[x]$ . This matter of restrictions on  $T$ ,  $A$ , and  $H$  deserves much more study.)

Concerning how questions: H&H suggests briefly that, if  $b$  is allowed to occur in  $T$ , then no covering law is obtainable in general and deriving  $P(b)$  from  $(T \wedge A)$  seems more like answering a how question than answering a why question. This suggestion deserves to be clarified and studied in detail.

Another type of question awaiting further study is the deliberative question, exemplified by,

'What shall I do?'

These seem to call for a decision or resolution (Wheatley [1955]). They might be analyzed as genuine questions, but of a type peculiar to one-person decision-making situations. Alternatively they might be analyzed as having properties appropriate to the usual two-person question-and-answer situation, plus further pragmatic properties as well (so that they call for both an answer and a resolution). Alternatively, as suggested by Mayo [1956], they might be interpreted as calling for an imperative ('Do  $X$ !').

## 7.2 Other Types of Reply

As noted in Section 2.1, Hamblin's Postulate (1) has not been universally accepted. In the first place, as already noted, some logicians and many linguists have argued for allowing noun phrases to count as direct answers. There is much unfinished business here, e.g. no one has yet provided a comprehensive formal system for connecting noun phrases with the system of completeness-claims and distinctness-claims. Natural language seems to have at least a rudimentary system; e.g. to questions like

'Who were all the students who passed?'

we give answers like

'Only two: Shane and Mark.'

The problems for research are: What exactly is the system in natural language? and How can we formalize analogs for artificial languages?

In the second place, we might accept Hamblin's assumption that every direct answer is a sentence, but question whether it must be a statement. The suggestion of Mayo (see Section 7.1) that imperatives like 'Do  $X$ !' should count as direct answers presumably applies to deliberative questions and their close relatives. In the case of most other kinds of question, however, it often seems appropriate to reply with certain kinds of imperative, as in

‘Ask someone who knows the Lexitron!’

or interrogative, as in

‘How detailed an answer do you want?’

Also, of course, there are nonanswer declaratives like

‘I don’t know’ and

‘That’s a long story’

It might be argued that some of these nonanswer sentences are evasions (or possibly corrections) of the given question, but some should count as *illuminations* or at least *helps*. In any case the field awaits formalization. For a beginning, on ‘I don’t know’ replies at least, see Todt and Schmidt-Radefeldt [1979], p. 15.

There is much work to be done on the general topic of appropriate response, and in particular on the topic of finding or constructing an appropriate response. See, e.g., Lehnert [1978].

The notions of incomplete answer and partial answer invite further study. In the usual conception a partial answer is one that is implied by a direct answer. Belnap and Steel [1976] emphasize that this conception is relative to the type of implication assumed, and that we may refine the type of partial answerhood by refining the type of implication. For other problems and other perspectives on partial answerhood, see Cresswell [1965], Kubiński [1967], Groenendijk and Stokhof [1984], pp. 233 – 236, Higginbotham [1993], and Wiśniewski [1995], pp. 114 – 115, 179 – 180.

Finally, there is the challenge of Hamblin’s Postulate (3). If it is interpreted as saying that, for every question, exactly one answer is true, then it seems obviously false. Is there some other plausible interpretation under which it is true? Consider Higginbotham’s concept of proper partition (noted in 6.6 above), in which exactly one cell corresponds to the true state of nature. This might provide an intensional-analytic rationale for Postulate (3). On this and related matters, see Groenendijk and Stokhof [1984].

### 7.3 *Implying, Raising, and Suppressing*

Hamblin’s *containment* (recall 2.1 above) may be thought of as a kind of implication between questions. When that relation holds, each answer to the first question implies an answer to the second, so the first covers the second, so, if you ask the first, you need not ask the second. Belnap’s *propositional implication* (recall 4.9) is a simple generalization from implication between statements to, loosely speaking, implication between statements and questions. When that relation holds, truth of the implying statements and questions guarantees truth of the implied statements and questions; so,



if you know that the first question is good, you know that the second is good.

Wiśniewski [1994a and 1995] defines a cluster of concepts that are less like the Hamblin and Belnap concepts and more like the informal notion of *raising*, as in

- ‘Your statement raises some hard questions.’  
 ‘Your question raises a more basic question.’

Loosely speaking, Wiśniewski’s concept of implication expresses the idea that, if you need to ask the first question, then you would be well advised to ask the second also. The remainder of this section, except for the final three paragraphs, summarizes part of Wiśniewski’s analysis.

Wiśniewski defines his concepts for a very wide class of languages. This class includes more than the usual first-order languages, but it will be helpful here for the reader to think of a first-order language with the usual extensional semantics. Wiśniewski uses ‘d-wff’ for ‘declarative well-formed formula’,  $Q$  for questions, and  $dQ$  for the set of direct answers to  $Q$ . Some interpretations are distinguished as *normal*. A set  $X$  *entails* a d-wff  $A$  iff  $A$  is true in every normal interpretation in which all the d-wffs in  $X$  are true.  $X$  *logically entails*  $A$  iff  $A$  is true in every interpretation in which all the d-wffs in  $X$  are true.  $X$  *multiple-conclusion entails*  $Y$  (or,  $X$  *mc-entails*  $Y$ ) iff, for each normal interpretation  $I$  where all the d-wffs in  $X$  are true, at least one d-wff in  $Y$  is true in  $I$ .

Assumptions about questions:

1. Each question has at least two direct answers.
2. Direct answers are sentences (*d-wffs* without free variables). (Hence each  $dQ$  is at most denumerable.)
3. Each set of sentences that is finite and has at least two members is the  $dQ$  for some  $Q$ .

$Q$  is *sound in* an interpretation  $I$  iff some  $A$  in  $dQ$  is true in  $I$ .  $Q$  is *safe* iff  $Q$  is sound in every normal  $I$ .  $Q$  is *sound relative to*  $X$  iff  $X$  mc-entails  $dQ$ .  $Q$  is *risky* iff  $Q$  is not safe.

$X$  *evokes*  $Q$  iff (i)  $X$  mc-entails  $dQ$ , (ii) for each  $A$  in  $dQ$ ,  $X$  does not entail  $A$ .

$X$  *generates*  $Q$  iff  $X$  evokes  $Q$  and  $Q$  is risky.

$Q$  *implies*  $Q'$  *on the basis of* a set  $X$  of d-wffs [or  $\text{Im}(Q, X, Q')$ ] iff (i) for each  $A$  in  $dQ$ :  $X + A$  mc-entails  $dQ'$ , and (ii) for each  $B$  in  $dQ'$ : there is a nonempty proper subset  $Y$  of  $dQ$  such that  $X + B$  mc-entails  $Y$ . (Roughly, the implying question  $Q$  raises the implied  $Q'$  because  $Q'$  helps to answer  $Q$ ;  $Q'$  helps because each  $B$  in  $dQ'$  directs attention to a proper subset of  $dQ$ .)  $Q$  *implies*  $Q'$  [or  $\text{Im}(Q, Q')$ ] iff  $\text{Im}(Q, \Lambda, Q')$ .

It is not the case that every safe question is implied by every  $Q$  on the basis of every  $X$ ; safety guarantees that clause (i) holds but does not guarantee that clause (ii) holds. On the other hand, if  $\text{Im}(Q, Q')$ , then  $Q'$  is safe iff  $Q$  is safe.

**THEOREM 3.** *If  $Q$  is sound relative to  $X$ , then there exists a sequence  $Z$  of simple yes-no questions (i.e., questions of the form  $\{A, \neg A\}$ ) such that:*

1. *each question in  $Z$  is implied by  $Q$  on the basis of  $X$ ,*
2. *each set consisting of direct answers to the questions in  $Z$  that contains exactly one direct answer to each question in  $Z$  entails along with  $X$  some  $A$  in  $dQ$ , and*
3. *each nonlogical constant that occurs in some direct answer to a question in  $Z$  occurs in some  $A$  in  $dQ$ .*

(Note: If  $dQ$  is finite, then  $Z$  can be finite.)

Let  $X$  be a finite nonempty set of d-wffs. Then the pair  $\langle X, Q \rangle$  is an  $e_1$ -argument, and is *valid* iff  $X$  evokes  $Q$ . The triple  $\langle Q, X, Q' \rangle$  is an  $e_2$ -argument, and is *valid* iff  $Q$  implies  $Q'$  on the basis of  $X$ . For any given language  $L$  having d-wffs and questions, we can construct a metalanguage  $ML$  that has statements asserting that evocation and implication hold between particular  $X$ 's and  $Q$ 's. Then, for the given  $L$ , the set of all of these  $ML$  statements that are true can be regarded as the logic of questions of  $L$ . (The basic idea for this conception was suggested by Kubiński.)

In addition to establishing his general framework as outlined above, Wiśniewski studies several particular languages of the usual kinds (propositional, first-order, etc.) and gives many results and examples for these. Many of these results and examples are for the case where everything has a name — i.e., where every entity in the universe of the assumed interpretation  $I$  is denoted by some closed term in the language (under  $I$ ). A problem for future research is to explore in detail the cases in which this is not so — i.e., where there are ‘real answers’ (in Belnap’s sense) that are not expressed by nominal answers.

Another area that awaits development is the case of questions in logic and mathematics, questions whose direct answers are logical or normal truths. Wiśniewski’s definition of evocation is tailored to fit the case of factual questions, where no direct answer is normally or logically true. The challenge is to extend the analysis to capture the concept of evocation for the wider class of questions.

Questions can be raised. Can questions be suppressed? Is suppression a dual of evocation? Is it sufficient to say simply that  $X$  *suppresses*  $Q$  iff  $X$  entails  $\neg A$  for all  $A$  in  $dQ$ , and that  $Q'$  *suppresses*  $Q$  iff every  $B$  in  $dQ'$  suppresses  $Q$ ? Or are there other conditions that are sufficient for suppression? Why don’t we know more about suppression? Don’t ask.

### 7.4 Reduction of Questions

As noted in earlier sections, some theorists hold that questions are reducible to entities of other kinds. E.g., some hold that each interrogative is semantically or pragmatically equivalent to an imperative sentence. We might say that this sort of reduction is *inter-categorial*. In contrast, as common speech has long recognized, there is what we might call *intra-categorial* reduction, where one interrogative is construed as equivalent to some other interrogative, as indicated by locutions like

‘Let me rephrase my question; what I am really asking is . . .’

There are several problems of interest to logicians. One is to develop precise concepts for the general notions of equivalence and reduction, another is to find techniques for demonstrating reducibility, and another is to establish particular results.

Wiśniewski [1994b and 1995] has suggested some general concepts and established some results. The basic definition is this: [Concerning notation, see 7.3 above.] A question  $Q$  is *reducible to* a nonempty set  $S$  of questions iff:

1. for each  $A$  in  $dQ$ , for each question  $Q'$  in  $S$ ,  $A$  mc-entails  $dQ'$ ,
2. each set consisting of direct answers to questions in  $S$  that contains exactly one direct answer to each question in  $S$  entails some  $A$  in  $dQ$ , and
3. no question in  $S$  has more direct answers than  $Q$ .

Some of Wiśniewski’s theorems are these: [Concerning terminology see 7.3.]

1. A question  $Q$  is safe iff  $Q$  is reducible to some set of simple yes-no questions that are implied by  $Q$ .
2. If  $dQ$  is finite, then  $Q$  is safe iff  $Q$  is reducible to some finite set of simple yes-no questions that are implied by  $Q$ .
3. If  $Q$  is risky but  $dQ$  is finite, then  $Q$  is reducible to a finite set of conditional yes-no questions [i.e., questions of the form  $\{A \wedge B, A \wedge \neg B\}$ ] that are implied by  $Q$ .
4. If  $Q$  is risky but there is a d-wff  $B$  such that (i)  $B$  is entailed by every  $A$  in  $dQ$  and (ii)  $B$  mc-entails  $dQ$ , then  $Q$  is reducible to some set of conditional yes-no questions that are implied by  $Q$ .

Wiśniewski proves these and other theorems using straightforward model-theoretic arguments, and it might seem that such arguments are the only

means available for establishing reducibility results. Without making any claims, we conjecture that there might also be exotic methods available, perhaps different methods for different types of question. We mention two examples:

The first is the set of techniques suggested in the paper of Todt and Schmidt-Radefeldt [1979]. Among the primitive signs of the language adopted are  $\top$  (for the true) and  $\perp$  (for the false). This allows interrogatives of the form

$$?v^b(v^b \leftrightarrow A)$$

which literally say

‘Which Boolean truth-value is equivalent to the statement  $A$ ?’

or colloquially

‘Is it the case that  $A$ ?’.

It remains to be seen how much can be done with this sort of apparatus, and in general what its advantages are.

The second is the methodology suggested by Leszko [1980]. According to that work certain types of question can be represented via graphs. (Leszko concentrates on the Kubiński questions noted in Section 4.12 above, questions like ‘For which  $n$   $x$ ’s and  $m$   $y$ ’s is it the case that ...?’.) Once a question type has been represented via graphs, we can study the questions by studying the matrices associated with the graphs. See also Leszko [forthcoming].

### 7.5 *Sequencing and Programming*

The general problems here are to evaluate sequences of questions with respect to their answer-yield, their safety, or other properties, and to compare sequences with respect to various concepts of containment, implication, and equivalence. One special class of problems concerns question trees, including the case of trees in which the nodes represent questions and the branches represent answers. Such a question tree can be used to represent a strategy or plan for asking questions one at a time, where at any time the choice of question depends on what previous questions have been asked and what answers have been given. Some problems for study are:

Under what conditions does a question tree represent a safe plan?

Under what conditions are two trees equivalent?

Under what conditions is a tree equivalent to a single question?

Not much has been done on these matters in general, but a few useful concepts have been developed. In Belnap and Steel [1976], p. 138, a sequence of interrogatives  $(I_1, \dots, I_n)$  is said to be a *direct partition* of an interrogative  $I$  if the conjunctive interrogative  $(I_1 \wedge \dots \wedge I_n)$  is erotetically equivalent to  $I$ , and a *subdirect partition* if  $(I_1 \wedge \dots \wedge I_n)$  is erotetically equivalent to  $(I \wedge \dots \wedge I)$ , i.e.  $I$  conjoined  $n$  times.

One special kind of sequence that has received some attention is the ‘corrections-accumulating’ sequence. The basic idea was presented in Stahl [1962]. Consider the three questions:

1. Which are the two primes between 13 and 17?
2. Which are at least two primes between 13 and 17?
3. Is there any prime between 13 and 17, or not?

Stahl pointed out that in natural language we occasionally put the sequence 1-2-3 by saying ‘1, or 2, or 3,’ where the *or* is understood to be noncommutative.

Stahl generalized and formalized as follows [recall Section 2.2 above]: An *inferential question series* is a series in which the first question is relative to some  $S$  and the  $n + 1$ st question is relative to  $S \cup \{A_n\}$ , where  $A_n$  is a sufficient answer to the  $n$ th question,  $A_n$  is not a consequence of  $S$ , and all the sufficient answers to the  $n$ th question which neither imply  $A_n$  nor are consequences of  $S$  are not compatible with  $A_n$ . A *sufficient answer of the  $n$ th degree* is a wff which either is a consequence of  $S$  or implies a conjunction  $(A_1 \wedge A_2 \wedge \dots \wedge A_{n-1} \wedge B)$  which is consistent with  $S$ , where  $B$  is a sufficient answer to the  $n$ th question which does not imply  $A_n$ . The intention is to yield the theorem: A sufficient answer of  $n$ th degree is incompatible with sufficient answers of lower degree, unless these are consequences of  $S$ .

Åqvist develops this idea within his framework as follows. Let  $Q$  be a QIE-question, and let  $\{p_1, \dots, p_n\}$  be a finite set of ordinary statements. Define  $Q[p_1, \dots, p_n]$  as  $(p_1 \wedge \dots \wedge p_n \wedge Q)$ , define its Core as  $(p_1 \wedge \dots \wedge p_n \wedge \text{Core}Q)$ , and define its Pres as  $(p_1 \wedge \dots \wedge p_n \wedge \text{Pres}Q)$ . Recall that  $\text{Correc}X = \neg \text{Pres}X$ .

Now let  $S = \{Q_1, \dots, Q_n\}$  be any finite set of QIE-questions. Form out of  $S$  the  $n!$  distinct  $n$ -termed sequences such that each member of  $S$  occurs exactly once in the sequence. Next, name these sequences and arrange them in some fixed order:

$$S_1 = \langle Q_{1_1}, \dots, Q_{n_1} \rangle, S_2 = \langle Q_{1_2}, \dots, Q_{n_2} \rangle, \dots, \\ S_{n!} = \langle Q_{1_{n!}}, \dots, Q_{n_{n!}} \rangle.$$

Then, for each  $S_j$ , define the *simplest corrections-accumulating sequence associated with  $S_j$*  (or  $\text{sca}(S_j)$  for short) as

$$\langle Q_{1_j}, Q_{2_j}[\text{Correc}Q_{1_j}], \dots, \dots \rangle$$

$$Q_{n_j}[\text{Correc}Q_{1_j}, \text{Correc}Q_{2_j}, \dots, \text{Correc}Q_{(n-1)_j}].$$

Finally, for each  $S_j$ , let its QIE-translation = the QIE-translation of  $\text{sca}(S_j) =$

$$\begin{aligned} &!(\text{Core}Q_{1_j} \vee \text{Core}(Q_{2_j}[\text{Correc}Q_{1_j}]) \vee \dots \vee \\ &\text{Core}(Q_{n_j}[\text{Correc}Q_{1_j}, \dots, \text{Correc}Q_{(n-1)_j}])). \end{aligned}$$

Let  $S = \langle Q_1, \dots, Q_n \rangle$  be a question sequence. We say that  $S$  is *corrections-accumulable* iff, for all  $1 \leq i < j \leq n$ ,  $\text{Pres}Q_j$  does not entail  $\text{Pres}Q_i$ . It is *successively presupposition-containing* iff, for all  $1 \leq i < j \leq n$ ,  $\text{Pres}Q_i$  entails  $\text{Pres}Q_j$ . It is *quite reasonable* iff all the disjuncts inside the ! in the QIE-translation of  $\text{sca}(S)$  are consistent.

It turns out that, for each finite question-set  $S = \{Q_1, \dots, Q_n\}$ , there is at most one sequence  $S_j$  ( $1 \leq j \leq n!$ ) formable out of  $S$  that is both corrections-accumulable and successively presupposition-containing. In the example above, the sequence would be  $\langle 1, 2, 3 \rangle$ . Also, for a sequence to be quite reasonable it is necessary that it be corrections-accumulable.

One other fact: A corrections-accumulable sequence can have a safe question  $Q$  (i.e. with valid  $\text{Pres}Q$ ) only in its final position. (For all of the above, see Åqvist [1969].)

Similar concepts and constructions can be developed in Belnap's framework by using his conditional and given-that questions. We form the appropriate sequence of interrogatives and then construct their union. See Belnap [1969].

Picard [1980] studies question sequences in the light of practical considerations like probability, cost, and utility. The general problem is how to replace a single complex and costly question  $Q$  by a questionnaire  $Q'$  (which is a sequence of simple which and whether questions), such that  $Q'$  will yield the true answer to  $Q$  but asking  $Q'$  will be more efficient and economical than asking  $Q$ . Questionnaires are represented as weighted finite circuitless graphs meeting certain conditions. (Think of a questionnaire as a bush — starting from one node — whose non-terminal nodes are questions and whose branches are the direct answers.) Each answer is assigned a probability. Answers and questions can be assigned utilities and costs. For more in this area see, e.g., Kampé de Fériet and Picard (eds.) [1974].

## 7.6 Comparison of Approaches

Is there a correct approach to the theory and logic of questions? It is not yet clear that we have a correct approach to clarifying and answering that question; especially, it is not yet clear that we have a correct criterion for recognizing a correct answer.

What is clear is that much more work remains to be done in comparing different approaches and evaluating their respective advantages. A little

work of this sort has been done, but most of it has been done on small aspects and points of detail. What we offer below are some rough surmises. We don't claim that these are accurate or correct; we do hope that they are clear enough to stimulate further study.

1. The systems that can provide the most questions are those that assume questions as metaphysical or intensional entities — as in the theories of Tichý [1978] or Higginbotham [1993].
2. Is one approach better than others at providing interrogatives? There is some appeal in the idea noted in 6.8 above, that we can adopt a very rich language and thus have interrogatives that specify a detailed description of what is wanted and how to search for it. On the other hand, interrogatives are instruments for communication, and our choice of interrogative system is influenced by the purposes at hand. Thus it is meaningful or useful to compare interrogative systems relative to specific motivations (e.g. the motivation to model the question-and-answer system of the Danish people, or the motivation to construct an information-retrieval system for the Yale Medical School Library), but not useful to make comparisons otherwise.
3. The systems that would be most useful in machine-assisted interactions, and especially in formal systems of information retrieval, will have the effectiveness properties emphasized in Belnap and Steel [1976].
4. For empirical models of the question-and-answer process in natural language several different kinds of system will be needed, including not only those of the MMK approach (designed to fit the 'standard' situation, as noted in 5.1 above) but also others (designed to fit other types of situation).
5. Theorizing about questions requires theorizing about interrogatives. We can be confident that an intensional theory of questions is complete and correct only if we are confident that we have a complete and correct theory of human concepts and intentions, and we can be confident of the latter only if we are confident that we have a complete and correct theory of human language, including a complete and correct theory of interrogatives.

### *7.7 General Erotetic Logic: Motivations*

There are several motivations for generalizing from erotetic logic in the narrow sense, concerned with question and answer, to erotetic logic in a broader sense, concerned with all the kinds of expression that call for reply.

First consider mixed sentences — e.g.:

1. ‘The old machine is broken, or does it need fuel?’
2. ‘The new one is missing, but do we need it?’
3. ‘What is wrong with the old one? or find the new one.’

To provide for such sentences we want a logic that will specify for declaratives, imperatives, interrogatives, . . . (?) what compounds are permissible, what expressions call for replies, and what replies are called for.

Second, consider vectored sentences such as

‘As Provost, I ask you, Dean Smith, will the plan be approved?’

This may be construed as a sentence that, at some level of analysis, consists of two parts:

1. the *body* (‘Will the plan be approved?’), and
2. the *vector* (indicating that the message comes from ‘I’ *qua* Provost and is for Smith *qua* Dean).

For discussions of vectored sentences see Harrah [1994]. To provide for such sentences we want a logic that will specify what expressions count as vectored sentences, and what expressions (vectored or unvectored) count as replies.

Third, consider vectored messages such as the formal memos used in large organizations and the formal letters used in commercial and legal correspondence. These have a vector that specifies a *to*, a *from*, a *when*, and possibly other parameters, and a body that may contain any number of sentences of various kinds. For such messages we want a logic that specifies what expressions count as messages and, for each message, what counts as a sufficient reply.

In 7.8 below we outline a way of developing a logic that provides for mixed sentences, vectored sentences, and vectored messages. This logic may be viewed as a system of general erotetic logic, and our sketch of it should serve to indicate what a general erotetic logic is. From another perspective it may be viewed as a logic of message and reply, or a *communicational logic*. Perhaps the concepts of general erotetic logic and communicational logic coincide, for in both cases the essential concern is with (a) a set of expressions and (b) for each expression, its set of sufficient replies.

### 7.8 *General Erotetic Logic: Systems*

In this section we outline a way of developing a particular system of general erotetic logic. It will be clear that we are in effect describing a class of systems, and indeed a fairly wide approach. We don’t claim that this is the only correct or fruitful approach. The motivation for this approach is



both empirical and engineering; the aim is to construct systems that will be useful in connection with human communication. (For more on motivation, empirical grounding, and details of development, see Harrah [1985, 1987, 1994].) In the paragraphs below we first describe the part of the system that handles unvectored sentences, and then the part for vectored sentences and vectored messages.

We begin with a standard first-order system having identity, descriptions, and some nonlogical axioms for set theory and syntax. We write  $U$  and  $E$  for its quantifiers and  $F, G, H, \dots$  for its wffs, which we call  $d$ -wffs. We add an infinite stock of *speech act operators*  $O, O', O'', \dots$ . A *basic speech act wff* (or *bsa-wff*) is an expression  $OVY$  such that  $O$  is a speech act operator,  $V$  is a (possibly empty) string of distinct variables, and  $Y$  is a list  $(Y_1, \dots, Y_n)$  in which each  $Y_i$  is a term or  $d$ -wff.

The *basic wffs* (or *b-wffs*) are the  $d$ -wffs and *bsa-wffs*. To every  $b$ -wff  $F$  we assign a  $d$ -wff  $CA$ , a  $d$ -wff  $CP$ , a set  $IR$ , and a set  $WR$ , such that:

1.  $IR$  consists of  $d$ -wffs (the *indicated replies* to  $F$ ).
2.  $CA$  (the *core assertion* in  $F$ ) is implied by every  $d$ -wff in  $IR$ .
3.  $WR$  (the *wanted replies* to  $F$ ) is a subset of  $IR$ .
4.  $CP$  (the *core projection* in  $F$ ) is implied by every  $d$ -wff in  $WR$ .
5.  $CA$  is implied by  $CP$ .

Given a *bsa-wff*  $F$ , with its  $CA$  and  $CP$ , we say that:

The *negative reply* to  $F$  is  $\neg CP$ .

The *corrective reply* to  $F$  is  $\neg CA$ .

The *direct replies* to  $F$  are:

1. the wanted replies, and  $\neg CP$ , if  $WR$  is nonempty;
2. the indicated replies, and  $\neg CA$ , if  $WR$  is empty but  $IR$  is not;
3.  $CP$ ,  $(CA \wedge \neg CP)$ , and  $\neg CA$ , if  $IR$  is empty.

The *full replies* to  $F$  are the  $d$ -wffs that imply direct replies.

The *partial replies* to  $F$  are the  $d$ -wffs that are implied by direct replies.

The *relevant replies* to  $F$  are the full replies plus the partial replies.

To give examples, we use the signs

$!w, !c, !d, .d, .as, .an, ?w, ?!$

to refer to distinct speech act operators (expressing respectively ultimatum, command, directive, declaration, assertion, announcement, whether-question, one-example question). To the eight kinds of *bsa*-wffs at the left below, content may be assigned as follows:

	<i>CA</i>	<i>CP</i>	<i>IR</i>	<i>WR</i>
$!^u(G)$	$(G \vee G')$	$G$	$\{G\}$	$\{G\}$
$!^c(G)$	$(G \vee G')$	$G$	$\{G\}$	$\Lambda$
$!^d(G)$	$(G \vee G')$	$G$	$\Lambda$	$\Lambda$
$:^d(G)$	$G$	$G$	$\{G\}$	$\{G\}$
$:^{as}(G)$	$G$	$G$	$\{G\}$	$\Lambda$
$:^{an}(G)$	$G$	$G$	$\Lambda$	$\Lambda$
$?^w(G, G')$	$(G \vee G')$	$(G \vee G')$	$\{G, G'\}$	$\{G, G'\}$
$?^1x(Gx)$	$ExGx$	$ExGx$	$\{Ga, \dots\}$	$\{Ga, \dots\}$

For smoothness, in the case of each *d*-wff  $F$ , we say that  $CP(F) = CA(F) = F$ , and the  $WR(F) = IR(F) = \Lambda$ . (Note: occasionally, as here, we use ‘*CA*’, . . . , ‘*WR*’ as functors.)

An *erotetic wff* (or *e*-wff) is a *bsa*-wff  $F$  such that  $WR(F) \neq \Lambda$ . The *speech act wffs* (or *sa*-wffs) are defined recursively:

1. Every *b*-wff is an *sa*-wff.
2. If  $F$  and  $G$  are *sa*-wffs and  $x$  is a variable, then  $(F \wedge G)$ ,  $(F \vee G)$ ,  $Ux F$ ,  $Ex F$  are *sa*-wffs, and  $(F \rightarrow G)$  is an *sa*-wff if  $F$  is a *d*-wff.

A *proper sa*-wff is an *sa*-wff that is not a *d*-wff, and a *non-basic sa*-wff is an *sa*-wff that is not a *b*-wff.

If  $F$  is an *sa*-wff, then  $G$  is the *core assertion in F* (or  $CA(F)$  for short) iff  $G$  is like  $F$  except that, wherever  $F$  contains a *bsa*-wff  $H$ ,  $G$  contains  $CA(H)$ . Similarly for  $CP(F)$ .

To any non-basic *sa*-wff  $F$ :

1. The *negative* reply is  $\neg CP(F)$ .
2. The *corrective* reply is  $\neg CA(F)$ .
3. The *direct* replies are  $CP(F)$ ,  $(CA(F) \wedge \neg CP(F))$ , and  $\neg CA(F)$ .
4. The *full* replies are the *d*-wffs that imply direct replies.
5. The *partial* replies are the *d*-wffs implied by direct replies.
6. The *relevant* replies are the full replies plus the partial replies.

(Note that the non-basic *sa*-wffs do not have indicated or wanted replies.)

We assume that for the *d*-wffs we have a first-order predicate logic of the usual kind, and that in addition there might be axioms for set theory, syntax, or the like. To provide for analysis of *sa*-wffs we add the following *rules of ca-derivation*:

- (1)  $(F \wedge G) \vdash F$
- (2)  $(F \wedge G) \vdash (G \wedge F)$
- (3)  $((F \wedge G) \wedge H) \vdash (F \wedge (G \wedge H))$
- (4)  $F, G \vdash (F \wedge G)$
- (5)  $(F \vee F) \vdash F$
- (6)  $(F \vee G) \vdash (G \vee F)$
- (7)  $((F \vee G) \vee H) \vdash (F \vee (G \vee H))$
- (8)  $(F \vee G) \vdash (CA(F) \rightarrow F)$
- (9)  $\neg G, (G \vee F) \vdash F$
- (10)  $G, (G \rightarrow F) \vdash F$
- (11)  $Ux Fx \vdash Ft$
- (12)  $Ex Fx \vdash (CA(Ft) \rightarrow Ft)$
- (13)  $F \vdash G$ , where  $G$  is any one-step alphabetic variant of  $F$
- (14)  $F \vdash CA(F)$

$Z$  is a *ca-derivation from  $S$*  iff  $Z$  is a finite nonempty sequence of *sa-wffs* such that, for every member  $F$  of  $Z$ , either  $F$  is an axiom,  $F$  is a member of  $S$ , or  $F$  comes from preceding members of  $Z$  by a rule of *ca-derivation*. If  $F$  is the last member of  $Z$ , we say that  $Z$  is a *ca-derivation of  $F$  from  $S$*  and that  $F$  is *ca-derivable from  $S$* , and we write  $S \vdash_{ca} F$ .

The following theorem shows that *ca-derivation* is conservative with respect to the derivation of *d-wffs*. Let  $F$  be any *d-wff*, let  $S$  be any set of *sa-wffs*, and let  $CA(S)$  be the set of *d-wffs* that are the core assertions in the members of  $S$ . Let us say that  $F$  is *standardly derivable from a set  $S'$*  just in case there is a finite nonempty sequence  $Z$  of *d-wffs*  $G$  (including  $F$ ) such that every  $G$  either is in  $S'$  or is an axiom or comes from preceding members of  $Z$  by some rule of first-order predicate logic. Then:

**THEOREM 4.**  *$F$  is ca-derivable from  $S$  iff  $F$  is standardly derivable from  $CA(S)$ .*

Various types of content are now definable. E.g. the *assertive commitment* of  $S$  is the set of all *d-sentences*  $F$  such that  $S \vdash_{ca} F$ ; the *projective commitment* of  $S$  is the set of all *d-sentences*  $F$  such that  $S' \vdash_{ca} F$  (where  $S'$  is the union of  $S$  and the set of core projections in members of  $S$ ); and the *erotetic commitment* of  $S$  is the set of all *e-sentences*  $F$  such that  $S \vdash_{ca} F$ .

Let  $Z$  be a *ca-derivation from  $S$* , and let  $S'$  be a finite set of closed terms. Then  $Z$  is *ca-complete for  $S$  relative to  $S'$*  iff all the members of  $S$  have been put into  $Z$  and all the rules that can be applied have been applied. More precisely:

1. For any proper *sa-wff*  $(F \wedge G)$ , if it is in  $Z$ , so is  $F$ . [and analogously for *ca-rules* 2, 3, 5, 6, 7, 9, 10, and 14]
2. For any proper *sa-wff*  $(F \vee G)$ , if it is in  $Z$ , and some  $X$  is such that  $X$  is a proper *sa-wff*,  $X$  is either  $F$  or  $G$ , and  $S \vdash_{ca} CA(X)$ , then, for at least one such  $X$ ,  $(CA(X) \rightarrow X)$  is in  $Z$ .

3. For any proper *sa*-wff  $(G \vee F)$ , if it is in  $Z$ , and  $S \vdash_{ca} \neg G$ , then  $\neg G$  is in  $Z$ .
4. For any proper *sa*-wff  $(G \rightarrow F)$ , if it is in  $Z$ , and  $S \vdash_{ca} G$ , then  $G$  is in  $Z$ .
5. For any proper *sa*-wff  $Ux Fx$ , if it is in  $Z$ , and  $t$  is a closed term in  $S'$ , then  $Ft$  is in  $Z$ .
6. For any proper *sa*-wff  $Ex Fx$ , if it is in  $Z$ , and some closed term  $t$  in  $S'$  is such that  $S \vdash_{ca} CA(Ft)$ , then, for at least one such term  $t$ ,  $(CA(Ft) \rightarrow Ft)$  is in  $Z$ .

Where  $S$  is a finite set of *sa*-wffs, a *sufficient reply to S* is constructed in the following way: First find a *ca*-derivation  $Z$  from  $S$  that is *ca*-complete for  $S$  relative to the set of closed terms that occur in members of  $S$ . Choose *b*-sentences  $F_1, \dots, F_n$  that occur in  $Z$ , provided that all the *e*-sentences in  $Z$  are included among  $F_1, \dots, F_n$ . Then choose  $G_1, \dots, G_n$  such that each  $G_i$  is a direct reply to  $F_i$ . Then  $(G_1 \wedge \dots \wedge G_n)$  is a sufficient reply to  $S$ .

Unfortunately *ca*-complete derivations are not effectively recognizable as such, so sufficient replies are not effectively recognizable as such. On the other hand, by making certain additions and changing some details, we can make the reply process more effective in certain respects. The key is to extend the language by adding a stock of reply indicators  $r_i$  and *r*-wffs of the form

$$(r_{j_1} F_1 : G_1) \wedge \dots \wedge (r_{j_n} F_n : G_n)$$

Roughly, each  $(r_i F : G)$  says that  $G$  is a reply to  $F$  of the kind  $i$ . In particular, a sufficient reply to  $S$  would have the form displayed above, where each  $r_i$  would be an indicator for direct reply. For details, see Harrah [1985].

Concerning vectored sentences and vectored messages: Each such expression  $X$  consists of a body  $B$  and a vector  $V$ ; the content of  $X$  is a function of the content of  $B$  and the content of  $V$ . To simplify here we assume that the body of a vectored message is a finite nonempty string of *sa*-sentences; thus each vectored sentence counts as a vectored message, but a vectored sentence cannot occur inside a vectored message.

Vectors are expressions of various kinds, and each kind of vector brings certain *presumptions*. Example:

‘To: Jane Smith, Dean of the College’

brings the presumption

‘Jane Smith is Dean of the College’

(For discussion of vectors and presumptions, see Harrah [1994].) The content of the vector is determined by these presumptions. We assume that each vectored  $X$  has finitely many presumptions, that the presumptions are  $d$ -sentences, and that each presumption of  $X$  is effectively recognizable from ‘ $X$ ’.

For message analysis: Let  $M$  be a vectored message, and let  $S(M)$  be the set consisting of (1) the  $sa$ -sentences in the body of  $M$  and (2) the presumptions of  $M$ . Then  $Z$  is an *ma-derivation from  $M$*  iff  $Z$  is a *ca-derivation from  $S(M)$* , and  $Z$  is *ma-complete for  $M$*  iff  $Z$  is *ca-complete for  $S(M)$*  relative to the set of terms that occur in  $M$ .

We construct a *sufficient reply to  $M$*  in either of three ways:

Option I: First find an *ma-derivation  $Z$  from  $M$*  that is *ma-complete for  $M$* . Choose  $b$ -sentences  $F_1, \dots, F_n$  that occur in  $Z$  and include all the  $e$ -sentences in  $Z$ . Then choose direct replies  $G_i$  to these  $F_i$  and form  $(G_1 \wedge \dots \wedge G_n)$  as a sufficient reply to  $M$ .

Option II: Find a  $d$ -sentence  $F$  that is *ca-derivable* from the set of presumptions of  $M$  (an  $F$  that you believe is false). Then the negation  $\neg F$  is a *vector-challenge to  $M$*  and may be given as a sufficient reply to  $M$ .

Option III: Find a  $d$ -sentence  $F$  such that (1)  $F$  is *ma-derivable from  $M$* , and (2) every *ma-derivation of  $F$  from  $M$*  contains at least one presumption of  $M$ . Then the negation  $\neg F$  is a *vector-challenge* and is a sufficient reply to  $M$ .

*University of California, Riverside, USA.*

## BIBLIOGRAPHY

- [Åqvist, 1965] L. Åqvist. *A New Approach to the Logical Theory of Interrogatives*. Almqvist & Wiksell, Uppsala, 1965.
- [Åqvist, 1969] L. Åqvist. Scattered topics in interrogative logic. In J. Davis, D. J. Hockney, and W. K. Wilson, editors, *Philosophical Logic*, pages 114 – 121. D. Reidel, Dordrecht, 1969.
- [Åqvist, 1971] L. Åqvist. Revised foundations for imperative epistemic and interrogative logic. *Theoria*, 37:33 – 73, 1971.
- [Åqvist, 1983] L. Åqvist. On the “Tell Me Truly” approach to the analysis of interrogatives. In F. Kiefer, editor, *Questions and Answers*, pages 9 – 14. D. Reidel, Dordrecht, 1983.
- [Belnap and Steel, 1976] N. Belnap and T. Steel. *The Logic of Questions and Answers*. Yale, New Haven, 1976.
- [Belnap, 1963] N. Belnap. An analysis of questions: preliminary report. Technical Report 7 1287 1000/00, System Development Corporation, Santa Monica, CA, 1963.
- [Belnap, 1969] N. Belnap. Åqvist’s corrections-accumulating question-sequences. In J. Davis, D. J. Hockney, and W. K. Wilson, editors, *Philosophical Logic*, pages 122 – 134. D. Reidel, Dordrecht, 1969.
- [Belnap, 1981] N. Belnap. Questions and answers in Montague grammar. In S. Peters and E. Saarinen, editors, *Processes, Beliefs, and Questions. Essays on Formal Semantics of Natural Language and Natural Language Processing*, pages 165 – 198. D. Reidel, Dordrecht, 1981.

- [Belnap, 1983] N. Belnap. Approaches to the semantics of questions in natural language. In R. Bäuerle et al., editors, *Meaning, Use, and Interpretation of Language*, pages 21 – 29. Walter de Gruyter, Berlin, 1983.
- [Berkov, 1979] V. Berkov. *Nauchnaya problema (logiko-metodologicheskii aspekt)*. Izdatel'stvo BGY im. V.I. Lenina, Minsk, 1979.
- [Bogdan, 1987] R. Bogdan, editor. *Jaakko Hintikka*. D. Reidel, Dordrecht, 1987.
- [Bolinger, 1978] D. Bolinger. Yes-no questions are not alternative questions. In H. Hiz, editor, *Questions*, pages 87 – 105. D. Reidel, Dordrecht, 1978.
- [Bromberger, 1992] S. Bromberger. *On What We Know We Don't Know: Explanation, Theory, Linguistics, and How Questions Shape Them*. University of Chicago Press, Chicago, 1992.
- [Carlson, 1983] L. Carlson. *Dialogue Games*. D. Reidel, Dordrecht, 1983.
- [Cohen, 1929] F. Cohen. What is a question? *The Monist*, 39:350 – 364, 1929.
- [Cresswell, 1965] M. Cresswell. On the logic of incomplete answers. *The Journal of Symbolic Logic*, 30:65 – 68, 1965.
- [Dacey, 1981] R. Dacey. An interrogative account of the dialectical inquiring system based upon the economic theory of information. *Synthese*, 47:43 – 55, 1981.
- [Davis et al., 1969] J. Davis, D. J. Hockney, and W. K. Wilson, editors. *Philosophical Logic*. D. Reidel, Dordrecht, 1969.
- [Ficht, 1978] H. Ficht. Supplement to a bibliography on the theory of questions and answers. *Linguistische Berichte*, 55:92 – 114, 1978.
- [Finn, 1974] V. Finn. K logiko-semioticheskoy teorii informatsionnogo poiska. In *Informatsionnye voprosy semiotiki, lingvistiki i avtomaticheskogo perevoda*, volume 5. Vsesoyuznyy Institut Nauchnoy i Tekhnicheskoy Informatsii, ANSSSR, Moscow, 1974.
- [Gornstein, 1967] I. Gornstein. The logical analysis of questions: a historical survey. 1967.
- [Grewendorf, 1983] G. Grewendorf. What answers can be given? In F. Kiefer, editor, *Questions and Answers*, pages 45 – 84. D. Reidel, Dordrecht, 1983.
- [Groenendijk and Stokhof, 1984] J. Groenendijk and M. Stokhof. *Studies on the Semantics of Questions and the Pragmatics of Answers*. Academisch Proefschrift, Amsterdam, 1984.
- [Hamblin, 1958] C. Hamblin. Questions. *The Australasian Journal of Philosophy*, 36:159 – 168, 1958.
- [Hamblin, 1967] C. Hamblin. Questions. In P. Edwards, editor, *The Encyclopedia of Philosophy*. Macmillan, New York, 1967.
- [Hand, editor, 1994] M. Hand, editor. Game theoretical semantics. *Synthese*, 99:311 – 456, 1994.
- [Hand, 1988] M. Hand. Game-theoretical semantics, Montague semantics, and questions. *Synthese*, 74:207 – 222, 1988.
- [Harrah, 1961] D. Harrah. A logic of questions and answers. *Philosophy of Science*, 28:40 – 46, 1961.
- [Harrah, 1963a] D. Harrah. *Communication: A Logical Model*. MIT, Cambridge, MA, 1963a.
- [Harrah, 1963b] D. Harrah. Review of Stahl [1962]. *The Journal of Symbolic Logic*, 28:259, 1963b.
- [Harrah, 1969] D. Harrah. On completeness in the logic of questions. *American Philosophical Quarterly*, 6:158 – 164, 1969.
- [Harrah, 1975] D. Harrah. A system for erotetic sentences. In A. Anderson et al., editors, *The Logical Enterprise*, pages 235 – 245. Yale, New Haven, 1975.
- [Harrah, 1979] D. Harrah. Critical study of Hintikka [1976]. *Noûs*, 13:95 – 99, 1979.
- [Harrah, 1980] D. Harrah. On speech acts and their logic. *Pacific Philosophical Quarterly*, 61:204 – 211, 1980.
- [Harrah, 1981] D. Harrah. The semantics of question sets. In D. Krallmann and G. Stickel, editors, *Zur Theorie der Frage*. Gunter Narr Verlag, Tübingen, 1981.
- [Harrah, 1981a] D. Harrah. On the complexity of texts and text theory. *Text*, 1:83 – 95, 1981a.

- [Harrah, 1982] D. Harrah. Guarding what we say. In T. Pauli, editor, *Philosophical Essays dedicated to Lennart Åqvist*, pages 119 – 131. University of Uppsala, Uppsala, 1982.
- [Harrah, 1985] D. Harrah. A logic of message and reply. *Synthese*, 63:275 – 294, 1985.
- [Harrah, 1987] D. Harrah. Hintikka's theory of questions. In R. Bogdan, editor, *Jaakko Hintikka*, pages 199 – 213. D. Reidel, Dordrecht, 1987.
- [Harrah, 1994] D. Harrah. On the vectoring of speech acts. In S. Tsohatzidis, editor, *Foundations of Speech Act Theory: Philosophical and Linguistic Perspectives*, pages 374 – 390. Routledge, London, 1994.
- [Harris, 1994] S. Harris. GTS and interrogative tableaux. *Synthese*, 99:329 – 343, 1994.
- [Higginbotham, 1993] J. Higginbotham. Interrogatives. In K. Hale and S. Keyser, editors, *The View from Building 20: Essays in Linguistics in Honor of Sylvain Bromberger*, pages 195 – 227. The MIT Press, Cambridge, MA, 1993.
- [Hintikka and Halonen, 1995] J. Hintikka and I. Halonen. Semantics and pragmatics for why-questions. *The Journal of Philosophy*, 92:636 – 657, 1995.
- [Hintikka, editor, 1988] J. Hintikka, editor. Knowledge-seeking by questioning. *Synthese*, 74:1 – 262, 1988.
- [Hintikka, 1976] J. Hintikka. *The Semantics of Questions and the Questions of Semantics: Case Studies in the Interrelations of Logic, Semantics and Syntax*. North-Holland, Amsterdam, 1976.
- [Hintikka, 1983] J. Hintikka. New foundations for a theory of questions and answers. In F. Kiefer, editor, *Questions and Answers*, pages 159 – 190. D. Reidel, Dordrecht, 1983.
- [Hintikka, 1992] J. Hintikka. The interrogative model of inquiry as a general theory of argumentation. *Communication and Cognition*, 25:221 – 242, 1992.
- [Hiž, 1962] H. Hiž. Questions and answers. *The Journal of Philosophy*, 59:253 – 265, 1962.
- [Hiž, 1978] H. Hiž, editor. *Questions*. D. Reidel, Dordrecht, 1978.
- [Hoepelman, 1983] J. Hoepelman. On questions. In F. Kiefer, editor, *Questions and Answers*, pages 191 – 227. D. Reidel, Dordrecht, 1983.
- [Kampé de Fériet and Picard, 1974] J. Kampé de Fériet and C. Picard, editors. *Théories de l'Information (Lecture notes in Mathematics, Volume 398)*. Springer-Verlag, Berlin, 1974.
- [Karttunen, 1978] L. Karttunen. Syntax and semantics of questions. In H. Hiž, editor, *Questions*, pages 165 – 210. D. Reidel, Dordrecht, 1978.
- [Keenan and Hull, 1973] E. Keenan and R. Hull. The logical presuppositions of questions and answers. In J. Petöfi and D. Franck, editors, *Präsuppositionen in Philosophie und Linguistik*, pages 441 – 466. Athenäum, Frankfurt/M., 1973.
- [Kiefer, 1983] F. Kiefer, editor. *Questions and Answers*. D. Reidel, Dordrecht, 1983.
- [Koj and Wiśniewski, 1989] L. Koj and A. Wiśniewski. Inquiries into the generating and proper use of questions. *Wydawnictwo Naukowe UMCS, Realizm Racjonalność Relatywizm*, 12, 1989.
- [Koura, 1988] A. Koura. An approach to why-questions. *Synthese*, 74:191 – 206, 1988.
- [Krallmann and Stickel, 1981] D. Krallmann and G. Stickel, editors. *Zur Theorie der Frage*. Gunter Narr Verlag, Tübingen, 1981.
- [Kripke, 1975] S. Kripke. Outline of a theory of truth. *The Journal of Philosophy*, 72:690 – 716, 1975.
- [Kubiński, 1960] T. Kubiński. An essay in the logic of questions. *Atti del XII Congr. Intern. di Filosofia (Firenze)*, 5:315 – 322, 1960.
- [Kubiński, 1967] T. Kubiński. Some observations about a notion of incomplete answer. *Studia Logica*, 21:39 – 42, 1967.
- [Kubiński, 1980] T. Kubiński. *An Outline of the Logical Theory of Questions*. Akademie-Verlag, Berlin, 1980.
- [Lehnert, 1978] W. Lehnert. *The Process of Question Answering*. Wiley, New York, 1978.
- [Leszko, 1980] R. Leszko. *Wyznaczenie Pewnych Klas Pytań Przez Grafy i Ich Macierze*. Wyższa Szkoła Pedagogiczna, Zielona Góra, 1980.

- [Leszko, forthcoming] R. Leszko. Graphs and matrices of compound numerical questions. *Acta Universitatis Wratislaviensis*, forthcoming.
- [Lewis and Lewis, 1975] D. Lewis and S. Lewis. Review of Olson and Paul, *contemporary philosophy in scandinavia*. *Theoria*, 41:39 – 60, 1975.
- [Materna, 1981] P. Materna. Question-like and non-question-like imperative sentences. *Linguistics and Philosophy*, 4:393 – 404, 1981.
- [Mayo, 1956] B. Mayo. Deliberative questions: a criticism. *Analysis*, 16:58 – 63, 1956.
- [Meyer, 1988] M. Meyer, editor. *Questions and Questioning*. Walter de Gruyter, Berlin, 1988.
- [Picard, 1980] C. Picard. *Graphs and Questionnaires*. North-Holland, Amsterdam, 1980.
- [Prior and Prior, 1955] M. Prior and A. Prior. Erotetic logic. *Philosophical Review*, 64:43 – 59, 1955.
- [Shoemith and Smiley, 1978] D. Shoemith and T. Smiley. *Multiple-conclusion Logic*. Cambridge University Press, Cambridge, 1978.
- [Stahl, 1956] G. Stahl. La logica de las preguntas. *Anales de la Universidad de Chile*, 102:71 – 75, 1956.
- [Stahl, 1962] G. Stahl. Fragenfolgen. In M. Käsbaauer and F. Kutschera, editors, *Logik und Logikkalkül*, pages 149 – 157. Alber, Freiburg/Munich, 1962.
- [Szaniawski, 1973] K. Szaniawski. Questions and their pragmatic value. In R. Bogdan and I. Niiniluoto, editors, *Logic, Language, and Probability*, pages 121 – 123. D. Reidel, Dordrecht, 1973.
- [Tichý, 1978] P. Tichý. Questions, answers, and logic. *American Philosophical Quarterly*, 15:275 – 284, 1978.
- [Todt and Schmidt-Radefeldt, 1979] G. Todt and J. Schmidt-Radefeldt. Wissensfragen und direkte Antworten in der Fragelogik LA<sup>2</sup>. *Linguistische Berichte*, 62:1 – 24, 1979.
- [Todt and Schmidt-Radefeldt, 1981] G. Todt and J. Schmidt-Radefeldt. Review of Belnap and Steel [1976]. *The Journal of Pragmatics*, 5:95 – 101, 1981.
- [Wachowicz, 1978] K. Wachowicz. Q-morpheme hypothesis. In H. Hiž, editor, *Questions*. D. Reidel, Dordrecht, 1978.
- [Wheatley, 1955] J. Wheatley. Deliberative questions. *Analysis*, 15:49 – 60, 1955.
- [Wiśniewski, 1994a] A. Wiśniewski. Erotetic implications. *Journal of Philosophical Logic*, 23:173 – 195, 1994a.
- [Wiśniewski, 1994b] A. Wiśniewski. On the reducibility of questions. *Erkenntnis*, 40:265 – 284, 1994b.
- [Wiśniewski, 1995] A. Wiśniewski. *The Posing of Questions: Logical Foundations of Erotetic Inferences*. Kluwer, Dordrecht, 1995.
- [Wiśniewski, editor, 1997] A. Wiśniewski and J. Zygmunt, editors. *Erotetic Logic, Deontic Logic, and Other Logical Matters: Essays in Memory of Tadeusz Kubiński*. Wydawnictwo Uniwersytetu Wrocławskiego, Wrocław, 1997.



## SEQUENT SYSTEMS FOR MODAL LOGICS

## INTRODUCTION

[T]he framework of ordinary sequents is not capable of handling all interesting logics. There are logics with nice, simple semantics and obvious interest for which no decent, cut-free formulation seems to exist . . . . Larger, but still satisfactory frameworks should, therefore, be sought. A. Avron [1996, p. 3]

This chapter surveys the application of various kinds of sequent systems to modal and temporal logic, also called tense logic. The starting point are ordinary Gentzen sequents and their limitations both technically and philosophically. The rest of the chapter is devoted to generalizations of the ordinary notion of sequent. These considerations are restricted to formalisms that do not make explicit use of semantic parameters like possible worlds or truth values, thereby excluding, for instance, Gabbay's labelled deductive systems, indexed tableau calculi, and Kanger-style proof systems from being dealt with. Readers interested in these types of proof systems are referred to [Gabbay, 1996], [Goré, 1999] and [Pliuškeviene, 1998]. Also Orłowska's [1988; 1996] Rasiowa-Sikorski-style relational proof systems for normal modal logics will not be considered in the present chapter. In relational proof systems the logical object language is associated with a language of relational terms. These terms may contain subterms representing the accessibility relation in possible-worlds models, so that semantic information is available at the same level as syntactic information. The derivation rules in relational proof systems manipulate finite sequences of relational formulas constructed from relational terms and relational operations. An overview of *ordinary* sequent systems for non-classical logics is given in [Ono, 1998], and for a general background on proof theory the reader may consult [Troelstra and Schwichtenberg, 2000]. In this chapter we shall pay special attention to *display logic*, a general proof-theoretic approach developed by Belnap [1982]. Two applications of the modal display calculus are included as case studies: the formulas-as-types notion of construction for temporal logic and a display calculus for propositional bi-intuitionistic logic (also called Heyting-Brouwer logic). This logic comprises both constructive implication and coimplication (see, for example, [Goré, 2000], [Rauszer, 1980], [Wolter, 1998]), and its sequent-calculus presentation to be given is based on a modal translation into the temporal propositional logic **S4t**.<sup>1</sup>

<sup>1</sup>The chapter consists of revised and expanded material from [Wansing, 1998] and includes the contents of the unpublished report [Wansing, 2000] on formulas-as-types for temporal logics. Moreover, the sequent calculus for bi-intuitionistic logic and subsystems of bi-intuitionistic logics in Section 3.8 and the translation of multiple-sequent systems into higher-arity sequent systems in Section 4.1 are new.

*A note on notation.* In the present chapter, both classical and constructive logics will be considered. Therefore it makes sense to reflect this distinction in the notation for the logical operations. In particular, the following symbols will be used:  $\triangleright$  (constructive, intuitionistic implication),  $\blacktriangleleft$  (coimplication),  $\supset$  (Boolean implication),  $\frown$  (intuitionistic negation),  $\smile$  (conegation),  $\neg$  (Boolean negation).

## 1 ORDINARY SEQUENT SYSTEMS

The presentation of normal modal logics as ordinary (standard) sequent systems has turned out to be problematic for both technical and philosophical reasons. The technical problems chiefly result from a lack of flexibility of the ordinary notion of sequent for dealing with the multitude of interesting and important modal logics in a uniform and perspicuous way. In this section a number of standard Gentzen systems for normal modal propositional logics is reviewed in order to give an impression of what has been and what can be done to present normal modal logics as ordinary Gentzen calculi. An ordinary Gentzen system is a collection of rule schemata for manipulating *Gentzen sequents*; these are derivability statements of the form  $\Delta \rightarrow \Gamma$ , where  $\Delta$  and  $\Gamma$  are finite, possibly empty sets of formulas. The set terms ‘ $\Delta$ ’ and ‘ $\Gamma$ ’ are called the antecedent and the succedent of  $\Delta \rightarrow \Gamma$ , respectively. Often, a sequent

$$\{A_1, \dots, A_m\} \rightarrow \{B_1, \dots, B_n\}$$

is written as  $A_1, \dots, A_m \rightarrow B_1, \dots, B_n$ . This notation supports viewing the ‘,’ (the comma) as a *structure connective* in the language of sequents. Indeed, the sequent arrow in Gentzen’s [1934] denotes a derivability relation between finite *sequences* of formulas separated by the comma. Gentzen, however, postulated structural rules that justify thinking of antecedents and succedents as denoting sets:

$$\text{(permutation)} \quad \frac{\Delta, A, B, \Gamma \rightarrow \Sigma}{\Delta, B, A, \Gamma \rightarrow \Sigma} \quad \frac{\Delta \rightarrow \Sigma, A, B, \Gamma}{\Delta \rightarrow \Sigma, B, A, \Gamma}$$

$$\text{(contraction)} \quad \frac{\Delta, A, A, \Gamma \rightarrow \Sigma}{\Delta, A, \Gamma \rightarrow \Sigma} \quad \frac{\Delta \rightarrow \Sigma, A, A, \Gamma}{\Delta \rightarrow \Sigma, A, \Gamma}$$

Gentzen also postulated

$$\text{(monotonicity)} \quad \frac{\Delta, \Gamma \rightarrow \Sigma}{\Delta, A, \Gamma \rightarrow \Sigma} \quad \frac{\Delta \rightarrow \Gamma, \Sigma}{\Delta \rightarrow \Gamma, A, \Sigma}$$

These three rules are structural in the sense of exhibiting no operation from an underlying logical object language. If the polymorphic comma is interpreted as a binary structure connective that may or may not be associative, the antecedent and the succedent of a sequent are *Gentzen terms*,

and in generalized sequent calculi, the sequents display Gentzen terms or other, much more complex data structures. We shall use ‘ $\vdash$ ’ to denote the derivability relation in a given axiomatic system or a consequence relation between finite sets of sequents and single sequents satisfying identity, cut, and monotonicity. In other words, if  $\Delta$  and  $\Gamma$  are finite sets of sequents and  $s, s'$  are sequents, then we assume that  $\{s\} \vdash s$ ,

$$\frac{\Delta \vdash s}{\Delta \cup \{s'\} \vdash s} \quad \text{and} \quad \frac{\Delta \vdash s \quad \Gamma \cup \{s\} \vdash s'}{\Delta \cup \Gamma \vdash s'}.$$

### 1.1 Ordinary Gentzen systems for normal modal logics

The syntax of modal propositional logic (in Backus-Naur form, see for example [Goldblatt, 1992, p. 3]) is given by:

$$A ::= p \mid \mathbf{t} \mid \mathbf{f} \mid \neg A \mid A \wedge B \mid A \vee B \mid A \supset B \mid A \equiv B \mid \diamond A \mid \square A.$$

The smallest normal modal propositional logic **K** admits a simple presentation as an ordinary Gentzen system (see, for instance, [Leivant, 1981], [Mints, 1990], [Sambin and Valentini, 1982]). In the language with  $\square$  (“necessarily”) as the only primitive modal operator and  $\diamond A$  (“possibly  $A$ ”) being defined as  $\neg \square \neg A$ , one may just add the rule

$$(\rightarrow \square)_1 \quad \Delta \rightarrow A \vdash \square \Delta \rightarrow \square A$$

to the standard sequent system **LCPL** for classical propositional logic **CPL**, where  $\square \Delta = \{\square A \mid A \in \Delta\}$ . A sequent calculus **LK4** for **K4** can be obtained by supplementing **LCPL** with the rule

$$(\rightarrow \square)_2 \quad \Delta, \square \Delta \rightarrow A \vdash \square \Delta \rightarrow \square A$$

(see [Sambin and Valentini, 1982]). In [Goble, 1974] it is shown that the pair of modal sequent rules  $(\rightarrow \square)_1$  and

$$(\square \rightarrow)_1 \quad \Delta, A \rightarrow \emptyset \vdash \square \Delta, \square A \rightarrow \emptyset$$

yields a sequent system for **KD** (where ‘ $\emptyset$ ’ denotes the empty set) and that a sequent calculus for **KD4** is obtained, if  $(\rightarrow \square)_1$  is replaced by the rule

$$(\rightarrow \square)_3 \quad \Delta' \rightarrow A \vdash \square \Delta \rightarrow \square A,$$

where  $\Delta'$  results from  $\Delta$  by prefixing zero or more formulas in  $\Delta$  by  $\square$ . Shvarts [1989] gives a sequent calculus formulation of **KD45** by adjoining to **LCPL** the following rule for  $\square$ :

$$[\square] \quad \square \Delta_1, \Delta_2 \rightarrow \square \Gamma_1, \Gamma_2 \vdash \square \Delta_1, \square \Delta_2 \rightarrow \square \Gamma_1, \square \Gamma_2,$$

where  $\Gamma_2$  contains at most one formula. If in addition  $\Gamma_1$  and  $\Gamma_2$  are required to be non-empty, this results in a sequent system for **K45**.

Among the most important modal logics are the almost ubiquitous systems **S4** and **S5**. Standard sequent systems for the axiomatic calculi **S4** (= **KT4**) and **S5** (= **KT5** = **KT4B**) were studied by Ohnishi and Matsumoto [1957]. They considered the following schematic sequent rules for  $\Box$  and  $\Diamond$ :

$$\begin{aligned} (\rightarrow \Box)_0 & \quad \Box\Delta \rightarrow \Box\Gamma, A \vdash \Box\Delta \rightarrow \Box\Gamma, \Box A; \\ (\Box \rightarrow)_0 & \quad \Delta, A \rightarrow \Gamma \vdash \Delta, \Box A \rightarrow \Gamma; \\ (\rightarrow \Diamond)_0 & \quad \Delta \rightarrow \Gamma, A \vdash \Delta \rightarrow \Gamma, \Diamond A; \\ (\Diamond \rightarrow)_0 & \quad \Diamond\Gamma, A \rightarrow \Diamond\Delta \vdash \Diamond\Gamma, \Diamond A \rightarrow \Diamond\Delta; \end{aligned}$$

where  $\Diamond\Delta = \{\Diamond A \mid A \in \Delta\}$ . If either the rules  $(\rightarrow \Box)_0$  and  $(\Box \rightarrow)_0$  or the rules  $(\rightarrow \Diamond)_0$  and  $(\Diamond \rightarrow)_0$  are adjoined to **LCPL**, then the result is a sequent calculus **LS5\*** for **S5**. If  $\Gamma$  is empty in  $(\rightarrow \Box)_0$  or  $(\Diamond \rightarrow)_0$ , this yields a sequent calculus **LS4** for **S4**. Several other modal logics can be obtained by imposing suitable constraints on the structures exhibited in  $(\rightarrow \Box)_0$  and  $(\Diamond \rightarrow)_0$ , respectively. Ohnishi and Matsumoto show that if  $(\rightarrow \Box)_0$  and  $(\Diamond \rightarrow)_0$  are replaced by  $(\rightarrow \Box)_1$  and

$$(\Diamond \rightarrow)_1 \quad A \rightarrow \Gamma \vdash \Diamond A \rightarrow \Diamond\Gamma,$$

one obtains a Gentzen-system **LKT** for **KT** (= **T**). Kripke [1963] noted that the equivalences between  $\Box A$  and  $\neg\Diamond\neg A$  and between  $\Diamond A$  and  $\neg\Box\neg A$  cannot be proved by means of Ohnishi's and Matsumoto's rules. In the case of **S4**, Kripke suggested remedying this by using sequent rules which exhibit both  $\Box$  and  $\Diamond$ , namely in addition to  $(\Box \rightarrow)_0$  and  $(\rightarrow \Diamond)_0$  the rules

$$\begin{aligned} (\rightarrow \Box)' & \quad \Box\Gamma \rightarrow A, \Diamond\Delta \vdash \Box\Gamma \rightarrow \Box A, \Diamond\Delta \\ \text{and } (\Diamond \rightarrow)' & \quad A, \Box\Gamma \rightarrow \Diamond\Delta \vdash \Diamond A, \Box\Gamma \rightarrow \Diamond\Delta. \end{aligned}$$

Such rules fail to give a separate account of the inferential behaviour of  $\Box$  and  $\Diamond$ , since only the combined use of these operations is specified. Another problem with Ohnishi's and Matsumoto's sequent rules for **S5** is that the cut-rule

$$\Delta \rightarrow \Sigma, A; \quad \Gamma, A \rightarrow \Theta \vdash \Gamma, \Delta \rightarrow \Sigma, \Theta$$

cannot be eliminated: the system without cut allows proving less formulas than the full system containing cut. Ohnishi and Matsumoto [1959] give the following counter-example to cut-elimination:

$$\frac{\frac{\frac{\Box p \rightarrow \Box p}{\emptyset \rightarrow \neg\Box p, \Box p}}{\emptyset \rightarrow \Box\neg\Box p, \Box p} \quad \frac{p \rightarrow p}{\Box p \rightarrow p}}{\emptyset \rightarrow \Box\neg\Box p, p}$$

A solution to the problem of defining a cut-free ordinary Gentzen system for **S5** has been given in [Bräuner, 2000].<sup>2</sup> The logic **S5** can be faithfully

<sup>2</sup>Another, perhaps less convincing solution has been presented by Ohnishi [1982]. Define the degree  $\text{deg}(A)$  of a modal formula in the language with  $\Box$  primitive as follows:

embedded into monadic predicate logic, the first-order logic of unary predicates, under a translation  $\mathfrak{t}$  employing a single individual variable  $x$ , see for instance [Mints, 1992]. The translation  $\mathfrak{t}$  assigns to every propositional variable  $p$  an atomic formula  $P(x)$ , and for compound formulas it is defined as follows:

$$\begin{aligned} \mathfrak{t}(t) &= t, \\ \mathfrak{t}(\neg A) &= \neg \mathfrak{t}(A), \\ \mathfrak{t}(A \sharp B) &= \mathfrak{t}(A) \sharp \mathfrak{t}(B), \text{ for } \sharp \in \{\supset, \wedge, \vee\}, \\ \mathfrak{t}(\Box A) &= \forall x \mathfrak{t}(A), \\ \mathfrak{t}(\Diamond A) &= \exists x \mathfrak{t}(A). \end{aligned}$$

**THEOREM 1.** *A modal formula  $A$  is provable in **S5** if and only if  $\mathfrak{t}(A)$  is provable in monadic predicate logic.*

The familiar cut-free sequent calculus for monadic predicate logic can serve as a starting point for defining a cut-free ordinary sequent system for **S5** with side-conditions on the introduction rules for  $\Box$  on the right and  $\Diamond$  on the left of the sequent arrow. The side conditions are simple, though their precise formulation requires some terminology that will be useful also in other contexts. An inference *inf* is a pair  $(\Delta, s)$ , where  $\Delta$  is a set of sequents (the premises of *inf*) and  $s$  is a single sequent (the conclusion of *inf*). A rule of inference  $R$  is a set of inferences. If  $\text{inf} \in R$ , then *inf* is said to be an instantiation of  $R$ . The rule  $R$  is an axiomatic rule, if  $\Delta = \emptyset$  for every  $(\Delta, s) \in R$ . We assume that inference rules are stated by using variables for structures (in the present case finite sets of formulas) and formulas. Every structure occurrence in an inference *inf* (a sequent  $s$ ) is called a constituent of *inf* ( $s$ ). The *parameters* of  $\text{inf} \in R$  are those constituents which occur as substructures of structures assigned to structure variables in the statement of  $R$ . Constituents of *inf* are defined as *congruent* in *inf* if and only if (iff) they are occupying similar positions in occurrences of structures assigned to the same structure variable, in the present case iff they belong to a set assigned to the same set variable.

**DEFINITION 2.** Two formula occurrences are immediately connected in a proof  $\Pi$  iff  $\Pi$  contains an inference *inf* such that one of the following

1.  $\text{deg}(p) = 0$ , for every propositional variable  $p$ ;
2.  $\text{deg}(\neg A) = \text{deg}(A)$ ;
3.  $\text{deg}(A \wedge B) = \max(\text{deg}(A), \text{deg}(B))$ ;
4.  $\text{deg}(\Box A) = \text{deg}(A) + 1$ .

Ohnishi adds to  $(\Box \rightarrow)_0$  and  $(\rightarrow \Box)_0$  two further rules that deviate considerably from familiar introduction schemata:

$$\Gamma, A^*, \Delta \rightarrow \Sigma \vdash \Gamma, A, \Delta \rightarrow \Sigma \quad \text{and} \quad \Gamma \rightarrow \Delta, A^*, \Sigma \vdash \Gamma \rightarrow \Delta, A, \Sigma,$$

where the formula  $A^*$  is defined in such a way that (i)  $A$  and  $A^*$  are equivalent in **S5** and (ii)  $\text{deg}(A^*) \leq 1$ .

conditions holds:

1. both occurrences are non-parametric, one in the conclusion and the other in a premise of *inf*;
2. *inf* belongs to an axiomatic sequent rule and both occurrences are non-parametric in *inf*;
3. *inf*  $\in$  cut and both occurrences are non-parametric in *inf*;
4. the occurrences are parametric and congruent in *inf*.

A list of formula occurrences  $A_1, \dots, A_n$  in a proof  $\Pi$  is called a connection between  $A_1$  and  $A_n$  in  $\Pi$  iff for every  $i \in \{1, \dots, n-1\}$ , the occurrences  $A_i$  and  $A_{i+1}$  are immediately connected in  $\Pi$ . A formula is said to be modally closed if every propositional variable in the formula occurs in the scope of an occurrence  $\diamond$  or  $\square$ .

**DEFINITION 3.** Two formula occurrences in a proof  $\Pi$  are said to be dependent on each other in  $\Pi$  iff there exists a connection between these occurrences that does not contain any modally closed formula.

The sequent system **LS5** extends **LCPL** by  $(\square \rightarrow)_0$ ,  $(\rightarrow \diamond)_0$  and the rules:

$$\begin{array}{l} (\rightarrow \square)'' \quad \Gamma \rightarrow \Delta, A \vdash \Gamma \rightarrow \Delta, \square A \\ \text{and } (\diamond \rightarrow)'' \quad \Gamma, A \rightarrow \Delta \vdash \Gamma, \diamond A \rightarrow \Delta, \end{array}$$

where applications of  $(\rightarrow \square)''$  and  $(\diamond \rightarrow)''$  in a proof  $\Pi$  must be such that in  $\Pi$  none of the formula occurrences in  $\Gamma$  and  $\Delta$  depends on the displayed occurrence of  $A$ . A cut-free proof of the notorious sequent  $\emptyset \rightarrow \square \neg \square p, p$  is then easily available (as it is also in Ohnishi's [1982] calculus):

$$\frac{\frac{\frac{p \rightarrow p}{\square p \rightarrow p}}{\emptyset \rightarrow \neg \square p, p}}{\emptyset \rightarrow \square \neg \square p, p}$$

**THEOREM 4.** ([Braüner, 2000]) *A sequent  $\Delta \rightarrow \Gamma$  is provable in **LS5** iff  $\bigwedge \Delta \supset \bigvee \Gamma$  is provable in **S5**.*

Avron [1984] (see also [Shimura, 1991]) presents a sequent calculus **LS4Grz** for **S4Grz** (= **KGrz**). He replaces the rule  $(\rightarrow \square)_0$  in Ohnishi and Matsumoto's sequent calculus for **S4** by the rule

$$(\rightarrow \square)_4 \quad \square(A \supset \square A), \square \Delta \rightarrow A \vdash \square \Delta \rightarrow \square A$$

exhibiting both  $\square$  and  $\supset$ . In [Takano, 1992], Takano defines sequent calculi **LKB**, **LKTB**, **LKDB**, and **LK4B** for **KB**, **KTB** (= **B**), **KDB**, and **K4B**.

The systems **LKB** and **LK4B** are obtained from **LCPL** by including the rules

$$\begin{aligned} (\rightarrow \Box)_B \quad & \Gamma \rightarrow \Box\Theta, A \vdash \Box\Gamma \rightarrow \Theta, \Box A \\ \text{and } (\rightarrow \Box)_{ABE} \quad & \Gamma, \Box\Gamma \rightarrow \Box\Theta, \Box\Delta, A \vdash \Box\Gamma \rightarrow \Box\Theta, \Delta, \Box A \end{aligned}$$

respectively. **LKTB** and **LKDB** result from **LKB** by adjoining  $(\Box \rightarrow)_0$  and

$$(\Box \rightarrow)_D \quad \Gamma \rightarrow \Box\Delta \vdash \Box\Gamma \rightarrow \Delta$$

respectively. Standard sequent systems for several other modal logics can be found in [Goré, 1992] and [Zeman, 1973]. The sequent calculus for **S4.3** ( $= \mathbf{S4} + \Box(\Box A \supset B) \vee \Box(\Box B \supset A)$ ) in [Zeman, 1973] results from **LS4** by the addition of the *axiomatic sequent*

$$\Box(A \vee \Box B), \Box(\Box A \vee B) \rightarrow \Box A, \Box B.$$

Shimura [1991] obtains a cut-free sequent system **LS4.3** by adding to **LCPL** the rules  $(\Box \rightarrow)_0$  and

$$(\rightarrow \Box)_5 \quad \Box\Gamma \rightarrow (\Box\Delta) \setminus \{\Box A_1\} \dots \Box\Gamma \rightarrow (\Box\Delta) \setminus \{\Box A_n\} \vdash \Box\Gamma \rightarrow \Box\Delta,$$

where  $\Delta = \{A_1, \dots, A_n\}$  and  $\setminus$  is set-theoretic difference.

## 1.2 Ordinary Gentzen systems for normal temporal logics

The syntax of temporal propositional logic is given by:

$$\begin{aligned} A ::= & p \mid \mathbf{t} \mid \mathbf{f} \mid \neg A \mid A \wedge B \mid A \vee B \mid A \supset B \mid A \equiv \\ & B \mid \langle P \rangle A \mid [P]A \mid \langle F \rangle A \mid [F]A. \end{aligned}$$

Also a number of normal temporal propositional logics have been presented as ordinary sequent calculi. Nishimura [1980], for example, defines sequent systems **LKt** and **LK4t** for the minimal normal temporal logic **Kt** and the tense-logical counterpart **K4t** of **K4**. The sequent calculus **LKt** comprises the following introduction rules for forward-looking necessity  $[F]$  (“always in the future”) and backward-looking necessity  $[P]$  (“always in the past”):<sup>3</sup>

$$\begin{aligned} (\rightarrow [F]) \quad & \Gamma \rightarrow A, [P]\Delta \vdash [F]\Gamma \rightarrow [F]A, \Delta; \\ (\rightarrow [P]) \quad & \Gamma \rightarrow A, [F]\Delta \vdash [P]\Gamma \rightarrow [P]A, \Delta, \end{aligned}$$

where  $[F]\Delta = \{[F]A \mid A \in \Delta\}$  and  $[P]\Delta = \{[P]A \mid A \in \Delta\}$ . In **K4t**, these rules are replaced by the following pair of rules:

$$\begin{aligned} (\rightarrow [F])_4 \quad & [F]\Gamma, \Gamma \rightarrow A, [P]\Delta, [P]\Sigma \vdash [F]\Gamma \rightarrow [F]A, \Delta, [P]\Sigma; \\ (\rightarrow [P])_4 \quad & [P]\Gamma, \Gamma \rightarrow A, [F]\Delta, [F]\Sigma \vdash [P]\Gamma \rightarrow [P]A, \Delta, [F]\Sigma. \end{aligned}$$

<sup>3</sup>Nishimura allows infinite sets in antecedent and succedent position. It is proved, however, that if a sequent  $\Gamma \rightarrow \Delta$  is provable, then there are finite sets  $\Gamma' \subseteq \Gamma$  and  $\Delta' \subseteq \Delta$  such that the sequent  $\Gamma' \rightarrow \Delta'$  is provable.

In both systems,  $\langle P \rangle$  (“sometimes in the past”) and  $\langle F \rangle$  (“sometimes in the future”) are treated not as primitive but as defined by  $\langle P \rangle A := \neg[P]\neg A$  and  $\langle F \rangle A := \neg[F]\neg A$ . Note also that this approach gives completely parallel rules for  $[F]$  and  $[P]$  and that these rules do not exploit the interrelation between the backward and the forward-looking modalities, that shows up, for instance, in the provability of  $A \supset [F]\langle P \rangle A$  and  $A \supset [P]\langle F \rangle A$ .

In summary, it may be said that many normal modal and temporal logics are presentable as ordinary Gentzen calculi, and that in some cases suitable constraints on the structures exhibited in the statement of the sequent rules for the modal operators allow for a number of variations. However, no uniform way of presenting only the most important normal modal and temporal propositional logics as ordinary Gentzen calculi is known. Further, the standard approach fails to be *modular*: in general it is not the case that a single axiom schema is captured by a single sequent rule (or a finite set of such rules). In the following section a more philosophical critique of ordinary Gentzen systems is advanced.

### 1.3 Introduction schemata and the meaning of the logical operations

The philosophical (and methodological) problems with applying the notion of a Gentzen sequent to modal logics have to do with the idea of *defining* the logical operations by means of introduction schemata (together with structural assumptions about derivability formulated in terms of structural rules). This ‘anti-realistic’ conception of the meaning of the logical operations is often traced back to a certain passage on natural deduction from Gentzen’s *Investigations into Logical Deduction* [Gentzen, 1934, p. 80]:

[I]ntroductions represent, as it were, the ‘definitions’ of the symbols concerned, and the eliminations are no more, in the final analysis, than the consequences of these definitions.

To qualify as a definition of a logical operation, an introduction schema must satisfy certain adequacy criteria. Such conditions are discussed, for instance, by Hacking [1994]. Following Hacking, if introduction rules are to be regarded as defining logical operations, these rules must be such that the structural rules monotonicity (also called weakening, thinning, or dilution), reflexivity, and cut can be eliminated. Hacking claims that

[i]t is not provability of cut-elimination that excludes modal logic, but dilution-elimination . . . The serious modal logics such as **T**, **S4** and **S5** have cut-free sequent-calculus formalizations, but the rules place restrictions on side formulas. Gentzen’s rules for sentential connections are all ‘local’ in that they concern



only the components from which the principal formula is built up, and place no restrictions on the side formulas. Gentzen's own first-order rules, though not strictly local, are equivalent to local ones. That is why dilution-elimination goes through for first-order logic but not for modal logics ([Hacking, 1994, p. 24]).

By dilution-elimination Hacking means that the monotonicity rules

$$\Delta \rightarrow \Gamma \vdash \Delta, A \rightarrow \Gamma, \quad \Delta \rightarrow \Gamma \vdash \Delta \rightarrow \Gamma, A$$

may be replaced by atomic thinning rules

$$\Delta \rightarrow \Gamma \vdash \Delta, p \rightarrow \Gamma, \quad \Delta \rightarrow \Gamma \vdash \Delta \rightarrow \Gamma, p.$$

without changing the set of provable sequents. Similarly, reflexivity-elimination amounts to replaceability of  $\vdash A \rightarrow A$  by  $\vdash p \rightarrow p$ . The term “cut-elimination” is reserved for something stronger than replaceability of cut by the atomic cut-rule

$$\Delta \rightarrow \Sigma, p; \quad \Gamma, p \rightarrow \Theta \vdash \Gamma, \Delta \rightarrow \Sigma, \Theta.$$

A cut-elimination proof shows the admissibility of cut: the rule has no effect on the set of provable sequents.

The introduction rules for  $\Box$  in **LS4** prevent dilution-elimination. Obviously, the sequent  $\Box B, \Box A \rightarrow \Box A$ , for example, cannot be proved using only these rules and atomic thinning. A problem with the requirement of dilution-elimination is the weak status monotonicity has acquired as a defining characteristic of logical deduction. In view of the substantial work on relevance logic, many other substructural logics, and a plethora of non-monotonic reasoning formalisms extending a monotonic base system, monotonicity of inference is not generally viewed as a touchstone of logicity anymore. Moreover, also reflexivity and cut have been questioned. Unrestricted transitivity of deduction as expressed by the cut-rule does not hold, for instance, in Tennant's intuitionistic relevant logic [1994], and both reflexivity and cut fail to be validated by Update-to-Test semantic consequence as defined in Dynamic Logic, see [van Benthem, 1996]. Reflexivity-elimination and cut-elimination are, however, important. According to Belnap [1982, p. 383], the provability of  $A \rightarrow A$  constitutes

half of what is required to show that the “meaning” of formulas . . . is not context-sensitive, but that instead formulas “mean the same” in both antecedent and consequent position. (The [Cut] Elimination Theorem . . . is the other half of what is required for this purpose).

A similar remark can be found in [Girard, 1989, p. 31]. Cut-elimination is indispensable, because it amounts to the familiar non-creativity requirement

for definitions (see, for instance, [Hacking, 1994], [von Kutschera, 1968]). If one adds introduction rules for a (finitary) operation  $f$  to a sequent calculus, this addition ought to be conservative, so that in the extended formalism, every proof of an  $f$ -free formula  $A$  is convertible into a proof of  $A$  without any application of an introduction rule for  $f$ .

There are other reasons why the eliminability of cut is a desirable property. Usually, cut-elimination implies the subformula property: every cut-free proof of a sequent  $s$  contains only subformulas of formulas in  $s$ . In sequent calculi for decidable logics, the subformula property can often be used to give a syntactic proof of decidability. According to Sambin and Valentini [1982, p. 316], it

is usually not difficult to choose suitable [sequent] rules for each modal logic if one is content with completeness of rules. The real problem however is to find a set of rules also satisfying the subformula-property.

The sequent calculi for **S5** in [Mints, 1970], [Sato, 1977], and [Sato, 1980], although admitting cut-elimination, do not have the subformula property. In a sequent calculus with an enriched structural language, the subformula property need not be accompanied by a *substructure property*. In such systems the subformula property for the logical vocabulary need neither imply nor be of direct use for syntactic decidability proofs. Avron [1996, p. 2] requires of a decent sequent calculus simplicity of the structures employed and a ‘real’ subformula property. But even without the substructure property, the subformula property may be useful, for instance in proving conservative extension results, see also Section 3.8.

It is well-known that cut-elimination itself does not guarantee efficient proof search (see [D’Agostino and Mondadori, 1994], [Boolos, 1984]), so that it may be attractive to work with an analytic, subformula property preserving cut-rule, if possible. An application of cut

$$\Delta \rightarrow \Sigma, A \quad \Gamma, A \rightarrow \Theta \vdash \Gamma, \Delta \rightarrow \Sigma, \Theta$$

is *analytic* (see [Smullyan, 1968]), if the cut-formula  $A$  is a subformula of some formula in the conclusion sequent  $\Gamma, \Delta \rightarrow \Sigma, \Theta$ . Let  $\text{Sub}(\Delta)$  denote the set of all subformulas of formulas in  $\Delta$ . Applications of the sequent rules

$$\begin{aligned} (\rightarrow \Box)_B & \quad \Gamma \rightarrow \Box\Theta, A \vdash \Box\Gamma \rightarrow \Theta, \Box A \\ (\rightarrow \Box)_{ABE} & \quad \Gamma, \Box\Gamma \rightarrow \Box\Theta, \Box\Delta, A \vdash \Box\Gamma \rightarrow \Box\Theta, \Delta, \Box A \\ \text{and } (\Box \rightarrow)_D & \quad \Gamma \rightarrow \Box\Delta \vdash \Box\Gamma \rightarrow \Delta \end{aligned}$$

may be said to be analytic if  $\Box\Theta \subseteq \text{Sub}(\Gamma \cup \{A\})$ ,  $\Box\Delta \subseteq \text{Sub}(\Box\Gamma \cup \Box\Theta \cup \{A\})$ , and  $\Box\Delta \subseteq \text{Sub}(\Gamma)$ , respectively. Takano [1992] shows that the cut-rule in

**LS5\***, **LKB**, **LKTB**, **LKDB**, **LK4B**, **LKt** and **LK4t** can be replaced by the analytic cut-rule: every proof in these sequent calculi can be transformed into a proof of the same sequent such that every application of cut (and, moreover, every application of the rules  $(\rightarrow \Box)_B$ ,  $(\rightarrow \Box)_{4BE}$ , and  $(\rightarrow \Box)_D$ ) in this proof is analytic.

Although admissibility of analytic cut is a welcome property, in general, unrestricted cut-elimination is to be preferred over elimination of analytic cut. Admissibility of cut has great conceptual significance. The cut-rule justifies certain substitutions of data; in particular it justifies the use of previously proved formulas. Moreover, if the cut-rule is assumed, the non-creativity requirement for definitions implies that cut must be eliminable.

There are other nice properties of introduction schemata as definitions in addition to enabling cut-elimination and reflexivity-elimination. The assignment of meaning to the logical operations should, for instance, be non-holistic, and hence sequent rules like the above  $(\rightarrow \Box)'$  and  $(\diamond \rightarrow)'$  are unsuitable. If (the statement of) an introduction rule for a logical operation  $f$  exhibits no connective other than  $f$ , the rule is called *separated*, see [Zucker and Tragesser, 1978]. An even stronger condition is *segregation*, requiring that the antecedent (succedent) of the conclusion sequent in a left (right) introduction rule must not exhibit any structure operation. Segregation has been suggested (although not under this name) by Belnap [1996] who explains that

[t]he nub is this. If a rule for  $\supset$  only shows how  $A \supset B$  behaves *in context*, then that rule is not *merely* explaining the meaning of  $\supset$ . It is also and inextricably explaining the meaning of the context. Suppose we give sufficient conditions for

$$A \supset B, \Delta \rightarrow \Gamma$$

in part by the rule

$$\frac{\Delta \rightarrow A \quad B \rightarrow \Gamma}{A \supset B, \Delta \rightarrow \Gamma}$$

Then we are not explaining  $A \supset B$  alone. We are simultaneously involving the comma not just in our explicans (that would surely be all right), but in our explicandum. We are explaining two things at once. There is no way around this. You do not have to take it as a defect, but it is a fact. ... If you are a 'holist', probably you will not care; but then there is not much about which holists much care. [Belnap, 1996, p. 81 f.] (notation adjusted)

Moreover, the rules for  $f$  may be required to be *weakly symmetrical* in the sense that every rule should either belong to a set of rules  $(f \rightarrow)$  which introduce  $f$  on the left side of  $\rightarrow$  in the conclusion sequent or to a set of rules  $(\rightarrow f)$  which introduce  $f$  on the right side of  $\rightarrow$  in the conclusion sequent. The introduction rules for  $f$  are called *symmetrical*, if they are weakly symmetrical and both  $(\rightarrow f)$  and  $(f \rightarrow)$  are non-empty. The sequent rules for  $f$  are called *weakly explicit*, if the rules  $(\rightarrow f)$  and  $(f \rightarrow)$  exhibit  $f$  in their conclusion sequents only, and they are called *explicit*, if in addition to being weakly explicit, the rules in  $(\rightarrow f)$  and  $(f \rightarrow)$  exhibit only one occurrence of  $f$  on the right, respectively the left side of  $\rightarrow$ . Separation, symmetry, and explicitness of the rules imply that in a sequent calculus for a given logic  $\Lambda$ , every connective that is explicitly definable in  $\Lambda$  also has separate, symmetrical, and explicit introduction rules. These rules can be found by decomposition of the defined connective, if it is assumed that the deductive role of  $f(A_1, \dots, A_n)$  only depends on the deductive relationships between  $A_1, \dots, A_n$ . It is therefore desirable to have introduction rules for  $\Box$ ,  $\Diamond$ ,  $\langle P \rangle$ ,  $[P]$ ,  $\langle F \rangle$  and  $[F]$  as primitive operations, so that the familiar mutual definitions are derivable.

A further desirable property, reminiscent of implicit definability in predicate logic, is the unique characterization of  $f$  by its introduction rules. Suppose that  $\Lambda$  is a logical system with a syntactic presentation  $S$  in which  $f$  occurs. Let  $S^*$  be the result of rewriting  $f$  everywhere in  $S$  as  $f^*$ , and let  $\Lambda\Lambda^*$  be the system presented by the union  $SS^*$  of  $S$  and  $S^*$  in the combined language with both  $f$  and  $f^*$ . Let  $A_f$  denote a formula (in this language) that contains a certain occurrence of  $f$ , and let  $A_{f^*}$  denote the result of replacing this occurrence of  $f$  in  $A$  by  $f^*$ . The connectives  $f$  and  $f^*$  are said to be *uniquely characterized* in  $\Lambda\Lambda^*$  iff for every formula  $A_f$  in the language of  $\Lambda\Lambda^*$ ,  $A_f$  is provable in  $SS^*$  iff  $A_{f^*}$  is provable in  $SS^*$ . Došen [1985] has proved that unique characterization is a non-trivial property and that the connectives in his higher-level systems  $S4p/D$  and  $S5p/D$  for **S4** and **S5**, respectively, are uniquely characterized.

As we have seen, the standard sequent-style proof-theory for normal modal and temporal logic fails to be modular. The idea that modularity can be achieved by systematically varying structural features of the derivability relation while keeping the introduction rules for the logical operations untouched can be traced back to Gentzen [1934] and has been referred to as Došen's Principle in [Wansing, 1994]. In [Došen, 1988, p. 352], Došen suggests that "the rules for the logical operations are never changed: all changes are made in the structural rules." This methodology is adopted, for example in Došen's [1985] higher-level sequent systems for **S4** and **S5**, Blamey and Humberstone's [1991] higher-arity sequent calculi for certain extensions of **K**, Nishimura's [1980] higher-arity sequent systems for **Kt** and **K4t**, and the presentation of normal modal and temporal logics as cut-free display sequent calculi.

Another methodological aspect is generality. Is there a type of sequent system that allows not only a uniform treatment of the most important modal and temporal logics but also a treatment of substructural logics, other non-classical logics and systems combining operations from different families of logics and that, moreover, is rich enough to suggest important, hitherto unexplored logics? The framework of display logic to be presented in the next section has been devised explicitly as an instrument for combining logics (see [Belnap, 1982]), and has been suggested, for example, as a tool for defining subsystems of classical predicate logic (see [Wansing, 1999]). In addition to generality, a ‘real’ subformula property, and Došen’s principle, Avron [1996] requires of a good sequent calculus framework also *semantics independence*. The framework should not be so closely tied to a particular semantics that one can more or less read off the semantic structures in question. Moreover, the proof systems instantiating the framework should lead to a better understanding of the respective logics and the differences between them.

Note that each of the ordinary sequent systems presented in the present section fails to satisfy some of the more philosophical requirements mentioned. The same holds true for the ordinary sequent systems for various non-normal, classical modal logics investigated in [Lavendhomme and Lucas, 2000]. There are thus not only technical but also methodological and philosophical reasons for investigating generalizations of the notion of a Gentzen sequent.

## 2 GENERALIZED SEQUENT SYSTEMS

In this section the application of a number of generalizations of the ordinary notion of sequent to normal modal propositional and temporal logics is surveyed.

### 2.1 Higher-level sequent systems

Došen [1985] has developed certain non-standard sequent systems for **S4** and **S5**. In these Gentzen-style systems one is dealing with sequents of arbitrary finite level. Sequents of level 1 are like ordinary sequents, whereas sequents of level  $n + 1$  ( $0 < n < \omega$ ) have finite sets of sequents of level  $n$  on both sides of the sequent arrow. The main sequent arrow in a sequent of level  $n$  carries the superscript  $n$ , and  $\emptyset$  is regarded as a set of any finite level. The rules for logical operations are presented as *double-line* rules. A double-line rule

$$\frac{\underline{\underline{s_1, \dots, s_n}}}{s_0}$$

involving sequents  $s_0, \dots, s_n$ , denotes the rules

$$\frac{s_1, \dots, s_n}{s_0}, \frac{s_0}{s_1}, \dots, \frac{s_0}{s_n}.$$

Došen gives the following double-line sequent rules for  $\Box$  and  $\Diamond$ :

$$\frac{X + \{\emptyset \rightarrow^1 \{A\}\} \rightarrow^2 X_2 + \{X_3 \rightarrow^1 X_4\}}{X_1 \rightarrow^2 X_2 + \{X_3 + \{\Box A\} \rightarrow^1 X_4\}}$$

$$\frac{X_1 + \{\{A\} \rightarrow^1 \emptyset\} \rightarrow^2 X_2 + \{X_3 \rightarrow^1 X_4\}}{X_1 \rightarrow^2 X_2 + \{X_3 \rightarrow^1 X_4 + \{\Diamond A\}\}},$$

where  $+$  refers to the union of disjoint sets. If these rules are added to Došen's higher-level sequent calculus  $\text{Cp}/D$  for **CPL**, this results in the sequent system  $\text{S5p}/D$  for **S5**. The sequent calculus  $\text{S4p}/D$  for **S4** is then obtained by imposing a structural restriction on the monotonicity rule of level 2:

$$X \rightarrow^2 Y \vdash X \cup Z_1 \rightarrow^2 Y \cup Z_2.$$

The restriction is this: if  $Y = \emptyset$ , then  $Z_2$  must be a singleton or empty; if  $Y \neq \emptyset$ , then  $Z_2$  must be empty. If the same restriction is applied to monotonicity of level 1 in  $\text{Cp}/D$ , then this gives a higher-level sequent system for intuitionistic propositional logic **IPL**.

Note that  $\Diamond$  and  $\Box$  are interdefinable in  $\text{S4p}/D$  and  $\text{S5p}/D$ . The double-line rules for  $\Box$  and  $\Diamond$ , however, do not satisfy weak symmetry and weak explicitness, but the upward directions of these rules can be replaced by:

$$\emptyset \rightarrow^1 \{A\} \vdash \emptyset \rightarrow^1 \{\Box A\} \quad \text{and} \quad \{A\} \rightarrow^1 \emptyset \vdash \{\Diamond A\} \rightarrow^1 \emptyset.$$

Whereas cut can be eliminated at levels 1 and 2, cut of all levels fails to be eliminable [Došen, 1985, Lemma 1]. Moreover, in Došen's higher-level framework it is not clear how restrictions similar to the one used to obtain  $\text{S4p}/D$  from  $\text{S5p}/D$  would allow to capture further axiomatic systems of normal modal propositional logic.

## 2.2 Higher-dimensional sequent systems

A 'higher-dimensional' proof theory for modal logics has been developed by Masini [1992; 1996]. This approach is based on the notion of a *2-sequent*. In order to define this notion, various preparatory definitions are useful. Any finite sequence of modal formulas is called a 1-sequence. The empty 1-sequence is denoted by  $\epsilon$ . A 2-sequence is an infinite 'vertical' succession of 1 sequences,  $\Gamma = \{\alpha_i\}_{0 < i < \omega}$  such that  $\exists j \geq 1, \forall k \geq j : \alpha_k = \epsilon$ . For each  $i$ ,  $\alpha_i$  is said to be at level  $i$ . The depth of  $\Gamma$  ( $\text{hp}\Gamma$ ) is defined as  $\min\{i \mid i \geq 0, \forall k > i : \alpha_k = \epsilon\}$ . A *2-sequent* is an expression  $\Gamma \rightarrow \Delta$ , where  $\Gamma$  and  $\Delta$

are 2-sequences. The depth of  $\Gamma \rightarrow \Delta$  ( $\natural(\Gamma \rightarrow \Delta)$ ) is defined as  $\max(\natural\Gamma, \natural\Delta)$ . If  $\Gamma \rightarrow \Delta$  is a 2-sequent and  $A$  an occurrence of a modal formula in  $\Gamma \rightarrow \Delta$ , then  $A$  is said to be maximal in  $\Gamma \rightarrow \Delta$ , if  $A$  is at level  $k$  in  $\Gamma$  or in  $\Delta$  and  $k = \natural(\Gamma \rightarrow \Delta)$ .  $A$  is the maximum in  $\Gamma \rightarrow \Delta$ , if  $A$  is the unique maximal formula in  $\Gamma \rightarrow \Delta$ . The sequent rules for  $\Box$  are based on the idea of “internalizing the level structure of 2-sequents” [Masini, 1992, p. 231]:

$$\begin{array}{c}
 \begin{array}{c}
 \Gamma \\
 \alpha \\
 \beta, A \quad \rightarrow \Delta \\
 \Gamma' \\
 \hline
 \Gamma \\
 \alpha, \Box A \quad \rightarrow \Delta \\
 \beta \\
 \Gamma'
 \end{array}
 \quad
 (\Box \rightarrow)
 \quad
 \begin{array}{c}
 \Delta \\
 \Gamma \rightarrow \mu \\
 A \\
 \hline
 \Gamma \rightarrow \Delta \\
 \mu, \Box A
 \end{array}
 \quad
 (\rightarrow \Box)
 \end{array}$$
  

$$\begin{array}{c}
 \begin{array}{c}
 \Gamma \\
 \alpha \quad \rightarrow \Delta \\
 A \\
 \hline
 \Gamma \\
 \alpha, \Diamond A \quad \rightarrow \Delta
 \end{array}
 \quad
 (\Diamond \rightarrow)
 \quad
 \begin{array}{c}
 \Delta \\
 \Gamma \rightarrow \mu \\
 \pi, A \\
 \Delta' \\
 \hline
 \Gamma \rightarrow \Delta \\
 \mu, \Diamond A \\
 \pi \\
 \Delta'
 \end{array}
 \quad
 (\rightarrow \Diamond)
 \end{array}$$

where  $\alpha$ ,  $\beta$ ,  $\pi$ , and  $\mu$  denote arbitrary 1-sequences, and  $A$  must be the maximum of the premise 2-sequent in  $(\rightarrow \Box)$  and  $(\Diamond \rightarrow)$ . According to Masini, these introduction rules give rise to a “general basic proof theory of modalities” [Masini, 1992, p. 232]. If added to a 2-sequent calculus for **CPL**, the above rules result, however, in a sequent calculus for **KD** instead of the basic system **K**. This sequent system for **KD** admits cut-elimination,  $\Box$  and  $\Diamond$  are interdefinable, and the introduction rules are separate, symmetrical, and explicit, but no indication is given of how to present axiomatic extensions of **KD** as higher-dimensional sequent systems. Moreover, it is not clear how Masini’s framework may be modified in order to obtain a 2-sequent calculus for **K**.

### 2.3 Higher-arity sequent systems

In search of generalizations of the standard Gentzen-style sequent format, it is a natural move to consider consequence relations with an arity greater than 2. It seems that the first higher-arity sequent calculus was formulated by Schröter [1955], see also [Gottwald, 1989]. This formalism is a natural generalization of Gentzen’s sequent calculus for **CPL** to truth-functional

$n$ -valued logic. The intended truth-functional reading of a Gentzen sequent  $s = \Delta \rightarrow \Gamma$  is given by a translation  $\sigma$  of  $s$  into a formula:

$$\sigma(\Delta \rightarrow \Gamma) = \bigwedge \Delta \supset \bigvee \Gamma,$$

The sequent  $s$  thus is true under a given interpretation if either some formula in  $\Delta$  is false, or some formula in  $\Gamma$  is true, and the two places of the sequent arrow correspond to the two truth-values of classical logic. In general, in  $n$ -valued logic (with  $2 \leq n$ ) one obtains  $n$ -place sequents  $s = \Delta_1; \Delta_2; \dots; \Delta_n$ , with the understanding that  $s$  is true under an interpretation if for every  $i \leq n$ , some formula in  $\Delta_i$  has truth-value  $i$ ; for a comprehensive treatment of sequent calculi for truth-functional many-valued logics see [Zach, 1993]. We shall here briefly review some relevant parts of the work of Blamey and Humberstone [1991], who investigate an application of three-place and, ultimately, four-place sequent arrows to normal modal logic. This approach is congenial to display logic with respect to a realization of the Došen-Principle insofar as Blamey and Humberstone emphasize that distinctions between various well-known normal modal logics can “be reflected at the purely structural level, if an appropriate notion of *sequent*” is adopted [Blamey and Humberstone, 1991, p. 763]. Let  $\Gamma, \Delta, \Theta$ , and  $\Sigma$  range over finite sets of formulas in the modal propositional language with  $\Box$  as primitive. The four-place sequent

$$\Gamma \rightarrow_{\Sigma}^{\Theta} \Delta$$

has the following heuristic reading:

$$(\bigwedge \Gamma \wedge \bigwedge \Box \Sigma) \supset (\bigvee \Delta \vee \bigvee \Box \Theta).$$

This kind of sequent had independently been used by Sato [1977], where a cut-free sequent calculus for **S5** is presented containing a left introduction rule for  $\Box$  that fails to be weakly explicit. Blamey’s and Humberstone’s introduction rules for  $\Box$  are:

$$(\Box \downarrow)_0 \quad \vdash \emptyset \rightarrow_A^{\emptyset} \Box A \quad (\Box \uparrow)_0 \quad \vdash \Box A \rightarrow_{\emptyset}^A \emptyset.$$

In order to obtain a sequent calculus for **K** the following structural rules are assumed:

$$\begin{aligned} (R) \quad & \vdash A \rightarrow_{\emptyset}^{\emptyset} A & (\text{vertical } R) \quad & \vdash \emptyset \rightarrow_A^A \emptyset \\ (M) \quad & \Gamma \rightarrow_{\Sigma}^{\Theta} \Delta \vdash \Gamma, \Gamma' \rightarrow_{\Sigma, \Sigma'}^{\Theta, \Theta'} \Delta, \Delta' \\ (\text{undercut}) \quad & \Sigma \rightarrow_{\emptyset}^{\emptyset} A \quad \Gamma \rightarrow_{\Sigma', A}^{\Theta} \Delta \vdash \Gamma \rightarrow_{\Sigma, \Sigma'}^{\Theta} \Delta \\ (T) \quad & \Gamma, A \rightarrow_{\Sigma}^{\Theta} \Delta \quad \Gamma \rightarrow_{\Sigma}^{\Theta} A, \Delta \vdash \Gamma \rightarrow_{\Sigma}^{\Theta} \Delta \\ (\text{vertical } T) \quad & \Gamma \rightarrow_{\Sigma, A}^{\Theta} \Delta \quad \Gamma \rightarrow_{\Sigma}^{\Theta, A} \Delta \vdash \Gamma \rightarrow_{\Sigma}^{\Theta} \Delta. \end{aligned}$$



Against the background of these rules, the introduction rules  $(\Box \downarrow)_0$  and  $(\Box \uparrow)_0$  are interplaceable with the following rules, respectively:

$$\begin{aligned} (\Box \downarrow) \quad & \Gamma, \Box A \rightarrow_{\Sigma}^{\Theta} \Delta \vdash \Gamma \rightarrow_{\Sigma, A}^{\Theta} \Delta \\ (\Box \uparrow) \quad & \Gamma \rightarrow_{\Sigma, A}^{\Theta} \Delta \vdash \Gamma, \Box A \rightarrow_{\Sigma}^{\Theta} \Delta. \end{aligned}$$

The introduction rules for the Boolean operations are adaptations of the familiar rules to the higher-arity case. Here is a simple example of a derivation in this formalism (using some obvious notational simplifications):

$$\frac{\frac{\frac{A \wedge B \rightarrow A}{A \wedge B \rightarrow A, B} \quad \frac{\Box A, \Box B \rightarrow \Box A \wedge \Box B}{\Box A \rightarrow_B \Box A \wedge \Box B} (\Box \downarrow)}{\emptyset \rightarrow_{A, B} \Box A \wedge \Box B} (\text{undercut}) \text{ twice}}{\frac{\emptyset \rightarrow_{A \wedge B} \Box A \wedge \Box B}{\Box(A \wedge B) \rightarrow \Box A \wedge \Box B} (\Box \uparrow)}$$

The axiom schemata  $D$ ,  $T$ ,  $\not\downarrow$ , and  $B$  are captured by purely structural rules not exhibiting any logical operations:

$$\begin{aligned} D \quad & \Sigma \rightarrow_{\emptyset}^{\emptyset} \emptyset \vdash \emptyset \rightarrow_{\Sigma}^{\emptyset} \emptyset \\ T \quad & \vdash \emptyset \rightarrow_A^{\emptyset} A \\ \not\downarrow \quad & \Sigma \rightarrow_{\Sigma}^{\Theta} A \quad \Gamma \rightarrow_{\Sigma', A}^{\Theta} \Delta \vdash \Gamma \rightarrow_{\Sigma, \Sigma'}^{\Theta} \Delta \\ B \quad & \Sigma \rightarrow_{\emptyset}^{\Delta} A \quad \Gamma \rightarrow_{\Sigma, A}^{\Theta} \Delta \vdash \Gamma \rightarrow_{\Sigma}^{\Theta} \Delta. \end{aligned}$$

Since Blamey and Humberstone are primarily interested in semantical aspects of their sequent systems, they do not consider cut-elimination. Although their calculi satisfy Došen's Principle, it remains unclear whether their approach is fully modular for the most important systems of normal modal propositional logic. They do not present a structural equivalent of the 5-axiom schema, but rather treat **S5** as **KTB4**.

In [Nishimura, 1980], Nishimura uses six-place sequents

$$\Theta_1; \Gamma; \Theta_2 \rightarrow \Sigma_1; \Delta; \Sigma_2.$$

These higher-arity sequents can intuitively be read as follows:

$$(\bigwedge [P]\Theta_1 \wedge \bigwedge \Gamma \wedge \bigwedge [F]\Theta_2) \supset (\bigvee [P]\Sigma_1 \vee \bigvee \Delta \vee \bigvee [F]\Sigma_2).$$

Nishimura defines introduction rules for the tense logical operations  $[F]$  and  $[P]$ , which are explicit in the sense of Section 1.3:

$$\begin{aligned} (\rightarrow [F])' \quad & \frac{\Theta_1; \Gamma; \Theta_2 \rightarrow \Sigma_1; \Delta; A; \Sigma_2}{\Theta_1; \Gamma; \Theta_2 \rightarrow \Sigma_1; \Delta; [F]A; \Sigma_2} \\ ([F] \rightarrow)' \quad & \frac{\Theta_1; \Gamma; A; \Theta_2 \rightarrow \Sigma_1; \Delta; \Sigma_2}{\Theta_1; \Gamma; [F]A; \Theta_2 \rightarrow \Sigma_1; \Delta; \Sigma_2} \\ (\rightarrow [P])' \quad & \frac{\Theta_1; \Gamma; \Theta_2 \rightarrow \Sigma_1; A; \Delta; \Sigma_2}{\Theta_1; \Gamma; \Theta_2 \rightarrow \Sigma_1; [P]A, \Delta \Sigma_2} \\ ([P] \rightarrow)' \quad & \frac{\Theta_1, A; \Gamma; \Theta_2 \rightarrow \Sigma_1; \Delta; \Sigma_2}{\Theta_1; [P]A, \Gamma; \Theta_2 \rightarrow \Sigma_1; \Delta; \Sigma_2} \end{aligned}$$

In accordance with the Došen Principle, these rules are held constant in sequent systems for **Kt** and **K4t**. The difference between these logics is accounted for by different structural rules, namely

$$\begin{array}{c} \text{(r-trans)} \quad \frac{\emptyset; \Gamma; \emptyset \rightarrow \Delta; A; \emptyset}{\emptyset; \emptyset; \Gamma \rightarrow \emptyset; \Delta; A} \quad \text{(l-trans)} \quad \frac{\emptyset; \Gamma; \emptyset \rightarrow \emptyset; A; \Delta}{\Gamma; \emptyset; \emptyset \rightarrow A; \Delta; \emptyset} \end{array}$$

in the case of **Kt** and

$$\text{(r-trans)}_4 \quad \frac{\emptyset; \Gamma; \Gamma \rightarrow \Delta, \Sigma; A; \emptyset}{\emptyset; \emptyset; \Gamma \rightarrow \Sigma; \Delta; A} \quad \text{(l-trans)}_4 \quad \frac{\Gamma; \Gamma; \emptyset \rightarrow \emptyset; A; \Delta, \Sigma}{\Gamma; \emptyset; \emptyset \rightarrow A; \Delta; \Sigma}$$

in the case of **K4t**. Nishimura observes that although in the introduction rules for  $\langle F \rangle$  and  $\langle P \rangle$  subformulas are preserved from premise sequent to conclusion sequent, cut-elimination fails to hold in the six-place sequent systems for **Kt** and **K4t**. There is, for instance, no cut-free proof of  $; p; \rightarrow ; [F] \neg [P] \neg p$ .<sup>4</sup>

## 2.4 Multiple-sequent systems

Indrzejczak, in [1997; 1998], suggested non-standard sequent systems for certain extensions of the minimal regular modal logic **C** using three sequent arrows  $\rightarrow$ ,  $\Box \rightarrow$ , and  $\Diamond \rightarrow$ . These sequent arrows denote binary relations between finite sets of  $S$ -formulas, where the set of  $S$ -formulas is defined as the union of the set of modal formulas and  $\{-A \mid A \text{ is a modal formula}\}$ . As before, we shall use  $A, B, C, \dots$  to denote modal formulas. The symbol ‘ $\neg$ ’ is a unary structure connective that may not be nested, and the sequent arrows  $\Box \rightarrow$  and  $\Diamond \rightarrow$  are auxiliary in the sense that they fail to represent consequence relations, because (in general) neither  $\vdash A \Box \rightarrow A$  nor  $\vdash A \Diamond \rightarrow A$ . The logics presented by such multiple-sequent systems are given by the set of provable sequents  $\Delta \rightarrow \Gamma$ . The intended meaning of a sequent is captured by a translation  $\sigma$  from sequents into ordinary sequents using a translation  $\delta$  from  $S$ -formulas to modal formulas. For every modal formula  $A$ ,  $\delta(\neg A) := \neg A$  and  $\delta(A) := A$ . The translation  $\sigma$  is defined as follows:

$$\begin{aligned} \sigma(\Gamma \rightarrow \Delta) &= \bigwedge \delta(\Gamma) \rightarrow \bigvee \delta(\Delta) \\ \sigma(\Gamma \Box \rightarrow \Delta) &= \bigwedge \delta(\Gamma) \rightarrow \Box \bigvee \delta(\Delta) \\ \sigma(\Gamma \Diamond \rightarrow \Delta) &= \Diamond \bigwedge \delta(\Gamma) \rightarrow \bigvee \delta(\Delta) \end{aligned}$$

Here  $\delta(\Gamma) := \{A \mid A \in \Gamma\} \cup \{\neg A \mid \neg A \in \Gamma\}$ . For every modal formula  $A$ ,  $A^*$  is defined as  $\neg A$  and  $\neg A^*$  as  $A$ . If  $\Delta$  is a set of  $S$ -formulas,  $\Delta^* :=$

<sup>4</sup>Note that Nishimura allows infinite sets in antecedent and succedent position. It is, however, shown that if a sequent  $\Theta_1; \Gamma; \Theta_2 \rightarrow \Sigma_1; \Delta; \Sigma_2$  is provable, then there are finite sets  $\Theta'_i \subseteq \Theta_i$ ,  $\Sigma'_i \subseteq \Sigma_i$ , ( $i = 1, 2$ ),  $\Gamma' \subseteq \Gamma$ , and  $\Delta' \subseteq \Delta$  such that the sequent  $\Theta'_1; \Gamma'; \Theta'_2 \rightarrow \Sigma'_1; \Delta'; \Sigma'_2$  is provable.

$\{A \mid \neg A \in \Delta\} \cup \{\neg A \mid A \in \Delta\}$ . Let  $(\rightarrow)$  be any of  $\rightarrow, \Box\rightarrow, \Diamond\rightarrow$ . The following reflexivity and monotonicity rules are assumed:

$$\vdash A \rightarrow A; \quad \Delta (\rightarrow) \Gamma \vdash \Delta (\rightarrow) \Gamma, A; \quad \Delta (\rightarrow) \Gamma \vdash \Delta, A (\rightarrow) \Gamma.$$

Next, there are further structural rules called shifting rules:

$$\begin{array}{ll} [\rightarrow^*] & A, \Delta \rightarrow \Gamma \vdash \Delta \rightarrow \Gamma, A^* \quad [* \rightarrow] \quad \Delta \rightarrow \Gamma, A \vdash \Delta, A^* \rightarrow \Gamma \\ [\text{TR}] & \Delta \Box\rightarrow \Gamma \vdash \Gamma^* \Diamond\rightarrow \Delta^* \quad \Delta \Diamond\rightarrow \Gamma \vdash \Gamma^* \Box\rightarrow \Delta^* \end{array}$$

The introduction rules for  $\wedge, \vee, \supset$  and  $\neg$  are formulated for arbitrary sequent arrows. Whereas the rules for  $\wedge$  and  $\vee$  are versions of the familiar introduction rules, the rules for  $\neg$  and  $\supset$  can be formulated such that they make use of the structure connective  $\neg$ :

$$\begin{array}{l} \Delta, \neg A (\rightarrow) \Gamma \vdash \Delta, \neg A (\rightarrow) \Gamma \\ \Delta (\rightarrow) \Gamma, \neg A \vdash \Delta (\rightarrow) \Gamma, \neg A \\ \Delta, \neg A (\rightarrow) \Gamma \quad \Sigma, B (\rightarrow) \Theta \vdash \Delta, \Sigma, A \supset B (\rightarrow) \Gamma, \Theta \\ \Delta (\rightarrow) \Gamma, \neg A, B \vdash \Delta (\rightarrow) \Gamma, A \supset B \end{array}$$

The introduction rules for the modal operators are not formulated for arbitrary sequent arrows:

$$\begin{array}{ll} [\Box\Box\rightarrow] & A \rightarrow \Delta \vdash \Box A \Box\rightarrow \Delta \quad [\rightarrow \Box] \quad \Delta \Box\rightarrow A \vdash \Delta \rightarrow \Box A \\ [\Diamond\Diamond\rightarrow] & A \Diamond\rightarrow \Delta \vdash \Diamond A \rightarrow \Delta \quad [\Diamond\rightarrow \Diamond] \quad \Delta \rightarrow A \vdash \Delta \Diamond\rightarrow \Diamond A \\ [\Diamond\rightarrow \Diamond] & \neg A, \Delta \Diamond\rightarrow \Gamma \vdash \Delta \Diamond\rightarrow \Gamma, \Diamond A \quad [\Box\Box\rightarrow] \quad \Delta \Box\rightarrow \Gamma, \neg A \vdash \Delta \Box A \Box\rightarrow \Gamma \end{array}$$

The above collection of sequent rules forms a multiple-sequent calculus **MC** for the system **C**. An axiomatization of **C** can be obtained by replacing the necessitation rule in the familiar axiomatization of **K** by the weaker rule

$$(\text{RR}) \quad \text{if } (A \wedge B) \supset C \text{ is provable, then so is } (\Box A \wedge \Box B) \supset \Box C,$$

see [Chellas, 1980]. The necessitation rule and the modal axiom schemata  $D, T,$  and  $4$  can be captured in a modular fashion by pairs of sequent rules:

$$\begin{array}{ll} [\text{nec}] & \Delta \rightarrow \emptyset \vdash \Delta \Diamond\rightarrow \emptyset \quad \emptyset \rightarrow \Delta \vdash \emptyset \Box\rightarrow \Delta \\ [D] & \Delta \Box\rightarrow \emptyset \vdash \Delta \rightarrow \emptyset \quad \emptyset \Diamond\rightarrow \Delta \vdash \emptyset \rightarrow \Delta \\ [T] & \Delta \Box\rightarrow \Gamma \vdash \Delta \rightarrow \Gamma \quad \Delta \Diamond\rightarrow \Gamma \vdash \Delta \rightarrow \Gamma \\ [4] & \Delta \rightarrow \Sigma \vdash \Delta \Box\rightarrow \Sigma \quad \Theta \rightarrow \Gamma \vdash \Theta \Diamond\rightarrow \Gamma, \end{array}$$

where in rule  $[4]$ , every  $S$ -formula in  $\Delta$  has the shape  $\Box A$  or  $\neg \Diamond A$  and every  $S$ -formula in  $\Gamma$  has the shape  $\Diamond A$  or  $\neg \Box A$ . All sequent systems obtained in this way satisfy a generalized subformula property: for every modal formula  $A$ , it holds that if  $A$  or  $\neg A$  is used in a proof of  $\Delta \rightarrow \Gamma$ , then  $A$  is a subformula of  $\Delta \cup \Gamma$  (where the notion of a subformula of an  $S$ -formula is defined in the obvious way). Indrzejczak does not investigate the admissibility of

cut for  $\rightarrow$  or the admissibility of cut for  $\Box\rightarrow$  and  $\Diamond\rightarrow$  in extensions of **CT** (where  $\vdash A\Box\rightarrow A$  and  $\vdash A\Diamond\rightarrow A$ ). Note that the introduction rules for the modal operators fail to be symmetrical, since there are no introduction rules for  $\Box$  on the left and  $\Diamond$  on the right of  $\rightarrow$ . Moreover, the side conditions on [4] are such that the status of this rule as a purely structural rule is doubtful. The multiple-sequent systems for extensions of **KB** make use of denumerably many sequent arrows  $\xrightarrow{n}$  ( $n \geq 0$ ), where logics are defined by the provable sequents  $\Delta \xrightarrow{0} \Gamma$ . The introduction rules

$$\begin{array}{ll} A \xrightarrow{n} \Delta \vdash \Box A \xrightarrow{n+1} \Delta & \Delta \xrightarrow{n+1} A \vdash \Delta \xrightarrow{n} \Box A \\ A \xrightarrow{n+1} \Delta \vdash \Diamond A \xrightarrow{n} \Delta & \Delta \xrightarrow{n} A \vdash \Delta \xrightarrow{n+1} \Diamond A \end{array}$$

fail to introduce  $\Box$  on the left and  $\Diamond$  on the right of  $\xrightarrow{0}$ , so that also these rules are not symmetrical.

In Section 4.1, we shall point to a simple relation between Indrzejczak's multiple-sequent systems and higher-arity sequent systems for modal logics.

### 2.5 Hypersequents

Hypersequents were introduced into the literature by Pottinger [1983], and have later systematically been studied by Avron [1991; 1991a; 1996]. A *hypersequent* is a sequence

$$\Gamma_1 \rightarrow \Delta_1 \mid \Gamma_2 \rightarrow \Delta_2 \mid \dots \mid \Gamma_n \rightarrow \Delta_n$$

of ordinary sequents (or, more generally, sequents in which  $\Delta_i$  and  $\Gamma_i$  are sequences of formula occurrences) as their *components*. The symbol ' $\mid$ ' in the statement of a hypersequent enriches the language of sequents and is intuitively to be read as disjunction. This expressive enhancement “makes it possible to introduce *new* types of structural rules, and . . . to allow greater versatility in developing interesting logical systems” [Avron, 1996, p. 6]. In particular, a distinction may be drawn between internal and external versions of structural rules. The internal rules deal with formulas within a certain component, whereas the external rules deal with components within a hypersequent. Let  $G, H, H_1, H_2$  etc. be schematic letters for possibly empty hypersequents. External monotonicity, for instance, can be contrasted with internal monotonicity:

$$H_1 \mid H_2 \vdash H_1 \mid G \mid H_2 \quad \text{vs.} \quad H_1 \mid \Gamma \rightarrow \Delta \mid H_2 \vdash H_1 \mid A, \Gamma \rightarrow \Delta \mid H_2.$$

Cut only has an internal version:

$$\frac{G_1 \mid \Gamma_1 \rightarrow \Delta_1, A \mid H_1 \quad G_2 \mid A, \Gamma_2 \rightarrow \Delta_2 \mid H_2}{G_1 \mid G_2 \mid \Gamma_1, \Gamma_2 \rightarrow \Delta_1, \Delta_2 \mid H_1 \mid H_2}$$

The use of hypersequents allows a cut-free presentation **GS5** of **S5** satisfying the subformula property. The system **GS5** consists of hypersequential

versions of the rules of **LS4**, in particular, external and internal versions of contraction and monotonicity, the above cut-rule, and a structural rule of a new kind, namely the *modalized splitting rule*:

$$(MS) \quad G \mid \Box\Gamma_1, \Gamma_2 \rightarrow \Box\Delta_1, \Delta_2 \mid H \vdash G \mid \Box\Gamma_1 \rightarrow \Box\Delta_1 \mid \Gamma_2 \rightarrow \Delta_2 \mid H.$$

In the next section we shall define display sequents, and in Section 4.2 we shall define a translation of hypersequents into display sequents.

### 3 DISPLAY LOGIC

We shall develop display logic only to the extent needed to cover a variety of normal modal and temporal logics based on classical or intuitionistic logic. A more comprehensive presentation of display logic and its application to modal and non-classical logics can be found in [Belnap, 1982], [Belnap, 1990], [Belnap, 1996], [Goré, 1998], [Kracht, 1996], [Restall, 1998], [Wansing, 1998]. Note that except for the substructure property, all requirements examined in the previous sections are satisfied by the display sequent systems to be presented.

#### 3.1 Introduction rules through residuation

Whereas the ordinary sequent systems for temporal logics presented in Section 1.2 fail to exploit the interaction between the backward and the forward looking modalities, the modal display calculus is based on observing that the operators  $\langle P \rangle$  and  $[F]$  form a residuated pair. The following definition is taken from Dunn [1990, p. 32]:

DEFINITION 5. Consider two partially ordered sets  $\mathcal{A} = (\mathbf{A}, \leq)$  and  $\mathcal{B} = (\mathbf{B}, \leq')$  with functions  $f: \mathbf{A} \rightarrow \mathbf{B}$  and  $g: \mathbf{B} \rightarrow \mathbf{A}$ . The pair  $(f, g)$  is called

$$\begin{aligned} & \textit{residuated} && \text{iff} && (fa \leq' b \text{ iff } a \leq gb); \\ & \textit{a Galois connection} && \text{iff} && (b \leq' fa \text{ iff } a \leq gb); \\ & \textit{a dual Galois connection} && \text{iff} && (fa \leq' b \text{ iff } gb \leq a); \\ & \textit{a dual residuated pair} && \text{iff} && (b \leq' fa \text{ iff } gb \leq a). \end{aligned}$$

Obviously,  $(\langle P \rangle, [F])$  forms a residuated pair with respect to the provability relation in normal extensions of **Kt**, and  $(\neg\langle F \rangle, \neg\langle P \rangle)$  is a Galois connection.<sup>5</sup> These ideas of residuation and Galois connection can be generalized. In [Dunn, 1990], [Dunn, 1993], Dunn has defined an abstract law of

<sup>5</sup>The fact that  $\langle P \rangle$  and  $[F]$  form a residuated pair is also used in Kashima's [1994] sequent calculi for various normal temporal logics. The approach of Kashima is similar to the modal display calculus and the modal signs approach developed by Cerrato [1993; 1996] insofar as the structural language of sequents is extended by unary structure operations. Whereas nesting of these operations is not allowed in Cerrato's sequent systems for normal modal propositional logics, Kashima allows iteration. Kashima inductively defines a notion of sequent as follows:

residuation for  $n$ -place connectives  $f$  and  $g$ . The formulation of this principle refers to *traces* of operations and assumes the presence (or definability) of a truth constant  $t$  and a falsity constant  $f$ . We shall use  $A \dashv\vdash B$  to express that  $A$  and  $B$  are interderivable in a given axiom system.

DEFINITION 6. An  $n$ -place connective  $f$  ( $n \geq 0$ ) has a trace  $(\rho_1, \dots, \rho_n) \mapsto +$  (in symbols  $T(f) = (\rho_1, \dots, \rho_n) \mapsto +$ ) iff

- $f(A_1, \dots, t, \dots, A_n) \dashv\vdash t$ , if  $\rho_i = +$  (the indicated  $t$  is in position  $i$ );
- $f(A_1, \dots, f, \dots, A_n) \dashv\vdash t$ , if  $\rho_i = -$  (the indicated  $f$  is in position  $i$ );
- if  $A \vdash B$  and  $\rho_i = +$ , then  $f(A_1, \dots, A, \dots, A_n) \vdash f(A_1, \dots, B, \dots, A_n)$ ;
- if  $A \vdash B$  and  $\rho_i = -$ , then  $f(A_1, \dots, B, \dots, A_n) \vdash f(A_1, \dots, A, \dots, A_n)$ .

The operation  $f$  has a trace  $(\rho_1, \dots, \rho_n) \mapsto -$  ( $T(f) = (\rho_1, \dots, \rho_n) \mapsto -$ ) iff

- $f(A_1, \dots, f, \dots, A_n) \dashv\vdash f$ , if  $\rho_i = +$  (the indicated  $f$  is in position  $i$ );
- $f(A_1, \dots, t, \dots, A_n) \dashv\vdash f$ , if  $\rho_i = -$  (the indicated  $t$  is in position  $i$ );
- if  $A \vdash B$  and  $\rho_i = +$ , then  $f(A_1, \dots, B, \dots, A_n) \vdash f(A_1, \dots, A, \dots, A_n)$ ;
- if  $A \vdash B$  and  $\rho_i = -$ , then  $f(A_1, \dots, A, \dots, A_n) \vdash f(A_1, \dots, B, \dots, A_n)$ .

In  $\mathbf{Kt}$ ,  $\neg$  has traces  $- \mapsto +$  and  $+ \mapsto -$ , whereas  $[F]$  has trace  $+ \mapsto +$  and  $\langle P \rangle$  has trace  $- \mapsto -$ .

DEFINITION 7. Two  $n$ -place operations  $f$  and  $g$  are *contrapositives in place  $j$*  iff  $T(f) = (\rho_1, \dots, \rho_j, \dots, \rho_n) \mapsto \rho$  implies  $T(g) = (\rho_1, \dots, -\rho, \dots, \rho_n) \mapsto -\rho_j$ , where  $-+ = -$  and  $-- = +$ .

DEFINITION 8. Let

$$\underline{S(f, A_1, \dots, A_n, B)} \text{ iff } \begin{cases} B \vdash f(A_1, \dots, A_n) & \text{if } T(f) = (\dots) \mapsto + \\ f(A_1, \dots, A_n) \vdash B & \text{if } T(f) = (\dots) \mapsto - \end{cases}$$

1. every temporal formula is a sequent;
2. if  $\Gamma$  is a sequent, then so is  $^P\{\Gamma\}$  and  $^F\{\Gamma\}$ ;
3. if  $n \geq 0$  and every  $\Gamma_i$  ( $1 \leq i \leq n$ ) is a sequent, then so is  $\Gamma_1, \dots, \Gamma_n$ .

The intuitive meaning of a sequent is given by the following inductively defined translation  $(\cdot)^*$  from sequents into formulas:

1.  $(\Gamma)^* = A$ , if  $\Gamma$  is the formula  $A$ ;
2.  $(^P\{\Gamma\})^* = [P]^P(\Gamma)^*$ ;  $(^F\{\Gamma\})^* = [F]^F(\Gamma)^*$ ;
3. if  $n > 0$ , then  $(\Gamma_1, \dots, \Gamma_n)^* = \bigvee\{(\Gamma_1)^*, \dots, (\Gamma_n)^*\}$ ;
4.  $(\cdot)^* = (p \wedge \neg p)$ , for some atom  $p$ .

Residuation then shows up in Kashima's "turn rules":

$$\Gamma, ^F\{\Delta\} \vdash ^P\Gamma, \Delta; \quad \Gamma, ^P\{\Delta\} \vdash ^F\Gamma, \Delta.$$

Most of Kashima's sequent rules used to capture various structural properties of accessibility either fail to be explicit or separated in the sense of Section 1.3. Cut-elimination for these systems is shown semantically, i.e., in a non-constructive way.

A pair of  $n$ -place connectives  $f$  and  $g$  satisfies the abstract law of residuation just in case for some  $j$  ( $1 \leq j \leq n$ ),  $f$  and  $g$  are contrapositives in place  $j$ , and

$$S(f, A_1, \dots, A_j, \dots, A_n, B) \text{ iff } S(g, A_1, \dots, B, \dots, A_n, A_j).$$

**OBSERVATION 9.** The abstract law of residuation holds for the pairs  $(\mathbf{t}, \mathbf{f})$ ,  $(\neg, \neg)$ ,  $(\langle P \rangle, [F])$ ,  $(\wedge, \triangleright)$ ,  $(\blacktriangleleft, \vee)$ ,  $(\wedge, \neg \dots \vee \dots)$ , and  $(\dots \wedge \neg \dots, \vee)$ , where  $\triangleright$  is intuitionistic implication and  $\blacktriangleleft$  is coimplication.

Coimplication  $\blacktriangleleft$  is characterized by

$$A \vdash B \vee C, \Delta \text{ iff } A \blacktriangleleft B \vdash C, \Delta.$$

In classical logic, the residual of disjunction is definable, since

$$A \vdash B \vee C, \Delta \text{ iff } A \wedge \neg B \vdash C, \Delta \text{ iff } \neg(A \supset B) \vdash C,$$

but in bi-intuitionistic logic it is not, see Section 3.8. For each of the pairs  $(\mathbf{t}, \mathbf{f})$ ,  $(\neg, \neg)$ ,  $(\langle P \rangle, [F])$ ,  $(\wedge, \triangleright)$ ,  $(\blacktriangleleft, \vee)$ , the structural language of display sequents contains one structure connective. Since in classical logic  $\wedge$  and  $\vee$  are interdefinable using  $\neg$ , the pairs  $(\wedge, \neg \dots \vee \dots)$  and  $(\dots \wedge \neg \dots, \vee)$  require only a single structure connective in addition to the unary structure operation associated with  $(\neg, \neg)$ . We shall use  $X, Y, Z$  (possibly with subscripts) as variables for structures. A display sequent is an expression  $X \rightarrow Y$ ;  $X$  is called the antecedent and  $Y$  is called the succedent of  $X \rightarrow Y$ . The structures are defined by:

$$X ::= A \mid \mathbf{I} \mid *X \mid \bullet X \mid X \circ Y \mid X \bowtie Y \mid X \times Y.$$

The association of structure connectives with pairs of operations satisfying the abstract law of residuation is accomplished by the following translations  $\tau_1$  of antecedents and  $\tau_2$  of succedents into formulas:

$$\begin{array}{ll} \tau_1(A) = A & \tau_2(A) = A \\ \tau_1(\mathbf{I}) = \mathbf{t} & \tau_2(\mathbf{I}) = \mathbf{f} \\ \tau_1(*X) = \neg \tau_2(X) & \tau_2(*X) = \neg \tau_1(X) \\ \tau_1(\bullet X) = \langle P \rangle \tau_1(X) & \tau_2(\bullet X) = [F] \tau_2(X) \\ \tau_1(X \bowtie Y) = \tau_1(X) \wedge \tau_1(Y) & \tau_2(X \bowtie Y) = \tau_2(X) \triangleright \tau_2(Y) \\ \tau_1(X \times Y) = \tau_1(X) \blacktriangleleft \tau_1(Y) & \tau_2(X \times Y) = \tau_2(X) \vee \tau_2(Y) \\ \tau_1(X \circ Y) = \tau_1(X) \wedge \tau_1(Y) & \tau_2(X \circ Y) = \tau_2(X) \vee \tau_2(Y) \end{array}$$

Under these translations, the following basic structural rules are valid ((1)–(4) in normal temporal logic; (5) and (6) in bi-intuitionistic logic) if  $\rightarrow$  is

understood as provability:

Basic structural rules

- (1)  $X \circ Y \rightarrow Z \dashv\vdash X \rightarrow Z \circ *Y \dashv\vdash Y \rightarrow *X \circ Z$
- (2)  $X \rightarrow Y \circ Z \dashv\vdash X \circ *Z \rightarrow Y \dashv\vdash *Y \circ X \rightarrow Z$
- (3)  $X \rightarrow Y \dashv\vdash *Y \rightarrow *X \dashv\vdash X \rightarrow **Y$
- (4)  $X \rightarrow \bullet Y \dashv\vdash \bullet X \rightarrow Y$
- (5)  $X \rtimes Y \rightarrow Z \dashv\vdash Y \rightarrow X \rtimes Z \dashv\vdash X \rightarrow Y \rtimes Z$
- (6)  $X \rightarrow Y \rtimes Z \dashv\vdash X \rtimes Y \rightarrow Z \dashv\vdash X \rtimes Z \rightarrow Y,$

where  $X_1 \rightarrow Y_1 \dashv\vdash X_2 \rightarrow Y_2$  abbreviates  $X_1 \rightarrow Y_1 \vdash X_2 \rightarrow Y_2$  and  $X_2 \rightarrow Y_2 \vdash X_1 \rightarrow Y_1$ . If two sequents are interderivable by means of (1)–(6), then these sequents are said to be *structurally* or *display equivalent*. The following pairs of sequents, for example, are display equivalent on the strength of (1)–(3):

$$\begin{array}{llll} X \circ Y \rightarrow Z & *Z \rightarrow *Y \circ *X; & X \rightarrow Y \circ Z & *Z \circ *Y \rightarrow *X; \\ X \rightarrow Y & *Y \rightarrow X; & X \rightarrow *Y & Y \rightarrow *X; \\ X \rightarrow Y & **X \rightarrow Y. & & \end{array}$$

The name ‘display logic’ derives from the fact that any substructure of a given display sequent  $s$  may be *displayed* as the entire antecedent or succedent of a structurally equivalent sequent  $s'$ . In order to state this fact precisely, we define the notion of a polarity vector and antecedent and succedent part of a sequent (cf. [Goré, 1998]).

**DEFINITION 10.** To each  $n$ -place structure connective  $c$  we assign two *polarity vectors*  $ap(c, \pm_1, \dots, \pm_n)$  and  $sp(c, \pm_1, \dots, \pm_n)$ , where  $\pm_i \in \{+, -\}$  and  $1 \leq i \leq n$ :

$$\begin{array}{lllll} ap(*, -) & ap(\bullet, +) & ap(\circ, +, +) & ap(\rtimes, +, +) & ap(\rtimes, +, -) \\ sp(*, -) & sp(\bullet, +) & sp(\circ, +, +) & sp(\rtimes, -, +) & sp(\rtimes, +, +) \end{array}$$

We write  $ap(c, j, \pm)$  and  $sp(c, j, \pm)$  to express that  $c$  has antecedent, respectively succedent polarity  $\pm$  at place  $j$ .

**DEFINITION 11.** Let  $s = X \rightarrow Y$ . The exhibited occurrence of  $X$  is an antecedent part of  $s$ , and the exhibited occurrence of  $Y$  is a succedent part of  $s$ . If  $c(X_1, \dots, X_n)$  is an antecedent [succedent] part of  $s$ , then the substructure occurrence  $X_j$  ( $1 \leq j \leq n$ ) is

1. an antecedent [succedent] part of  $s$  if  $ap(c, j, +)$  [ $sp(c, j, +)$ ];
2. a succedent [antecedent] part of  $s$  if  $ap(c, j, -)$  [ $sp(c, j, -)$ ].

**THEOREM 12.** (Display Theorem, Belnap) *For each display sequent  $s$  and each antecedent [succedent] part  $X$  of  $s$  there exists a display sequent  $s'$  structurally equivalent to  $s$  such that  $X$  is the entire antecedent [succedent] of  $s'$ .*



**Proof.** The theorem was first proved in [Belnap, 1982]; we shall follow the proof in [Restall, 1998]. A *context* results from a structure by replacing one occurrence of a substructure by the ‘Void’ (in symbols ‘ $-$ ’). If  $f$  is a context and  $X$  is a structure, then  $f(X)$  is the result of substituting  $X$  for the Void in  $f$ . A context  $f$  is called *antecedent positive (negative)* if the indicated  $X$  is an antecedent part (a succedent part) of  $f(X) \rightarrow Y$ ;  $f$  is said to be *succedent positive (negative)* if the indicated  $X$  is a succedent part (an antecedent part) of  $Y \rightarrow f(X)$ . A contextual sequent has the shape  $f \rightarrow Z$  or  $Z \rightarrow f$ , and a pair of contextual sequents is said to be structurally equivalent if the sequents are interderivable by means of rules (1)–(6). The Display Theorem then follows from the following lemma.

LEMMA 13. (i) *Suppose  $f$  is a context in antecedent position. If  $f$  is antecedent positive, then  $f(X) \rightarrow Y$  is structurally equivalent to  $X \rightarrow f^a(Y)$ , where  $f^a$  is a context obtained by unraveling the Void in  $f$ . If  $f$  is antecedent negative, then  $f(X) \rightarrow Y$  is structurally equivalent to  $f^a(Y) \rightarrow X$ .* (ii) *Suppose  $f$  is a context in succedent position. If  $f$  is succedent positive, then  $Y \rightarrow f(X)$  is structurally equivalent to  $f^c(Y) \rightarrow X$ , where  $f^c$  is a context obtained by unraveling the Void in  $f$ . If  $f$  is succedent negative, then  $Y \rightarrow f(X)$  is structurally equivalent to  $X \rightarrow f^c(Y)$ .*

The proof is by induction on the complexity of contexts.

Case 1:  $f = -$ . Then  $f$  is antecedent and succedent positive, and  $f^a(Y) = f^c(Y) = Y$ .

Case 2:  $f = \bullet g$ . Then  $f(X) \rightarrow Y$  is structurally equivalent to  $g(X) \rightarrow \bullet Y$ , and  $Y \rightarrow f(X)$  is equivalent to  $\bullet Y \rightarrow g(X)$ . By the induction hypothesis, these sequents are equivalent to  $X \rightarrow f^a(\bullet Y)$ ,  $f^a(\bullet Y) \rightarrow X$ ,  $f^c(\bullet Y) \rightarrow X$ , or  $X \rightarrow f^c(\bullet Y)$ . Hence  $f^a = g^a(\bullet -)$  and  $f^c = g^c(\bullet -)$ .

Case 3:  $f = *g$ . Then  $f(X) \rightarrow Y$  is equivalent to  $*Y \rightarrow g(X)$ . Depending on whether  $g$  is succedent positive or negative,  $f(X) \rightarrow Y$  is structurally equivalent to  $g^c(*Y) \rightarrow X$  or to  $X \rightarrow g^c(*Y)$ . Therefore, by the induction hypothesis,  $f^a = g^c(*-)$ . Similarly,  $f^c = g^a(*-)$ .

Case 4:  $f = Z \circ g$ . Then  $f(X) \rightarrow Y$  is equivalent to  $g(X) \rightarrow *Z \circ Y$ . By the induction hypothesis, this sequent is equivalent to  $X \rightarrow g^a(*Z \circ Y)$  or  $g^a(*Z \circ Y) \rightarrow X$ , and hence  $f^a = g^a(*Z \circ -)$ . Similarly,  $f^c = g^a(- \circ *Z)$ .

Case 5:  $f = g \circ Z$ . Similar to Case 4.

Case 6:  $f = g \times Z$ . Then  $Y \rightarrow f(X)$  is equivalent to  $g(X) \rightarrow Y \times Z$ , and by the induction hypothesis, the latter is equivalent to  $X \rightarrow g^a(Y \times Z)$  or to  $g^a(Y \times Z) \rightarrow X$ . Thus  $f^c = g^a(- \times Z)$ . Similarly,  $f^a = g^c(Z \times -)$ .

Case 7:  $f = Z \times g$ . Analogous to the previous case.

Cases 8 and 9:  $f = g \times Z$  and  $f = Z \times g$ . Analogous to Cases 6 and 7. ■

If (for suitable notions of structural equivalence, antecedent part, and succedent part) a sequent calculus satisfies the Display Theorem, it is said to enjoy the *display property*. Note that the set of rules (1)–(6) is not the only

<i>truth and falsity rules</i>	
$(\rightarrow \mathbf{f})$	$X \rightarrow \mathbf{I} \vdash X \rightarrow \mathbf{f}$
$(\mathbf{f} \rightarrow)$	$\vdash \mathbf{f} \rightarrow \mathbf{I}$
$(\rightarrow \mathbf{t})$	$\vdash \mathbf{I} \rightarrow \mathbf{t}$
$(\mathbf{t} \rightarrow)$	$\mathbf{I} \rightarrow X \vdash \mathbf{t} \rightarrow X$
<i>Boolean introduction rules</i>	
$(\rightarrow \neg)$	$X \rightarrow *A \vdash X \rightarrow \neg A$
$(\neg \rightarrow)$	$*A \rightarrow X \vdash \neg A \rightarrow X$
$(\rightarrow \wedge)$	$X \rightarrow A \quad Y \rightarrow B \vdash X \circ Y \rightarrow A \wedge B$
$(\wedge \rightarrow)$	$A \circ B \rightarrow X \vdash A \wedge B \rightarrow X$
$(\rightarrow \vee)$	$X \rightarrow A \circ B \vdash X \rightarrow A \vee B$
$(\vee \rightarrow)$	$A \rightarrow X \quad B \rightarrow Y \vdash A \vee B \rightarrow X \circ Y$
$(\rightarrow \supset)$	$X \circ A \rightarrow B \vdash X \rightarrow A \supset B$
$(\supset \rightarrow)$	$X \rightarrow A \quad B \rightarrow Y \vdash A \supset B \rightarrow *X \circ Y$
$(\rightarrow \equiv)$	$X \circ A \rightarrow B \quad X \circ B \rightarrow A \vdash X \rightarrow A \equiv B$
$(\equiv \rightarrow)$	$X \rightarrow A \quad B \rightarrow Y \quad X \rightarrow B \quad A \rightarrow Y \vdash A \equiv B \rightarrow *X \circ Y$
<i>tense logical introduction rules</i>	
$(\rightarrow [F])$	$\bullet X \rightarrow A \vdash X \rightarrow [F]A$
$([F] \rightarrow)$	$A \rightarrow X \vdash [F]A \rightarrow \bullet X$
$(\rightarrow \langle F \rangle)$	$X \rightarrow A \vdash * \bullet * X \rightarrow \langle F \rangle A$
$(\langle F \rangle \rightarrow)$	$* \bullet * A \rightarrow Y \vdash \langle F \rangle A \rightarrow Y$
$(\rightarrow [P])$	$X \rightarrow * \bullet * A \vdash X \rightarrow [P]A$
$([P] \rightarrow)$	$A \rightarrow X \vdash [P]A \rightarrow * \bullet * X$
$(\rightarrow \langle P \rangle)$	$X \rightarrow A \vdash \bullet X \rightarrow \langle P \rangle A$
$(\langle P \rangle \rightarrow)$	$A \rightarrow \bullet X \vdash \langle P \rangle A \rightarrow X$
<i>nonclassical introduction rules</i>	
$(\rightarrow \wedge)'$	$X \rightarrow A \quad Y \rightarrow B \vdash X \times Y \rightarrow A \wedge B$
$(\wedge \rightarrow)'$	$A \times B \rightarrow X \vdash A \wedge B \rightarrow X$
$(\rightarrow \triangleright)$	$X \rightarrow A \times B \vdash X \rightarrow A \triangleright B$
$(\triangleright \rightarrow)$	$X \rightarrow A \quad B \rightarrow Y \vdash A \triangleright B \rightarrow X \times Y$
$(\rightarrow \vee)'$	$X \rightarrow A \times B \vdash X \rightarrow A \vee B$
$(\vee \rightarrow)'$	$A \rightarrow X \quad B \rightarrow Y \vdash A \vee B \rightarrow X \times Y$
$(\rightarrow \blacktriangleleft)$	$X \rightarrow A \quad B \rightarrow Y \vdash X \times Y \rightarrow A \blacktriangleleft B$
$(\blacktriangleleft \rightarrow)$	$A \times B \rightarrow X \vdash A \blacktriangleleft B \rightarrow X$

Table 1. Introduction rules.

$(\mathbf{I}_+^\circ)$	$X \rightarrow Z \vdash \mathbf{I} \circ X \rightarrow Z$	$X \rightarrow Z \vdash X \circ \mathbf{I} \rightarrow Z$
	$X \rightarrow Z \vdash X \rightarrow Z \circ \mathbf{I}$	$X \rightarrow Z \vdash X \rightarrow \mathbf{I} \circ Z$
$(\mathbf{I}_-^\circ)$	$\mathbf{I} \circ X \rightarrow Z \vdash X \rightarrow Z$	$X \circ \mathbf{I} \rightarrow Z \vdash X \rightarrow Z$
	$X \rightarrow Z \circ \mathbf{I} \vdash X \rightarrow Z$	$X \rightarrow \mathbf{I} \circ Z \vdash X \rightarrow Z$
$(\mathbf{I})$	$\mathbf{I} \rightarrow X \vdash Z \rightarrow X$	$X \rightarrow \mathbf{I} \vdash X \rightarrow Z$
$(\mathbf{I}^*)$	$\mathbf{I} \rightarrow X \dashv\vdash * \mathbf{I} \rightarrow X$	$X \rightarrow \mathbf{I} \dashv\vdash X \rightarrow * \mathbf{I}$
$(\mathbf{P}\circ)$	$X_1 \circ X_2 \rightarrow Z \vdash X_2 \circ X_1 \rightarrow Z$	$Z \rightarrow X_1 \circ X_2 \vdash Z \rightarrow X_2 \circ X_1$
$(\mathbf{C}\circ)$	$X \circ X \rightarrow Z \vdash X \rightarrow Z$	$Z \rightarrow X \circ X \vdash Z \rightarrow X$
$(\mathbf{E}\circ)$	$X \rightarrow Z \vdash X \circ X \rightarrow Z$	$Z \rightarrow X \vdash Z \rightarrow X \circ X$
$(\mathbf{M}\circ)$	$X_1 \rightarrow Z \vdash X_1 \circ X_2 \rightarrow Z$	$X_1 \rightarrow Z \vdash X_2 \circ X_1 \rightarrow Z$
	$Z \rightarrow X_1 \vdash Z \rightarrow X_1 \circ X_2$	$Z \rightarrow X_1 \vdash Z \rightarrow X_2 \circ X_1$
$(\mathbf{A}\circ)$	$X_1 \circ (X_2 \circ X_3) \rightarrow Z \dashv\vdash (X_1 \circ X_2) \circ X_3 \rightarrow Z$	
	$Z \rightarrow X_1 \circ (X_2 \circ X_3) \dashv\vdash (X_1 \circ X_2) \circ X_3 \rightarrow Z$	
$(\mathbf{MN})$	$\mathbf{I} \rightarrow X \vdash \mathbf{I} \rightarrow \bullet X$	$X \rightarrow \mathbf{I} \vdash X \rightarrow \bullet \mathbf{I}$
	$\mathbf{I} \rightarrow X \vdash \mathbf{I} \rightarrow * \bullet * X$	$X \rightarrow \mathbf{I} \vdash X \rightarrow * \bullet * \mathbf{I}$

Table 2. Additional structural rules.

possible choice of display rules warranting the display property, see [Belnap, 1996] and [Goré, 1998].<sup>6</sup> The display property allows an “‘essentials-only’ proof of cut elimination relying on easily established and maximally general properties of structural and connective rules” [Belnap, 1996, p. 80]. Further, the display property enables a statement of the introduction rules that satisfies the segregation requirement. Belnap emphasizes that the display property may be used to keep certain proof-theoretic components as separate as possible. In a sequent calculus enjoying the display property, the behaviour of the structural elements can be described by the structural rules, and the right (left) introductions rules for an  $n$ -place logical operation  $f$  can be formulated with  $f(A_1, \dots, A_n)$  standing alone as the entire succedent (antecedent) of the conclusion sequent. Since  $f(A_1, \dots, A_n)$  plays no inferential roles beyond being derived and allowing to derive, these left and right rules provide a complete explanation of the inferential meaning of  $f$ . The constant  $\mathbf{I}$  induces introduction rules for  $t$  and  $f$ . The operations  $*$  and  $\circ$  give rise to introduction rules for the Boolean connectives. The structure operation  $\bullet$  permits formulating introduction rules for the modal-

<sup>6</sup>Goré [Goré, 1998] introduces binary structure connectives  $<$  and  $>$  to be interpreted as directional versions of implication in succedent position and coimplication in antecedent position. The display property is guaranteed by the following structural rules (notation adjusted):

$$\begin{aligned} X \rightarrow Z < Y \dashv\vdash X \circ Y \rightarrow Z \dashv\vdash Y \rightarrow X > Z \\ Z < Y \rightarrow X \dashv\vdash Z \rightarrow X \circ Y \dashv\vdash X > Z \rightarrow Y. \end{aligned}$$

ities, whereas  $\times$  and  $\leftarrow$  give rise to introduction schemata for conjunction, disjunction, implication, and coimplication in bi-intuitionistic logic. These introduction rules are assembled in Table 1. The further structural rules in Table 2 contain many redundancies when they are assumed as a set. Such a rich inventory of structural inference rules is, however, an advantage in a treatment of substructural subsystems of normal modal and temporal logics, see [Goré, 1998]. In addition to a set of structural rules and a set of introduction rules, every display sequent system contains two *logical* rules exhibiting neither structural nor logical operations, namely reflexivity for atoms (alias identity) and cut:

$$\text{(id)} \quad \vdash p \rightarrow p \quad \text{and} \quad \text{(cut)} \quad X \rightarrow A \quad A \rightarrow Y \vdash X \rightarrow Y.$$

The identity rule (id) can be generalized to arbitrary formulas from temporal or bi-intuitionistic logic.

OBSERVATION 14. For every formula  $A$ ,  $\vdash A \rightarrow A$ .

**Proof.** The proof is by induction on the complexity of  $A$ . For example,

$$\frac{A \rightarrow A}{[P]A \rightarrow * \bullet * A} \quad \frac{A \rightarrow A}{\bullet A \rightarrow \langle P \rangle A} \quad \frac{A \rightarrow A \quad B \rightarrow B}{A \times B \rightarrow A \leftarrow B}$$

$$\frac{[P]A \rightarrow * \bullet * A}{[P]A \rightarrow [P]A} \quad \frac{A \rightarrow \bullet \langle P \rangle A}{\langle P \rangle A \rightarrow \langle P \rangle A} \quad \frac{A \times B \rightarrow A \leftarrow B}{A \leftarrow B \rightarrow A \leftarrow B}.$$

■

DEFINITION 15. The display sequent system **DCPL** is given by (id), (cut), the Boolean rules, and the structural rules exhibiting **I**, **\***, and **o**. The system **DKt** consists of **DCPL** plus the tense logical rules and the structural rules exhibiting **•**. The system **DK** results from **DKt** by removing the introduction rules for  $[P]$  and  $\langle P \rangle$ .

A sequent rule is invertible if every premise sequent can be derived from the conclusion sequent.

OBSERVATION 16. The following holds in every purely structural extension of **DKt** and **DK**. (i) The logical operations are uniquely characterized. (ii) The introduction rules for  $\neg$ ,  $\wedge$ , and  $\vee$ , the left introduction rules for  $t$ ,  $\langle P \rangle$ , and  $\langle F \rangle$ , and the right introduction rules for  $f$ ,  $\supset$ ,  $\equiv$ ,  $[P]$ , and  $[F]$  are invertible. (iii) The modalities  $[F]$  and  $\langle F \rangle$  ( $[P]$  and  $\langle P \rangle$ ) are interdefinable using  $\neg$ .

Note that there exist various duality and symmetry transformations on proofs in display logic, see [Goré, 1998], [Kracht, 1996].

### 3.2 Completeness

We shall first consider weak completeness of **DKt** and **DK**, that is, the coincidence of **Kt** (**K**) and **DKt** (**DK**) with respect to provable formulas. We shall then strengthen this result and in Section 3.4 turn to axiomatic extensions of **K** and **Kt**.

**THEOREM 17.** (i) If  $\vdash A$  in **Kt**, then  $\vdash \mathbf{I} \rightarrow A$  in **DKt**. (ii) If  $\vdash X \rightarrow Y$  in **DKt**, then  $\tau_1(X) \vdash \tau_2(Y)$  in **Kt**.

**Proof.** (i) We may take any axiomatization of **Kt** and show that the axiom schemata are provable in **DKt**, and the proof rules preserve provability in **DKt**. The following is a cut-free proof of the *K* axiom schema for  $[F]$ ; the proof for  $[P]$  is analogous:

$$\begin{array}{c}
 \frac{A \rightarrow A}{[F]A \rightarrow \bullet A} \\
 \frac{[F](A \supset B) \circ [F]A \rightarrow \bullet A}{\bullet([F](A \supset B) \circ [F]A) \rightarrow A \quad B \rightarrow B} \\
 \frac{A \supset B \rightarrow * \bullet ([F](A \supset B) \circ [F]A) \circ B}{[F](A \supset B) \rightarrow \bullet (* \bullet ([F](A \supset B) \circ [F]A) \circ B)} \\
 \frac{[F](A \supset B) \circ [F]A \rightarrow \bullet (* \bullet ([F](A \supset B) \circ [F]A) \circ B)}{[F](A \supset B) \circ [F]A \rightarrow \bullet (* \bullet ([F](A \supset B) \circ [F]A) \circ B)} \\
 \frac{\bullet([F](A \supset B) \circ [F]A) \rightarrow * \bullet ([F](A \supset B) \circ [F]A) \circ B}{\bullet([F](A \supset B) \circ [F]A) \circ \bullet([F](A \supset B) \circ [F]A) \rightarrow B} \\
 \frac{\bullet([F](A \supset B) \circ [F]A) \rightarrow B}{[F](A \supset B) \circ [F]A \rightarrow [F]B} \\
 \frac{[F](A \supset B) \rightarrow [F]A \supset [F]B}{\mathbf{I} \circ [F](A \supset B) \rightarrow [F]A \supset [F]B} \\
 \mathbf{I} \rightarrow [F](A \supset B) \supset [F]A \supset [F]B
 \end{array}$$

Necessitation for  $[F]$  and  $[P]$  is taken care of by the (MN) rules. It remains to derive the tense logical interaction schemata  $A \supset [F]\langle P \rangle A$  and  $A \supset [P]\langle F \rangle A$ :

$$\begin{array}{c}
 \frac{A \rightarrow A}{\bullet A \rightarrow \langle P \rangle A} \\
 \frac{\bullet A \rightarrow \langle P \rangle A}{A \rightarrow [F]\langle P \rangle A} \\
 \frac{A \rightarrow A}{* \bullet * A \rightarrow \langle F \rangle A} \\
 \frac{* \bullet * A \rightarrow \langle F \rangle A}{* \langle F \rangle A \rightarrow \bullet * A} \\
 \frac{* \langle F \rangle A \rightarrow \bullet * A}{A \rightarrow * \bullet * \langle F \rangle A} \\
 \frac{A \rightarrow * \bullet * \langle F \rangle A}{A \rightarrow [P]\langle F \rangle A}
 \end{array}$$

(ii) By induction on the complexity of proofs in **DKt**. ■

**COROLLARY 18.** (i) In **Kt**,  $\vdash A$  iff  $\vdash \mathbf{I} \rightarrow A$  in **DKt**. (ii) In **K**,  $\vdash A$  iff  $\vdash \mathbf{I} \rightarrow A$  in **DK**.

**Proof.** (i) By the previous theorem. (ii) This follows from the fact that every frame complete normal propositional tense logic is a conservative extension of its modal fragment. ■

LEMMA 19. *In every extension of  $\mathbf{DKt}$  by structural inference rules, it holds that  $\vdash X \rightarrow \tau_1(X)$  and  $\vdash \tau_2(X) \rightarrow X$ .*

**Proof.** By induction on the complexity of  $X$ . ■

This lemma allows one to prove strong completeness.

THEOREM 20. *In  $\mathbf{DKt}$ ,  $\vdash X \rightarrow Y$  iff  $\tau_1(X) \vdash \tau_2(Y)$  in  $\mathbf{Kt}$ .*

**Proof.** ( $\Rightarrow$ ): This is Theorem 17, (ii). ( $\Leftarrow$ ): Suppose that in  $\mathbf{Kt}$ ,  $\tau_1(X) \vdash \tau_2(Y)$ . Hence  $\vdash_{\mathbf{Kt}} \tau_1(X) \supset \tau_2(Y)$ . By Corollary 18,  $\vdash_{\mathbf{DKt}} \mathbf{I} \rightarrow \tau_1(X) \supset \tau_2(Y)$  and thus  $\vdash_{\mathbf{DKt}} \tau_1(X) \rightarrow \tau_2(Y)$ . Since by Lemma 19,  $\vdash X \rightarrow \tau_1(X)$  and  $\vdash \tau_2(Y) \rightarrow Y$  in  $\mathbf{DKt}$ , an application of cut gives  $\vdash X \rightarrow Y$ . ■

COROLLARY 21.  *$\mathbf{DK}$  is strongly sound and complete with respect to  $\mathbf{K}$ .*

COROLLARY 22.  *$\mathbf{DCPL}$  is strongly sound and complete with respect to  $\mathbf{CPL}$ .*

### 3.3 Strong cut-elimination

A remarkable quality of display logic is that a strong cut-elimination theorem holds for every properly displayable and every displayable logic. Proper displayability and displayability are easily checkable properties. A *proper display calculus* is a calculus of sequents whose rules of inference satisfy the following eight conditions (recall the terminology from Section 1.1):

- C1 *Preservation of formulas.* Each formula which is a constituent of some premise of *inf* is a subformula of some formula in the conclusion of *inf*.
- C2 *Shape-alikeness of parameters.* Congruent parameters are occurrences of the same structure.
- C3 *Non-proliferation of parameters.* Each parameter of *inf* is congruent to at most one constituent in the conclusion of *inf*.
- C4 *Position-alikeness of parameters.* Congruent parameters are either all antecedent or all succedent parts of their respective sequents.
- C5 *Display of principal constituents.* A principal formula of *inf* is either the entire antecedent or the entire succedent of the conclusion of *inf*.
- C6 *Closure under substitution for consequent parts.* Each rule is closed under simultaneous substitution of arbitrary structures for congruent formulas which are consequent parts.

C7 *Closure under substitution for antecedent parts.* Each rule is closed under simultaneous substitution of arbitrary structures for congruent formulas which are antecedent parts.

C8 *Eliminability of matching principal formulas.* If there are inferences  $inf_1$  and  $inf_2$  with respective conclusions (1)  $X \rightarrow A$  and (2)  $A \rightarrow Y$  with  $A$  principal in both inferences, and if cut is applied to obtain (3)  $X \rightarrow Y$ , then either (3) is identical to one of (1) or (2), or there is a proof of (3) from the premises of  $inf_1$  and  $inf_2$  in which every cut-formula of any application of cut is a proper subformula of  $A$ .

Obviously, every display calculus satisfying C1 enjoys the subformula property, that is, every cut-free proof of any sequent  $s$  contains no formulas which are not subformulas of constituents of  $s$ . If a logical system can be presented as a proper display calculus, it is said to be *properly displayable*. Belnap [1982] showed that in every properly displayable logic, a proof of a sequent  $s$  can be converted into a proof of  $s$  not containing any application of cut

$$\frac{(1) X \rightarrow A \quad (2) A \rightarrow Y}{(3) X \rightarrow Y}.$$

The proof of strong cut-elimination reveals that every sufficiently long sequence of steps in a certain process of cut-elimination terminates with a cut-free proof. The elimination process consists of various kinds of actions, *principal moves*, *parametric moves*, and a combination of parametric and principal moves. If the cut-formula  $A$  is principal in the final inference in the proofs of both (1) and (2), a principal move is performed. Otherwise, if there is no previous application of cut, a parametric move or a combination of parametric and principal moves is executed. According to this distinction we define primitive reductions of proofs  $\Pi$  ending in an application of cut. Recently, Jeremy Dawson and Rajeev Goré discovered a gap in the proof of strong normalization presented in [Wansing, 1998]. To avoid the problem, the primitive reduction steps have to be redefined. Let  $\Pi_i$  be the proof of (i) we are dealing with, ( $i = 1, 2$ ).

*Principal moves.* By C8, there are two subcases:

Case 1. (3) is the same as (i):  $\frac{\Pi_1 \quad \Pi_2}{(3)} \rightsquigarrow \Pi_i$

Case 2. There is a proof  $\Pi$  of (3) from the premises  $s_1, \dots, s_n$  of (1) and  $s'_1, \dots, s'_m$  of (2) in which every cut-formula of any application of cut is a proper subformula of  $A$ :

$$\frac{\frac{\Pi^1 \quad \Pi^2}{s_1, \dots, s_n \quad s'_1, \dots, s'_m} \quad (1) \quad (2)}{(3)} \rightsquigarrow \frac{\Pi^1 \quad \Pi^2}{\Pi} (3)$$

*Parametric moves.* The parametric moves modify proofs on a larger scale than the principal moves. The parametric moves show that applications of structural rules need never immediately precede applications of cut. Suppose that  $A$  is parametric in the inference ending in (1). The case for (2) is completely symmetrical. In order to define the parametric moves, we inductively define a set  $Q$  of occurrences of  $A$ , called the set of ‘parametric ancestors’ of  $A$  (in  $\Pi_1$ ), cf. [Belnap, 1982, p. 394]. We start with putting the displayed occurrence of  $A$  in (1) into  $Q$ . Then, by working up  $\Pi_1$ , we add for every inference *inf* in  $\Pi_1$  each constituent of a premise of *inf* which is congruent (with respect to *inf*) to a constituent of the conclusion of *inf* already in  $Q$ . What we obtain is a finite tree of parametric ancestors of  $A$  rooted in the displayed occurrence of  $A$  in (1). This tree and the tree of parametric ancestors of the displayed occurrence of  $A$  in (2) either contain an application of cut or not. If so, we do *not* perform a reduction, but instead consider one of these applications of cut above (1) or (2) for reduction. If not, that is, if there is no application of cut in the trees of parametric ancestors, then for each path of parametric ancestors of  $A$  in  $\Pi_1$ , we distinguish two subcases. Let  $A_u$  be the uppermost element of the path and let *inf* be the inference ending in the sequent  $s$  which contains  $A_u$ .

Case 1.  $A_u$  is not parametric in *inf*. By C4 and C5, it is the entire consequent of  $s$ . We cut with  $\Pi_2$  and replace every parametric ancestor of  $A$  below  $A_u$  in the path by  $Y$ .

Case 2.  $A_u$  is parametric in *inf*. Then, with respect to *inf*,  $A_u$  is congruent only to itself, and we just replace every parametric ancestor of  $A$  below  $A_u$  in the path by  $Y$ . Moreover, we delete  $\Pi_2$ , which is now superfluous.

Call the result of simultaneously carrying out these operations for every path of parametric ancestors of  $A$  in  $\Pi_1$  and removing the initial occurrence of (3) (since now (2) = (3))  $\Pi^l$ . If the tree of parametric ancestors of the displayed occurrence of  $A$  in (1) contains at most one element  $A_u$  that is not parametric in *inf*,  $\Pi$  reduces to  $\Pi^l$ :  $\Pi \rightsquigarrow \Pi^l$ . Typically we have the situation of Figure 1.

By C3 and the bottom-up definition of  $Q$ , for every inference *inf* in  $\Pi_1$ ,  $Q$  must contain the whole congruence class of *inf*, if  $Q$  is inhabited at all. By C4,  $Q$  only consists of consequent parts. Hence, by C2 and C6, the result of such a reduction is in fact a proof of (3), since on the path from (1) to  $Z \rightarrow A$  we have the same sequence of inference rules being applied as on the path from (3) to  $Z \rightarrow Y$ . If the cut-formula  $A$  is parametric in the inference ending in (2), we rely on C7 instead of C6 and obtain a proof  $\Pi^r$ .

If the tree of parametric ancestors of the displayed occurrence of  $A$  in (1) contains more than one element  $A_u$  that is not parametric in *inf*, parametric and principal moves have to be combined. If  $A$  is non-parametric in the final inference of  $\Pi_2$ , we apply to  $\Pi^l$  a principal move on every cut with  $\Pi_2$ . Call



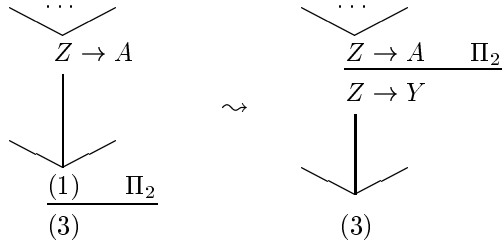


Figure 1.

the resulting proof  $\Pi^{l*}$ :  $\Pi \rightsquigarrow \Pi^{l*}$ . If  $A$  is parametric in the final inference of  $\Pi_2$ , consider  $\Pi^{lr}$ . We apply to  $\Pi^{lr}$  a principal move on every cut with any subproof of  $\Pi_2$  ending in a sequent containing a parametric ancestor  $A_u$ . Call the resulting proof  $\Pi^{lr*}$ :  $\Pi \rightsquigarrow \Pi^{lr*}$ . Thus, if the tree of parametric ancestors of the displayed occurrence of  $A$  in (1) contains more than one element  $A_u$  that is not parametric in *inf*, the primitive reduction of  $\Pi$  gives a proof that is calculated via some intermediate steps. Moreover, instead of a cut with cut-formula  $A$ , we obtain several cuts with subformulas of  $A$  as the cut-formula. Here is a worked out example:

$$\Pi = \frac{\frac{\frac{\frac{\frac{\frac{\Pi^1}{*(A \circ B) \circ X \rightarrow (A \circ B)}}{*(A \circ B) \circ X \rightarrow (A \vee B)}}{*(A \vee B) \circ X \rightarrow (A \circ B)}}{*(A \vee B) \circ X \rightarrow (A \vee B)}}{X \rightarrow (A \vee B) \circ (A \vee B)}}{X \rightarrow (A \vee B)}}{\frac{\frac{\frac{\frac{\Pi^{2_1}}{A \rightarrow Y} \quad \frac{\Pi^{2_2}}{B \rightarrow Z}}{(A \vee B) \rightarrow (Y \circ Z)}}{(A \vee B) \rightarrow (Y \circ Z) \circ W}}{X \rightarrow (Y \circ Z) \circ W}}{X \rightarrow (Y \circ Z) \circ W}}$$

$$\Pi^l = \frac{\frac{\frac{\frac{\frac{\frac{\Pi^1}{*(A \circ B) \circ X \rightarrow (A \circ B)}}{*(A \circ B) \circ X \rightarrow (A \vee B)}}{*(A \circ B) \circ X \rightarrow (Y \circ Z) \circ W}}{*(A \circ B) \circ X \rightarrow (Y \circ Z) \circ W}}{*((Y \circ Z) \circ W) \circ X \rightarrow (A \circ B)}}{*((Y \circ Z) \circ W) \circ X \rightarrow (A \vee B)}}{*((Y \circ Z) \circ W) \circ X \rightarrow (Y \circ Z) \circ W}}{X \rightarrow ((Y \circ Z) \circ W) \circ ((Y \circ Z) \circ W)}}{X \rightarrow ((Y \circ Z) \circ W)}$$

$$\begin{array}{c}
\begin{array}{c}
\Pi^1 \\
\frac{*(A \circ B) \circ X \rightarrow (A \circ B)}{*(A \circ B) \circ X \rightarrow (A \vee B)} \quad \frac{\Pi^{2_1} \quad \Pi^{2_2}}{A \rightarrow Y \quad B \rightarrow Z} \\
\frac{*(A \circ B) \circ X \rightarrow (A \vee B)}{*(A \circ B) \circ X \rightarrow (Y \circ Z)} \\
\frac{*(A \circ B) \circ X \rightarrow (Y \circ Z) \circ W}{*(Y \circ Z) \circ W \circ X \rightarrow (A \circ B)} \quad \frac{\Pi^{2_1} \quad \Pi^{2_2}}{A \rightarrow Y \quad B \rightarrow Z} \\
\frac{*(Y \circ Z) \circ W \circ X \rightarrow (A \vee B)}{*(Y \circ Z) \circ W \circ X \rightarrow (Y \circ Z)} \quad \frac{A \rightarrow Y \quad B \rightarrow Z}{(A \vee B) \rightarrow (Y \circ Z)} \\
\frac{*(Y \circ Z) \circ W \circ X \rightarrow (Y \circ Z)}{*((Y \circ Z) \circ W) \circ X \rightarrow (Y \circ Z) \circ W} \\
\frac{*((Y \circ Z) \circ W) \circ X \rightarrow (Y \circ Z) \circ W}{X \rightarrow ((Y \circ Z) \circ W) \circ ((Y \circ Z) \circ W)} \\
\frac{X \rightarrow ((Y \circ Z) \circ W) \circ ((Y \circ Z) \circ W)}{X \rightarrow ((Y \circ Z) \circ W)}
\end{array} \\
\Pi^{lr} =
\end{array}$$

$$\begin{array}{c}
\begin{array}{c}
\Pi^1 \\
\frac{*(A \circ B) \circ X \rightarrow (A \circ B)}{(* (A \circ B) \circ X) \circ *B \rightarrow A} \quad \frac{\Pi^{2_1}}{A \rightarrow Y} \\
\frac{(* (A \circ B) \circ X) \circ *B \rightarrow Y}{*Y \circ (* (A \circ B) \circ X) \rightarrow B} \quad \frac{\Pi^{2_2}}{B \rightarrow Z} \\
\frac{*Y \circ (* (A \circ B) \circ X) \rightarrow Z}{*(A \circ B) \circ X \rightarrow (Y \circ Z)} \\
\frac{*(A \circ B) \circ X \rightarrow (Y \circ Z)}{*(A \circ B) \circ X \rightarrow (Y \circ Z) \circ W} \\
\frac{*(A \circ B) \circ X \rightarrow (Y \circ Z) \circ W}{*((Y \circ Z) \circ W) \circ X \rightarrow A \circ B} \quad \frac{\Pi^{2_1}}{A \rightarrow Y} \\
\frac{*((Y \circ Z) \circ W) \circ X \circ *B \rightarrow A}{(* ((Y \circ Z) \circ W) \circ X) \circ *B \rightarrow Y} \quad \frac{\Pi^{2_2}}{B \rightarrow Z} \\
\frac{*Y \circ (* ((Y \circ Z) \circ W) \circ X) \rightarrow B}{*Y \circ (* ((Y \circ Z) \circ W) \circ X) \rightarrow Z} \\
\frac{*Y \circ (* ((Y \circ Z) \circ W) \circ X) \rightarrow Z}{*((Y \circ Z) \circ W) \circ X \rightarrow (Y \circ Z)} \\
\frac{*((Y \circ Z) \circ W) \circ X \rightarrow (Y \circ Z)}{*((Y \circ Z) \circ W) \circ X \rightarrow ((Y \circ Z) \circ W)} \\
\frac{*((Y \circ Z) \circ W) \circ X \rightarrow ((Y \circ Z) \circ W)}{X \rightarrow ((Y \circ Z) \circ W) \circ ((Y \circ Z) \circ W)} \\
\frac{X \rightarrow ((Y \circ Z) \circ W) \circ ((Y \circ Z) \circ W)}{X \rightarrow ((Y \circ Z) \circ W)}
\end{array} \\
\Pi \rightsquigarrow \Pi^{lr*} =
\end{array}$$

**THEOREM 23.** *Every proper display calculus enjoys strong cut-elimination.*

**Proof.** See Appendix A. ■

**COROLLARY 24.** *Cut is an admissible rule of every proper display calculus.*

Theorem 23 can straightforwardly be applied to **DK** and **DKt**. It can easily be checked that in these systems conditions C1 – C7 are satisfied.

Verification of C8 is also a simple exercise. We have for instance:

$$\begin{array}{c}
 \frac{\frac{\bullet X \rightarrow A \quad A \rightarrow Y}{X \rightarrow [F]A} \quad [F]A \rightarrow \bullet Y}{X \rightarrow \bullet Y} \\
 \\
 \frac{X \rightarrow A \quad \frac{* \bullet * A \rightarrow Y}{* \bullet * X \rightarrow \langle A \rangle} \quad \langle A \rangle \rightarrow Y}{* \bullet * X \rightarrow Y} \\
 \\
 \frac{\bullet X \rightarrow A \quad A \rightarrow Y}{\bullet X \rightarrow Y} \\
 X \rightarrow \bullet Y \\
 \\
 \frac{\frac{X \rightarrow A \quad \frac{* \bullet * A \rightarrow Y}{* \bullet * X \rightarrow \langle A \rangle} \quad \langle A \rangle \rightarrow Y}{* \bullet * X \rightarrow Y}}{X \rightarrow A} \quad \frac{\frac{\frac{* \bullet * A \rightarrow Y}{* Y \rightarrow \bullet * A} \quad \bullet * Y \rightarrow * A}{A \rightarrow * \bullet * Y}}{X \rightarrow * \bullet * Y} \\
 \frac{\bullet * Y \rightarrow * X}{* Y \rightarrow \bullet * X} \\
 * \bullet * X \rightarrow Y.
 \end{array}
 \quad \rightsquigarrow$$

**THEOREM 25.** *Strong cut-elimination holds for  $DK$  and  $DKt$ .*

**COROLLARY 26.**  *$DKt$  is a conservative extension of  $DK$ .*

We shall now briefly consider generalizations of Theorem 23. By conditions C6 and C7, the inference rules of a proper display calculus are closed under simultaneous substitution of arbitrary structures for congruent formulas. The proof of strong normalization can be generalized to logics which for formulas of a certain shape satisfy closure under substitution either only for congruent formulas (of this shape) which are consequent parts or only for congruent formulas (of this shape) which are antecedent parts. In order to extend the proof of strong cut-elimination to such systems, C6 and C7 have to be replaced by the more general condition of *regularity*, see [Belnap, 1990]. A formula  $A$  is defined as *cons-regular* if the following holds: (i) if  $A$  occurs as a consequent parameter of an inference *inf* in a certain rule  $R$ , then  $R$  contains also the inference resulting by replacing every member of the congruence class of  $A$  in *inf* with an arbitrary structure  $X$ , and (ii) if  $A$  occurs as an antecedent parameter of an inference *inf* in a certain rule  $R$ , then  $R$  contains also the inference resulting by replacing every member of the congruence class of  $A$  in *inf* with any structure  $X$  such that  $X \rightarrow A$  is the conclusion of an inference in which  $A$  is *not* parametric. The notion of *ant-regularity* is defined in exactly the dual way. The new condition on rules then is

**C6/C7 Regularity.** Every formula is regular.

A display calculus *simpliciter* is a calculus of sequents satisfying C1 - C5, C6/7, and C8. If a logic can be presented as a display calculus, then it is said to be displayable. Obviously, every properly displayable logic is displayable. Also the parametric moves must be redefined. Suppose in what follows that the cut-formula  $A$  is parametric in both the final inference of  $\Pi_1$  and the final inference of  $\Pi_2$ . Moreover, suppose that the trees of parametric ancestors

of  $A$  in  $\Pi_1$  and in  $\Pi_2$  do not contain any application of cut. If  $A_u$  is the tip of a path of parametric ancestors of  $A$  in  $\Pi_i$ , let  $inf$  be the inference ending in the sequent which contains  $A_u$ . Let us call  $A_u$  significant, if it is not parametric in  $inf$ . Then, in a *proper* display calculus we may choose whether we cut every significant tip  $A_u$  in the tree of parametric ancestors of  $A$  in  $\Pi_1$  with  $\Pi_2$  or whether we cut every significant tip  $A_u$  in the tree of parametric ancestors of  $A$  in  $\Pi_2$  with  $\Pi_1$  to obtain  $\Pi'$  or  $\Pi''$ . Both operations form an essential part in the definition of certain primitive reductions. In a display calculus simpliciter this indeterministic choice has to be abandoned. If the cut-formula is cons-regular, we cut with  $\Pi_2$ , and if the cut-formula is ant-regular, we cut with  $\Pi_1$ . This further restriction on parametric moves does not affect the proof of strong cut-elimination.

**THEOREM 27.** *Every displayable logic enjoys strong cut-elimination.*

A further strengthening of the strong cut-elimination theorem has recently been proved in [Demri and Goré, 1999], where it is shown that condition C8 may be relaxed. A proof  $\Pi$  ending in a principal application of cut may also be replaced by a proof  $\Pi'$  of the same sequent if the degree of any application of cut in  $\Pi'$  is the same as the degree of the cut-formula in  $\Pi$ , and in  $\Pi'$ , every inference except possibly one falls under a structural rule with a single premise. Moreover, in [Demri and Goré, 1999] a display sequent calculus for the minimal nominal tense logic is defined, and it is shown that every extension of this calculus by structural rules satisfying conditions C1 – C7 enjoys strong cut-elimination.

### 3.4 Kracht's algorithm

The class of all properly displayable normal propositional tense logics has been characterized by Kracht [1996]. The idea is to obtain a canonical way of capturing axiomatic extensions of  $\mathbf{Kt}$  by purely structural inference rules over  $\mathbf{DKt}$ .

**DEFINITION 28.** Let  $\mathbf{Kt} + \alpha$  be an extension of  $\mathbf{Kt}$  by a tense logical axiom schema  $\alpha$ , and let  $\mathbf{DKt} + \alpha'$  be an extension of  $\mathbf{DKt}$  by a set  $\alpha'$  of purely structural inference rules.  $\mathbf{Kt} + \alpha$  is said to be *properly displayed* by  $\mathbf{DKt} + \alpha'$  if (i)  $\mathbf{DKt} + \alpha'$  is a proper display calculus and (ii) every derived rule of  $\mathbf{Kt} + \alpha$  is the  $\tau$ -translation of a sequent rule derivable in  $\mathbf{DKt} + \alpha'$ .

Now, every axiom schema is equivalent to a schema of the form  $A \supset B$ , where  $A$  and  $B$  are implication-free. The schema  $A \supset B$  has the same deductive strength as the rule

$$B \rightarrow X \vdash A \rightarrow X.$$

Moreover, if  $A$  and  $B$  are only built up from propositional variables,  $t$ ,  $\wedge$ ,

$\vee$ ,  $\langle F \rangle$ , and  $\langle P \rangle$ , then by classical logic and distribution of  $\langle F \rangle$  and  $\langle P \rangle$  over disjunction, we have

$$A \equiv \bigvee_{i \leq m} C_i \quad \text{and} \quad B \equiv \bigvee_{j \leq n} D_j,$$

where every  $C_i$  and  $D_j$  is only built up from  $t$ ,  $\wedge$ ,  $\langle F \rangle$ , and  $\langle P \rangle$ . Therefore  $A \supset B$  may as well be replaced by the rule schemata

$$\frac{D_1 \rightarrow Y \quad \dots \quad D_n \rightarrow Y}{C_i \rightarrow Y}$$

These rule schemata can now be translated into purely structural display sequent rules, using the following translation  $\eta$  from formulas of the fragment under consideration into structures:

$$\begin{aligned} \eta(p) &= p & \eta(t) &= \mathbf{I} \\ \eta(\langle F \rangle A) &= * \bullet * \eta(A) & \eta(\langle P \rangle A) &= \bullet \eta(A) \\ \eta(A \wedge B) &= \eta(A) \wedge \eta(B) \end{aligned}$$

The resulting structural rules

$$\frac{\eta(D_1) \rightarrow Y \quad \dots \quad \eta(D_n) \rightarrow Y}{\eta(C_i) \rightarrow Y}$$

may still violate condition C3. In order to avoid this obstruction of proper display, it must be required that in the inducing schema  $A \supset B$ , the schematic formula  $A$  contains each formula variable *only once*. A tense logical formula schema is then said to be *primitive* if it has the form  $A \supset B$ ,  $A$  contains each formula variable only once, and  $A$ ,  $B$  are built up from  $t$ ,  $\wedge$ ,  $\vee$ ,  $\langle F \rangle$ , and  $\langle P \rangle$ .

**LEMMA 29.** *Every extension of  $\mathbf{Kt}$  by primitive axiom schemata can be properly displayed.*

Next, if  $\mathbf{DKt} + \alpha'$  properly displays  $\mathbf{Kt} + \alpha$ , by condition (ii) of Definition 28, the structural rules in  $\alpha'$  may all have the form

$$\frac{X_1 \rightarrow Y \quad \dots \quad X_n \rightarrow Y}{Z \rightarrow Y}$$

This rule has the same deductive strength as the axiom schema

$$\tau_1(Z) \supset \bigvee_i \tau_1(X_i),$$

which is a primitive formula schema.

**THEOREM 30.** (Kraicht) *An axiomatic extension of  $\mathbf{Kt}$  can be properly displayed in precisely the case that it is axiomatizable by a set of primitive axiom schemata.*

The question whether an axiomatically presented normal temporal logic  $\Lambda$  is properly displayable thus reduces to the question whether  $\Lambda$  can be axiomatized by primitive axioms over  $\mathbf{Kt}$ . The implicit use of tense logic in the structural language of sequents may help to find simple structural sequent rules expressing less simple modal axiom schemata. The following example is taken from [Kracht, 1996]. The .3 axiom schema  $\Box(\Box A \supset \Box B) \vee \Box(\Box B \supset \Box A)$  has the primitive modal equivalent

$$(\Diamond A \wedge \Diamond B) \supset ((\Diamond(A \wedge \Diamond B) \vee \Diamond(B \wedge \Diamond A)) \vee \Diamond(A \wedge B)),$$

which in tense logic is equivalent to the simpler primitive schema

$$\langle P \rangle \langle F \rangle A \supset ((\langle F \rangle A \vee A) \vee \langle P \rangle A).$$

Application of Kracht's algorithm results in the following structural rule:

$$X \rightarrow Y \quad \bullet X \rightarrow Y \quad * \bullet * X \rightarrow Y \vdash \bullet * \bullet * X \rightarrow Y.$$

Kracht also proves a semantic characterization of the properly displayable tense logics. Let  $\mathcal{F}$  be a class of Kripke frames  $\langle W, \mathcal{R}, \mathcal{R}^{-1} \rangle$  for temporal logics, where  $\mathcal{R}^{-1}$  is the inverse of  $\mathcal{R}$  (i.e.,  $\mathcal{R} = \{(x, y) \mid (y, x) \in \mathcal{R}\}$ ). A first-order sentence (open formula) over two binary relation symbols  $R$  and  $R^{-1}$  is said to be *primitive* if it has the form  $(\forall)(\exists)A$ , where every quantifier is restricted with respect to  $R$  or  $R^{-1}$ , and  $A$  is built up from  $\wedge$ ,  $\vee$ , and atomic formulas  $x = y$ ,  $xRy$ ,  $xR^{-1}y$ , where at least one of  $x, y$  is not in the scope of an existential quantifier.

**THEOREM 31.** (Kracht) *A class  $\mathcal{F}$  of Kripke frames for temporal logics is describable by a set of primitive first-order sentences iff the tense logic of  $\mathcal{F}$  can be properly displayed.*

The characteristic axiom schemata of quite a few fundamental systems of modal and tense logic are equivalent to primitive schemata, and therefore these systems can be presented as proper display calculi, cf. Table 3.<sup>7</sup> A set of structural sequent rules  $\alpha'$  is said to *correspond* to a property of an accessibility relation  $\mathcal{R}$  (with a modal or tense logical axiom schema  $\alpha$ ) iff under the  $\tau$ -translation the rules in  $\alpha'$  are admissible just in the event that

<sup>7</sup>Goré recently observed that Theorem 20 in [Kracht, 1996] is incorrect. This theorem states that an axiomatic extension of  $\mathbf{K}$  can be properly displayed iff it is axiomatizable by a set of primitive *modal* axiom schemata. There are, however, first-order frame properties that correspond to a primitive tense logical schema but fail to correspond to a primitive *modal* axiom schema. An example of such a frame property is weak directedness:

$$\forall s \forall t \forall u (sRt \wedge sRu \supset \exists v (tRv \wedge uRv)).$$

Weak directedness corresponds to the .2 schema  $\Diamond \Box A \supset \Box \Diamond A$  (alias  $\langle F \rangle [F] A \supset [F] \langle F \rangle A$ ). Although .2 has no primitive modal equivalent, it has a primitive tense logical equivalent, namely  $\langle P \rangle \langle F \rangle A \supset \langle F \rangle \langle P \rangle A$ . The latter schema induces a structural rule that may be added to display calculi for (extensions of)  $\mathbf{K}$ . Therefore,  $\mathbf{K.2}$  is properly displayable, although .2 is not primitive.

$\mathcal{R}$  enjoys the property (the rules in  $\alpha'$  have the same deductive strength as  $\alpha$ ). Every axiom schema  $\alpha$  in Table 3 corresponds to a purely structural sequent rule  $\alpha'$  which can directly be determined from  $\alpha$ , see Table 4.

<i>schema</i>	<i>primitive equivalent</i>
$D$ $[F]A \supset \langle F \rangle A$	$\mathbf{t} \supset \langle F \rangle \mathbf{t}$
$T$ $[F]A \supset A$	$A \supset \langle F \rangle A$
$4$ $[F]A \supset [F][F]A$	$\langle F \rangle \langle F \rangle A \supset \langle F \rangle A$
$5$ $\langle F \rangle A \supset [F]\langle F \rangle A$	$\langle P \rangle \langle F \rangle A \supset \langle F \rangle A$
$B$ $A \supset [F]\langle F \rangle A$	$(A \wedge \langle F \rangle B) \supset \langle F \rangle (B \wedge \langle F \rangle A)$
$Alt1$ $\langle F \rangle A \supset [F]A$	$(\langle F \rangle A \wedge \langle F \rangle B) \supset \langle F \rangle (A \wedge B)$
$T^c$ $A \supset [F]A$	$\langle F \rangle A \supset A$
$4^c$ $[F][F]A \supset [F]A$	$\langle F \rangle A \supset \langle F \rangle \langle F \rangle A$
$.2$ $\langle F \rangle [F]A \supset [F]\langle F \rangle A$	$\langle P \rangle \langle F \rangle A \supset \langle F \rangle \langle P \rangle A$
$.3$ $[F]([F]A \supset [F]B) \vee [F]([F]B \supset [F]A)$	$\langle P \rangle \langle F \rangle A \supset (((F)A \vee A) \vee \langle P \rangle A)$
$\bar{l}inf$ $\langle F \rangle A \supset [F](((F)A \vee A) \vee \langle P \rangle A)$	$\langle P \rangle \langle F \rangle A \supset (((F)A \vee A) \vee \langle P \rangle A)$
$\bar{l}inp$ $\langle P \rangle A \supset [P](((P)A \vee A) \vee \langle F \rangle A)$	$\langle F \rangle \langle P \rangle A \supset (((P)A \vee A) \vee \langle F \rangle A)$
$V$ $[F]A$	$\langle P \rangle \mathbf{t} \supset A$
$D_p$ $[P]A \supset \langle P \rangle A$	$\mathbf{t} \supset \langle P \rangle \mathbf{t}$
$T_p$ $[P]A \supset A$	$A \supset \langle P \rangle A$
$4_p$ $[P]A \supset [P][P]A$	$\langle P \rangle \langle P \rangle A \supset \langle P \rangle A$
$5_p$ $\langle P \rangle A \supset [P]\langle P \rangle A$	$\langle F \rangle \langle P \rangle A \supset \langle P \rangle A$
$B_p$ $A \supset [P]\langle P \rangle A$	$(A \wedge \langle P \rangle B) \supset \langle P \rangle (B \wedge \langle P \rangle A)$
$Alt1_p$ $\langle P \rangle A \supset [P]A$	$(\langle P \rangle A \wedge \langle P \rangle B) \supset \langle P \rangle (A \wedge B)$
$T_p^c$ $A \supset [P]A$	$\langle P \rangle A \supset A$
$4_p^c$ $[P][P]A \supset [P]A$	$\langle P \rangle A \supset \langle P \rangle \langle P \rangle A$
$V_p$ $[P]A$	$\langle F \rangle \mathbf{t} \supset A$

Table 3. Axioms and primitive axioms.

Let  $\Gamma(\Theta)$  be the set of all (all purely modal) axiom schemata from Table 3,  $\bar{\Gamma} \subseteq \Gamma$ ,  $\bar{\Theta} \subseteq \Theta$ ,  $\Gamma' = \{\alpha' \mid \alpha \in \bar{\Gamma}\}$ , and  $\Theta' = \{\alpha' \mid \alpha \in \bar{\Theta}\}$ .

**THEOREM 32.** *In  $\mathbf{DKt}\cup\Gamma'$ ,  $\vdash X \rightarrow Y$  iff  $\vdash \tau_1(X) \supset \tau_2(Y)$  in  $\mathbf{Kt}\cup\Gamma$ . In  $\mathbf{DK}\cup\Theta'$ ,  $\vdash X \rightarrow Y$  iff  $\vdash \tau_1(X) \supset \tau_2(Y)$  in  $\mathbf{K}\cup\Theta$ .*

**Proof.** This follows from axiomatizability by primitive schemata. ■

**THEOREM 33.** *Strong cut-elimination holds for  $\mathbf{DKt}\cup\Gamma'$  and  $\mathbf{DK}\cup\Theta'$ .*

**Proof.** The rules in  $\Gamma'$  and  $\Theta'$  satisfy conditions C2 – C7. ■

**COROLLARY 34.**  *$\mathbf{DKt}\cup\Gamma'$  is a conservative extension of  $\mathbf{DK}\cup\Gamma'$ .*

Kracht's algorithm can be dualized. Every schema  $A \supset B$  is interreplaceable with the rule

$$X \rightarrow A \vdash X \rightarrow B.$$

$D'$	$* \bullet * \mathbf{I} \rightarrow Y \vdash \mathbf{I} \rightarrow Y$
$T'$	$* \bullet * X \rightarrow Y \vdash X \rightarrow Y$
$A'$	$* \bullet * X \rightarrow Y \vdash * \bullet \bullet * X \rightarrow Y$
$\bar{5}'$	$* \bullet * X \rightarrow Y \vdash \bullet * \bullet * X \rightarrow Y$
$B'$	$* \bullet *(X \circ * \bullet * Y) \rightarrow Z \vdash Y \circ * \bullet * X \rightarrow Z$
$AltI'$	$* \bullet *(X \circ Y) \rightarrow Z \vdash * \bullet * X \circ * \bullet * Y \rightarrow Z$
$T^{c'}$	$X \rightarrow Y \vdash * \bullet * X \rightarrow Y$
$4^{c'}$	$* \bullet \bullet * X \rightarrow Y \vdash * \bullet * X \rightarrow Y$
$.2'$	$* \bullet * \bullet X \rightarrow Y \vdash \bullet * \bullet * X \rightarrow Y$
$.3'$	$X \rightarrow Y \quad \bullet X \rightarrow Y \quad * \bullet * X \rightarrow Y \vdash \bullet * \bullet * X \rightarrow Y$
$linf'$	$= .3'$
$linp'$	$X \rightarrow Y \quad \bullet X \rightarrow Y \quad * \bullet * X \rightarrow Y \vdash \bullet * \bullet * X \rightarrow Y$
$V'$	$X \rightarrow Y \vdash \bullet \mathbf{I} \rightarrow Y$
$D_p'$	$\bullet \mathbf{I} \rightarrow Y \vdash \mathbf{I} \rightarrow Y$
$T_p'$	$\bullet X \rightarrow Y \vdash X \rightarrow Y$
$A_p'$	$\bullet X \rightarrow Y \vdash \bullet \bullet X \rightarrow Y$
$\bar{5}_p'$	$\bullet X \rightarrow Y \vdash * \bullet \bullet \bullet X \rightarrow Y$
$B_p'$	$\bullet (X \circ \bullet Y) \rightarrow Z \vdash Y \circ \bullet X \rightarrow Z$
$AltI_p'$	$\bullet (X \circ Y) \rightarrow Z \vdash \bullet X \circ \bullet Y \rightarrow Z$
$T_p^{c'}$	$X \rightarrow Y \vdash \bullet X \rightarrow Y$
$4_p^{c'}$	$\bullet \bullet X \rightarrow Y \vdash \bullet X \rightarrow Y$
$V_p'$	$X \rightarrow Y \vdash * \bullet * \mathbf{I} \rightarrow Y$

Table 4. Structural rules corresponding to axiom schemata.

If  $A$  and  $B$  are only built up from propositional variables,  $\mathbf{f}$ ,  $\wedge$ ,  $\vee$ ,  $[F]$ , and  $[P]$ , then by classical logic and distribution of  $[F]$  and  $[P]$  over conjunction, we have

$$A \equiv \bigwedge_{i \leq m} C_i \quad \text{and} \quad B \equiv \bigwedge_{j \leq n} D_j,$$

where every  $C_i$  and  $D_j$  is only built up from  $\mathbf{f}$ ,  $\vee$ ,  $[F]$ , and  $[P]$ . Therefore  $A \supset B$  may be replaced by the rule schemata

$$\frac{X \rightarrow C_1 \quad \dots \quad X \rightarrow C_m}{X \rightarrow D_j}.$$

These schemata are translatable into purely structural sequent rules using the following translation  $\eta'$  from formulas of the fragment under consideration into structures:

$$\begin{aligned} \eta'(p) &= p & \eta'(\mathbf{f}) &= \mathbf{I} \\ \eta'([F]A) &= \bullet \eta'(A) & \eta'([P]A) &= * \bullet * \eta'(A) \\ \eta'(A \vee B) &= \eta'(A) \vee \eta'(B) \end{aligned}$$



The resulting structural rules

$$\frac{X \rightarrow \eta'(C_1) \ \dots \ X \rightarrow \eta'(C_m)}{X \rightarrow \eta'(D_j)}$$

again may still violate condition C3. In order to avoid the obstruction of proper display, it must be required that in the inducing schema  $A \supset B$ , the schematic formula  $B$  contains each formula variable only once. A tense logical formula schema is then said to be *dually primitive* if it has the form  $A \supset B$ ,  $B$  contains each formula variable only once, and  $A, B$  are built up from  $\mathbf{f}$ ,  $\wedge$ ,  $\vee$ ,  $[F]$ , and  $[P]$ .

**THEOREM 35.** *An axiomatic extension of  $\mathbf{Kt}$  can be properly displayed iff it is axiomatizable by a set of dually primitive axiom schemata.*

For instance, rule  $T'$  is equivalent to  $X \rightarrow \bullet Y \vdash X \rightarrow Y$  and  $4'$  with  $X \rightarrow \bullet Y \vdash X \rightarrow \bullet \bullet Y$ . Moreover,  $D'$  is equivalent to  $\bullet X \circ \bullet Y \rightarrow \bullet \mathbf{I} \vdash X \rightarrow \bullet Y$ ,  $AltI'$  with  $X \rightarrow Y \vdash X \rightarrow \bullet \bullet \bullet Y$ , and  $V'$  with  $\vdash \bullet \mathbf{I} \rightarrow X$ , see [Wansing, 1994].

The properly displayable modal and tense logics satisfy Došen's Principle. They are all based on the same set of left and right introduction rules, so that the logical operations indeed have the same proof-theoretic, operational meaning in each of these systems. Kracht's characterization results show that many interesting and important intensional logics admit a cut-free display sequent calculus presentation. In Sections 3.8 and 4 other applications of the display calculus are pointed out. Display sequent systems for various non-normal modal logics may be found in [Belnap, 1982].

### 3.5 Formulas-as-types for temporal logics

It is well-known that every derivation in Gentzen's natural deduction calculus for intuitionistic implicational logic can be encoded by a typed  $\lambda$ -term, and vice versa [Howard, 1980]. In particular, every natural deduction proof can be encoded by a closed term, and every closed term encodes a proof. It is also well-known that every pair of non-convertible typed  $\lambda$ -terms defines different functionals of finite type [Friedman, 1975]. Every type  $A$  is associated with an infinite set  $D^A$ , every term variable  $x^A$  of type  $A$  denotes an element from  $D^A$ , and every term  $M^{(A \triangleright B)}$  of type  $A \triangleright B$  denotes an element from the set  $(D^B)^{D^A}$  of all functions from  $D^A$  to  $D^B$ . Together with the encoding, this interpretation results in a set-theoretic semantics of proofs in intuitionistic implicational logic. In this section, we shall develop a set-theoretic interpretation of sequent proofs in the  $\{t, [F], \langle P \rangle, \triangleright, \wedge\}$ -fragment of the smallest normal temporal intuitionistic (or, for that purpose, minimal) logic  $\mathbf{IntKt}$ . The interpretation is based on the observation that the modalities  $\langle P \rangle$  and  $[F]$  form a residuated pair with respect to derivability. The encoding of proofs by typed terms should be such that

proof-simplification (or normalization) corresponds with a suitable reduction relation on terms, and therefore the set-theoretic semantics of terms has to validate the equalities underlying the reduction rules. The principal cut-elimination steps for  $\langle P \rangle$  and  $[F]$  reveal that two pairs of term forming operations  $o_1$  and  $o_2$  are needed such that  $o_1(o_2(M)) = M$ . We shall use the following identities:

$$\bigcup \mathcal{P}a = a \quad \text{and} \quad \bigcap \mathcal{S}a = a,$$

where  $\mathcal{P}$  is the familiar powerset operation and  $\mathcal{S}a =_{\text{def}} \{b \mid a \subseteq b\}$ . Since in general  $\mathcal{S}a$  is a proper class, we shall restrict the denotations of terms to the universe  $V_{\omega_1}$ . This is enough to accommodate the sets used as domains of the intended models in Section 3.7.

We shall first define a display sequent system **DIntKt** for the fragment of **IntKt** under consideration, and then present an extension  $\lambda_t$  of the typed  $\lambda$ -calculus. The set of types in  $\lambda_t$  is the set of all formulas in the language  $\mathcal{L} = \{t, [F], \langle P \rangle, \triangleright, \wedge\}$  based on a denumerable set *Atom* of propositional variables. In Section 3.6 it is proved that term reduction is a homomorphic image of proof-simplification. Next, an encoding of terms by proofs is presented. A set-theoretic semantics of proofs in **DIntKt** is obtained in Section 3.7 by showing that every pair of non-convertible  $\lambda_t$ -terms defines different sets in the set-theoretic universe under consideration. In particular, every term  $M^{[F]A}$  denotes an element from  $\{\mathcal{P}a \mid a \in D^A\}$ , and every term  $M^{\langle P \rangle A}$  denotes an element from  $\{\mathcal{S}a \mid a \in D^A\}$ . Also the formulas-as-types notion of construction for various extensions of **DIntKt** is dealt with and remarks on some related work about formulas-as-types for modal logics are made.

First, we shall define the sequent system **DIntKt**. We assume the following language of structures:

$$X ::= A \mid \mathbf{I} \mid \bullet X \mid X \times Y.$$

A sequent now is an expression  $X \rightarrow Y$ , provided  $Y \neq \mathbf{I}$ . The declarative meaning of the structure connectives can be made explicit by a translation  $\tau$  from the set of sequents into the set of  $\mathcal{L}$ -formulas:

$$\tau(X \rightarrow Y) := \tau_1(X) \triangleright \tau_2(Y),$$

where  $\tau_i$  ( $i = 1, 2$ ) is defined as follows:

$$\begin{array}{ll} \tau_i(A) & = A & \tau_1(\mathbf{I}) & = t \\ \tau_1(X \times Y) & = \tau_1(X) \wedge \tau_1(Y) & \tau_2(X \times Y) & = \tau_1(X) \triangleright \tau_2(Y) \\ \tau_1(\bullet X) & = \langle P \rangle \tau_1(X) & \tau_2(\bullet X) & = [F] \tau_2(X) \end{array}$$

Given this understanding of the structure connectives, the basic structural rules (4) and (5) from Section 3.1 are assumed. Clearly, the Display Theorem holds for this structural language and calculus.

DEFINITION 36. The display sequent calculus **DIntKt** is given by the logical rules (id) and (cut), the basic structural rules (4) and (5), the introduction rules for  $\mathbf{t}$ ,  $\triangleright$ ,  $\langle P \rangle$ ,  $[F]$ , and the rules  $(\rightarrow \wedge)'$  and  $(\wedge \rightarrow)'$ , together with the following structural rules:

$$\begin{array}{l}
 \text{(empty structure)} \quad X \rightarrow Y \vdash \mathbf{I} \times X \rightarrow Y, \quad X \rightarrow Y \vdash X \times \mathbf{I} \rightarrow Y \\
 \quad \quad \quad \mathbf{I} \times X \rightarrow Y \vdash X \rightarrow Y, \quad X \times \mathbf{I} \rightarrow Y \vdash X \rightarrow Y \\
 \text{(associativity)} \quad (X_1 \times X_2) \times X_3 \rightarrow Y \vdash X_1 \times (X_2 \times X_3) \rightarrow Y \\
 \text{(permutation)} \quad X \times Y \rightarrow Z \vdash Y \times X \rightarrow Z \\
 \text{(contraction)} \quad X \times X \rightarrow Y \vdash X \rightarrow Y \\
 \text{(expansion)} \quad X \rightarrow Y \vdash X \times X \rightarrow Y \\
 \text{(monotonicity)} \quad X \rightarrow Z \vdash X \times Y \rightarrow Z, \quad X \rightarrow Z \vdash Y \times X \rightarrow Z \\
 \text{(necessitation)} \quad \mathbf{I} \rightarrow X \vdash \bullet \mathbf{I} \rightarrow X.
 \end{array}$$

To show that **DIntKt** is a display calculus for **IntKt**, we define an axiomatic calculus **HIntKt**.

DEFINITION 37. The system **HIntKt** consists of the axiom schemata and rules of the  $\{\mathbf{t}, \wedge, \triangleright\}$ -fragment of positive intuitionistic logic, together with

1.  $([F]A \wedge [F]B) \triangleright [F](A \wedge B)$
2.  $[F]\mathbf{t}$
3.  $A \triangleright [F]\langle P \rangle A$
4.  $\frac{\vdash A \triangleright B}{\vdash [F]A \triangleright [F]B}$
5.  $\frac{\vdash A \triangleright B}{\vdash \langle P \rangle A \triangleright \langle P \rangle B}$

The relational semantics to be presented is a straightforward adaptation of the semantics developed by Bošić and Došen [1984]. A comprehensive survey of intuitionistic modal logics and their algebraic and relational semantics is [Wolter and Zakharyashev, 1999]. A temporal frame is defined as a structure  $\langle W, R_I, R_T \rangle$ , where  $W$  is a non-empty set (of states),  $R_I$  and  $R_T$  are binary relations on  $W$ ,  $R_I$  is both reflexive and transitive, and, moreover, (i)  $R_I R_T \subseteq R_T R_I$  (i.e. the composition of  $R_T$  and  $R_I$  is a subset of the composition of  $R_I$  and  $R_T$ ) and (ii)  $R_I^{-1} R_T^{-1} \subseteq R_T^{-1} R_I^{-1}$ . If  $\mathcal{F} = \langle W, R_I, R_T \rangle$  is a temporal frame, the temporal model based on  $\mathcal{F}$  is the structure  $\langle \mathcal{F}, v \rangle$ , where  $v$  is a function from  $Atom \times W$  into  $\{0, 1\}$  satisfying:

$$\text{(Heredity)} \quad (v(p, u) = 1 \text{ and } u R_I t) \text{ implies } v(p, t) = 1.$$

Let  $\mathcal{M} = \langle W, R_I, R_T, v \rangle$  be a temporal model. Verification of a formula  $A$  at a state  $u \in W$  ( $\mathcal{M}, u \models A$ ) is inductively defined as follows:

$$\begin{aligned}
\mathcal{M}, u \models p & \quad \text{iff} \quad v(p, u) = 1 \\
\mathcal{M}, u \models t & \\
\mathcal{M}, u \models A \wedge B & \quad \text{iff} \quad \mathcal{M}, u \models A \text{ and } \mathcal{M}, u \models B \\
\mathcal{M}, u \models A \triangleright B & \quad \text{iff} \quad (\forall t \in W) uR_I t \text{ implies } [\mathcal{M}, t \not\models A \text{ or } \mathcal{M}, t \models B] \\
\mathcal{M}, u \models [F]A & \quad \text{iff} \quad (\forall t \in W) uR_T t \text{ implies } \mathcal{M}, t \models A \\
\mathcal{M}, u \models \langle P \rangle A & \quad \text{iff} \quad (\exists t \in W) tR_T u \text{ and } \mathcal{M}, t \models A
\end{aligned}$$

For every formula  $A$ , if  $A$  is verified at state  $u$  and  $uR_I t$ , then  $A$  is also verified at  $t$ . Condition (i) ensures this general heredity property for formulas  $[F]A$ , and condition (ii) ensures it for formulas  $\langle P \rangle A$ . A formula  $A$  is true in a model  $\langle W, R_T, R_I, v \rangle$  if  $A$  is verified at every  $u \in W$ , and  $A$  is said to be true on a frame  $\mathcal{F}$ , if  $A$  is valid in every model based on  $\mathcal{F}$ . If  $\mathcal{K}$  is a class of models (frames),  $A$  is said to be valid in  $\mathcal{K}$  iff  $A$  is valid in every model (valid on every frame) in  $\mathcal{K}$ .

**THEOREM 38.** ***HIntKt** is sound and complete with respect to the class of all temporal frames, i.e. for every  $\mathcal{L}$ -formula  $A$ ,  $A$  is provable in **HIntKt** iff  $A$  is valid in the class of all temporal frames.*

**Proof.** Soundness is shown by induction on proofs in **HIntKt**; for completeness see Appendix B.  $\blacksquare$

**LEMMA 39.** (1) *If  $\vdash A$  in **HIntKt**, then  $\vdash \mathbf{I} \rightarrow A$  in **DIntKt**, and (2) *If  $\vdash X \rightarrow Y$  in **DIntKt**, then  $\vdash \tau(X \rightarrow Y)$  in **HIntKt**.**

**Proof.** (1) By induction on proofs in **HIntKt**. We shall consider only two example cases:

$$\begin{array}{c}
\frac{A \rightarrow A}{\bullet A \rightarrow \langle P \rangle A} \\
\frac{A \rightarrow [F] \langle P \rangle A}{A \times \mathbf{I} \rightarrow [F] \langle P \rangle A} \\
\frac{\mathbf{I} \rightarrow A \times [F] \langle P \rangle A}{\mathbf{I} \rightarrow A \triangleright [F] \langle P \rangle A}
\end{array}
\qquad
\begin{array}{c}
\frac{A \rightarrow A}{[F]A \rightarrow \bullet A} \qquad \frac{B \rightarrow B}{[F]B \rightarrow \bullet B} \\
\frac{[F]A \rightarrow \bullet A}{[F]A \times [F]B \rightarrow \bullet A} \qquad \frac{[F]B \rightarrow \bullet B}{[F]A \times [F]B \rightarrow \bullet B} \\
\frac{\bullet([F]A \times [F]B) \rightarrow \bullet A \qquad \bullet([F]A \times [F]B) \rightarrow \bullet B}{\bullet([F]A \times [F]B) \rightarrow A \wedge B} \\
\frac{\bullet([F]A \times [F]B) \rightarrow A \wedge B}{([F]A \times [F]B) \rightarrow [F](A \wedge B)} \\
\frac{([F]A \times [F]B) \rightarrow [F](A \wedge B)}{([F]A \wedge [F]B) \rightarrow [F](A \wedge B)} \\
\frac{([F]A \wedge [F]B) \rightarrow [F](A \wedge B)}{([F]A \wedge [F]B) \times \mathbf{I} \rightarrow [F](A \wedge B)} \\
\frac{([F]A \wedge [F]B) \times \mathbf{I} \rightarrow [F](A \wedge B)}{\mathbf{I} \rightarrow ([F]A \wedge [F]B) \times [F](A \wedge B)} \\
\frac{\mathbf{I} \rightarrow ([F]A \wedge [F]B) \times [F](A \wedge B)}{\mathbf{I} \rightarrow ([F]A \wedge [F]B) \triangleright [F](A \wedge B)}
\end{array}$$

(2) By induction on proofs in **DIntKt**.  $\blacksquare$

COROLLARY 40. In  $\mathbf{HIntKt}$ ,  $\vdash A$  iff  $\vdash \mathbf{I} \rightarrow A$  in  $\mathbf{DIntKt}$ .

By induction on the complexity of  $X$ , one can prove the following

LEMMA 41. In every extension of  $\mathbf{DIntKt}$  by structural rules, it holds that  $\vdash X \rightarrow \tau_1(X)$  and  $\vdash \tau_2(X) \rightarrow X$ .

THEOREM 42. In  $\mathbf{DIntKt}$ ,  $\vdash X \rightarrow Y$  iff  $\vdash \tau(X \rightarrow Y)$  in  $\mathbf{HIntKt}$ .

**Proof.** Analogous to the proof of Theorem 20. ■

Since  $\mathbf{DIntKt}$  is a proper display calculus, we have the following

THEOREM 43.  $\mathbf{DIntKt}$  enjoys strong cut-elimination.

Take any terminating cut-elimination algorithm  $elim_c$  for  $\mathbf{DIntKt}$ . We may also define a binary relation  $\rightsquigarrow_s$  on the set of proofs in  $\mathbf{DIntKt}$  by the following stipulations:

$$\frac{\frac{A \rightarrow A \quad B \rightarrow B}{A \times B \rightarrow A \wedge B}}{A \wedge B \rightarrow A \wedge B} \rightsquigarrow_s A \wedge B \rightarrow A \wedge B$$

$$\frac{\frac{A \rightarrow A \quad B \rightarrow B}{A \triangleright B \rightarrow A \times B}}{A \triangleright B \rightarrow A \triangleright B} \rightsquigarrow_s A \triangleright B \rightarrow A \triangleright B$$

If  $\Pi \rightsquigarrow_s \Pi'$ , we say that in  $\Pi'$  a redundant part of  $\Pi$  has been removed. Let  $elim_r$  denote the terminating algorithm that removes redundant parts of a proof in top-down left to right order, so that a redundant part is removed only if it has no redundant part above it. Obviously, in any extension of  $\mathbf{DIntKt}$ , every proof of a sequent  $s$  can be converted into a proof of  $s$  containing no redundant part. Let  $elim$  denote  $elim_r \circ elim_c$ , i.e. the composition of  $elim_r$  and  $elim_c$ . The algorithm  $elim$  is the process of proof simplification to be considered. We assume that  $elim(\Pi) = \Pi$  if  $\Pi$  contains no application of (cut) and no redundant part.

### 3.6 The typed $\lambda$ -calculus $\lambda_t$

The set  $T$  of type symbols (or just types) is the set of all  $\mathcal{L}$ -formulas. The set  $V$  of term variables is defined as  $\{v_i^A \mid 0 < i \in \omega, A \in T\}$ .

DEFINITION 44. The set Term of typed terms is defined as the smallest set  $\Delta$  such that

1.  $V \subseteq \Delta$ ;
2. if  $M^A, N^B \in \Delta$ , then  $\langle M^A, N^B \rangle^{(A \wedge B)} \in \Delta$ ;
3.  $M^{(A \wedge B)} \in \Delta$ , then  $(M^{(A \wedge B)})_0^A, (M^{(A \wedge B)})_1^B \in \Delta$ ;

4. if  $x^A \in V$  and  $M^B \in \Delta$ , then  $(\lambda x^A M^B)^{(A \triangleright B)} \in \Delta$ ;
5. if  $M^{(A \triangleright B)}, N^A \in \Delta$ , then  $(M^{(A \triangleright B)}, N^A)^B \in \Delta$ ;
6. if  $M^A \in \Delta$ , then  $(\mathcal{P}M)^{[F]A}, (\mathcal{S}M)^{\langle P \rangle A} \in \Delta$ ;
7. if  $M^{[F]A} \in \Delta$ , then  $(\cup M^{[F]A})^A \in \Delta$ ;
8. if  $M^{\langle P \rangle A} \in \Delta$ , then  $(\cap M^{\langle P \rangle A})^A \in \Delta$ .

A term  $M^A$  is said to be a term of type  $A$ ; obviously, every term has a unique type. If confusion is unlikely to arise, we shall often write  $M$  instead of  $M^A$  and omit parentheses not needed for disambiguation. The set  $fv(M)$  of free variables of  $M$ , the set of subterms of  $M$ , and  $M[x^A := N^A]$ , the result of substituting term  $N$  of type  $A$  for every occurrence of  $x^A$  in  $M$  are inductively defined in the obvious way. If a variable  $x$  in  $M$  is not an element of  $fv(M)$ ,  $x$  is said to be a bound variable of  $M$ . The set of bound variables of  $M$  is denoted as  $bv(M)$ . We shall also write  $M(x_1^{A_1}, \dots, x_n^{A_n})$  to express that  $x_1, \dots, x_n \in fv(M)$ . If  $M(x_1^{A_1}, \dots, x_n^{A_n})$  and  $N_1, \dots, N_n$  are terms of types  $A_1, \dots, A_n$ , then  $M(N_1, \dots, N_n)$  is the result of substituting in  $M$  the variables  $x_i$  by  $N_i$ . We shall use ‘ $\equiv$ ’ to denote syntactic identity between term.

DEFINITION 45. The typed  $\lambda$ -calculus  $\lambda_t$  consists of the following rules and axiom schemata:

1.  $\lambda x^A M = (\lambda y^A M[x := y])$ , if  $y \notin (fv(M) \cup bv(M))$ ;
2.  $\lambda x(M, x) = M$ , if  $x \notin fv(M)$ ;
3.  $(\lambda x M)N = M[x := N]$ , if  $bv(M) \cup fv(N) = \emptyset$ ;
4.  $\langle (M_0, M_1) \rangle_i = M_i$ ;
5.  $\langle (M)_0, (M)_1 \rangle = M$ ;
6.  $\cup \mathcal{P}M = M$ ;
7.  $\cap \mathcal{S}M = M$ ;
8.  $M^A = M^A$ ;
9.  $M = N \vdash N = M$ ;  $M = N, N = G \vdash M = G$ ;
10.  $M = N \vdash (G, M) = (G, N)$ ;  $M = N \vdash (M, G) = (N, G)$ ;
11.  $M = N \vdash \lambda x M = \lambda x N$ ;
12.  $M = N \vdash \mathcal{P}M = \mathcal{P}N$ ;  $M = N \vdash \cup M = \cup N$ .

DEFINITION 46. The binary relations on Term,  $\rightarrow_r$  (one-step reduction),  $\rightarrow_r$  (reduction), and  $=_r$  (equality) are defined as follows:

1.
  - $\lambda x(Mx) \rightarrow_r M$ , if  $x \notin fv(M)$ ;
  - $(\lambda xM)N \rightarrow_r M[x := N]$ , if  $bv(M) \cup fv(N) = \emptyset$ ;
  - $(\langle M, N \rangle)_0 \rightarrow_r M$ ;  $(\langle M, N \rangle)_1 \rightarrow_r N$ ;
  - $\langle (M)_0, (M)_1 \rangle \rightarrow_r M$ ;
  - $\cup \mathcal{P}M \rightarrow_r M$ ;  $\cap \mathcal{S}M \rightarrow_r M$ ;
  - if  $M^{A \triangleright B} \rightarrow_r N^{A \triangleright B}$ , then  $(M, G^A) \rightarrow_r (N, G)$ ;
  - if  $M^{A \wedge B} \rightarrow_r N^{A \wedge B}$ , then  $(M)_i \rightarrow_r (N)_i$ ;
  - if  $M^A \rightarrow_r N^A$ , then  $\lambda xM \rightarrow_r \lambda xN$ ,  $(G^{A \triangleright B}M) \rightarrow_r (GN)$ ,  $\langle M, G \rangle \rightarrow_r \langle N, G \rangle$ ,  $\langle G, M \rangle \rightarrow_r \langle G, N \rangle$ ,  $\mathcal{P}M \rightarrow_r \mathcal{P}N$ ,  $\mathcal{S}M \rightarrow_r \mathcal{S}N$ ,  $\cap M \rightarrow_r \cap N$ ,  $\cup M \rightarrow_r \cup N$ .
2.  $\rightarrow_r$  is the reflexive transitive closure of  $\rightarrow_r$ ;
3.  $=_r$  is the equivalence relation generated by  $\rightarrow_r$ .

DEFINITION 47.  $\lambda_t$ -terms  $\lambda x(Mx)$  (where  $x \notin fv(M)$ ),  $(\lambda xM)N$  (where  $bv(M) \cup fv(N) = \emptyset$ ),  $(\langle M, N \rangle)_0$ ,  $(\langle M, N \rangle)_1$ ,  $\langle (M)_0, (M)_1 \rangle$ ,  $\cup \mathcal{P}M$ , and  $\cap \mathcal{S}M$  are called redexes. A term  $M$  is a normal form (*nf*) if it has no redex as a subterm, and  $M$  has a *nf* if there is a *nf*  $N$  such that  $M =_r N$ .  $M$  is said to be strongly normalizable with respect to  $\rightarrow_r$  ( $sn(M)$ ) if every sequence of reduction steps starting at  $M$  is finite.

THEOREM 48. *Every  $M \in \text{Term}$  is strongly normalizable with respect to  $\rightarrow_r$ .*

**Proof.** See Appendix C. ■

Let  $norm(M)$  refer to the iterated contraction of the leftmost redex in  $M$ . Since by the previous theorem, every reduction starting at  $M$  is finite,  $norm$  is a terminating normalization algorithm with respect to  $\rightarrow_r$ .

We shall now encode proofs by giving recipes for building up constructions of sequents. Every formula occurring in an antecedent part of a sequent  $s$  is said to be an antecedent formula component of  $s$ .

DEFINITION 49. A construction of a sequent  $s$  is a term  $M^A$  such that an occurrence of  $A$  is the succedent part of  $s$ , and every type of a free variable of  $M$  is an antecedent formula component of  $s$ .

This notion of construction is a straightforward adaptation of the notion of construction for ordinary natural deduction and sequent calculi. The *set* of types of the free variables occurring in the term encoding a derivation  $\Pi$  is a subset of the set of assumptions on which  $\Pi$  depends. Therefore applications

of structural inference rules are not reflected by term modifications, and variations of structural rules are captured by imposing conditions on variable binding and occurrences of free variables in the encoding terms (see, for instance, [van Benthem, 1986, Chapter 7], [van Benthem, 1991], [Helman, 1977], [Wansing, 1992]).

**OBSERVATION 50.** Given a proof in *DIntKt* of a sequent  $s$ , one can find a construction  $M$  of  $s$ .

**Proof.** We define a function  $f$  from the set  $\Pi\mathbf{DIntKt}$  of proofs in  $\mathbf{DIntKt}$  to *Term* such that  $f(\Pi)$  is a construction of the conclusion sequent of  $\Pi$ . The pairs of sequent rules and terms or term construction rules in Table 5 amount to an inductive definition of  $f$ . The variables newly introduced into the conclusion of a term construction rule are the numerically first variables of the types indicated not occurring in the premise term. ■

Clearly, *norm* is a function on *Term*. Let  $\Pi^+\mathbf{DIntKt}$  denote the set of all proofs in  $\mathbf{DIntKt}$  containing an application of (cut) or a redundant part, and let  $\Pi^-\mathbf{DIntKt}$  denote the set of all cut-free proofs in  $\mathbf{DIntKt}$  containing no redundant part. Let  $+Term$  denote the set of all terms that are not normal forms, and let  $-Term$  denote the set of all terms that are normal forms.

**THEOREM 51.** *Let  $\mathcal{A} = \langle \Pi\mathbf{DIntKt}, elim \rangle$  and  $\mathcal{B} = \langle Term, norm \rangle$ . The function  $f$  defined in the proof of Observation 50 is a homomorphism from  $\mathcal{A}$  to  $\mathcal{B}$ .*

**Proof.** See Appendix D. ■

Under the encoding of proofs by terms, surjective pairing ( $\langle (M)_0, (M)_1 \rangle \rightarrow_r M$ ) and  $\eta$ -reduction ( $\lambda x(Mx) \rightarrow_r M$ , if  $x \notin fv(M)$ ) correspond with replacing proofs

$$\frac{A \rightarrow A \quad B \rightarrow B}{A \times B \rightarrow A \wedge B} \quad \text{and} \quad \frac{A \rightarrow A \quad B \rightarrow B}{A \triangleright B \rightarrow A \times B}$$

$$A \wedge B \rightarrow A \wedge B \quad \quad \quad A \triangleright B \rightarrow A \triangleright B$$

by the axiomatic sequents  $A \wedge B \rightarrow A \wedge B$  and  $A \triangleright B \rightarrow A \triangleright B$ , respectively. Note that there are no analogues of surjective pairing and  $\eta$ -reduction that correspond with a replacement of proofs of  $[F]A \rightarrow [F]A$  and  $\langle P \rangle A \rightarrow \langle P \rangle A$  from  $A \rightarrow A$  by the axiomatic sequents  $[F]A \rightarrow [F]A$  and  $\langle P \rangle A \rightarrow \langle P \rangle A$ . Moreover, since in the encoding applications of structural rules are not reflected by term formation steps, it is in general *not* the case that if  $M = f(\Pi)$ ,  $\Pi$  can be uniquely reconstructed from  $M$ .



<i>Logical rules</i>	
$A \rightarrow A$	$v_1^A$
$\frac{X \rightarrow A \quad A \rightarrow Y}{X \rightarrow Y}$	$\frac{M^A \quad N(x^A)}{N[x := M]}$
<i>Structural rules</i>	
$\frac{s}{s'}$	$\frac{M}{\overline{M}}$
<i>Intuitionistic connective rules</i>	
$\mathbf{I} \rightarrow t$	$v_1^t$
$\frac{\mathbf{I} \rightarrow X}{t \rightarrow \overline{X}}$	$\frac{M}{\overline{M}}$
$\frac{X \rightarrow A \quad Y \rightarrow B}{X \times Y \rightarrow A \wedge B}$	$\frac{M^A \quad N^B}{\langle M, N \rangle}$
$\frac{A \times B \rightarrow X}{A \wedge B \rightarrow \overline{X}}$	$\frac{M(x^A, y^B)}{\overline{M((z^{A \wedge B})_0, (z^{A \wedge B})_1)}}$
$\frac{X \rightarrow A \times B}{\overline{X} \rightarrow A \triangleright B}$	$\frac{M(x^A)}{\lambda x^A M}$
$\frac{X \rightarrow A \quad B \rightarrow Y}{A \triangleright B \rightarrow X \times Y}$	$\frac{M^A \quad N(x^B)}{N[x := (y^{(A \triangleright B)}, M)]}$
<i>Modal connective rules</i>	
$\frac{\bullet X \rightarrow A}{X \rightarrow [F]A}$	$\frac{M}{\overline{PM}}$
$\frac{A \rightarrow X}{[F]A \rightarrow \bullet X}$	$\frac{M(x^A)}{\overline{M(\cup y^{[F]A})}}$
$\frac{X \rightarrow A}{\bullet X \rightarrow \langle P \rangle A}$	$\frac{M}{\overline{SM}}$
$\frac{A \rightarrow \bullet X}{\langle P \rangle A \rightarrow \overline{X}}$	$\frac{M(x^A)}{\overline{M(\cap y^{\langle P \rangle A})}}$

Table 5. Sequent rules and term construction rules.

### 3.7 A denotational semantics of proofs

We shall now define models for  $\lambda_t$ . The completeness proof to be given straightforwardly extends H. Friedman's [1975] completeness proof for typed  $\lambda$ -calculus. The plan of the proof is as follows: first it is shown that  $\lambda_t$  is sound and complete with respect to the class of all models. This is achieved by defining a canonical model that itself characterizes  $\lambda_t$ . Then a notion of intended model is defined. In such models the typed terms have their intended set-theoretic interpretation. In order to characterize provable equality of terms in  $\lambda_t$  by validity in all intended models, it is shown that for every intended model  $\mathcal{M}$ , there exists a 'partial homomorphism' from  $\mathcal{M}$  onto the canonical model. Since such partial homomorphisms turn out to preserve validity,  $\lambda_t$  is sound and complete with respect to the class of all intended models.

**DEFINITION 52.** A structure  $\mathcal{F} = \langle \{D^A\}, \{\text{AP}_{A,B}\}, \{\text{PRO}_{A,B}^0\}, \{\text{PRO}_{A,B}^1\}, \{\text{PAIR}_{A,B}\}, \{\text{P}_A\}, \{\text{S}_A\}, \{\text{P}\downarrow_A\}, \{\text{S}\downarrow_A\} \rangle$  is called a type structure frame (or just a frame) iff for all types  $A, B$ :

1.  $D^A$  (the domain of type  $A$ ) is a non-empty set;
2.  $\text{AP}_{A,B} : D^{(A \triangleright B)} \times D^A \longrightarrow D^B$ ,  
 $\text{PRO}_{A,B}^0 : D^{(A \wedge B)} \longrightarrow D^A$ ,  
 $\text{PRO}_{A,B}^1 : D^{(A \wedge B)} \longrightarrow D^B$ ,  
 $\text{PAIR}_{A,B} : D^A \times D^B \longrightarrow D^{(A \wedge B)}$ ,  
 $\text{P}_A : D^A \longrightarrow D^{[F]A}$ ,  
 $\text{S}_A : D^A \longrightarrow D^{(P)A}$ ,  
 $\text{P}\downarrow_A : D^{[F]A} \longrightarrow D^A$ ,  
 $\text{S}\downarrow_A : D^{(P)A} \longrightarrow D^A$ ;
3. (*extensionality*) if  $a, b \in D^{(A \triangleright B)}$  and  $(\forall c \in D^A)$  we have  $(\text{AP}_{A,B}(a, c) = \text{AP}_{A,B}(b, c))$ , then  $a = b$ ;
4. (*pro*) for all  $a \in D^A, b \in D^B$ :  
 $\text{PRO}_{A,B}^0(\text{PAIR}_{A,B}(a, b)) = a$ ,  $\text{PRO}_{A,B}^1(\text{PAIR}_{A,B}(a, b)) = b$ ;
5. (*pair*) for all  $a \in D^{A \wedge B}$ :  $\text{PAIR}_{A,B}(\text{PRO}_{A,B}^0(a), \text{PRO}_{A,B}^1(a)) = a$ ;
6. (*future*) for all  $a \in D^A$ :  $\text{P}\downarrow(\text{P}a) = a$ ;
7. (*past*) for all  $a \in D^A$ :  $\text{S}\downarrow(\text{S}a) = a$ .

An assignment in a frame  $\langle \{D^A\}, \{\text{AP}_{A,B}\}, \{\text{PRO}_{A,B}^0\}, \{\text{PRO}_{A,B}^1\}, \{\text{PAIR}_{A,B}\}, \{\text{P}_A\}, \{\text{S}_A\}, \{\text{P}\downarrow_A\}, \{\text{S}\downarrow_A\} \rangle$  is a function  $f$  defined on the set  $V$  of term variables such that  $f(x^A) \in D^A$ . The set of all assignments in a given frame is denoted by  $\text{Asg}$ . If  $y \in V$ , then  $f_a^y$  is defined by  $f_a^y(x) = f(x)$ , if  $x \neq y$ ,  $f_a^y(y) = a$ .

DEFINITION 53. Suppose that  $\mathcal{F} = \langle \{D^A\}, \{\text{AP}_{A,B}\}, \{\text{PRO}_{A,B}^0\}, \{\text{PRO}_{A,B}^1\}, \{\text{PAIR}_{A,B}\}, \{\text{P}_A\}, \{\text{S}_A\}, \{\text{P}\downarrow_A\}, \{\text{S}\downarrow_A\} \rangle$  is a frame. Then  $\langle \mathcal{F}, \text{val} \rangle$  is said to be a type structure model (or just a model) based on  $\mathcal{F}$  iff  $\text{val}$  is the valuation function from  $\text{Term} \times \text{Asg}$  to  $\bigcup_{A \in T} D^A$  such that:

1.  $\text{val}(x, f) = f(x)$ ;
2.  $\text{AP}_{A,B}(\text{val}((\lambda x M), f), a) = \text{val}(M, f_a^x), \forall a \in D^A$ ;
3.  $\text{val}((M^{A \triangleright B}, N^B), f) = \text{AP}_{A,B}(\text{val}(M, f), \text{val}(N, f))$ ;
4.  $\text{val}(\langle M^A, N^B \rangle, f) = \text{PAIR}_{A,B}(\text{val}(M, f), \text{val}(N, f))$ ;
5.  $\text{val}((M^{A \wedge B})_i, f) = \text{PRO}_{A,B}^i(\text{val}(M, f)), i = 0, 1$ ;
6.  $\text{val}((\mathcal{P}M^A)^{[F]A}, f) = \text{P}_A(\text{val}(M, f))$ ;
7.  $\text{val}((\mathcal{S}M^A)^{\langle P \rangle A}, f) = \text{S}_A(\text{val}(M, f))$ ;
8.  $\text{val}((\bigcup M^{[F]A})^A, f) = \text{P}\downarrow_A(\text{val}(M, f))$ ;
9.  $\text{val}((\bigcap M^{\langle P \rangle A})^A, f) = \text{S}\downarrow_A(\text{val}(M, f))$ .

Let  $\mathcal{M} = \langle \mathcal{F}, \text{val} \rangle$  be a model.

LEMMA 54. (1)  $\text{val}(M[x := N], f) = \text{val}(M, f_{\text{val}(N, f)}^x)$ , if  $\text{bv}(M) \cap \text{fv}(N) = \emptyset$ . (2)  $\text{val}(M[x := y], f_a^y) = \text{val}(M, f_a^x)$ , if  $y \notin \text{bv}(M) \cup \text{fv}(M)$ .

**Proof.** (1) By induction on  $M$ , for fixed  $N$ ; (2) by (1). ■

The equality  $M = N$  is said to hold in  $\mathcal{M}$  under assignment  $f$  ( $\mathcal{M}, f \models M = N$ ) iff  $\text{val}(M, f) = \text{val}(N, f)$ .  $M = N$  is called valid in  $\mathcal{M}$  ( $\mathcal{M} \models M = N$ ) iff  $\mathcal{M}, f \models M = N$ , for all  $f \in \text{Asg}$ .  $M = N$  is said to be valid in a class  $\mathcal{K}$  of models, if  $\mathcal{M} \models M = N$ , for each  $\mathcal{M} \in \mathcal{K}$ .

OBSERVATION 55. (Soundness) If  $M = N$  is provable in  $\lambda_t$ , then  $M = N$  is valid in the class of all models.

**Proof.** By induction on proofs in  $\lambda_t$ . We must show that every axiom is valid in every model, and that the rules of inference preserve validity. We shall consider two cases not already dealt with in [Friedman, 1975].

$$\begin{aligned}
& \langle (M)_0, (M)_1 \rangle = M: \\
& \text{val}(\langle (M, N) \rangle_0, \langle (M, N) \rangle_1, f) \\
= & \text{PAIR}(\text{val}(\langle (M, N) \rangle_0, f), \text{val}(\langle (M, N) \rangle_1, f)) \\
= & \text{PAIR}(\text{PRO}^0(\text{PAIR}(\text{val}(M, f), \text{val}(N, f))), \\
& \text{PRO}^1(\text{PAIR}(\text{val}(M, f), \text{val}(N, f)))) \\
= & \text{PAIR}(\text{val}(M, f), \text{val}(N, f)) = \text{val}(\langle M, N \rangle, f). \\
\\
& \cap SM = M: \\
& \text{val}(\cap SM, f) \\
= & \text{S}\downarrow(\text{val}(SM, f)) \\
= & \text{S}\downarrow(\text{S}(\text{val}(M, f))) \\
= & \text{val}(M, f) \quad \blacksquare
\end{aligned}$$

Next, we define the frame  $\mathcal{F}_0$  on which the canonical model is based. Let  $|M| = \{N \mid \vdash_{\lambda_t} M = N\}$ ;  $|M|$  is the equivalence class of  $M$  with respect to provable equality in  $\lambda_t$ .

DEFINITION 56.  $\mathcal{F}_0 = \langle \{D^A\}, \{\text{AP}_{A,B}\}, \{\text{PRO}_{A,B}^0\}, \{\text{PRO}_{A,B}^1\}, \{\text{PAIR}_{A,B}\}, \{\text{P}_A\}, \{\text{S}_A\}, \{\text{P}\downarrow_A\}, \{\text{S}\downarrow_A\} \rangle$  is defined as follows:

- $D^A = \{|M| \mid M \text{ is of type } A\}$ ;
- $\text{AP}_{A,B}(|M^{A \triangleright B}|, |N^A|) = |(M, N)|$ ;
- $\text{PRO}_{A,B}^0(|M^{A \wedge B}|) = |(M)_0|$ ;
- $\text{PRO}_{A,B}^1(|M^{A \wedge B}|) = |(M)_1|$ ;
- $\text{PAIR}_{A,B}(|M^A|, |N^B|) = |\langle M, N \rangle|$ ;
- $\text{P}_A(|M^A|) = |\mathcal{P}M|$ ;
- $\text{S}_A(|M^A|) = |SM|$ ;
- $\text{P}\downarrow_A(|M^A|) = |\cup M|$ ;
- $\text{S}\downarrow_A(|M^A|) = |\cap M|$ .

LEMMA 57.  $\mathcal{F}_0$  is a frame.

**Proof.** Clearly,  $D^A$  is a non-empty set, and  $\text{AP}_{A,B}$ ,  $\text{PRO}_{A,B}^0$ ,  $\text{PRO}_{A,B}^1$ ,  $\text{PAIR}_{A,B}$ ,  $\text{P}_A$ ,  $\text{S}_A$ ,  $\text{P}\downarrow_A$ , and  $\text{S}\downarrow_A$  are functions with appropriate domain and range, for all types  $A$  and  $B$ . For (extensionality) see [Friedman, 1975]. For (pro), (pair), (future), and (past), use the obvious equalities.  $\blacksquare$

A function  $g : V \rightarrow \text{Term}$  is called a substitution, if  $g(x)$  and  $x$  are of the same type. A substitution is called regular, if for pairwise distinct variables  $x, y$ ,  $fv(g(x)) \cap fv(g(y)) = \emptyset$ . Let  $M(g)$  denote the result of simultaneously

replacing in  $M$  every free occurrence of each variable  $x$  by  $g(x)$ . It can easily be shown that if  $M \in \text{Term}$  and  $\Gamma$  is a finite set of variables, then there is an  $N$  such that  $\vdash_{\lambda_t} M = N$ ,  $fv(M) = fv(N)$ , and  $bv(N) \cap \Gamma = \emptyset$ .

**DEFINITION 58.** Suppose  $f$  is an assignment in  $\mathcal{F}_0$  and  $g$  is a regular substitution such that  $f(x) = |g(x)|$ , for every  $x \in V$ . For a given term  $M$ , choose a term  $N$  such that  $\vdash_{\lambda_t} M = N$  and for every  $x \in fv(N)$ ,  $bv(N) \cap fv(g(x)) = \emptyset$ . Then  $val(M, f)$  is defined by  $val(M, f) = |N(g)|$ .

It can be shown that  $val : \text{Term} \times \text{Asg} \rightarrow \bigcup_A D^A$ , and  $\vdash_{\lambda_t} M = N$  implies  $val(M, f) = val(N, f)$ , cf. [Friedman, 1975].

**LEMMA 59.**  $\mathcal{M}_0 = \langle \mathcal{F}_0, val \rangle$  is a type structure model.

**Proof.** We consider those conditions not already assumed in Friedman's paper. Let  $g$  be a regular substitution and  $f(x) = |g(x)|$ , for  $f \in \text{Asg}$ . Choose  $M_1, N_1$  such that  $\vdash_{\lambda_t} M = M_1$ ,  $\vdash_{\lambda_t} N = N_1$ , and  $bv(M_1) \cap fv(g(x)) = bv(N_1) \cap fv(g(x)) = \emptyset$ , for every  $x \in fv(M_1) \cup fv(N_1)$ .

- 4 :  $val(\langle M, N \rangle, f) = | \langle M_1, N_1 \rangle(g) | = \text{PAIR}(|M_1(g)|, |N_1(g)|) = \text{PAIR}(val(M, f), val(N, f))$ .
- 5 :  $val((M)_i, f) = |(M_1)_i(g)| = \text{PRO}^i(|M_1(g)|) = \text{PRO}^i(val(M, f))$ .
- 6 :  $val((\mathcal{P}M^A)^{[F]A}, f) = |\mathcal{P}M_1(g)| = \mathbf{P}_A(|M_1(g)|) = \mathbf{P}_A(val(M, f))$ .
- 8 :  $val((\cup M^{[F]A})^A, f) = |\cup M_1(g)| = \mathbf{P}\downarrow_A(|M_1(g)|) = \mathbf{P}\downarrow_A(val(M, f))$ .
- 7 and 9 : analogous to the previous two cases. ■

**THEOREM 60.** (Completeness) *If  $M = N$  is valid in the class of all models, then  $\vdash_{\lambda_t} M = N$ .*

**Proof.** Suppose  $\not\vdash_{\lambda_t} M = N$ . Choose  $M_1, N_1$  such that  $\vdash_{\lambda_t} M = N_1$ ,  $\vdash_{\lambda_e} N = N_1$ , and  $bv(M_1) \cap fv(M_1) = bv(N_1) \cap fv(N_1) = \emptyset$ . Then  $val(M, f) = |M_1| \neq |N_1| = val(N, f)$ , for  $f(x) = |\text{id}(x)|$ , for all  $x \in V$ , where  $\text{id}$  is the identity function on  $V$ . Thus,  $\mathcal{M}_0 \not\models M = N$ . ■

We now define the intended models. Following the terminology of Friedman, we shall call the frames underlying an intended model 'full temporal type structures over infinite sets'.

**DEFINITION 61.** A type structure frame  $\mathcal{F} = \langle \{\mathbf{D}^A\}, \{\mathbf{AP}_{A,B}\}, \{\mathbf{PRO}_{A,B}^0\}, \{\mathbf{PRO}_{A,B}^1\}, \{\mathbf{PAIR}_{A,B}\}, \{\mathbf{P}_A\}, \{\mathbf{S}_A\}, \{\mathbf{P}\downarrow_A\}, \{\mathbf{S}\downarrow_A\} \rangle$  is said to be a full temporal type structure over infinite sets, if

- $\mathbf{D}^t$  is infinite, and for every  $p \in \text{Atom}$ ,  $\mathbf{D}^p$  is infinite;
- $\mathbf{D}^{A \wedge B} = \mathbf{D}^A \times \mathbf{D}^B$ ;
- $\mathbf{D}^{A \triangleright B} = (\mathbf{D}^B)^{\mathbf{D}^A}$ ;

- $\mathbf{D}^{[F]A} = \{\mathcal{P}a \mid a \in \mathbf{D}^A\}$ ;  $\mathbf{D}^{(P)A} = \{\mathcal{S}a \mid a \in \mathbf{D}^A\}$ ;
- $\mathbf{AP}_{A,B}(a, b) = a(b)$ ;
- $\mathbf{PRO}_{A,B}^0(\langle a, b \rangle) = a$ ;  $\mathbf{PRO}_{A,B}^1(\langle a, b \rangle) = b$ ;
- $\mathbf{PAIR}_{A,B}(a, b) = \langle a, b \rangle$ ;
- $\mathbf{P}_A(a) = \mathcal{P}a$ ;  $\mathbf{S}_A(a) = \mathcal{S}a$ ;
- $\mathbf{P}\downarrow_A(a) = \cup a$ ;  $\mathbf{S}\downarrow_A(a) = \cap a$ .

DEFINITION 62. Let  $\mathcal{F} = \langle \{D^A\}, \{\mathbf{AP}_{A,B}\}, \{\mathbf{PRO}_{A,B}^0\}, \{\mathbf{PRO}_{A,B}^1\}, \{\mathbf{PAIR}_{A,B}\}, \{\mathbf{P}_A\}, \{\mathbf{S}_A\}, \{\mathbf{P}\downarrow_A\}, \{\mathbf{S}\downarrow_A\} \rangle$ ,  $\mathcal{F}^* = \langle \{D^{*A}\}, \{\mathbf{AP}_{A,B}^*\}, \{\mathbf{PRO}_{A,B}^{*0}\}, \{\mathbf{PRO}_{A,B}^{*1}\}, \{\mathbf{PAIR}_{A,B}^*\}, \{\mathbf{P}_A^*\}, \{\mathbf{S}_A^*\}, \{\mathbf{P}\downarrow_A^*\}, \{\mathbf{S}\downarrow_A^*\} \rangle$  be frames, and let  $\mathcal{M} = \langle \mathcal{F}, \text{val} \rangle$  and  $\mathcal{M}^* = \langle \mathcal{F}^*, \text{val}^* \rangle$  be models. A family of functions  $\{f_A\}$  is called a partial homomorphism from  $\mathcal{M}$  onto  $\mathcal{M}^*$  iff

1. for each type  $A$ ,  $f_A$  is a partial function from  $D^A$  onto  $D^{*A}$ ;
2. if  $f_{A \triangleright B}(a)$  exists, then  $f_B(\mathbf{AP}_{A,B}(a, b)) = \mathbf{AP}_{A,B}^*(f_{A \triangleright B}(a), f_A(b))$ , for all  $b$  in the domain of  $f_A$ ,
3. if  $f_A(a)$ ,  $f_B(b)$  exist, then  $f_{A \wedge B}(\mathbf{PAIR}_{A,B}(a, b)) = \mathbf{PAIR}_{A,B}^*(f_A(a), f_B(b))$ ;
4. if  $f_{A \wedge B}(a)$  exists, then  $f_A(\mathbf{PRO}_{A,B}^0(a)) = \mathbf{PRO}_{A,B}^{*0}(f_{A \wedge B}(a))$ ;
5. if  $f_{A \wedge B}(a)$  exists, then  $f_B(\mathbf{PRO}_{A,B}^1(a)) = \mathbf{PRO}_{A,B}^{*1}(f_{A \wedge B}(a))$ ;
6. if  $f_A(a)$  exists, then  $f_{[F]A}(\mathbf{P}_A(a)) = \mathbf{P}_A^*(f_A(a))$ ;  $f_{(P)A}(\mathbf{S}_A(a)) = \mathbf{S}_A^*(f_A(a))$ ;
7. if  $f_{[F]A}(a)$ ,  $f_{(P)A}(b)$  exist, then  $f_A(\mathbf{P}\downarrow_A(a)) = \mathbf{P}\downarrow_A^*(f_{[F]A}(a))$ ;  $f_A(\mathbf{S}\downarrow_A(b)) = \mathbf{S}\downarrow_A^*(f_{(P)A}(b))$ .

LEMMA 63. Let  $\mathcal{M}$ ,  $\mathcal{M}^*$  be as in the previous definition, and let  $\{f_A\}$  be a partial homomorphism from  $\mathcal{M}$  onto  $\mathcal{M}^*$ . If  $g$ ,  $g^*$  are assignments in  $\mathcal{F}$  and  $\mathcal{F}^*$  respectively, and  $f_A(g(x^A)) = g^*(x)$ , then  $f_A(\text{val}(M^A, g)) = \text{val}^*(M, g^*)$ .

**Proof.** By induction on  $M$ . We consider the cases not already dealt with in [Friedman, 1975]. Note that we may assume  $f_A(g(x^A)) = g^*(x)$ , since  $f_A$  is onto.

- $M \equiv \langle N^A, G^B \rangle$ :  $f_{A \wedge B}(\text{val}(\langle N, G \rangle, g))$   
 $= f_{A \wedge B}(\mathbf{PAIR}(\text{val}(N, g), \text{val}(G, g)))$   
 $= \mathbf{PAIR}_{A,B}^*(f_A(\text{val}(N, g)), f_B(\text{val}(G, g)))$   
 $= \mathbf{PAIR}_{A,B}^*(\text{val}^*(N, g^*), \text{val}^*(G, g^*))$  by the induction hypothesis  
 $= \text{val}^*(\langle N, G \rangle, g^*)$ .

- $M \equiv (N^{A \wedge B})_i$ :  $f(\text{val}((N)_i, g)) = f(\text{PRO}^i(\text{val}(N, g)))$   
 $= \text{PRO}^{*i}(f_{A \wedge B}(\text{val}(N, g)))$   
 $= \text{PRO}^{*i}(\text{val}^*(N, g^*))$  by the induction hypothesis  
 $= \text{val}^*((N)_i, g^*)$ .
- $M \equiv \mathcal{P}N^A$ :  $f_{[F]A}(\text{val}(\mathcal{P}N, g)) = f_{[F]A}(\mathbf{P}_A(\text{val}(N, g)))$   
 $= \mathbf{P}_A^*(f_A(\text{val}(N, g))) = \mathbf{P}_A^*(\text{val}^*(N, g^*)) = \text{val}^*(\mathcal{P}N, g^*)$ .
- $M \equiv \cup N^{[F]A}$ :  $f_A(\text{val}(\cup N, g)) = f_A(\mathbf{P}_{\downarrow A}(\text{val}(N, g)))$   
 $= \mathbf{P}_{\downarrow A}^*(f_{[F]A}(\text{val}(N, g))) = \mathbf{P}_{\downarrow A}^*(\text{val}^*(N, g^*)) = \text{val}^*(\cup N, g^*)$ .
- $M \equiv \mathcal{S}N, \cap N$ : analogous to the previous two cases. ■

**COROLLARY 64.** *Let  $\mathcal{M} = \langle \mathcal{F}, \text{val} \rangle$ ,  $\mathcal{M}^* = \langle \mathcal{F}^*, \text{val}^* \rangle$  be models. If there is a partial homomorphism from  $\mathcal{M}$  onto  $\mathcal{M}^*$ , then  $\mathcal{M} \models M = N$  implies  $\mathcal{M}^* \models M = N$ .*

**Proof.** Suppose  $\mathcal{M} \models M^B = N^B$ ,  $\{f_A\}$  is a partial homomorphism from  $\mathcal{M}$  onto  $\mathcal{M}^*$ , and  $g^*$  is an assignment in  $\mathcal{M}^*$ . We choose an assignment  $g$  in  $\mathcal{M}$  such that for every  $A \in T$ ,  $g^*(x) = f_A(g(x^A))$ . By the previous lemma,  $\text{val}^*(M, g^*) = f_B(\text{val}(M, g)) = f_B(\text{val}(N, g)) = \text{val}^*(N, g^*)$  ■

**THEOREM 65.** *Let  $\mathcal{M}$  be a model based on a full temporal type structure over infinite sets. Then  $\vdash_{\lambda_t} M = N$  iff  $\mathcal{M} \models M = N$ .*

**Proof.** It suffices to show that  $\mathcal{M} \models M = N$  implies  $\mathcal{M}_0 \models M = N$ . To prove this, we define by induction on  $A$  a partial homomorphism  $\{f_A\}$  from  $\mathcal{M}$  onto  $\mathcal{M}_0$  as follows:

- $A = p, A = t, p \in \text{Atom}$ :  
 $f_A$  is any function from  $\mathbf{D}^A$  onto  $\mathcal{M}_0$ 's domain  $D^A$ .  
 (Such a function exists, since  $\mathbf{D}^A$  is infinite and  $D^A$  is denumerable.)
- $A = (B \wedge C)$ :  
 If  $f_B(b), f_C(c)$  exist, then  $f_{B \wedge C}(\langle b, c \rangle) = f_{B \wedge C}(\mathbf{PAIR}(b, c))$  is defined as  $\mathbf{PAIR}_{B,C}(f_B(b), f_C(c))$ .
- $A = (B \triangleright C)$ :  
 $f_{B \triangleright C}(a)$  is defined as the unique member of  $D^{(B \triangleright C)}$  (if it exists) such that  $f_C(a(b)) = \mathbf{AP}_{B,C}(f_{B \triangleright C}(a), f_B(b))$ , for all  $b$  in the domain of  $f_B$ .
- $A = [F]A$ :  
 $f_{[F]A}(a) = f_{[F]A}(\mathbf{P}_A(b))$  for some  $b \in \mathbf{D}^A$  is defined as  $\mathbf{P}_A(f_A(b))$  if  $f_A(b)$  exists.
- $A = \langle P \rangle A$ :  
 $f_{\langle P \rangle A}(a) = f_{\langle P \rangle A}(\mathbf{S}_A(b))$  for some  $b \in \mathbf{D}^A$  is defined as  $\mathbf{S}_A(f_A(b))$  if  $f_A(b)$  exists.

That  $\{f_A\}$  is a partial homomorphism follows from the definition of  $\{f_A\}$  and the following equations:

$$\begin{array}{ll}
f_A(\mathbf{PRO}_{A,B}^0(\langle a, b \rangle)) & = f_B(\mathbf{PRO}_{A,B}^1(\langle a, b \rangle)) \\
= f_A(a) & = f_B(b) \\
= \mathbf{PRO}_{A,B}^0(\mathbf{PAIR}_{A,B}(f_A(a), f_B(b))) & = \mathbf{PRO}_{A,B}^1(\mathbf{PAIR}_{A,B}(f_A(a), f_B(b))) \\
= \mathbf{PRO}_{A,B}^0(f_{A \wedge B}(\mathbf{PAIR}_{A,B}(a, b))) & = \mathbf{PRO}_{A,B}^1(f_{A \wedge B}(\mathbf{PAIR}_{A,B}(a, b))) \\
= \mathbf{PRO}_{A,B}^0(f_{A \wedge B}(\langle a, b \rangle)) & = \mathbf{PRO}_{A,B}^1(f_{A \wedge B}(\langle a, b \rangle)) \\
\\
f_A(\mathbf{P} \downarrow_A (\mathcal{P}a)) & = f_A(\mathbf{S} \downarrow_A (\mathcal{S}a)) \\
= f_A(a) & = f_A(a) \\
= \mathbf{P} \downarrow_A (f_{[F]A}(\mathcal{P}a)) & = \mathbf{S} \downarrow_A (f_{\langle P \rangle A}(\mathcal{S}a))
\end{array}$$

It remains to be shown that  $f_A$  is onto, for every type  $A$ . For  $A = t$  and  $A = p \in \mathit{Atom}$ , this follows from the definitions of  $f_t$ ,  $f_p$  and  $\mathcal{F}_0$ . For the remaining cases we consider two examples.  $A = [F]B$ . Assume  $d = |\mathcal{P}M| \in D^{[F]B}$ . Choose  $a \in \mathbf{D}^{[F]B}$  such that  $a = \mathcal{P}b$  for  $b \in \mathbf{D}^B$  and  $b = f_B^{-1}(|M^B|)$ . Since  $f_B$  is onto, such an element  $a$  from  $\mathbf{D}^{[F]B}$  exists. Then  $f_{[F]B}(a) = f_{[F]B}(\mathbf{P}_B(b)) = \mathbf{P}_B(f_B(b)) = |\mathcal{P}M| = d$ . Consider now  $A = (B \triangleright C)$ , and assume  $d \in D^{(B \triangleright C)}$ . Choose  $a \in \mathbf{D}^{(B \triangleright C)}$  such that for every  $b$  in the domain of  $f_B$ ,  $a(b) \in f_C^{-1}(\mathbf{Ap}(d, f_B(b)))$ . Then  $f_{(B \triangleright C)}(a) = d$ . Since  $f_C$  and  $f_B$  may be assumed to be onto, the set of such  $a \in \mathbf{D}^{(B \triangleright C)}$  is non-empty.  $\blacksquare$

Whereas the encoding of substructural subsystems of  $\mathbf{DIntKt}$  obtained by giving up all or part of  $\mathbf{DIntKt}$ 's structural rules will require modifications of the notion of construction, in order to encode structural extensions of  $\mathbf{DIntKt}$ , the notion of construction need not be altered. Various extensions of  $\mathbf{HIntKt}$  can be presented as structural extensions of  $\mathbf{DIntKt}$ . The following axiom schemata are those schematic axioms from Table 3, which are in  $\mathcal{L}$ . Each axiom schema  $\mathbf{Ax}$  in this table corresponds with the associated structural rule  $\mathbf{Ax}'$  in the sense that an  $\mathcal{L}$ -formula  $A$  is provable in  $\mathbf{HIntKt} + \mathbf{Ax}$  iff  $\mathbf{I} \rightarrow A$  is provable in  $\mathbf{DIntKt} + \mathbf{Ax}'$ .

In the literature, several proposals have been made to extend the formulas-as-types notion of construction from positive logic to modal logics based on it. We shall here briefly point to five such approaches.

**1.** Gabbay and de Queiroz [1992] interpret the necessity modality  $\Box$  “as a sort of second-order universal quantification (quantification over structured collections of formulas)” [Gabbay and de Queiroz, 1992, p. 1359]. Using the framework of Labelled Natural Deduction [de Queiroz and Gabbay, 1999], proofs in various modal logics are encoded by imposing conditions on abstraction over possible-world variables [de Queiroz and Gabbay, 1997]. However, Gabbay and de Queiroz do not consider a Friedman-style completeness proof for the  $\lambda$ -calculus under consideration.



<i>name</i>	<i>axiom schema</i>	<i>name</i>	<i>structural rule</i>
$T$	$[F]A \triangleright A$	$T'$	$X \rightarrow \bullet Y \vdash X \rightarrow Y$
$4$	$[F]A \triangleright [F][F]A$	$4'$	$X \rightarrow \bullet Y \vdash X \rightarrow \bullet \bullet Y$
$V$	$[F]A$	$V'$	$X \rightarrow Y \vdash \bullet \mathbf{I} \rightarrow Y$
$T^c$	$A \triangleright [F]A$	$T^{c'}$	$X \rightarrow Y \vdash X \rightarrow \bullet Y$
$4^c$	$[F][F]A \triangleright [F]A$	$4^{c'}$	$X \rightarrow \bullet \bullet Y \vdash X \rightarrow \bullet Y$
$D_p$	$t \triangleright \langle P \rangle t$	$D_p'$	$\bullet \mathbf{I} \rightarrow Y \vdash \mathbf{I} \rightarrow Y$
$T_p$	$A \triangleright \langle P \rangle A$	$T_p'$	$\bullet X \rightarrow Y \vdash X \rightarrow Y$
$4_p$	$\langle P \rangle \langle P \rangle A \triangleright \langle P \rangle A$	$4_p'$	$\bullet X \rightarrow Y \vdash \bullet \bullet X \rightarrow Y$
$B_p$	$(A \wedge \langle P \rangle B) \triangleright \langle P \rangle (B \wedge \langle P \rangle A)$	$B_p'$	$\bullet (X \times \bullet Y) \rightarrow Z \vdash Y \times \bullet X \rightarrow Z$
$Alt1_p$	$(\langle P \rangle A \wedge \langle P \rangle B) \triangleright \langle P \rangle (A \wedge B)$	$Alt1_p'$	$\bullet (X \times Y) \rightarrow Z \vdash \bullet X \times \bullet Y \rightarrow Z$
$T_p^c$	$\langle P \rangle A \triangleright A$	$T_p^{c'}$	$X \rightarrow Y \vdash \bullet X \rightarrow Y$
$4_p^c$	$\langle P \rangle A \triangleright \langle P \rangle \langle P \rangle A$	$4_p^{c'}$	$\bullet \bullet X \rightarrow Y \vdash \bullet X \rightarrow Y$

 Table 6. Axioms in  $\mathcal{L}$ .

2. Borghuis [1993; 1994; 1998] investigates the formulas-as-types-notion of construction for several normal modal propositional logics based on **CPL**. Fitch-style natural deduction proofs in these modal logics are interpreted in a second-order  $\lambda$ -calculus. In this approach, unary type-forming operators are introduced to encode applications of import and export rules for  $\square$  in Fitch-style natural deduction. The operations  $\hat{k}$  and  $\check{k}$  encoding the export and import rules for  $\square$  in the smallest normal modal logic **K**, for example, satisfy the following reduction rule:  $\hat{k}(\check{k}M) \rightarrow_r M$ . Borghuis proves strong normalization results for the modal typed  $\lambda$ -calculi under consideration. However, the term-forming operations used to encode applications of import and export rules for  $\square$  are not provided with a set-theoretic interpretation.

3. Martini and Masini [1996] consider formulas-as-types for 2-sequent calculi, cf. Section 2.2. They introduce two unary term-forming operations  $\text{gen}$  and  $\text{ungen}$  to encode applications of  $\square$ -introduction and  $\square$ -elimination rules. A strong normalization theorem is proved for the typed  $\lambda$ -calculus encoding proofs in the 2-sequent calculus for the modal logic **S4**. However, the typed terms do not receive a set-theoretic interpretation.

4. Recently, Sasaki [1999] suggested understanding a  $\lambda$ -term of type  $\square A$  as either denoting an element from the domain associated with  $A$ , or being undefined. A term  $M^{A \triangleright \square B}$  would then denote a partial function from  $D^A$  to  $D^B$ . Sasaki defines an extended typed  $\lambda$ -calculus with various formation rules for obtaining terms of type  $\square A$ . Moreover, natural deduction proofs in the extension of the intuitionistic modal logic **IntK** by the axiom schemata

$$T_c \quad A \triangleright \square A \quad \text{and} \quad 4_c \quad \square \square A \triangleright \square A$$

are encoded by terms in the extended typed  $\lambda$ -calculus. Unfortunately, no denotational semantics for this  $\lambda$ -calculus is developed.

5. The approach that comes closest to the one presented here is Restall's [1999, Chapter 7], who also applies Belnap's display calculus. Introductions of  $[F]$  on the right (left) of the sequent arrow are encoded using a unary operator *up* (*down*), lifting (lowering) terms of type  $A$  ( $[F]A$ ) to terms of type  $[F]A$  ( $A$ ), just like the operation  $\mathcal{P}$  ( $\cup$ ). Backward-looking possibility is treated quite differently. Introductions of  $\langle P \rangle$  (in Restall's notation  $\diamond$ ) on the right are encoded using a unary type-lifting operation  $\bullet$  (not to be confused with the structure connective  $\bullet$ ). Introductions on the left are encoded by a unary term-forming operation turning terms  $N^B$ ,  $M^{(P)A}$  into the term *let*  $M$  *be*  $\bullet x$  *in*  $N$  of type  $B$ . Whereas the term *down up*  $N$  reduces in one step to  $N$ , *let*  $\bullet G$  *be*  $\bullet x$  *in*  $N$  reduces in one step to  $N[x := G]$ . Restall proves normalization for the extended typed  $\lambda$ -calculus under consideration, however, no set-theoretic interpretation of *up*, *down*,  $\bullet$ , and *let*  $M$  *be*  $\bullet x$  *in* is suggested.

In the literature on functional programming there are various proposals for providing an *operational* semantics of proofs in modal logics, notably in intuitionistic **S4**. Natural deduction in the framework of Martin-Löf's type theory is considered in [Davis and Pfenning, 2000] and [Pfenning, 2000]. Also, further references can be found in these papers.

### 3.8 Bi-intuitionistic logic

Suppose a connective  $f_1$  is introduced in a finite-set-to-formula sequent calculus, whereas another connective  $f_2$  is introduced in a formula-to-finite-set sequent system. Then the right introduction rules for  $f_1$  and the left introduction rules for  $f_2$  satisfy the segregation condition. However, if we just combine the sets of rules of both sequent calculi, neither  $Af_1B$  nor  $Af_2B$  is introduced in the most general context, namely in an arbitrary finite set of formulas, because there are no structure operations like in display logic that allow keeping track of succedent (antecedent) formulas on the left (right) of  $\rightarrow$ . This leads to a problem encountered in formulating an ordinary sequent calculus for bi-intuitionistic logic **BiInt**, the combination of intuitionistic logic and dual-intuitionistic logic. It can be shown that in the ordinary finite-set-to-formula sequent calculus no binary operation  $\sharp$  is definable such that  $\sharp$  satisfies (in the finite-set-to-formula setting) the dual Deduction Theorem characteristic of coimplication:  $A \rightarrow B$  iff  $A\sharp B \rightarrow \emptyset$ , see [Goré, 2000]. Bi-intuitionistic logic extends the language of intuitionistic logic by *coimplication*, the residual of disjunction, and *conegation*. The syntax of **BiInt** is given by:

$$A ::= p \mid \neg A \mid \sim A \mid A \wedge B \mid A \vee B \mid A \triangleright B \mid A \triangleleft B.$$

In the presence of a falsity constant  $\mathbf{f}$ , intuitionistic negation  $\neg$  can be defined by  $\neg A := (A \triangleright \mathbf{f})$ , and in the presence of a truth constant  $\mathbf{t}$ , conegation  $\smile$  can be defined by  $\smile A := (\mathbf{t} \blacktriangleleft A)$ .

Bi-intuitionistic logic has a natural algebraic and possible-worlds semantics, see [Rauszer, 1980]. The possible-worlds semantics adds to Kripke models for intuitionistic logic evaluation clauses for conegation and coimplication. A frame is a pair  $\langle I, \sqsubseteq \rangle$ , where  $I$  is a non-empty set (of states), and  $\sqsubseteq$  is a reflexive and transitive binary relation on  $I$ . A structure  $\langle I, \sqsubseteq, v \rangle$  is a bi-intuitionistic model if  $v$  is a function assigning to every propositional variable  $p$  a subset  $v(p)$  of  $I$  and, moreover, for every  $t, u \in I$ , if  $t \sqsubseteq u$  and  $t \in v(p)$ , then  $u \in v(p)$ . Verification of a formula  $A$  in the model  $\mathcal{M} = \langle I, \sqsubseteq, v \rangle$  at state  $t$  (in symbols  $\mathcal{M}, t \models A$ ) is inductively defined as follows:

$$\begin{aligned} \mathcal{M}, t \models p & \quad \text{iff } t \in v(p), \text{ for every propositional variable } p; \\ \mathcal{M}, t \models \neg A & \quad \text{iff for all } u \in I, t \sqsubseteq u \text{ implies } \mathcal{M}, u \not\models A \\ \mathcal{M}, t \models \smile A & \quad \text{iff there exists } u \in I, u \sqsubseteq t, \text{ and } \mathcal{M}, u \not\models A \\ \mathcal{M}, t \models A \wedge B & \quad \text{iff } \mathcal{M}, t \models A \text{ and } \mathcal{M}, t \models B; \\ \mathcal{M}, t \models A \vee B & \quad \text{iff } \mathcal{M}, t \models A \text{ or } \mathcal{M}, t \models B; \\ \mathcal{M}, t \models A \triangleright B & \quad \text{iff for all } u \in I, \text{ if } t \sqsubseteq u \text{ then } \mathcal{M}, u \not\models A \text{ or } \mathcal{M}, u \models B; \\ \mathcal{M}, t \models A \blacktriangleleft B & \quad \text{iff there is a } u \in I, u \sqsubseteq t \text{ } \mathcal{M}, u \models A \text{ and } \mathcal{M}, u \not\models B; \end{aligned}$$

where  $\mathcal{M}, t \not\models A$  is the (classical) negation of  $\mathcal{M}, t \models A$ . A formula  $A$  is valid in  $\mathcal{M} = \langle I, \sqsubseteq, v \rangle$  if for every  $t \in I$ ,  $\mathcal{M}, t \models A$ ; and  $A$  is valid on a frame  $\mathcal{F} = \langle I, \sqsubseteq \rangle$  if  $A$  is valid in every model  $\langle \mathcal{F}, v \rangle$  based on  $\mathcal{F}$ . A formula  $A$  is said to be valid in a class  $\mathcal{K}$  of models (frames) if  $A$  is valid in every model (frame) from  $\mathcal{K}$ .

The axiomatic system **HBiInt** consists of axiom schemata for intuitionistic logic **Int**, modus ponens, the rule

$$\text{from } A \text{ infer } \neg \smile A$$

and the following axiom schemata:

1.  $A \triangleright (B \vee (A \blacktriangleleft B))$
2.  $(A \blacktriangleleft B) \triangleright \smile(A \triangleright B)$
3.  $((A \blacktriangleleft B) \blacktriangleleft C) \triangleright (A \blacktriangleleft (B \vee C))$
4.  $\neg(A \blacktriangleleft B) \triangleright (A \triangleright B)$
5.  $(A \triangleright (B \blacktriangleleft B)) \triangleright \neg A$
6.  $\neg A \triangleright (A \triangleright (B \blacktriangleleft B))$
7.  $((B \triangleright B) \blacktriangleleft A) \triangleright \smile A$
8.  $\smile A \triangleright ((B \triangleright B) \blacktriangleleft A)$

**THEOREM 66.** *A formula  $A$  in the language of  $\mathbf{BiInt}$  is valid in the class of all models iff  $A$  is provable in  $\mathbf{HBiInt}$ .*

In the present section, we shall apply the modal display calculus and use a modal translation of  $\mathbf{BiInt}$  into  $\mathbf{S4t}$  to give a display sequent calculus for  $\mathbf{BiInt}$  based on the structure connectives  $\mathbf{I}$ ,  $*$ ,  $\circ$ , and  $\bullet$ , cf. [Goré, 1995], [Wansing, 1998, Chapter 10]. A direct display sequent system for  $\mathbf{BiInt}$  not relying on a modal translation has been presented in [Goré, 2000]. Sometimes making a detour via a modal translation may be useful. In [Wansing, 1999], a modal translation into  $\mathbf{S4}$  has been used to give a cut-free display sequent calculus for a certain constructive modal logic of consistency, for which no other proof system is known. In view of the possible-worlds semantics for  $\mathbf{BiInt}$  and the familiar modal translation of  $\mathbf{Int}$  into  $\mathbf{S4}$  (see [Gödel, 1933]), a faithful modal translation  $m$  of  $\mathbf{BiInt}$  into  $\mathbf{S4t}$  can be straightforwardly defined as follows:

1.  $m(p) = [F]p$ , for every propositional variable  $p$ ;
2.  $m(t) = t$ ;
3.  $m(f) = f$ ;
4.  $m(A \sharp B) = m(A) \sharp m(B)$ ,  $\sharp \in \{\wedge, \vee\}$ ;
5.  $m(A \triangleright B) = [F](m(A) \supset m(B))$ ;
6.  $m(A \blacktriangleleft B) = \langle P \rangle \neg(m(A) \supset m(B))$ .

**THEOREM 67.** ([Lukowski, 1996]) *A formula  $A$  in the language of  $\mathbf{BiInt}$  is provable in  $\mathbf{HBiInt}$  iff  $m(A)$  is provable in  $\mathbf{S4t}$ .*

**DEFINITION 68.** The display sequent system  $\mathbf{DBiInt}$  consists of (id), (cut), the basic structural rules (1) – (4) of Section 1.3, rules  $(\rightarrow t)$ ,  $(t \rightarrow)$ ,  $(\rightarrow f)$ ,  $(f \rightarrow)$ ,  $(\rightarrow \wedge)$ ,  $(\wedge \rightarrow)$ ,  $(\rightarrow \vee)$ ,  $(\vee \rightarrow)$ , the structural rules from Table 2 and:

$$\begin{array}{ll}
(\rightarrow \frown) & \bullet X \rightarrow *A \vdash X \rightarrow \frown A \\
(\frown \rightarrow) & *A \rightarrow X \vdash \frown A \rightarrow \bullet X \\
(\rightarrow \smile) & X \rightarrow *A \vdash \bullet X \rightarrow \smile A \\
(\smile \rightarrow) & *A \rightarrow \bullet X \vdash \smile A \rightarrow X \\
(\rightarrow \triangleright)^m & \bullet X \circ A \rightarrow B \vdash X \rightarrow A \triangleright B \\
(\triangleright \rightarrow)^m & X \rightarrow A \quad B \rightarrow Y \vdash A \triangleright B \rightarrow \bullet(*X \circ Y) \\
(\rightarrow \blacktriangleleft)^m & X \rightarrow A \quad B \rightarrow *X \vdash \bullet X \rightarrow A \blacktriangleleft B \\
(\blacktriangleleft \rightarrow)^m & * \bullet X \circ A \rightarrow B \vdash A \blacktriangleleft B \rightarrow X \\
(\text{persistence}) & p \rightarrow X \vdash \bullet p \rightarrow X \\
(\text{reflexivity}) & X \rightarrow \bullet Y \vdash X \rightarrow Y \\
(\text{transitivity}) & X \rightarrow \bullet Y \vdash X \rightarrow \bullet \bullet Y
\end{array}$$

It can be shown that the persistence rule for arbitrary formulas is an admissible rule of **DBiInt**. This can be used to prove weak completeness of **DBiInt** with respect to **HBiInt**.

LEMMA 69. *In **DBiInt**,  $A \rightarrow X \vdash \bullet A \rightarrow X$ .*

**Proof.** By induction on  $A$ ; for example:

$$\begin{array}{c}
 \frac{}{A \rightarrow A} \\
 \frac{}{*A \rightarrow *A} \\
 \frac{}{\wedge A \rightarrow \bullet * A} \\
 \frac{}{\wedge A \rightarrow \bullet \bullet * A} \\
 \frac{}{\bullet \wedge A \rightarrow \bullet * A} \\
 \frac{}{\bullet \bullet \wedge A \rightarrow *A} \\
 \frac{\bullet \wedge A \rightarrow \wedge A \quad \wedge A \rightarrow X}{\bullet \wedge A \rightarrow X} \text{ (cut)}
 \end{array}$$

$$\begin{array}{c}
 \frac{}{A \rightarrow A} \\
 \frac{}{*A \rightarrow *A} \\
 \frac{}{*A \rightarrow *A \circ B} \quad \frac{}{B \rightarrow B} \\
 \frac{}{*A \rightarrow *A \circ B} \quad \frac{}{B \rightarrow *A \circ B} \\
 \frac{}{*(A \circ B) \rightarrow A} \quad \frac{}{B \rightarrow **(*A \circ B)} \\
 \frac{}{\bullet **(*A \circ B) \rightarrow A \blacktriangleleft B} \\
 \frac{}{*(A \circ B) \rightarrow \bullet(A \blacktriangleleft B)} \\
 \frac{}{*(A \circ B) \rightarrow \bullet \bullet(A \blacktriangleleft B)} \\
 \frac{}{* \bullet \bullet(A \blacktriangleleft B) \rightarrow *A \circ B} \\
 \frac{}{A \circ * \bullet \bullet(A \blacktriangleleft B) \rightarrow B} \\
 \frac{}{* \bullet \bullet(A \blacktriangleleft B) \circ A \rightarrow B} \\
 \frac{}{A \blacktriangleleft B \rightarrow \bullet(A \blacktriangleleft B)} \\
 \frac{}{\bullet(A \blacktriangleleft B) \rightarrow A \blacktriangleleft B} \quad \frac{}{A \blacktriangleleft B \rightarrow X} \\
 \frac{}{\bullet(A \blacktriangleleft B) \rightarrow X}
 \end{array}$$

■

THEOREM 70. *In **DBiInt**  $\vdash \mathbf{I} \rightarrow A$  iff in **HBiInt**  $\vdash A$ .*

**Proof.**  $\Leftarrow$ : By induction on proofs in **HBiInt**. As an example, we here consider only the proof of one axiom schema of **HBiInt**:

$$\begin{array}{c}
\frac{B \rightarrow B}{* \bullet * \bullet \mathbf{I} \circ B \rightarrow B} \\
\frac{A \rightarrow A \quad B \triangleleft B \rightarrow * \bullet \mathbf{I}}{A \triangleright (B \triangleleft B) \rightarrow \bullet (*A \circ * \bullet \mathbf{I})} \\
\frac{A \triangleright (B \triangleleft B) \rightarrow \bullet (*A \circ * \bullet \mathbf{I})}{A \triangleright (B \triangleleft B) \rightarrow *A \circ * \bullet \mathbf{I}} \text{ (reflexivity)} \\
\frac{A \triangleright (B \triangleleft B) \rightarrow * \bullet \mathbf{I} \circ *A}{\bullet \mathbf{I} \circ (A \triangleright (B \triangleleft B)) \rightarrow *A} \\
\frac{\bullet \mathbf{I} \circ (A \triangleright (B \triangleleft B)) \rightarrow * \bullet \mathbf{I} \circ *A}{\bullet (\bullet \mathbf{I} \circ (A \triangleright (B \triangleleft B))) \rightarrow \bullet * A} \text{ (persistence)} \\
\frac{\bullet (\bullet \mathbf{I} \circ (A \triangleright (B \triangleleft B))) \rightarrow \bullet * A}{\bullet \mathbf{I} \circ (A \triangleright (B \triangleleft B)) \rightarrow \frown A} \\
\frac{\bullet \mathbf{I} \circ (A \triangleright (B \triangleleft B)) \rightarrow \frown A}{\mathbf{I} \rightarrow (A \triangleright (B \triangleleft B)) \triangleright \frown A}
\end{array}$$

$\Rightarrow$ : We define the translations  $\tau_1$  and  $\tau_2$  from structures into tense logical formulas as in Section 1.3, except that now  $\tau_1(A) = \tau_2(A) = m(A)$ . By induction on proofs in **DBiInt**, it can be shown that  $\vdash X \rightarrow Y$  in **DBiInt** implies  $\vdash \tau_1(X) \supset \tau_2(Y)$  in **S4t**. Therefore,  $\vdash \mathbf{I} \rightarrow A$  in **DBiInt** implies  $\vdash m(A)$  in **S4t**. By the previous theorem we have  $\vdash A$  in **HBiInt**. ■

**THEOREM 71.** *Strong cut-elimination holds for **DBiInt**.*

**Proof.** **DBiInt** is a proper display calculus. As to the fulfillment of condition C8, the derivation on the left, for example, reduces to the derivation on the right, using contraction:

$$\begin{array}{c}
\frac{X \rightarrow A \quad B \rightarrow *X \quad * \bullet Y \circ A \rightarrow B}{\bullet X \rightarrow A \triangleleft B \quad A \triangleleft B \rightarrow Y} \\
\frac{\bullet X \rightarrow A \triangleleft B \quad A \triangleleft B \rightarrow Y}{\bullet X \rightarrow Y} \\
\frac{* \bullet Y \circ A \rightarrow B}{X \rightarrow A \quad A \rightarrow \bullet Y \circ B} \\
\frac{X \rightarrow \bullet Y \circ B}{* \bullet Y \circ X \rightarrow B \quad B \rightarrow *X} \\
\frac{* \bullet Y \circ X \rightarrow B \quad B \rightarrow *X}{* \bullet Y \circ X \rightarrow *X} \\
\frac{* \bullet Y \circ X \rightarrow *X}{X \rightarrow \bullet Y \circ *X} \\
\frac{X \rightarrow \bullet Y \circ *X}{X \circ X \rightarrow \bullet Y} \\
\frac{X \circ X \rightarrow \bullet Y}{X \rightarrow \bullet Y} \\
\frac{X \rightarrow \bullet Y}{\bullet X \rightarrow Y}
\end{array}$$

■

**COROLLARY 72.** ***DBiInt**  $\cup$  **DS4t** is a conservative extension of both **DBiInt** and **DS4t**.*

As in Section 3.1, let for modal formulas  $A$  the translations  $\tau_i$  ( $i = 1, 2$ ) be defined by  $\tau_i(A) = A$ .

**LEMMA 73.** *In **DBiInt**  $\cup$  **DS4t**, (i)  $\vdash X \rightarrow \tau_1(X)$  and (ii)  $\vdash \tau_2(X) \rightarrow X$ .*

**Proof.** Both (i) and (ii) are proved simultaneously by induction on  $X$ . In particular we have to verify that for every formula of the language of **BiInt**,  $\vdash A \rightarrow m(A)$  and  $\vdash m(A) \rightarrow A$ . But this is the case, see for example:

$$\begin{array}{c}
 \frac{A \rightarrow m(A) \quad m(B) \rightarrow B}{m(A) \supset m(B) \rightarrow *A \circ B} \\
 \frac{*(*A \circ B) \rightarrow *(m(A) \supset m(B))}{*(*A \circ B) \rightarrow \neg(m(A) \supset m(B))} \\
 \frac{\bullet *(*A \circ B) \rightarrow \langle P \rangle \neg(m(A) \supset m(B))}{*\bullet \langle P \rangle \neg(m(A) \supset m(B))} \\
 \frac{*\bullet \langle P \rangle \neg(m(A) \supset m(B)) \rightarrow *A \circ B}{*\bullet \langle P \rangle \neg(m(A) \supset m(B)) \rightarrow B \circ *A} \\
 \frac{*\bullet \langle P \rangle \neg(m(A) \supset m(B)) \circ A \rightarrow B}{A \blacktriangleleft B \rightarrow \langle P \rangle \neg(m(A) \supset m(B))}
 \end{array}$$

$$\begin{array}{c}
 \frac{m(A) \rightarrow A}{m(A) \rightarrow A \circ m(B)} \quad \frac{B \rightarrow m(B)}{B \circ m(A) \rightarrow m(B)} \\
 \frac{*A \circ m(A) \rightarrow m(B)}{*A \rightarrow m(A) \supset m(B)} \quad \frac{B \rightarrow m(A) \supset m(B)}{*(m(A) \supset m(B)) \rightarrow *B} \\
 \frac{*(m(A) \supset m(B)) \rightarrow A}{\neg(m(A) \supset m(B)) \rightarrow A} \quad \frac{\neg(m(A) \supset m(B)) \rightarrow *B}{B \rightarrow *\neg(m(A) \supset m(B))} \\
 \frac{\bullet \neg(m(A) \supset m(B)) \rightarrow A \blacktriangleleft B}{\neg(m(A) \supset m(B)) \rightarrow \bullet(A \blacktriangleleft B)} \\
 \frac{\neg(m(A) \supset m(B)) \rightarrow \bullet(A \blacktriangleleft B)}{\langle P \rangle \neg(m(A) \supset m(B)) \rightarrow A \blacktriangleleft B}
 \end{array}$$

■

**THEOREM 74.** *In **DBiInt**  $\vdash X \rightarrow Y$  iff  $\tau_1(X) \supset \tau_2(Y)$  is valid on every frame (understood as a frame for **S4t**).*

**Proof.** ( $\Rightarrow$ ): This follows by induction on proofs in **DBiInt**. ( $\Leftarrow$ ): Suppose that  $\tau_1(X) \supset \tau_2(Y)$  is valid on every frame. Hence  $\tau_1(X) \supset \tau_2(Y)$  is a theorem of **S4t** and hence  $\vdash \tau_1(X) \rightarrow \tau_2(Y)$  in **DBiInt**  $\cup$  **DS4t**. By the previous lemma,  $\vdash X \rightarrow Y$  in **DBiInt**  $\cup$  **DS4t** and by Corollary 72,  $\vdash X \rightarrow Y$  in **DBiInt**. ■

One advantage of the translation-based sequent system **DBiInt** is that by abandoning combinations of the structural rules (*persistence*), (*reflexivity*), and (*transitivity*), one obtains cut-free sequent calculus presentations of the subsystems of **BiInt** that arise from giving up the corresponding semantic requirements: persistence of atomic information, reflexivity, and transitivity of the relation  $\sqsubseteq$ . Also seriality of  $\sqsubseteq$ , a weakening of reflexivity, is expressible by a purely structural sequent rule, see condition  $D'$  in Table 4.

## 4 INTERRELATIONS AND EXTENSIONS

While the existence of a rich inventory of types of proof systems for modal and other logics may be welcomed, for instance, from the point of view of designing and combining logics, there also exists the need of comparing different approaches and investigating their interrelations and their relative advantages and disadvantages. Mints [1997], for example, presents cut-free systems of indexed sequents for certain extensions of  $\mathbf{K}$  and defines a translation of these sequent systems into equivalent display calculi. In this final section a translation of multiple-sequent systems into higher-arity sequent systems and a translation of hypersequents into display sequents are defined, showing that multiple-sequent systems can be simulated within higher-arity proof systems and that the method of hypersequents can be simulated within display logic. Moreover, one interesting aspect of extending the sequent-style proof systems for modal and temporal *propositional* logics to sequent calculi for modal and temporal *predicate* logics is considered, namely avoiding the provability of the Barcan formula and its converse. We also briefly refer to recent work on display calculi for extended modal languages. Finally, the relation between display logic and Dunn's Gaggles Theory is pointed out.

## 4.1 Translation of multiple-sequent systems

The translation  $\sigma$  in Section 2.4 reveals a straightforward relation between Indrzejczak's multiple-sequent systems and higher-arity sequent systems for modal logics. The intended meaning of the multiple-sequents can be expressed by four-place sequents using a translation  $\mu$ :

$$\begin{aligned}\mu(\Gamma \rightarrow \Delta) &= \delta(\Gamma) \rightarrow_{\emptyset}^{\emptyset} \delta(\Delta) \\ \mu(\Gamma \square \rightarrow \Delta) &= \delta(\Gamma) \rightarrow_{\emptyset}^{\bigvee \delta(\Delta)} \emptyset \\ \mu(\Gamma \diamond \rightarrow \Delta) &= \emptyset \rightarrow_{\emptyset}^{\neg \wedge \delta(\Gamma)} \delta(\Delta).\end{aligned}$$

If  $\mathbf{S}$  is a multiple-sequent system, then let  $\mu(\mathbf{S})$  be the result of the  $\mu$ -translation of the rules of  $\mathbf{S}$ . Let  $\mu^*$  denote the translation of four-place sequents into modal formulas stated in Section 2.3. If  $s_1, \dots, s_n/s$  is a rule of  $\mathbf{MC}$ , then  $\mu^*(\mu(s_1)), \dots, \mu^*(\mu(s_n))/\mu^*(\mu(s))$  is validity preserving in  $\mathbf{C}$ . For the rule  $[TR]$ , for instance, we have  $\mu^*(\mu([TR])) =$

$$\frac{\bigwedge \delta(\Delta) \supset \square \bigvee \delta(\Gamma)}{\diamond \neg \bigvee \delta(\Gamma) \supset \neg \bigwedge \delta(\Delta)} = \frac{\bigwedge \delta(\Delta) \supset \square \bigvee \delta(\Gamma)}{\diamond \bigwedge \delta(\Gamma^*) \supset \bigvee \delta(\Delta^*)}$$

Moreover, (RR) is derivable and  $\mathbf{CPL}$  is contained in  $\mu(\mathbf{MC})$ . Hence,

**OBSERVATION 75.** The system  $\mu(\mathbf{MC})$  is sound and complete with respect to  $\mathbf{C}$ :  $\vdash \Gamma \rightarrow \Delta$  in  $\mu(\mathbf{MC})$  iff  $\mu^*(\mu(\Gamma \rightarrow \Delta))$  is valid in  $\mathbf{C}$ .



The translation  $\mu$  is also faithful for the extension of **MC** by the rules [nec], [D], [T], and [4] and extensions of **C** by the necessitation rule and the axiom schemata *D*, *T* and 4.

#### 4.2 Translation of hypersequents

In order to characterize various non-classical logics by means of hypersequential calculi, Avron [1996] uses different semantical readings of hypersequents. Basically a distinction can be drawn between interpreting the sequent arrow of a component in a hypersequent as material implication or as a constructive implication not definable in terms of Boolean negation and disjunction. This difference in interpretation requires different translations of hypersequents into display sequents. If the sequent arrow is interpreted constructively, a suitable translation may, for example, exploit a faithful embedding of the logic under consideration into a normal modal or temporal logic. In such a case, the sequent arrow is interpreted as strict material implication. In [Wansing, 1998, Chapter 11] translations of hypersequents into display sequents are defined that simulate hypersequents in Avron's hypersequential calculi **GL3**, **GS5**, and **GLC** for Łukasiewicz 3-valued logic **L3**, **S5**, and Dummett's superintuitionistic logic **LC**, also called Gödel-Dummett logic. We shall here consider only the translations suitable for **S5** and **LC**. The treatment of **GL3** is slightly more involved, because **L3** comprises connectives from different 'families' of logical operations. To deal with this composite character of **L3** in display logic, the structure connective  $\circ$  is replaced by two binary structure operations  $\circ_c$  and  $\circ_i$ , see [Wansing, 1998]. If  $\Delta = \{A_1, \dots, A_n\}$ , let  $*\Delta = \{*A_1, \dots, *A_n\}$ . Since  $\circ$  is assumed to be associative and commutative, we may put  $(\circ\Delta) = A_1 \circ \dots \circ A_n$ . If  $\Delta = \emptyset$ , let  $*\Delta = (\circ\Delta) = \mathbf{I}$ . Recall the notion of hypersequent from Section 2.5.

**DEFINITION 76.** The translation  $\eta_0$  of ordinary sequents into display structures is defined by

$$\eta_0(\Delta \rightarrow \Gamma) = \bullet((\circ * \Delta) \circ (\circ \Gamma)),$$

and the translation  $\eta$  of non-empty hypersequents into display sequents is defined by

$$\eta(s_1 \mid \dots \mid s_n) = \mathbf{I} \rightarrow \eta_0(s_1) \circ \dots \circ \eta_0(s_n).$$

**THEOREM 77.** *For every hypersequent  $H$ ,  $\vdash \eta(H)$  in **DS5** iff  $\vdash H$  in **GS5**.*

In the hypersequential system **GLC** the components of a hypersequent are restricted to be ordinary Gentzen sequents with at most a single conclusion. Dummett's **LC** is the logic of linearly ordered intuitionistic Kripke

models. An axiomatization of **LC** is obtained from an axiomatization **HInt** of **Int** by adding the axiom schema  $(A \triangleright B) \vee (B \triangleright A)$ . It is well-known that the modal translation  $m$  defined in Section 3.8 (restricted to the language of intuitionistic logic, i.e. the language of **LC**) is a faithful embedding of **LC** into **S4.3**, the logic of linearly ordered modal Kripke models.

**THEOREM 78.** *For every formula  $A$  in the language of **LC**,  $\vdash A$  in **LC** iff  $\vdash m(A)$  in **S4.3**.*

**DEFINITION 79.** The translation  $\zeta_0$  of a single-conclusion ordinary sequent  $s = A_1, \dots, A_n \rightarrow B$  is defined by

$$\zeta_0(s) = \bullet(*A_1 \circ \bullet(*A_2 \circ \dots \bullet(*A_n \circ B) \dots)).$$

If  $s = A_1, \dots, A_n \rightarrow \emptyset$ , then  $\zeta(s) = \bullet(*A_1 \circ \bullet(*A_2 \circ \dots \bullet(*A_n \circ \mathbf{I}) \dots))$ . If  $s = \emptyset \rightarrow B$ , then  $\zeta(s) = \bullet(*\mathbf{I} \circ B)$ , and if  $s = \emptyset \rightarrow \emptyset$ ,  $\zeta(s) = \bullet(*\mathbf{I} \circ \mathbf{I})$ . The translation  $\zeta$  of hypersequents with at most single-conclusion components into display sequents is defined by

$$\zeta(s_1 \mid \dots \mid s_n) = \mathbf{I} \rightarrow \zeta_0(s_1) \circ \dots \circ \zeta_0(s_n).$$

**THEOREM 80.** *For every hypersequent  $H$  with at most single-conclusion components,  $\vdash \zeta(H)$  in **DLC** iff  $\vdash H$  in **GLC**.*

### 4.3 Predicate logics and other logics

Modal predicate logic is still a largely unexplored area. As to sequent systems for modal predicate logics, one notorious problem is providing introduction rules for the modal operators and the quantifiers such that neither the Barcan formula (BF)  $\forall x \Box A \supset \Box \forall x A$  nor its converse (BFc)  $\Box \forall x A \supset \forall x \Box A$  are provable on the strength of only these rules. It is well-known that (BF) corresponds to the assumption of constant domains and (BFc) to the persistence of individuals along the accessibility relation; cf. for example [Fitting, 1993]. One way of avoiding the provability of the Barcan formula and its converse is described in [Wansing, 1998, Chapter 12]. The idea is to exploit the well-known similarity between  $\Box [\diamond]$  and  $\forall x [\exists x]$  to develop display introduction rules for  $\forall x [\exists x]$ ; i.e., instead of thinking of the modal operators as quantifiers, one thinks of the quantifiers as modal operators, see also [Andreka *et al.*, 1998]. The addition of quantifiers to display logic is briefly discussed in [Belnap, 1982]:

Quantifiers may be added with the obvious rules:

$$(UQ) \quad \frac{Aa \vdash X}{(x)Ax \vdash X} \quad \frac{X \vdash Aa}{X \vdash (x)Ax}$$

provided, for the right rule, that  $a$  does not occur free in the conclusion. . . . The rule for the existential quantifier would be

dual. ... [A]s yet this addition provides no extra illumination. I think that is because these rules for quantifiers are “structure free” (no structure connectives are involved; ...). One upshot is that adding these quantifiers to modal logic brings along Barcan and its converse ... willy-nilly, which is an indication of an unrefined account; alternatives therefore need investigating. [Belnap, 1982, p. 408 f.]

Using the structure-independent rules (UQ), we would have the following proofs of (BF) and (BFc):

$$\begin{array}{l}
 \frac{A \rightarrow A}{\Box A \rightarrow \bullet A} \text{ (UQ)} \\
 \frac{\Box A \rightarrow \bullet A}{\forall x \Box A \rightarrow \bullet A} \\
 \frac{\forall x \Box A \rightarrow \bullet A}{\bullet \forall x \Box A \rightarrow A} \text{ (UQ)} \\
 \frac{\bullet \forall x \Box A \rightarrow A}{\bullet \forall x \Box A \rightarrow \forall x A} \\
 \frac{\bullet \forall x \Box A \rightarrow \forall x A}{\forall x \Box A \rightarrow \Box \forall x A} \\
 \frac{\forall x \Box A \rightarrow \Box \forall x A}{\mathbf{I} \circ \forall x \Box A \rightarrow \Box \forall x A} \\
 \mathbf{I} \rightarrow \forall x \Box A \supset \Box \forall x A
 \end{array}
 \qquad
 \begin{array}{l}
 \frac{A \rightarrow A}{\forall x A \rightarrow A} \text{ (UQ)} \\
 \frac{\forall x A \rightarrow A}{\Box \forall x A \rightarrow \bullet A} \\
 \frac{\Box \forall x A \rightarrow \bullet A}{\bullet \Box \forall x A \rightarrow A} \\
 \frac{\bullet \Box \forall x A \rightarrow A}{\Box \forall x A \rightarrow \Box A} \text{ (UQ)} \\
 \frac{\Box \forall x A \rightarrow \Box A}{\Box \forall x A \rightarrow \forall x \Box A} \\
 \frac{\Box \forall x A \rightarrow \forall x \Box A}{\mathbf{I} \circ \Box \forall x A \rightarrow \forall x \Box A} \\
 \mathbf{I} \rightarrow \Box \forall x A \supset \forall x \Box A
 \end{array}$$

Structure-dependent introduction rules for  $\forall x$  and  $\exists x$  are, however, available. For every binary relation  $\mathcal{R}_x$  on a non-empty set  $S$  of states, we may define the following functions on the powerset of  $S$ :

$$\begin{aligned}
 \forall x A &:= \{a \mid \forall b (a\mathcal{R}_x b \Rightarrow b \in A)\}, & \exists x^\checkmark A &:= \{a \mid \exists b (b\mathcal{R}_x a \ \& \ b \in A)\}, \\
 \forall x^\checkmark A &:= \{a \mid \forall b (b\mathcal{R}_x a \Rightarrow b \in A)\}, & \exists x A &:= \{a \mid \exists b (a\mathcal{R}_x b \ \& \ b \in A)\}.
 \end{aligned}$$

We then have

$$\exists x^\checkmark A \subseteq B \text{ iff } A \subseteq \forall x B, \quad \exists x A \subseteq B \text{ iff } A \subseteq \forall x^\checkmark B,$$

and for every individual variable  $x$ , we may introduce a structure connective  $\bullet_x$ , which in succedent position is to be understood as  $\forall x$  and in antecedent position as a backward-looking existential quantifier  $\exists x^\checkmark$ . Semantically, what is required to account for these quantifiers is a generalization of the Tarskian semantics for first-order logic, see [Andreka *et al.*, 1998]. Let  $\mathcal{M}$  be any first-order model and let  $\alpha, \beta, \dots$  range over variable assignments in  $\mathcal{M}$ . Tarski’s truth definition for the existential quantifier is:

$$\begin{aligned}
 \mathcal{M} \models \exists x A[\alpha] \text{ iff } & \text{for some assignment } \beta \text{ on } |\mathcal{M}|: \\
 & \alpha =_x \beta \text{ and } \mathcal{M} \models A[\beta],
 \end{aligned}$$

where  $\alpha =_x \beta$  means that  $\alpha$  and  $\beta$  differ at most with respect to the object assigned to  $x$ . In the more general semantics the concrete relations  $=_x$  between variable assignments are replaced by abstract binary relations  $\mathcal{R}_x$  of ‘variable update’ between ‘states’  $\alpha, \beta, \gamma, \dots$  from a set of states  $S$ .

Assuming an interpretation of atoms containing free variables, the truth definition for the existential quantifier becomes:

$$\mathcal{M}, \alpha \models \exists x A \quad \text{iff for some } \beta \in S : \alpha \mathcal{R}_x \beta \text{ and } \mathcal{M}, \beta \models A$$

Thus, to every individual variable  $x$  there is associated a transition relation  $\mathcal{R}_x$  on states. The resulting minimal predicate logic, **KFOL**, is nothing but the  $\omega$ -modal version of the minimal normal modal logic **K**. In order to obtain an axiomatization of **KFOL**, one may just take any axiomatic presentation of **K** and replace every occurrence of  $\diamond$  and  $\square$  by one of  $\exists x$  and  $\forall x$ , respectively. The basic structural rules for the structure connective  $\bullet_x$  are:

$$X \rightarrow \bullet_x Y \dashv\vdash \bullet_x X \rightarrow Y.$$

In analogy to the case for  $\square$  and  $\diamond$ , we obtain the following structure-dependent introduction rules for  $\forall x$  and  $\exists x$ :

$$\begin{array}{ll} (\rightarrow \forall x) & \bullet_x X \rightarrow A \vdash X \rightarrow \forall x A \\ (\forall x \rightarrow) & A \rightarrow X \vdash \forall x A \rightarrow \bullet_x X \end{array} \quad \begin{array}{ll} (\rightarrow \exists x) & X \rightarrow A \vdash * \bullet_x * X \rightarrow \exists x A \\ (\exists x \rightarrow) & * \bullet_x * A \rightarrow X \vdash \exists x A \rightarrow X \end{array}$$

In addition to these introduction rules we need further structural assumption in order to take care of the necessitation rules in axiomatic presentations of normal modal and tense logics:

$$(MN\bullet_x) \quad \mathbf{I} \rightarrow X \vdash \mathbf{I} \rightarrow \bullet_x X \quad X \rightarrow \mathbf{I} \vdash X \rightarrow \bullet_x \mathbf{I}$$

The structural account of the quantifiers as modal operators blocks the above proofs of (BF) and (BFc). In the presence of additional structural sequent rules, however, these schemata become derivable:

OBSERVATION 81. BF and BFc correspond to the structural rules

$$\text{rBF} \quad X \rightarrow \bullet_x \bullet_x Y \vdash X \rightarrow \bullet_x Y; \quad \text{rBFc} \quad X \rightarrow \bullet_x X \vdash X \rightarrow \bullet_x Y.$$

The apparatus of display logic has also been applied to other extensions of normal modal propositional logic. A result of Kracht concerns the undecidability of decidability of display calculi. Consider the fusion or ‘independent sum’ of **Kf** and **Kf**, i.e. the bimodal logic **Kf**  $\otimes$  **Kf** of two functional accessibility relations  $\mathcal{R}_1, \mathcal{R}_2$ . In this system there are two pairs of modal operators, say,  $[1], \langle 1 \rangle$  and  $[2], \langle 2 \rangle$  each satisfying the *D* and the *Alt1* axiom schemata. The structural language of sequents for this logic comes with two unary operations  $\bullet_1$  and  $\bullet_2$  satisfying the display equivalence

$$\bullet_i X \rightarrow Y \dashv\vdash X \rightarrow \bullet_i Y,$$

$i = 1, 2$ . Clearly, **Kf**  $\otimes$  **Kf** has many properly displayable extensions. Using an encoding of Thue-processes into frames of **Kf**  $\otimes$  **Kf**, Grefe and Kracht [1996] have proved a theorem about the undecidability of decidability.

**THEOREM 82.** (Grefe and Kracht) *It is undecidable whether or not a display calculus is decidable.*

According to Kracht, Theorem 82 indicates a serious weakness of display logic. In any case, the theorem provides insight into the expressive power of display logic; it shows that the subformula property and the strong cut-elimination theorem for displayable logics fail to guarantee decidability. Undecidability of the decidability of properly displayable extensions of  $\mathbf{Kf} \otimes \mathbf{Kf}$  is a remarkable property of this particular family of bimodal logics, but is *not* a defect of the modal display calculus, at least insofar as the proof of the theorem also shows that it is undecidable whether or not a finite axiomatic calculus is decidable. Would it be desirable to have a proof-theoretic framework in which only decidable logics can be presented? A weakness of display logic is that it does not lend itself easily to obtain decidability proofs. Restall [1998] uses a display presentation to prove, among other things, decidability of certain relevance logics which are not known to have the finite model property. In [Wansing, 1998, Chapter 6] display logic is used to prove decidability of  $\mathbf{Kf}$  and deterministic dynamic propositional logic without Kleene star.

Display calculi for logics with relative accessibility relations can be found in [Demri and Goré, 2000] and for nominal tense logics in [Demri and Goré, 1999]. In both cases the calculi are obtained using modal translations.

#### 4.4 Gaggle Theory

The generality of display logic has been highlighted by Restall [1995], who observes a close relation between display logic and J. Michael Dunn's *Gaggle Theory* [1990; 1993; 1995]. The relation between gaggle theory and display logic has also been investigated and worked out by Goré [1998]. A *gaggle* is an algebra  $\mathcal{G} = \langle \mathbf{G}, \leq, OP \rangle$ , where  $\leq$  is a distributive lattice ordering on  $\mathbf{G}$ , and  $OP$  is a founded family of operations. The latter means that there is an  $f \in OP$  such that for every  $g \in OP$ ,  $f$  and  $g$  satisfy the abstract law of residuation, see Section 3. If one only requires that  $\leq$  is a partial order, and every  $f \in OP$  has a trace, then  $\mathcal{G}$  is said to be a *tonoid*. Restall defines the notion of *mimicing structure*. An  $n$ -place logical operation  $f$  mimics antecedent structure if there is a possibly complex  $n$ -place structure connective  $\sharp$  such that the following rules are admissible:

$$\begin{aligned} s = \sharp(A_1, \dots, A_n) \rightarrow X \vdash f(A_1, \dots, A_n) \rightarrow X \\ \mathcal{C}(X_1, A_1) \dots \mathcal{C}(X_n, A_n) \vdash \sharp(A_1, \dots, A_n) \rightarrow f(A_1, \dots, A_n) \end{aligned}$$

where  $\sharp(A_1, \dots, A_n)$  is an antecedent part of  $s$ ,  $\mathcal{C}(X_i, A_i) = X_i \rightarrow A_i$ , if  $A_i$  is an antecedent part of  $\sharp(A_1, \dots, A_n)$ , and  $\mathcal{C}(X_i, A_i) = A_i \rightarrow X_i$ , if  $A_i$  is a succedent part of  $\sharp(A_1, \dots, A_n)$ . Dually,  $f$  mimics succedent structure

if there is a possibly complex  $n$ -place structure connective  $\sharp$  such that the following rules are admissible:

$$\begin{aligned} s = X \rightarrow \sharp(A_1, \dots, A_n) \vdash X \rightarrow f(A_1, \dots, A_n) \\ \mathcal{C}(X_1, A_1) \dots \mathcal{C}(X_n, A_n) \vdash f(A_1, \dots, A_n) \rightarrow \sharp(A_1, \dots, A_n) \end{aligned}$$

where  $\sharp(A_1, \dots, A_n)$  is a succedent part of  $s$ ,  $\mathcal{C}(X_i, A_i) = X_i \rightarrow A_i$ , if  $A_i$  is an antecedent part of  $\sharp(A_1, \dots, A_n)$ , and  $\mathcal{C}(X_i, A_i) = A_i \rightarrow X_i$ , if  $A_i$  is a succedent part of  $\sharp(A_1, \dots, A_n)$ .

**THEOREM 83.** (Restall [1995]) *If a logical operation  $f$  in a display calculus presentation  $\mathbf{DA}$  of a logic  $\Lambda$  mimics structure, then  $f$  is a tonoid operator on the Lindenbaum algebra of  $\Lambda$ .*

If every logical operation of  $\mathbf{DA}$  mimics structure, mutual provability is a congruence relation and  $\Lambda$  has an algebraic semantics. Dunn's representation theorem for tonoids supplies also a Kripke-style relational semantics.

## 5 APPENDICES

### 5.1 Appendix A

The proof of Theorem 23 takes its pattern from the proof of strong normalization for typed  $\lambda$ -calculus (see for instance [Hindley and Seldin, 1986, Appendix 2]) and follows the argument given in [Roorda, 1991, Chapter 2, reprinted in [Troelstra, 1992]]. This proof has been extracted from the proof of strong cut-elimination for classical predicate logic in [Dragalin, 1988, Appendix B]. Suppose that  $\Pi$  is a proof containing an application of cut. A (one-step) reduction of  $\Pi$  is the proof  $\Sigma$  resulting by applying a primitive reduction to a subproof of  $\Pi$ . If  $\Pi$  reduces to  $\Sigma$ , this is denoted by  $\Pi > \Sigma$  (or  $\Sigma < \Pi$ ).  $\Pi$  is said to be reducible iff there is a  $\Sigma$  such that  $\Pi > \Sigma$ .

**LEMMA 84.** *If a proof cannot be reduced, then it is cut-free.*

**Proof.** Since the case distinction in the definition of primitive reductions is exhaustive, every proof that contains an application of cut is reducible. ■

**DEFINITION 85.** We inductively define the set of *inductive* proofs.

- a** Every instantiation of an axiomatic rule is an inductive proof.
- b** If  $\Pi$  ends in an inference *inf* different from cut, and every premise  $s_i$  of *inf* has an inductive proof  $\Pi_i$  in  $\Pi$ , then  $\Pi$  is inductive.
- c**  $\Pi = \frac{\Pi_1 \quad \Pi_2}{(3)} \text{ cut}$  is inductive, if every  $\Sigma$  such that  $\Pi > \Sigma$  is inductive.

LEMMA 86. *If  $\Pi$  is inductive, and  $\Pi > \Sigma$ , then  $\Sigma$  is inductive.*

**Proof.** By induction on the construction of  $\Pi$ . If  $\Pi$  is inductive by **a**, then no reduction can be performed. If  $\Pi$  is inductive by **b**, then every reduction on  $\Pi$  takes place in the  $\Pi_i$ 's, which are inductive. Hence, by the induction hypothesis,  $\Sigma$  is inductive due to **b**. If  $\Pi$  is inductive by **c**, then  $\Sigma$  is inductive by definition. ■

DEFINITION 87. Let  $\Pi$  be an inductive proof. The size  $ind(\Pi)$  of  $\Pi$  is inductively defined as follows (the clauses correspond to those in the previous definition):

**a**  $ind(\Pi) = 1$ ;

**b**  $ind(\Pi) = \sum_i ind(\Pi_i) + 1$ ;

**c**  $ind(\Pi) = \sum_{\Sigma < \Pi} ind(\Sigma) + 1$ .

A proof  $\Pi$  is said to be *strongly normalizable* iff every sequence of reductions starting at  $\Pi$  terminates.

LEMMA 88. *Every inductive proof is strongly normalizable.*

**Proof.** By induction on  $ind(\Pi)$ . If  $ind(\Pi) = 1$ , no reduction is feasible. If  $\Pi$  is inductive by **b**, then every reduction is in the premises  $\Pi_i$ , and we can apply the induction hypothesis. If  $\Pi$  is inductive by **c**, then every proof to which  $\Pi$  reduces is inductive and therefore every such proof is strongly normalizable, by the induction hypothesis. But then  $\Pi$  is also strongly normalizable. ■

LEMMA 89. *Let  $\Pi$  be an inductive proof and let  $inf$  be the final inference of  $\Pi$ . If  $\Pi > \Pi'$  by reducing a proof  $\Pi_j$  of a premise sequent of  $inf$ , then  $ind(\Pi) > ind(\Pi')$ .*

**Proof.** By induction on  $ind(\Pi)$ . If  $ind(\Pi) = 1$ , then  $\Pi$  cannot be reduced. Whence  $\Pi$  is inductive by **b** or **c**. If  $\Pi$  is inductive by **c**, then by definition,  $ind(\Pi) > ind(\Pi')$ . If  $\Pi$  is inductive by **b**, then  $\Pi_j$  is inductive by definition. If  $\Pi_j$  is inductive by **a**, it cannot be reduced. If  $\Pi_j$  is inductive by **b**, then the reduction of  $\Pi_j$  to  $\Pi'_j$  takes place in the proof of some premise sequent of the final inference of  $\Pi_j$ . By the induction hypothesis,  $ind(\Pi_j) > ind(\Pi'_j)$ . Hence  $ind(\Pi) > ind(\Pi')$ . If  $\Pi_j$  is inductive by **c**, then by definition,  $ind(\Pi_j) > ind(\Pi'_j)$  and thus  $ind(\Pi) > ind(\Pi')$ . ■

LEMMA 90. *Suppose  $\Pi$  ends in an application  $inf$  of cut, and  $\Pi_1$  and  $\Pi_2$  are the proofs of the premises of  $inf$ . If  $\Pi_1$  and  $\Pi_2$  are inductive, then so is  $\Pi$ .*

**Proof.** We must show that every  $\Sigma < \Pi$  is inductive. For this purpose, we define two complexity measures for  $\Pi$ :  $r(\Pi)$ , the rank of  $\Pi$ , and  $h(\Pi)$ , the height of  $\Pi$ .  $r(\Pi)$  is the number of symbols in the cut-formula.  $h(\Pi)$  is defined by:

$$h(\Pi) = \text{ind}(\Pi_1) + \text{ind}(\Pi_2).$$

We use induction on  $r(\Pi)$  and, for fixed rank, induction on  $h(\Pi)$ .

Case 1.  $\Sigma$  is obtained by reduction in  $\Pi_1$  or  $\Pi_2$ , say  $\Pi_1 > \Pi'_1$ . It follows from Lemma 89 that  $\text{ind}(\Pi'_1) < \text{ind}(\Pi_1)$ . Then  $h(\Sigma) < h(\Pi)$ . Since  $\Pi_1$  and  $\Pi_2$  are inductive, by Lemma 86,  $\Sigma$  has inductive premises, and by the induction hypothesis for  $h(\Pi)$ ,  $\Sigma$  is inductive.

Case 2.  $\Sigma$  is obtained by reducing *inf*. Then this reduction was either a principal or a parametric move.

**Principal move.**

Case 1. Since  $\Sigma$  proves one of (1) or (2),  $\Sigma$  is inductive by assumption.

Case 2. Since for every new proof  $\Pi'$  ending in an application of cut,  $r(\Pi) > r(\Pi')$ ,  $\Sigma$  is inductive by the induction hypothesis for  $r(\Pi)$ .

**Parametric move.** Suppose  $A$  is parametric in the inference ending in (1) (the case for (2) is analogous). If the tree of parametric ancestors of the displayed occurrence of  $A$  in (1) contains at most one element  $A_u$  that is not parametric in *inf*, we have Figure 1, and we may assume that there is no application of cut on the path from (1) to  $Z \rightarrow A$ .

$$\text{Let } \Pi' = \frac{\frac{\Pi^1}{Z \rightarrow A} \quad \Pi_2}{Z \rightarrow Y} \quad \text{and } \Pi'' = \frac{\Pi^1}{Z \rightarrow A}.$$

Consider  $\Pi$  and  $\Pi'$ . Clearly,  $r(\Pi) = r(\Pi')$ , hence we use induction on the height. Since both  $\Pi_1$  and  $\Pi''$  are inductive by **b**,  $\text{ind}(\Pi'') < \text{ind}(\Pi_1)$ . Hence  $h(\Pi') < h(\Pi)$ . By the induction hypothesis for  $h(\Pi)$ ,  $\Pi'$  is inductive, and thus  $\Sigma$  is inductive by definition. If the primitive reduction of  $\Pi$  to  $\Sigma$  requires cutting with  $\Pi_2$  more than once, analogously every new  $\Pi'$  and hence  $\Sigma$  can be shown to be inductive.

If the tree of parametric ancestors of the displayed occurrence of  $A$  in (1) contains *more than one* element  $A_u$  that is not parametric in *inf*,  $\Sigma = \Pi'^*$  or  $\Sigma = \Pi'^{r*}$ . Since for every new proof  $\Pi'$  ending in an application of cut,  $r(\Pi) > r(\Pi')$ ,  $\Sigma$  is inductive by the induction hypothesis for  $r(\Pi)$ . ■

**COROLLARY 91.** *Every proof is inductive.*

Now Theorem 23 follows by Lemma 88 and Corollary 91, and cut is an admissible rule by Lemma 84.



## 5.2 Appendix B

To prove completeness of **HIntKt** with respect to the class of all temporal models we shall adopt completely standard methods as applied, for example, in [Schütte 1969, pp. 48–51]. Suppose  $\Delta$  and  $\Gamma$  are finite sets of formulas, where  $\Gamma$  is empty or a singleton, and let  $p$  be a new propositional variable not already in *Atom*. The formula  $\Delta \triangleright \Gamma$  is defined as follows:

$$\Delta \triangleright \Gamma = \begin{cases} \bigwedge \Delta \triangleright B & \text{if } \Delta \neq \emptyset, \Gamma = \{B\} \\ \mathbf{t} \triangleright B & \text{if } \Delta = \emptyset, \Gamma = \{B\} \\ \bigwedge \Delta \triangleright p & \text{if } \Delta \neq \emptyset, \Gamma = \emptyset \\ \mathbf{t} \triangleright p & \text{if } \Delta = \Gamma = \emptyset \end{cases}$$

The pair  $(\Delta, \Gamma)$  is said to be consistent if  $\Delta \triangleright \Gamma$  is unprovable in **HIntKt** based on  $\mathcal{L}^+ = \mathcal{L} \cup \{p\}$ . In what follows, let  $A \in \mathcal{L}$ . Let  $\text{sub}(A)$  denote the finite set of all subformulas of  $A$ . If  $C = (A_1 \triangleright \dots (A_{n-1} \triangleright A_n) \dots)$ , then  $\text{sub}^*(\{C\}) = (\bigcup_{1 \leq i \leq n} \text{sub}(A_i)) \setminus \{p\}$ ;  $\text{sub}^*(\emptyset) = \emptyset$ . The pair  $(\Delta, \Gamma)$  is called  $A$ -complete, if  $\Delta \cup \text{sub}^*(\Gamma) = \text{sub}(A)$ . A pair  $(\Delta^*, \Gamma^*)$  is called an expansion of  $(\Delta, \Gamma)$ , if  $\Delta^*$  is a finite superset of  $\Delta$ , and either  $\Gamma^* = \Gamma$  or  $\Gamma^*$  has the shape  $(A_1 \triangleright \dots (A_{n-1} \triangleright A_n) \dots)$  and  $n > 1$ .

**LEMMA 92.** *If  $(\Delta, \Gamma)$  is consistent, then so is  $(\Delta \cup \{A\}, \Gamma)$  or  $(\Delta, \{A \triangleright B\})$ , where  $B = p$  if  $\Gamma = \emptyset$ , and  $\Gamma = \{B\}$  otherwise.*

**Proof.** Suppose neither  $(\Delta \cup \{A\}, \Gamma)$  nor  $(\Delta, \{A \triangleright B\})$  are consistent. Then both  $(\bigwedge \Delta \wedge A) \triangleright B$  and  $\bigwedge \Delta \triangleright (A \triangleright B)$  are derivable in **HIntKt** based on  $\mathcal{L}^+$ . But then also  $\bigwedge \Delta \triangleright B$  is derivable, and hence  $(\Delta, \Gamma)$  is not consistent; a contradiction. ■

**COROLLARY 93.** *Every consistent pair  $(\Delta, \Gamma)$  such that  $\Delta, \text{sub}^*(\Gamma) \subseteq \text{sub}(A)$  can be expanded to an  $A$ -complete consistent pair.*

Let  $\Delta \subseteq \text{sub}(A)$ . Then  $\Delta$  is said to be  $A$ -designated, if some  $A$ -complete pair  $(\Delta, \Gamma)$ , where  $\text{sub}^*(\Gamma) = \text{sub}(A) \setminus \Delta$  is consistent. By soundness of **HIntKt** based on  $\mathcal{L}^+$ , the formula  $\mathbf{t} \triangleright p$  fails to be provable. Therefore  $(\emptyset, \emptyset)$  is consistent. By the previous corollary, for every formula  $A$ ,  $(\emptyset, \emptyset)$  can be expanded to an  $A$ -complete consistent pair. Hence, for every  $A$ , the set  $\mathcal{D}(A)$  of all  $A$ -designated subsets of  $\text{sub}(A)$  is non-empty.

**LEMMA 94.** *If  $C \in \text{sub}(A)$ , then  $C$  belongs to an  $A$ -designated set  $\Delta$  iff  $\Delta \triangleright \{C\}$  is provable in **HIntKt**.*

**Proof.** If  $C \in \Delta$ , then clearly  $\Delta \triangleright \{C\}$  is provable in **HIntKt**. If  $C \notin \Delta$ , then  $C \in \text{sub}(A) \setminus \Delta$ , and since  $\Delta$  is  $A$ -designated,  $(\Delta, \{C\})$  is consistent. In other words,  $\Delta \triangleright \{C\}$  is not provable in **HIntKt**. ■

DEFINITION 95. For every formula  $A$ , the structure  $\mathcal{M}^A = \langle W^A, R_I^A, R_T^A, v^A \rangle$  is called the canonical model for  $A$  if

$$\begin{aligned} W^A &= \mathcal{D}(A) \\ R_I^A &= \subseteq \\ uR_T^A t &\text{ iff } [F]B \in u \text{ implies } B \in t \\ v^A(p, u) = 1 &\text{ iff } p \in u. \end{aligned}$$

As we have seen, the set  $W^A$  is non-empty, and it can easily be shown that  $\mathcal{M}^A$  is indeed a temporal model.

LEMMA 96. *Let  $u, t \in \mathcal{D}(A)$ . For every formula  $B$ , ( $[F]B \in u$  implies  $B \in t$ ) iff for every formula  $C$ , ( $C \in u$  implies  $\langle P \rangle C \in t$ ).*

**Proof.** First, suppose (i) for all  $B$ ,  $[F]B \in u$  implies  $B \in t$  but (ii) there is a formula  $C \in u$  such that  $\langle P \rangle C \notin t$ . By (i),  $[F]\langle P \rangle C \notin u$ . By the previous lemma,  $u \triangleright [F]\langle P \rangle C$  is not provable in **HintKt**. Since  $C \triangleright [F]\langle P \rangle C$  is provable, also  $u \triangleright C$  fails to be provable. But then, by the previous lemma,  $C \notin u$ , which contradicts (ii). Suppose now (iii) for all  $C$ ,  $C \in u$  implies  $\langle P \rangle C \in t$  but (iv) there is a formula  $[F]B \in u$  such that  $B \notin t$ . By (iii),  $\langle P \rangle [F]B \in t$ , and by the previous lemma,  $t \triangleright \langle P \rangle [F]B$  is provable in **HintKt**. Since  $\langle P \rangle [F]B \triangleright B$  is provable, also  $t \triangleright B$  is provable. Hence  $B \in t$ , a contradiction with (iv).  $\blacksquare$

LEMMA 97. (Verification Lemma) *Consider  $\mathcal{M}^A = \langle W^A, R_I^A, R_T^A, v^A \rangle$ . For every  $C \in \text{sub}(A)$  and every  $u \in \mathcal{D}(A)$ ,  $\mathcal{M}^A, u \models C$  iff  $C \in u$ .*

**Proof.** By induction on  $C$ . We shall consider only two cases. Let  $\bigwedge u$  denote  $t$ , if  $u = \emptyset$ , and note that for all  $B \in u$ ,  $\vdash \langle P \rangle \bigwedge u \triangleright \langle P \rangle B$ . Hence for every  $u, t \in W^A$  we have: (\*) if  $\langle P \rangle \bigwedge u \in t$ , then for every  $B \in u$ ,  $\langle P \rangle B \in t$ .

1.  $C = [F]B$ .

$\Rightarrow$ : Suppose  $[F]B \notin u$ . This is the case iff

$$\begin{aligned} &\bigwedge u \triangleright [F]B \text{ cannot be proved} \\ \text{iff} &\quad \langle P \rangle \bigwedge u \triangleright B \text{ cannot be proved} \\ \text{iff} &\quad (\langle P \rangle \bigwedge u, \{B\}) \text{ is consistent} \\ \text{iff} &\quad (\exists t \in \mathcal{D}(A)) u \subseteq t, \langle P \rangle \bigwedge u \in t, B \notin t \quad \text{by Corollary 93} \\ \text{only if} &\quad (\exists t \in \mathcal{D}(A)) uR_T^A t, B \notin t \quad \text{by Lemma 96 and (*)} \\ \text{iff} &\quad \mathcal{M}, u \not\models [F]B \quad \text{by the ind. hyp.} \end{aligned}$$

$\Leftarrow$ : Suppose  $[F]B \in u$ . Then for all  $t \in W^A$ ,  $uR_T^A t$  implies  $B \in t$ . By the induction hypothesis,  $\mathcal{M}^c, u \models [F]B$ .

2.  $C = \langle P \rangle B$ .

$\Rightarrow$ : Suppose  $\mathcal{M}^A, u \models \langle P \rangle B$ . This is the case iff

only if  $(\exists t \in W^A) tR_T^A u$  and  $\mathcal{M}^A, t \models B$   
 $(\exists t \in W^A) (B \in t \text{ implies } \langle P \rangle B \in u), B \in t$  by Lem. 96,  
 ind. hyp.  
 only if  $\langle P \rangle B \in u$ .

$\Leftarrow$ : Suppose  $\langle P \rangle B \in u$ . Put  $t' := \{C \mid \langle P \rangle C \in u\}$ . Clearly, the pair  $(t', \emptyset)$  is consistent. Hence

only if  $(\exists t \in W^A) t' \subseteq t, \bigwedge t' \in t$  by Corollary 85  
 $(\exists t \in W^A) tR_T^A u$  and  $\mathcal{M}^A, t \models B$  by Lemma 88 and the  
 ind. hyp.  
 iff  $\mathcal{M}^A, u \models \langle P \rangle B$

■

**COROLLARY 98.** *If  $A$  is valid in every temporal model, then  $A$  is provable in **HIntKt**.*

**Proof.** Suppose  $A$  is not provable in **HIntKt**. Then the pair  $(\emptyset, \{A\})$  is consistent, and, by the previous corollary, there exists a  $u \in \mathcal{D}(A)$  such that  $A \notin u$ . By the Verification Lemma,  $\mathcal{M}^A, u \not\models A$ . ■

**COROLLARY 99.** **HIntKt** is decidable.

**Proof.** This follows easily by the fact that  $sub(A)$  is finite. ■

### 5.3 Appendix C

In order to prove strong normalization for  $\lambda_t$ , we shall follow R. de Vri-  
 jer's [1987] proof of strong normalization for typed  $\lambda$ -calculus with pairing  
 and projections satisfying surjective pairing. Let  $h(M)$  (the height of the  
 reduction tree of  $M$ ) be the length of a reduction sequence of  $M$  that has  
 maximal length.

**DEFINITION 100.**  $M^A \in Term$  is said to be computable iff

1.  $sn(M)$ ;
2. if  $A = B \triangleright C$ ,  $M \rightarrow_r N_1$ , and  $N_2^B$  is computable, then  $(N_1, N_2)^C$  is  
 computable;
3. if  $A = B \wedge C$  and  $M \rightarrow_r \langle N_1, N_2 \rangle$ , then  $N_1^B, N_2^C$  are computable;
4. if  $A = [F]B$  and  $M \rightarrow_r \mathcal{P}N$ , then  $N^B$  is computable;
5. if  $A = \langle P \rangle B$  and  $M \rightarrow_r \mathcal{S}N$ , then  $N^B$  is computable.

The set of all computable terms is denoted by  $\mathbf{C}$ .

By this definition, every computable term is strongly normalizable. The aim is to show that every term is computable.

LEMMA 101.

- (a) If  $M \in \mathbf{C}$  and  $M \rightarrow_r N$ , then  $N \in \mathbf{C}$ .
- (b)  $\mathbf{C}$  is closed under repeated formation of application terms  $(M, N)$ .
- (c) If  $x \in V$ , then  $x \in \mathbf{C}$ .
- (d) If for every  $N^A \in \mathbf{C}$ ,  $(M^{(A \triangleright B)}, N) \in \mathbf{C}$ , then  $M \in \mathbf{C}$ .
- (e) If  $(M^{A \wedge B})_0 \in \mathbf{C}$  and  $(M^{A \wedge B})_1 \in \mathbf{C}$ , then  $M \in \mathbf{C}$ .
- (f) If  $N_1, N_2 \in \mathbf{C}$ , and  $G \in \mathbf{C}$ , for every  $G$  such that  $\langle N_1, N_2 \rangle \rightarrow_r G$ , then  $\langle N_1, N_2 \rangle \in \mathbf{C}$ .
- (g) If  $N \in \mathbf{C}$  and  $G \in \mathbf{C}$ , for all  $G$  such that  $(N)_i \rightarrow_r G$ , then  $(N)_i \in \mathbf{C}$ .
- (h) If  $N \in \mathbf{C}$ , and  $G \in \mathbf{C}$ , for all  $G$  such that  $\mathcal{P}N \rightarrow_r G$ , then  $\mathcal{P}N \in \mathbf{C}$ .
- (i) If  $N \in \mathbf{C}$ , and  $G \in \mathbf{C}$ , for all  $G$  such that  $SN \rightarrow_r G$ , then  $SN \in \mathbf{C}$ .
- (j) If  $N \in \mathbf{C}$ , and  $G \in \mathbf{C}$ , for all  $G$  such that  $\cup N \rightarrow_r G$ , then  $\cup N \in \mathbf{C}$ .
- (k) If  $N \in \mathbf{C}$ , and  $G \in \mathbf{C}$ , for all  $G$  such that  $\cap N \rightarrow_r G$ , then  $\cap N \in \mathbf{C}$ .

**Proof.** (a): By induction on  $h(M)$ . (b) By reflexivity of  $\rightarrow_r$  and Clause 2 in the definition of  $\mathbf{C}$ . (c): By induction on  $A \in T$ . If  $A = B \triangleright C$ , the claim follows by (b). (d): If for every  $N^A \in \mathbf{C}$ ,  $(M, N) \in \mathbf{C}$ , then  $sn(M)$ , since by (c) and the assumption  $(M, x^B) \in \mathbf{C}$ . Now suppose  $M \rightarrow_r N_1, N_2 \in \mathbf{C}$ , and for every  $N$ ,  $(M, N) \in \mathbf{C}$ . Then  $(M, N_2) \rightarrow (N_1, N_2)$  and, by (a),  $(N_1, N_2) \in \mathbf{C}$ . Thus  $M \in \mathbf{C}$ . (e): Since  $sn((M)_i)$ , also  $sn(M)$ . Suppose  $M \rightarrow_r \langle N_0, N_1 \rangle$ . Then  $(M)_i \rightarrow_r ((N_0, N_1))_i \rightarrow_r N_i$ . Since  $(M)_i \in \mathbf{C}$  and  $\mathbf{C}$  is closed under  $\rightarrow_r$ , also  $N_i \in \mathbf{C}$ . (f): Obviously, for every  $M$ ,  $sn(M)$  iff  $sn(N)$ , for each  $N$  such that  $M \rightarrow_r N$ . Moreover, suppose that  $\langle N_1, N_2 \rangle \rightarrow_r \langle G_1, G_2 \rangle$ . This is the case iff  $\langle N_1, N_2 \rangle \equiv \langle G_1, G_2 \rangle$  or there is a term  $M^*$  such that  $\langle N_1, N_2 \rangle \rightarrow_r M^*$ , and  $M^* \rightarrow_r \langle G_1, G_2 \rangle$ . In both cases  $G_1, G_2 \in \mathbf{C}$ . (g): By induction on the type  $A$  of  $(N)_i$ . If  $A$  is atomic, Clauses 2–5 in the definition of  $\mathbf{C}$  hold trivially.  $A = \langle P \rangle B$ : Suppose  $(N)_i \rightarrow_r SM$ . If  $N \equiv \langle M_1, M_2 \rangle$ , then  $(N)_i \rightarrow_r M_i$ , and  $M_i \in \mathbf{C}$ . If  $SM \not\equiv M_i$ , then  $M_i \rightarrow_r SM$ , and  $SM \in \mathbf{C}$ , by closure of  $\mathbf{C}$  under  $\rightarrow_r$ . If  $N \not\equiv \langle M_1, M_2 \rangle$ , then there is a term  $M^* \in \mathbf{C}$  such that  $(N)_i \rightarrow_r M^*$  and  $M^* \rightarrow_r SM$ . In each subcase,  $M \in \mathbf{C}$ . The cases  $A = [F]B$  and  $A = B \wedge C$  are analogous. If  $A = B \triangleright C$ , we may use closure of  $\mathbf{C}$  under application. (h): Suppose  $\mathcal{P}N \rightarrow_r \mathcal{P}G$ . This holds iff  $N \equiv G$  or there is a term  $M^*$  such that  $\mathcal{P}N \rightarrow_r M^*$  and  $M^* \rightarrow_r \mathcal{P}G$ . In both cases  $G \in \mathbf{C}$ . (i): Analogous to (h). (j): By induction

on the type  $A$  of  $\cup N$ . The only interesting case is  $A = [F]B$ . Suppose  $\cup N \rightarrow_r \mathcal{P}M$ . If  $N \equiv \mathcal{P}N_1$ , then  $\cup N \rightarrow_r N_1$  and  $N_1 \in \mathbf{C}$ . If  $\mathcal{P}M \not\equiv \mathcal{P}N_1$ , then  $N_1 \rightarrow_r \mathcal{P}M$ , and  $\mathcal{P}M \in \mathbf{C}$ . In each case  $M \in \mathbf{C}$ . (k): Analogous to (j). ■

**THEOREM 102.** *If  $M \in \text{Term}$  is  $\lambda$ -free, then  $M \in \mathbf{C}$ .*

**Proof.** By induction on  $M$ . (1):  $M$  is a variable: Lemma 101 (c). (2)  $M \equiv \langle N_1, N_2 \rangle$ : Lemma 101 (b) and the induction hypothesis. (3)  $M = \langle N_1^A, N_2^B \rangle$ : In view of Lemma 101 (f), it is enough to show that  $G \in \mathbf{C}$ , for every  $G$  such that  $\langle N_1, N_2 \rangle \rightarrow_r G$ . There are two subcases. (i):  $N_1 \equiv (G)_0$  and  $N_2 \equiv (G)_1$ . Then the claim follows by (e). (ii):  $G \equiv \langle N_1, N^* \rangle$  and  $N_2 \rightarrow_r N^*$  or  $G \equiv \langle M^*, N_2 \rangle$  and  $N_1 \rightarrow_r M^*$ . We may use induction on  $h(N_1) + h(N_2)$ . (4)  $M \equiv (N)_i$ . In view of Lemma 101 (g), it is enough to show that  $G \in \mathbf{C}$ , for every  $G$  such that  $M \rightarrow_r G$ . There are two cases. (i)  $N \equiv \langle N_0, N_1 \rangle$  and  $G \equiv N_i$ . Then we may use the induction hypothesis. (ii)  $G \equiv (N^*)_i$ ,  $N \rightarrow_r N^*$ , and we may use induction on  $h(N)$ . (5)  $M \equiv \mathcal{P}N$ : In view of Lemma 101 (h), it is enough to show that  $G \in \mathbf{C}$ , for every  $G$  such that  $M \rightarrow_r G$ . If  $M \rightarrow_r G$ , then  $G \equiv \mathcal{P}N^*$ ,  $N \rightarrow_r N^*$ , and we may use induction on  $h(N)$ . (6)  $M \equiv \mathcal{S}N$ : Analogous to (5), using Lemma 101 (i). (7)  $M \equiv \cap N$ : Given Lemma 101 (k), it suffices to show that  $G \in \mathbf{C}$ , for every  $G$  such that  $M \rightarrow_r G$ . There are two cases. (i)  $N \equiv \mathcal{S}G_1$  and  $G \equiv G_1$ . Then we may use the induction hypothesis. (ii)  $G \equiv \cap N^*$ ,  $N \rightarrow_r N^*$ , and we may use induction on  $h(N)$ . (8)  $M \equiv \cup N$ : Analogous to (7), using Lemma 101 (j). ■

Strong normalizability of all terms is derived from computability of all terms under substitution.

**DEFINITION 103.**  $M^A \in \text{Term}$  is said to be computable under substitution iff any substitution of free variables in  $M$  by computable terms of suitable type results in a computable term.

Let  $\mathbf{C}^s$  denote the set of all terms computable under substitution.

**THEOREM 104.** *Every  $\lambda_t$ -term  $M$  is computable under substitution.*

**Proof.** By induction on  $M$ . For term variables the claim is obvious. Moreover, since  $\mathbf{C}$  is closed under application,  $\mathbf{C}^s$  is also closed under application. If  $M \equiv \langle N_1, N_2 \rangle$ ,  $M \equiv (N)_i$ ,  $M \equiv \mathcal{P}N$ , or  $M \equiv \mathcal{S}N$ , the claim follows by the induction hypothesis. If  $M \equiv \lambda x^A N$ , it must be show that  $\lambda x N \in \mathbf{C}^s$  if  $N \in \mathbf{C}^s$ . Suppose that  $\lambda x N^*$  is the result of substituting a computable term for a free variable in  $\lambda x N$ , and suppose that  $G^A$  is a computable term such that  $(M, G)$  does not have a type  $B \triangleright C$ . Then, by Lemma 101 (f) – (k),  $((\lambda x N^*)G) \in \mathbf{C}$ , if for every term  $H$ ,  $((\lambda x N^*)G) \rightarrow_r H$  implies  $H \in \mathbf{C}$ . Since by assumption  $N \in \mathbf{C}^s$ , we have  $N^* \in \mathbf{C}$ . Therefore we may use induction on  $h(N^*) + h(G)$  to show that  $((\lambda x N^*)G) \in \mathbf{C}$ . There are three

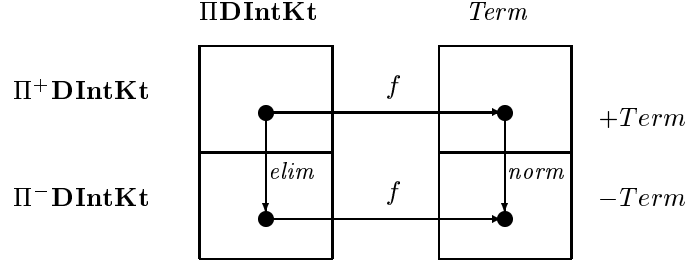


Figure 2. Normalization as a homomorphic image of proof-simplification.

subcases. (i)  $H \equiv N^*[x := G]$  and  $x \in fv(N^*)$ . Then  $N^* \in C^*$  implies  $H \in C$ . (ii)  $H \equiv N^*[x := G]$  and  $x \notin fv(N^*)$ . Then  $H \equiv N^* \in C$ . (iii)  $H$  is obtained from  $((\lambda x N^*)G)$  by executing one reduction step either in  $N^*$  or  $G$ . In this case we may use the induction hypothesis. ■

COROLLARY 105. *If  $M$  is a  $\lambda_t$ -term, then  $M$  is strongly normalizable.*

#### 5.4 Appendix D

It has to be shown that  $f$  is a homomorphism from  $\mathcal{A}$  to  $\mathcal{B}$ , i.e., for every  $\Pi \in \Pi^+ \mathbf{DIntKt}$ , we have  $f(elim(\Pi)) = norm(f(\Pi))$ , see Figure 2. The proof is by induction on  $\Pi$ . If the rule applied to obtain the conclusion sequent  $s_c$  of  $\Pi$  is an axiomatic sequent  $A \rightarrow A$ , then  $f(elim(\Pi)) = f(\Pi)$ , and  $f(\Pi)$  is a  $nf$ . If the rule applied to obtain  $s_c$  is such that the term construction step associated with it cannot generate a redex, we may apply the induction hypothesis. We shall consider the remaining cases.

$$\text{Case 1. } \Pi = \frac{\Pi' \quad A \times B \rightarrow X}{A \wedge B \rightarrow X}$$

A redex could be generated if the free variables  $x^A, y^B$  in the construction of  $A \times B \rightarrow X$  occur in the context  $\langle x, y \rangle$ . But then  $X = A \wedge B, A \times B \rightarrow X$  has been derived from  $\{A \rightarrow A, B \rightarrow B\}$ , and  $elim(\Pi) = A \wedge B \rightarrow A \wedge B$ . The claim holds, since  $\langle (v_1^{A \wedge B})_0, (v_1^{A \wedge B})_1 \rangle \rightarrow_r v_1^{A \wedge B}$ .

$$\text{Case 2. } \Pi = \frac{\Pi' \quad X \rightarrow A \times B}{X \rightarrow A \triangleright B}$$

A redex could be generated if the free variable  $x^A$  in the construction of  $X \rightarrow A \times B$  occurs in the context  $(N^{A \triangleright B}, x^A)$ . But then  $X = A \triangleright B$ ,

$X \rightarrow A \times B$  has been derived from  $\{A \rightarrow A, B \rightarrow B\}$ , and  $elim(\Pi) = A \triangleright B \rightarrow A \triangleright B$ . The claim holds, since  $\lambda v_1^A(v_1^{A \triangleright B}, v_1^A) \rightarrow_r v_1^{A \triangleright B}$ .

$$\text{Case 3. } \Pi = \frac{\frac{\Pi_1 \quad \Pi_2}{X \rightarrow A \quad A \rightarrow Y}}{X \rightarrow Y}$$

Suppose the exhibited application of cut in  $\Pi$  is not principal. If this application is reduced in one step, either the  $f$ -images of the resulting proof and  $\Pi$  are the same, or some principal cuts have been performed on subformulas of  $A$ . Thus, there are five remaining cases to be considered.

Case 3.1 ( $t$ ):

$$\begin{array}{ccc} \frac{\frac{\Pi}{\mathbf{I} \rightarrow X} \quad \frac{\mathbf{I} \rightarrow X \quad \mathbf{I} \rightarrow X}{t \rightarrow X}}{\mathbf{I} \rightarrow X} & \text{is converted into} & \frac{\Pi}{\mathbf{I} \rightarrow X} \\ \downarrow f & & \downarrow f \\ \frac{v_1^t \quad M}{M} & \rightarrow_r & M \end{array}$$

Case 3.2 ( $\wedge$ ):

$$\begin{array}{ccc} \frac{\frac{\Pi_1 \quad \Pi_2 \quad \Pi_3}{X \rightarrow A \quad Y \rightarrow B \quad A \times B \rightarrow Z} \quad \frac{A \wedge B \rightarrow Z}{A \wedge B \rightarrow Z}}{X \times Y \rightarrow Z} & \text{is conv. into} & \frac{\frac{\Pi_1 \quad \Pi_2 \quad \Pi_3}{X \rightarrow A \quad A \times B \rightarrow Z} \quad \frac{X \rightarrow B \times Z}{X \times B \rightarrow Z} \quad \frac{Y \rightarrow B \quad B \rightarrow X \times Z}{B \rightarrow X \times Z}}{Y \rightarrow X \times Z} \\ \downarrow f & & \downarrow f \\ \frac{M_1^A \quad M_2^B \quad N(x^A, y^B)}{\langle M_1, M_2 \rangle \quad N((z^{A \wedge B})_0, (z^{A \wedge B})_1)} & \rightarrow_r & \frac{M_1^A \quad \frac{N(x^A, y^B)}{N}}{N(M_1)} \\ & & \frac{M_2^B \quad N(M_1)}{N(M_1, M_2)} \\ & & N(M_1, M_2) \end{array}$$

Case 3.3 ( $\triangleright$ ):

$$\begin{array}{ccc}
\frac{\frac{\Pi_1}{X \rightarrow A \times B} \quad \frac{\Pi_2 \quad \Pi_3}{Y \rightarrow A \quad B \rightarrow Z}}{X \rightarrow A \triangleright B} & \text{is conv. into} & \frac{\frac{\Pi_1}{X \rightarrow A \times B} \quad \frac{\Pi_3}{B \rightarrow Z}}{X \times A \rightarrow B} \\
\frac{\Pi_2}{A \triangleright B \rightarrow Y \times Z} & & \frac{\Pi_2}{Y \rightarrow A} \\
\hline
X \rightarrow Y \times Z & & \frac{X \times A \rightarrow Z}{A \rightarrow X \times Z} \\
& & \frac{Y \rightarrow X \times Z}{X \times Y \rightarrow Z} \\
& & \frac{X \times Y \rightarrow Z}{X \rightarrow Y \times Z} \\
\downarrow f & & \downarrow f \\
\frac{\frac{M^B(x^A)}{\lambda x^A M} \quad \frac{N_1^A \quad N_2(y^B)}{N_2(z^A \triangleright B), N_1}}{N_2(\lambda x^A M, N_1)} & \rightarrow_r & \frac{\frac{M^B(x^A)}{M^B} \quad \frac{N_2(y^B)}{N_2(M)}}{N_2(M)} \\
& & \frac{N_1^A \quad \frac{N_2(M(x^A))}{N_2(M(N_1))}}{N_2(M(N_1))} \\
& & \frac{N_2(M(N_1))}{N_2(M(N_1))}
\end{array}$$

Case 3.4 ( $[F]$ ):

$$\begin{array}{ccc}
\frac{\frac{\Pi_1}{\bullet X \rightarrow A} \quad \frac{\Pi_2}{A \rightarrow Y}}{X \rightarrow [F]A} & \text{is converted into} & \frac{\frac{\Pi_1}{\bullet X \rightarrow A} \quad \frac{\Pi_2}{A \rightarrow Y}}{\frac{\bullet X \rightarrow Y}{X \rightarrow \bullet Y}} \\
\frac{\bullet X \rightarrow \bullet Y}{X \rightarrow \bullet Y} & & \\
\downarrow f & & \downarrow f \\
\frac{\frac{M^A}{\mathcal{P}M} \quad \frac{N(x^A)}{N(\cup y[F]A)}}{N(\cup \mathcal{P}M)} & \rightarrow_r & \frac{\frac{M^A}{N(M)} \quad \frac{N(x^A)}{N(M)}}{N(M)}
\end{array}$$

Case 3.5 ( $\langle P \rangle$ ): analogous to the previous case. ■

## ACKNOWLEDGEMENT

I would like to thank Dov Gabbay for inviting me to contribute this chapter to the Second Edition of the Handbook of Philosophical Logic.

*Dresden University of Technology, Germany.*

## BIBLIOGRAPHY

- [D'Agostino and Mondadori, 1994] M. D'Agostino and M. Mondadori, The Taming of the Cut. Classical Refutations with Analytic Cut, *Journal of Logic and Computation* 4 (1994), 285–319.



- [Andréka *et al.*, 1998] H. Andréka, I. Németi and J. van Benthem, Modal Languages and Bounded Fragments of Predicate Logic, *Journal of Philosophical Logic* 27 (1998), 217–274.
- [Avron, 1984] A. Avron, On modal systems having arithmetical interpretations, *Journal of Symbolic Logic* 49 (1984), 935–942.
- [Avron, 1991] A. Avron, Using Hypersequents in Proof Systems for Non-classical Logics, *Annals of Mathematics and Artificial Intelligence* 4 (1991), 225–248.
- [Avron, 1991a] A. Avron, Natural 3-valued Logics—Characterization and Proof Theory, *Journal of Symbolic Logic*, (56) 1991, 276–294.
- [Avron, 1996] A. Avron, The Method of Hypersequents in Proof Theory of Propositional Non-Classical Logics, in: W. Hodges *et al.* (eds.), *Logic: From Foundations to Applications*, Oxford University Press, Oxford, 1996, 1–32.
- [Belnap, 1982] N.D. Belnap, Display Logic, *Journal of Philosophical Logic* 11 (1982), 375–417. Reprinted with minor changes as §62 of A.R. Anderson, N.D. Belnap, and J.M. Dunn, *Entailment: the logic of relevance and necessity*. Vol. 2, Princeton University Press, Princeton, 1992.
- [Belnap, 1990] N.D. Belnap, Linear Logic Displayed, *Notre Dame Journal of Formal Logic* 31 (1990), 14–25.
- [Belnap, 1996] N.D. Belnap, The Display Problem, in: H. Wansing (ed.), *Proof Theory of Modal Logic*, Kluwer Academic Publishers, Dordrecht, 1996, 79–92.
- [van Benthem, 1986] J. van Benthem, *Essays in Logical Semantics*, Kluwer Academic Publishers, Dordrecht, 1986.
- [van Benthem, 1991] J. van Benthem, *Language in Action*. North-Holland, Amsterdam, 1991.
- [van Benthem, 1996] J. van Benthem, *Exploring Logical Dynamics*, CSLI Publications, Stanford, 1996.
- [Blamey and Humberstone, 1991] S. Blamey and L. Humberstone, A Perspective on Modal Sequent Logic, *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 27 (1991), 763–782.
- [Boolos, 1984] G. Boolos, Don't eliminate cut, *Journal of Philosophical Logic* 13 (1984), 373–378.
- [Borghuis, 1993] T. Borghuis, Interpreting modal natural deduction in type theory, in: M. de Rijke (ed.), *Diamonds and Defaults*, Kluwer Academic Publishers, Dordrecht, 1993, 67–102.
- [Borghuis, 1994] T. Borghuis, *Coming to Terms with Modal Logic: On the interpretation of modalities in typed  $\lambda$ -calculus*, PhD thesis, Department of Computer Science, University of Eindhoven, 1994.
- [Borghuis, 1998] T. Borghuis, Modal Pure Type Systems, *Journal of Logic, Language and Information* 7 (1998), 265–296.
- [Bošić and Došen, 1984] M. Bošić and K. Došen, Models for normal intuitionistic modal logics, *Studia Logica* 43 (1984), 217–245.
- [Braüner, 2000] T. Braüner, A Cut-Free Gentzen Formulation of the Modal Logic **S5**, *Logic Journal of the IGPL* 8 (2000), 629–643.
- [Bull and Segerberg, 1984] R. Bull and K. Segerberg, Basic Modal Logic. In: D. Gabbay and F. Guenther (eds), *Handbook of Philosophical Logic*, Vol. II, *Extensions of Classical Logic*, Reidel, Dordrecht, 1984, 1–88.
- [Cerrato, 1993] C. Cerrato, Modal sequents for normal modal logics, *Mathematical Logic Quarterly* 39 (1993), 231–240.
- [Cerrato, 1996] C. Cerrato, Modal sequents, in: H. Wansing (ed), *Proof Theory of Modal Logic*, Kluwer Academic Publishers, Dordrecht, 1996, 141–166.
- [Chellas, 1980] B. Chellas, *Modal Logic: An Introduction*, Cambridge University Press, Cambridge, 1980.
- [Davis and Pfenning, 2000] R. Davies and F. Pfenning, A Modal Analysis of Staged Computation, 2000, to appear in: *Journal of the ACM*.
- [Demri and Goré, 1999] S. Demri and R. Goré, Cut-free display calculi for nominal tense logics. In *Proc Tableaux '99*, pp. 155–170. Lecture Notes in AI, Springer-Verlag, Berlin, 1999.

- [Demri and Goré, 2000] S. Demri and R. Goré, Display calculi for logics with relative accessibility relations, *Journal of Logic, Language and Information* 9 (2000), 213–236.
- [Došen, 1985] K. Došen, Sequent-systems for modal logic, *Journal of Symbolic Logic* 50 (1985), 149–159.
- [Došen, 1988] K. Došen, Sequent systems and groupoid models I, *Studia Logica* 47 (1988), 353–389.
- [Dragalin, 1988] A. Dragalin, *Mathematical Intuitionism. Introduction to Proof Theory*, American Mathematical Society, Providence, 1988.
- [Dunn, 1990] J. M. Dunn, Gaggles Theory: An Abstraction of Galois Connections and Residuation with Applications to Negation and Various Logical Operations, in: J. van Eijk (ed.), *Logics in AI, Proc. European Workshop JELIA 1990*, Lecture Notes in Computer Science 478, Springer-Verlag, Berlin, 1990, 31–51.
- [Dunn, 1993] J.M. Dunn, Partial-Gaggles Applied to Logics with Restricted Structural Rules, in: P. Schroeder-Heister and K. Došen (eds.), *Substructural Logics*, Clarendon Press, Oxford, 1993, 63–108.
- [Dunn, 1995] J.M. Dunn, Gaggles Theory Applied to Modal, Intuitionistic, and Relevance Logics, in: I. Max and W. Stelzner (eds.), *Logik und Mathematik: Frege-Kolloquium 1993*, de Gruyter, Berlin, 1995, 335–368.
- [Fitting, 1993] M. Fitting, Basic Modal Logic, in: D. Gabbay et al. (eds), *Handbook of Logic in Artificial Intelligence and Logic Programming, Vol. 1, Logical Foundations*, Oxford UP, Oxford, 1993, 365–448.
- [Friedman, 1975] H. Friedman, Equality between functionals, in: R. Parikh (ed.), *Logic Colloquium Boston 1972–73*, Springer Lecture Notes in Mathematics Vol. 453, Springer-Verlag, Berlin, 1975, 22–37.
- [Gabbay, 1996] D. Gabbay, *Labelled Deductive Systems: Volume 1. Foundations*, Oxford University Press, Oxford, 1996.
- [Gabbay and de Quieroz, 1992] D. Gabbay and R. de Quieroz, Extending the Curry-Howard interpretation to linear, relevant and other resource logics, *Journal of Symbolic Logic* 57 (1992), 1319–1365.
- [Gentzen, 1934] G. Gentzen. Investigations into Logical Deduction, in: M. E. Szabo (ed.), *The Collected Papers of Gerhard Gentzen*, North Holland, Amsterdam, 1969, 68–131. English translation of: Untersuchungen über das logische Schließen, *Mathematische Zeitschrift* 39 (1934), I 176–210, II 405–431.
- [Girard, 1989] J.-Y. Girard, Y. Lafont, and P. Taylor, *Proofs and Types*, Cambridge University Press, Cambridge, 1989.
- [Goble, 1974] L. Goble, Gentzen systems for modal logics, *Notre Dame Journal of Formal Logic* 15 (1974), 455–461.
- [Gödel, 1933] K. Gödel, Eine Interpretation des intuitionistischen Aussagenkalküls, *Ergebnisse eines mathematischen Kolloquiums* 4 (1933), 39–40. Reprinted and translated in: S. Feferman et al. (eds.), *Kurt Gödel. Collected Works. Vol. 1*, Oxford University Press, Oxford, 1986, 300–303.
- [Goldblatt, 1992] R. Goldblatt, *Logics of Time and Computation*, CSLI Lecture Notes 7, Stanford, CSLI Publications, 2nd revised and expanded edition, 1992.
- [Goré, 1992] R. Goré., *Cut-Free Sequent and Tableau Systems for Propositional Normal Modal Logics*, PhD thesis, University of Cambridge Computer Laboratory, Technical Report No. 257, 1992.
- [Goré, 1995] R. Goré, Intuitionistic Logic Redisplayed, Technical Report TR-ARP-1-1995, Australian National University, 1995.
- [Goré, 1998] R. Goré, Substructural Logics on Display, *Logic Journal of the IGPL* 6 (1998), 451–504.
- [Goré, 1998] R. Goré, Gaggles, Gentzen and Galois: How to display your favourite substructural logic, *Logic Journal of the IGPL* 6 (1998), 669–694.
- [Goré, 1999] R. Goré, Tableau Methods for Modal and Temporal Logics, in: M. D’Agostino, D. Gabbay, R. Hähnle, and J. Posegga (eds), *Handbook of Tableau Methods*, Kluwer Academic Publishers, Dordrecht, 1999, 297–396.
- [Goré, 2000] R. Goré, Dual Intuitionistic Logic Revisited, in: R. Dyckhoff (ed.), *Proceedings Tableaux 2000*, LNAI 1847, Springer-Verlag Berlin, 2000, 252–267.
- [Gottwald, 1989] S. Gottwald, *Mehrwertige Logik*, Akademie-Verlag, Berlin, 1989.

- [Hacking, 1994] I. Hacking, What is Logic?, *The Journal of Philosophy* 76 (1979), 285–319. Reprinted in: D. Gabbay (ed.), *What is a Logical System?*, Oxford University Press, Oxford, 1994, 1–33. (Cited after the reprint.)
- [Helman, 1977] G. Helman, *Restricted Lambda-abstraction and the Interpretation of Some Non-classical Logics*, PhD thesis, Department of Philosophy, University of Pittsburgh, 1977.
- [Hindley and Seldin, 1986] J.R. Hindley and J.P. Seldin, *Introduction to Combinators and  $\lambda$ -Calculus*, Cambridge UP, Cambridge, 1986.
- [Howard, 1980] W.A. Howard, The formulae-as-types notion of construction, in: J.R. Hindley and J.P. Seldin (eds.), *To H.B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, Academic Press, London, 1980, 479–490.
- [Indrzejczak, 1997] A. Indrzejczak, Generalised Sequent Calculus for Propositional Modal Logics, *Logica Trianguli* 1 (1997), 15–31.
- [Indrzejczak, 1998] A. Indrzejczak, Cut-free Double Sequent Calculus for S5, *Logic Journal of the IGPL* 6 (1998), 505–516.
- [Kashima, 1994] R. Kashima, Cut-free sequent calculi for some tense logics, *Studia Logica* 53 (1994), 119–135.
- [Kracht, 1996] M. Kracht, Power and Weakness of the Modal Display Calculus, in: H. Wansing (ed.), *Proof Theory of Modal Logic*, Kluwer Academic Publishers, Dordrecht, 1996, 93–121.
- [Kripke, 1963] S. Kripke, Semantical analysis of modal logic I: Normal modal propositional calculi, *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 11 (1968), 3–16.
- [von Kutschera, 1968] F. von Kutschera, Die Vollständigkeit des Operatorensystems  $\{\neg, \wedge, \vee, \supset\}$  für die intuitionistische Aussagenlogik im Rahmen der Gentzensemantik, *Archiv für Mathematische Logik und Grundlagenforschung*, 11 (1968), 3–16.
- [Lavendhomme and Lucas, 2000] R. Lavendhomme and T. Lucas, Sequent calculi and decision procedures for weak modal systems, *Studia Logica* 65 (2000), 121–145.
- [Leivant, 1981] D. Leivant, On the proof theory of the modal logic for arithmetic provability, *Journal of Symbolic Logic* 46 (1981), 531–538.
- [Lukowski, 1996] P. Lukowski, Modal interpretation of Heyting-Brouwer Logic, *Bulletin of the Section of Logic* 25 (1996), 80–83.
- [Martini and Masini, 1996] A. Martini and A. Masini, A Computational Interpretation of Modal Proofs, in: H. Wansing (ed.), *Proof Theory of Modal Logic*, Kluwer Academic Publishers, Dordrecht, 1996, 213–241.
- [Masini, 1992] A. Masini, 2-Sequent calculus: a proof theory of modalities, *Annals of Pure and Applied Logic* 58 (1992), 229–246, 1992.
- [Mints, 1970] G. Mints, Cut-free calculi of the S5 type, *Studies in constructive mathematics and mathematical logic. Part II. Seminars in Mathematics* 8 (1970), 79–82.
- [Mints, 1990] G. Mints, Gentzen-type systems and resolution rules. Part I. Propositional Logic, in: P. Martin-Löf and G. Mints (eds.), *COLOG-88*, Lecture Notes in Computer Science 417, Springer-Verlag, Berlin, 198–231, 1990.
- [Mints, 1992] G. Mints, *A Short Introduction to Modal Logic*, CSLI Lecture Notes 30, CSLI Publications, Stanford, 1992.
- [Mints, 1997] G. Mints, Indexed systems of sequents and cut-elimination. *Journal of Philosophical Logic* 26 (1997), 671–696.
- [Nishimura, 1980] H. Nishimura, A Study of Some Tense Logics by Gentzen's Sequential Method; *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 16 (1980), 343–353.
- [Ohnishi and Matsumoto, 1957] M. Ohnishi and K. Matsumoto, Gentzen Method in Modal Calculi, *Osaka Mathematical Journal* 9 (1957), 113–130.
- [Ohnishi and Matsumoto, 1959] M. Ohnishi and K. Matsumoto, Gentzen Method in Modal Calculi, II, *Osaka Mathematical Journal* 11 (1959), 115–120.
- [Ohnishi, 1982] M. Ohnishi, A New Version to Gentzen Decision Procedure for Modal Sentential Calculus S5, *Mathematical Seminar Notes* 10 (1982), Kobe University, 161–170.
- [Ono, 1998] H. Ono, Proof-Theoretic Methods in Nonclassical Logic — an Introduction, *MSJ Memoirs* 2, Mathematical Society of Japan, 1998, 207–254.

- [Orlowska, 1988] E. Orlowska, Relational interpretation of modal logics, in: H. Andreka, D. Monk and I. Nemeti (eds.), *Algebraic Logic. Colloquia Mathematica Societatis Janos Bolyai* 54, North Holland, Amsterdam, 443–471, 1988.
- [Orlowska, 1996] E. Orlowska, Relational Proof Systems for Modal Logics, in: H. Wansing (ed.), *Proof Theory of Modal Logic*, Kluwer Academic Publishers, Dordrecht, 55–77, 1996.
- [Pfenning, 2000] F. Pfenning and R. Davies, A Judgmental Reconstruction of Modal Logic, Department of Computer Science, Carnegie Mellon University, Pittsburgh, 2000.
- [Pliuškevičienė, 1998] A. Pliuškevičienė, Cut-free Calculus for Modal Logics Containing the Barcan Axiom, in: M. Kracht et al. (eds.), *Advances in Modal Logic '96*, CSLI Publications, Stanford, 1998, 157–172.
- [Pottinger, 1983] G. Pottinger, Uniform, cut-free formulations of  $T$ ,  $S4$  and  $S5$  (Abstract), *Journal of Symbolic Logic* 48 (1983), 900–901.
- [de Queiroz and Gabbay, 1997] R. de Queiroz and D. Gabbay, The functional interpretation of modal necessity, in: M. de Rijke (ed.), *Advances in Intensional Logic*, Kluwer Academic Publishers, Dordrecht, 1997, 61–91.
- [de Queiroz and Gabbay, 1999] R. de Queiroz and D. Gabbay, An introduction to labelled natural deduction, in: H.J. Ohlbach and U. Reyle (eds.), *Logic Language and Reasoning. Essays in Honour of Dov Gabbay*, Kluwer Academic Publishers, Dordrecht, 1999.
- [Rauszer, 1980] C. Rauszer, *An algebraic and Kripke-style approach to a certain extension of intuitionistic logic*, *Dissertationes Mathematicae*, vol. CLXVII, Warsaw, 1980.
- [Restall, 1995] G. Restall, Display Logic and Gaggle Theory, *Reports on Mathematical Logic* 29 (1995), 133–146.
- [Restall, 1998] G. Restall, Displaying and Deciding Substructural Logics 1: Logics with Contraposition, *Journal of Philosophical Logic* 27 (1998), 179–216.
- [Restall, 1999] G. Restall, *An Introduction to Substructural Logics*, Routledge, London, 1999.
- [Roorda, 1991] D. Roorda, *Resource Logics*. PhD thesis, Institute for Logic, Language and Computation, University of Amsterdam, 1991.
- [Sambin and Valentini, 1982] G. Sambin and S. Valentini, The modal logic of provability. The sequential approach, *Journal of Philosophical Logic* 11 (1982), 311–342.
- [Sasaki, 1999] K. Sasaki, On intuitionistic modal logic corresponding to extended typed  $\lambda$ -calculus for partial functions, Manuscript, Department of Computer Science, Leipzig University, 1999.
- [Sato, 1977] M. Sato. A Study of Kripke-type Models for Some Modal Logics by Gentzen's Sequential Method. *Publications of the Research Institute for Mathematical Sciences, Kyoto University* 13 (1977), 381–468.
- [Sato, 1980] M. Sato, A cut-free Gentzen-type system for the modal logic  $S5$ , *Journal of Symbolic Logic* 45 (1980), 67–84.
- [Schröeter, 1955] K. Schröeter, Methoden zur Axiomatisierung beliebiger Aussagen- und Prädikatenkalküle, *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 1 (1955), 214–251.
- [Schütte, 1968] K. Schütte, *Vollständige Systeme modaler und intuitionistischer Logik*, Springer-Verlag, Berlin, 1968.
- [Shimura, 1991] T. Shimura, Cut-Free Systems for the Modal Logic  $S4.3$  and  $S4.3Grz$ , *Reports on Mathematical Logic* 25 (1991), 57–73.
- [Shvarts, 1989] G. Shvarts, Gentzen style systems for  $K45$  and  $K45D$ , in: A. Meyer and M. Taitslin (eds.), *Logic at Botik '89*, Lecture Notes in Computer Science 363, Springer-Verlag, Berlin, 1989, 245–256.
- [Smullyan, 1968] R. Smullyan, Analytic cut, *Journal of Symbolic Logic* 33 (1968), 560–564.
- [Takano, 1992] M. Takano, Subformula property as a substitute for cut-elimination in modal propositional logics, *Mathematica Japonica* 37 (1992), 1129–1145.

- [Tennant, 1994] N. Tennant, The Transmission of Truth and the Transitivity of Deduction, in: D. Gabbay (ed.), *What is a Logical System?*, Oxford University Press, Oxford, 1994, 161–178.
- [Troelstra, 1992] A. Troelstra. *Lectures on Linear Logic*, CSLI Lecture Notes 29, CSLI Publications, Stanford, 1992.
- [Troelstra and Schwichtenberg, 2000] A. Troelstra and H. Schwichtenberg, *Basic Proof Theory*, Cambridge Tracts in Theoretical Computer Science 43, Second Edition, Cambridge University Press, Cambridge, 2000.
- [de Vrijer, 1987] R. de Vrijer, Strong normalization in  $N - HA_p^\omega$ . *Proceedings of the Koninklijke Nederlandse Akademie van Wetenschappen* 90 (1987), 473–478.
- [Wansing, 1992] H. Wansing, Formulas-as-types for a hierarchy of sublogics of intuitionistic propositional logic, in: H. Wansing and D. Pearce (eds.), *Nonclassical Logics and Information Processing*, Springer Lecture Notes in Artificial Intelligence 619, Springer-Verlag, Berlin, 1992, 125–145.
- [Wansing, 1994] H. Wansing, Sequent systems for normal modal propositional logics, *Journal of Logic and Computation* 4 (1994), 125–142.
- [Wansing, 1998] H. Wansing, *Displaying Modal Logic*, Kluwer Academic Publishers, Dordrecht, 1998.
- [Wansing, 1999] H. Wansing, Displaying the modal logic of consistency, *Journal of Symbolic Logic* 64 (1999), 1573–1590.
- [Wansing, 1999] H. Wansing, Predicate Logics on Display, *Studia Logica* 62 (1999), 49–75.
- [Wansing, 2000] H. Wansing, Formulas-as-types for temporal logic, Report, Dresden University of Technology, Institute of Philosophy, 2000.
- [Wittgenstein, 1953] L. Wittgenstein, *Philosophical Investigations*, Blackwell, Oxford, 1953.
- [Wolter, 1998] F. Wolter, On Logics With Coimplication, *Journal of Philosophical Logic* 27 (1998), 353–387.
- [Wolter and Zakharyashev, 1999] F. Wolter and M. Zakharyashev. Intuitionistic Modal Logic, in: A. Cantini et al. (eds.), *Logic and Foundations of Mathematics*, Kluwer Academic Publishers, Dordrecht, 1999, 227–238.
- [Zach, 1993] R. Zach, *Proof Theory of Finite-valued Logics*, Diplomarbeit, Institut für Computersprachen, Technische Universität Wien, 1993.
- [Zeman, 1973] J.J. Zeman, *Modal Logic. The Lewis-Modal Systems*, Oxford University Press, Oxford, 1973.
- [Zucker and Tragesser, 1978] J. Zucker and R. Tragesser, The adequacy problem for inferential logic, *Journal of Philosophical Logic* 7 (1978), 501–516.



LENNART ÅQVIST

## DEONTIC LOGIC

### I. INTRODUCTION

#### 1 THE PROTAGORAS PARADOX: AN EXERCISE IN ELEMENTARY LOGIC FOR LAWYERS AND MORALISTS

An ancient paradox is about the famous Greek law teacher Protagoras and goes like this: Protagoras and Euathlus agree that the former is to instruct the latter in rhetoric and is to receive a certain fee which is to be paid *if and only if* Euathlus wins his first court-case (in some versions: *as soon as* he has won his first case). Well, Euathlus completed his course but did not take any law cases. Some time elapsed and Protagoras sued his student for the sum. The following arguments were presented to the judge in court.

*Protagoras*: If I win this case, then Euathlus has to pay me by virtue of your verdict. On the other hand, if he wins the case, then he will win his first case, hence he has to pay me, this time by virtue of our agreement. In either case, he has to pay me. Therefore, he is obliged to pay me my fee.

*Euathlus*: If I win this case, then, by your verdict, I don't have to pay. If, however, Protagoras wins the case, then I will not yet have won my first case, so, by our agreement, I don't have to pay. Hence I am not obliged to pay the fee.

Let us now raise two questions:

Who was right?

Could deontic logic, in the sense of the logical theory of norms and normative systems, be helpful in providing a solution to this problem, or kind of problem?

In this chapter we shall not attempt to answer the first question, but just refer the reader to the attempts made by Lenzen [1977], Smullyan [1978] and Åqvist [1981]. But we shall indeed argue for an affirmative answer to the second question, agreeing with the following statement made by Bertrand Russell in 'On Denoting':

A logical theory may be tested by its capacity for dealing with puzzles, and it is a whole-some plan, in thinking about logic, to stock the mind with as many puzzles as possible, since these serve much the same purpose as is served by experiments in physical science.

As we shall see in Section 9 below, however, our affirmative answer will have to be carefully qualified as a result of the examination we undertake

in Part II of a number of paradoxes and dilemmas that have beset recent developments in deontic logic in the last thirty years. First of all, though, let us have a somewhat closer look at the subject as it stands nowadays.

## 2 THE IMPORTANCE OF VON WRIGHT'S AND ANDERSON'S WORK

What is deontic logic? It is tempting to answer with Rescher [1966], *à propos* the closely related area of the logic of imperatives and commands, that it is a field with the property that there is virtually no single issue in it upon which a settled consensus has been reached. Resisting that temptation, though, we say that deontic logic, broadly conceived, is the logical study of the normative use of language and that its subject matter is a variety of normative concepts, notably those of *obligation* (prescription), *prohibition* (forbiddance), *permission* and *commitment*. The first one among these concepts is often expressed by such words as 'shall', 'ought' and 'must', the second by 'shall not', 'ought not' and 'must not', and the third one by 'may'; the fourth notion amounts to an idea of *conditional obligation*, expressible by 'if..., then it shall (ought, must) be the case that \_\_\_'.

A powerful trend of research in the area was initiated by the famous contribution of Von Wright [1951], where the formal properties of *monadic* ('unconditional', 'absolute') normative concepts were systematically explored. Certain paradoxical results were seen to arise in Von Wright's monadic deontic logic, however, which led him to propose systems for and permission *dyadic* ('conditional', 'relative') normative concepts, where the notions of obligation, permission etc. are made relative to, or conditional on, certain circumstances. Thus, the dyadic deontic logic of Von Wright [1956] was proposed as a reaction to the Prior [1954] paradoxes of commitment ('derived obligation'), and that of Von Wright [1964; 1965] as a reaction to the Chisholm [1963] contrary-to-duty imperative paradox. One major problem-area, with which we shall deal in this chapter, concerns the mathematical structure and interpretation of the Von Wright-type deontic logics just mentioned, whether they be of the monadic kind or the dyadic one.

In Anderson [1956] the author interestingly argued that the study of normative concepts undertaken by deontic logic could profit a good deal from our considering their behavior in the context of *normative systems*, like systems of ethics (moral theories) and systems of positive law. He then, naturally, noticed and emphasized the role played by *sanctions* or *penalties* in actual normative systems, and went on to *define* the deontic or normative notions of obligation, forbiddance etc. along the following lines: letting  $S$  be a constant proposition, describing a situation which will count as a penalty or sanction relatively to the normative system under investigation, we say that a state-of-affairs  $p$  is obligatory if and only if (iff) the absence



of  $p$  entails the sanction  $S$ ; that  $p$  is forbidden iff  $p$  (itself) entails the sanction  $S$ ; and that  $p$  is permitted iff it is possible that  $p$  obtains without the sanction  $S$  being realized. (For graphic pictures of essentially these definitions, in the style of traditional ‘squares of opposition’, see Anderson [1968] and Åqvist [1987, Section 5.3].) Anderson [1956; 1958] then sets out to add these definitions to various systems of alethic modal logic (i.e. the logic of ‘ordinary’ necessity and possibility), whereby he achieves a kind of *reduction of monadic* deontic logic to alethic modal logic, provided only that the alethic system is supplemented with the constant  $S$  and (possibly) with some axiom governing  $S$ . Anderson [1956; 1959] also suggests a definition in terms of  $S$  and alethic modal notions of the *dyadic* concept of commitment.

A second major problem-area which will occupy us in the present work, concerns the mathematical structure and interpretation of Anderson-style systems of alethic modal logic with a propositional constant added to their basic machinery. Also, we shall be highly interested in the relation of such Anderson-style systems to Von Wright-type deontic logics of the two sorts mentioned above.

Let me now briefly say something about the current state of the subject of deontic logic. I think it is only fair to claim with Von Wright [1977] (his introduction to the proceedings of the international and interdisciplinary Bielefeld Colloquium in March 1975 — presumably the first one of its kind) that the widespread and intense interest aroused by deontic logic indicates that we have to deal with a new logical discipline, which has come to stay and is not just ‘*eine vorübergehende Erscheinung*’. On the other hand, he points out, it is still a relatively poorly developed branch of exact research for the following reasons:

- (i) The number of open problems is very big.
- (ii) There is a good deal of controversy and disagreement about fundamental matters in the area, e.g. about the interpretation and the validity of its basic principles.
- (iii) The high expectations as to the applicability of deontic logic to actual and potential normative systems, notably in the areas of ethics and legal theory, can hardly be said to have been satisfied to more than a very slight and modest degree.

The energy and dedication with which Von Wright himself has, since the ‘fifties, labored to improve and refine on his originally proposed systems, bear out conclusions (i) and (ii) very clearly in my opinion. And there is no doubt that the optimism in regard to applications — so characteristic of Anderson — has turned into pessimism.

My third main concern in this chapter will then be to do something to remove that pessimism. Suppose we want to achieve a *logical* analysis [Oppenheim, 1944] or a *rational reconstruction* [Wedberg, 1951] of some system

of positive law, say, some relevant part of any existing commercial or criminal code. It is then clear that the languages of the current systems of deontic logic, which we will encounter below, are almost totally inadequate for the formulation of even very simple rules of the system. A main reason for this being so is that these languages are just propositional and thus lack quantificational resources of expression. What is even worse, they lack explicit *temporal* resources, which fact makes them especially useless from the legal point of view. In Section 9 below, we follow Van Eck [1981] in arguing that, for any serious purposes of application, the expressive resources of deontic languages must be enriched so as to include temporal and quantificational ones. If this is done, the hope of deontic logicians and others concerned to be able to contribute substantively to ethical and legal theory might be regained. Maybe also to a linguistically important branch such as speech act theory.

Let me close this introductory section with a cursory historical note. Suggestions about a logic of normative concepts and sentences (including one for imperatives and commands) may be found in Aristotle, in the Stoics (see Rescher [1966]), in medieval philosophers (see Knuuttila [1981]), in Leibniz, as well as in Bentham and his followers in legal philosophy (see Lindahl [1977]). The first systematic attempt to build a formal theory of normative concepts is due to Mally [1926] (for a nice exposition of Mally's *Deontik*, see Føllesdal and Hilpinen [1971], who also cover later twentieth-century developments in an exemplary way).

### 3 PLAN OF THIS CHAPTER

The present work can be divided into two. The first deals with certain much-discussed difficulties in connection with the application of formal systems of deontic logic to a natural language such as English (Part II: Paradoxes and Dilemmas), and the other half is devoted to the purely mathematical presentation and elaboration of a number of formal systems of deontic logic (Parts III–VI).

In Part III we deal with ten systems of monadic deontic logic, which go back to the work of Timothy J. Smiley in [1963], to that of William H. Hanson in [1965], and, in the three cases of  $\mathbf{OM}^+$ ,  $\mathbf{OS4}^+$  and  $\mathbf{OS5}^+$ , to the mixed alethic-deontic systems  $\mathbf{OM}$ ,  $\mathbf{OM}'$  and  $\mathbf{OM}''$  of Anderson [1956]; among other things, Smiley [1963] proved the three former systems to be identical with the deontic fragments of the three latter ones. The ten Smiley–Hanson systems, as I pertinently call them, are studied both from a *proof-theoretical* or *axiomatic* point of view (Section 10.2) and from that of *model-theoretical* ‘possible worlds’ *semantics* in a sense deriving from Hintikka [1957], Kanger [1957], Montague [1960] and, above all, Kripke [1963] (Section 10.3). In Section 11 we prove the semantic soundness and

completeness of the ten Smiley–Hanson systems, replacing the method of *semantic tableaux* used by Kripke [1963] and Hanson [1965] with the Henkin technique of *maximal consistent* (‘saturated’) sets, as transferred to modal logic by Makinson [1966] (see also Lemmon and Scott [1966]).

In Part IV we introduce ten Anderson-style systems of alethic modal logic with a propositional constant  $Q$ , interpreted as the negation of the Andersonian sanction  $S$  and due to Kanger [1957]. Each of these systems is supplemented by the famous Anderson-style definitions of monadic deontic operators, expressing *obligation*, *permission* and *prohibition*, respectively:

$$\begin{aligned} OA &= df & \Box(Q \rightarrow A) \\ PA &= df & \Diamond(Q \wedge A) \\ FA &= df & \Box(Q \rightarrow \neg A). \end{aligned}$$

Several of these ‘mixed’ alethic-deontic systems were considered by Smiley [1963], to whom, essentially, we owe what I take to be one of the main mathematical results on propositional monadic deontic logic: the Translation Theorem stated in Theorem 45 and proved (in very broad outline) in the proof thereafter. The gist of that result is that the ten Smiley–Hanson systems are, in the well defined sense of Definition 44, the *deontic fragments* of the corresponding alethic systems, as supplemented with  $Q$  and with the above definitions of  $O$ ,  $P$  and  $F$ .

In Parts V and VI we try to pursue the very same line of thought in the area of propositional dyadic deontic logic, for which a good deal of motivation was provided in Part II, notably by Prior’s paradoxes of commitment and Chisholm’s contrary-to-duty imperative paradox (Sections 7 and 8). Thus, we are concerned about logics of *conditional obligation* and *permission*, expressed by such dyadic forms as  $O_B A$  and  $P_B A$  (Section 15), and also about the possibility of *representing* such dyadic logics in systems of alethic modal logic to which a *monadic*, or one-place,  $Q$ -connective is added as well as the definitions given in Section 15.1:

$$\begin{aligned} O_B A &= df & \Box(QB \rightarrow A) \\ P_B A &= df & \Diamond(QB \wedge A) \\ F_B A &= df & \Box(QB \rightarrow \neg A). \end{aligned}$$

We are then able, in Part V, to extend the Smileyan Translation Theorem to the following pairs of logical systems:

- (1)  $\left. \begin{array}{ll} \mathbf{O}_{dy}\mathbf{S4} & \text{and } \mathbf{S4}_{Qmo} \\ \mathbf{O}_{dy}\mathbf{S5} & \text{and } \mathbf{S5}_{Qmo} \end{array} \right\}$  (Section 16)
- (2)  $\mathbf{O}_{dy}\mathbf{S5}^N$  and  $\mathbf{S5}_{Qmo}^N$  (Section 17)

where the first member in each pair is a system of (propositional) dyadic deontic logic and where the second one is a (propositional) alethic modal

logic supplemented with the monadic  $Q$ -operator and with the above Section 15.1 definitions of dyadic deontic operators.

In Part VI we consider certain axiomatic extensions of the systems  $\mathbf{O}_{dy}\mathbf{S5}^N$  (Section 18) and  $\mathbf{S5}_{Qmo}^N$  (Section 19). In Section 20 we then obtain a weakened version of the projected Smileyan Translation theorem, but have to leave open the problem of establishing it fully.

In Sections 21–22 we make an attempt to reconstruct and identify three dyadic deontic logics due to Bengt Hansson [1969] on the basis of certain further extensions of the calculi  $\mathbf{O}_{dy}\mathbf{S5}^N$  and  $\mathbf{S5}_{Qmo}^n$ . A highly interesting feature of those extensions is that, in their semantics, we work with an explicitly specified *preference relation* with which we connect the remaining items in the models considered. Again, in Section 23, we deal with the completeness problem for the most important one among the extensions of  $\mathbf{O}_{dy}\mathbf{S5}^N$ , called the ‘strongly normal’ *core* system  $\mathbf{G}$ , and offer a positive solution to that problem. Finally, in three concluding sections (Sections 24–26) we present some further quite recent results on that ‘core’ system  $\mathbf{G}$ , which were obtained in Åqvist [1996; 1997].

The idea of basing Dyadic Deontic Logic, or the Logic of Conditional Obligation/Permission, on some kind of *preference* theory was proposed by several writers in the late 1960s and early 1970s. Pioneering contributions are due to Sven Danielsson [1968] and to Bengt Hansson [1969]; of these two, [Hansson, 1969] is more easily accessible from a mathematical standpoint, whence our aforementioned attempt to reconstruct his systems in our own framework. Danielsson’s work is considered in the Appendix of [Åqvist, 1987], where it is compared with that of Bas C. van Fraassen, Franz von Kutschera and David Lewis.

So much for the logical technicalities of various systems of formal deontic logic. Going back to Part II, then, we start out with certain preliminary considerations of the relation between formal languages (like that of the Smiley–Hanson systems of monadic deontic logic) and natural languages such as English. Thus, in Section 5, we explain in some detail how a system of formal deontic logic can be supplemented with *definitions of locutions in ordinary English* in much the same way as systems of alethic modal logic were extended with definitions of deontic modalities by Anderson; e.g. we stipulate the following:

- (D6) It is obligatory that  $A = df \quad OA$   
 (D7) It is permitted that  $A = df \quad PA$   
 (D8) You post the letter  $\quad = df \quad p$

where  $p$  is the first proposition letter in the formal language, which, on the basis of definition-theoretical considerations, we regard as a propositional *constant* and hence as a logical symbol. We then end up with a ‘mixed’ formal-English system which contains, on the basis of the definitions added,

certain expressions that count as reasonably good English sentences. In Section 5.1 and 5.2 we then show how our approach leads to (i) a *fragment* of normative English *generated* by those definitions, and (ii) a *translation* (formalization, symbolization) *of* that fragment *into* the original formal language — a translation which is in an obvious sense *induced* by those definitions. After having generalized these ideas in a straightforward way, we go on to introduce (in Section 5.3) the conception of a natural deontic logic *over* the English fragment just generated; this conception is based on an admittedly vague and imprecise notion of *logical validity (truth)*, which is supposedly applicable to some major part of the English language or, at the very least, to the sentences of the fragment. In Section 5.5 we consider *one* well known attempt to make such an intuitive notion of validity more precise, viz. the so called Bolzano criterion.

The whole argument of the crucially important Section 5 provides a basis for comparing and contrasting our natural deontic logic with systems of formal deontic logic, like the Smiley–Hanson ones. As appears from Section 5.3, such a comparison will, in general, lead to one of two results: *either* (i) the natural deontic logic is ‘perfectly matched’ by the formal one, *or* (ii) it is not so ‘perfectly matched’, because there are ‘clashes’, or ‘discrepancies’ between the two. In turn, such a clash may be of two different kinds, at least (as is explained under (I) and (II) in Section 5.3). All this leads up to the three notions of adequacy and faithful representation defined in Section 5.4. That section and its definitions are the main outcome of the discussion of formalization and translation in Section 5. We regard that preparatory discussion as indispensable to any orderly presentation of the deontic paradoxes: it gives us a framework, viz. *definitional extensions of formal theories*, which enables us to present those paradoxes with a sufficient degree of mathematical precision and, at the same time, is of some independent methodological interest to the study of the connections between natural and formal languages in general.

In Section 6–8 we proceed to an exposition of the familiar puzzles known respectively as Alf Ross’ paradox, Arthur N. Prior’s paradoxes of commitment, and Roderick M. Chisholm’s contrary-to-duty imperative paradox. A vital reason for dealing with the latter two is to show how the idea of dyadic deontic logic, originally proposed by Von Wright [1956], naturally suggests itself in any attempt to overcome them. In Section 9, however, we follow Van Eck [1981] in giving a survey of various problems which, as it seems, cannot be satisfactorily handled on the dyadic approach, nor, for that matter, on the monadic one. In Section 9.2 we diagnose the current state of the subject as a whole by agreeing with Van Eck [1981] that there is an urgent need in the area for explicitly temporal and quantificational resources in the basic languages of workable and useful deontic logics; as far as the latter sort of resources are concerned, the point has been made by quite a few writers, of course, e.g. Anderson [1956, p. 200]. In Section

9.2.1, we comment on a powerful research trend in present-day deontic logic, which combines our subject with the logic of action.

As for problems not dealt with in this essay, but which are nevertheless of a very fundamental nature, let us just mention the following two:

- (i) *The problem of truth-or-falsity of deontic (normative) sentences*: are there any true-or-false deontic sentences, expressing obligations, permissions or prohibitions? If not, why? If yes, what are the conditions under which such sentences are true/false, respectively? The discussion in the early twentieth century of this basic issue led some moral and legal philosophers as well as precursors of modern Von Wright-style deontic logic to emphasize a highly important distinction, which may be found in Bentham and which is nowadays usually credited to Ingemar Hedenius [1941]. So our second problem is:
  
- (ii) *The problem of explicating formally the Hedenius [1941] distinction between 'genuine' and 'spurious' deontic sentences*. According to Hedenius [1941], a sentence like 'You shall not kill!' normally directly expresses a prohibition against killing and is then a *genuine* deontic sentence. But the very same sentence, when uttered, e.g. by a Swedish lawyer, may well be interpreted as an elliptical formulation for 'According to Swedish law in 1982 you shall not kill' and function as a *spurious* deontic sentence, which just *asserts the existence of a norm* prohibiting killing *within* a specified legal *system* (without 'directly expressing' that prohibition, as it were). So the present problem concerns the formal explication of the Hedenius genuine vs. spurious distinction. We note here that Wedberg [1951] draws a similar distinction in dividing deontic sentences relatively to a given legal system into such as are *internal* and such as are *external* to the system. Again, Stenius [1963] stresses a *modal vs. factual* distinction applicable to interpretations of normative sentences, Hansson [1969] an analogous *imperative vs. descriptive* one. Finally, Von Wright [1963] distinguishes *norms* from *norm-propositions*.

As a quick reaction to this second problem, I recommend that deontic logicians consider more seriously the scarce attempts in the literature to construct logics of *commanding* as opposed to logics of *commands*, e.g. Fisher [1961a], Hanson [1966] and Bailhache [1981], where authorities and addressees are explicitly brought to the fore. These attempts look very promising indeed for future developments of our subject.

#### 4 ELEMENTARY PROPERTIES OF SOME VON WRIGHT-TYPE DEONTIC LOGICS

In this section we shall first introduce two notions of a *normal* Von Wright-type deontic logic, secondly, comment on our suggested definition of those notions and, thirdly, list some obvious properties of these logics. Consider the formal language of the ten Smiley–Hanson systems of propositional monadic deontic logic studied in Parts III and IV below. Its set  $\Sigma$  of we formed *sentences* is defined as the smallest set which (i) has every proposition letter as an element, and (ii) is closed under the usual truth-functional connectives including the constants *verum* and *falsum* as well as under the two primitive monadic deontic operators **O** (for obligation) and **P** (for permission). We then propose the following:

DEFINITION 1.

Let  $\mathcal{L}$  be any subset of  $\Sigma$ . Then:

- (I)  $\mathcal{L}$  is a *normal propositional monadic Von Wright-type deontic logic* iff
  - (a) every thesis, i.e. provable sentence, of the system **OK** (Section 10.2) is a member of  $\mathcal{L}$ , and
  - (b)  $\mathcal{L}$  is closed under uniform substitution for proposition letters, detachment for material implication and the rule of *O*-necessitation (Section 10.2).
- (II)  $\mathcal{L}$  is a *strongly normal* (propositional monadic) Von Wright-type deontic logic iff
  - (a) every thesis of the system **OK**<sup>+</sup> (Section 10.2) is a member of  $\mathcal{L}$ , and
  - (b)  $\mathcal{L}$  is closed under substitution, detachment and *O*-necessitation.

REMARK 2.

- (i) The system **OK** is determined as follows:

*Axiom schemata:*

- (A0) All tautologies over  $\Sigma$
- (A1)  $PA \Leftrightarrow \neg O\neg A$
- (A2)  $O(A \rightarrow B) \rightarrow (OA \rightarrow OB)$ .

*Rules of proof:*

$$(R1) \quad \frac{A, A \rightarrow B}{B} \quad (\text{modus ponens, detachment}).$$

$$(R2) \quad \frac{A}{OA} \quad (O\text{-necessitation}).$$

In the spirit of Section 10.2 we then define the set of **OK**-*provable* sentences (of **OK**-*theses*) as the smallest set which contains every instance of the schemata (A0)–(A2) as its member and which is closed under the rules (R1) and (R2). Since in this work we usually identify a logic(al system) with the set of its theses, we may even say that the system **OK** is identical to that set. Note also that our use of axiom *schemata* instead of single axioms guarantees that **OK** is closed under (uniform) substitution (for proposition letters), so no primitive rule of proof to that effect is needed.

The system **OK**<sup>+</sup> results from **OK** by adding to the latter every instance of the schema

$$(A3) \quad OA \rightarrow PA \quad (\text{whatever is obligatory is also permitted})$$

as a new axiom; for sure, **OK**<sup>+</sup> is to remain closed under (R1) and (R2).

- (ii) Our definition of ‘normality’ is meant to harmonize with the definition of a *normal* (alethic) *modal* logic given, e.g. by Makinson [1966], Segerberg [1971a] and Hansson and Gärdenfors [1973]. Many writers, including Von Wright [1951], Prior [1955] and Anderson [1956], are likely to regard this notion of normality (i.e. the one defined in (I)) as too weak, since not every instance of the schema (A3) is provable in **OK**; they are then likely to prefer, other things being equal, **OK**<sup>+</sup> to **OK** and our concept of *strong* normality as the better notion. But quite a few authors, say, Erik Stenius [1963, (interesting argument on p. 254)] and Manfred Moritz [1963] (to mention just one specimen from a large production), have reacted against accepting (A3) (or its equivalent **OK**<sup>+</sup>**1** stated in Section 4.1 below) as a valid principle of deontic logic which is satisfied by every *existing* system of norms (at best, according to Stenius, (A3) is satisfied by every system of norms which is *possible to obey*); such authors are likely to prefer our weaker notion of normality.

Our definition of normality might still seem to be objectionable on the ground that every normal logic is required to be closed under the rule (R2) of *O*-necessitation. Roughly speaking, accepting this rule commits us to the position that every logically true proposition



(e.g. every tautology) is obligatory and, dually, every logical falsehood (contradictory state-of-affairs) is forbidden. Prior [1955] finds ‘no evident reasonableness’ in this position, and Von Wright [1951] explicitly rejects it when he proposes a *principle of deontic contingency* to the effect that such schemata as

$$O(A \vee \neg A) \quad \text{and} \quad \neg P(A \wedge \neg A)$$

should *not* be accepted as valid.

It seems to me, however, that the rule of *O*-necessitation (or something very much like it) has been successfully defended by Stenius [1963, p. 253], and by Anderson [1956, pp. 181–183]. Also, in Anderson [1956, Section IX], he outlines an interesting way of doing justice to the intuitions underlying Von Wright’s principle of deontic contingency. Essentially, his method amounts to this: we may accept a system admitting rule R2 as our *basic* (monadic) deontic logic; then, if such a system has resources for expressing alethic *contingency* in the sense of absence of necessity and of impossibility, we could define in it *new* concepts of obligation, permission and prohibition, which will apply only to contingent propositions (state-of-affairs). Clearly, the logic of these new concepts will not be normal in the sense of our definition above; but it may be developed *as a definitional extension of* a normal deontic logic in our sense. And this, I take it, is highly advantageous from a methodological standpoint.

- (iii) Somewhat cautiously, we speak in the definition above of normal (strongly normal) *propositional* monadic Von Wright-*type* deontic logics for the following reasons. Although our notions are ultimately inspired by the pioneering contribution of Von Wright [1951], the main difference between his original system and those discussed in this essay is that for Von Wright *O* and *P* are deontic *predicates*, which form sentences when applied to *names of acts* (in the sense of ‘act-types’), whereas for us, and many others, *O* and *P* are deontic *modalities* (modal operators or connectives), which form sentences when applied to *sentences* (which may possibly assert that such and such an act is performed, though). An advantage of viewing *O* and *P* as modalities is that questions concerning the status and acceptability of ‘mixed formulae’, like  $OA \rightarrow A$ , and of formulae involving iterated modalities, like  $O(OA \rightarrow A)$ , can be meaningfully raised and discussed; on Von Wright’s original approach, such questions were ruled out at the outset, since those kinds of formulae did not even count as well formed sentences.

(iv) Our proposed concepts of normality deviate from the Andersonian notion of a *normal deontic logic*, defined in Anderson [1956, p. 168], in the following respects:

- (a) Whereas we require closure under the rule of *O*-necessitation, Anderson only requires closure under a rule allowing for the intersubstitutability of provably equivalent expressions from the classical two-valued propositional calculus.
- (b) As was pointed out under (ii) above, he wants the schema (A3) to be provable in any normal deontic logic; we make this a condition of *strong* normality.
- (c) In addition to requiring certain schemata to be *provable* in any normal deontic logic Anderson also requires that certain schemata should *not* be provable in such logics, e.g..

$PA \rightarrow A$  (whatever is permitted is the case)

$A \rightarrow PA$  (whatever is the case is permitted).

Now, I think our considerations under (ii) above explain sufficiently well why we prefer to deviate from the Andersonian concept of normality in the respects (a) and (b). As far as (c) goes, I take his suggestion that certain schemata be *unprovable* in normal deontic logics to be perfectly sound; *negative requirements* of the sort may well be used to define new and stronger notions of normality, if properly defended, that is to say. Anderson's list of 'unprovables' could even be extended with a huge number of difficult items; let me just pick a few from the vast literature:

- (1)  $OA \Leftrightarrow A$  (whatever is obligatory is the case and conversely).
- (2)  $O(A \vee B) \rightarrow (OA \vee OB)$  (if *A-or-B* is obligatory, then *A* is obligatory or *B* is obligatory).
- (3)  $(PA \wedge PB) \rightarrow P(A \wedge B)$  (if *A* is permitted and *B* is permitted, then *A-and-B* is permitted).
- (4)  $(OA \wedge (A \rightarrow OB)) \rightarrow OB$  (if *A* is obligatory and if *A* then it is obligatory that *B*, then *B* is obligatory).

The strange result (1) is provable in the Mally [1926] system, see also Føllesdal and Hilpinen [1971, p. 4]. For interesting comments on the Mally system from the fresh standpoint of deontic *temporal*

logic, see Van Eck [1981, p. 81 f]. The unacceptable results 2 and 3 would seem to be provable in natural extensions of the Kalinowski [1953] systems  $\mathbf{K}_1$  and  $\mathbf{K}_2$ , which are based on a suggestive (2 + 3) valued matrix; see Prior [1956]. Again, (4), due to Prior [1955], is criticized by Hintikka [1957] and [1971]; but it receives an interesting vindication in terms of the Hintikka notion of *deontic* (as opposed to *logical*) *consequence*. Note that this notion of Hintikka's is very definitely based on a conception of  $O$  and  $P$  as modalities as opposed to predicates (see remark (iii) above).

- (v) We readily verify that the ten Smiley–Hanson systems (Section 10) are normal propositional monadic Von Wright-type deontic logics in the sense of clause (I) of our definition; and that the five +-systems (starting with  $\mathbf{OK}^+$ ) are strongly normal ones. The question now arises: can we extend the notions of normality to the systems of dyadic deontic logic studied in Parts V and VI? My answer will be a bit tentative, because dyadic deontic logic does not yet appear to be a sufficiently well established discipline; as will be clear from the Appendix of Åqvist [1987], there is still too much controversy and disagreement about fundamentals to justify a firm answer. Nevertheless, I believe that our argument in Parts V and VI shows that the basic language of dyadic deontic logic must contain the operators  $N$  and  $M$  of *universal necessity* and *universal possibility* (for this terminology, see Scott [1970, p. 157]). The set of sentences of such a language will then be  $\Sigma_{O,N}^2$  (Section 17) rather than  $\Sigma_O^2$  (Section 15), and the weakest dyadic logic over  $\Sigma_{O,N}^2$  is the system  $\mathbf{O}_{dy}\mathbf{S5}^N$  (see again Section 17). So we propose to define a *normal propositional dyadic Von Wright-type deontic logic* as any subset of  $\Sigma_{O,N}^2$  which contains every thesis of  $\mathbf{O}_{dy}\mathbf{S5}^N$  and which is closed under its rules of inference (detachment and  $N$ -necessitation; as usual, closure under substitution is guaranteed by the use of axiom schemata). Again, a *strongly normal* dyadic logic of this sort will, we propose, have to contain the system  $\mathbf{G}$  (Sections 22 and 23) and to be closed under the above rules. Note that  $\mathbf{O}_{dy}\mathbf{S5}^N$  and  $\mathbf{G}$  are much richer theories than, e.g. the Smiley–Hanson systems, in point of expressive and deductive power.

#### 4.1 Theorems and rules of $\mathbf{OK}$ and $\mathbf{OK}^+$

We now list some theorem-schemata, or thesis-schemata, of  $\mathbf{OK}$ , i.e. schemata of which each instance (in  $\Sigma$ ) is provable in  $\mathbf{OK}$ . By our definition of normality, they will then be provable in every normal *monadic* deontic logic as well.

- OK1.**  $OA \Leftrightarrow \neg P\neg A.$   
**OK2.**  $\neg OA \Leftrightarrow P\neg A.$   
**OK3.**  $O\neg A \Leftrightarrow \neg PA.$   
**OK4.**  $O\top.$   
**OK5.**  $O(A \wedge B) \Leftrightarrow (OA \wedge OB).$   
**OK6.**  $P(A \vee B) \Leftrightarrow (PA \vee PB).$   
**OK7.**  $OA \wedge PB \rightarrow P(A \wedge B).$   
**OK8.**  $OA \vee OB \rightarrow O(A \vee B).$   
**OK9.**  $P(A \wedge B) \rightarrow (PA \wedge PB).$   
**OK10.**  $(OA \wedge O(A \rightarrow B)) \rightarrow OB.$   
**OK11.**  $(PA \wedge O(A \rightarrow B)) \rightarrow PB.$   
**OK12.**  $(O\neg B \wedge O(A \rightarrow B)) \rightarrow O\neg A$   
**OK13.**  $(O(A \rightarrow (B \vee C)) \wedge (O\neg B \wedge O\neg C)) \rightarrow O\neg A.$   
**OK14.**  $OB \rightarrow O(A \rightarrow B).$   
**OK15.**  $O\neg A \rightarrow O(A \rightarrow B).$

Suggested readings of many of these items can be found in Anderson [1956, pp. 180 ff]. Note that the compound operator  $O\neg$  may be read as ‘it is forbidden that’.

Furthermore, **OK** (and hence any normal monadic logic) is closed under the following rules of proof:

$$\mathbf{OKa.} \quad \frac{A \rightarrow B}{OA \rightarrow OB}.$$

$$\mathbf{OKb.} \quad \frac{A \Leftrightarrow B}{OA \Leftrightarrow OB}.$$

$$\mathbf{OKc.} \quad \frac{A \rightarrow B}{PA \rightarrow PB}.$$

$$\mathbf{OKd.} \quad \frac{A \Leftrightarrow B}{PA \Leftrightarrow PB}.$$

Again, the following schemata are provable in **OK**<sup>+</sup> and, hence, in any *strongly* normal monadic calculus:

- OK<sup>+</sup>1.**  $OA \rightarrow \neg O\neg A$ .  
**OK<sup>+</sup>2.**  $\neg(OA \wedge O\neg A)$ .  
**OK<sup>+</sup>3.**  $PA \vee P\neg A$ .  
**OK<sup>+</sup>4.**  $P\top$ .  
**OK<sup>+</sup>5.**  $\neg(O(A \vee B) \wedge (O\neg A \wedge O\neg B))$ .

Schema **OK<sup>+</sup>5** may be taken to assert that it is impossible to be obliged to choose between forbidden alternatives (see, e.g. Von Wright [1951]). Note that **OK<sup>+</sup>5** is not a theorem-schema of **OK**; hence, it is not forthcoming in every normal monadic calculus.

Finally, **OK<sup>+</sup>** is closed under the following rule of proof:

$$\mathbf{OK}^+ \mathbf{a.} \quad \frac{A}{PA}$$

whereas **OK** fails to be closed under that rule.

## II. PARADOXES AND DILEMMA'S

### 5 PRELIMINARIES ON FORMALIZATION AND TRANSLATION

Consider the *formal*, or *symbolic language* common to the ten Smiley–Hanson systems of monadic deontic logic to be studied in Part III. That formal language can be conceived of as a structure

$$L = \langle \text{Bas}, \text{LogCon}, \text{Aux}, \text{Sent} \rangle$$

where:

- (i) Bas (= the set of *basic sentences* of L) is a denumerable set Prop of *proposition letters*  $p, q, r, p_1, p_2, \dots$
- (ii) LogCon (= the set of primitive *logical connectives* or *constants* of L) is the set  $\{\top, \perp, \neg, \wedge, \vee, \rightarrow, \Leftrightarrow, O, P\}$ .
- (iii) Aux (= the set of *auxiliary* symbols of L) is the set consisting of the left parenthesis and the right parenthesis; thus,  $\text{Aux} = \{(, )\}$ .
- (iv) Sent (= the set of all well formed *sentences* of L) is identical to the set  $\Sigma$  as defined Section 10.1 i.e. the smallest set  $S$  such that
  - (a) every proposition letter in Prop is in  $S$ ,
  - (b)  $\top$  and  $\perp$  are in  $S$ ,
  - (c) if  $A$  is in  $S$ , then so are  $\neg A, OA$  and  $PA$ ,

- (d) if  $A, B$  are in  $S$ , then so are  $(A \wedge B)$ ,  $(A \vee B)$ ,  $(A \rightarrow B)$  and  $(A \Leftrightarrow B)$ .

This ‘recursive’ or ‘inductive’ definition of  $\Sigma$  may be summarized by saying that the set of sentences of  $L$  is the smallest set which (i) contains every basic sentence in  $\text{Bas}$  as an element, this according to clause (a), and (ii) is *closed* under every logical connective in  $\text{LogCon}$ , this being the joint effect of clauses (b)–(d). Note that these connectives are of different *degrees*, which are revealed, so to speak, by the way they are used to form new sentences : by clause (c),  $\neg$ ,  $O$  and  $P$  do so when applied to *one single* sentence as their argument and, hence, are said to be of degree 1; by clause (d),  $\wedge$ ,  $\vee$ ,  $\rightarrow$  and  $\Leftrightarrow$  do so when applied to *any two* sentences as their arguments, and are, consequently, said to be of degree 2; finally, by clause (b),  $\top$  and  $\perp$  require *no* argument *at all* when used to form sentences, hence, they are said to be of degree 0 and to be *propositional constants* (as opposed to so called *propositional variables*).

Now, assume that we are interested in *theories*, or *logics*, formulated in the language  $L$  as just described; for instance, any of the Smiley–Hanson systems to be studied below. Generally speaking, such a logic is a subset of  $\Sigma$  determined by a finite number of *classes of axioms* having a common and peculiar form or *Gestalt* (these classes are usually known as ‘axiom schemata’) as well as by certain *rules of inference* (perhaps more appropriately: *rules of proof*). Now, why have deontic logicians and moral philosophers alike paid so much attention to systems of the Smiley–Hanson kind rather than to other subsets of  $\Sigma$  that might just as well have been selected for attention? An obvious reason appears to be this: certain definite *ordinary language readings* are associated with the logical connectives in  $L$  and, *on these readings*, the principles of these logics have more or less good claims to being true, valid, correct, acceptable, or whatnot, in a pre-systematic, informal or intuitive sense.

Which English readings are we then to associate with the connectives in  $\text{LogCon}$ ? Here is a familiar list:

- $\neg$ : not (more fully: it is not the case that)
- $\wedge$ : and (more fully: both \_\_\_\_\_, and...)
- $\vee$ : or (more fully: either \_\_\_\_\_, or...)
- $\rightarrow$ : if \_\_\_\_\_, then...
- $\Leftrightarrow$ : if and only if (alternatively: if and only if \_\_\_\_\_, then...)
- $O$ : it is obligatory that (alternatively: it ought to be that)
- $P$ : it is permitted that (alternatively: it may be that).

Thus,  $\neg$ ,  $\wedge$ ,  $\vee$ ,  $\rightarrow$  and  $\Leftrightarrow$  are to be symbols for the five best known so-called *truth-functions* of classical propositional logic, viz. negation, con-

junction, disjunction, material implication and material equivalence, respectively. Again,  $O$  and  $P$  are to symbolize the *normative* (or *deontic*) notions of obligation and permission. Finally,  $\top$  (*verum* in Latin) symbolizes some arbitrary, but fixed logical truth or tautology, and  $\perp$  (*falsum* in Latin) some arbitrary, but fixed logical falsehood, contradiction or absurdity; their precise reading in English is less important.

We may think of the list above as presenting items in a little logical dictionary or lexicon: they tell us both

- (i) how to translate connectives in LogCon into plain English, and
- (ii) how to translate certain English locutions, viz. those appearing to the right in the list, ‘back’ into the formal language L.

In this way, we suggest, we get an idea about the intended interpretation of the language L or, at least, of its logical connectives. But it still remains unclear how to understand the list of lexical items (the ‘logical dictionary’) itself; what is its status in a formal theory, like any one of the Smiley–Hanson systems of monadic deontic logic? What does it mean to assign an English reading to a connective in L on the basis of a lexical item in the list?

In answer to these questions, we propose to equate that list with the following series of *definitions*, applicable to any L-sentences  $A$  and  $B$ :

- (D1) It is not the case that  $A = \text{df } \neg A$ .
- (D2) Both  $A$  and  $B = \text{df } (A \wedge B)$ .
- (D3) Either  $A$  or  $B = \text{df } (A \vee B)$ .
- (D4) If  $A$  then  $B = \text{df } (A \rightarrow B)$ .
- (D5) If and only if  $A$  then  $B = \text{df } (A \leftrightarrow B)$ .
- (D6) It is obligatory that  $A = \text{df } OA$ .
- (D7) It is permitted that  $A = \text{df } PA$ .

Let us now distinguish some effects of adding this series D1–D7 to our language L or to any theory formulated in L.

- (I) We obtain a new language, call it L(D1–D7), which is *like* L in having (i) the same set Bas of basic sentences, viz. the set Prop, and (ii) the same set Aux consisting of the two brackets. But L(D1–D7) *differs from* L in having (i) a larger set LogCon than L and (ii) a larger set of sentences than L; because the seven defined English connectives ‘it is not the case that’, ‘both \_\_\_\_\_ and...’ etc. will be among the logical connectives of L(D1–D7), though not among those of L; and because

the set of sentences of  $L(D1-D7)$ , call it  $\Sigma(D1-D7)$ , will be closed, not only under the old, ‘symbolic’ connectives of  $L$ , but under the new, defined English ones as well. For example, the following strings:

$$(3) \quad \left. \begin{array}{l} \text{it is not the case that } p \\ \text{it is obligatory that } q \end{array} \right\} \quad (\text{where } p, q \text{ are in Prop})$$

are sentences of  $L(D1-D7)$ , though not of  $L$ . We shall call the new language  $L(D1-D7)$  a *definitional enrichment* of  $L$  (reserving the more familiar label ‘definitional extension’ for theories, or logics, formulated in  $L$ ).

REMARK 3. A description of the first sentence above, which is agreeable to fans of use and mention, is this: the result of writing the prefix ‘it is not the case that’ immediately in front of the proposition letter in Prop that is denoted by ‘ $p$ ’. In this work we mostly stick to the familiar convention of using logical connectives *autonomously*, i.e. as names of themselves.

(II) Usually, one conceives of a definition as something added to a *theory* rather than to a ‘bare’ language (see e.g. Suppes [1957, p. 152 f.]). Suppose, then, that we add the series D1–D7 to any of the Smiley–Hanson systems below, e.g. to the logic **OK** as described in Section 10.2. How are we to understand D1–D7 within **OK**, then? In particular, what does the symbol ‘=df’ mean? We suggest here that ‘=df’ can throughout be replaced by the sign  $\Leftrightarrow$  for material *equivalence* or ‘biconditionality’, and that every resulting sentence is to be regarded as a *new axiom* which is added to those of the logic **OK**. (We assume then that the outermost pair of parentheses has been dropped from D1–D7, complying with the customary convention.) From the standpoint of the logic **OK**, D1–D7 differ from its ‘proper’ axiom schemata A0–A2 chiefly in respect of introducing into it *new* symbols that are not already in its language  $L$ , viz. the seven English connectives defined by D1–D7 (they also differ from A0–A2 in other respects not commented upon here). Thus, the logic **OK** *supplemented with* D1–D7 *as here understood*, call it **OK(D1–D7)**, will be a certain subset of the larger sentence-set  $\Sigma(D1-D7)$ , *not* of  $\Sigma$ , and is determined not only by A0–A2 and the rules of inference R1 and R2, but by the ‘definitional axiom schemata’ D1–D7 as well. Finally, just as  $L(D1-D7)$  was said above to be a definitional enrichment of the language  $L$ , we now say that the logic **OK(D1–D7)** is a *definitional extension* of the logic **OK**; the same thing goes for any theory over, i.e. formulated in,  $L$ , to which D1–D7 are added.

(III) We observed a moment ago that such strings as

$$(3) \quad \begin{array}{l} \text{it is not the case that } p \\ \text{it is obligatory that } q \end{array} \quad (\text{with } p, q \text{ in Prop})$$



count as sentences in the enrichment  $L(D1-D7)$  of  $L$ . Now, these strings are hardly acceptable as sentences of *English*, although they surely involve nice English components, viz. the *definienda* of D1 and D6, respectively. At best, they could count as sentences of pseudo-English or quasi-English. The reason for this being so is that, as such,  $p$  and  $q$  fail to make any sense in English; as one often says, they are just empty place-holders for genuine English sentences. But couldn't we begin to assign English readings to these letters, just as we did to the logical connectives in  $L$  by means of D1–D7? If this were possible, we might be able to *generate* some genuinely English sentences within some definitional enrichment of  $L$ . Let us try. Consider the English sentence figuring in the Alf Ross paradox, viz.

(0) You post the letter.

Again, let  $p$  be the first proposition letter in an assumed enumeration, or ordering, of the set Prop. We now lay down the definition:

(D8) You post the letter =df  $p$

which, in accordance with the decisions taken under (II) above, is to be understood as a *single axiom*:

You post the letter  $\Leftrightarrow p$

(or, perhaps, as a degenerate axiom schema having that single axiom as its only instance).

We must now raise a vitally important question: is D8, thus understood, acceptable from a standpoint of the theory of definitions? Well, its answer depends on which of the following alternatives applies:

- (A) In D8,  $p$  is, a propositional *variable*, and any theory  $T$  formulated in the language  $L$  enriched by D8, call it  $L(D8)$ , is closed under a *rule of substitution* for propositional variables.
- (B) In D8,  $p$  is, and indeed has to be, a propositional *constant*, which is syntactically and grammatically on a par with the logical 0-place connectives  $\top$  and  $\perp$  of  $L$ . Otherwise, D8 is altogether unacceptable from the point of view of definition theory.

Let us explore these two alternatives in turn.

*Alternative A.* What does it mean to say that any theory over  $L(D8)$  is closed under substitution for propositional variables? At least the following, I suggest:

- (1) For each sentence  $A$  of  $L(D8)$ :

You post the letter  $\Leftrightarrow A$

is *provable* (as a 'thesis' or 'theorem') in any such theory.

More leisurely expressed: D8 is an axiom of any theory  $T$  over  $L(D8)$ . Hence, D8 is provable in  $T$ . And so is any result of substituting a sentence  $A$  of  $L(D8)$  for the variable  $p$  in D8.

Now, an obvious consequence of adding D8 to  $L$  is this:

(2) The strings:

You post the letter

$\neg$  You post the letter

are sentences of  $L(D8)$ .

Therefore, by (1) and (2), we conclude that the string

(£) You post the letter  $\Leftrightarrow \neg$  You post the letter

is not only a sentence of  $L(D8)$  but also *provable* in any theory over  $L(D8)$ , since (£) is the result of substituting the  $L(D8)$ -sentence

$\neg$  You post the letter

for  $p$  in D8. But (£) is obviously a contradiction in any such theory (of any interest), hence, any such theory is inconsistent.

Hence, alternative  $A$  must be rejected.

Another way of explaining its failure is as follows. On alternative  $A$ , D8 violates a basic definition-theoretic principle according to which no proper definition of a sentence-forming connective is allowed to use any *free* (i.e. not bound to any quantifier) variable in the *definiens* which does not occur already in the *definiendum*. Well, in D8, on this alternative,  $p$  is a free propositional variable which obviously does not occur in the English sentence (or sentence-forming connective of degree 0) ‘You post the letter’ (although, for sure, the 16th letter ‘ $p$ ’ of the English alphabet does). For a statement of the definition-theoretic principle in a somewhat different context, see Suppes [1957, p. 156 f.]; for its motivation in the present context, recall that we just proved every theory over  $L(D8)$  to be inconsistent. Note also that the principle is a consequence of more general criteria in definition theory known as those of non-creativity and relative consistency (Suppes [1957, pp. 154–5]).

*Alternative B.* Having rejected  $A$ , this is the alternative we must accept in order to protect D8. Let us pay attention to two consequences of regarding  $p$  in D8 as a *constant*, and indeed a *logical* one. (Note that in the context we disregard any subdivision of constants into logical vs. non-logical (descriptive) ones; this is a deliberate, but remediable oversimplification.)

(i) The disastrous sentence (£) will not be provable in any *consistent* theory over  $L(D8)$ , or over  $L(D1–D7, D8)$ , because D8 is now

a proper definition which conforms to the rules and criteria of definition theory. In particular, there is nothing like a ‘rule of substitution’ for logical constants.

- (ii) Let us continue to think of the set Prop of ‘proposition letters’ as a set of *free* propositional *variables*, which form the stock of *non-logical* symbols of L and its enrichments. But then, obviously, the first letter  $p$ , used in D8 as now understood, is wrongly classified, being a constant symbol. So in any enrichment of L containing D8,  $p$  has to be *re-classified* as a logical constant, i.e. *moved from Prop into* the set LogCon of the enrichment. Similarly, the English sentence ‘You post the letter’, defined by D8, should be classified as a constant symbol and placed in the set LogCon of such enrichments.

EXERCISE 4. Consider the language  $L(D1-D7, D8)$  and think of it as a structure  $L(D1-D7, D8) = \langle \text{Bas}, \text{LogCon}, \text{Aux}, \text{Sent} \rangle$ . Specify the components of that structure, observing the instructions just given under (ii) above!

REMARK 5. One might object to our argument concerning the status of  $p$  in D8 that we have overlooked a third alternative:  $p$  is a propositional *parameter*. Here is my answer: either a parameter is a free variable and the present alternative amounts to Alternative A and is ‘out’; or a parameter is a constant, so the present alternative amounts to Alternative B and we gladly accept it. Is the objection thereby met?

### 5.1 *On the fragment of normative English constructible in $L(D1-D7, D8)$ .*

Consider the language  $L(D1-D7, D8)$ . The following are sentences of  $L(D1-D7, D8)$ :

It is obligatory that you post the letter.

Either it is not the case that you post the letter or it is permitted that you post the letter.

If it is obligatory that you post the letter then it is permitted that you post the letter.

which are also, indeed, reasonably good *sentences of English*. (We don’t deny that they can be improved from the standpoint of *stylistic elegance*, using devices like better punctuation, inversion of word order, pronominalization, attaching some negation- or obligation-expressing phrase to the main verb in the infinitive, and so on.) Now, is it possible somehow to characterize rigorously the set of those *English* sentences which we can *generate*

in the language  $L(D1-D7, D8)$ ? Yes, easily, as follows: it is the smallest set containing (i) the English sentence (0) (= ‘You post the letter’), defined in D8, and which (ii) is closed under the seven English connectives defined in the series D1–D7. We now propose the following, somewhat more comprehensive notion:

DEFINITION 6 (FNE(D1–D7, D8)). Let  $L(D1-D7, D8)$  be the *definitional enrichment* of  $L$  that arises from adding D1–D7, D8 to  $L$ . Think of it as a structure whose components are specified as in the Exercise above! Then, by the *fragment of normative English constructible in  $L(D1-D7, D8)$*  we shall mean the structure

$$\text{FNE}(D1-D7, D8) = \langle \text{Bas}, \text{LogCon}, \text{Aux}, \text{Sent} \rangle$$

where:

- (i) Bas = {You post the letter}.
- (ii) LogCon = {it is not the case that, both \_\_\_\_ and..., either \_\_\_\_ or..., if \_\_\_\_ then..., if and only if \_\_\_\_ then..., it is obligatory that, it is permitted that}.
- (iii) Aux =  $\emptyset$  (i.e. the empty set).
- (iv) Sent = (as usual) the smallest set which contains every basic sentence in Bas as an element and is closed under every logical connective in LogCon.

REMARK 7.

- (a) The basic sentence as well as the logical connectives are used autonomously in this description, i.e. as names of themselves. Hence, no quotation marks are needed.
- (b) Note that, although the English sentence ‘You post the letter’ was classified as a logical connective (of degree 0) in  $L(D1-D7, D8)$ , it is not so classified in the fragment  $\text{FNE}(D1-D7, D8)$  but as a basic sentence. The reason for this decision is that we want Sent to be definable as the result of closing a *non-empty* set Bas under certain connectives.
- (c) There is no need for parentheses in the fragment. This is due to the ‘smart’ reading in English of the *binary* connectives that is codified by the definitions D2–D5; the point being that every English sentence in the fragment, which has any of the *definienda* of D2–D5 as its *principal sign* or *main connective*, begins with that very connective.

Thus, no ambiguities of grouping are forthcoming in the fragment, and its notation is Polish or Łukasiewiczian.

### 5.2 *The translation induced by D1–D7, D8; An extension of that translation*

Let us quickly rehearse what has been done so far. We started out by considering the purely *symbolic*, or *formal* language  $L$  common to the Smiley–Hanson systems of monadic deontic logic and its set  $\Sigma$  of purely symbolic sentences. Then, we added to  $L$  eight definitions D1–D7, D8, thereby obtaining a richer language  $L(D1–D7, D8)$ , whose set  $\text{LogCon}$  contained eight fresh *English* connectives, viz. the *definienda* of D1–D8; also, the first proposition letter in an assumed ordering of  $\text{Prop}$ , i.e.  $p$ , was moved into that set  $\text{LogCon}$ . The set of sentences of that richer language, call it  $\text{Sent}_{L(D1–D7, D8)}$ , was then seen not only to be larger than that of  $L$ , but indeed to contain as a subset a certain class of ‘reasonably good’ *sentences of English*, viz. the set of sentences of the *fragment*  $\text{FNE}(D1–D7, D8)$  of *Normative English constructible in*  $L(D1–D7, D8)$ , as defined above. Call that class either by the fancy name of

$$\text{Sent}_{\text{FNE}(D1–D7, D8)} \quad (= \text{Sent}, \text{ as defined by clause (iv)})$$

or call it simply  $\Phi$ .

Now, can we say anything informative about the relation of the original sentence-set  $\Sigma$  to that of the fragment  $\text{FNE}$ , i.e.  $\Phi$ , apart from the claim that the fragment and its sentences were somehow *generated* by the addition of D1–D7, D8 to the original language  $L$ ? (Assume, for safety, that the letter  $p$  is reclassified and moved into  $\text{LogCon}$  already in  $L$ , so that we are free to add D8 to  $L$ .) Well, we can obviously assert the following:

There exists a *translation* (formalization, symbolization) of the fragment *into* the original formal language  $L$  in the sense of a *function*  $t$  which *maps*  $\Phi$  *one-to-one into*  $\Sigma$ , where  $t$  is defined by the following recursive conditions:

- (i)  $t(\text{‘You post the letter’}) = p$ ; cf. D8.

Again, assume that  $t$  has been defined for any English sentences  $A, B$  in  $\Phi$ . Then:

- (ii)  $t(\text{it is not the case that } A) = \neg t(A)$ ; cf. D1.  
 (iii)  $t(\text{both } A \text{ and } B) = (t(A) \wedge t(B))$ ; cf. D2.  
 (iv)  $t(\text{either } A \text{ or } B) = (t(A) \vee t(B))$ ; cf. D3.

- (v)  $t(\text{if } A \text{ then } B) = (t(A) \rightarrow t(B))$ ; cf. D4.
- (vi)  $t(\text{if and only if } A \text{ then } B) = (t(A) \Leftrightarrow t(B))$ ; cf. D5.
- (vii)  $t(\text{it is obligatory that } A) = (Ot(A))$ ; cf. D6.
- (viii)  $t(\text{it is permitted that } A) = (Pt(A))$ ; cf. D7.

It is obvious that, as defined,  $t$  is a function *from*  $\Phi$  *into*  $\Sigma$ . It is almost just as obvious that  $t$  is *one-one* in the sense that for any two distinct sentences  $A, B$  in  $\Phi$  we have that  $t(A)$  and  $t(B)$  are distinct sentences in  $\Sigma$ ; the inductive proof of this fact is rather tedious, although basically easy to grasp.

We see that each clause in the definition of  $t$  corresponds to exactly one member in the series D1–D8. Hence, it is appropriate to speak of the translation  $t$  as being *induced by* that series. Again, in the present case of  $\Phi$  and  $\Sigma$ , we speak alternatively of  $t$  as a *formalization* or a *symbolization*, because  $\Phi$  is the set of sentences of a fragment of a *natural* language, viz. English, and  $\Sigma$  is the set of sentences of a *purely symbolic* language, viz. L, and  $t$  translates the former into the latter. Not every translation function has this special character, of course.

Suppose now that we add further definitions in the style of D8 to the language L(D1–D7, D8), of the general form:

- (D9)        \_\_\_\_\_ =df  $p_1$ .
- (D10)      \_\_\_\_\_ =df  $p_2$ .
- .
- .
- .
- (D8+k.)    \_\_\_\_\_ =df  $p_k$ .

where:

- (i)  $k$  is a positive integer  $\geq 1$ .
- (ii)  $p_1, \dots, p_k$  are distinct proposition letters that have been reclassified already in the original language L and moved into its set LogCon, where we also find  $p$  which is distinct from all of them.
- (iii) The blanks are filled by distinct English sentences considered as *un-analyzed wholes* ('ungetrenntes Ganzes' in the terminology of Hilbert and Ackermann [1928]), and all distinct from 'You post the letter'.

Then, the *fragment of normative English constructible in* the enrichment L(D1–D7, D8–D8+k) is easy to identify: its set Bas of basic sentences contains exactly  $k + 1$  English members and, as usual, its total set Sent of sentences is the result of closing Bas under the by now familiar seven

English connectives. Now, let  $\Phi$  be that total set, this time, and let  $\Sigma$  still be the set of purely symbolic sentences of  $L$  (with  $p, p_1, \dots, p_k$  reclassified as indicated above). How are we to define the extended translation, let us still call it  $t$ , which is induced by the series D1–D7, D8–D8+k? Very easy: just add to our earlier definition of  $t$  the following  $k$  clauses in the recursion (or induction) basis:

- $$\begin{aligned} (i_1) \quad t(\text{definiendum of D9}) &= p_1; \quad \text{cf. D9.} \\ (i_2) \quad t(\text{definiendum of D10}) &= p_2; \quad \text{cf. D10.} \\ &\vdots \\ &\vdots \\ (i_k) \quad t(\text{definiendum of D8+k}) &= p_k; \quad \text{cf. D8+k.} \end{aligned}$$

EXERCISE 8. Verify that, as just defined, the extended translation  $t$  remains a one-one mapping of  $\Phi$ , as presently understood, into  $\Sigma$ !

### 5.3 On the natural deontic logic over $FNE(D1-D7, D8-D8+k)$

Let  $FNE(D1-D7, D8-D8+k)$ , or  $FNE_k$  for short, be the fragment of normative English constructible in the enrichment  $L(D1-D7, D8-D8+k)$  of  $L$ , as described above. Let  $\Phi$  still be the set of sentences of  $FNE_k$ ; thus, all members of  $\Phi$  are English sentences (some of them, though, less than fully elegant from the stylistic point of view). Then, by the *Natural Deontic Logic over  $FNE_k$*  we shall mean a certain subset  $NDL$  of  $\Phi$ , which is vaguely characterized as follows:

$NDL$  = the set of sentences of  $FNE_k$  which are *logically valid* or *logically correct* (or, if you have no qualms about applying that notion to normative sentences: *logically true*).

Suppose, at least for the time being and for the sake of argument, that this admittedly vague characterization of  $NDL$  makes ‘some reasonable’ sense to you and to me. Let  $A$  be any sentence in English. We then have the following immediate result:

$A \in NDL$  iff  $A$  belongs to  $\Phi$  and  $A$  is logically valid.

Writing ‘ $\Vdash A$ ’ for ‘ $A$  is logically valid’, we may alternatively express the result as a set-theoretical identity:

$$NDL = \{A \in \Phi : \Vdash A\}$$

Some further notions can now be defined:

DEFINITION 9 (Logical consequence, inconsistency and consistency in NDL). Let  $\Gamma$  be a *set* of sentences in  $\Phi$ , and let  $A$  be a *sentence* in  $\Phi$ . Then we say:

- (i)  $A$  is a *logical consequence in NDL of  $\Gamma$*  (in symbols:  $\Gamma \Vdash_{NDL} A$ ) iff there are sentences  $B_1, B_2, \dots, B_n$  in  $\Gamma$ , with  $n \geq 0$ , such that

$\Vdash$  If both  $B_1$  and both  $B_2$  and ... and both  $B_{n-1}$  and  $B_n$ , then  $A$

(i.e. that sentence is to be logically valid).

- (ii)  $\Gamma$  is *inconsistent in NDL* iff there is a sentence  $B$  in  $\Phi$  such that the sentence

Both  $B$  and it is not the case that  $B$

is a logical consequence in NDL of  $\Gamma$ .

- (iii)  $\Gamma$  is *consistent in NDL* iff  $\Gamma$  is not inconsistent in NDL.

The three notions just introduced should be compared with the *proof-theoretical* concepts of derivability, inconsistency and consistency in certain formal deontic logics  $\mathcal{L}$  (see Section 10.2.1 below). Also, we observe that the usefulness of the definition above rests on our having recourse to a viable notion of *logical validity*, which is *applicable to English sentences* in  $\Phi$  and, hopefully, even to larger sets of English sentences. We now face the difficulty that such an intuitive, natural-language-oriented conception is vague, imprecise and the source of endless disputes and disagreement; in this respect it differs unfavorably from the well-defined notions of *provability* and *validity* in a specified *formal system* of deontic logic. The difficulty is interestingly illustrated by the so-called ‘paradoxes of deontic logic’ in the following way.

Let  $\mathcal{L}$  be some formal system of deontic logic over our language  $L$ , to which definitions D1–D7, D8–D8+ $k$  are added, so that the fragment FNE $_k$  and its sentence-set  $\Phi$  are available as well as the natural deontic logic NDL over that fragment. Consider the set  $\Delta$  of all sentences in  $\Phi$  whose  $t$ -translations are provable in  $\mathcal{L}$ ; thus

$$\Delta = \{A \in \Phi : t(A) \text{ is provable in } \mathcal{L}\}.$$

Now, suppose we start to compare NDL and  $\Delta$ . If we find that they are identical, then the notion of provability in  $\mathcal{L}$  ‘matches perfectly’ the intuitive conception of logical validity determining NDL and everything is fine; there is no clash between the well-defined formal notion and the intuitive conception. This possibility is not very likely to arise. It is much more likely that we face one of the following possibilities:



- (I) We find a sentence in  $\Phi$  which is such that (a) it is *not* in NDL because we feel that it is *not* logically valid, but (b) its  $t$ -translation is indeed provable in  $\mathcal{L}$  (which is usually easy to verify); so the sentence is in  $\Delta$ . Then we have a clash between the formal notion of provability in  $\mathcal{L}$  and the intuitive concept of validity of a kind I call below failure of *right-to-left adequacy*. Roughly speaking, the clash is due to the fact that  $\mathcal{L}$  sanctions *more* English sentences as valid than our ‘logical intuition’ is willing to accept as members of NDL. Most deontic paradoxes will be seen to be of this kind in subsequent sections.
- (II) We find a sentence in  $\Phi$  which is such that (a) it is logically valid from an intuitive viewpoint and hence is in NDL, but (b) its  $t$ -translation fails to be provable in  $\mathcal{L}$  (as is usually easy to verify); so the sentence is *not* in  $\Delta$ . This situation illustrates an opposite sort of clash which I call below failure of *left-to-right adequacy*. The import of such a clash is then that there are intuitive validities in NDL, which fail to be representable in the formal system  $\mathcal{L}$ . In Section 9.2.2 below I argue that the Good Samaritan paradox illustrates this failure with respect to the Smiley–Hanson systems of monadic deontic logic; admittedly, it is usually cited as an instance of the former kind of failure, just as the majority of paradoxes are.

These considerations were designed to show how the vagueness and impreciseness of an intuitive, natural-language-oriented notion of logical validity leads to ‘clashes’ when confronted with formal systems of deontic logic. In order to eliminate to some extent the vagueness from which our characterization of NDL thus suffers we propose below (Section 5.5) a criterion of validity known as the Bolzano Criterion, which will be seen ‘at work’ in Section 7. Other criteria are possible, however, as will be seen in Section 6.

Since the concepts of (failure of) *adequacy* met with in (I) and (II) above are, we suggest, highly important to any orderly discussion and even presentation of the deontic paradoxes, we shall now introduce them in a rigorous way.

#### 5.4 *Some notions of adequate translation and faithful representation*

DEFINITION 10 (Three concepts of adequacy). Let  $\mathcal{L}$ , to begin with, be any of the ten Smiley–Hanson systems of monadic deontic logic, characterized in Section 10.2 below in proof-theoretical terms. We write ‘ $\vdash_{\mathcal{L}} A$ ’ to indicate that the formal sentence  $A$  (in  $\Sigma$ ) is *provable* in  $\mathcal{L}$ . Consider the Natural Deontic Logic NDL ( $\subseteq \Phi$ ) over the English fragment  $\text{FNE}_k$  as well as the extended translation  $t$  from  $\Phi$  into  $\Sigma$ . We then say the following:

- (i)  $t$  is *left-to-right adequate* with respect to NDL and  $\mathcal{L}$ , iff, for each  $A$  in  $\Phi$ , if  $A \in \text{NDL}$  then  $\vdash_{\mathcal{L}} t(A)$ .
- (ii)  $t$  is *right-to-left adequate* with respect to NDL and  $\mathcal{L}$ , iff, for each  $A$  in  $\Phi$ , if  $A \notin \text{NDL}$  then  $\not\vdash_{\mathcal{L}} t(A)$ ; where ' $\not\vdash_{\mathcal{L}}$ ' means 'not provable in  $\mathcal{L}$ '.
- (iii)  $t$  is *fully adequate* with respect to NDL and  $\mathcal{L}$ , iff, for each  $A$  in  $\Phi$ ,  $A \in \text{NDL}$  if and only if  $\vdash_{\mathcal{L}} t(A)$ .

REMARK 11.

- (a) Clearly,  $t$  is fully adequate w.r.t. NDL and  $\mathcal{L}$ , just in case  $t$  is both left-to-right and right-to-left adequate w.r.t. NDL and  $\mathcal{L}$ . Again, in order to show that  $t$  is not fully adequate w.r.t. NDL and  $\mathcal{L}$ , it is enough to show that *either*  $t$  fails to be left-to-right adequate *or*  $t$  fails to be right-to-left adequate w.r.t. NDL and  $\mathcal{L}$ .
- (b) We may use clauses(i)–(iii) in the present definition as a basis for introducing the following notions of *faithful representation*, applicable to  $\mathcal{L}$ :
  - (i')  $\mathcal{L}$  is a *left-to-right faithful representation* of NDL *under*  $t$ , iff,  $t$  is left-to-right adequate w.r.t. NDL and  $\mathcal{L}$ .
  - (ii')  $\mathcal{L}$  is a *right-to-left faithful representation* of NDL *under*  $t$ , iff,  $t$  is right-to-left adequate w.r.t. NDL and  $\mathcal{L}$ .
  - (iii')  $\mathcal{L}$  is a *faithful representation* of NDL *under*  $t$ , iff,  $t$  is fully adequate w.r.t. NDL and  $\mathcal{L}$ .

### 5.5 *On the Bolzano criterion of logical validity for sentences in natural languages*

Consider the fragment  $\text{FNE}_k$  and its sentence-set  $\Phi$ . As applied to members of  $\Phi$ , what I shall call the Bolzano criterion of logical validity asserts the following:

- BCLV. Let  $A \in \Phi$ . Then,  $A$  is *logically valid* ( $\Vdash A$ ) iff (i)  $A$  is true, and (ii) every result of uniformly substituting a sentence in  $\Phi$  for any *basic* sentence in  $A$  is true as well.

For examples of this criterion 'in use', see Section 7 below.

The present version of the Bolzano criterion is, of course, a bit restrictive, since it only applies to members of  $\Phi$ . Inspired by Quine [1963], Føllesdal and Hilpinen [1971, p. 1] suggest a less restrictive version:

A deontic sentence is a truth of deontic logic if it is true and remains true for all variations of its non-logical and non-deontic words (that is, expressions which are not logical or deontic words).

Again, Kanger [1957, p. 50] gives the following version of the Bolzano criterion:

By a logically true statement we understand a statement  $A$  such that the result of generalizing all extralogical constants in  $A$  is true.

If we assume all deontic words in the sense of Føllesdal and Hilpinen to be logical constants in the sense of Kanger, we could perhaps prove their respective versions of the Bolzano criterion to be equivalent.

For our purposes in the following Sections, however, the version BCLV given above should hopefully turn out to be sufficient.

## 6 ALF ROSS'S PARADOX

Consider the English sentences:

- (0) You post the letter.
- (1) You burn the letter.

Consider also the enrichment  $L(D1-D7, D8, D9)$  of  $L$ , where  $D8$  and  $D9$  are as follows:

- (D8) You post the letter = df  $p$ .
- (D9) You burn the letter = df  $p_1$ .

Again, let  $FNE_1$  be the fragment of normative English constructible in that enrichment, where, specifically,  $Bas = \{(0), (1)\}$  and  $\Phi$  = the smallest superset of  $Bas$  closed under our seven English connectives defined in  $D1-D7$ . The translation  $t$  induced by the series  $D1-D9$  is then defined as in Section 5.2 above, where, in particular, we have the following clauses in the recursion basis:

- (i)  $t(\text{'You post the letter'}) = t((0)) = p$ .
- (i<sub>1</sub>)  $t(\text{'You burn the letter'}) = t((1)) = p_1$ .

The Alf Ross paradox, first presented in Ross [1944], is to be taken, we suggest, as an argument against  $t$  being fully adequate with respect to the natural deontic logic  $NDL$  over  $FNE_k$  and *any* Smiley–Hanson system  $\mathcal{L}$  of monadic deontic logic. In order to state the argument let us first consider the sentence

(1.1) It is obligatory that either you post the letter or you burn the letter which is in  $\Phi$ ; a nicer *stylistic variant* of (1.1) in the sense of Kalish and Montague [1964, pp. 10f.] is perhaps the following:

(1.2) You ought to post the letter or burn it

Furthermore, we consider the sentences:

(0.1) It is obligatory that you post the letter,

(0.2) If it is obligatory that you post the letter then it is obligatory that either you post the letter or you burn the letter,

which are both in  $\Phi$ . Note that (0.2) is the conditional having (0.1) as its antecedent and (1.1) as its consequent.

Now, the ‘paradox’ starts out with the following claim, which we shall treat as an *hypothesis* in the proof-theoretical sense by so indicating its status to the right:

1. The sentence (1.1) is not a logical ‘claim’ or hypothesis consequence in NDL of the unit set of (0.1).

In symbols:  $\{(0.1)\} \not\|_{NDL} (1.1)$

By our definition of logical consequence in NDL and the fact that (0.1) is in  $\Phi$  and is the only sentence in  $\{(0.1)\}$ , we then obtain from the hypothesis 1:

2. (0.2) is not logically valid. from 1 by the definition of  $\|_{NDL}$  and the fact etc.

In symbols:  $\not\vdash (0.2)$ .

where we indicate to the right how this line 2 is obtained. Hence:

3. (0.2)  $\notin$  NDL from 2 by the definition of NDL.

Again, let  $\mathcal{L}$  be *any* of the Smiley–Hanson systems of monadic deontic logic. The following is a demonstration that the L-sentence  $Op \rightarrow O(p \vee p_1)$  is provable in  $\mathcal{L}$ ; where we write ‘ $\vdash_{\mathcal{L}}$ ’ for ‘provable in  $\mathcal{L}$ ’:

4.  $\vdash_{\mathcal{L}} p \rightarrow (p \vee p_1)$  since all tautologies over L are provable in  $\mathcal{L}$  by virtue of axiom schema A0.

5.  $\vdash_{\mathcal{L}} Op \rightarrow O(p \vee p_1)$  from 4 by the fact that the set of  $\mathcal{L}$ -provable L-sentences is closed under the rule of inference

$$\frac{A \rightarrow B}{OA \rightarrow OB}$$

Here, to say that the set of  $\mathcal{L}$ -provable L-sentences is closed under that rule of inference means that for all L-sentences  $A, B$ : if  $\vdash_{\mathcal{L}} A \rightarrow B$ , then  $\vdash_{\mathcal{L}} OA \rightarrow OB$ .

We continue the argument by making the following straightforward observation:

6.  $t((0.2)) = Op \rightarrow O(p \vee p_1)$  by the definition of  $t$ , clauses (i),(i<sub>1</sub>),(iv),(vii) and (v).

Therefore:

7.  $\vdash_{\mathcal{L}} t((0.2))$  from 5 and 6 by the logic of  $=$ .
8. There is a sentence  $A$  in  $\Phi$ , viz. from 3 and 7 by adjunction and (0.2), such that  $A \notin \text{NDL}$  but existential generalization.  
 $\vdash_{\mathcal{L}} t(A)$
9.  $t$  is not right-to-left adequate with respect to NDL and  $\mathcal{L}$  from 8 by the definition of right-to-left adequacy.
10.  $t$  is not fully adequate w.r.t. NDL and  $\mathcal{L}$  from 9 by the definition of full adequacy.

Thus, on the basis of the hypothesis 1, we have used the conditional sentence (0.2) to show that  $t$  fails to be right-to-left, and hence fully, adequate w.r.t. NDL and any Smiley–Hanson system  $\mathcal{L}$  of monadic deontic logic. In other words, no such system is a faithful representation of NDL under  $t$ . Before embarking on a discussion of this argument, we stick in a little exercise:

#### EXERCISE 12.

- (i) Prove that the set of  $\mathcal{L}$ -provable L-sentences is closed under the rule  $A \rightarrow B/OA \rightarrow OB$ , for any of the ten Smiley–Hanson systems  $\mathcal{L}$ , as described in Section 10 below!
- (ii) Give a rigorous proof in full detail of line 6 above!

Let us now try to assess the above argument. Clearly, the proof of lines 9 and 10 rests, or is conditional, on the hypothesis 1, which is to the effect that (1.1) is not a logical consequence (in NDL) of (the unit set of) (0.1). What motivation or reason could then be given for this claim? The following line of thought is indicated and discussed by Wedberg [1969, p. 217], and by Hansson [1969, p. 383].

A way for the addressee of (1.1) to *obey* the command expressed by it is surely that he burns the letter. By so doing, however, he does not obey the

command expressed by (0.1), which he even positively *disobeys*. So there is a way of obeying (1.1), which is at the same time a way of disobeying (0.1) and, hence, a way of *not* obeying it. Therefore, (1.1) cannot be a logical consequence of (0.1). (The following terminological replacements are all right with me: ‘norm’ for ‘command’, ‘satisfy’ or ‘fulfil’ for ‘obey’, ‘dissatisfy’ or ‘violate’ for ‘disobey’.)

As is clearly enough indicated by the aforementioned authors, this argument is rather confused. I would like to add the following point to their valuable discussion. The criterion of logical consequence to which the argument tacitly appeals appears to be this:

LC0. Let  $A, B$  be any command-expressing sentences in English. Then,  $B$  is a logical consequence of  $A$  iff every way of obeying  $B$  is a way of obeying  $A$ .

Putting (1.1) =  $B$  and (0.1) =  $A$  in LC0, we indeed arrive at the strange result under debate. But hasn’t LC0 got things turned upside down? According to criteria in terms of obedience or satisfaction (suggested, e.g. by Von Wright [1955] and Rescher [1966]), we should rather have something like:

LC1.  $B$  is a logical consequence of  $A$  iff every way of obeying  $A$  (= the *implicans*) is a way of obeying  $B$  (= the *implicatum*).

Using LC1 in the place of LC0, we cannot derive hypothesis 1 any longer. On the contrary, using LC1, we have that (1.1) is indeed a logical consequence of (0.1), since every way of posting the letter is a way of either-posting-or-burning it.

Are we then entitled to dismiss the Ross Paradox on the basis of having removed one basic confusion that seems to underlie it? I don’t think so, nor does, e.g. Von Wright [1968, pp. 21 ff], where an interesting argument in favor of hypothesis 1 is indicated, having the form of a *reductio ad absurdum* of its negation, which is to the effect that (1.1) *is* a logical consequence of (0.1). Von Wright’s *reductio* is presented and discussed in some detail in Åqvist [1987, Section 5]. Here we just call the reader’s attention to some main conclusions emerging from that discussion.

- (i) The argument of Von Wright [1968, pp. 21 ff], is seen to depend crucially on a so-called principle of free choice permission, the status and acceptability of which has since been the object of a still lively and intensive debate. We mention the following contributions from the literature: Von Wright [1971], Føllesdal and Hilpinen [1971], Kamp [1973; 1979], Lewis [1978], Hilpinen [1979; 1981] as well as Nute [1981].
- (ii) Although the principle of free choice permission fails to be valid for the weak notion of permission reflected by the ‘standard’ P-operator

in normal monadic Von-Wright-type deontic logics (Section 4 above), there remains the possibility of defining notions of *strong* permission for which that principle is indeed valid — after all, permissive phrases are well known to be ambiguous in ordinary discourse. An interesting attempt to define such a notion of strong permission was made by Von Wright [1971]; the idea is based on the Andersonian reduction of deontic logic to alethic modal logic with a propositional constant. We also mention that Anderson [1968] uses the very same idea to develop what he calls *eubouliatic logic*, in the sense of a logic of *prudence*, *safety*, *risk* and related concepts of a decision-theoretic brand.

- (iii) Contrary to the view of its originator, the Alf Ross paradox does not seem to be a serious threat to the very possibility of constructing a viable deontic logic. But it usefully directs our attention to the ambiguity of normative phrases in natural language as a possible source of error and confusion — in viable deontic logics we should be able to express, to do justice to, and to pinpoint such ambiguities. For this reason I agree with Von Wright in claiming that the puzzle deserves serious consideration.

## 7 ARTHUR N. PRIOR'S PARADOXES OF DERIVED OBLIGATION (‘COMMITMENT’)

Consider the English sentences:

- (2) John Doe impregnates Suzy Mae.  
(3) John Doe marries Suzy Mae.

as well as the enrichment  $L(D1-D11)$  of  $L$ , where D10 and D11 as follows:

- (D10) John Doe impregnates Suzy Mae =df  $p_2$ .  
(D11) John Doe marries Suzy Mae =df  $p_3$ .

We let  $FNE_3$  be the fragment of normative English constructible in  $L(D1-D11)$ , where Bas contains (2) and (3) as new members and where  $\Phi$  is defined as usual. Fresh clauses in the basis of the recursive definition of the translation  $t$ :

- (i<sub>2</sub>)  $t(\text{‘John Doe impregnates Suzy Mae’}) = p_2$ .  
(i<sub>3</sub>)  $t(\text{‘John Doe marries Suzy Mae’}) = p_3$ .

The paradoxes of commitment, or derived obligation, go back at least to Prior [1954] and can be viewed as arguments against  $t$  being fully adequate with respect to NDL and any Smiley–Hanson system  $\mathcal{L}$  of monadic deontic logic. Consider the sentences:

- (2.1) If it is not the case that John Doe impregnates Suzy Mae then if John Doe impregnates Suzy Mae then it is obligatory that John Doe marries Suzy Mae.
- (2.2) If it is obligatory that John Doe marries Suzy Mae then if John Doe impregnates Suzy Mae then it is obligatory that John Doe marries Suzy Mae.
- (2.3) If it is obligatory that it is not the case that John Doe impregnates Suzy Mae then it is obligatory that if John Doe impregnates Suzy Mae then John Doe marries Suzy Mae.
- (2.4) If it is obligatory that John Doe marries Suzy Mae then it is obligatory that if John Doe impregnates Suzy Mae then John Doe marries Suzy Mae.

Although they are all in  $\Phi$ , the sentences (2.1)–(2.4) give a somewhat queer impression and may be difficult to understand. So here are their  $t$ -translations into L:

$$t((2.1)) \neg p_2 \rightarrow (p_2 \rightarrow Op_3).$$

$$t((2.2)) Op_3 \rightarrow (p_2 \rightarrow Op_3).$$

$$t((2.3)) O\neg p_2 \rightarrow O(p_2 \rightarrow p_3).$$

$$t((2.4)) Op_3 \rightarrow O(p_2 \rightarrow p_3).$$

We now state four paradoxes of commitment in ‘one fell swoop’:

1. None of the sentences (2.*i*), for ‘claim’ or hypothesis  $i = 1, 2, 3, 4$  are logically valid.  
In symbols:  $\not\vdash(2.i)$ .

Now, let  $\mathcal{L}$  be any of the Smiley–Hanson systems of Monadic Deontic Logic. Then:



- |    |   |  |
|----|---|--|
| 2. | $\vdash_{\mathcal{L}} t((2.i))$ , for all $i = 1, \dots, 4$   | exercise   |
| 3. | There are sentences $A_i$ (with $i = 1, \dots, 4$ ) in $\Phi$ , viz. (2.i), such that $A_i \notin \text{NDL}$ but $\vdash_{\mathcal{L}} t(A_i)$ | from 1 and 2 by the definition of NDL, adjunction and existential generalization |
| 4. | $t$ is not right-to-left adequate with respect to NDL and $\mathcal{L}$   | from 3 by the definition of right-to-left adequacy                               |
| 5. | $t$ is not fully adequate w.r.t. NDL and $\mathcal{L}$  | from 4 by the definition of full adequacy  |

Before trying to assess this argument, we stick in a little exercise:

EXERCISE 13.

- (i) Prove assertion 2 in the above argument!
- (ii) Find more idiomatic stylistic variants of the English (?) sentences (2.i), for  $i = 1, 2, 3, 4$ !

What motivation could reasonably be given for the claim 1 in this argument? I suggest the following: the common consequent of (2.1) and (2.2) as well as that of (2.3) and (2.4) may both be said to involve the notion of *commitment* or *conditional* ('derived') *obligation* in such a way that

(2.5) Impregnating Suzy Mae commits John to marrying her

is a stylistic variant (in the sense of Kalish and Montague [1964]) of both these common consequents. Let us now consider four cases in turn, with a view to illustrate the Bolzano criterion of logical validity (see above Section 5.5).

*Case I.* Suppose that (2.1) is logically valid. Then, by the Bolzano criterion, not only is (2.1) itself true, but so is every result of uniformly substituting a sentence of the fragment  $\text{FNE}_3$  for any *basic* sentence in (2.1). Therefore, the following sentence in  $\Phi$  must be true:

- (2.1.1) If it is not the case that John Doe impregnates Suzy Mae, then if John Doe impregnates Suzy Mae then it is obligatory that *it is not the case that* John Doe marries Suzy Mae

the  $t$ -translation of which is

$$t((2.1.1)) = \neg p_2 \rightarrow (p_2 \rightarrow O\neg p_3).$$

But this result is unpalatable for the following reason: on the basis of it we may infer from the mere fact that John does not impregnate Suzy that impregnating her commits John both to the act of marrying her as well as to the act of not marrying her. There are certainly legal systems where such an inference is rejected as absurd. Hence, (2.1) cannot be logically valid.

*Case II.* Here we consider an extended fragment  $FNE_4$ , whose set  $\text{Bas}$  contains the new sentence

(4) John Doe kills Suzy Mae

which is introduced by a definition D12 in the obvious way so as to yield the fresh clause for  $t$ :

(i<sub>4</sub>)  $t(\text{'John Doe kills Suzy Mae'}) = p_4$ .

Now, suppose that (2.2) is logically valid. Then, by the Bolzano criterion, not only is (2.2) true itself, but so is the result of substituting (4) for (2) in (2.2). Call this result (2.2.1) and observe that

$$t((2.2.1)) = Op_3 \rightarrow (p_4 \rightarrow Op_3)$$

Again, (2.2.1) being true is a strange result, because it seemingly entitles us to say that *if* John has a duty to marry Suzy, *then*, by logic alone, he has such a duty (even) if he kills her. But, we are told by some plain man in the street, when she is dead, John cannot marry her and so is not obliged or committed to marry her. For, as the saying goes, *ought* implies *can*. Hence, to sum up, if (2.2) is logically valid, then (2.2.1) is true. But (2.2.1) is not true (for the reasons just given); therefore, by *modus tollens*, (2.2) is not logically valid.

*Case III.* Suppose that (2.3) is logically valid. Then, by the Bolzano criterion, not only is (2.3) true itself, but so is the result of inserting the word 'it is not the case that' immediately in front of (3) in (2.3). Call that result (2.3.1) and note that

$$t((2.3.1)) = O\neg p_2 \rightarrow O(p_2 \rightarrow \neg p_3).$$

The argument against (2.3.1) being true is similar to the one used in Case I: on the basis of it we may infer from the mere fact that it is forbidden for John to impregnate Suzy that impregnating her commits John both to the act of marrying her and to that of not marrying her. But this is absurd for the same reason as in Case I.

*Case IV.* The point of departure here is the assumption that (2.4) is logically valid. The refutation of that assumption proceeds (rightly or wrongly) along

the lines followed in Case II. The result analogous to (2.2.1) is called (2.4.1) where

$$t((2.4.1)) = Op_3 \rightarrow O(p_4 \rightarrow p_3).$$

We have now got an idea about what kind of counterexamples people are apt to give against the alleged logical validity of the sentences (2.i). We may then try to pin down certain patterns of ‘intuitive reaction’ to these paradoxes of commitment, which are fairly well discernible or recognizable in the vast literature on the subject. (Incidentally, they will also be seen to apply to the Alf Ross paradox.) Let me distinguish the following two tendencies: (i) the deprivation-of-counterintuitive-force tendency, and (ii) the improved-formalization tendency, and briefly comment on them in turn.

### 7.1 *The deprivation-of-counterintuitive-force tendency*

One argues as follows. Admittedly, formulations like (2.i) are ambiguous in ordinary English, so there are different ways of understanding or interpreting (2.i). Now, each of the Smiley–Hanson systems  $\mathcal{L}$  is based on a clearcut model-theoretic semantics, which provides a mathematically precise interpretation of its logical constants in terms of *truth conditions* stated relatively to certain set-theoretical structures called *models* (see Section 10.3 below). Via our definitions D1–D7 etc. generating fragments of normative English, this precise interpretation is automatically transferred to every sentence of such an English fragment. So, if we just stick to *that* interpretation of (2.i) and, most importantly, do not ‘read into’ them anything ‘beyond’ it, their counterintuitive appearance will simply vanish.

To illustrate a bit: since every ‘if ... then \_\_\_\_\_’-connective in any of (2.i) is intended to mean the same as the arrow  $\rightarrow$  of material implication, as ordinarily understood in the classical propositional calculus (on which all the Smiley–Hanson systems are based), we have the following results on the *t*-translations of (2.i):

$$\vdash_{\mathcal{L}} t((2.1)) \Leftrightarrow (\neg p_2 \rightarrow (\neg p_2 \vee Op_3))$$

$$\vdash_{\mathcal{L}} t((2.2)) \Leftrightarrow (Op_3 \rightarrow (\neg p_2 \vee Op_3))$$

$$\vdash_{\mathcal{L}} t((2.3)) \Leftrightarrow (O\neg p_2 \rightarrow O(\neg p_2 \vee p_3))$$

$$\vdash_{\mathcal{L}} t((2.4)) \Leftrightarrow (Op_3 \rightarrow O(\neg p_2 \vee p_3))$$

According to the strategy of interpretation just outlined, we are to understand the sentences (2.i) to mean *exactly* what is meant by the corresponding

right members in these equivalences according to the semantics for  $\mathcal{L}$ ; most importantly, we are *not* to read anything *more* into them. We then observe the following:

- (i) Sentences (2.1) and (2.2) as well as their *t*-translations are instances of familiar tautologies, which cannot fail to be true by virtue of the usual truth-table test. Hence, their counterintuitive force disappears as soon as this fact is grasped and strictly adhered to.
- (ii) Sentences (2.3) and (2.4) as well as their *t*-translations are not instances of any truth-table tautologies. Instead, their validity in  $\mathcal{L}$  depends on and is due to the model-theoretic truth condition for the *O*-operator (in terms of ‘possible worlds’, the relation of ‘deontic accessibility’ etc. ; see Section 10.3 below). Once this fact is grasped and strictly adhered to, their counter-intuitive force will vanish. Note here that, on the basis of the last two of the equivalences above, the paradoxical conditionals (2.3) and (2.4) both reduce to *Ross-paradoxical* sentences, with which we have already dealt.

Let us also note, finally, that the present Deprivation-of-Counterintuitive-Force Tendency is nicely illustrated, e.g. by Von Wright [1956, p. 508] and by Anderson [1956, p.185], as well as by Prior [1955, p. 224]. This tendency, however, is not the only one emerging from their valuable discussion, as we shall now see.

## 7.2 *The improved-formalization tendency*

Even if the argument just reported is successful in ‘explaining away’ the counterintuitiveness of the sentences (2.*i*), it is difficult to remain satisfied with it ‘as giving the full story of the matter’. As we said above, formulations like (2.*i*) are admittedly ambiguous in ordinary language, so the possibility of understanding their consequents as expressing some notion of *commitment* (as illustrated by (2.5)) remains an interesting fact of linguistic usage. But then the problem arises how to formalize the notion adequately, which is surely a problem in view of the difficulties pinpointed in Cases I–IV above. Let us now quickly find out what attitudes Von Wright and Anderson took on this issue.

In the second part of Von Wright [1956] we are warned against interpreting the form

$$O(A \rightarrow B)$$

to mean ‘doing *A* *commits* us (morally) to do *B*’ (p. 508). Von Wright says that if we do so

... then a ‘paradox’ instantly arises. For then we should have to say that a forbidden act commits us to any other act (whether obligatory, permitted, or forbidden). And this, obviously, conflicts with our ‘intuitions’ in the matter.

Perhaps we can say that Von Wright’s reason for his *caveat* amounts essentially to Case III. Again, Anderson [1956, p. 185], expresses a similar warning: “... we should be wary in interpreting  $OCpq$  as meaning ‘ $p$  commits us to  $q$ ’.”

Being thus dissatisfied with  $O(A \rightarrow B)$  (and with  $A \rightarrow OB$ , we may add in the case of Anderson) as adequately reflecting the everyday notion of commitment, our authors are naturally led to look for *new ways of formalizing* that notion. They tried different approaches, though, which we are now going to describe.

Von Wright [1956] introduces a new *primitive* symbol  $P(p/c)$ , to be read as:  $p$  is permitted under conditions  $c$ . He then defines  $O(p/c)$  as  $\neg P(\neg p/c)$ , for which he suggests the reading:  $p$  is obligatory under conditions  $c$ , or:  $c$  commits us to (do)  $p$ . Furthermore, he gives two axioms for the new primitive operator, on the basis of which, to the best of my knowledge, the first known system of dyadic deontic logic was developed. Further additions to and refinements of the system were suggested in Von Wright [1964] and [1965]. This idea of using *binary* (i.e. two-place) *primitives*, expressing conditional or ‘relative’ permission, prohibition and obligation, initiated a very fruitful line of research in the history of modern deontic logic. In effect, dyadic deontic logic as originating with Von Wright [1956], can be said to have dominated recent work in the field up to this date. For the moment, we just remind the reader of the following contributions: Rescher [1958; 1962], Powers [1967], Danielsson [1968], Hansson [1969], Segerberg [1971], Føllesdal and Hilpinen [1971], Van Fraassen [1972], Von Kutschera [1973; 1974], Lewis [1974], Chellas [1974] and Spohn [1975].

Let us now turn to Anderson. He was able to make a different suggestion, because he had at his disposal a more powerful logical apparatus on which he wished to base the theory of deontic notions, including commitment: that of alethic modal logic with a propositional constant  $S$  symbolizing some penalty or sanction. So already in Anderson [1956] we find him suggesting (p. 185) that an alternative (to  $OCpq$ ) candidate for the formal analogue of commitment is  $C'pOq$ : ‘ $p$  entails that  $q$  is obligatory’. Here ‘ $C'$ ’ denotes *strict* implication; using the symbolism adopted in the present essay, we write in the place of  $C'pOq$ :

$$\Box(A \rightarrow OB).$$

In Anderson [1959], written as a reaction to the Rescher [1958] attempt to elaborate further the Von Wright [1956] proposal about dyadic deontic

logic, Anderson repeats his suggestion and, highly importantly in my opinion, proceeds to lay down a number of *adequacy criteria* for a theory of commitment (or a viable analysis of that notion).

Let us pause for a while to consider the import of the Anderson proposal. To start with, let us replace the Von Wright [1956] dyadic notation  $O(p/c)$  with the one adopted in the present essay, viz.  $O_B A$  to be read: ‘if  $B$  then it is obligatory that  $A$ ’ or: ‘ $B$  commits us to  $A$ ’. For this notation, see Section 15 below. The Anderson proposal is then to the following effect: consider the following

DEFINITION 14.

$$\text{Def}^{\text{com}}. O_B A = \text{df } \Box(B \rightarrow OA).$$

Then, add  $\text{Def}^{\text{com}}$  to any suitable system  $\mathcal{K}$  of alethic modal logic with the constant  $S$  (or with Kanger’s  $Q$ ), which already has definitions of ‘standard monadic’ obligation and permission (see below Sections 12 and 13).

Suppose that  $\mathcal{K}$  is well determined: we are then able to investigate the logic of commitment (i.e. the laws governing locutions of the form  $O_B A$ ) within  $\mathcal{K}$  supplemented with  $\text{Def}^{\text{com}}$ .

Having now broadly outlined the respective approaches of Von Wright and Anderson to the problem of formalizing a ‘reasonable’ concept of commitment or conditional obligation, we have to face the obvious questions: What expectations are to guide us in this enterprise? What properties do we expect that concept to have (and not to have)? I call this the problem of *adequacy criteria* and will devote a special section to it.

### 7.3 Adequacy criteria for a theory of commitment or conditional obligation

Consider the language of Dyadic Deontic Logic as described in Section 17 below. Its set of sentences is called  $\Sigma_{0,N}^2$  and is such that, for any  $A, B$  in  $\Sigma_{0,N}^2$ ,  $O_B A$  and  $P_B A$  are in  $\Sigma_{0,N}^2$  as well. We now adhere to the just adopted reading of  $O_B A$  as ‘ $B$  commits us to  $A$ ’. We are looking for a theory  $\mathcal{L}$  over this language in the sense of a proper subset of  $\Sigma_{0,N}^2$  (why *proper?*), which is to serve as a viable and plausible logic of commitment. Let provability and unprovability in  $\mathcal{L}$  be denoted by  $\vdash_{\mathcal{L}}$  and  $\not\vdash_{\mathcal{L}}$ , respectively.

Adequacy criteria for  $\mathcal{L}$  may now be divided into (i) those to the effect that certain sentence schemata are to be *unprovable* in  $\mathcal{L}$ , and (ii) those to the effect that certain schemata are to be *provable* in  $\mathcal{L}$ . Let us begin by giving some examples of the first category.

In the light of our discussion of the sentences (2.i) ( $i = 1, \dots, 4$ ) we consider certain generalized sentence schemata arising from the translations  $t((2.i))$  as follows. First, generalize  $p_2$  to an arbitrary formal sentence  $A$

and  $p_3$  to an arbitrary formal sentence  $B$ . We then obtain, e.g. from  $t((2.1))$  the schema

$$\neg A \rightarrow (A \rightarrow OB)$$

Secondly, replace the consequent in this schema by the appropriate formal analogue of commitment available in our language of dyadic deontic logic, viz.  $O_A B$ . We thus obtain in our present example:

$$\neg A \rightarrow O_A B$$

Performing the same operations with all the  $t(2.i)$ , we obtain three schemata which, in the light of the difficult Cases I–IV, we expect all to be unprovable in  $\mathcal{L}$ . In other words, we expect  $\mathcal{L}$  to satisfy the following three adequacy criteria belonging to the unprovability category:

- (C1)  $\not\vdash_{\mathcal{L}} \neg A \rightarrow O_A B$ .
- (C2)  $\not\vdash_{\mathcal{L}} OB \rightarrow O_A B$ .
- (C3)  $\not\vdash_{\mathcal{L}} O\neg A \rightarrow O_A B$ .

where the monadic, i.e. without subscript,  $O$ -operator is defined by

$$OB = \text{df } O_{\top} B$$

where the constant  $\top$ , known as *verum*, denotes some arbitrary tautologous condition. This proposal for handling monadic or ‘absolute’ obligations in dyadic deontic logic was made already by Von Wright [1956, p. 509].

I shall now make two observations.

- (i) All systems of Dyadic Deontic Logic dealt with in Sections 17–23 satisfy the criteria C1–C3. In particular, this is true of the strongly normal “core” system **G**. (see Section 23 below).
- (ii) Von Wright [1956] claims that his new deontic logic satisfies C2 and C3. And Anderson [1959] explicitly adopts C3 (writing  $Fp$  for  $O\neg p$ , where  $F$  means ‘it is forbidden that’).

#### EXERCISE 15.

- (1) Prove assertion (i) just made above!
- (2) Suppose Von Wright were to *prove* the claim reported in (ii) above; how should he go about doing so?

As a further adequacy criterion for  $\mathcal{L}$ , we consider

$$(C4) \quad \not\vdash_{\mathcal{L}} O_B A \rightarrow O_{B \wedge C} A.$$

A classical intuitive counterexample to the validity of the schema in C4 is provided by the following instance of it:

$$O_{p_2 p_3} \rightarrow O_{p_2 \wedge p_4 p_3}$$

which is read: if impregnating Suzy Mae commits John Doe to marrying her, then both-impregnating-and-killing-her commits John to marrying her. This is absurd, though, since the antecedent may be accepted as true while the consequent is rejected as false. This counterexample, or argument for C4, apparently originated with Powers [1967] and is elaborated by various subsequent writers, notably Danielsson [1968, p. 66 f], Hansson [1969, p. 392], Van Fraassen [1972, p. 418 f] and Van Eck [1981, p. 8]. The objectionable schema in C4 is pertinently called a *principle of augmentation* by Chellas [1974, p. 31].

EXERCISE 16. (after Danielsson [1968, p.67]).

- (1) Suppose that our desired theory  $\mathcal{L}$  satisfies one of the following adequacy criteria belonging to the provability category:

$$(C5) \quad \vdash_{\mathcal{L}} O_B A \Leftrightarrow O(B \rightarrow A).$$

$$(C6) \quad \vdash_{\mathcal{L}} O_B A \Leftrightarrow (B \rightarrow OA).$$

Show that in each case  $\mathcal{L}$  will violate C4! Show also that in each case  $\mathcal{L}$  violates at least one of C1–C3!

- (2) Let  $\mathcal{K}$  be a system of alethic modal logic with the constant  $S$  (or  $Q$ ) to which  $\text{Def}^{\text{com}}$  is added so that  $\mathcal{K}$  satisfies the following criterion:

$$(C7) \quad \vdash_{\mathcal{K}} O_B A \Leftrightarrow \Box(B \rightarrow OA).$$

Then, show that  $\mathcal{K}$  or, strictly speaking, its deontic fragment, violates C4!

- (3) What assumptions concerning deducibility in  $\mathcal{L}(\mathcal{K})$  have *minimally* to be made in order for your proofs to work?
- (4) Suppose that C1–C4 are accepted as reasonable adequacy criteria for  $\mathcal{L}$ . What conclusion as to the status of C5–C7 are we to draw in the light of the above results?

EXERCISE 17. Show that all systems of dyadic deontic logic considered in Sections 17–23 satisfies the criterion C4! Hint: it is enough to prove that our strongly normal system  $\mathbf{G}$  (Section 23 below) has this property.



In Anderson [1959] it is suggested that our desired theory  $\mathcal{L}$  of commitment is to meet the following six adequacy criteria, which are all to the effect that certain schemata should be *provable* in  $\mathcal{L}$  (and thus belong to our second category):

$$(C8) \quad \vdash_{\mathcal{L}}(A \wedge O_A B) \rightarrow OB.$$

$$(C9) \quad \vdash_{\mathcal{L}}(OA \wedge O_A B) \rightarrow OB.$$

$$(C10) \quad \vdash_{\mathcal{L}}(PA \wedge O_A B) \rightarrow PB.$$

$$(C11) \quad \vdash_{\mathcal{L}}(O_A B \wedge O_B C) \rightarrow O_A C.$$

$$(C12) \quad \vdash_{\mathcal{L}}O_A B \rightarrow O(A \rightarrow B).$$

$$(C13) \quad \vdash_{\mathcal{L}}O_{\neg A} A \rightarrow OA.$$

#### EXERCISE 18.

- (1) Show that the system  $\mathbf{G}$ , as described in Section 23 below, satisfies the criteria C9,C10,C12 (which is a weakened version of C5) and C13!
- (2) Consider any system of dyadic deontic logic which is an axiomatic extension of  $\mathbf{O}_{dy}\mathbf{S5}^N$  (see Section 18 below) and is dealt with in Part VI. Determine, for each of the four criteria just mentioned, which is the *weakest* system satisfying that criterion!
- (3) Consider again our strongly normal system  $\mathbf{G}$  (Section 23). Show that it neither satisfies C8 nor C11!

In Åqvist [1963] the following criticism was levelled against the Andersonian set C8–C13 of criteria. (On page 25, note 2 of that paper, the gist of the argument was credited to T. Dahlquist.) Suppose that  $\mathcal{L}$  satisfies both C8 and C9. Then, provided only that  $\mathcal{L}$  possesses a certain minimal deductive power,  $\mathcal{L}$  will satisfy the following condition C14:

$$(C14) \quad \vdash_{\mathcal{L}}(O_A B \wedge O_{\neg A} \neg B \wedge OA \wedge \neg A) \rightarrow (OB \wedge O\neg B).$$

Suppose further that  $\mathcal{L}$  meets this condition:

$$(C15) \quad \vdash_{\mathcal{L}}\neg(OB \wedge O\neg B).$$

Then, as we easily show using *modus tollens*,  $\mathcal{L}$  will also meet this condition:

$$(C16) \quad \vdash_{\mathcal{L}}\neg(O_A B \wedge O_{\neg A} \neg B \wedge OA \wedge \neg A).$$

The refutability in  $\mathcal{L}$  of the schema inside the negation-sign in C16 now amounts to this: whatever be meant by  $A$  and  $B$  here, the following conjunction (or set) of assumptions is logically impossible or provably false

in  $\mathcal{L}$ :

$A$  commits us to  $B$   
(not:  $A$ ) commits us to (not  $B$ )

*and*

it is obligatory that  $A$   
it is not the case that  $A$ .

But, the argument goes on, this result is counterintuitive, because we can find  $A$  and  $B$ , as well as English readings of them, for which that conjunction (or set) appears to be perfectly possible or consistent logically. A famous case in point is the so-called Chisholm contrary-to-duty imperative paradox, first stated in Chisholm [1963] and later discussed by a number of authors, e.g. Von Wright [1964; 1965], Sellars [1967], Åqvist [1966; 1967], Powers [1967], Hansson [1969], Føllesdal and Hilpinen [1971], Mott [1973], al-Hibri [1978], Tomberlin [1981], Van Eck [1981], and presumably several others. We now address ourselves to that puzzle.

## 8 RODERICK M. CHISHOLM'S CONTRARY-TO-DUTY IMPERATIVE PARADOX

Several versions of the puzzle are known in the literature. Following Van Eck [1981] I shall consider a Suzy Mae version of it, which explicitly involves the notion of commitment:

- I. It ought to be that John does not impregnate Suzy Mae.
- II. Not-impregnating Suzy Mae commits John to not marrying her.
- III. Impregnating Suzy Mae commits John to marrying her.
- IV. John impregnates Suzy Mae.

Let  $\mathcal{C} = \{I, II, III, IV\}$ . We note that the set  $\mathcal{C}$  is, from an intuitive standpoint, both *consistent* in the sense that no contradiction follows from it and *non-redundant* in the sense that none of its members follows from the remainder of the set. We then expect any adequate formalization of I–IV to *preserve both these properties*. Let us call this adequacy criterion our *requirement of consistency and non-redundancy*.

We now consider three attempts to formalize the sentences I–IV, using only the resources of Monadic Deontic Logic.

*First attempt:*

- Ia.  $O\neg p_2$ .
- IIa.  $O(\neg p_2 \rightarrow \neg p_3)$ .
- IIIa.  $O(p_2 \rightarrow p_3)$ .
- IVa.  $p_2$ .

*Objection:* Let  $\mathcal{L}$  be any of the ten Smiley–Hanson systems. We have that IIIa is  $\mathcal{L}$ -derivable ( $\mathcal{L}$ -deducible) from Ia, although III does not follow logically from I. Hence, the non-redundancy part of our requirement is violated by this proposal.

*Second attempt:*

- Ib(=Ia).  $O\neg p_2$ .
- IIb.  $\neg p_2 \rightarrow Op_3$ .
- IIIb.  $p_2 \rightarrow Op_3$ .
- IVb(=IVa).  $p_2$ .

*Objection:* IIb is  $\mathcal{L}$ -derivable from IVb, although II is not a logical consequence of IV. Therefore, non-redundancy is not preserved by this formalization either, contrary to our requirement.

*Third attempt:*

- Ic(=Ia).  $O\neg p_2$ .
- IIc(=IIa).  $O(\neg p_2 \rightarrow \neg p_3)$ .
- IIIc(=IIIb).  $p_2 \rightarrow Op_3$ .
- IVc(=IVa).  $p_2$ .

*Objection:* Let  $\mathcal{L}$  be any of the Smiley–Hanson +-systems, having the characteristic axiom schema A3:  $OA \rightarrow PA$ , which, by A1, is equivalent to

$$OA \rightarrow \neg O\neg A \quad (\text{see Section 10.2 below}).$$

We then observe that  $\{Ic, IIc\} \vdash_{\mathcal{L}} O\neg p_2$  and that  $\{IIIc, IVc\} \vdash_{\mathcal{L}} Op_3$ . Hence  $\{Ic, IIc, IIIc, IVc\} \vdash_{\mathcal{L}} \perp$ , so the present formalization fails to preserve the consistency of  $\mathcal{C}$ , contrary to our requirement.

#### EXERCISE 19.

- (i) The objections to the three attempts just considered rest on claims about derivability in all, or certain, Smiley–Hanson systems  $\mathcal{L}$ . Give *careful* proofs (in full detail) of these claims!

- (ii) What happens to the objection to the third attempt, if the restriction to the Smiley–Hanson +-systems is dropped?
- (iii) Consider the formal sentences Ia–IVa, IIb, IIIb, which are all in  $\Sigma$ . Which English sentences in  $\Phi$ , i.e. the set of sentences of the fragment  $\text{FNE}_3$  (see Section 7 above), are such that these formal sentences are the  $t$ -translations of the latter English sentences, respectively? What relation do those English sentences bear to I–IV above?

Let us now pass to consideration of certain attempts to formalize the sentences I–IV, using the stronger resources of dyadic deontic logic, i.e. the language described in Section 17 below and its set of sentences  $\Sigma_{0,N}^2$ .

*Fourth attempt:*

$$\begin{aligned} \text{Id}(=\text{Ia}). & \quad O\neg p_2. \\ \text{IId}(=\text{IIa}). & \quad O(\neg p_2 \rightarrow \neg p_3). \\ \text{IIIId}. & \quad O_{p_2} p_3. \\ \text{IVd}(=\text{IVa}). & \quad p_2. \end{aligned}$$

*Fifth attempt:*

$$\begin{aligned} \text{Ie}(=\text{Ia}). & \quad O\neg p_2. \\ \text{IIe}. & \quad O_{\neg p_2} \neg p_3. \\ \text{IIIe}(=\text{IIIId}). & \quad O_{p_2} p_3. \\ \text{IVe}(=\text{IVa}). & \quad p_2. \end{aligned}$$

The fourth and fifth attempts to deal with Chisholm’s puzzle give rise to the following result:

**THEOREM 20** (Contrary-to-duty imperative paradox). *Let  $\mathcal{L}$  be any of the systems of dyadic deontic logic presented in Sections 15–23. Let*

$$\begin{aligned} \mathcal{C}d &= \{\text{Id}, \text{IId}, \text{IIIId}, \text{IVd}\} \\ \mathcal{C}e &= \{\text{Ie}, \text{IIe}, \text{IIIe}, \text{IVe}\}. \end{aligned}$$

*Then:*

- (i)  $\mathcal{C}d$  is  $\mathcal{L}$ -consistent in the sense that  $\perp$  (falsum) is not  $\mathcal{L}$ -derivable from  $\mathcal{C}d$ . In symbols:  $\mathcal{C}d \not\vdash_{\mathcal{L}} \perp$ .
- (ii)  $\mathcal{C}e$  is  $\mathcal{L}$ -consistent in the same sense, i.e.  $\mathcal{C}e \not\vdash_{\mathcal{L}} \perp$ .

- (iii)  $Cd$  is  $\mathcal{L}$ -non-redundant in the sense that none of its members is  $\mathcal{L}$ -derivable from the remainder of  $Cd$ .
- (iv)  $Ce$  is  $\mathcal{L}$ -non-redundant in the same sense.

In short, the intuitive content of the theorem is to the effect that the fourth and fifth attempts both satisfy our requirement of consistency and non-redundancy with respect to any dyadic system  $\mathcal{L}$  of a certain kind.

**Proof.**[Sketch] It is enough to prove the points (i)–(iv) for the case where  $\mathcal{L}$  = our strongly normal system  $\mathbf{G}$  (why?). To begin with, let us have a look at the diagram shown in Figure 1. The meaning of this is that it represents a set  $W$  of *possible worlds* (or *situations*), consisting of four distinct members  $x, y, z, u$ , which are ranked by a binary relation of *strict preference* or *strict betterness*. That relation is represented by  $\succ$ , and the ranking order is from left to right, so that  $u$  is the *best* member of  $W$ ,  $x$  the *second best* etc. Moreover, the proposition letters  $p_2, p_3$  (read in accordance with D10 and D11, Section 7 above) are taken to be true/false at different worlds as shown by the diagram:

$p_2$  is true at  $x$  and  $y$ , but false at  $z$  and  $u$   
 $p_3$  is true at  $x$  and  $z$ , but false at  $y$  and  $u$ .

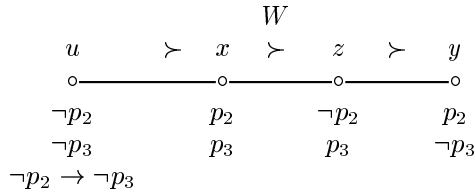


Figure 1.

Molecular Boolean compounds of these ‘atoms’ receive truth-values according to the familiar tables. But how do we handle sentences of the forms  $OA$  and  $O_BA$  (expressing ‘absolute’ and ‘conditional’ obligations, respectively)? The main suggestion embodied in the dyadic approach to deontic logic (and perhaps most clearly stated by Hansson [1969]) comes down to this. Letting  $B(A)$  be any sentence in  $\Sigma_{0,N}^2$ , we mean by a  $B(A)$ -world any world in  $W$  at which  $B(A)$  is true; then, we propose the following *truth condition* for any sentences of the form  $O_BA$ , relatively to any world  $w$  in  $W$ :

TC.  $O_BA$  is true at  $w$  iff all the best  $B$ -worlds are  $A$ -worlds.

Consider now the point  $y$  in  $W$ , i.e. the ‘*worst*’ of all possible worlds in  $W$ ’ according to the ranking  $\succ$ . For each member of  $\mathcal{C}d$  we want to figure out whether it is true at  $y$  or not:

Id? Remembering the Von Wright [1956] type definition of the monadic  $O$ -operator (Section 7.3 above)

$$OB = \text{df } O_{\top}B$$

we obtain by TC that  $O_{\top}\neg p_2$  is true at  $y$  iff all the best  $\top$ -worlds are  $\neg p_2$ -worlds. But the set of  $\top$ -worlds =  $W$  (why?) and the set of *best*  $\top$ -worlds, according to  $\succ$ , =  $\{u\}$ , i.e. the unit set of  $u$ . Now,  $\neg p_2$  is true at  $u$ , so all the best  $\top$ -worlds are  $\neg p_2$ -worlds. Hence, by TC, Id is true at  $y$ .

IId? The same kind of argument is helpful in establishing the truth at  $y$  of IId.

IIId? By TC we have that  $O_{p_2}p_3$  is true at  $y$  iff all the best  $p_2$ -worlds are  $p_3$ -worlds as well. Now the set of  $p_2$ -worlds =  $\{x, y\}$  and the set of *best*  $p_2$ -worlds =  $\{x\}$ .  $p_3$  is true at  $x$ , so all the best  $p_2$ -worlds are  $p_3$ -worlds. Hence, by TC, IIId is true at  $y$ .

IVd? By our diagram,  $p_2$  (=IVd) is true at  $y$ .

Upshot so far: every sentence in the set  $\mathcal{C}d$  is true at  $y$ .

We can now proceed to establish point (i) of our theorem. First of all, we claim that the ‘model’ pictured in Figure 1 can be used to define a *strong deontic  $H_3$ -model*  $\mathcal{U}$ , in the strict sense introduced in Section 22 below, which is such that all members of  $\mathcal{C}d$  are true at  $y$  in  $\mathcal{U}$ . (The construction, or definition of  $\mathcal{U}$  is left as an exercise to the reader.) Suppose then, contrary to (i), that  $\mathcal{C}d$  is not  $\mathbf{G}$ -consistent. By the definition of the latter notion (see already Section 10.2.1) we obtain:

$$\vdash_{\mathbf{G}} (\text{Id} \wedge \text{IId} \wedge \text{IIId} \wedge \text{IVd}) \rightarrow \perp$$

or simply:

$$\vdash_{\mathbf{G}} \neg (\text{Id} \wedge \text{IId} \wedge \text{IIId} \wedge \text{IVd})$$

which is to the effect that the *negation* of the conjunction of the members of  $\mathcal{C}d$  is provable in  $\mathbf{G}$ .

However, since that negation is provable in  $\mathbf{G}$ , then, by the Soundness Theorem for  $\mathbf{G}$  (Theorem 72 below), it is strongly deontically  $H_3$ -valid, which means in particular that it is true at  $y$  in our model  $\mathcal{U}$  just constructed. But, as all members of  $\mathcal{C}d$  are true at  $y$  in  $\mathcal{U}$ , their conjunction must be true at  $y$  as well. Contradiction. This proves point (i) of the theorem.

To deal with (ii) we only have to show that the sentence  $\text{IIe}$  in  $\mathcal{C}e$  is true at  $y$  in the intuitive model picture in Figure 1. Well,  $O_{\neg p_2} \neg p_3$  is, by TC, true at  $y$  iff all the best  $\neg p_2$ -worlds are  $\neg p_3$ -worlds as well. Now, the set of  $\neg p_2$ -worlds  $= \{u, z\}$  and the set of *best*  $\neg p_2$ -worlds  $= \{u\}$ .  $\neg p_3$  is true at  $u$  (by our diagram), so all the best  $\neg p_2$ -worlds are  $\neg p_3$ -worlds. Hence, by TC,  $\text{IIe}$  is true at  $y$ . The remainder of the proof of point (ii) parallels that of (i).

Our strategy in dealing with the non-redundancy points (iii) and (iv) is the following. Suppose we want to show that in  $\mathbf{G}$   $\text{IVd}(=p_2)$  is *independent in*  $\mathcal{C}d$  in the sense that  $\{\text{Id}, \text{IID}, \text{IIId}\} \not\vdash_G \text{IVd}$ . Suppose that we find, i.e. are able to construct, a strong deontic  $H_3$ -model  $\mathcal{U}$  and a world  $w$  in  $\mathcal{U}$  such that  $\text{Id}, \text{IID}, \text{IIId}$  as well as the *negation* of  $\text{IVd}$  are all true at  $w$  in  $\mathcal{U}$ . We then use the Soundness Theorem for  $\mathbf{G}$  to conclude (exactly how?) that  $\text{IVd}$  is not  $\mathbf{G}$ -derivable from  $\{\text{Id}, \text{IID}, \text{IIId}\}$ .

Following this strategy, the proof of (iii) and (iv) is almost routine and can be left to the reader. Just a few hints:

The case of the independence of  $\text{IVd}(=\text{IVe})$  in  $\mathcal{C}d$  and  $\mathcal{C}e$  is particularly easy: use the same model as above, but consider the point  $z$ , at which  $p_2$  is false, instead of  $y$ .

For the case of  $\text{Id}(=\text{Ie})$ : stick to  $y$  as the ‘point of evaluation’ in the above model, but change  $\succ$  in such a way that  $x$  is ranked above  $u$ .

For the cases of  $\text{IID}$  and  $\text{IIe}$ : stick to  $y$  in the original model, but assume  $p_3$  to be true at  $u$ .

And so on.

This completes the outline of a proof of the Theorem on the contrary-to-duty imperative paradox. ■

### 8.1 On the choice between the fourth and the fifth attempt

Suppose we grant that the formalizations of I–IV codified in the fourth and fifth attempts are superior to their predecessors, because they preserve *both* the intuitive consistency *and* the intuitive non-redundancy of the set  $\mathcal{C}$ . Which of  $\mathcal{C}d$  and  $\mathcal{C}e$  are we then to choose? I shall now offer an argument for taking a neutral position on this issue: it does not matter which one we choose, we may leave the choice open.

Let us ask, to begin with: as  $\mathcal{C}d$  and  $\mathcal{C}e$  differ only with respect to their second member, what is the logical relation of  $\text{IID}$  to  $\text{IIe}$ ? In answer to that question we state and prove the following result:

LEMMA 21 ( $\text{IID}$  and  $\text{IIe}$ ). *Let  $\mathcal{L}$  be any of the dyadic systems dealt with in Part VI below. Then:*

- (i) *If  $\mathcal{L}$  contains the system  $\mathbf{E}$  (see Section 22 at the end), then*

$$\vdash_{\mathcal{L}} O_{\neg p_2} \neg p_3 \rightarrow O(\neg p_2 \rightarrow \neg p_3).$$

(ii) If  $\mathcal{L}$  contains the system  $\mathbf{G}$  (Section 22 at the end), then

$$\vdash_{\mathcal{L}} P_{\top} \neg p_2 \rightarrow O(\neg p_2 \rightarrow \neg p_3) \Leftrightarrow O_{\neg p_2} \neg p_3.$$

**Proof.**

*Ad (i):*

- |  |   |
|--|---|
| 1. $O_{\neg p_2} \neg p_3$   | hypothesis  |
| 2. $N(\neg p_2 \Leftrightarrow (\top \wedge \neg p_2))$                                    | $\mathbf{E}$ contains <b>S5</b> for $N$   |
| 3. $O_{\neg p_2} \neg p_3 \Leftrightarrow O_{\top \wedge \neg p_2} \neg p_3$               | from 2 by $\alpha 0$ (Section 18), which is an axiom schema in $\mathbf{E}$ , using <i>modus ponens</i> |
| 4. $O_{\top \wedge \neg p_2} \neg p_3$   | from 1 and 3 by propositional logic   |
| 5. $O_{\top \wedge \neg p_2} \neg p_3 \rightarrow O_{\top}(\neg p_2 \rightarrow \neg p_3)$ | instance of $\alpha 2$ (Section 18), which is an axiom schema in $\mathbf{E}$                           |
| 6. $O_{\top}(\neg p_2 \rightarrow \neg p_3)$   | from 4,5 by <i>modus ponens</i>   |

The sequence 1–6 is a deduction in  $\mathbf{E}$  and hence in any  $\mathcal{L}$  of the sort under consideration. Rewriting  $O_{\top}$  as  $O$  in 6, we obtain the desired result (i) by the rule of conditional proof (or, if you like, the Deduction Theorem), which is valid in (for)  $\mathbf{E}$ , of course.

*Ad (ii):*

- |  |  |
|--|--|
| 1. $P_{\top} \neg p_2 \rightarrow (O_{\neg p_2} \neg p_3 \rightarrow O_{\top}(\neg p_2 \rightarrow \neg p_3))$             | from (i) by propositional logic, since $\mathbf{G}$ contains $\mathbf{E}$                              |
| 2. $P_{\top} \neg p_2 \rightarrow (O_{\top}(\neg p_2 \rightarrow \neg p_3) \rightarrow O_{\top \wedge \neg p_2} \neg p_3)$ | instance of $\alpha 4$ (Section 18), which is one of the characteristic axiom schemata in $\mathbf{G}$ |
| 3. $P_{\top} \neg p_2 \rightarrow (O_{\top}(\neg p_2 \rightarrow \neg p_3) \rightarrow O_{\neg p_2} \neg p_3)$             | from 2 and line 3 in the proof of (i) above, using propositional logic                                 |
| 4. $P_{\top} \neg p_2 \rightarrow (O_{\top}(\neg p_2 \rightarrow \neg p_3) \Leftrightarrow O_{\neg p_2} \neg p_3)$         | from 1 and 3 by propositional logic  |



The present sequence 1–4 is a rather fragmentary proof in  $\mathbf{G}$  of its last line, 4. Rewriting  $O_{\top}$  as  $O$  in 4, we obtain the desired result (ii). ■

COROLLARY 22. *Suppose that  $\mathcal{L}$  contains the system  $\mathbf{G}$ . Then:*

(iii)  $Ce \vdash_{\mathcal{L}} \text{IIId}$

(iv)  $Cd \vdash_{\mathcal{L}} \text{IIe}$

**Proof.** Here, (iii) follows from (i) and the fact that  $\mathbf{G}$  contains  $\mathbf{E}$ . Again, (iv) is obtained from (ii) as follows: using  $\alpha 3$  (Section 18), which is a somewhat controversial schema of  $\mathbf{G}$ , together with the fact that  $\vdash_{\mathbf{G}} M\top$  ( $\mathbf{G}$  contains  $\mathbf{S5}$  for  $M$ ), we see that the sentence  $\text{Id}$  entails  $P_{\top} \neg p_2$ , i.e. the antecedent of 4 above. So, using that result together with 4 and  $\text{IIId}$ , we obtained the desired conclusion  $\text{IIe}$ . ■

Clearly, in spite of  $\text{IIId}$  being in general weaker than  $\text{IIe}$ , (iii) and (iv) are jointly to the effect that it does not matter which of  $Cd$  and  $Ce$  we choose as the ‘correct’ formalization of  $\mathcal{C}$ ; *provided*, however, that  $\mathbf{G}$  can be accepted as a satisfactory dyadic system. I have nothing against assuming this to be the case. Bearing in mind that  $\mathbf{G}$  is a generalized version of Hansson’s  $\mathbf{DSDL3}$ , this attitude of mine should be shared, e.g. by Hansson [1969], Føllesdal and Hilpinen [1971] and Spohn [1975].

## 9 PROBLEMS UNSOLVED BY THE DYADIC APPROACH; THE NEED FOR TEMPORAL AND QUANTIFICATIONAL RESOURCES IN THE BASIC LANGUAGE OF SATISFACTORY DEONTIC LOGICS; ON THE LOGIC OF ACTION; FAILURE OF LEFT-TO-RIGHT ADEQUACY

Our discussion of Prior’s paradoxes of commitment and Chisholm’s contrary-to duty imperative paradox was mainly designed to show how the idea of dyadic deontic logic naturally arises as an attempt to cope with these difficulties. The Exercises on such adequacy criteria for a logic of commitments as C1–C13 (Section 7.3 above) as well as our Theorem on the Contrary-to-Duty Imperative Paradox and Lemma on  $\text{IIId}$  and  $\text{IIe}$  should have given the reader a fairly clear opinion of the virtues of the dyadic approach and a nice explanation why it has proved to be such a powerful trend of thought in the development of modern deontic logic. It is now time to turn to its vices, i.e. to certain problems or problem-areas which the dyadic approach appears unable to handle. Mainly following the admirable survey given in Van Eck [1981], we present a list of such problems or problem-areas:

(I.) The dilemma of commitment and detachment.

- (II.) *Prima facie vs.* actual obligation; the *ceteris paribus* proviso.
- (III.) The ought-implies-can problem.
- (IV.) The Good Samaritan paradox and the Jephta dilemma.
- (V.) The problem of the relationship of act-utilitarianism to deontic logic.

I shall now briefly state these difficulties. After having done so, I shall then, without going into details, indicate what I take to be the proper attitude to them and what conclusions are in my opinion ‘reasonably’ to be drawn from them.

### 9.1 Survey of difficulties (after Van Eck [1981])

*I. The dilemma of commitment and detachment.* Suppose we were to accept a dyadic system  $\mathcal{L}$  satisfying the Andersonian criterion C8 (Exercise 18 above):

$$(C8) \quad \vdash_{\mathcal{L}} (A \wedge O_A B) \rightarrow OB$$

so that  $\mathcal{L}$  allows a *principle of detachment* to be valid for commitment-expressing formulae of the type  $O_A B$ . Then, provided only that  $\mathcal{L}$  is sufficiently strong in other respects (which?), we quickly obtain results like

$$\mathcal{C}d(e) \vdash_{\mathcal{L}} Op_3$$

$$\mathcal{C}d(e) \vdash_{\mathcal{L}} O\neg p_3 \wedge Op_3$$

$$\mathcal{C}d(e) \vdash_{\mathcal{L}} \perp$$

in violation of our requirement that the consistency of  $\mathcal{C}$  should be preserved. Hence, if  $\mathcal{L}$  is of this kind, the fourth and fifth attempts both break down as solutions to the Chisholm puzzle.

Now, these attempts may be defended by claiming that any ‘correct’ dyadic system must, like our  $\mathbf{G}$  and others, not allow detachment for commitment; it must definitely not satisfy the criterion C8. In short, in order for the dyadic solutions to work, detachment should not be possible.

However, detachment is not so easily given up from an intuitive standpoint. Here are some voices from the literature:

In nothing like schema (i) (sc. the one in C8) is valid, how can conditional obligation-sentences play the important role in normative argumentation which they seem to play? (Danielsson [1968, p. 66]).

How can we take seriously a conditional obligation if it cannot, by way of detachment, lead to an unconditional obligation? (Van Eck [1981, p. 23]).

So, on the other hand, we seem to feel that detachment should be possible after all. But we cannot have things both ways, can we? This is the dilemma on commitment and detachment.

*II. Prima Facie vs. Actual Obligation; the Ceteris Paribus Proviso.* Suppose that at a given time  $t$  John promises Suzy to marry her at a certain later time  $t + 7$ . Assume that the promise gives rise to an obligation for John to marry Suzy at  $t + 7$ , and that this obligation comes into force at time  $t$ . Now, in the meantime between  $t$  and  $t + 7$ , various things might happen that make it impossible for John to fulfill his obligation. For instance, he learns that his mother in Australia is dying, whence there arises, say at time  $t + 3$ , an obligation for John to go and visit her in Australia immediately. John takes off, but is then unable to marry Suzy at  $t + 7$ , so he breaks his promise and violates his first obligation.

Following Hintikka [1971] we may characterize this situation as one where an earlier obligation, due to the promise, is *overruled* by a stronger obligation, which arises in the meantime between the moment at which the first obligation comes into force ( $= t$ ) and the moment of its fulfillment ( $= t + 7$ ). Hintikka [1971] goes on to suggest that the famous *prima facie* vs. actual duty distinction (Sir David Ross [1930; 1939], Richard Price [1948]) should somehow be applicable to this situation. Following Van Eck [1981], then, I think we may say that at time  $t + 3$  the earlier obligation, though still in force at that time, is a mere *prima facie* duty, whereas the later and stronger obligation has acquired the status of an *actual* duty of John's. Our present problem concerns the explication of this distinction and its formal representation in systems of deontic logic; apparently, the issue was first raised by Hintikka [1971] and further discussed by Purtill [1973], Bergström [1974] and Van Eck [1981]. Furthermore, if we try to bring out the distinction by saying that John's first *prima facie* duty carries an implicit or tacit *ceteris paribus* rider 'other things being equal', whereas his second *actual* duty does not, we face the problem of analyzing the import of *ceteris paribus* provisos, in general as well as in the particular case at hand.

*III. The Ought-Implies-Can Problem.* My remarks on the celebrated Kantian principle will by necessity be very brief and will fail to do justice to the impressive richness of the literature on it. We ask: does 'ought' imply 'can'? then, if the answer is Yes, in what sense of (at least) (i) 'can' and (ii) 'imply'? Again, there are at least two alternatives under each heading here: (i.i) 'can' means logical possibility, and (i.ii) 'can' means some stronger possibility of a more 'practical' or 'real' kind, which might be explicated as a *temporally dependent* possibility in a sense that seemingly originates with

Montague [1968] and is further developed, e.g. by Chellas [1969] and Van Eck [1981]; moreover, ‘imply’ could mean (ii.i) ordinary logical consequence, or (ii.ii) some different form of consequence, say, the interesting notion of *deontic* consequence proposed in Hintikka [1957] and [1971].

Having thus surveyed some candidates in the area, I just think Van Eck [1981, II. Section 2.2], has given excellent reasons for regarding the combination (i.ii) with (ii.i) as providing the most viable and interesting interpretation of the Kantian principle: the alternative (i.i) seems to make it trivial, and the reasons against (ii.i) and for (ii.ii) are far from clear. But there are independent positive reasons as well for interpreting ‘can’ as temporally dependent, or historical, possibility and ‘imply’ as ordinary logical consequence.

*IV. The Good Samaritan Paradox ant the Jephtha Dilemma.* The first puzzle here goes back at least to Prior [1958] and has been discussed in a number of contributions, of which we only mention Danielsson [1968], Wedberg [1969], Van Fraassen [1972], Castañeda [1968a; 1974], Tomberlin and McGuinness [1977] and Van Eck [1981]. Interestingly, though, Knuuttila [1981] points out that in the fourteenth century versions of the paradox were known to and dealt with by Roger Rosetus in his *Commentary on the Sentences*.

Again, the Jephtha Dilemma (see the *Book of Judges*) was taken by Von Wright [1965] to be an interesting problem case for deontic logic, which illustrates such notions as those of a *predicament* and a *conflict of duty*. It has later been discussed extensively by Van Eck [1981].

The following version of the Good Samaritan paradox is presented in Tomberlin and McGuinness [1977] and goes back to Castañeda [1974]; consider the argument:

- (5) If Bob pays \$500 to the man he will murder one week hence, then Bob will murder a man one week hence.
- (6) It ought to be that Bob pays \$500 to the man he will murder one week hence (because Bob owes that amount of money to the latter).

*Therefore:*

- (7) It ought to be that Bob will murder a man one week hence.

In this argument, the first premiss, (5), may be taken to be, not only true, but even *logically true*. As for the second, (6), let us just assume it to be true. On the other hand, (7), is plainly false, in spite of the fact that the premisses (5) and (6) are both true. Hence, the argument must be invalid. But, if we translate it into the language of any of the Smiley–Hanson systems  $\mathcal{L}$  of Monadic Deontic Logic, which are all closed under the

rule of inference:

$$\frac{A \rightarrow B}{OA \rightarrow OB} \text{ (see Exercise 12 above)}$$

we find that the translation of (7) is  $\mathcal{L}$ -derivable from the translations of the premisses (5) and (6). What has gone wrong here?

The Jephtha Dilemma is reminiscent of the famous Morning Star Paradox in quantified modal logic with identity (see, e.g. Kanger [1957a]) and can be stated as follows. Consider the argument:

- (8) Miriam (i.e. the daughter of Jephtha) is identical to the first being that will meet Jephtha on his return home.
- (9) It ought to be that Jephtha immolates the first being that will meet him on his return home (because he has promised God to do so).

*Therefore:*

- (10) It ought to be that Jephtha immolates Miriam (his own daughter).

Again, here, as it seems, the premisses are true while the conclusion is false. So the argument must be invalid. But if we formalize it in a suitable system  $\mathcal{L}$  of deontic logic, it may well turn out that the inference is countenanced as valid in  $\mathcal{L}$ . How are we to account for this?

*V. The problem of the relationship of Act-Utilitarianism to deontic logic.* In Castañeda [1967; 1968], Castañeda points out the following intriguing difficulty for the familiar ethical theory known as Act-Utilitarianism. My statement of the problem will involve some amount of ‘precization’. Let  $X$  be any moral agent, let  $C$  be any situation or set of circumstances, and let  $A$  be any act open to  $X$  in  $C$  (in the sense that it is possible for  $X$  to do  $A$  in  $C$ ). Then, the following is a central thesis of Act-Utilitarianism:

- (U)  $X$  ought to do  $A$  in  $C$  iff for each act  $A'$  such that (i)  $A'$  is open to  $X$  in  $C$  and (ii)  $A'$  is an alternative to  $A$  in  $C$  and (iii)  $A'$  is distinct from  $A$

we have that the consequences of  $X$ 's doing  $A$  in  $C$  are better than those of  $X$ 's doing  $A'$  in  $C$ .

It may be that condition (iii) in this formulation of (U) is redundant, because entailed by (ii). Next, consider this set of assumptions:

- (11)  $X$  ought to do  $P \wedge Q$  in  $C$  (where  $P \wedge Q$  is a certain conjunctive act open to  $X$  in  $C$ )
- (12) The three acts  $P \wedge Q, P, Q$  are all (i) open to  $X$  in  $C$ , (ii) alternatives to each other in  $C$ , and (iii) distinct from one another.
- (13) If  $X$  ought to do  $P \wedge Q$  in  $C$ , then  $X$  ought to do  $P$  in  $C$  and  $X$  ought to do  $Q$  in  $C$ .

**THEOREM 23.** (after Castañeda [1968]): *The set  $\{(11), (12), (13), (U)\}$  is inconsistent.*

**Proof.** Exercise. (If you are unable to do it, see Castañeda [1968]!) What assumptions about the preference relation *better than* do we minimally have to make in order to establish the present result? ■

We should note here that (13) is reminiscent of the principle asserting that  $O$  is distributive over  $\wedge$ , which is valid in all the Smiley–Hanson systems of monadic deontic logic. Now, suppose we stick to the assumptions (11) and (12): then we face the tough choice, described by Wedberg [1969], between

- (i) maintaining (13) and rejecting the utilitarian thesis (U) already on grounds of deontic logic; and
- (ii) maintaining (U) and rejecting the deontic-logical principle (13).

Perhaps this is a ‘false dilemma’, though: why not abandon (12) and try to save both (13) and (U)?

I shall not discuss here this proposal and others that might be or have been made. I just like to point out that in Åqvist [1969] an attempt was made to relate Castañeda’s problem to the interesting work done by Bergström [1966] on utilitarian and teleological ethics. The notion of the *alternatives* to an action is seen to play a crucial role in Bergström’s analysis and was further scrutinized by Prawitz [1968; 1970], Bergström [1968] and by other contributors. The discussion quickly turned out to be surprisingly complex and it is difficult to give a fair assessment of its outcome.

## 9.2 Diagnosis

I have now finished my survey of the problem areas I–V. In my opinion, the existence of these difficulties, together with various unsuccessful attempts to overcome them, give considerable support to the diagnosis that

the languages of the current systems of deontic logic are far too poor to function as a satisfactory medium for formulating cues for the moral agent (Van Eck [1981, p. 1]).

I think this conclusion is born out in an especially clear way by the Van Eck treatment of problem II and his fascinating account of how *prima facie* obligations *pass into* actual duties as time elapses. The account is based on a system of temporal logic, more specifically: a system of *temporally relative modal* and *deontic predicate* logic. In my view, the virtues of this system are shown by its capacity to handle, not only the paradoxes of commitment and the contrary-to-duty imperative, but, as Van Eck also shows, the four problem areas I–IV as well. Handle them in a more convincing way than has so far been done up to this date, that is to say. As Van Eck observes in the preface to his [1981], though, he is not alone in having conceived of the idea of constructing a semantics for a notion of *temporally relative necessity* and basing a semantics of *temporal deontic* notions upon it: similar ideas can be found in Åqvist and Hoepelman [1981], going back to Chellas [1969] and Montague [1968] (see problem III above). Also, Thomason [1981] (deriving from an original 1970 version) and [1981a] should be mentioned in the present context. Now, the Van Eck framework appears to be richer than these rival ones, because it uses temporal variables, constants and quantifiers in the *object-language*. And this, I think, makes it more useful for philosophical applications; an illustration of this claim is perhaps the little paper Åqvist [1981], where precisely a Van Eck-type framework is applied to the ancient so-called Protagoras paradox (see also Lenzen [1977] and Smullyan [1978]).

Among recent contributions in the same vein, those of Bailhache [1991] (going back to work done in the early eighties) and [1993] strike me as particularly valuable. See also Åqvist [1997a].

There are strong reasons, then, for enriching the basic language of satisfactory deontic logics with explicit temporal resources. Moreover, problems IV and V nicely show, I take it, the need for *quantificational* resources in that language as well; how could we otherwise even begin to state those problems in an intelligible way? The indispensability of quantifiers in deontic logic was, on general grounds, very well argued by Hintikka already in his [1957] (and later in his [1971]), when he comments on the fundamental work done by Von Wright [1951; 1951a] as well as by Prior [1955]. A system of deontic predicate logic (quantification theory) is also presented in Kanger [1957]; his system is one *with identity* and, interestingly, blocks deontic analogues of the Morning Star paradox such as the Jephta dilemma in the model-theoretic semantics given for it.

Again, the need for quantifiers in deontic logic seems to be one of the main tenets of Castañeda's in an impressively large number of contributions, of which we mention here only Castañeda [1954; 1959; 1981].

We must briefly touch on the following question: given the indispensability of quantifiers in deontic logic, *exactly over what* sort of entities do we have to quantify? agents, patients, times, places, circumstances or what? Hintikka [1957] suggested *individual acts*; in Makinson [1981] this suggestion

is shown to give rise to considerable interpretational difficulties. See also Robison [1964] for further suggestions. In answer to this question I would like here to recommend a very broad, liberal and open-minded attitude: as deontic logicians, we should be prepared to quantify over whatever entities the ethical theory or normative (e.g. legal) system requires us to consider seriously, and to adjust our deontic predicate logic accordingly. We should not, I contend, worry too much about ontological commitments; in today's research situation the important thing is to get the right kind of structure going.

### 9.2.1 *On the logic of action*

Having now stressed the importance of temporal and quantificational machinery to viable deontic logics, there is a third research trend in our area, to which I like to draw the reader's attention. This trend claims that such logics ought to be combined with a *logic of action*; it is usually taken to have been initiated in Von Wright [1963] and followed up, e.g. in Von Wright [1967] (with comments by Chisholm [1967]) and Von Wright [1974]. However, if we consider the distinguishing mark of a logic of action to be the presence in its basic language of a special 'causal' operator of *agency* (expressing that an agent *brings it about*, *sees to it*, *makes it true* that so-and-so is the case), we might just as well credit Kanger [1957], Anderson [1962] and Kanger and Kanger [1966] with the idea. Anyway, the latter authors apply it in attempts to reconstruct and to extend the Hohfeldian system of jural relationships as set forth by Hohfeld [1919]. In this endeavor they are followed by, notably, Pörn [1970], Anderson [1971] and Lindahl [1977]. This movement is highly interesting and promising for the future, and any account of present-day deontic logic would be seriously incomplete if it did not mention it.

Finally, I like to close this Part by making a remark on the Good Samaritan paradox, which is intended to illustrate the notion of *left-to-right adequacy* (Section 5.4 above) and its failure in connection with the Smiley–Hanson systems of monadic deontic logic.

### 9.2.2 *Remark on the Good Samaritan paradox*

Consider the definitional enrichment L(D1–D14) of L, where D13 and D14 are as follows:

(D13) Bob pays \$500 to the man he will murder one week hence =df  $p_5$

(D14) Bob will murder a man one week hence =df  $p_6$

Extending the translation  $t$  in the obvious way, we obtain the following formalization of the sentence (5) in the paradox:



$$t(5) = (p_5 \rightarrow p_6).$$

Note, then, that (5) is in  $\Phi$  (= the sentence-set of the fragment  $\text{FNE}_6$ ) and that (5) is logically true (presumably). Hence, (5) is in NDL. On the other hand,  $t(5)$  is *not* provable in any of the Smiley–Hanson systems  $\mathcal{L}$  (why?). So we have a counterexample to the *left-to-right* adequacy of  $t$  with respect to NDL and  $\mathcal{L}$ .

The import of this counterexample is that there are validities in natural deontic logic (NDL), or relations of ‘natural’ logical consequence, which fail to be representable in certain formal deontic logics, such as  $\mathcal{L}$ . The reason is, of course, that we need a *quantificational* formal framework with *definite descriptions* in order adequately to formalize such a sentence as (5); the propositional language  $L$  is simply not expressive enough.

We should contrast this counterexample with those met with above, which purported to show that  $t$  was not right-to-left adequate with respect to NDL and  $\mathcal{L}$ , in the sense that the latter sanctions *more* logical validities than the former. The present situation is precisely the opposite one. The Good Samaritan paradox is usually cited as an instance of failing right-to-left adequacy, just as those of commitment etc. I think it is of some interest to note that it could also be used to illustrate the opposite failure.

### III. TEN SMILEY–HANSON SYSTEMS OF MONADIC DEONTIC LOGIC

#### 10 LANGUAGE, PROOF THEORY AND SEMANTICS

##### 10.1 Language

###### 10.1.1 Alphabet

Our alphabet consists of

- (i) a denumerable set Prop of *proposition letters*  $p, q, r, p_1, p_2, \dots$ ;
- (ii) the primitive *logical connectives*  $\top$  (*verum*),  $\perp$  (*falsum*),  $\neg$  (negation),  $O$  (obligation),  $P$  (permission),  $\wedge$  (conjunction),  $\vee$  (disjunction),  $\rightarrow$  (material implication) and  $\Leftrightarrow$  (material equivalence); and
- (iii) the parentheses ( ).

###### 10.1.2 Sentences

(well formed formulas, wffs): The set  $\Sigma$  of all sentences of our language is defined as the smallest set  $S$  such that

- (a) every proposition letter in Prop is in  $S$ ,
- (b)  $\top$  and  $\perp$  are in  $S$ ,
- (c) if  $A$  is in  $S$ , then so are  $\neg A$ ,  $OA$  and  $PA$ ,
- (d) if  $A, B$  are in  $S$ , then so are  $(A \wedge B)$ ,  $(A \vee B)$ ,  $(A \rightarrow B)$  and  $(A \Leftrightarrow B)$ .

The sentences under (a) and (b) are the *atomic* sentences of the language.

### 10.1.3 Degrees of logical connectives

$\top$  and  $\perp$  are of degree 0;  $\neg, O, P$  are of degree 1; and the remaining connectives are all of degree 2.

### 10.1.4 Definition

$$FA = \text{df } \neg PA \text{ (alternatively: } O\neg A).$$

### 10.1.5 Conventions for dropping brackets

Brackets are omitted in accordance with these canons:

- (i) Connectives of degree 1 bind more strongly than connectives of degree 2.
- (ii) Among the latter,  $\wedge$  and  $\vee$  bind more strongly than  $\rightarrow$  and  $\Leftrightarrow$ .
- (iii) Outer brackets are mostly dropped around sentences.

## 10.2 Proof theory

The following two *rules of inferences* are common to all the ten Smiley–Hanson systems of monadic deontic logic to be dealt with:

$$(R1) \frac{A, A \rightarrow B}{B} \text{ (modus ponens).}$$

$$(R2) \frac{A}{OA} \text{ (O-necessitation).}$$

Consider next the following list A0–A7 of *axiom schemata*:

- (A0) All truth-functional tautologies (over our present language)
- (A1)  $PA \Leftrightarrow \neg O\neg A$
- (A2)  $O(A \rightarrow B) \rightarrow (OA \rightarrow OB)$
- (A3)  $OA \rightarrow PA$
- (A4)  $OA \rightarrow OOA$
- (A5)  $POA \rightarrow OA$
- (A6)  $O(OA \rightarrow A)$
- (A7)  $O(POA \rightarrow A)$

The ten logics to be studied here are called **OK**, **OM**, **OS4**, **OB**, **OS5**, **OK<sup>+</sup>**, **OM<sup>+</sup>**, **OS4<sup>+</sup>**, **OB<sup>+</sup>**, and **OS5<sup>+</sup>**. They are defined as follows (where R1 and R2 are assumed for all):

- OK** = A0–A2
- OM** = A0–A2, A6
- OS4** = A0–A2, A4, A6
- OB** = A0–A2, A6, A7
- OS5** = A0–A2, A4, A5 (note that A6 and A7 are derivable in **OS5**)

Again, let  $\mathcal{L}$  be any of these five systems. Then:

$$\mathcal{L}^+ = \mathcal{L}, A3$$

Of these ten deontic logics, **OK** is (apart from unessential differences) identical to the system **F** of Hanson [1965], **OK<sup>+</sup>** to his **D**, **OB<sup>+</sup>** to his **DB**, whereas **OB** is discussed neither by Hanson [1965] nor by Smiley [1963]. The remaining six systems are named exactly as in Smiley [1963].

### 10.2.1 Provability and consistency

Let  $\mathcal{L}$  be any of the ten systems just defined. Then, the set of  $\mathcal{L}$ -provable sentences (or the set of  $\mathcal{L}$ -theses) is the smallest set  $S \subseteq \Sigma$  such that (i) each instance of every axiom schema of  $\mathcal{L}$  is in  $S$ , and (ii)  $S$  is closed under the rules R1 and R2. We write ' $\vdash_{\mathcal{L}} A$ ' to indicate that  $A$  is  $\mathcal{L}$ -provable. Also, a set  $S$  of sentences is  $\mathcal{L}$ -inconsistent iff there are  $B_1, \dots, B_n$  in  $S$  ( $n \geq 1$ ) such that  $\vdash_{\mathcal{L}} (B_1 \wedge \dots \wedge B_n) \rightarrow \perp$ ; and  $S$  is  $\mathcal{L}$ -consistent otherwise.

Again, we say that a sentence  $A$  is  $\mathcal{L}$ -derivable from a set  $S$  of sentences, in symbols:  $S \vdash_{\mathcal{L}} A$ , just in case  $S \cup \{\neg A\}$  is  $\mathcal{L}$ -inconsistent. Clearly,  $\vdash_{\mathcal{L}} A$  iff

$\emptyset \mid_{\mathcal{L}} A$ , i.e. the  $\mathcal{L}$ -provable sentences are exactly those that are  $\mathcal{L}$ -derivable from the empty set.

### 10.3 Semantics

#### 10.3.1 Models

By a *model* we mean a triple  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$  where:

- (i)  $W$  is a non-empty set (heuristically, of ‘possible worlds’ or ‘possible situations’).
- (ii)  $R \subseteq W \times W$  (a binary relation on  $W$ , heuristically, of ‘deontic alternativeness’ or ‘co-permissibility’).
- (iii)  $V$  is an assignment, which associates a truth-value 1 or 0 with each ordered pair  $\langle p, x \rangle$  where  $p$  is a proposition letter and  $x$  is an element of  $W$ ; in technical jargon,  $V: \text{Prop} \times W \rightarrow \{1, 0\}$ .

#### 10.3.2 Truth conditions

Let  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$  be any model, let  $x$  be any member of  $W$ , and let  $A$  be in  $\Sigma$ . We want to define what it means for  $A$  to be *true at  $x$  in  $\mathcal{U}$* , in symbols:  $\mid_x^{\mathcal{U}} A$ . As usual, the definition is recursive on the length of  $A$ :

$$\mid_x^{\mathcal{U}} p \quad \text{iff} \quad V(p, x) = 1 \quad (\text{for any } p \text{ in Prop}).$$

$$\mid_x^{\mathcal{U}} \top.$$

$$\text{not } \mid_x^{\mathcal{U}} \perp.$$

$$\mid_x^{\mathcal{U}} \neg A \quad \text{iff} \quad \text{not } \mid_x^{\mathcal{U}} A.$$

$$\mid_x^{\mathcal{U}} OA \quad \text{iff} \quad \text{for every } y \text{ in } W \text{ such that } xRy, \mid_y^{\mathcal{U}} A.$$

$$\mid_x^{\mathcal{U}} PA \quad \text{iff} \quad \text{for some } y \text{ in } W \text{ such that } xRy, \mid_y^{\mathcal{U}} A.$$

$$\mid_x^{\mathcal{U}} (A \wedge B) \quad \text{iff} \quad \mid_x^{\mathcal{U}} A \quad \text{and} \quad \mid_x^{\mathcal{U}} B.$$

$$\frac{\mathcal{U}}{x} (A \vee B) \text{ iff } \frac{\mathcal{U}}{x} A \text{ or } \frac{\mathcal{U}}{x} B \text{ (or both).}$$

$$\frac{\mathcal{U}}{x} (A \rightarrow B) \text{ iff if } \frac{\mathcal{U}}{x} A, \text{ then } \frac{\mathcal{U}}{x} B.$$

$$\frac{\mathcal{U}}{x} (A \Leftrightarrow B) \text{ iff } (\frac{\mathcal{U}}{x} A \text{ iff } \frac{\mathcal{U}}{x} B).$$

### 10.3.3 Conditions on $R$ in a model

Corresponding to the five axiom schemata A3–A7 we now list five conditions on the relation  $R$  in a model (where we assume the variables ‘ $x$ ’, ‘ $y$ ’, ‘ $z$ ’ to range over  $W$ , and where we use the symbols  $\&$ ,  $\supset$ ,  $\forall$  and  $\exists$  as a shorthand notation in the metalanguage in the obvious way):

- (R3)  $R$  is *serial* in  $W$  :  $\forall x \exists y (xRy)$
- (R4)  $R$  is *transitive* in  $W$  :  $\forall x, y, z (xRy \& yRz \supset xRz)$
- (R5)  $R$  is *Euclidean* in  $W$  :  $\forall x, y, z (xRy \& xRz \supset yRz)$
- (R6)  $R$  is *almost reflexive* in  $W$  :  $\forall x, y (xRy \supset yRy)$
- (R7)  $R$  is *almost symmetric* in  $W$  :  $\forall x, y, z (xRy \supset (yRz \supset zRy))$

### 10.3.4 Classification of models

We now use the restrictions on  $R$  just listed to obtain a subcategorization of the set of all models into various kinds. Thus, we stipulate that:

- The class of **OK**-models = the class of *all* models (no condition on  $R$  being imposed).
- The class of **OM**-models = the class of all models with almost reflexive  $R$ .
- The class of **OS4**-models = the class of all models with transitive and almost reflexive  $R$ .
- The class of **OB**-models = the class of all models with almost symmetric and almost reflexive  $R$ .
- The class of **OS5**-models = the class of all models with Euclidean and transitive  $R$ .
- The class of **OK**<sup>+</sup>-models = the class of all models with serial  $R$ .
- The class of **OM**<sup>+</sup>-models = the class of all models with serial and almost reflexive  $R$ .
- The class of **OS4**<sup>+</sup>-models = the class of all models with serial, transitive and almost reflexive  $R$ .

The class of **OB**<sup>+</sup>-models = the class of all models with serial, almost symmetric and almost reflexive  $R$ .  
 The class of **OS5**<sup>+</sup>-models = the class of all models with serial, Euclidean and transitive  $R$ .

In this chain of definitions, we always take the relevant restrictions on  $R$  to be relative to the world-set  $W$  in a model, so that ‘serial’ means ‘serial in  $W$ ’, and so on.

### 10.3.5 Validity and satisfiability

Let  $\mathcal{L}$  be any of the ten systems **OK, OM, OS4, OB, OS5, OK<sup>+</sup>, OM<sup>+</sup>, OS4<sup>+</sup>, OB<sup>+</sup>, OS5<sup>+</sup>**. We say that a sentence  $A$  is  $\mathcal{L}$ -valid (in symbols:  $\models_{\mathcal{L}} A$ ) iff  $\models_x^{\mathcal{U}} A$  for all  $\mathcal{L}$ -models  $\mathcal{U}$  and for all  $x$  in  $W$ . Also, we say that a set  $S$  of sentences is  $\mathcal{L}$ -satisfiable iff there is an  $\mathcal{L}$ -model  $\mathcal{U}$  and member  $x$  of  $W$  such that for all sentences  $A$  in  $S$ ,  $\models_x^{\mathcal{U}} A$ . Clearly, we have that  $\models_{\mathcal{L}} A$  iff the unit set  $\{\neg A\}$  is not  $\mathcal{L}$ -satisfiable.

Again, we may introduce a semantic notion parallel to that of (proof-theoretic) derivability: we say that a sentence  $A$  is *semantically  $\mathcal{L}$ -entailed* by a set  $S$  of sentences (in symbols:  $S \models_{\mathcal{L}} A$ ) iff  $S \cup \{\neg A\}$  is not  $\mathcal{L}$ -satisfiable.

We then have that  $\models_{\mathcal{L}} A$  iff  $\emptyset \models_{\mathcal{L}} A$ .

## 11 SEMANTIC SOUNDNESS AND COMPLETENESS OF THE SMILEY–HANSON SYSTEMS

**THEOREM 24** (Soundness Theorem). *Let  $\mathcal{L}$  be any of the systems **OK, OM, OS4, . . . OS5<sup>+</sup>**. Then, for all  $A \in \Sigma$ , if  $\vdash_{\mathcal{L}} A$ , then  $\models_{\mathcal{L}} A$ . In other words, all  $\mathcal{L}$ -provable sentences are  $\mathcal{L}$ -valid.*

**Proof.** [Outlined] For each system  $\mathcal{L}$  we must show that (i) every instance of every axiom schema of  $\mathcal{L}$  is  $\mathcal{L}$ -valid, and that (ii) the rules R1 and R2 preserve  $\mathcal{L}$ -validity. Then, we can verify by inductions on the length of proof in  $\mathcal{L}$  that if  $\vdash_{\mathcal{L}} A$ , then  $\models_{\mathcal{L}} A$ . To do this is a bit tedious, for sure, but

entirely routine. Let us give just one example here in order to illustrate the methodology, or strategy, of argument. ■

**EXAMPLE 25.** Suppose we want to check that all instances of A5 are indeed **OS5**-valid. Assume otherwise, then, i.e. that there is a sentence  $A$  such that, for some **OS5**-model  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$  and some  $x$  in  $W$ , we have:

(1) it is not the case that  $\frac{\mathcal{U}}{x} POA \rightarrow OA$

Applying relevant truth conditions sufficiently many times to (1), we reduce it to

(2)  $\frac{\mathcal{U}}{x} POA$  and  $\frac{\mathcal{U}}{x} P\neg A$ .

Applying the truth condition for  $P$ , we obtain from (2):

(3)  $\frac{\mathcal{U}}{y} OA$ , for some  $y$  in  $W$  with  $xRy$

as well as

(4)  $\frac{\mathcal{U}}{z} \neg A$ , for some  $z$  in  $W$  with  $xRz$ .

Since  $\mathcal{U}$  is an **OS5**-model,  $R$  is Euclidean in  $W$ ; hence we obtain

(5)  $yRz$  (because, by (3) and (4),  $xRy$  and  $xRz$ ).

Then, applying the truth condition for  $O$  to (3), we get from (5):

(6)  $\frac{\mathcal{U}}{z} A$

which result contradicts (4), as the latter gives us

(7) not  $\frac{\mathcal{U}}{z} A$

by the truth condition for  $\neg$ . Contradiction.

**COROLLARY 26.** *Let  $\mathcal{L}$  be as usual and let  $S \subseteq \Sigma$  be any set of sentences. Then, if  $S$  is  $\mathcal{L}$ -satisfiable, then  $S$  is  $\mathcal{L}$ -consistent.*

**Proof.** Assume otherwise, i.e., that some  $S$  is  $\mathcal{L}$ -satisfiable but not  $\mathcal{L}$ -consistent. Then, by the definition of  $\mathcal{L}$ -inconsistency, there are  $B_1, \dots, B_n$  in  $S$  such that  $\vdash_{\mathcal{L}} (B_1 \wedge \dots \wedge B_n) \rightarrow \perp$ . Hence, by the Soundness Theorem,  $\frac{\mathcal{U}}{\mathcal{L}} (B_1 \wedge \dots \wedge B_n) \rightarrow \perp$ . But this means that for some  $x$  in the model  $\mathcal{U}$ , whose

existence is guaranteed by  $S$  being  $\mathcal{L}$ -satisfiable, we have  $\frac{\mathcal{U}}{x} B_1 \wedge \dots \wedge B_n$

as well as  $\frac{\mathcal{U}}{x} (B_1 \wedge \dots \wedge B_n) \rightarrow \perp$ , hence  $\frac{\mathcal{U}}{x} \perp$ . Contradiction.  $\blacksquare$

**THEOREM 27 (Completeness Theorem).** *Version I (strong completeness). Let  $\mathcal{L}$  be as usual and let  $S \subseteq \Sigma$ . Then, if  $S$  is  $\mathcal{L}$ -consistent, then  $S$  is  $\mathcal{L}$ -satisfiable. Version II (weak completeness). Let  $\mathcal{L}$  be as usual and let*

$A \in \Sigma$ . Then, if  $\models_{\mathcal{L}} A$ , then  $\vdash_{\mathcal{L}} A$ . In other words, all  $\mathcal{L}$ -valid sentences are  $\mathcal{L}$ -provable.

**Proof.** Let us first see how the weak version can be obtained as a corollary of the strong one. Assume, contrary to the weak version, that for some sentence  $A$ ,  $\models_{\mathcal{L}} A$  but not  $\vdash_{\mathcal{L}} A$ . Then,  $\{\neg A\}$  must be  $\mathcal{L}$ -consistent (otherwise, we would have  $\{\neg A\} \vdash_{\mathcal{L}} \perp$ ,  $\vdash_{\mathcal{L}} \neg A \rightarrow \perp$ , and  $\vdash_{\mathcal{L}} A$ ; but we assumed: not  $\vdash_{\mathcal{L}} A$ ). Therefore, by the strong version I,  $\{\neg A\}$  is  $\mathcal{L}$ -satisfiable, i.e. for some  $\mathcal{L}$ -model  $\mathcal{U}$  and for some  $x$  in  $W$ ,  $\models_x^{\mathcal{U}} \neg A$ , so not  $\models_x^{\mathcal{U}} A$ . But this result conflicts with  $\models_{\mathcal{L}} A$ . Contradiction. ■

We are then justified in concentrating our efforts on establishing the strong version I of the Completeness Theorem. We begin by calling attention to the following definitions and lemmata.

**DEFINITION 28** ( $\mathcal{L}$ -saturated sets). Let  $\mathcal{L}$  be as usual and let  $x \subseteq \Sigma$  be any set of sentences. We say that  $x$  is  $\mathcal{L}$ -saturated iff

- (i)  $x$  is  $\mathcal{L}$ -consistent, and
- (ii) for each sentence  $A$ , either  $A \in x$  or  $\neg A \in x$ .

**LEMMA 29** ( $\mathcal{L}$ -saturated sets). Let  $x$  be any  $\mathcal{L}$ -saturated set of sentences. Then, for all sentences  $A, B$ :

- (i) Every  $\mathcal{L}$ -provable sentence is in  $x$ .
- (ii)  $x$  is closed under modus ponens (if  $A \in x$  and  $A \rightarrow B \in x$ , then  $B \in x$ ).
- (iii)  $T \in x$ .
- (iv)  $\perp \notin x$ .
- (v)  $\neg A \in x$  iff  $A \notin x$ .
- (vi)  $A \wedge B \in x$  iff  $A \in x$  and  $B \in x$ .
- (vii)  $A \vee B \in x$  iff  $A \in x$  or  $B \in x$ .
- (viii)  $A \rightarrow B \in x$  iff if  $A \in x$  then  $B \in x$ .
- (ix)  $A \Leftrightarrow B \in x$  iff  $A \in x$  if and only if  $B \in x$ .

**Proof.** Familiar. ■



LEMMA 30 (Lindenbaum's Lemma). *Any  $\mathcal{L}$ -consistent set  $x$  of sentences can be extended to an  $\mathcal{L}$ -saturated set  $x^+$  with  $x \subseteq x^+$  ( $\mathcal{L}$  being as usual).*

**Proof.** See, e.g. Makinson [1966, p. 381 f]. ■

LEMMA 31 (Makinson's Lemma). *Let  $\mathcal{L}$  be as usual and let  $x$  be any  $\mathcal{L}$ -saturated set of sentences. Let  $A$  be any sentence such that  $\neg OA \in x$ . Let  $x_A = \{B \in \Sigma : OB \in x\} \cup \{\neg A\}$ . Then  $x_A$  is  $\mathcal{L}$ -consistent.*

**Proof.** (see Makinson [1966, p. 382]). Suppose  $x_A$  is not  $\mathcal{L}$ -consistent. Then there are sentences  $B_1, \dots, B_n (n \geq 0)$  such that each  $OB_i \in x$  and such that  $\vdash_{\mathcal{L}}(B_1 \wedge \dots \wedge B_n \wedge \neg A) \rightarrow \perp$ ; by virtue of the fact that axiom schema A0 is in every  $\mathcal{L}$ , then, such that

$$\vdash_{\mathcal{L}}(B_1 \wedge \dots \wedge B_n) \rightarrow A.$$

Consider first the case where  $n = 0$ . This means that  $\vdash_{\mathcal{L}}A$ . Then by the rule R2 of  $O$ -necessitation (common to all our  $\mathcal{L}$ ), we have  $\vdash_{\mathcal{L}}OA$ . Hence, by the Lemma on  $\mathcal{L}$ -saturated sets,  $OA \in x$ . Thus,  $OA$  and  $\neg OA$  are both in  $x$ , so  $x$  is  $\mathcal{L}$ -inconsistent. Contradiction.

Consider next the case where  $n \geq 1$ . Since  $\vdash_{\mathcal{L}}(B_1 \wedge \dots \wedge B_n) \rightarrow A$ , we have by a tautology under A0 and R1 that  $\vdash_{\mathcal{L}}B_1 \rightarrow (B_2 \rightarrow \dots (B_n \rightarrow A) \dots)$ . Hence, by R2,  $\vdash_{\mathcal{L}}OB_1 \rightarrow (B_2 \rightarrow \dots (B_n \rightarrow A) \dots)$ . Hence, using axiom schema A2 (common to all our  $\mathcal{L}$ )  $n$  times, together with R1 and appropriate tautologies under A0, we obtain that  $\vdash_{\mathcal{L}}OB_1 \rightarrow (OB_2 \rightarrow \dots (OB_n \rightarrow OA) \dots)$ . Hence, by the Lemma on  $\mathcal{L}$ -saturated sets, that sentence is in  $x$ . But each  $OB_i \in x$ , so, by the same Lemma (clause (ii)) applied  $n$  times,  $OA \in x$ . Thus,  $OA$  and  $\neg OA$  are both in  $x$ , so  $x$  is  $\mathcal{L}$ -inconsistent. Contradiction. ■

DEFINITION 32 (Canonical  $\mathcal{L}$ -models). Let  $\mathcal{L}$ , as usual, be any of our ten monadic deontic logics, and let  $S$  be any  $\mathcal{L}$ -consistent set of sentences, so that, by Lindenbaum's Lemma,  $S^+$  is  $\mathcal{L}$ -saturated and  $S \subseteq S^+$ . By the *canonical  $\mathcal{L}$ -model generated by  $S$*  we mean the structure

$$\mathcal{U}_{\mathcal{L}} = \langle W_{\mathcal{L}}, \mathcal{R}_{\mathcal{L}}, \mathcal{V}_{\mathcal{L}} \rangle$$

where:

- (i)  $W_{\mathcal{L}}$  = the smallest collection  $U$  of  $\mathcal{L}$ -saturated sets such that:
  - (a)  $S^+$  is in  $U$ .
  - (b) If  $x$  is in  $U$ , and  $A$  is a sentence with  $\neg OA \in x$ , then  $(x_A)^+$  is in  $U$  (where  $x_A$  is defined as in Makinson's Lemma).
- (ii)  $R_{\mathcal{L}}$  = the binary relation on  $W_{\mathcal{L}}$  such that for all  $x, y$  in  $W_{\mathcal{L}}$  :  $xR_{\mathcal{L}}y$  iff for all sentences  $A$ , whenever  $OA \in x$  then  $A \in y$ .

- (iii)  $V_{\mathcal{L}}$  = the assignment defined as follows: for each proposition letter  $p$  and each  $x$  in  $W_{\mathcal{L}}$ ,  $V(p, x) = 1$  iff  $p \in x$ .

LEMMA 33 (Verification Lemma). *As just defined,  $\mathcal{U}_{\mathcal{L}} = \langle W_{\mathcal{L}}, \mathcal{R}_{\mathcal{L}}, \mathcal{V}_{\mathcal{L}} \rangle$  is an  $\mathcal{L}$ -model.*

LEMMA 34 (Coincidence Lemma). *For each sentence  $A$  and for each  $x$  in  $W_{\mathcal{L}}$  (as defined above),  $\frac{\mathcal{U}_{\mathcal{L}}}{x} A$  iff  $A \in x$ .*

We shall wait a little with the proofs of these two lemmata. Instead, let us see how they together yield Version I of the Completeness Theorem.

**Proof.** [Completeness Theorem] (Version I): Letting  $\mathcal{L}$  be as usual, assume  $S$  to be any  $\mathcal{L}$ -consistent set of sentences. We are to show that  $S$  is  $\mathcal{L}$ -satisfiable. Well, by the Verification Lemma,  $\mathcal{U}_{\mathcal{L}}$  (as just defined) is an  $\mathcal{L}$ -model. By the Coincidence Lemma, we obtain in particular that for each sentence  $A$ ,  $\frac{\mathcal{U}_{\mathcal{L}}}{S^+} A$  iff  $A \in S^+$  (as  $S^+$ , by definition, belongs to  $W_{\mathcal{L}}$ ). Hence, since  $S \subseteq S^+$ , we have  $\frac{\mathcal{U}_{\mathcal{L}}}{S^+} A$  for every  $A$  in  $S$ . In other words, assuming  $S$  to be any  $\mathcal{L}$ -consistent set of sentences, we have constructed an  $\mathcal{L}$ -model, viz.  $\mathcal{U}_{\mathcal{L}}$ , such that for some  $x$  in  $W_{\mathcal{L}}$ , viz.  $S^+$ ,  $\frac{\mathcal{U}_{\mathcal{L}}}{x} A$  for each  $A$  in  $S$ ; i.e. we have shown  $S$  to be  $\mathcal{L}$ -satisfiable. ■

We still need one more lemma before being able to establish the Verification Lemma and the Coincidence Lemma (the proofs of which have not yet been given):

LEMMA 35 (Saturation Lemma for canonical  $\mathcal{L}$ -models). *Let  $\mathcal{L}$  be as usual, let  $S$  be any  $\mathcal{L}$ -consistent set of sentences, and let  $\mathcal{U}_{\mathcal{L}}$  be defined as above. Then,  $W_{\mathcal{L}}$  is such that for all sentences  $A$  and all  $x$  in  $W_{\mathcal{L}}$ :*

- (i)  $OA \in x$  iff for all  $y$  in  $W_{\mathcal{L}}$  with  $xR_{\mathcal{L}}y$ ,  $A \in y$ .
- (ii)  $PA \in x$  iff there is a  $y$  in  $W_{\mathcal{L}}$  such that  $xR_{\mathcal{L}}y$  and  $A \in y$ .

**Proof.** (From now on we shall use  $\&$ ,  $\supset$ ,  $\forall$ ,  $\exists$  etc. as metalinguistic shorthands with their familiar meanings and use ‘ $x$ ’, ‘ $y$ ’, ‘ $z$ ’ as variables over  $W_{\mathcal{L}}$ .)

Ad(i): The ‘only if’ part is easy — assume, for any  $x, y$  in  $W_{\mathcal{L}}$ :

1.  $OA \in x$  hypothesis
2.  $xR_{\mathcal{L}}y$  hypothesis

Then:

- |    |  |  |
|----|--|--|
| 3. | $\forall B(OB \in x \supset B \in y)$                                | from 2 by the definition of $R_{\mathcal{L}}$          |
| 4. | $A \in y$  | 1,3, universal instantiation,<br>modus ponens          |
| 5. | $OA \in x \supset (xR_{\mathcal{L}}y \supset A \in y)$               | 1–4, rule of conditional proof,<br>discharging 1 and 2 |
| 6. | $\forall x, y(OA \in x \supset (xR_{\mathcal{L}}y \supset A \in y))$ | $x, y$ any members of $W_{\mathcal{L}}$                |

Number 6 can easily be rewritten as the ‘only if’ half of (i)

Again, to do the ‘if’ part, assume for any  $x$  in  $W_{\mathcal{L}}$ :

- |    |               |            |
|----|---------------|------------|
| 1. | $OA \notin x$ | hypothesis |
|----|---------------|------------|

Then:

- |     |  |  |
|-----|--|--|
| 2.  | $\neg OA \in x$  | from 1 by the Lemma on $\mathcal{L}$ -<br>saturated sets                         |
| 3.  | $\neg A \in x_A$   | by the definition of $x_A$ in<br>Makinson’s Lemma                                |
| 4.  | $\neg A \in (x_A)^+$   | $x_A \subseteq (x_A)^+$ by Lindenbaum  |
| 5.  | $A \notin (x_A)^+$   | from 4 by the Lemma on $\mathcal{L}$ -<br>saturated sets                         |
| 6.  | $(x_A)^+ \in W_{\mathcal{L}}$                                    | by the definition of $W_{\mathcal{L}}$ and<br>2                                  |
| 7.  | $\forall B(OB \in x \supset B \in (x_A)^+)$                      | by the definition of $x_A$ in<br>Makinson’s Lemma                                |
| 8.  | $xR_{\mathcal{L}}(x_A)^+$  | from 7 by the definition of<br>$R_{\mathcal{L}}$                                 |
| 9.  | $\exists y(xR_{\mathcal{L}}y \& A \notin y)$                     | from 5,6,8 by existential<br>generalization, the $y$ at issue<br>being $(x_A)^+$ |
| 10. | $OA \notin x \supset \exists y(xR_{\mathcal{L}}y \& A \notin y)$ | 1–9, conditional proof, dis-<br>charging 1                                       |
| 11. | $\forall y(xR_{\mathcal{L}}y \supset A \in y) \supset OA \in x$  | from 10 by contraposition  |

where 11 is the desired ‘if’ half of (i).

Ad (ii): The verification of (ii) can be left to the reader. *Hint*: appeal to the fact that every instance of A1, i.e.  $PA \Leftrightarrow \neg O\neg A$ , is in every  $\mathcal{L}$ -saturated

set.

The proof of the Saturation Lemma for Canonical  $\mathcal{L}$ -Models is complete. ■

Let us now deal with our unproven lemmas and start with the easiest one (or, at least, the one with the shortest proof):

**Proof.** [The Coincidence Lemma] For each wff  $A$  and each  $x$  in  $W_{\mathcal{L}}$ ,  $\frac{\mathcal{U}_{\mathcal{L}}}{x}A$  iff  $A \in x$ .

The proof proceeds by induction on the length of  $A$ .

*Basis.*  $A$  is either (a)  $\top$ , or (b)  $\perp$ , or (c) some proposition letter  $p$ .

- (a)  $\frac{\mathcal{U}_{\mathcal{L}}}{x}\top$  and  $\top \in x$  (by the truth condition for  $\top$  and by the Lemma on  $\mathcal{L}$ -saturated sets),
- (b) not  $\frac{\mathcal{U}_{\mathcal{L}}}{x}\perp$  and  $\perp \notin x$  (correspondingly),
- (c)  $\frac{\mathcal{U}_{\mathcal{L}}}{x}p$  iff  $V_{\mathcal{L}}(p, x) = 1$  iff  $p \in x$  (by the truth condition for proposition letters and by the definition of  $V_{\mathcal{L}}$ ).

*Induction Step.* The inductive cases for  $\neg, \wedge, \vee, \rightarrow$  and  $\Leftrightarrow$  are trivial, using the Lemma on  $\mathcal{L}$ -saturated sets. Consider then *Case*  $A = OB$  (for some wff  $B$ ): We would like to argue that  $\frac{\mathcal{U}_{\mathcal{L}}}{x}OB$  iff for all  $y$  in  $W_{\mathcal{L}}$  such that  $xR_{\mathcal{L}}y, \frac{\mathcal{U}_{\mathcal{L}}}{y}B$ , iff, for all  $y$  in  $W_{\mathcal{L}}$  such that  $xR_{\mathcal{L}}y, B \in y$ , iff,  $OB \in x$ . Well, the first ‘iff’ holds by the definition of truth, the second is guaranteed by the inductive hypothesis, and the third ‘iff’ is simply clause (i) of the Saturation Lemma for canonical  $\mathcal{L}$ -models. Thus, we are done. *Case*

$A = PB$ : The reasoning is perfectly analogous, the third ‘iff’ being clause (ii) of that lemma.

The proof of the Coincidence Lemma is complete. ■

**Proof.** *Missing Proof of the Verification Lemma.* As defined in the Definition of Canonical  $\mathcal{L}$ -Models,  $\mathcal{U}_{\mathcal{L}} = \langle W_{\mathcal{L}}, R_{\mathcal{L}}, V_{\mathcal{L}} \rangle$  is an  $\mathcal{L}$ -model.

We have to consider various cases in the proof, depending on how we identify the logic  $\mathcal{L} \in \{\mathbf{OK}, \mathbf{OM}, \mathbf{OS4}, \mathbf{OB}, \mathbf{OS5}, \mathbf{OK}^+, \mathbf{OM}^+, \mathbf{OS4}^+, \mathbf{OB}^+, \mathbf{OS5}^+\}$ . The detailed demonstration in each case will not be given here; instead the reader is referred to Åqvist [1987, Section 10.1.11] for desired details. This completes our account of the Completeness Theorem for the ten Smiley–Hanson systems of monadic deontic logic. ■

IV. REPRESENTABILITY OF MONADIC DEONTIC LOGICS IN  
SYSTEMS OF ALETHIC MODAL LOGIC WITH A PROPOSITIONAL  
CONSTANT

12 TEN ALETHIC MODAL LOGICS WITH A PROPOSITIONAL  
CONSTANT

In this section we define ten systems  $\mathbf{K}_Q, \mathbf{M}_Q, \mathbf{S4}_Q, \mathbf{B}_Q, \mathbf{S5}_Q, \mathbf{K}_Q^+, \mathbf{M}_Q^+, \mathbf{S4}_Q^+, \mathbf{B}_Q^+$  and  $\mathbf{S5}_Q^+$  of alethic modal logic with a *prohairesis*, i.e. preference-theoretical, propositional constant  $Q$  (after Kanger [1957]). They are all based on a common formal language, which we are now going to describe. Its *alphabet* is like that of the language of the Smiley–Hanson systems except that:

- (i)  $\Box$  (necessity) and  $\Diamond$  (possibility) replace  $O$  and  $P$ , respectively, among the primitive logical connectives of degree 1.
- (ii) A propositional constant  $Q$  (for ‘optimality’ or ‘admissibility’) is added to the primitive logical connectives of degree 0.

The set  $\Sigma$  of all *sentences* of our new language is then defined as in the old language except that clause (b) reads:

- (b)  $\top, \perp$  and  $Q$  are in  $S$

and clause (c) reads:

- (c) if  $A$  is in  $S$ , then so are  $\neg A, \Box A$  and  $\Diamond A$ .

We must point out here explicitly that the set Prop of our new alethic language is assume to be identical to the set Prop of our old, deontic language.

DEFINITION 36.

$$OA = \text{df } \Box(Q \rightarrow A)$$

$$PA = \text{df } \Diamond(Q \wedge A)$$

$$FA = \text{df } \Box(Q \rightarrow \neg A)$$

As for the *proof theory* of our ten alethic systems, the following two *rules of inference* are common to all of them:

$$(R1) \quad \frac{A, A \rightarrow B}{B} \quad (\textit{modus ponens})$$

$$(R2') \quad \frac{A}{\Box A} \quad (\Box\text{-necessitation})$$

Consider then the following list B0–B7 of *axiom schemata*:

- (B0) All truth functional tautologies (over our new language)
- (B1)  $\Diamond A \Leftrightarrow \neg \Box \neg A$ ,
- (B2)  $\Box(A \rightarrow B) \rightarrow (\Box A \rightarrow \Box B)$ ,
- (B3)  $\Diamond Q$ ,
- (B4)  $\Box A \rightarrow \Box \Box A$ ,
- (B5)  $\Diamond \Box A \rightarrow \Box A$ ,
- (B6)  $\Box A \rightarrow A$ ,
- (B7)  $\Diamond \Box A \rightarrow A$ ,

Assuming R1 and R2' for all ten alethic modal logics with  $Q$ , then, we define them as follows:

- $\mathbf{K}_Q = \text{B0–B2}$
- $\mathbf{M}_Q = \text{B0–B2, B6}$
- $\mathbf{S4}_Q = \text{B0–B2, B4, B6}$
- $\mathbf{B}_Q = \text{B0–B2, B6, B7}$
- $\mathbf{S5}_Q = \text{B0–B2, B5, B6}$  (note that B4 and B7 are derivable in  $\mathbf{S5}_Q$ )

Again let  $\mathcal{K}$  be any of these five systems. Then:

$$\mathcal{K}^+ = \mathcal{K}, \text{B3}$$

We observe that, apart from the presence of  $Q$  in the alethic language, the first five systems are familiar from literature on basic modal logic (see e.g. Kripke [1963], Makinson [1966], Lemmon and Scott [1966], Hughes and Cresswell [1968]). The remaining five logics, the + systems, are then formed by adding a consistency postulate for  $Q$ , viz. B3, just as in Smiley [1963].

Let  $\mathcal{K}$  be any of the ten systems just defined. We introduce the notions of  *$\mathcal{K}$ -provability*,  *$\mathcal{K}$ -inconsistency*,  *$\mathcal{K}$ -consistency* and  *$\mathcal{K}$ -derivability* in perfect analogy with the corresponding  $\mathcal{L}$ -notions defined for the Smiley–Hanson deontic logics. We write ' $\vdash_{\mathcal{K}} A$ ' and ' $S \vdash_{\mathcal{K}} A$ ' to indicate, respectively, that the sentence  $A$  is  $\mathcal{K}$ -provable and that  $A$  is  $\mathcal{K}$ -derivable from a set  $S$  of sentences.

Turning then to the *semantics* for our ten alethic systems, we obviously need a fresh notion of model. So, by an *alethic model* (perhaps we should even say 'alethic prohairetic model', but it is too long) we shall mean an ordered quadruple

$$U = \langle \mathcal{W}, \mathcal{R}^{\S}, \text{opt}, \mathcal{V} \rangle$$

where:

- (i)  $W$  is a non-empty set.
- (ii)  $R^{\S} \subseteq W \times W$  is a binary relation on  $W$  (of ‘alethic alternativeness’ or ‘alethic accessibility’).
- (iii)  $\text{opt} \subseteq W$  (heuristically,  $\text{opt}$  is to be the set of ‘optimal’, ‘best’ or ‘sufficiently good’ elements of  $W$  according to some unspecified preference ordering on  $W$ ).
- (iv)  $V : \text{Prop} \times W \rightarrow \{1, 0\}$  (as usual).

Now, let  $\mathcal{U}$  be any alethic model, let  $x$  be any member of  $W$ , and let  $A$  be in  $\Sigma$ . The following changes in the definition of *truth at  $x$  in  $\mathcal{U}$*  are then called for: replace the clauses for  $O$  and  $P$  by the following, respectively:

$$\begin{aligned} \frac{\mathcal{U}}{x} \Box A & \text{ iff for every } y \text{ in } W \text{ with } xR^{\S}y, \frac{\mathcal{U}}{y} A. \\ \frac{\mathcal{U}}{x} \Diamond A & \text{ iff for some } y \text{ in } W \text{ with } xR^{\S}y, \frac{\mathcal{U}}{y} A. \end{aligned}$$

Moreover, we add a clause governing the constant  $Q$ :

$$\frac{\mathcal{U}}{x} Q \text{ iff } x \in \text{opt}.$$

*Conditions on  $R^{\S}$  and  $\text{opt}$  in alethic models.* Corresponding to the five axiom schemata B3–B7 we now list five conditions on  $R^{\S}$  and  $\text{opt}$  in an alethic model (adhering to previously adopted notational conventions):

- r3.  $R^{\S}$  is ‘opt-serial’ in  $W$ :  $\forall x \exists y (xR^{\S}y \ \& \ y \in \text{opt})$
- r4.  $R^{\S}$  is transitive in  $W$ .
- r5.  $R^{\S}$  is Euclidean in  $W$ .
- r6.  $R^{\S}$  is reflexive in  $W$ :  $\forall x (xR^{\S}x)$
- r7.  $R^{\S}$  is symmetric in  $W$ :  $\forall x, y (xR^{\S}y \supset yR^{\S}x)$

*Classification of alethic models.* We summarize our categorization of alethic models in the self-explanatory Table 1:

*Validity and satisfiability.* Let  $\mathcal{K} \in \{\mathbf{K}_Q, \mathbf{M}_Q, \mathbf{S4}_Q, \mathbf{B}_Q, \mathbf{S5}_Q, \mathbf{K}_Q^+, \mathbf{M}_Q^+, \mathbf{S4}_Q^+, \mathbf{B}_Q^+, \mathbf{S5}_Q^+\}$ . The notions of  $\mathcal{K}$ -*validity*,  $\mathcal{K}$ -*satisfiability*, and *semantic  $\mathcal{K}$ -entailment* are then defined in perfect analogy with the corresponding  $\mathcal{L}$ -notions, and the notations  $\frac{\mathcal{U}}{\mathcal{K}} A$  and  $S \frac{\mathcal{U}}{\mathcal{K}} A$  will be used with their obvious meaning.

Table 1.

Kind of alethic model	Condition on $R^{\S}$ and opt
$K_Q$	No restriction on $R^{\S}$ or on opt
$M_Q$	$R^{\S}$ reflexive (in $W$ )
$S4_Q$	$R^{\S}$ transitive and reflexive
$B_Q$	$R^{\S}$ symmetric and reflexive
$S5_Q$	$R^{\S}$ Euclidean and reflexive
$K_Q^+$	$R^{\S}$ opt-serial (in $W$ )
$M_Q^+$	$R^{\S}$ opt-serial and reflexive
$S4_Q^+$	$R^{\S}$ opt-serial, transitive and reflexive
$B_Q^+$	$R^{\S}$ opt-serial, symmetric and reflexive
$S5_Q^+$	$R^{\S}$ opt-serial, Euclidean and reflexive (in $W$ )

### 13 SEMANTIC SOUNDNESS AND COMPLETENESS OF THE TEN ALETHIC SYSTEMS

**THEOREM 37.** *Let  $\mathcal{K}$  be any of the ten systems  $\mathbf{K}_Q, \mathbf{M}_Q, \dots, \mathbf{S5}_Q^+$ . Then, all  $\mathcal{K}$ -provable sentences are  $\mathcal{K}$ -valid.*

**Proof.**[Outlined] Proceed just as in the case of the  $\mathcal{L}$ -systems of monadic deontic logic. ■

**EXAMPLE 38.** Suppose we want to check that the axiom  $\diamond Q (=B3)$  is indeed  $\mathbf{K}_Q^+$ -valid. Assume otherwise then, i.e. that for some  $\mathbf{K}_Q^+$ -model  $\mathcal{U} = \langle W, R^{\S}, \text{opt}, \mathcal{V} \rangle$  and some  $x$  in  $W$ , we have:

$$(1) \quad \text{not } \frac{\mathcal{U}}{x} \diamond Q.$$

By the truth conditions for  $\diamond$  and  $Q$ , (1) amounts to:

$$(2) \quad \text{not } \exists y (xR^{\S}y \ \& \ y \in \text{opt}).$$



But, by the opt-seriality of  $R^{\S}$  in  $\mathbf{K}_Q^+$ -models (see the table above), we have:

$$(3) \quad \exists y(xR^{\S}y \ \& \ y \in \text{opt}).$$

Contradiction. Hence B3 is  $\mathbf{K}_Q^+$ -valid.

**THEOREM 39** (Completeness Theorem).

Version I (strong completeness). *Let  $\mathcal{K}$  be as usual and let  $S \subseteq \Sigma$ . Then, if  $S$  is  $\mathcal{K}$ -consistent, then  $S$  is  $\mathcal{K}$ -satisfiable.*

Version II (weak completeness). *Let  $\mathcal{K}$  be as usual. Then, all  $\mathcal{K}$ -valid sentences are  $\mathcal{K}$ -provable.*

**Proof** (Outlined). Obtaining the weak version as a corollary of the strong one, we concentrate on the latter. The Definition of and the Lemma (Lss) on  $\mathcal{L}$ -saturated sets are restated for the  $\mathcal{K}$ -systems without any significant changes. Similarly for Lindenbaum's Lemma. In Makinson's Lemma we replace every reference to  $O$  by a reference to  $\Box$  and use R2' and B2 in the place of R2 and A2; the Lemma then goes through nicely for the  $\mathcal{K}$ -systems as well.

We come next to canonical models:

**DEFINITION 40** (Canonical  $\mathcal{K}$ -models). Let  $\mathcal{K}$  be as usual and let  $S$  be any  $\mathcal{K}$ -consistent set of sentences. By the *canonical  $\mathcal{K}$ -model generated by  $S$*  we mean the structure

$$\mathcal{U}_{\mathcal{K}} = \langle W_{\mathcal{K}}, \mathcal{R}_{\mathcal{K}}^{\S}, \text{opt}_{\mathcal{K}}, \mathcal{V}_{\mathcal{K}} \rangle$$

where:

(i)  $W_{\mathcal{K}}$  = the smallest collection  $U$  of  $\mathcal{K}$ -saturated sets such that:

(a)  $S^+$  is in  $U$ .

(b) If  $x$  is in  $U$ , and  $A$  is a sentence with  $\neg\Box A \in x$ , then  $(x_A)^+$  is in  $U$  (where  $x_A$  is defined as in our reformulated Makinson's Lemma).

(ii)  $\mathcal{R}_{\mathcal{K}}^{\S}$  = the binary relation on  $W_{\mathcal{K}}$  such that for all  $x, y$  in  $W_{\mathcal{K}}$ :

$$xR_{\mathcal{K}}^{\S}y \text{ iff } \forall A(\Box A \in x \supset A \in y).$$

(iii)  $\text{opt}_{\mathcal{K}} = \{x \in W_{\mathcal{K}} : Q \in x\}$ .

(iv)  $\mathcal{V}_{\mathcal{K}}$  = the assignment defined as follows:  $\mathcal{V}_{\mathcal{K}}(p, x) = 1$  iff  $p \in x$  (for all  $p \in \text{Prop}$  and  $x \in W_{\mathcal{K}}$ ).

LEMMA 41 (Verification Lemma.). *As just defined,  $\mathcal{U}_\mathcal{K} = \langle \mathcal{W}_\mathcal{K}, \mathcal{R}_\mathcal{K}^\S, \text{opt}_\mathcal{K}, V_\mathcal{K} \rangle$  is a  $\mathcal{K}$ -model.*

LEMMA 42 (Coincidence Lemma.). *For each sentence  $A$  and each  $x$  in  $W_\mathcal{K}$ :  $\frac{\mathcal{U}_\mathcal{K}}{x} A$  iff  $A \in x$ .*

Waiting a little with the proofs of these lemmata, we establish that they together yield the strong completeness of the systems  $\mathcal{K}$  by an argument perfectly analogous to the one used in connection with the  $\mathcal{L}$ -systems.

Again, the crucial clauses of the *Saturation Lemma for Canonical  $\mathcal{K}$ -Models* are as follows:

- (i)  $\Box A \in x$  iff for all  $y$  in  $W_\mathcal{K}$  with  $xR_\mathcal{K}^\S y$ ,  $A \in y$ .
- (ii)  $\Diamond A \in x$  iff there is a  $y$  in  $W_\mathcal{K}$  such that  $xR_\mathcal{K}^\S y$  and  $A \in y$ .

The proof of this lemma parallels the one given in the  $\mathcal{L}$ -case; just replace  $O$  by  $\Box$ ,  $P$  by  $\Diamond$ , and so on.

**Proof.** *Proof of the Coincidence Lemma.* There is a new case in the induction basis, viz.

*Case  $A = Q$ .* We are to show that  $\frac{\mathcal{U}_\mathcal{K}}{x} Q$  iff  $Q \in x$ . Well, we have that  $\frac{\mathcal{U}_\mathcal{K}}{x} Q$  iff  $x \in \text{opt}_\mathcal{K}$  iff  $Q \in x$ ; where the first ‘iff’ comes from the truth condition for  $Q$  and the second from the definition of  $\text{opt}_\mathcal{K}$  in canonical  $\mathcal{K}$ -models (clause (iii)). So we are done.

The novel cases in the induction step are those where  $A = \Box B$  and  $A = \Diamond B$ ; they are handled in perfect analogy with the cases  $A = OB$  and  $A = PB$  in the corresponding proof for the  $\mathcal{L}$ -systems. For the critical ‘iff’s, appeal to the Saturation Lemma for Canonical  $\mathcal{K}$ -Models. ■

**Proof.** *Proof of the Verification Lemma.* The cases where  $\mathcal{K} \in \{\mathbf{K}_Q, \mathbf{M}_Q, \mathbf{S4}_Q, \mathbf{B}_Q, \mathbf{S5}_Q\}$  are familiar from the literature on basic modal logic (see, e.g. Makinson [1966] and Lemmon and Scott [1966]). The only new thing to be verified in these five cases is that  $\text{opt}_\mathcal{K}$ , as we have defined it, is a subset of  $W_\mathcal{K}$ ; which is a completely trivial point in view of clause (iii) of our Definition of Canonical  $\mathcal{K}$ -Models. As for the remaining five alethic systems, consider:

*Case  $\mathcal{K} = \mathbf{K}_Q^+$ .* We are required to show that the relation  $R_{K_Q^+}^\S$  is opt-serial in  $W_{K_Q^+}$  in the sense that  $\forall x \exists y (xR_{K_Q^+}^\S y \ \& \ y \in \text{opt}_{K_Q^+})$ . Well, in regard of any  $x$  in  $W_{K_Q^+}$ , we have by Lss:

- |    |  |   |
|----|--|---|
| 1. | $\Diamond Q \in x$   | B3 for $\mathbf{K}_Q^+$   |
| 2. | $\exists y(xR_{K_Q^+}^{\S}y \ \& \ Q \in y)$                           | from 1 by clause (ii) of the Saturation Lemma for Canonical $\mathcal{K}$ -Models |
| 3. | $\exists y(xR_{K_Q^+}^{\S}y \ \& \ y \in \text{opt}_{K_Q^+})$          | from 2 by the definition of $\text{opt}_{K_Q^+}$                                  |
| 4. | $\forall x\exists y(xR_{K_Q^+}^{\S}y \ \& \ y \in \text{opt}_{K_Q^+})$ | from 3 by universal generalization, $x$ being <i>any</i> member of $W_{K_Q^+}$    |

where 4 = Q.E.D. The remaining cases present no novelties, so the proof of the Verification Lemma is complete. The Completeness Theorem for the ten alethic  $\mathcal{K}$ -systems is thereby fully proved. ■

## 14 THE PROBLEM OF ISOLATING THE DEONTIC FRAGMENT OF THE $\mathcal{K}$ -SYSTEMS

### 14.1 Problem

Let  $\Sigma$  be the set of sentences of our *alethic* language (common to the  $\mathcal{K}$ -systems) and let  $\Sigma_0$  be the set of sentences of our *deontic* language (common to the Smiley–Hanson  $\mathcal{L}$ -systems); we now need different labels for the two sets. Let  $\mathcal{K}$ , as usual, be any of the ten alethic systems with the constant  $Q$ . Then, *exactly* which sentences in  $\Sigma_0$  are provable in  $\mathcal{K}$ , using the  $\mathcal{K}$ -definitions of  $O$  and  $P$ ? where the latter are:

$$OA = \text{df } \Box(Q \rightarrow A), \quad PA = \text{df } \Diamond(Q \wedge A).$$

In other words, the problem is to characterize, for each  $\mathcal{K}$ , the set of deontic sentences which are provable in  $\mathcal{K}$  on the basis of those two definitions (meaning by ‘deontic sentence any member of  $\Sigma_0$ )  $A$  third formulation of our task: for each  $\mathcal{K}$ , isolate the deontic fragment of  $\mathcal{K}$ !

Now, the locution ‘sentence in  $\Sigma_0$  provable in  $\mathcal{K}$  using the  $\mathcal{K}$ -definitions of  $O$  and  $P$ ’, which crops up in these formulations, is not entirely clear, or, at least, could be made more precise. To that purpose, we suggest that the  $\mathcal{K}$ -definitions of  $O$  and  $P$  in effect amount to there being a certain function which maps our deontic language into the alethic one in the following way:

DEFINITION 43 (The translation  $\phi$  from  $\Sigma_0$  into  $\Sigma$ ). For each sentence  $A$  in  $\Sigma_0$ , define  $\phi(A) \in \Sigma$  by the following recursive conditions:

- (i)  $\phi(p) = p$ , for each proposition letter  $p$ ,
- (ii)  $\phi(\top) = \top$ ,
- (iii)  $\phi(\perp) = \perp$ ,
- (iv)  $\phi(\neg A) = \neg\phi(A)$ ,
- (v)  $\phi(A \wedge B) = \phi(A) \wedge \phi(B)$ .

Similarly for  $\phi(A \vee B)$ ,  $\phi(A \rightarrow B)$  and  $\phi(A \leftrightarrow B)$ .

- (vi)  $\phi(OA) = \Box(Q \rightarrow \phi(A))$
- (vii)  $\phi(PA) = \Diamond(Q \wedge \phi(A))$

Clearly, (vi) and (vii) are the only interesting clauses in this definition, because we easily verify by induction on the length of  $A$  that  $\phi(A) = A$ , provided that  $A$  does not contain  $O$  or  $P$ . Note how (vi) and (vii) correspond to the  $\mathcal{K}$ -definitions of  $O$  and  $P$ . Note also the importance of our assumption that the alethic and the deontic language have *the same* set Prop of proposition letters (why is that assumption important in the present context?)

In the sequel we shall often write  $\phi A$  instead of  $\phi(A)$ .

We need one more definition in order to be able to give a precise formulation of our problem.

**DEFINITION 44** (The deontic fragment of  $\mathcal{K}$  under  $\phi$ ). Let  $\mathcal{K}$  be as usual, and let  $\phi$  be the translation from  $\Sigma_0$  into  $\Sigma$  as just defined. By *the deontic fragment of  $\mathcal{K}$  under  $\phi$*  (in symbols:  $\text{DF}(\mathcal{K}, \phi)$ ) we mean the set of sentences  $A$  in  $\Sigma_0$  such that  $\phi A$  is provable in  $\mathcal{K}$ ; more compactly expressed:

$$\text{DF}(\mathcal{K}, \phi) = \{A \in \Sigma_0 : \vdash_{\mathcal{K}} \phi A\}.$$

Since the translation  $\phi$  is fixed, we may drop the reference to it and speak simply of the deontic fragment of  $\mathcal{K}$ ,  $\text{DF}(\mathcal{K})$ , in accordance with the convention

$$\text{DF}(\mathcal{K}) = \text{DF}(\mathcal{K}, \phi)$$

for  $\mathcal{K}$  as usual.

A precise version of the problem raised at the beginning of this section is then the following:

#### 14.2 *The problem restated*

Let  $\mathcal{K}$  be any of our ten alethic systems. Let  $\mathcal{L}$  be any of the ten Smiley–Hanson deontic logics and let us identify  $\mathcal{L}$  with the set of its theses so that  $\mathcal{L} = \{A \in \Sigma_0 : \vdash_{\mathcal{L}} A\}$ . Then, for which  $\mathcal{L}$ , if any, do we have that  $\mathcal{L} = \text{DF}(\mathcal{K})$ ?

Let us illustrate the import of the restated problem a little. Suppose that we claim that (the set of theses of) **OM** is in fact identical to the deontic fragment of **M<sub>Q</sub>**. What are we then claiming? According to our definitions, the following:

$$(1) \quad \mathbf{OM} = \{A \in \Sigma_0 : \vdash_{\mathbf{OM}} A\} = \{A \in \Sigma_0 : \vdash_{\mathbf{M}_Q} \phi A\} = \text{DF}(\mathbf{M}_Q).$$

Fortunately, a more intelligible rendering of (1) is available:

$$(2) \quad \text{For each sentence } A \text{ in } \Sigma_0 : \vdash_{\mathbf{OM}} A \text{ iff } \vdash_{\mathbf{M}_Q} \phi A$$

i.e. in plain language,  $A$  is provable in **OM** iff its translation  $\phi A$  is provable in **M<sub>Q</sub>** (for any deontic sentence  $A$ ).

Smiley [1963] in effect proved this result (2), i.e. that  $\mathbf{OM} = \text{DF}(\mathbf{M}_Q)$ . He also proved, among other things, that  $\mathbf{OS4} = \text{DF}(\mathbf{S4}_Q)$ ,  $\mathbf{OS5} = \text{DF}(\mathbf{S5}_Q)$ ,  $\mathbf{OM}^+ = \text{DF}(\mathbf{M}_Q^+)$ ,  $\mathbf{OS4}^+ = \text{DF}(\mathbf{S4}_Q^+)$  and  $\mathbf{OS5}^+ = \text{DF}(\mathbf{S5}_Q^+)$ , using algebraic techniques. We shall now restate these Smileyan results and extend them so as to obtain a full solution to the problem raised above. We do so by indicating how to prove a Translation Theorem for monadic deontic logic, applying the Henkin-style model-theoretic technique of saturated sets instead of the matrix method used by Smiley. We think that, by doing so, we not only facilitate the understanding of monadic deontic logics as such, but also will be able to see more clearly their connection with dyadic deontic logics (logics of conditional obligation and permission) and to understand better the transition from the former to the latter.

First of all, let us correlate our alethic systems to the deontic ones by defining a one-one function  $c$  from the former onto the latter. The definition of  $c$  appears from the self-explanatory Table 2.

We can now state a nice result on deontic logic:

**THEOREM 45** (Translation theorem for monadic deontic logic). *(After Smiley [1963]). Let  $\mathcal{K}$  be any of the ten alethic systems  $\mathbf{K}_Q, \mathbf{M}_Q, \dots, \mathbf{S5}_Q^+$ , and let  $c(\mathcal{K})$  be its correlate among the ten Smiley–Hanson systems according to the above table. We identify  $c(\mathcal{K})$  with the set of its theses. Then,  $c(\mathcal{K}) = \text{DF}(\mathcal{K})$ ; i.e. for each sentence  $A$  in  $\Sigma_0 : \vdash_{c(\mathcal{K})} A$  iff  $\vdash_{\mathcal{K}} \phi A$ .*

The proof, which provides a solution to our present problem, is a bit lengthy, so we devote a special section to it.

**Proof. (The Translation Theorem)** (very broad outline)

*Case  $\mathcal{K} = \mathbf{K}_Q$  and  $c(\mathcal{K}) = \mathbf{OK}$ .*

*‘Only if’ part:* We are to show that  $\vdash_{\mathbf{OK}} A$  only if  $\vdash_{\mathbf{K}_Q} \phi A$ , for  $A \in \Sigma_0$ . We do so by induction on the length of the supposed **OK**-proof of  $A$ .

*Basis.* The length of the supposed **OK**-proof = 1, so  $A$  is an instance of one or other of the axiom schemata A0–A2.

Suppose  $A$  is an axiom under A0 so that  $A$  is a tautology over the deontic language. Then  $\phi A$  is a tautology over the alethic language (the detailed

Table 2.

Alethic system $\mathcal{K}$	Deontic system $c(\mathcal{K})$
$K_Q$	OK
$M_Q$	OM
$S4_Q$	OS4
$B_Q$	OB
$S5_Q$	OS5
$K_Q^+$	OK <sup>+</sup>
$M_Q^+$	OM <sup>+</sup>
$S4_Q^+$	OS4 <sup>+</sup>
$B_Q^+$	OB <sup>+</sup>
$S5_Q^+$	OS5 <sup>+</sup>

proof of this is left to the reader), hence  $\phi A$  is an axiom under B0, hence  $\overline{\mathbf{K}_Q} \phi A$ .

Suppose  $A$  is an axiom under A1 so that  $A = PB \Leftrightarrow \neg O \neg B$  and  $\phi A = \Diamond(Q \wedge \phi B) \Leftrightarrow \neg \Box(Q \rightarrow \phi(\neg B))$ , for some  $B \in \Sigma_0$ . The following is then a  $\mathbf{K}_Q$ -proof of  $\phi A$ :

1.  $\Diamond(Q \wedge \phi B) \Leftrightarrow \neg \Box \neg(Q \wedge \phi B)$  B1
2.  $\neg \Box \neg(Q \wedge \phi B) \Leftrightarrow \neg \Box(Q \rightarrow \neg \phi B)$  from B0, B2, R1, R2' by various elementary steps
3.  $\Diamond(Q \wedge \phi B) \Leftrightarrow \neg \Box(Q \rightarrow \phi(\neg B))$  1, 2, B0, R1, definition of  $\phi$

where 3 =  $\phi A$ . Hence,  $\overline{\mathbf{K}_Q} \phi A$ , as desired.

Again, suppose that  $A$  is an instance of A2 so that  $\phi A = \Box(Q \rightarrow (\phi B \rightarrow \phi C)) \rightarrow (\Box(Q \rightarrow \phi B) \rightarrow \Box(Q \rightarrow \phi C))$ , for some  $B$  and  $C$  in  $\Sigma_0$ . The desired result to the effect that  $\overline{\mathbf{K}_Q} \phi A$  is readily obtained from B0 and B2 by R2' and R1.

*Induction Step.* There is an **OK**-proof of  $A$  of length  $> 1$ , and either (i)

$A$  is got by applying R1 to some **OK**-thesis  $B$  and  $B \rightarrow A$ , or (ii)  $A$  is of the form  $OB$  and is obtained by applying R2 to some **OK**-thesis  $B$ .

*Case (i):* By the induction hypothesis  $\phi B$  and  $\phi(B \rightarrow A)$  are both provable in  $\mathbf{K}_Q$ . But, by the definition of  $\phi$ ,  $\phi(B \rightarrow A) = \phi B \rightarrow \phi A$ , so that  $\phi A$  follows by R1. Hence,  $\vdash_{\mathbf{K}_Q} \phi A$ .

*Case (ii):* By the induction hypothesis we have  $\vdash_{\mathbf{K}_Q} \phi B$  in this case. We then obtain  $\vdash_{\mathbf{K}_Q} \phi(OB)$  as follows:

1.  $Q \rightarrow \phi B$   $\vdash_{\mathbf{K}_Q} \phi B, B0, R1$
2.  $\Box(Q \rightarrow \phi B)$  from 1 by R2'
3.  $\phi(OB)$  from 2 by the definition of  $\phi$

where 3 =  $\phi A$ . Hence  $\vdash_{\mathbf{K}_Q} \phi A$ , as desired.

This completes the proof of the ‘only if’ part. ■

*‘If’ part:* We must show that if  $\vdash_{\mathbf{K}_Q} \phi A$ , then  $\vdash_{\mathbf{OK}} A$ , or, contrapositively, that if  $\not\vdash_{\mathbf{OK}} A$ , ( $A$  is *not* **OK**-provable), then  $\not\vdash_{\mathbf{K}_Q} \phi A$  ( $\phi A$  is *not*  $\mathbf{K}_Q$ -provable), for any sentence  $A$  in the deontic language. This part is harder, because proof-theoretical methods seem to be less natural here; however, in view of our soundness and completeness results for the  $\mathcal{L}$ - and the  $\mathcal{K}$ -systems, the problem is not too difficult to cope with.

*Strategy of argument.* We would like to argue as follows:

1.  $\not\vdash_{\mathbf{OK}} A$  hypothesis
2.  $\not\vdash_{\mathbf{OK}} A$  from 1 by the completeness of **OK**
3.  $\not\vdash_x^{\mathcal{U}} A$ , for some **OK**-model  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$  and some  $x$  in  $\mathcal{W}$  from 2 by the definition of **OK**-validity.

Consider that **OK**-model  $\mathcal{U}$ . We claim that we can construct from it a corresponding  $\mathbf{K}_Q$ -model  $\mathcal{U}^* = \langle \mathcal{W}, \mathcal{R}^*, \text{opt}, \mathcal{V} \rangle$  with the property that for all  $B \in \Sigma_0$  and all  $y$  in  $\mathcal{W}$  :  $\vdash_y^{\mathcal{U}} B$  iff  $\vdash_y^{\mathcal{U}^*} \phi B$ . Then:

- |    |                                       |  |
|----|---------------------------------------|--|
| 4. | $\not\vdash_x^{\mathcal{U}^*} \phi A$ | from 3 by the fact that $\mathcal{U}^*$ exists and has the property mentioned  |
| 5. | $\not\vdash_{\mathbf{K}_Q} \phi A$    | from 4 by the definition of $\mathbf{K}_Q$ -validity, $\mathcal{U}^*$ being a $\mathbf{K}_Q$ -model where $\phi A$ fails to be true at some $x$ in $W$ |
| 6. | $\not\vdash_{\mathbf{K}_Q} \phi A$    | from 5 by the soundness of $\mathbf{K}_Q$  |

where 6 is our desired conclusion.

The crux of this argument is obviously isolated at a single point, viz. the construction of the  $\mathbf{K}_Q$ -model  $\mathcal{U}^*$  from the given  $\mathbf{OK}$ -model  $\mathcal{U}$ , and the proof that  $\mathcal{U}^*$  has the desired property indicated above. On the basis of that construction and proof, the crucial step from 3 to 4 is fully justified and the ‘if’ part is seen to go through in the present case. What remains to be done, then, is to state a definition and to prove a couple of nice lemmata.

**DEFINITION 46 (of  $\mathcal{U}^*$ ).** Let  $\mathcal{U} = \langle W, R, \mathcal{V} \rangle$  be *any*  $\mathbf{OK}$ -model. We define  $\mathcal{U}^*$  to be the structure  $\langle W, R^*, \text{opt}, V \rangle$  where:

- (i)  $R^* = R$ .
- (ii)  $\text{opt} = \{y \in W : \text{for some } x \text{ in } W, xRy\}$ .

Note that  $W$  and  $V$  are common to  $\mathcal{U}$  and  $\mathcal{U}^*$ . As for  $V$ , this is made possible by our assumption that the alethic and the deontic language have *the same* set Prop of proposition letters. We may also remark that  $\text{opt}$  is here defined to be what is known as ‘the converse domain’ of the relation  $R$  in  $\mathbf{OK}$ -models.

**LEMMA 47 (Easy Lemma).** *As defined,  $\mathcal{U}^*$  is a  $\mathbf{K}_Q$ -model.*

**Proof.** Appealing to the definition of a  $\mathbf{K}_Q$ -model, we see that it is enough to show (i) that  $R^* \subseteq W \times W$ , and (ii) that  $\text{opt} \subseteq W$ , there being no further restrictions on  $R^{\S}$  and  $\text{opt}$  in such alethic models. These points are immediate in view of (i) and (ii) in the definition of the structure  $\mathcal{U}^*$ . ■

**LEMMA 48 (On relations).** *Let  $\mathcal{U}$  and  $\mathcal{U}^*$  be as in the Definition of  $\mathcal{U}^*$ . Then, for all  $x, y$ , in  $W$ :*

$$xRy \text{ iff } xR^*y \text{ and } y \in \text{opt}.$$

**Proof.**

*Left-to-right:* Assume, for any  $x, y$  in  $W$ :



1.  $xRy$  hypothesis
- Then:
2.  $\exists x(xRy)$  from 1 by existential generalization
  3.  $y \in \text{opt}$  from 2 by the definition of  $\text{opt}$  in  $\mathcal{U}^*$
  4.  $xR^*y$  from 1 by definition of  $R^*$  in  $\mathcal{U}^*$
  5.  $xR^*y \& y \in \text{opt}$  3,4, adjunction

where 5 is the desired conclusion.

*Right-to-left.* This direction is immediate by the definition of  $R^*$ . ■

LEMMA 49 (Crucial Lemma). *Let  $\mathcal{U}$  and  $\mathcal{U}^*$  be as in the Definition of  $\mathcal{U}^*$ . Then, for all  $A \in \Sigma_0$  and for all  $x$  in  $W$ :  $\frac{\mathcal{U}}{x} A$  iff  $\frac{\mathcal{U}^*}{x} \phi A$ .*

**Proof.** By induction on the length of  $A$ . By the definition of the translation  $\phi$ , the three cases in the induction basis are seen to be trivial. For the same reason, the inductive cases involving truth-functional connectives go through easily. Consider then:

*Case  $A = OB$ .* We are required to show that  $\frac{\mathcal{U}}{x} OB$  iff  $\frac{\mathcal{U}^*}{x} \phi(OB)$ .  
Well, for any  $B \in \Sigma_0$  and any  $x$  in  $W$ , we clearly have:

1.  $\frac{\mathcal{U}}{x} OB$  iff  $\forall y(xRy \supset \frac{\mathcal{U}}{y} B)$  by the definition of truth in  $\mathcal{U}$
2.  $\frac{\mathcal{U}^*}{x} \Box(Q \rightarrow \phi B)$  iff  $\forall y(xR^*y \& y \in \text{opt} \supset \frac{\mathcal{U}^*}{y} \phi B)$  by the definition of truth in  $\mathcal{U}^*$
3.  $\frac{\mathcal{U}}{y} B$  iff  $\frac{\mathcal{U}^*}{y} \phi B$  by the induction hypothesis,  $y$  being any member of  $W$

Hence:

4. (Right member of 1) iff (right member of 2) from 3 and the Lemma on Relations by elementary steps
5.  $\frac{\mathcal{U}}{x} OB$  iff  $\frac{\mathcal{U}^*}{x} \Box(Q \rightarrow \phi B)$  from 1,2,4 by the transitivity of 'iff'

where 5 yields the desired result by the definition of  $\phi$ .

*Case  $A = PB$ .* The reasoning parallels that of the preceding case; we just make the necessary switches from  $O, \Box, \rightarrow, \forall, \supset$  to  $P, \Diamond, \wedge, \exists$  and  $\&$ , respectively.

The proof of the Crucial Lemma is complete.

Armed with the definition of the  $\mathbf{K}_Q$ -model  $\mathcal{U}^*$  and our three lemmas on it, we have fully justified the decisive step from 3 to 4 in our strategic argument given above. This completes the proof of the ‘if’ part of *Case*  $\mathcal{K} = \mathbf{K}_Q$  and  $c(\mathcal{K}) = \mathbf{OK}$ . That case is thereby fully proved. ■

*Remaining Cases.* For the details of the remaining nine cases, see Åqvist [1987, Section 13.5.2–10]. In the present survey I only want to indicate the main novelty in each case, which appears in the proof of the ‘if’ part at the juncture where, given any  $c(\mathcal{K})$ -model  $\mathcal{U}$ , we construct a corresponding  $\mathcal{K}$ -model  $\mathcal{U}^*$  with the ‘right’ properties. Thus, in general, we lay down a definition of this form:

DEFINITION 50 (Definition of  $\mathcal{U}$ ). Let  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$  be any  $c(\mathcal{K})$ -model, so that  $R$  satisfies the appropriate restriction. Define  $\mathcal{U}^*$  to be the structure  $\langle W, R^*, \text{opt}, V \rangle$  where:

- (i)  $R^*$  = the binary relation on  $W$  such that for all  $x, y$  in  $W$  :  $xR^*y$  iff
- (ii)  $\text{opt} = \{y \in W : \text{for some } x \text{ in } W, xRy\} =$  the converse domain of  $R$  in  $W$ .

Having filled in the blank in (i) for each case, one goes on to state and prove a

LEMMA 51 (Easy). *As defined,  $\mathcal{U}^*$  is a  $\mathcal{K}$ -model*

as well as the

LEMMA 52 (On relations). *Let  $\mathcal{U}$  and  $\mathcal{U}^*$  be as in the above Definition of  $\mathcal{U}^*$ . Then, for all  $x, y$  in  $W$  :  $xRy$  iff  $xR^*y$  and  $y \in \text{opt}$ .*

Using this lemma on Relations, we give an inductive proof of the

LEMMA 53 (Crucial).  $\frac{\mathcal{U}}{x}A$  iff  $\frac{\mathcal{U}^*}{x}\phi A$  for  $\mathcal{U}$  and  $\mathcal{U}^*$  as above, and for  $A$  and  $x$  as usual), just as in the case of  $\mathcal{K} = \mathbf{K}_Q$  and  $c(\mathcal{K}) = \mathbf{OK}$ .

Again, armed with the definition of the  $\mathcal{K}$ -model  $\mathcal{U}^*$  and our three lemmas on it, we justify the decisive step from 3 to 4 in the strategic argument for the ‘if’ part. This will then complete the proof in each of the remaining nine cases.

We now indicate how to fill in the blank in clause (i) of the Definition of  $\mathcal{U}^*$ , for various cases.

*Case*  $\mathcal{K} = \mathbf{M}_Q$  and  $c(\mathcal{K}) = \mathbf{OM}$ . Fill in the blank with this condition: ( $x = y$  or  $xRy$ ).

*Case*  $\mathcal{K} = \mathbf{S4}_Q$  and  $c(\mathcal{K}) = \mathbf{OS4}$ . Fill in the blank with the same condition!

*Case*  $\mathcal{K} = \mathbf{B}_Q$  and  $c(\mathcal{K}) = \mathbf{OB}$ . Fill in the blank with the condition ( $x = y$  or  $xRy$  or  $yRx$ )

*Case*  $\mathcal{K} = \mathbf{S5}_Q$  and  $c(\mathcal{K}) = \mathbf{OS5}$ . Fill in the blank with this condition:

For some natural number  $n \geq 1$  :  $xR^n y$  where  $\mathcal{R}$  is the relation on  $W$  defined by:  $xRy$  iff ( $x = y$  or  $xRy$  or  $yRx$ ) and where  $\mathcal{R}^n$  is the  $n$ th power of the relation  $\mathcal{R}$ , defined in the usual inductive way in terms of relative products. Thus, in the present case,  $R^*$  is defined as the *chain*, or *proper ancestral*, of the relation  $\mathcal{R}$ . Certain inductively provable additional lemmata are then needed to establish the Easy Lemma and the Lemma on Relations in this case, which is more complicated than the preceding ones.

*The five remaining + cases.* Define  $R^*$  just as in the corresponding case *without* +, i.e. the case where the axiomatic systems lack the schemata **B3** and **A3** and where the accessibility relations are not required to be opt-serial or serial in the relevant models.

Our broad outline of the proof of the Translation Theorem for Monadic Deontic Logic is complete. ■

## V. FIRST STEPS IN DYADIC DEONTIC LOGIC

### 15 TWO NEW LANGUAGES AND A PROBLEM

Consider the deontic language common to the Smiley–Hanson monadic systems and its set  $\Sigma_0$  of well formed sentences. Let us now think of  $O$  and  $P$  as *dyadic* (i.e. two-place) deontic connectives, expressing *conditional* obligation and permission, respectively. We then obtain a new deontic language, the set of sentences of which will be called  $\Sigma_0^2$  and is defined as the smallest set  $S$  such that:

- (a) Every proposition letter is in  $S$ .
- (b)  $\top$  and  $\perp$  are in  $S$ .
- (c) If  $A$  is in  $S$ , then so is  $\neg A$ .
- (d) If  $A, B$  are in  $S$ , then so are  $(A \wedge B)$ ,  $(A \vee B)$ ,  $(A \rightarrow B)$  and  $(A \Leftrightarrow B)$ .
- (e) If  $A, B$  are in  $S$ , then so are  $O_B A$  and  $P_B A$ .

**REMARK 54.** Apart from the fact that clause (c) has been curtailed, the new thing about the present deontic language and its set  $\Sigma_0^2$  is of course embodied in clause (e). So, note that we write  $O_B A$  and  $P_B A$ , where

many authors would write  $O(A/B)$  and  $P(A/B)$  and where still others would use  $(BOA)$  and  $(BPA)$ ; the former authors are obviously inspired by the notation familiar from *probability theory*, whereas the latter (e.g. Van Eck [1981]) stick to what might be labeled the *standard binary connective* notation (cf. clause (d) above). The motivation for our choice of notation will appear from what follows; it might be called a *relative necessity* or *sententially indexed modality* notation (cf. e.g. Chellas [1975, Section 5]).

DEFINITION 55.

*'Dyadic' Definitions of monadic deontic connectives:*

$$\begin{aligned} OA &=df \quad O_{\top}A \\ PA &=df \quad P_{\top}A \\ FA &=df \quad \neg P_{\top}A \quad (\text{alternatively: } O_{\top}\neg A) \end{aligned}$$

Again, consider the alethic language common to the systems  $\mathcal{K}$  and its set  $\Sigma$  of well formed sentences. In this language  $Q$  was thought of as a propositional *constant*, i.e. as a zero-place connective, thus of degree 0. Now, think of  $Q$  as a *monadic* (i.e. one-place) prohairetic connective, so that  $QA$  might be read as 'optimally  $A$ ', 'ideally  $A$ ', or what have you. We then obtain a new alethic language with an additional one-place connective  $Q$ ; its set of *sentences* will be called  $\Sigma^1$  and is defined as the smallest set  $S$  such that:

- (a) Every proposition letter is in  $S$ .
- (b)  $\top$  and  $\perp$  are in  $S$ .
- (c) If  $A$  is in  $S$ , then so are  $\neg A$ ,  $\Box A$ ,  $\Diamond A$  and  $QA$ .
- (d) As usual.

As compared to the old alethic language with  $Q$ , the new one has just the nullary connectives  $\top$  and  $\perp$  (clause (b)), whereas  $Q$  reappears among the monadic ones (clause (c)).

### 15.1 'Alethic' Definitions of dyadic deontic connectives

$$\begin{aligned} Q & \quad (\text{the old propositional constant}) =df \quad Q\top \\ O_BA & =df \quad \Box(QB \rightarrow A), \\ P_BA & =df \quad \Diamond(QB \wedge A), \\ F_BA & =df \quad \Box(QB \rightarrow \neg A). \end{aligned}$$

We can now state the problem announced in the title to this section.

### 15.2 Problem

Consider the two sets of sentences  $\Sigma^1$  and  $\Sigma_0^2$ . As usual, the new alethic language and the new deontic one are assumed to have *the same* set Prop of proposition letters. Corresponding to the definitions in  $\Sigma^1$  of the dyadic connectives  $O$  and  $P$ , define a translation  $\phi$  from  $\Sigma_0^2$  into  $\Sigma^1$  just as in the case of  $\Sigma_0$  and  $\Sigma$ , except for the following fresh clauses:

$$(vi) \quad \phi(O_B A) = \Box(Q\phi B \rightarrow \phi A).$$

$$(vii) \quad \phi(P_B A) = \Diamond(Q\phi B \wedge \phi A).$$

Then, find a system  $\mathcal{L}$  of dyadic deontic logic and a system  $\mathcal{K}$  of alethic modal logic with our new monadic connective  $Q$  such that:

- (i) The set of  $\mathcal{L}$ -theses is a proper subset of  $\Sigma_0^2$ .
- (ii) The set of  $\mathcal{K}$ -theses is a proper subset of  $\Sigma^1$ .
- (iii) The set of  $\mathcal{L}$ -theses = the dyadic deontic fragment of  $\mathcal{K}$  under  $\phi$  as just defined; i.e. we are to have for each sentence  $A$  in  $\Sigma_0^2$ :  $\vdash_{\mathcal{L}} A$  iff  $\vdash_{\mathcal{K}} \phi A$ .

The word ‘proper’ is inserted in requirements (i) and (ii) just in order to make sure that the logics  $\mathcal{L}$  and  $\mathcal{K}$  are consistent. We now consider, to start with, certain rather weak systems having the desired properties (i)–(iii).

## 16 THE SYSTEMS $O_{dy}S4$ , $O_{dy}S5$ , $S4_{Qmo}$ AND $S5_{Qmo}$

The system  $O_{dy}S4$  is determined as follows.

*Rules of proof:*

$$(R1) \quad \frac{A, A \rightarrow B}{B} \quad (\textit{modus ponens})$$

$$(R2) \quad \frac{A}{O_B A} \quad (O_B\text{-necessitation})$$

*Axiom schemata:*

- (a0) All truth functional tautologies over  $\Sigma_0^2$
- (a1)  $P_B A \Leftrightarrow \neg O_B \neg A$ ,
- (a2)  $O_B(A \rightarrow C) \rightarrow (O_B A \rightarrow O_B C)$ ,
- (a3)  $O_B(O_B A \rightarrow A)$ ,
- (a4)  $O_B A \rightarrow O_C O_B A$ ,

The axiomatic system  $\mathbf{O}_{dy}\mathbf{S5}$  results from  $\mathbf{O}_{dy}\mathbf{S4}$  by omitting schema a3 and by adding in its place the following new schema a5:

- (a5)  $P_C O_B A \rightarrow O_B A$ ,

Note that a3 will be derivable in  $\mathbf{O}_{dy}\mathbf{S5}$  as a thesis schema.

The notions of *provability*, consistency, derivability etc. are defined for the systems  $\mathbf{O}_{dy}\mathbf{S4}$  and  $\mathbf{O}_{dy}\mathbf{S5}$  in the usual straightforward way.

The axiomatic system  $\mathbf{S4}_{Qmo}$  is simply the familiar modal calculus  $\mathbf{S4}$  over the present alethic language whose set of sentences =  $\Sigma^1$ ; similarly, the system  $\mathbf{S5}_{Qmo}$  is  $\mathbf{S5}$  over that alethic language. See Section 12 above.

We now turn to the *semantics* of the four systems just described. By an  $\mathbf{O}_{dy}\mathbf{S4}$ -model we shall mean any ordered triple

$$U = \langle W, R, V \rangle$$

where:

- (i)  $W$  is a non-empty set and  $V$  is a function from  $\text{Prop} \times W$  into the set of truth-values  $\{1,0\}$ ; thus,  $W$  and  $V$  are as usual.
- (ii)  $R$  is a function from the set of sentences  $\Sigma_0^2$  into the set of all binary relations on  $W$ , in symbols:  $R : \Sigma_0^2 \rightarrow \mathcal{P}(W \times W)$ . In other words, then: for each sentence  $B$  in  $\Sigma_0^2$ ,  $R_B \subseteq W \times W$  so that  $R_B$  is a binary relation on  $W$ . Moreover,  $R$  is to satisfy the following two conditions, corresponding to axiom schemata a3 and a4:
  - (3) For each  $B$  in  $\Sigma_0^2$  and any  $x, y$  in  $W$  :  $xR_B y \supset yR_B y$ .
  - (4) For any  $B, C$  in  $\Sigma_0^2$  and  $x, y, z$  in  $W$  :  $xR_C y \ \& \ yR_B z \supset xR_B z$ .

Again, by an  $\mathbf{O}_{dy}\mathbf{S5}$ -model we mean any  $\mathbf{O}_{dy}\mathbf{S4}$ -model  $\langle W, R, V \rangle$  where  $R$  satisfies the following new restriction that corresponds to the schema a5:

- (5) For any  $B, C$  in  $\Sigma_0^2$  and  $x, y, z$  in  $W : xR_Cy \ \& \ xR_Bz \supset yR_Bz$ .

*Truth conditions for dyadic deontic connectives:* Let  $U = \langle W, \mathcal{R}, \mathcal{V} \rangle$  be any  $\mathbf{O}_{dy}\mathbf{S4}$ - or  $\mathbf{O}_{dy}\mathbf{S5}$ -model, let  $x$  be any member of  $W$ , and let  $A$  be in  $\Sigma_0^2$ . The only change in the definition of *truth at  $x$  in  $U$* , given for our monadic deontic logics, will concern sentences of the new forms  $O_BA$  and  $P_BA$ , for which we adopt the following clauses:

$$\begin{aligned} \frac{U}{x} O_B A & \text{ iff for every } y \text{ in } W \text{ such that } xR_B y, \frac{U}{y} A. \\ \frac{U}{x} P_B A & \text{ iff for some } y \text{ in } W \text{ such that } xR_B y, \frac{U}{y} A. \end{aligned}$$

Notions of *validity*, satisfiability and semantic entailment, which are relative to  $\mathbf{O}_{dy}\mathbf{S4}$  and  $\mathbf{O}_{dy}\mathbf{S5}$ , are then introduced in the usual way.

Furthermore, define an  $\mathbf{S4}_{Qmo}$ -model to be any ordered quadruple

$$U = \langle W, \mathcal{R}^\S, \text{opt}, \mathcal{V} \rangle$$

with  $W, R^\S, V$  as in an  $\mathbf{S4}_Q$ -model so that  $R^\S$  is reflexive and transitive relation on  $W$ , and where:

$$\text{opt} : \Sigma^1 \rightarrow \mathcal{P}W$$

i.e.  $\text{opt}$  is a function which to each sentence  $B$  in  $\Sigma^1$  assigns a subset  $\text{opt}(B)$  of  $W$  as its value.

*Truth condition for the monadic  $Q$ -connective:* The old truth condition for the Kanger constant  $Q$  will have to be replaced by the following:

$$\frac{U}{x} QB \text{ iff } x \in \text{opt}(B)$$

which then governs sentences of the form  $QB$ .

Again, an  $\mathbf{S5}_{Qmo}$ -model is any  $\mathbf{S4}_{Qmo}$ -model where  $R^\S$  has the additional property of being *symmetric* and, hence, an *equivalence relation*, on  $W$ . Validity etc. is defined in the usual way relatively to  $\mathbf{S4}_{Qmo}$  and  $\mathbf{S5}_{Qmo}$ .

**THEOREM 56** (Soundness and completeness).

- (i) For each  $A$  in  $\Sigma_0^2$  :  $A$  is provable in  $\mathbf{O}_{dy}\mathbf{S4}$  /  $\mathbf{O}_{dy}\mathbf{S5}$  / iff  $A$  is valid in  $\mathbf{O}_{dy}\mathbf{S4}$  /  $\mathbf{O}_{dy}\mathbf{S5}$  /.
- (ii) For each  $A$  in  $\Sigma^1$  :  $A$  is provable in  $\mathbf{S4}_{Qmo}$  /  $\mathbf{S5}_{Qmo}$  / iff  $A$  is valid in  $\mathbf{S4}_{Qmo}$  /  $\mathbf{S5}_{Qmo}$  /.

**Proof.** See Åqvist [1987, Sections 15 and 16]. ■

**THEOREM 57** (Translation). *For each  $A$  in  $\Sigma_0^2$ :  $A$  is provable in  $\mathbf{O}_{dy}\mathbf{S4}/\mathbf{O}_{dy}\mathbf{S5}/$  iff  $\phi A$  is provable in  $\mathbf{S4}_{Qmo}/\mathbf{S5}_{Qmo}/$ . In other words, the set of  $\mathbf{O}_{dy}\mathbf{S4}/\mathbf{O}_{dy}\mathbf{S5}/$  theses = the dyadic deontic fragment (under  $\phi$ ) of  $\mathbf{S4}_{Qmo}/\mathbf{S5}_{Qmo}/$ .*

**Proof.** See Åqvist [1987, Sections 15 and 16]. The demonstrations are a little bit tedious. ■

## 17 TWO NEW SYSTEMS: $\mathbf{O}_{dy}\mathbf{S5}^N$ AND $\mathbf{S5}_{Qmo}^N$

At this juncture we observe that in the systems  $\mathbf{S4}_{Qmo}$  and  $\mathbf{S5}_{Qmo}$  there are no special axioms governing the monadic operator  $Q$ ; correspondingly, there are no restrictions on the function  $\text{opt}$  in the models for these systems. So, the following is a natural expectation: if we start adding axioms for  $Q$  to these systems as well as matching restrictions on  $\text{opt}$  in their modellings, we should obtain a series of dyadic deontic logics as the deontic fragments of these extended alethic calculi with monadic  $Q$ ; and to begin with, we are particularly interested in dyadic deontic logics that are in relevant respects similar to the systems **DSDL1–DSDL3** proposed by Bengt Hansson [1969]. As it turns out, however,  $\mathbf{S4}_{Qmo}$  and  $\mathbf{S5}_{Qmo}$  are not quite fit to serve as adequate bases for a development of dyadic deontic logic along those lines. But they come close to the basic system we are looking for; only one further step has to be taken.

Consider the following alethic system  $\mathbf{S5}_{Qmo}^N$ : its set of *theses* is identical to that of  $\mathbf{S5}_{Qmo}$ , but we want  $\Box$  to be interpreted as what Scott [1970] calls *universal necessity* and what Kanger [1957] called ‘analytic’ necessity. This means simply that  $\Box$  is to express truth at *every* world (point) in the set (space)  $W$ , unconditionally. Technically, we easily achieve this by defining an  $\mathbf{S5}_{Qmo}^N$ -model as any structure  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}^{\S}, \text{opt}, \mathcal{V} \rangle$ , where  $W$ ,  $\text{opt}$ ,  $V$  are as usual *and where*  $R^{\S} = W \times W$  (i.e. the universal binary relation on  $W$ ). So, in contrast to the case of  $\mathbf{S5}_{Qmo}$ -models,  $R^{\S}$  is no longer an arbitrary equivalence relation on  $W$ , but is now identified with a *particular* equivalence relation on  $W$ , viz.  $W \times W$ ; and the set of  $\mathbf{S5}_{Qmo}^N$ -models becomes a proper subset of the set of  $\mathbf{S5}_{Qmo}$ -models. As for the *completeness* of the system  $\mathbf{S5}_{Qmo}^N$  as just characterized, a proof can be extracted from Åqvist [1973, Section 5]; note in particular the treatment and the role of the operator  $\Box$  in that essay.

Now, what is the dyadic deontic fragment under  $\phi$  of  $\mathbf{S5}_{Qmo}^N$ ? If you believe it to be  $\mathbf{O}_{dy}\mathbf{S5}$  (once again), then try to prove a translation theorem for  $\mathbf{O}_{dy}\mathbf{S5}$  and  $\mathbf{S5}_{Qmo}^N$  along the familiar lines! You will find that it doesn’t work, because we will be unable to establish either the Easy Lemma or the Lemma on Relations. Instead, the correct answer to the above question is: the following system  $\mathbf{O}_{dy}\mathbf{S5}^N$ , which we are now going to describe quickly.



Let us reconsider our dyadic deontic language with the set of sentences  $= \Sigma_0^2$ , and add a pair of one-place modal operators  $N$  and  $M$  to its stock of primitive logical connectives.  $N(M)$  is to express *universal* necessity (possibility) in the sense indicated above. The set of *sentences* of the dyadic deontic language thus enriched will be called  $\Sigma_{0,N}^2$ ; adjusting the formation rule (c) in the obvious way, we have that whenever  $A$  is in  $\Sigma_{0,N}^2$ , so are  $NA$  and  $MA$ .

As for the *proof theory* of  $\mathbf{O}_{dy}\mathbf{S5}^N$ , its set of theses is a proper subset of this new set of sentences  $\Sigma_{0,N}^2$ . More precisely, it is determined by the following rules of inference and axiom schemata:

$$(R1) \quad \frac{A, A \rightarrow B}{B} \quad (\textit{modus ponens})$$

$$(R2'') \quad \frac{A}{NA} \quad (N\text{-necessitation})$$

(a0) All truth functional tautologies over  $\Sigma_{0,N}^2$

$$(a1) \quad P_B A \Leftrightarrow \neg O_B \neg A$$

$$(a2) \quad O_B(A \rightarrow C) \rightarrow (O_B A \rightarrow O_B C)$$

$$(a6) \quad O_B A \rightarrow NO_B A$$

$$(a7) \quad NA \rightarrow O_B A$$

(a8) An appropriate set of **S5**-schemata for  $N$  and  $M$  (e.g. B1, B2, B5 and B6, with  $N, M$  respectively replacing  $\Box, \Diamond$ ).

EXERCISE 58. Derive the schema  $P_B A \rightarrow NP_B A$  in  $\mathbf{O}_{dy}\mathbf{S5}^N$  as just described! Derive the system  $\mathbf{O}_{dy}\mathbf{S5}$  as a subsystem of  $\mathbf{O}_{dy}\mathbf{S5}^N$ !

Proceeding to the *semantics* for  $\mathbf{O}_{dy}\mathbf{S5}^N$ , we define an  $\mathbf{O}_{dy}\mathbf{S5}^N$ -model as any structure  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$ , where  $W, V$  are as usual and where  $R : \Sigma_{0,N}^2 \rightarrow \mathcal{P}(\mathcal{W} \times \mathcal{W})$  is a function from our fresh set of sentences  $\Sigma_{0,N}^2$  into the set of all binary relations on  $W$ , satisfying the following condition that corresponds to schema (a6):

$$(6) \quad \text{For each } B \text{ in } \Sigma_{0,N}^2 \text{ and any } x, y, z \text{ in } W : xR_B y \supset zR_B y.$$

EXERCISE 59. Show that in any  $\mathbf{O}_{dy}\mathbf{S5}^N$ -model  $R$  satisfies the restrictions (4) and (5)!

Clearly, we must supplement our earlier definition of *truth at  $x$  in  $\mathcal{U}$* , where  $\mathcal{U}$  is any  $\mathbf{O}_{dy}\mathbf{S5}^N$ -model, with new clauses governing the operators  $N$  and  $M$ :

$$\begin{aligned} \frac{\mathcal{U}}{x} NA & \text{ iff for each } y \text{ in } W, \frac{\mathcal{U}}{y} A \\ \frac{\mathcal{U}}{x} MA & \text{ iff for some } y \text{ in } W, \frac{\mathcal{U}}{y} A \end{aligned}$$

The notions of validity, satisfiability and semantic entailment, pertaining to  $\mathbf{O}_{dy}\mathbf{S5}^N$ , are then defined in the usual way.

**THEOREM 60** (Soundness and completeness for  $\mathbf{O}_{dy}\mathbf{S5}^N$ ). *For each  $A$  in  $\Sigma_{0,N}^2$ :*

$$\frac{}{\mathbf{O}_{dy}\mathbf{S5}^N A} \text{ iff } \frac{}{\mathbf{O}_{dy}\mathbf{S5}^N A}$$

**Proof.** See Åqvist [1987, Section 17.2]. ■

### 17.1 Representation of $\mathbf{O}_{dy}\mathbf{S5}^N$ in $\mathbf{S5}_{Qmo}^N$

In this section we announce the result that the set of  $\mathbf{O}_{dy}\mathbf{S5}^N$ -theses is the dyadic deontic fragment under  $\phi$  of  $\mathbf{S5}_{Qmo}^N$ . Clearly, then,  $\phi$  should be a translation from the new sentence-set  $\Sigma_{0,N}^2$  into  $\Sigma^1$ , which is effected by adding the following clauses to our definition of  $\phi$ :

- (viii)  $\phi(NA) = \Box\phi A$ ,
- (ix)  $\phi(MA) = \Diamond\phi A$ ,

**THEOREM 61** (Translation for  $\mathbf{O}_{dy}\mathbf{S5}^N$  and  $\mathbf{S5}_{Qmo}^N$ ).

*For each  $A$  in  $\Sigma_{0,N}^2$ :*

$$\frac{}{\mathbf{O}_{dy}\mathbf{S5}^N A} \text{ iff } \frac{}{\mathbf{S5}_{Qmo}^N \phi A}$$

**Proof.** See Åqvist [1987, Section 17.3.1]. ■

## VI. DEVELOPMENT OF DYADIC DEONTIC LOGIC THROUGH AXIOMATIC ADDITIONS TO THE SYSTEMS $\mathbf{O}_{dy}\mathbf{S5}^N$ AND $\mathbf{S5}_{Qmo}^N$

### 18 THE DYADIC CALCULI $\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i$

In this part we shall take  $\mathbf{O}_{dy}\mathbf{S5}^N$  as our *basic* and, in a certain sense, *minimal* system of dyadic deontic logic and form new calculi by adding to it one or more axiom schemata  $\alpha i$  from the following list  $\alpha 0$ – $\alpha 4$ :

- $\alpha 0.$   $N(A \Leftrightarrow B) \rightarrow (O_A C \Leftrightarrow O_B C)$
- $\alpha 1.$   $O_A A$
- $\alpha 2.$   $O_{A \wedge B} C \rightarrow O_A (B \rightarrow C)$
- $\alpha 3.$   $MA \rightarrow (O_A B \rightarrow P_A B)$
- $\alpha 4.$   $P_A B \rightarrow (O_A (B \rightarrow C) \rightarrow O_{A \wedge B} C)$

To start with, we consider the five systems  $\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i$ , for  $i = 0, 1, \dots, 4$ ; where  $\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i$  is the calculus which results from  $\mathbf{O}_{dy}\mathbf{S5}^N$  by adding *just* the schema  $\alpha i$  to the latter.

Turning quickly to semantics, we define, for  $i = 0, 1, \dots, 4$ , an  $\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i$  model as any  $\mathbf{O}_{dy}\mathbf{S5}^N$ -model  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}, \mathcal{V} \rangle$  where  $R$ , in addition to meeting (6), satisfies the condition  $\rho i$  in the list  $\rho 0$ – $\rho 4$  below of restrictions on  $R$ ; where  $A, B$  are any members of  $\Sigma_{0,N}^2$ ,  $x, y$  any members of  $W$ , and where, for each  $A$  in  $\Sigma_{0,N}^2$ ,

$$\| A \|^\mathcal{U} = \{y \in W : \left| \frac{\mathcal{U}}{y} A \right\}$$

(in other words,  $\| A \|^\mathcal{U}$  is to be the *truth-set* or *extension in  $\mathcal{U}$*  of the wff  $A$ ):

- $\rho 0.$   $\| A \|^\mathcal{U} = \| B \|^\mathcal{U} \supset R_A = R_B$
- $\rho 1.$   $xR_A y \supset \left| \frac{\mathcal{U}}{y} A \right|$
- $\rho 2.$   $xR_A y \ \& \ \left| \frac{\mathcal{U}}{y} B \right| \supset xR_{A \wedge B} y$
- $\rho 3.$   $\| A \|^\mathcal{U} \neq \emptyset \supset \forall x \exists y (xR_A y)$
- $\rho 4.$   $\exists z (xR_A z \ \& \ \left| \frac{\mathcal{U}}{z} B \right|) \supset (xR_{A \wedge B} y \supset (xR_A y \ \& \ \left| \frac{\mathcal{U}}{y} B \right|))$

**THEOREM 62** (Soundness and completeness). *For each  $i = 0, 1, \dots, 4$  and for each  $A$  in  $\Sigma_{0,N}^2$ :  $\left| \frac{\mathcal{U}}{\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i} A \right|$  iff  $\left| \frac{\mathcal{U}}{\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i} A \right|$ . In other words, the sentences provable in  $\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i$  are exactly the sentences valid in that system.*

**Proof.** See Åqvist [1987, Section 18.0–1]. ■

### 19 THE ALETHIC CALCULI $\mathbf{S5}_{QMO}^N + \beta I$

Consider the following list  $\beta 0$ – $\beta 4$  of axiom schemata that may be added to the system  $\mathbf{S5}_{Qmo}^N$ ; they all govern our monadic operator  $Q$ :

- $\beta 0.$   $\Box(A \Leftrightarrow B) \rightarrow \Box(QA \Leftrightarrow QB),$
- $\beta 1.$   $QA \rightarrow A,$
- $\beta 2.$   $(QA \wedge B) \rightarrow Q(A \wedge B),$
- $\beta 3.$   $\Diamond A \rightarrow \Diamond QA,$
- $\beta 4.$   $\Diamond(QA \wedge B) \rightarrow \Box(Q(A \wedge B) \rightarrow (QA \wedge B)).$

By the system  $\mathbf{S5}_{Qmo}^N + \beta i$ , for  $i = 0, 1, \dots, 4$ , we mean the calculus which results from  $\mathbf{S5}_{Qmo}^N$  by adding *just* the schema  $\beta i$  to the latter. There are then five systems of this sort to be considered.

Moving on to semantics, we define, for  $i = 0, 1, \dots, 4$ , an  $\mathbf{S5}_{Qmo}^N + \beta i$  model as any structure  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}^{\S}, \text{opt}, V \rangle$  with  $W, V$  as usual, where  $R^{\S} = W \times W$  (just as in  $\mathbf{S5}_{Qmo}^N$ -models) and where  $\text{opt}: \Sigma^1 \rightarrow \mathcal{P}W$  satisfies the condition  $\sigma_i$  in the list  $\sigma 0$ – $\sigma 4$  below of restrictions on  $\text{opt}$  (where  $A, B$  are any sentences in  $\Sigma^1$ ):

- $\sigma 0.$   $\| A \|_{\mathcal{U}} = \| B \|_{\mathcal{U}} \supset \text{opt}(A) = \text{opt}(B)$
- $\sigma 1.$   $\text{opt}(A) \subseteq \| A \|_{\mathcal{U}}$
- $\sigma 2.$   $\text{opt}(A) \cap \| B \|_{\mathcal{U}} \subseteq \text{opt}(A \wedge B)$
- $\sigma 3.$   $\| A \|_{\mathcal{U}} \neq \emptyset \supset \text{opt}(A) \neq \emptyset$
- $\sigma 4.$   $\text{opt}(A) \cap \| B \|_{\mathcal{U}} \neq \emptyset \supset (\text{opt}(A \wedge B) \subseteq \text{opt}(A) \cap \| B \|_{\mathcal{U}})$

**THEOREM 63** (Soundness and completeness). *For each  $i = 0, 1, \dots, 4$  and each  $A$  in  $\Sigma^1$ :*

$$\frac{}{\mathbf{S5}_{Qmo}^N + \beta i} A \quad \text{iff} \quad \frac{}{\mathbf{S5}_{Qmo}^N + \beta i} A$$

**Proof.** See Åqvist [1987, Section 19.0]. ■

## 20 WEAK REPRESENTATION OF $O_{dy}\mathbf{S5}^N + \alpha i$ IN $\mathbf{S5}_{Qmo}^N + \beta i$ ; IS FULL REPRESENTABILITY LOST?

Bearing in mind that  $\phi$  is now a translation from  $\Sigma_{0,N}^2$  into  $\Sigma^1$  (with fresh clauses for  $N$  and  $M$ , see Section 17.1 above), we state the following result:

**THEOREM 64** (Weak translation for  $O_{dy}\mathbf{S5}^N + \alpha i$  and  $\mathbf{S5}_{Qmo}^N + \beta i$ ). *For each  $i = 0, 1, \dots, 4$  and each  $A$  in  $\Sigma_{0,N}^2$ :*

$$\frac{}{O_{dy}\mathbf{S5}^N + \alpha i} A \quad \text{only if} \quad \frac{}{\mathbf{S5}_{Qmo}^N + \beta i} \phi A.$$

In other words, the set of  $\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i$  theses  $\subseteq$  the dyadic deontic fragment under  $\phi$  of  $\mathbf{S5}_{Qmo}^N + \beta i$ . Note that in this theorem '=' has been weakened to ' $\subseteq$ ' and 'iff' to 'only if'.

**Proof.** Our task is to give an ordinary 'only if' part demonstration. The new thing here is to verify, for each  $i = 0, 1, \dots, 4$ , that if  $A$  is an axiom under the new schema  $\alpha i$ , then its translation  $\phi A$  is provable in  $\mathbf{S5}_{Qmo}^N + \beta i$ . And we easily accomplish this, appealing precisely to the schema  $\beta i$ . ■

As to the converse result, i.e. the 'if' part, I have not been able to establish it; nor do I know whether it holds good or not. But I am inclined to believe that it does; *if* it does, however, its proof will be harder than any one so far met with in this essay — at least, so I believe.

EXERCISE 65. Try to prove the converse of the weak translation theorem stated above! Explain why you got lost, or else: congratulations and many thanks!

We like to add that, even if the full representability of  $\mathbf{O}_{dy}\mathbf{S5}^N + \alpha i$  in  $\mathbf{S5}_{Qmo}^N + \beta i$  should turn out not to hold, this does not entail that the enterprise of developing dyadic deontic logic and alethic modal logic with monadic  $Q$  in a parallel fashion is without considerable heuristic value. In fact, I think, the contrary will prove to be the case.

## 21 AN ATTEMPTED RECONSTRUCTION AND IDENTIFICATION OF THE HANSSON DYADIC SYSTEMS DSDL1, DSDL2, AND DSDL3: ALETHIC PRELIMINARIES

The main idea proposed in Hansson [1969] is that the concept of *validity* in Von Wright-type deontic logic (Hansson is anxious to point out that he just deals with this type in his paper) can be semantically explained in terms of a *preference relation* 'is at least as ideal as' *among possible worlds*; this claim is to apply whether the Von Wright-type deontic logic be monadic or dyadic. Hansson himself thinks of possible worlds as Boolean valuations in the sense familiar from the teaching of elementary propositional calculus. We shall not follow him in making this identification, however, because our Kripkean semantical technique has already supplied us with an independent notion of (a set of) possible worlds. Again, given a preference relation (ordering, ranking)  $R$  on a set of possible worlds  $W$ , we are automatically equipped with the notion of the *R-maximal* ('best', 'optimal', under  $R$ ) elements of  $W$  and of various (perhaps all) subsets of  $W$ . We could then give the following informal characterization of the function *opt* in our models  $\mathcal{U}$  for alethic systems with monadic  $\mathbf{Q}(\mathbf{S4}_{Qmo}$  etc.):

$\text{opt}(\top) =$  the  $R$ -maximal elements of  $\| \top \|^{\mathcal{U}}$ , i.e. of  $W$  as a whole.

Remember here that the *set*  $\text{opt} \subseteq W$  in models for the alethic systems with the *propositional constant*  $Q$  is simply to be equated with this set  $\text{opt}(\top)$ . Moreover, in general, we should have for any sentence  $A$  in  $\Sigma^1$ :

$\text{opt}(A) =$  the  $R$ -maximal elements of  $\| A \|^{\mathcal{U}}$ , i.e. of the set of worlds in  $W$  where  $A$  is true (= the extension in  $\mathcal{U}$  of the wff  $A$ ).

In the light of these heuristic preliminaries we now supplement our alethic  $\mathbf{S5}_{Qmo}^N$ -models with a Hanssonian preference relation on  $W$ , consider some possible conditions on it, and see what happens when one interprets the sentences in  $\Sigma^1$  relatively to these new enriched structures. Later on, we are going to perform a similar operation on our dyadic deontic models and the sentence-set  $\Sigma_{0,N}^2$ .

**DEFINITION 66** (Various sorts of  $\succsim$ -supplemented alethic models). Let  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}^{\S}, \text{opt}, \succsim, V \rangle$  be any structure where

- (i)  $W \neq \emptyset$  (as usual),
- (ii)  $R^{\S} = W \times W$  (as has now become usual),
- (iii)  $\text{opt}: \Sigma^1 \rightarrow \mathcal{P}W$  (as usual),
- (iv)  $\succsim \subseteq W \times W$  (novelty),
- (v)  $V: \text{Prop} \times W \rightarrow \{1, 0\}$  (as usual).

If you like, call any ordered quintuple of this sort a *minimal alethic H-model*. Note that we place no further conditions on  $\text{opt}$  or on  $\succsim$  in minimal alethic  $H$ -models. Furthermore, we say:

- (b)  $\mathcal{U}$  is an *alethic H-model* iff  $\text{opt}$  and  $\succsim$  *jointly* satisfy the following condition (for each  $A$  in  $\Sigma^1$ ):
  - $\delta 0.$   $\text{opt}(A) = \{x \in \| A \|^{\mathcal{U}} : (\forall y \in \| A \|^{\mathcal{U}}) x \succsim y\}$ .
- (c)  $\mathcal{U}$  is an *alethic H<sub>1</sub>-model* iff  $\succsim$ , in addition to meeting  $\delta 0$  jointly with  $\text{opt}$ , satisfies this condition  $\delta 1$ :
  - $\delta 1.$   $\succsim$  is *reflexive* in  $W$  (i.e. for all  $x$  in  $W$ ,  $x \succsim x$ ).
- (d)  $\mathcal{U}$  is an *alethic H<sub>2</sub>-model* iff  $\succsim$ , in addition to meeting  $\delta 0$  and  $\delta 1$ , satisfies this condition  $\delta 2$  (for each  $A$  in  $\Sigma^1$ ):
  - $\delta 2.$  If  $\| A \|^{\mathcal{U}} \neq \emptyset$ , then  $\{x \in \| A \|^{\mathcal{U}} : (\forall y \in \| A \|^{\mathcal{U}}) x \succsim y\} \neq \emptyset$ .

(e)  $\mathcal{U}$  is an *alethic  $H_3$ -model* iff  $\succsim$ , in addition to meeting  $\delta 0$ ,  $\delta 1$  and  $\delta 2$ , satisfies this condition  $\delta 3$ :

$\delta 3.$   $\succsim$  is *transitive* in  $W$ .

(f)  $\mathcal{U}$  is an *alethic strong  $H_3$ -model* iff  $\succsim$ , in addition to meeting  $\delta 0$ ,  $\delta 1$ ,  $\delta 2$  and  $\delta 3$ , satisfies this condition  $\delta 4$ :

$\delta 4.$   $\succsim$  is *strongly connected* (total, complete) in  $W$ ;

i.e. for all  $x, y$  in  $W$  :  $x \succsim y$  or  $y \succsim x$  (or both).

LEMMA 67 (On  $\succsim$ -supplemented alethic models). (b) *In the inductive definition of truth at  $x$  in  $\mathcal{U}$ , where  $\mathcal{U}$  is an alethic  $H$ -models, the clause for  $Q$  is equivalent to the following:*

$$\frac{\mathcal{U}}{x}QA \quad \text{iff} \quad \frac{\mathcal{U}}{x}A \ \& \ \forall y(\frac{\mathcal{U}}{y}A \supset x \succsim y)$$

(b') *Let  $\mathcal{U}$  be any alethic  $H$ -model so that  $\text{opt}$  and  $\succsim$  jointly satisfy  $\delta 0$ . Then  $\text{opt}$  satisfies the three conditions  $\sigma 0$ ,  $\sigma 1$  and  $\sigma 2$ , given in Section 19 above.*

(c) *Let  $\mathcal{U}$  be any alethic  $H_1$ -model. Then  $\text{opt}$  satisfies the three conditions just mentioned.*

(d) *Let  $\mathcal{U}$  be any alethic  $H_2$ -model. Then  $\text{opt}$  satisfies these conditions as well as  $\sigma 3$ .*

(e-f) *Let  $\mathcal{U}$  be any alethic  $H_3$ -model or alethic strong  $H_3$ -model. Then  $\text{opt}$  satisfies all the five conditions  $\sigma 0$ – $\sigma 4$ .*

**Proof.** See Åqvist [1987, Section 21]. ■

THEOREM 68 (Soundness). *The present result concerns axiomatic extensions of the alethic system  $\mathbf{S5}_{Qmo}^N$ ; we are thus dealing with subsets of  $\Sigma^1$ . The result is presented in the right-most column of Table 3; in the head of that column, the locution ‘matching kind of validity’ means, of course, the same as ‘truth at all points in every  $\succsim$ -supplemented alethic model of the corresponding kind’. And the corresponding kinds of models are then to be found in the leftmost column, and the appropriate restrictions on  $\succsim$  in the middle one.*

**Proof.** See Åqvist [1987, Section 21]. ■

REMARK 69.

(i) Questions of completeness have been left open.

Table 3.

Kind of alethic model supplemented with $\succcurlyeq$	Restriction(s) on $\succcurlyeq$ in such a model	Axiomatic extension of $\mathbf{S5}_{Q_{mo}}^N$ , sound with respect to the matching kind of validity
minimal $H$ -	none	$\mathbf{S5}_{Q_{mo}}^N$ itself
$H$ -	$\delta 0$	$\mathbf{S5}_{Q_{mo}}^N +$ every $\beta i$ with $i = 0, 1, 2$
$H_1$ -	$\delta 0$ and $\delta 1$	ditto
$H_2$ -	$\delta 0, \delta 1$ and $\delta 2$	$\mathbf{S5}_{Q_{mo}}^N +$ every $\beta i$ with $i = 0, 1, 2, 3$
$H_3$ -	$\delta 0, \delta 1, \delta 2$ and $\delta 3$	$\mathbf{S5}_{Q_{mo}}^N +$ every $\beta i$ with $i = 0, 1, 2, 3, 4$
strong $H_3$ -	$\delta 0, \delta 1, \delta 2, \delta 3$ and $\delta 4$	ditto

- (ii) Certain possible kinds of  $\succcurlyeq$ -supplemented alethic models have been left out of consideration, e.g. alethic  $H$ -models where  $\succcurlyeq$  satisfies  $\delta 1$  and  $\delta 3$  (but not necessarily  $\delta 2$ ) as well as such where  $\succcurlyeq$  satisfies  $\delta 3$  and  $\delta 4$  (and hence  $\delta 1$ ; but not necessarily  $\delta 2$ ).

## 22 ATTEMPTED IDENTIFICATION OF HANSSON DYADIC SYSTEMS: $\succcurlyeq$ -SUPPLEMENTED DEONTIC MODELS

In this section we supplement our dyadic deontic  $\mathbf{O}_{dy}\mathbf{S5}^N$ -models with a Hanssonian preference relation  $\succcurlyeq$ , consider possible conditions on it, and see what happens when we interpret the sentences in the beautiful set  $\Sigma_{0,N}^2$  relatively to these enriched structures. The exposition will parallel the one given in the last section.

**DEFINITION 70** (Various sorts of  $\succcurlyeq$ -supplemented dyadic deontic models). Let  $\mathcal{U} = \langle \mathcal{W}, \mathcal{R}, \succcurlyeq, \mathcal{V} \rangle$  be any structure

- (i)  $W \neq \emptyset$  (as usual)



- (ii)  $R : \Sigma_{0,N}^2 \rightarrow \mathcal{P}(W \times W)$  (as has now become usual)
- (iii)  $\succcurlyeq \subseteq W \times W$  (novelty)
- (iv)  $V : \text{Prop} \times W \rightarrow \{1, 0\}$  (as usual)

Then we say:

- (a)  $\mathcal{U}$  is a *minimal deontic H-model* iff  $R$  meets condition (6) (see Section 17 above).
- (b)  $\mathcal{U}$  is a *deontic H-model* iff  $R$  and  $\succcurlyeq$  jointly satisfy the following condition  $\gamma 0$  (for each  $A$  in  $\Sigma_{0,N}^2$  and any  $x, y$  in  $W$ ):
 
$$\gamma 0. \quad xR_A y \quad \text{iff} \quad \frac{\mathcal{U}}{y} A \ \& \ \forall z (\frac{\mathcal{U}}{z} A \supset y \succcurlyeq z).$$
- (c)  $\mathcal{U}$  is a *deontic H<sub>1</sub>-model* iff  $\succcurlyeq$  meets the reflexivity condition  $\delta 1$  as well as  $\gamma 0$ .
- (d)  $\mathcal{U}$  is a *deontic H<sub>2</sub>-model* iff  $\succcurlyeq$  meets the ‘Limit Condition’  $\delta 2$  as well as  $\delta 1$  and  $\gamma 0$ ; where  $\delta 2$  now applies to any sentence in  $\Sigma_{0,N}^2$ .
- (e)  $\mathcal{U}$  is a *deontic H<sub>3</sub>-model* iff  $\succcurlyeq$  satisfies the transitivity condition  $\delta 3$  as well as  $\delta 2$ ,  $\delta 1$  and  $\gamma 0$ .
- (f)  $\mathcal{U}$  is a *deontic strong H<sub>3</sub>-model* iff  $\succcurlyeq$  satisfies the connectedness requirement  $\delta 4$  in addition to meeting  $\delta 3$ ,  $\delta 2$ ,  $\delta 1$  and  $\gamma 0$ .

LEMMA 71 ( $\succcurlyeq$ -supplemented dyadic deontic models).

- (b) Let  $\mathcal{U}$  be any deontic H-model. Then, by  $\gamma 0$ , the inductive clauses for  $O$  and  $P$  in the definition of truth at  $x$  in  $\mathcal{U}$  become equivalent to the following:

$$\begin{aligned} \frac{\mathcal{U}}{x} O_B A & \quad \text{iff} \quad \forall y ((\frac{\mathcal{U}}{y} B \ \& \ \forall z (\frac{\mathcal{U}}{z} B \supset y \succcurlyeq z)) \supset \frac{\mathcal{U}}{y} A) \\ \frac{\mathcal{U}}{x} P_B A & \quad \text{iff} \quad \exists y ((\frac{\mathcal{U}}{y} B \ \& \ \forall z (\frac{\mathcal{U}}{z} B \supset y \succcurlyeq z)) \ \& \ \frac{\mathcal{U}}{y} A). \end{aligned}$$

- (b') Let  $\mathcal{U}$  be any deontic H-model. Then  $R$  satisfies condition (6) (stated in Section 17 above) as well as the three conditions  $\rho 0, \rho 1$  and  $\rho 2$  (stated in Section 18 above).
- (c) Let  $\mathcal{U}$  be any deontic H<sub>1</sub>-model. Then  $R$  satisfies those four conditions just mentioned.

- (d) Let  $\mathcal{U}$  be any deontic  $H_2$ -model. Then  $R$  satisfies those four conditions as well as  $\rho_3$ .
- (e-f) Let  $\mathcal{U}$  be any deontic  $H_3$ -model or a deontic strong  $H_3$ -model. Then  $R$  satisfies (6) as well as all five conditions  $\rho_0$ – $\rho_4$ .

**Proof.** Straightforward and left as an exercise. ■

**THEOREM 72 (Soundness).** *The present result concerns axiomatic extensions of the dyadic deontic system  $\mathbf{O}_{dy}\mathbf{S5}^N$ ; we are thus dealing with subsets of  $\Sigma_{0,N}^2$ . We state the result in the form of a table (Table 4) analogous to the one at the end of the last theorem (Theorem 68).*

Table 4.

Kind of dyadic deontic model supplemented with $\succcurlyeq$	Restriction(s) on $\succcurlyeq$ and/or $R$ in such a model	Axiomatic extension of $\mathbf{O}_{dy}\mathbf{S5}^N$ , sound with respect to the matching kind of validity
minimal $H$ -	(6)	$\mathbf{O}_{dy}\mathbf{S5}^N$ itself
$H$ -	$\gamma_0$	$\mathbf{O}_{dy}\mathbf{S5}^N$ + every $\alpha_i$ with $i = 0, 1, 2$
$H_1$ -	$\gamma_0$ and $\delta_1$	$\mathbf{O}_{dy}\mathbf{S5}^N$ + every $\alpha_i$ with $i = 0, 1, 2$
$H_2$ -	$\gamma_0, \delta_1$ and $\delta_2$	$\mathbf{O}_{dy}\mathbf{S5}^N$ + every $\alpha_i$ with $i = 0, 1, 2, 3$
$H_3$ -	$\gamma_0, \delta_1, \delta_2$ and $\delta_3$	$\mathbf{O}_{dy}\mathbf{S5}^N$ + every $\alpha_i$ with $i = 0, 1, 2, 3, 4$
strong $H_3$ -	$\gamma_0, \delta_1, \delta_2, \delta_3$ and $\delta_4$	$\mathbf{O}_{dy}\mathbf{S5}^N$ + every $\alpha_i$ with $i = 0, 1, 2, 3, 4$

**Proof.** Use the soundness result for  $\mathbf{O}_{dy}\mathbf{S5}^N$  + every  $\alpha_i$ , for relevant choices of  $i$ , together with the Lemma on  $\succcurlyeq$ -supplemented dyadic deontic models. ■

REMARK 73. Our new  $\succ$ -supplemented deontic models obviously engender new ‘matching kinds of validity’, such as deontic  $H$ -validity, deontic  $H_1$ -validity etc. Are any of the axiomatic extensions of  $\mathbf{O}_{dy}\mathbf{S5}^N$ , which appear in the table above, *complete* with respect to the kind of validity associated with them by the table? If so, which?

We shall try to answer some of these completeness questions elsewhere. Meanwhile, we make a first attempt at an identification of the Hansson systems **DSDL1**, **DSDL2** and **DSDL3** and start with a few preliminary observations.

- (I) The *language* of these three systems is poorer than that of  $\mathbf{O}_{dy}\mathbf{S5}^N$  in the following respects: (a) it lacks the operators  $N$  and  $M$  of universal necessity and possibility; (b) its set of well formed *sentences* differs from  $\Sigma_{0,N}^2$  in not allowing iterations and overlapping of dyadic deontic operators, nor any mixed formulas, e.g. of the type  $O_B A \rightarrow A$ . Thus, in the language used by Hansson, there are, for one thing, no instances of any of our axiom schemata a3–a7. On the other hand, the set of sentences of Hansson’s language should clearly be a *subset* of  $\Sigma_{0,N}^2$  (inasmuch as we are at all able to discuss his systems in our present framework). We attempt the following characterization of it.

Let  $\Pi$  be the smallest set which contains as its members all proposition letters in our set Prop as well as  $\top$  and  $\perp$ , and which is closed under  $\neg, \wedge, \vee, \rightarrow$  and  $\Leftrightarrow$ . Thus,  $\Pi$  is simply the set of wffs of the familiar propositional calculus with *verum* and *falsum*. We then define  $\Theta$  (= the set of sentences of our reconstructed Hansson language) to be the smallest set  $S$  such that:

- (a) If  $A, B$  are in  $\Pi$ , then  $O_B A$  and  $P_B A$  are in  $S$ .
- (b) If  $C, D$  are in  $S$ , then so are  $\neg C, (C \wedge D), (C \vee D), (C \rightarrow D)$  and  $(C \Leftrightarrow D)$ .

Clearly, then, we have that  $\Theta \subseteq \Sigma_0^2 \subseteq \Sigma_{0,N}^2$ .

- (II) As for the *semantics* of the Hansson dyadic system, we pointed out at the beginning of the previous section that we were going to use Kripkean possible worlds rather than Boolean valuations in the basic set on which  $\succ$  is defined; this approach naturally leads to our present sort of deontic models. Now, as characteristic conditions on  $\succ$ , Hansson [1969, p. 395 f.] adopts the following:

For the system <b>DSDL1</b> :	reflexivity ( $\delta 1$ ).
For the system <b>DSDL2</b> :	limitedness ( $\delta 2$ ; in addition to $\delta 1$ ).
For the system <b>DSDL3</b> :	transitivity ( $\delta 3$ ) and strong connectedness ( $\delta 4$ ) (in addition to $\delta 1$ and $\delta 2$ ).

As for the notion of *truth* as applied to sentences in  $\Theta$  of the forms  $O_B A$  and  $P_B A$ , Hansson proposes analogues of the two conditions stated under (b) in the Lemma on  $\succsim$ -supplemented dyadic deontic models (we disregard here the fact that Hansson's own concept of  $\succsim$ -maximality does not completely coincide with ours).

### 22.1 Semantic identification of DSDL1, DSDL2 and DSDL3

Let  $A$  be in  $\Sigma_{0,N}^2$ , and let  $\overline{\overline{H_1(H_2, \text{strong } H_3)}} A$  mean that  $A$  is true at every world in every deontic  $H_1$ - ( $H_2$ -, strong  $H_3$ -)model. We then suggest the following characterization of the Hansson systems as subsets of  $\Sigma_{0,N}^2$ :

- (i) **DSDL1** =  $\{A \in \Theta : \overline{\overline{H_1}} A\}$
- (ii) **DSDL2** =  $\{A \in \Theta : \overline{\overline{H_2}} A\}$
- (iii) **DSDL3** =  $\{A \in \Theta : \overline{\overline{\text{strong } H_3}} A\}$

So, for  $i = 1, 2$ , **DSDL $i$**  is the set of deontically  $H_i$ -valid sentences in  $\Theta$ , and **DSDL3** is the set of deontically *strong*  $H_3$ -valid sentences in  $\Theta$ .

What about axiomatic characterizations of the Hansson systems as deductive calculi? If such were available, our identification of them would be more satisfactory from the proof theoretical point of view. As things stand right now, however, we only have the following somewhat insufficient result:

### 22.2 Partial syntactic identification of DSDL1, DSDL2 and DSDL3

Let us introduce the following abbreviations:

- E** = the set of theses of  $\mathbf{O}_{dy}\mathbf{S5}^N$  + every  $\alpha i$  with  $i = 0, 1, 2$
- F** = the set of theses of  $\mathbf{O}_{dy}\mathbf{S5}^N$  + every  $\alpha i$  with  $i = 0, 1, 2, 3$
- G** = the set of theses of  $\mathbf{O}_{dy}\mathbf{S5}^N$  + every  $\alpha i$  with  $i = 0, 1, 2, 3, 4$

Furthermore, let  $\mathbf{E}/\Theta$  ( $\mathbf{F}/\Theta, \mathbf{G}/\Theta$ ) =  $\mathbf{E}(\mathbf{F}, \mathbf{G})$  restricted to the set  $\Theta$ ; i.e.  $\mathbf{E} \cap \Theta$  etc. Then we have three groups of more or less obvious results:

I. Soundness results	II. Unsoundness results	III. Incompleteness results
(i) $\mathbf{E}/\Theta \subseteq \mathbf{DSDL1}$	(iv) $\mathbf{F}/\Theta \not\subseteq \mathbf{DSDL1}$	(vii) $\mathbf{DSDL2} \not\subseteq \mathbf{E}/\Theta$
(ii) $\mathbf{F}/\Theta \subseteq \mathbf{DSDL2}$	(v) $\mathbf{G}/\Theta \not\subseteq \mathbf{DSDL1}$	(viii) $\mathbf{DSDL3} \not\subseteq \mathbf{E}/\Theta$
(iii) $\mathbf{G}/\Theta \subseteq \mathbf{DSDL3}$	(vi) $\mathbf{G}/\Theta \not\subseteq \mathbf{DSDL2}$	(ix) $\mathbf{DSDL3} \not\subseteq \mathbf{F}/\Theta$

**Proof.** As for Group I, appeal to our last soundness theorem; as for Groups II and III, appeal in addition to deontic analogues of the Exercises on alethic  $H$ -,  $H_1$ -, and  $H_2$ -models studied in Åqvist [1987, Section 21]. ■

What makes the given syntactic identification only a partial one is, of course, the fact that as yet we do not know whether the *converses* of (i)–(iii) hold, so that we could strengthen ‘ $\subseteq$ ’ to ‘ $=$ ’. Also, if it should turn out that  $\mathbf{G}/\Theta$  is not complete with respect to strong deontic  $H_3$ -validity among the members of  $\Theta$ , perhaps  $\mathbf{G}/\Theta$  is complete with respect to deontic  $H_3$ -validity in  $\Theta$  *simpliciter*?

### 23 ON THE COMPLETENESS PROBLEM FOR THE DYADIC DEONTIC LOGIC $\mathbf{G}$

As in the last section we let  $\mathbf{G}$  be the (set of theses of the) system  $\mathbf{Q}_{dy}\mathbf{S5}^N$  + every  $\sigma i$  with  $i 0, 1, \dots, 4$ . Thus  $\mathbf{G} \subseteq \Sigma_{0,N}^2$ , its rules of proof are *modus ponens* (R1) and  $N$ -necessitation (R2''), its axiom schemata are a0–a2, a6–a8 (Section 17 above) as well as  $\sigma 0$ – $\sigma 4$  (Section 18).

The system  $\mathbf{G}$ , as just characterized in purely axiomatic terms, is by far the most important dyadic deontic logic dealt with in the present chapter—recall that in Section 4 *supra* we defined a *strongly normal dyadic* deontic logic as one containing precisely the system  $\mathbf{G}$ . Obviously, then, the problem of finding an adequate *semantical* characterization relative to which  $\mathbf{G}$  can be proved *complete* is an equally important one. In the original version of this chapter I conjectured that  $\mathbf{G}$  was complete with respect to strong  $H_3$ -validity in the sense of the last section; for a ‘possible’, rather complicated proof of this conjectured result I referred the reader to Åqvist [1987, Section 23.3].

I am now convinced that the completeness problem for the system  $\mathbf{G}$  has a much simpler solution than was suggested by my earlier conjectured proof. In the remainder of this section I shall then outline that simpler solution as first presented in two recent papers by Åqvist [1996; 1997]. Again, in the three final sections (Sections 24–26) of this chapter I shall present some further recent results on the crucial dyadic system  $\mathbf{G}$ .

### 23.1 Improved semantics for $G$

DEFINITION 74 ( $\mathbf{G}$ -structures and truth at a point in a  $\mathbf{G}$ -structure). By a  $\mathbf{G}$ -structure we mean any ordered triple

$$\mathcal{U} = \langle W, V, best \rangle$$

where

- (i)  $W \neq \emptyset$
- (ii)  $V : \text{Prop} \times W \rightarrow \{1, 0\}$
- (iii)  $best : \Sigma_{0,N}^2 \rightarrow \mathcal{P}W$ .

We can now tell what it means for any sentence  $A$  in  $\Sigma_{0,N}^2$  to be *true at* a point (“world”)  $x \in W$  in a  $\mathbf{G}$ -structure  $\mathcal{U}$  [in symbols:  $\frac{\mathcal{U}}{x} A$ ], starting out with obvious clauses like

$$\begin{aligned} \frac{\mathcal{U}}{x} p &\text{ iff } V(p, x) = 1 \text{ (for any } p \text{ in Prop)} \\ \frac{\mathcal{U}}{x} \top & \\ \text{not } \frac{\mathcal{U}}{x} \perp & \end{aligned}$$

and so on for molecular sentences having Boolean connectives as their main operator. We then handle sentences having modal and dyadic deontic operators as their main operator as follows:

$$\begin{aligned} \frac{\mathcal{U}}{x} NA &\text{ iff for each } y \text{ in } W : \frac{\mathcal{U}}{y} A \\ \frac{\mathcal{U}}{x} MA &\text{ iff for some } y \text{ in } W : \frac{\mathcal{U}}{y} A \\ \frac{\mathcal{U}}{x} O_B A &\text{ iff for each } y \text{ in } best(B) : \frac{\mathcal{U}}{y} A \\ \frac{\mathcal{U}}{x} P_B A &\text{ iff for some } y \text{ in } best(B) : \frac{\mathcal{U}}{y} A \end{aligned}$$

DEFINITION 75 ( $\mathbf{G}$ -models,  $\mathbf{G}$ -validity and  $\mathbf{G}$ -satisfiability). We now focus our attention on a special kind of  $\mathbf{G}$ -structures called “ $\mathbf{G}$ -models”. By a  $\mathbf{G}$ -model we mean any  $\mathbf{G}$ -structure satisfying the following five conditions

on the function  $best$ , where, as usual, we let  $\|A\|^{\mathcal{U}}$ , or  $\|A\|$  for short, be the extension in  $\mathcal{U}$  of  $A$ :

- $\sigma 0$   $\|A\| = \|B\|$  only if  $best(A) = best(B)$
- $\sigma 1$   $best(A) \subseteq \|A\|$
- $\sigma 2$   $best(A) \cap \|B\| \subseteq best(A \wedge B)$
- $\sigma 3$   $\|A\| \neq \emptyset$  only if  $best(A) \neq \emptyset$
- $\sigma 4$   $best(A) \cap \|B\| \neq \emptyset$  only if  $best(A \wedge B) \subseteq best(A) \cap \|B\|$

for any sentences  $A, B$  in  $\Sigma_{0,N}^2$ . The notions of  $\mathbf{G}$ -validity and  $\mathbf{G}$ -satisfiability are then defined in the usual way.

**REMARK 76.** Conditions  $\sigma 0$ – $\sigma 4$  are very much like the so-named conditions on the function  $opt$  in Section 19 above. But note that  $opt$  was defined for all sentences in the alethic language of  $\mathbf{S5}_{Qmo}^N$ , whereas  $best$  is defined for those in the dyadic deontic language of  $\mathbf{O}_{dy}\mathbf{S5}^N$  and  $\mathbf{G}$ .

**THEOREM 77** (Soundness and Completeness of the system  $\mathbf{G}$ ).

Weak version: for each  $A$  in  $\Sigma_{0,N}^2$ :  $A$  is  $\mathbf{G}$ -provable iff  $A$  is  $\mathbf{G}$ -valid.

Strong version: for each  $S \subseteq \Sigma_{0,N}^2$ :  $S$  is  $\mathbf{G}$ -consistent iff  $S$  is  $\mathbf{G}$ -satisfiable.

**Proof.** (sketchy). The soundness parts are unproblematic. So we concentrate on the ‘only if’ half of the strong version, from which the ‘if’ half of the weak one is immediate.

Let  $S$  be any  $\mathbf{G}$ -consistent set of sentences, and let  $S^+$  be a maximal  $\mathbf{G}$ -consistent extension of  $S$ , the existence of which is guaranteed by Lindenbaum’s Lemma. Form the *canonical  $\mathbf{G}$ -structure generated by  $S^+$*  in the sense of the structure

$$\mathcal{U}^{S^+} = \langle W, V, best \rangle$$

where

- (i)  $W$  = the set of maximal  $\mathbf{G}$ -consistent sets  $x$  of sentences such that for all  $A$  in  $\Sigma_{0,N}^2$ : if  $NA \in S^+$ , then  $A \in x$ .
- (ii)  $V$  = the assignment defined as follows: for each  $p$  in  $Prop$  and each  $x$  in  $W$ ,  $V(p, x) = 1$  iff  $p \in x$ .
- (iii)  $best$  = the function from  $\Sigma_{0,N}^2$  into  $\mathcal{P}(W)$  defined by setting, for all sentences  $B$ ,  $best(B) = \{x \in W : \text{for all sentences } A, \text{ if } O_B A \in S^+, \text{ then } A \in x\}$ .

Omitting details we can then prove that, as just defined, the generated canonical  $\mathbf{G}$ -structure  $\mathcal{U}^{S^+}$  satisfies all lemmata needed for our desired completeness result. The most interesting one among them is the Verification Lemma, to the effect that  $\mathcal{U}^{S^+}$  is a  $\mathbf{G}$ -model satisfying the five conditions  $\sigma_0$ – $\sigma_4$  on the function *best*. A proof to precisely that effect is easily obtained as follows: define our three-place relation  $R$  by the requirement

$$(*) \quad xR_B y \text{ iff } y \in \text{best}(B) \text{ (for all } x, y \text{ in } W \text{ and all } B \text{ in } \Sigma_{0,N}^2)$$

On the basis of this definition, we quickly verify that our present truth conditions for  $O_B A$  and  $P_B A$  are equivalent to those given in Section 16 *supra*, that  $R$  meets condition (6) laid down in Section 17 *supra*, and that for each  $i = 0, 1, \dots, 4$ , the condition  $\sigma_i$  on *best* is equivalent to the restriction  $\rho_i$  on  $R$  (Section 18). Again, this result enables us to transform the proof given in Åqvist [1987, Section 18.1, pp. 161–165] into a proof that the canonical  $\mathbf{G}$ -structure  $\mathcal{U}^{S^+}$  is indeed a  $\mathbf{G}$ -model, as desired. ■

#### 24 AN INFINITE HIERARCHY OF EXTENSIONS OF $\mathbf{G}$ : THE DYADIC DEONTIC LOGICS $\mathbf{G}m[m = 1, 2, \dots]$

In this and the following sections of the present chapter we intend to shed some more light on the system  $\mathbf{G}$  by studying an infinite hierarchy  $\mathbf{G}m[m = 1, 2, \dots]$  of extensions of that system. The soundness and completeness of every system in that hierarchy is then asserted in a Theorem, for the proof of which we refer the reader to Åqvist [1996]. The main semantical technical device employed in our study of these extensions of  $\mathbf{G}$  is this: in the models of each system  $\mathbf{G}m$  in the hierarchy, we work with a set

$$\{\text{opt}_1, \text{opt}_2, \dots, \text{opt}_m\}$$

which is to be a *partition* of the set  $W$  of ‘possible worlds’ into exactly  $m$  non-empty, pairwise disjoint and together exhaustive ‘optimality’ classes, viewed as so many *levels of perfection*. Intuitively, we think of  $\text{opt}_1$  as the set of ‘best’ [optimal] members of  $W$  as a whole,  $\text{opt}_2$  as the set of best members of  $W - \text{opt}_1$  [the ‘second best’ members of  $W$ ],  $\text{opt}_3$  as the set of best members of  $W - (\text{opt}_1 \cup \text{opt}_2)$  [the ‘third best’ members of  $W$ ]; and so on. Now, we shall represent each level of perfection in the object-language of the systems by a so-called *systematic frame constant*. The truth conditions and axioms governing those constants can then be seen to play a highly important, characteristic role in our axiomatization.

Thus, the primitive logical *vocabulary* of the systems  $\mathbf{G}m$  ( $m$  any positive integer) results from that of  $\mathbf{G}$  by adding to the latter an infinite family

$$\{Q_i\} i = 1, 2, \dots$$



of *systematic frame constants*, indexed by the set of positive integers. As just explained, the  $Q_i$  are to represent different ‘levels of perfection’ in the models of the system  $\mathbf{G}m$ . The set *Sent* of their well-formed sentences (formulas, wffs) is then defined in the straightforward way—we think of the  $Q_i$  as zero-place connectives on a par with  $\top$  and  $\perp$ .

We begin the presentation of the new dyadic deontic logics  $\mathbf{G}m$  by outlining their *proof theory*. The rules of proof *modus ponens* and *N-necessitation* are common to  $\mathbf{G}$  and the  $\mathbf{G}m[m = 1, 2, \dots]$ . In addition to the *axiom schemata* of  $\mathbf{G}$ , each system  $\mathbf{G}m$  has the following:

- $\alpha 5.$   $Q_i \rightarrow \neg Q_j$ , for all positive integers  $i, j$  with  $1 < i \neq j < \omega$
- $\alpha 6.$   $P_B Q_i \rightarrow ((Q_1 \vee \dots \vee Q_{i-1}) \rightarrow \neg B)$ , for all  $i$  with  $1 < i \leq m$
- $\alpha 7.$   $Q_1 \rightarrow (O_B A \rightarrow (B \rightarrow A))$
- $\alpha 8.$   $(Q_1 \wedge O_B A \wedge B \wedge \neg A) \rightarrow P_B (Q_1 \vee \dots \vee Q_{i-1})$ ,  
for all  $i$  with  $1 < i \leq m$
- $\alpha 9.$   $Q_1 \vee \dots \vee Q_m$
- $\alpha 10.$   $MQ_1 \wedge \dots \wedge MQ_m$ .

Then, the *axiomatic system*  $\mathbf{G}m[m = 1, 2, \dots]$  is determined by the axiom schemata  $\alpha 0$ – $\alpha 2$ ,  $\alpha 6$ – $\alpha 8$  and  $\alpha 9$ – $\alpha 10$  (and the usual rules of proof). Moreover, we define the notions of *provability*, *derivability*, [in] *consistency* and *maximal consistency* for the systems  $\mathbf{G}m$  in the straightforward way.

Turning next to the *semantics* of the logics  $\mathbf{G}m$ , we define, for any positive integer  $m$ , a  $\mathbf{G}m$ -*structure* as an ordered quintuple

$$U = (W, V, \{\text{opt}_i\}_{i=1,2,\dots,m}, \text{best})$$

where

- (i)  $W \neq \emptyset$  [ $W$  is a non-empty set of ‘possible worlds’].
- (ii)  $V: \text{Prop} \rightarrow \text{pow}(W)$  [ $V$  is a valuation function which to each propositional variable assigns a subset of  $W$ ].
- (iii)  $\{\text{opt}_i\}_{i=1,2,\dots}$  is an infinite sequence of subsets of  $W$ .
- (iv)  $m$  is the positive integer under consideration.
- (v)  $\text{best}: \text{Sent} \rightarrow \text{pow}(W)$  [ $\text{best}$  is a function which to each sentence in the  $\mathbf{G}m$ -language assigns a subset of  $W$ , heuristically, the set of *best* worlds in the extension (truth-set) of the sentence under consideration].

The definition of a sentence being *true at a point in* a  $\mathbf{G}m$ -structure remains as in the case of  $\mathbf{G}$ , or  $\mathbf{G}$ -structures, except for the following fresh clause governing our frame constants:

$$\frac{U}{x} Q_i \text{ iff } x \in \text{opt}_i \quad (\text{for all positive integers } i).$$

We now focus our attention on a special kind of  $\mathbf{G}m$ -structures called ‘ $\mathbf{G}m$ -models’. By a  $\mathbf{G}m$ -model we shall mean any  $\mathbf{G}m$ -structure  $\mathcal{U}$  where  $\{\text{opt}_i\}$ ,  $m$  and best satisfy the following additional conditions:

*Exactly  $m$  Non-empty levels of Perfection*

This condition requires the set  $\{\text{opt}_1, \text{opt}_2, \dots, \text{opt}_m\}$  to be a *partition* of  $W$  in the sense that

- (a)  $\text{opt}_i \cap \text{opt}_j = \emptyset$ , for all positive integers  $i, j$  with  $1 \leq i \neq j \leq m$
- (b)  $\text{opt}_1 \cup \dots \cup \text{opt}_m = W$
- (c)  $\text{opt}_i \neq \emptyset$ , for each  $i$  with  $1 \leq i \leq m$ .
- (d)  $\text{opt}_i = \emptyset$ , for each  $i$  with  $m < i < \omega$ .

The second condition is one on our ‘choice’ function best; it is intended to capture the intuitive meaning of that function:

$$\gamma^0. \quad x \in \text{best}(B) \quad \text{iff} \quad \begin{array}{l} \frac{\mathcal{U}}{x} B \text{ and for each } y \text{ in } W : \text{ iff} \\ \frac{\mathcal{U}}{y} B, \text{ then } x \succcurlyeq y. \end{array}$$

Here, the weak preference relation  $\succcurlyeq$ , ‘is at least as good (ideal) as’ is to be understood as follows. First of all, by clauses (a) and (b) in the condition *Exactly  $m$  Non-empty levels of Perfection*, we have that for each  $x$  in  $W$  there is *exactly one* positive integer  $i$  with  $1 \leq i \leq m$  such that  $x \in \text{opt}_i$ . We then define a ‘ranking’ function  $r$  from  $W$  into the closed interval  $[1, m]$  of integers by setting

$$r(x) = \text{the } i, \text{ with } 1 < i < m, \text{ such that } x \in \text{opt}_i.$$

Finally, we define  $\succcurlyeq$  as the binary relation on  $W$  such that for all  $x, y$  in  $W$ :

$$x \succcurlyeq y \text{ iff } r(x) \leq r(y).$$

Armed with the notion of a  $\mathbf{G}m$ -model, we define, for  $m = 1, 2, \dots$ , those of  $\mathbf{G}m$ -validity and  $\mathbf{G}m$ -satisfiability in the obvious way.

**THEOREM 78.** (Soundness and completeness of the systems  $\mathbf{G}m[m = 1, 2, \dots]$ )

*Weak version:* For each  $A$  in *Sent*:  $A$  is  $\mathbf{G}m$ -provable iff  $A$  is  $\mathbf{G}m$ -valid.

*Strong version:* For each  $S \subseteq \text{Sent}$ :  $S$  is  $\mathbf{G}m$ -consistent iff  $S$  is  $\mathbf{G}m$ -satisfiable.

**Proof.** See Åqvist [1996]. Also, we observe that a proof that the generated canonical  $\mathbf{G}m$ -structure  $\mathcal{U}^{S^+}$  satisfies our condition  $\gamma^0$  on the function best is easily extracted from the proof given in Åqvist [1993] and [1987, Ch VI, Section 23.3.6, pp. 187–191] that the three-place relation  $R$  satisfies essentially the same condition (use the definition (\*) in the last section!). ■

25 ON THE RELATION OF THE ‘CORE’ SYSTEM  $\mathbf{G}$  TO THE LOGICS  $\mathbf{G}m[m = 1, 2, \dots]$

In the present section we deal with the system  $\mathbf{G}$  and state a result to the effect that  $\mathbf{G}$  is the *intersection* of all the logics  $\mathbf{G}m[m = 1, 2, \dots]$  (identifying as usual a ‘logic’ with the set of its theses, or sentences provable in it). Thus, the result answers the question how our crucial system  $\mathbf{G}$  is related to the  $\mathbf{G}m$ .

**THEOREM 79.** *For each  $\mathbf{G}$ -sentence  $A$  (i.e. member of  $\Sigma_{0,N}^2$ ):*

$$\vdash_{\mathbf{G}} A \text{ iff for each positive integer } m, \vdash_{\mathbf{G}m} A.$$

**Proof.** See Åqvist [1997, Section 3]. Let us just outline the main structure of our proof as given in that paper. The non-trivial direction here is the right-to-left one, the contraposed version of which asserts the following: if  $A$  is not  $\mathbf{G}$ -provable, then there exists a positive integer  $m$  such that  $A$  is not  $\mathbf{G}m$ -provable (either). We then argue as follows:

1.  $\not\vdash_{\mathbf{G}} A$  hypothesis
2.  $\not\vdash_{\mathbf{G}} \frac{\mathcal{U}}{\quad}$  from 1 by the weak completeness of  $\mathbf{G}$
3.  $\not\vdash_x \frac{\mathcal{U}}{\quad}$ , for some  $\mathbf{G}$ -model  $\mathcal{U} = \langle W, V, \text{best} \rangle$  and some  $x$  in  $W$  = from 2 by the definition of  $\mathbf{G}$ -validity.

Let  $\mathcal{U}^* = \langle W^*, V^*, \text{best}^* \rangle$  be the *filtration* of  $\mathcal{U}$  through the set of sub-sentences of  $A$  (in the sense rigorously defined in my 1997 paper mentioned above), and let  $[x]$  be the equivalence class of  $x$  under a certain equivalence relation on  $W$  (also defined in the paper). We then obtain:

4.  $\not\vdash_{[x]} \frac{\mathcal{U}^*}{\quad}$  from 3 by the Filtration Lemma for  $\mathbf{G}$  proved in Åqvist [1997].

We now observe that the filtration  $\mathcal{U}^*$  is necessarily a *finite*  $\mathbf{G}$ -model, so that there can be at most a *finite* number of levels of perfection compatible with and definable on  $\mathcal{U}^*$ . Again, this means that we can construct, for some positive integer  $m$ , a  $\mathbf{G}m$ -models

$$\mathcal{U}^{*+} = \langle W^*, V^*, \{\text{opt}_i\}_{i=1,2,\dots,m}, \text{best}^* \rangle$$

with the property that

5.  $\not\vdash_{[x]} \frac{\mathcal{U}^{*+}}{\quad}$  from 4 by the fact that the new items  $\text{opt}_i$  and  $m$  do not affect the truth-value of the  $\mathbf{G}$ -sentence  $A$

and then argue:

- |    |                                   |  |
|----|-----------------------------------|--|
| 6. | $\frac{}{\vdash_{\mathbf{G}m}}$   | from 5 by the definition of $\mathbf{G}m$ -validity  |
| 7. | $\frac{}{\vdash_{\mathbf{G}m} A}$ | from 6 by the soundness of each system $\mathbf{G}m$ |

where 7 is our desired conclusion. In Åqvist [1997, Section 3] we then finish the proof by providing a detailed careful justification of the two crucial steps 4 and 5 in the above overall argument. Among other things, our proof is seen to profit from the fact that the system  $\mathbf{G}$  is known to be *deductively equivalent*, under appropriate definitions, to a certain logic  $\mathbf{PR}$  of *preference*, as shown in Åqvist [1987, Appendix, §33]. ■

## 26 REPRESENTABILITY OF DYADIC DEONTIC LOGICS IN ALETHIC MODAL LOGICS WITH SYSTEMATIC FRAME CONSTANTS

In this final section of the present chapter we point out that the dyadic deontic logics  $\mathbf{G}m$  are *representable* in a hierarchy of alethic modal logics  $\mathbf{H}m[m = 1, 2, \dots]$ , which lack deontic operators in their primitive vocabulary, but which are such that we can *define* those operators in them, somewhat in the spirit of the well known Andersonian reduction. The detailed proof of this result in effect forms the bulk of Åqvist [1996]; here, we just present enough material so as to enable us to state that result in an intelligible way.

Consider the result of banishing the dyadic deontic operators  $O$  and  $P$  from the *primitive* logical vocabulary of the systems  $\mathbf{G}m[m = 1, 2, \dots]$ . Then, for any positive integer  $m$ , let the *axiomatic system*  $\mathbf{H}m$  of alethic modal logic with frame constants be determined by the rules of proof *modus ponens* and  $N$ -necessitation, and the axiom schemata  $\mathbf{a0}$  (= all tautologies over the present reduced language),  $\mathbf{a8}$  (=  $\mathbf{S5}$ -schemata for  $N, M$ ),  $\mathbf{\alpha5}$ ,  $\mathbf{\alpha9}$  and  $\mathbf{\alpha10}$ ; i.e. by those axiom schemata in  $\mathbf{G}m$  that do not contain occurrences of  $O$  or  $P$ . Clearly each system  $\mathbf{G}m$  is an extension of  $\mathbf{H}m$ .

As to the *semantics* of the alethic modal logics  $\mathbf{H}m[m = 1, 2, \dots]$ , a  $\mathbf{H}m$ -*structure* will be the ordered quadruple that results from deleting the function *best* in a  $\mathbf{G}m$ -*structure*, and a  $\mathbf{H}m$ -*model* will be a  $\mathbf{H}m$ -structure satisfying (a)–(d) in the requirement *Exactly  $m$  Non-empty levels of Perfection* (whereas the condition  $\gamma^0$  on *best* vanishes altogether). We then have the following result:

**THEOREM 80.** (Soundness and completeness of the systems  $\mathbf{H}m[m = 1, 2, \dots]$ )

*Weak version:* For every sentence  $A$ :  $A$  is  $\mathbf{H}m$ -provable iff  $A$  is  $\mathbf{H}m$ -valid.

*Strong version:* For each set  $S$  of sentences:  $S$  is  $\mathbf{H}m$ -consistent iff  $S$  is  $\mathbf{H}m$ -satisfiable.

**Proof.** See Åqvist [1996]. ■

What is the interest of the just considered infinite hierarchy  $\mathbf{H}m$  of alethic modal logics? We take it to be this: although the operators (for conditional obligation) and  $P$  (for conditional permission) are not primitive in the language of the  $\mathbf{H}m$ , they can be *defined* in those systems as follows:

$$\begin{aligned} O_B A &= \text{df } [M(Q_1 \wedge B) \supset N((Q_1 \wedge B) \supset A)] \wedge \\ \text{Def O.} \quad & [(\neg M(Q_1 \wedge B) \wedge M(Q_2 \wedge B)) \supset N(Q_2 \wedge B \supset A)] \wedge \dots \wedge \\ & [(\neg M(Q_1 \wedge B) \wedge \dots \wedge \neg M(Q_{m-1} \wedge B) \wedge M(Q_m \wedge B)) \supset \\ & N(Q_m \wedge B \supset A)]. \end{aligned}$$

$$\begin{aligned} P_B A &= \text{df } M(Q_1 \wedge B \wedge A) \vee \\ \text{Def P.} \quad & (\neg M(Q_1 \wedge B) \wedge M(Q_2 \wedge B \wedge A)) \vee \dots \vee \\ & (\neg M(Q_1 \wedge B) \wedge \dots \wedge \neg M(Q_{m-1} \wedge B) \wedge M(Q_m \wedge B \wedge A)). \end{aligned}$$

We can then prove the following:

**THEOREM 81** (Deductive Equivalence for  $\mathbf{H}m$  and  $\mathbf{G}m$ ). *Let  $\mathbf{H}m + \text{Def O} + \text{Def P}$  be the result of adding the definitions Def O and Def P supra to the alethic system  $\mathbf{H}m$ . Then, for all  $m = 1, 2, \dots$ ,  $\mathbf{H}m + \text{Def O} + \text{Def P}$  is deductively equivalent to  $\mathbf{G}m$  in the sense that the following two conditions are satisfied:*

- (i)  $\mathbf{H}m + \text{Def O} + \text{Def P}$  contains  $\mathbf{G}m$ .
- (ii) Each of Def O and Def P is provable in the form of an equivalence in  $\mathbf{G}m$ .

**Proof.** See Åqvist [1996]. ■

An alternative, more ‘semantical’ method of representing the dyadic deontic systems  $\mathbf{G}m$  in the alethic modal logics  $\mathbf{H}m$  is this: define recursively

a certain *translation*  $\phi$  from the set of  $\mathbf{G}m$ -sentences into the set of  $\mathbf{H}m$ -sentences by the stipulations:

$$\begin{aligned}\phi(p) &= p, \text{ for each propositional variable } p \text{ in Prop} \\ \phi(\top) &= \top \\ \phi(\perp) &= \perp \\ \phi(Q_i) &= Q_i, \text{ for each positive integer } i \\ \phi(\neg A) &= \neg\phi(A) \\ \phi(A \wedge B) &= (\phi(A) \wedge \phi(B))\end{aligned}$$

and similarly for  $\mathbf{G}m$  sentences having  $\vee, \rightarrow, \Leftrightarrow$  as their principal sign.

$$\begin{aligned}\phi(NA) &= N\phi A \\ \phi(MA) &= M\phi A\end{aligned}$$

where we have written  $\phi A$  instead of  $\phi(A)$  to the right. Finally, we have two characteristic clauses corresponding to *Def O* and *Def P*:

$$\begin{aligned}\phi(O_B A) &= [M(Q_1 \wedge \phi B) \supset N((Q_1 \wedge \phi B) \supset \phi A)] \wedge \\ &\quad [(\neg M(Q_1 \wedge \phi B) \wedge M(Q_2 \wedge \phi B)) \supset N(Q_2 \wedge \phi B \supset \phi A)] \wedge \dots \wedge \\ &\quad [(\neg M(Q_1 \wedge \phi B) \wedge \dots \wedge \neg M((Q_{m-1} \wedge \phi B) \wedge M(Q_m \wedge \phi B)) \supset \\ &\quad N(Q_m \wedge \phi B \supset \phi A)]\end{aligned}$$

Similarly for  $\phi(P_B A)$ : write it out as an  $m$ -termed disjunction!

We then have the following result:

**THEOREM 82** (Translation for  $\mathbf{G}m$  and  $\mathbf{H}m$ ). *For each positive integer  $m$ , and for each  $\mathbf{G}m$ -sentence  $A$ :*

$$\vdash_{\mathbf{G}m} A \text{ iff } \vdash_{\mathbf{H}m} \phi A.$$

**Proof.** Again, see Åqvist [1996]. The left-to-right direction is more or less immediate from the proof of the Deductive Equivalence theorem *supra*. The proof of the right-to-left is reminiscent of that of the Theorem on the relation of  $\mathbf{G}$  to the  $\mathbf{G}m$  in respect of utilizing a relevant completeness result in the second step. ■

Combining the present Translation theorem with the Theorem on the relation of  $\mathbf{G}$  to the  $\mathbf{G}m$ , we obtain the obvious

**COROLLARY 83.** *For any  $\mathbf{G}$ -sentence  $A$ :  $\vdash_{\mathbf{G}} A$  iff for all  $m = 1, 2, \dots$ ,  $\vdash_{\mathbf{H}m} \phi A$ .*

**Proof.** Immediate from the two theorems just mentioned. ■

## ACKNOWLEDGEMENTS

Research carried out under the auspices of the Bank of Sweden Tercentenary Foundation project RJ 79/23 on criminal intent and negligence.

*Uppsala Universitet, Sweden.*

## BIBLIOGRAPHY

Within the boundaries of a reasonable-size bibliography it is now, i.e. in 1982, impossible to do justice to the existing literature on deontic logic, especially if one is to include such topics as the logic of imperatives, ethical and legal theory, and what have you. Laudable attempts have been made in the past, though: the bibliographies of Anderson [1956] (together with a *Supplementary Bibliography* (1966)), Rescher [1966], Von Wright [1968], Chellas [1969] and Føllesdal and Hilpinen [1971] are still of a manageable size. *Really* comprehensive bibliographies are: Berkemann and Strasser [1974] (around 800 items) and Conte and Di Bernardo [1977] (approaching 1500 items!). So, naturally, my own present attempt will be fairly selective; and it remains unclear to me exactly what principle of selection is being applied. Anyway, here we go:

- [Alchourrón and Bulygin, 1971] C. E. Alchourrón and E. Bulygin. *Normative Systems*. (Library of Exact Philosophy), Springer-Verlag, Vienna, New York, 1971.
- [Alchourrón and Makinson, 1981] C. E. Alchourrón and D. Makinson. Hierarchies of regulations and their logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, D. Reidel, Dordrecht, pp. 125–148, 1981.
- [al-Hibri, 1978] A. al-Hibri. *Deontic Logic: A Comprehensive Appraisal and a New Proposal*. Univ. Press of America, Washington DC. 1978.
- [Anderson, 1956] A. R. Anderson. The formal analysis of normative systems. In N. Rescher, editor, *The Logic of Decision and Action*, Univ. Pittsburgh, 1967, pp. 147–213. 1956.
- [Anderson, 1958] A. R. Anderson. A reduction of deontic logic to alethic modal logic. *Mind* 67:100–103, 1958.
- [Anderson, 1959] A. R. Anderson. On the logic of commitment. *Philosophical Studies* 10:23–27, 1959.
- [Anderson, 1962] A. R. Anderson. Logic, norms, and roles. *Ratio* 4:36–49, 1962.
- [Anderson, 1968] A. R. Anderson. A new square of opposition: Eubouliatic logic. In *Akten des XIV. Internationalen Kongresses für Philosophie (Wien, 2–9 Sept. 1968)*, Verlag Herder, Wien, pp. 271–284, 1968.
- [Anderson, 1971] A. R. Anderson. The logic of Hohfeldian propositions. *Univ. Pittsburgh Law Review* 33:29–38, 1971.
- [Apostel, 1960] L. Apostel. Game theory and the interpretation of deontic logic. *Logique et Analyse* 3:70–90, 1960.
- [Åqvist, 1963] L. Åqvist. A note on commitment. *Philosophical Studies* 14:22–25, 1963.
- [Åqvist, 1963a] L. Åqvist. Postulate sets and decision procedures for some systems of deontic logic. *Theoria* 29:154–175, 1963.
- [Åqvist, 1965] L. Åqvist. Choice-offering and alternative-presenting disjunctive commands. *Analysis* 25:182–184, 1965.
- [Åqvist, 1966] L. Åqvist. “Next” and “Ought”, alternative foundations for Von Wright’s tense-logic, with an application to deontic logic. *Logique et Analyse* 9:231–251, 1966.
- [Åqvist, 1967] L. Åqvist. Good Samaritans, contrary-to-duty imperatives, and epistemic obligations. *Noûs* 1:361–379, 1967.
- [Åqvist, 1969] L. Åqvist. Improved formulations of act-utilitarianism. *Noûs* 3:299–323, 1969.

- [Åqvist, 1973] L. Åqvist. Modal logic with subjunctive conditionals and dispositional predicates. *J. Philosophical Logic* 2:1–76, 1973.
- [Åqvist, 1981] L. Åqvist. The Protagoras case: an exercise in elementary logic for lawyers. In W. Rabinowicz, editor, *Tankar och Tankefel tillägnade Zalma Puterman*, Filosofiska Föreningen, Uppsala, pp. 211–224. Also in *Annales Academiae Regiae Scientiarum Upsaliensis* 24:1981–1982, Almqvist & Wiksell International, Stockholm, pp. 5–15, 1981.
- [Åqvist, 1987] L. Åqvist. *An Introduction to Deontic Logic and the Theory of Normative Systems*. In *Indices* (Monographs in Philosophical Logic and Formal Linguistics, Bibliopolis, Naples), 1987.
- [Åqvist, 1993] L. Åqvist. A completeness theorem in deontic logic with systematic frame constants. *Logique et Analyse*, 36:177–192, 1993.
- [Åqvist, 1996] L. Åqvist. Systematic frame constants in defeasible deontic logic: a new form of Andersonian reduction. Forthcoming in D. Nute, editor, *Defeasible Deontic Reasoning*, Kluwer, Dordrecht, 1996.
- [Åqvist, 1997] L. Åqvist. On certain extensions of von Kutschera's preference-based dyadic deontic logic. In W. Lenzen, editor, *Das weite Spektrum der analytischen Philosophie: Festschrift für Franz von Kutschera*, W. de Gruyter, Berlin, New York, 1997.
- [Åqvist, 1997a] L. Åqvist. Branching time in deontic logic: remarks on an example by Alchourrón and Bulygin. In E. Garzón Valdés *et al.*, editors, *Normative Systems in Legal and Moral Theory: Festschrift for Carlos E. Alchourrón and Eugenio Bulygin*, Duncker and Humblot, Berlin, 1997.
- [Åqvist and Hoepelman, 1981] L. Åqvist and J. Hoepelman. Some theorems about a "tree" system of deontic tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, D. Reidel, Dordrecht, pp. 187–221, 1981.
- [Bailhache, 1981] P. Bailhache. Analytical deontic logic: authorities and addressees. *Logique et Analyse* 24:65–80, 1981.
- [Bailhache, 1991] P. Bailhache. *Essai de logique déontique*. Librairie philosophique J. Vrin, Paris, 1991.
- [Bailhache, 1993] P. Bailhache. The deontic branching time: two related conceptions. *Logique et Analyse*, 36:156–175, 1993.
- [Berg, 1960] J. Berg. A note on deontic logic. *Mind* 69:566–567, 1960.
- [Bergström, 1966] L. Bergström. *The Alternatives and Consequences of Actions. An Essay on Certain Fundamental Notions in Teleological Ethics*. Almqvist and Wiksell, Stockholm, 1966.
- [Bergström, 1968] L. Bergström. Alternatives and utilitarianism. *Theoria* 34:163–170, 1968.
- [Bergström, 1974] L. Bergström. Hintikka on *prima facie* obligations. *Theoria* 40:163–165, 1974.
- [Berkemann and Strasser, 1974] J. Berkemann and P. Strasser. Bibliographie zur Normenlogik. In H. Lenk, editor, *Normenlogik*, Verlag Dokumentation, Pullach bei München, pp. 207–251, 1974.
- [Castañeda, 1954] H.-N. Castañeda. *La lógica general de las normas y la ética*. Universidad de San Carlos (Guatemala), 30:129–196, 1954.
- [Castañeda, 1959] H.-N. Castañeda. The logic of obligation. *Philosophical Studies* 10:17–23, 1959.
- [Castañeda, 1967] H.-N. Castañeda. Ethics and logic: Stevensonianism revisited. *J. Philosophy* 64:671–683, 1967.
- [Castañeda, 1968] H.-N. Castañeda. A problem for utilitarianism. *Analysis* 28:141–142, 1968.
- [Castañeda, 1968a] H.-N. Castañeda. Acts, the logic of obligation, and deontic calculi. *Philosophical Studies* 19:13–26, 1968.
- [Castañeda, 1974] H.-N. Castañeda. *The Structure of Morality*. Charles C Thomas, Springfield, Illinois, 1974.
- [Castañeda, 1981] H.-N. Castañeda. The paradoxes of deontic logic: the simplest solution to all of them in one fell swoop. In R. Hilpinen, editor, *New Studies in Deontic Logic*, D. Reidel, Dordrecht, pp. 37–85, 1981.



- [Chellas, 1969] B. F. Chellas. *The Logical Form of Imperatives*. Perry Lane Press, Stanford, California, 1969.
- [Chellas, 1974] B. F. Chellas. Conditional obligation. In S. Stenlund, editor, *Logical Theory and Semantic Analysis. Essays dedicated to Stig Kanger on his Fiftieth Birthday*, D. Reidel, Dordrecht, pp. 23–33, 1974.
- [Chellas, 1975] B. F. Chellas. Basic conditional logic. *J. Philosophical Logic* 4:133–153, 1975.
- [Chisholm, 1963] R. M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis* 24:33–36, 1963.
- [Chisholm, 1967] R. M. Chisholm. Comments on Von Wright's "The logic of action". In N. Rescher, editor, *The Logic of Decision and Action*, Univ. Pittsburgh Press, Pittsburgh, pp. 137–139, 1967.
- [Conte and Di Bernardo, 1977] A. G. Conte and G. Di Bernardo. Bibliografia. In G. Di Bernardo, editor, *Logica deontica e semantica*, Società editrice il Mulino, Bologna, pp. 349–447, 1977.
- [Danielsson, 1968] S. Danielsson. *Preference and Obligation. Studies in the Logic of Ethics*. Filosofiska föreningen, Uppsala, 1968.
- [Fisher, 1961] M. Fisher. A three-valued calculus for deontic logic. *Theoria* 27:107–118, 1961.
- [Fisher, 1961a] M. Fisher. A logical theory of commanding. *Logique et Analyse* 4:154–169, 1961.
- [Føllesdal and Hilpinen, 1971] D. Føllesdal and R. Hilpinen. Deontic logic: an introduction. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Reading*, D. Reidel, Dordrecht, pp. 1–35, 1971.
- [Gärdenfors, 1978] P. Gärdenfors. On the interpretation of deontic logic. *Logique et Analyse* 21:371–398, 1978.
- [Hanson, 1965] W. H. Hanson. Semantics for deontic logic. *Logique et Analyse* 8:177–190, 1965.
- [Hanson, 1966] W. H. Hanson. A logic of commands. *Logique et Analyse* 9:329–343, 1966.
- [Hansson, 1969] B. Hansson. An analysis of some deontic logics. *Noûs* 3:373–398. Reprinted in R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, Reidel, pp. 121–147, 1971.
- [Hansson, 1970] B. Hansson. Deontic logic and different levels of generality. *Theoria* 36:241–248, 1970.
- [Hansson and Gärdenfors, 1973] B. Hansson and P. Gärdenfors. A guide to intensional semantics. In *Modality, Morality and Other Problems of Sense and Nonsense: Essays dedicated to Sören Halldén*, CWK Gleerup, Lund, pp. 151–167, 1973.
- [Hedenius, 1941] I. Hedenius. *Om rätt och moral* (On Law and Morals). Stockholm, 1941.
- [Hilbert and Ackermann, 1928] D. Hilbert and W. Ackermann. *Grundzüge der theoretischen Logik*. Berlin, 1928. American edition: Dover Publications, New York, 1946.
- [Hilpinen, 1979] R. Hilpinen. Disjunctive permissions and conditionals with disjunctive antecedents. In *Proceedings of the Second Soviet-Finnish Logic Conference*, December 1979.
- [Hilpinen, 1981] R. Hilpinen. Conditionals and Possible Worlds. In Guttorm Fløistad, editor, *Contemporary Philosophy: A New Survey Vol. I: Philosophy of Language and Philosophical Logic*, Martinus Nijhoff, The Hague, 1981.
- [Hintikka, 1957] J. Hintikka. Quantifiers in deontic logic. In *Societas Scientiarum Fennica, Commentationes Humanarum Litterarum* 23:4, Helsinki, pp. 1–23, 1957.
- [Hintikka, 1971] J. Hintikka. Some main problems of deontic logic. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, Reidel, Dordrecht, pp. 59–104, 1971.
- [Hohfeld, 1919] W. N. Hohfeld. *Fundamental Legal Conceptions as Applied in Judicial Reasoning and Other Essays*, Yale Univ. Press, New Haven, 1919.
- [Hughes and Cresswell, 1968] G. E. Hughes and M. J. Cresswell. *An Introduction to Modal Logic*. Methuen, London, 1968.

- [Kalinowski, 1953] G. Kalinowski. Théorie des propositions normatives. *Studia Logica* 1:147–182, 1953. Reprinted in the next item.
- [Kalinowski, 1972] G. Kalinowski. *Études de logique déontique. I (1953–1969)*. Librairie générale de droit et de jurisprudence, Paris, 1972.
- [Kalish and Montague, 1964] D. Kalish and R. Montague. *LOGIC: Techniques of Formal Reasoning*. Harcourt, Brace & World, New York, 1964.
- [Kamp, 1973] H. Kamp. Free choice permission. *Proc. Aristotelian Society*, N.S. 74:57–74, 1973–74.
- [Kamp, 1979] H. Kamp. Semantics versus pragmatics. In F. Guenther and S.J. Schmidt, editors, *Formal Semantics and Pragmatics for Natural Languages*, D. Reidel, Dordrecht, pp. 255–287, 1979.
- [Kanger, 1957] S. Kanger. *New Foundations for Ethical Theory*. Stockholm, 1957. Reprinted in R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, Reidel, Dordrecht, pp. 36–58, 1971.
- [Kanger, 1957a] S. Kanger. The Morning Star paradox. *Theoria* 23:1–11, 1957.
- [Kanger and Kanger, 1966] S. Kanger and H. Kanger. Rights and parliamentarism. *Theoria* 32:85–115, 1966.
- [Knuuttila, 1981] S. Knuuttila. The emergence of deontic logic in the fourteenth century. In R. Hilpinen, editor, *New Studies in Deontic Logic*, D. Reidel, Dordrecht, pp. 225–248, 1981.
- [Kripke, 1963] S. A. Kripke. Semantical analysis of modal logic I: Normal modal propositional calculi. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 9:67–96, 1963.
- [Lemmon, 1965] E. J. Lemmon. Deontic logic and the logic of imperatives. *Logique et Analyse* 8:39–71, 1965.
- [Lemmon and Scott, 1966] E. J. Lemmon and D. Scott. *Intensional Logic*, preliminary draft of initial chapters by E.J. Lemmon, July 1966. Nowadays available as *An Introduction to Modal Logic* (American Philosophical Quarterly Monograph No. 11), edited by K. Segerberg, Basil Blackwell, Oxford, 1977.
- [Lenzen, 1977] W. Lenzen. Protagoras versus Euathlus: reflections on a so-called paradox. *Ratio* 19:176–180, 1977.
- [Lewis, 1974] D. K. Lewis. Semantic analyses for dyadic deontic logic. In S. Stenlund, editor, *Logical Theory and Semantic Analysis*, D. Reidel, Dordrecht, pp. 1–14, 1974.
- [Lewis, 1978] D. K. Lewis. A problem about permission. In E. Saarinen *et al.*, editors, *Essays in Honour of Jaakko Hintikka*, D. Reidel, Dordrecht, pp. 163–175, 1978.
- [Lindahl, 1977] L. Lindahl. *Position and Change. A Study in Law and Logic*, D. Reidel, Dordrecht, 1977.
- [Makinson, 1966] D. Makinson. On some completeness theorems in modal logic. *Zeitschrift für mathematische Logik und Grundlagen der Mathematik* 12:379–384, 1966.
- [Makinson, 1981] D. Makinson. Quantificational reefs in deontic waters. In R. Hilpinen, editor, *New Studies in Deontic Logic*, D. Reidel, Dordrecht, pp. 87–91, 1981.
- [Mally, 1926] E. Mally. *Grundgesetze des Sollens. Elemente der Logik des Willens*. Leuschner & Lubensky, Graz, 1926.
- [Menger, 1934] K. Menger. *Moral, Wille und Weltgestaltung. Grundlegung zur Logik der Sitten*. Verlag Julius Springer, Vienna, 1934.
- [Montague, 1960] R. Montague. Logical necessity, physical necessity, ethics, and quantifiers. *Inquiry* 4:259–269, 1960.
- [Montague, 1968] R. Montague. Pragmatics. In R. Klibansky, editor, *Contemporary Philosophy, Vol. I: Logic and the Foundations of Mathematics*, La Nuova Italia Editrice, Firenze, pp. 102–122, 1968.
- [Moritz, 1963] M. Moritz. Permissive Sätze, Erlaubnissätze und deontische Logik. In *Philosophical Essays Dedicated to Gunnar Aspelin*, CWK Gleerup, Lund, pp. 108–121, 1963.
- [Mott, 1973] P. L. Mott. On Chisholm's paradox. *J. Philosophical Logic* 2:197–211, 1973.
- [Nowell-Smith and Lemmon, 1960] P. H. Nowell-Smith and E.J. Lemmon. Escapism: the logical basis of ethics. *Mind* 69:289–300, 1960.
- [Nute, 1981] D. Nute. Permission. *J. Philosophical Logic*, forthcoming.

- [Oppenheim, 1944] F. E. Oppenheim. Outline of a logical analysis of law. *Philosophy of Science* 11:142–160, 1944.
- [Powers, 1967] L. Powers. Some deontic logicians. *Noûs* 1:381–400, 1967.
- [Prawitz, 1968] D. Prawitz. A discussion note on utilitarianism. *Theoria* 34:76–84, 1968.
- [Prawitz, 1970] D. Prawitz. The alternatives to an action. *Theoria* 36:116–126, 1970.
- [Price, 1948] R. Price. *A Review of the Principal Questions and Difficulties in Morals*. 1st ed. 1758. Edited by D.D. Raphael, Clarendon Press, Oxford, 1948.
- [Prior, 1954] A. N. Prior. The paradoxes of derived obligation. *Mind* 63:64–65, 1954.
- [Prior, 1955] A. N. Prior. *Formal Logic*. Clarendon Press, Oxford, 1955.
- [Prior, 1956] A. N. Prior. Review of G. Kalinowski: 'Théorie des propositions normatives' (see above). *J. Symbolic Logic* 21:191–192, 1956.
- [Prior, 1958] A. N. Prior. Escapism: the logical basis of ethics. In A. Melden, editor, *Essays in Moral Philosophy*, Univ. Washington Press, Seattle, pp. 135–146, 1958.
- [Purtil, 1973] R. L. Purtil. Deontically perfect worlds and prima facie obligations. *Philosophia* 3:429–438, 1973.
- [Pörn, 1970] I. Pörn. *The Logic of Power*. Basil Blackwell, Oxford, 1970.
- [Pörn, 1977] I. Pörn. Action Theory and Social Science. Some Formal Models. D. Reidel, Dordrecht, 1977.
- [Quine, 1963] W. V. O. Quine. Carnap and logical truth. In P.A. Schilpp, editor, *The Philosophy of Rudolph Carnap*, Open Court, La Salle, Illinois, pp. 385–406, 1963.
- [Rescher, 1958] N. Rescher. An axiom system for deontic logic. *Philosophical Studies* 9:24–30, 1958.
- [Rescher, 1962] N. Rescher. Conditional permission in deontic logic. *Philosophical Studies* 13:1–6, 1962.
- [Rescher, 1966] N. Rescher. *The Logic of Commands*. Routledge & Kegan Paul, London, 1966.
- [Rescher and Robison, 1964] N. Rescher and J. Robison. Can one infer commands from commands? *Analysis* 24:176–179, 1964.
- [Robison, 1964] J. Robison. Who, What, Where, and When: A Note on Deontic Logic. *Philosophical Studies* 15:89–92, 1964.
- [Ross, 1944] A. Ross. Imperatives and logic. *Theoria* 7:53–71, 1941. Also in *Philosophy of Science* 11:30–46, 1944.
- [Ross, 1930] W. D. Ross. *The Right and the Good*. Clarendon Press, Oxford, 1930.
- [Ross, 1939] W. D. Ross. *Foundations of Ethics*. Clarendon Press, Oxford, 1939.
- [Scott, 1970] D. Scott. Advice on modal logic. In K. Lambert, editor, *Philosophical Problems in Logic. Some Recent Developments*, D. Reidel, Dordrecht, pp. 143–173, 1970.
- [Seegerberg, 1971] K. Seegerberg. Some logics of commitment and obligation. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, D. Reidel, Dordrecht, pp. 148–158, 1971.
- [Seegerberg, 1971a] K. Seegerberg. *An Essay in Classical Modal Logic*. Filosofiska föreningen, Uppsala, 1971.
- [Sellars, 1967] W. Sellars. Reflections on contrary-to-duty imperatives. *Noûs* 1:303–344, 1967.
- [Smiley, 1963] T. J. Smiley. Relative necessity. *J. Symbolic Logic* 28:113–134, 1963.
- [Smiley, 1963a] T. J. Smiley. The logical basis of ethics. *Acta Philosophica Fennica* 16:237–246, 1963.
- [Smullyan, 1978] R. M. Smullyan. *What is the name of this book? — the riddle of Dracula and other logical puzzles*. Prentice-Hall, Englewood Cliffs, 1978.
- [Spohn, 1975] W. Spohn. An analysis of Hansson's dyadic deontic logic. *J. Philosophical Logic* 4:237–252, 1975.
- [Stenius, 1963] E. Stenius. The principles of a logic of normative systems. *Acta Philosophica Fennica* 16:247–260, 1963.
- [Suppes, 1957] P. Suppes. *Introduction to Logic*. Van Nostrand, Princeton, 1957.
- [Thomason, 1981] R. H. Thomason. Deontic logic as founded on tense logic. In R. Hilpinen, editor, *New Studies in Deontic Logic*, D. Reidel, Dordrecht, pp. 165–176, 1981.

- [Thomason, 1981a] R. H. Thomason. Deontic logic and the role of freedom in moral deliberation. In R. Hilpinen, editor, *New Studies in Deontic Logic*, D. Reidel, Dordrecht, pp. 177–186, 1981.
- [Tomberlin, 1981] J. E. Tomberlin. Contrary-to-duty imperatives and conditional obligations. *Noûs* 15:357–375, 1981.
- [Tomberlin and F. McGuinness, 1977] J. E. Tomberlin and McGuinness. “Because” and Good Samaritans. *Critica* 9:67–81, (México), 1977.
- [Van Eck, 1981] J. E. Van Eck. *A System of Temporally Relative Modal and Deontic Predicate Logic and its Philosophical Applications*. Department of Philosophy, University of Groningen, The Netherlands, 1981. Also in *Logique et Analyse*, 25:249–290 and 25:339–381, 1982.
- [Van Fraassen, 1972] B. C. Van Fraassen. The logic of conditional obligation. *J. Philosophical Logic* 1:417–438, 1972.
- [Von Kutschera, 1973] F. Von Kutschera. *Einführung in die Logik der Normen, Werte und Entscheidungen*. Alber, Freiburg, 1973.
- [Von Kutschera, 1974] F. Von Kutschera. Normative Präferenzen und bedingte Gebote. In H. Lenk, editor, *Normenlogik*, Verlag Dokumentation, Pullach bei München, pp. 137–165, 1974.
- [Von Wright, 1951] G. H. Von Wright. Deontic logic. *Mind* 60:1–15, 1951.
- [Von Wright, 1951a] G. H. Von Wright. *An Essay in Modal Logic*. North-Holland, Amsterdam, 1951.
- [Von Wright, 1955] G. H. Von Wright. Om s.k. praktiska slutledningarna (On so-called Practical Inferences). *Tidskrift för Retsvetenskap* 68:465–495, 1955.
- [Von Wright, 1956] G. H. Von Wright. A note on deontic logic and derived obligation. *Mind* 65:507–509, 1956.
- [Von Wright, 1963] G. H. Von Wright. *Norm and Action. A Logical Inquiry*. Routledge & Kegan Paul, London, 1963.
- [Von Wright, 1964] G. H. Von Wright. A new system of deontic logic. In *Danish Yearbook of Philosophy* 1:173–182, 1964.
- [Von Wright, 1965] G. H. Von Wright. A correction to a new system of deontic logic. *Danish Yearbook of Philosophy* 2:103–107, 1965.
- [Von Wright, 1967] G. H. Von Wright. The logic of action — A sketch. In N. Rescher, editor, *The Logic of Decision and Action*, Univ. Pittsburgh Press, Pittsburgh, pp. 121–136, 1967.
- [Von Wright, 1968] G. H. Von Wright. *An Essay on Deontic Logic and the General Theory of Action*. North-Holland, Amsterdam, 1968.
- [Von Wright, 1971] G. H. Von Wright. Deontic logic and the theory of conditions. In R. Hilpinen, editor, *Deontic Logic: Introductory and Systematic Readings*, D. Reidel, Dordrecht, pp. 159–177, 1971.
- [Von Wright, 1974] G. H. Von Wright. Handlungslogik. In H. Lenk, editor, *Normenlogik*, Verlag Dokumentation, Pullach bei München, pp. 9–24, 1974.
- [Von Wright, 1977] G. H. Von Wright. Zur Einführung. In A.G. Conte, R. Hilpinen and G.H. Von Wright, editors, *Deontische Logik und Semantik*, Athenaiion, Wiesbaden, pp. 7–8, 1977.
- [Wedberg, 1951] A. Wedberg. Some problems in the logical analysis of legal science. *Theoria* 17:246–275, 1951.
- [Wedberg, 1969] A. Wedberg. Den klassiska deontiska konsekvensprincipens paradoxer: några preformella reflexioner. In *Logik, rätt och moral: Filosofiska studier tillägnade Manfred Moritz*, Studentlitteratur, Lund, pp. 213–232, 1969.
- [Weinberger, 1970] O. Weinberger. *Rechtslogik. Versuch einer Anwendung moderner Logik auf das juristische Denken*. Springer-Verlag, Vienna, New York, 1970.

## DEONTIC LOGIC AND CONTRARY-TO-DUTIES

### 1 INTRODUCTION

Deontic logic is concerned with the logical analysis of such normative notions as *obligation*, *permission*, *right* and *prohibition*. Although its origins lie in systematic legal and moral philosophy, deontic logic has begun to attract the interest of researchers in other areas, particularly computer science, management science and organisation theory. Among the application areas which have already received some attention in the literature are: issues of knowledge representation in the design of legal expert systems; the formal specification of aspects of computer systems, for instance in regard to security and access control policies, fault tolerance, and database integrity constraints; the formal characterisation of aspects of organisational structure, pertaining for example to the responsibilities and powers which agents are required or authorised to exercise. The “ $\Delta$ EON” workshop proceedings provide some illustrations of work in these areas (see [ $\Delta$ EON91;  $\Delta$ EON94;  $\Delta$ EON96]).

Deontic logic is one of the formal tools needed in the design and specification of *normative systems*, where the latter are understood to be sets of agents (human or artificial) whose interactions can fruitfully be regarded as norm-governed; the norms prescribe how the agents ideally should and should not behave, what they are permitted to do, and what they have a right to do. Importantly, the norms allow for the possibility that actual behaviour may at times deviate from the ideal, i.e. that violations of obligations, or of agents’ rights, may occur.

In [Jones and Sergot, 1992; Jones and Sergot, 1993] Jones and Sergot argue that it is precisely when the possibility of norm violation is kept open that deontic logic has a potentially useful role to play. If agents can always be assumed to behave in conformity to norm, the normative dimension ceases to be of interest: the actual does not depart from the ideal, so nothing is lost by *merely* describing what the agents in fact do. Thus, although it is correct to say that deontic logic deals with the logic of *obligation*, *permission* and other normative notions, a more insightful characterisation, Jones and Sergot suggest, views deontic logic as essentially concerned with representing and reasoning about the distinction between the actual and the ideal. Systems for which that distinction is relevant are genuinely normative systems, and their specification will ordinarily include “secondary” norms which indicate what is to be done in circumstances in which actual behaviour has deviated from the ideal. The methodological guidelines proposed in [Jones and Sergot, 1992; Jones and Sergot, 1993] strongly suggest

that “secondary” norms of this kind (first dubbed “contrary-to-duty” in [Chisholm, 1963]) will be a prominent feature of normative systems, and thus that any adequate deontic logic must accommodate them. However, the analysis of contrary-to-duty obligation sentences has proved to be a task of some considerable complexity. And it is this issue — at the very core of deontic logic — which this chapter addresses.

The plan is as follows: we first (Section 2) describe Standard Deontic Logic, and a number of its defects, including problems regarding the representation of conditional obligation sentences. In the course of Section 3 we examine a number of different theories which have attempted to accommodate contrary-to-duty obligation sentences (CTDs), and in the course of this examination we identify several criteria — eight in all — which, we argue, an adequate treatment of Chisholm’s puzzle about CTDs should meet. Some of these criteria are not tied to Chisholm’s problem, but apply quite generally to the analysis of CTDs. Section 4 presents a revised, and in parts considerably modified version of the [Carmo and Jones, 1997] theory of CTDs; its application to a number of CTD “scenarios” is investigated in some detail in Section 5, and this provides a further impression of the broad range of representational and reasoning issues which a CTD theory must address. Section 6 examines some possible counter-examples to the proposed analysis, thereby relating its treatment of CTD problems to other well-known issues in deontic logic, concerning — in particular — the closure of deontic operators under logical consequence, and the representation of conflicts of obligations. Section 7 offers further observations on alternative approaches based on temporal logic, the logic of action, and preference orderings, respectively. The overall aim of the chapter is to supply a rather detailed overview of a group of problems at the heart of deontic logic, and a guide to existing attempts to solve them.

## 2 DEONTIC LOGIC: THE STANDARD APPROACH

### 2.1 *Standard Deontic Logic*

The standard approach to deontic logic takes it to be a branch of modal logic, interpreting the necessity operator  $\Box$  as expressing ethical/legal necessity, i.e. as meaning “it is obligatory that”, and denoting it by O; accordingly, the dual possibility operator  $\Diamond = \neg\Box\neg$  is interpreted as expressing “it is permitted that” (and is denoted by P), and the impossibility modal construction  $\Box\neg$  is interpreted as expressing “it is forbidden that” (and is often denoted by F).

Axiomatically, the weakest deontic logic (called standard deontic logic, SDL for short) is then obtained by replacing the modal necessity schema (T) ( $\Box A \rightarrow A$ : unacceptable for a deontic interpretation, since what is

obligatory may fail to be the case) by the (D) schema (which requires that what is obligatory is permitted). Thus, following the Chellas classification [Chellas, 1980], SDL is the weakest normal modal system of type KD; that is, its theorems can be characterized as the smallest set of formulas that includes all instances of the following axiom schemas, and that is closed under the O-necessitation rule and Modus Ponens (MP).

Axiom schemas

- (PC) All instances of tautologies  
(PC stands for Propositional Calculus)
- (K)  $O(A \rightarrow B) \rightarrow (O A \rightarrow O B)$
- (D)  $(O A \rightarrow P A)$

Rules

$$\text{O-necessitation: } \frac{A}{O A}$$

$$\text{Modus Ponens (MP): } \frac{A, A \rightarrow B}{B}$$

We here employ capital letters ( $A, B, C, \dots$ ) to stand for arbitrary formulas (well-formed sentences of the underlying propositional modal logic), and we use lower case letters ( $p, q, \dots$ ) for arbitrary atomic sentences, and  $\perp$  and  $\top$  to denote, respectively, a contradiction and a tautology; parentheses will be omitted following the usual precedence rules for the operators; the Boolean connectives will be denoted by  $\neg, \wedge, \vee, \rightarrow$  and  $\leftrightarrow$ ; in the meta-language we denote such connectives by “not”, “and”, “or”, “if ... then ...” (or “implies”) and “iff” (if and only if), and in the meta-language we also avail ourselves of the universal and the existential quantifiers (these do not appear in the object language: we are concerned only with propositional modal logics). Moreover, as usual, we will use  $\vdash A$  (respectively  $\not\vdash A$ ) to denote that  $A$  is a theorem (respectively,  $A$  is not a theorem) of the underlying logical system; and, following the traditional philosophical/logical approach to deduction, we say (cf. [Chellas, 1980; Hughes and Cresswell, 1984]) that  $A$  is deducible from a set of hypotheses  $\Gamma$ , written  $\Gamma \vdash A$  (or simply  $A_1, \dots, A_n \vdash A$  if  $\Gamma$  is finite), iff  $A$  belongs to the smallest set of formulas that contains  $\Gamma$  and the theorems and that is closed under (MP).<sup>1</sup>

---

<sup>1</sup>In this way we get a Boolean, compact, deductive system (see e.g. [Bull and Segerberg, 2001]). Non-Boolean axiomatic approaches to deduction, where non-tautological rules may also be applied to the hypotheses, and not only to the theorems, may be found in some works in the field of mathematical logic, such as [Hamilton, 1978; Mendelson, 1979].

Semantically, the models  $\mathcal{M}$  of SDL are standard models [Chellas, 1980]:  $\mathcal{M} = (W, R, V)$ , where  $W$  is a non-empty set (the set of worlds),  $R$  is a binary relation on  $W$  and  $V$  is an assignment to each atomic sentence of a set of worlds; informally,  $V(p)$  denotes the set of worlds where  $p$  is true. In order to validate the schema (D) we require that the accessibility relation  $R$  is serial, i.e.  $(\forall w)(\exists v)wRv$  (using  $w, v, \dots$  to denote worlds and, as usual, writing  $wRv$  instead of  $\langle w, v \rangle \in R$ ). The deontic interpretation of the accessibility relation is as follows:  $wRv$  iff  $v$  is a deontic alternative to, or an ideal version of,  $w$ . The truth of a formula  $A$  in a world  $w$  of a model  $\mathcal{M}$  is denoted by  $\mathcal{M} \models_w A$  and is defined as usual: for instance,  $\mathcal{M} \models_w \bigcirc A$  iff  $(\forall v)$  (if  $wRv$  then  $\mathcal{M} \models_v A$ ); thus, informally,  $\bigcirc A$  is true in a world  $w$  iff  $A$  is true in all ideal versions of  $w$ . A formula  $A$  is true in a model  $\mathcal{M}$ , written  $\mathcal{M} \models A$ , iff  $A$  is true in all the worlds of the model  $\mathcal{M}$ ; and a formula  $A$  is valid, written  $\models A$ , iff  $A$  is true in all models.

## 2.2 SDL and its problems

It is widely accepted that SDL is not adequate as a basic deontic logic. In fact, few systems of logic have been as heavily criticised as SDL; SDL gives rise to a set of “paradoxes” (theorems of SDL that many have deemed to be counter-intuitive) and there are some deontic concepts and constructions which apparently cannot be expressed in SDL in a consistent manner. Some of the main examples will be given below. We have essentially two aims here: first, without any claims to originality, we comment on the reasons underlying the so-called paradoxes; secondly, we indicate which of these problems have a counterpart in other areas of applied modal logic (e.g., epistemic, doxastic and action logics), and which seem to be particular to deontic logic.

A first group of paradoxes has its origin in the closure of the  $\bigcirc$ -operator under logical consequence (that is, in the fact that SDL, like any normal modal logic, contains the (RM)-rule: “if  $\vdash A \rightarrow B$  then  $\vdash \bigcirc A \rightarrow \bigcirc B$ ”). Some well known examples are:

- Ross paradox:  $(\vdash \bigcirc A \rightarrow \bigcirc (A \vee B))$

*“If it is obligatory to mail the letter, then it is obligatory to mail the letter or to burn it”*

The question of the significance of this paradox has been the subject of considerable dispute. Whereas some claim that the second obligation (the one in the consequent) is a counter-intuitive consequence of the first, since it seems to leave open to the agent a choice to mail or to burn the letter, others maintain that the consequent does not leave a choice of this kind, because burning the letter is clearly not a way of meeting the obligation expressed by the antecedent. Given



the perspective on deontic logic advocated in Section 1, however, we should also look at the problem from the point of view of *violation*: supposing that  $A$  is obligatory and that  $A$  is not the case, how many obligations have been violated? If we accept the Ross theorem, then not only has the obligation that  $A$  been violated, but — in addition — for *each* state of affairs  $B$  which actually fails to obtain, an obligation that  $A \vee B$  has *also* been violated. This is a peculiar result; it contrasts, of course, with how things look from a *fulfilment* perspective; for if the obligation that  $A$  is fulfilled, then so are all the other obligations which can be derived by application of the Ross theorem. (We are assuming, as is natural, that within SDL violation of an obligation  $OC$  is to be expressed as the conjunction  $OC \wedge \neg C$ .)

- Free Choice Permission paradox:

This paradox has to do with the fact that (in SDL)  $\not\vdash P(A \vee B) \rightarrow (PA \wedge PB)$ , whereas — ordinarily — if it is permitted that  $A$  or  $B$  this would be understood to imply that  $A$  is permitted and  $B$  is permitted. We include this paradox in this group, since the reason why we cannot add  $P(A \vee B) \rightarrow (PA \wedge PB)$ , as a new axiom, to SDL is the fact that, by the (RM)-rule,  $\vdash PA \rightarrow P(A \vee B)$ , which together with  $P(A \vee B) \rightarrow (PA \wedge PB)$  would imply  $PA \rightarrow (PA \wedge PB)$ ; so permission to go to the cinema would imply permission to kill the President! Moreover, from any permission we could then deduce  $P\perp$ , which is inconsistent with the fact that  $\vdash O\top$ . However, in common with some other researchers, we think that this “paradox” is a pseudo-problem: if what we want to express is that both  $A$  and  $B$  are permitted, then we should simply represent that formally by  $PA \wedge PB$  (instead of by  $P(A \vee B)$ ).

- Good Samaritan paradox:

*“If it is obligatory that Mary helps John who has had an accident, then it is obligatory that John has an accident”*

On the assumption that “Mary helps John who has had an accident” is represented as the conjunction “Mary helps John and John has had an accident”, then the antecedent of the above conditional takes the form “ $O(A \wedge B)$ ”. Since, tautologically, a conjunction implies each of its conjuncts, the (RM)-rule yields the SDL theorem:  $\vdash O(A \wedge B) \rightarrow OB$ . In our view, the formal concepts needed to deal with problems about contrary-to-duty obligations can also provide an appropriate analysis of the Good Samaritan problem. So we return to this issue below, in Section 6.

- Deontic/epistemic paradox:

*“If it is obligatory that Mr. X knows that his wife commits adultery, then it is obligatory that X’s wife commits adultery”*

We here assume, as is usual, that the (T)-schema holds for the epistemic operator. So this problem is again a result of the fact that, in SDL, any logical consequence of that which is obligatory is itself obligatory.

We note that the closure of the necessity operator under logical consequence is also a source of problems for other applications of modal logic, for instance epistemic and doxastic logics, where the assumption that every agent knows (believes) every logical consequence of what he knows (believes) is an extreme idealisation. In the logic of action, too, it is surely not acceptable to suppose that an agent brings about all the logical consequences of that which he brings about (cf. [Elgesem, 1993]).<sup>2</sup>

A second problem of SDL has to do with the O -necessitation rule itself, according to which any tautology (more generally, any theorem) is obligatory, which is incompatible with the idea that obligations should be possible to fulfill and possible to violate. Similar problems occur with this rule in the epistemic and doxastic logics, where it requires that an agent knows (or believes) all theorems (called in [Hintikka, 1975] the “logical omniscience problem”), and in the logic of action, where it is in general supposed that that which can be brought about must be avoidable (see, e.g., [Elgesem, 1993; Santos and Carmo, 1996]).

A third problem of SDL is that, because of the (D)-schema, it is not possible to express consistently a conflict of obligations, even though, as a matter of fact, normative systems may indeed contain conflicting obligations. We shall return to this issue later in this chapter. But, again, we note at this point that this is not a problem only of deontic logic: similar problems may appear, for instance, in the logic of belief.

However, it is fair to say that, for most deontic logicians, the problem of how to represent *conditional obligation sentences* has been their principal reason for seeking an alternative to SDL. Let us denote by  $O(B/A)$  the “conditional obligation of  $B$ , given  $A$ ”; so  $O(B/A)$  is intended to mean that “it is obligatory that  $B$ , if  $A$  is the case”. In SDL there are two possible ways to represent such sentences:

$$\text{(option1)} \quad O(B/A) =_{\text{df}} A \rightarrow O B$$

$$\text{and (option2)} \quad O(B/A) =_{\text{df}} O(A \rightarrow B)$$

---

<sup>2</sup>In [Konolige and Pollack, 1993] it is argued that this problem, called the “side-effect problem” in [Bratman, 1987], is even worse for the logic of intentions - a logic which, it has been suggested, has very close similarities to deontic logic (see [Pörn, 1977; Jones, 1991]).

Note first what these two options have in common. With both of them we get (within SDL):

- (UN)  $\vdash O B \leftrightarrow O (B/\top)$   
 (SA)  $\vdash O (B/A) \rightarrow O (B/A \wedge C)$

The first theorem is generally seen as a good property, and has been accepted by many authors on the grounds that an unconditional obligation is a particular (limiting) case of a conditional obligation, where the condition is a logical truth. Here, however, we shall adopt the opposite view, in line with the opinion expressed by Carlos Alchourrón in [1993, pp. 62], who argued that (UN) was one of the wrong steps followed by almost all researchers in deontic logic.

(SA) is known as the “principle of strengthening of the antecedent”; it is problematic, since it appears to make the expression of defeasible (conditional or unconditional<sup>3</sup>) obligations impossible; but of course it is a commonplace feature of obligations that they are subject to exceptions. Consider, for instance, a conditional obligation to the effect that, if your aged mother is sick, then you should help her. Such a conditional obligation might well leave room for exceptions, just as penguins might be the exception to the generalisation that birds fly; supposing for example that your young child has been injured in a car accident, and urgently needs you at the hospital, the obligation to help your sick, aged mother may well be deemed to have been defeated, or overturned. But again this problem (which has some connections with the problem of how to deal with conflicting obligations) is not a specific issue of deontic logic; the problem of how to deal with defeasible conditionals appears in many other areas and has been a source of intensive research.

So far we have not yet found a problem that sets deontic logic apart from other branches of modal logic. But we here return to the point emphasised in the introduction and suggest that the issue of how to represent contrary-to-duty obligation sentences (CTDs) — obligations which come into force when some other obligation is violated — seems to be a specific problem of deontic logic. It has sometimes been proposed, however, that CTD obligations may be seen as handling exceptions to (primary) obligations. Although we accept that there may be some connections between the problem of how to deal with CTDs and the problems concerning allowable exceptions and default reasoning, it should be stressed that there are also crucially important differences (cf. [Prakken and Sergot, 1994]): when a CTD obligation comes into force because of some violation, we do not want then to say that the violated obligation has been defeated; it has not been overturned, it has been violated! We need to be able to integrate in

<sup>3</sup>Note that combining the two previous theorems we get  $\vdash O B \rightarrow O (B/A)$ , and so an unconditional obligation remains obligatory under any condition.

a single logical framework the ability to make deductions at two different levels: on the level of what *ideally* should be the case, and on the level of what *actually* should be the case, given the circumstances (where, of course, the circumstances might include the fact that what has happened deviates from the ideal). The simultaneous specification of both ideal behavior and of what to do when actual behavior deviates from the ideal is a central task of deontic logic.

### 3 CONTRARY-TO-DUTIES

#### 3.1 Chisholm's CTD-paradox and SDL

Consider the following set of four sentences, formulated by Chisholm in 1963 [Chisholm, 1963]:

EXAMPLE 1.

- (a) It ought to be that a certain man go to help his neighbours.
- (b) It ought to be that if he goes he tell them he is coming.
- (c) If he does not go, he ought not to tell them he is coming.
- (d) He does not go.

There is widespread *agreement* in the literature that, from the intuitive point of view, this set is consistent, and its members are logically independent of each other; and there is a good deal of *disagreement* in the literature as regards which *further* requirements an adequate formal representation of the Chisholm set should meet. We start by discussing whether the Chisholm set can be represented in SDL in a way that meets this set of two minimum requirements, leaving the discussion of other further requirements to later.

It is straightforward to represent sentences (a) and (d) in SDL; the question is how to represent (b) and (c), since they express conditional obligations. Let us leave that open for the moment, and represent them by the use of our binary conditional obligation operator above; we then get (using “tell” and “help” in an obvious way as abbreviations of the sentences concerned):

- (a)  $O \text{ help}$  (or  $O (\text{help}/\top)$ , since in  $\text{SDL} \vdash O \text{ help} \leftrightarrow O (\text{help}/\top)$ )
- (b)  $O (\text{tell} / \text{help})$
- (c)  $O (\neg\text{tell} / \neg\text{help})$
- (d)  $\neg\text{help}$

In regard to the representation of conditional obligations in SDL, recall the two alternatives:

$$\text{(option1)} \quad \text{O}(B/A) =_{\text{df}} A \rightarrow \text{O} B$$

$$\text{and (option2)} \quad \text{O}(B/A) =_{\text{df}} \text{O}(A \rightarrow B)$$

With (option1) we get the following results:

$$\vdash \neg A \rightarrow \text{O}(B/A)$$

$$\text{(FD)} \quad \vdash A \wedge \text{O}(B/A) \rightarrow \text{O} B$$

Given that an expression of the form  $\text{O}(B/A)$  is intended to mean that, in circumstances  $A$ ,  $B$  is obligatory, the first of these two results is clearly problematic. From the fact that it is not raining we should not be able to deduce that, in circumstances where it is raining, it is obligatory that the President be assassinated. The other theorem has to do with the fundamental issue of how we can detach new (unconditional) obligations from conditional obligations, and it states a kind of “factual detachment” principle, allowing the deduction of the *actual obligations* of the agent, that is, the obligations which arise given the *actual facts* of the situation.

With (option2) we get the following results:

$$\vdash \text{O} \neg A \rightarrow \text{O}(B/A)$$

$$\text{(DD)} \quad \vdash \text{O} A \wedge \text{O}(B/A) \rightarrow \text{O} B$$

According to the first theorem everything is obligatory on the condition that some forbidden fact is the case: thus (option 2) clearly does not allow us to express CTDs. The second theorem represents a kind of “deontic detachment” principle, allowing the deduction of the *ideal obligations* of the agent, i.e. the further obligations which arise if he behaves in a way which conforms with some existing set of obligations.

The surface structures of lines (b) and (c) in the original Chisholm set might be taken to indicate that, within SDL, (option 2) should be chosen for (b), and (option 1) for (c), giving:

(a)  $\text{O help}$

(b)  $\text{O}(\text{help} \rightarrow \text{tell})$

(c)  $\neg \text{help} \rightarrow \text{O} \neg \text{tell}$

(d)  $\neg \text{help}$

This was Chisholm's choice, and he rightly went on to point out that this formalisation yields an inconsistency, since  $O$  tell is derivable from (a) and (b), whilst  $O \neg$  tell is derivable from (c) and (d) (and an inconsistency follows by the (D)-schema).

If, alternatively, we use (option 1) for both lines (c) and (b), then the resulting set is consistent, but logical independence is lost, since (b) will then be a consequence of (d). Likewise, if (option 2) were adopted for both (b) and (c), then (c) would be a consequence of (a) by the (RM)-rule. So, in SDL, the conclusion is that the Chisholm set cannot be represented in a way which satisfies both of the two minimum requirements of consistency and logical independence.

### 3.2 *Some further requirements on the representation of CTDs*

A number of deontic logicians have argued that the problems raised by CTDs involve in an essential way either a temporal dimension or actions. We shall have a good deal more to say about these lines of approach later on (especially in Section 7), but for the moment we just want to register agreement with Prakken and Sergot [Prakken and Sergot, 1994; Prakken and Sergot, 1996], who have indicated that there are examples of CTD scenarios where it is far from obvious how considerations of the temporal or action dimensions might be applicable. Consider:

EXAMPLE 2.

- (a) There ought to be no dog.
- (b) If there is no dog, there ought to be no warning sign.
- (c) If there is a dog, there ought to be a warning sign.
- (d) There is a dog.

EXAMPLE 3.

- (a) There must be no fence.
- (b) -
- (c) If there is a fence, then it must be a white fence.
- (d) There is a fence.

Examples of these kinds suggest that a treatment of CTD's which is tied to temporal or action aspects will not be sufficiently general in its scope.

A further question which existing treatments of CTD's raise is this: are lines (b) and (c) of the Chisholm set to be assigned fundamentally different

logical forms? The theory we develop below gives a negative answer to this question, and supplies a uniform treatment of deontic conditionals. Our view is that, in the absence of strong arguments to the contrary, the surface forms of (b) and (c) should be deemed to be merely stylistic variants of essentially the same type of underlying logical structure. In particular, we reject the position taken in [Prakken and Sergot, 1994; Prakken and Sergot, 1996], where it is argued that (b) and (c) should be given distinct logical representations just because (c), unlike (b), is a contrary-to-duty conditional, expressing as it does the obligation which comes into force when the obligation expressed by line (a) is violated. Prakken and Sergot's approach makes the assignment of logical form to deontic conditionals a highly context-dependent matter, with the consequence that any insertion or deletion of a norm may require that some revision then has to be made to the formalisation of some other norm in the set; (e.g., deleting line (a) of the Chisholm set would require, on their approach, a change in the formalisation of line (c)). Likewise, the form initially assigned to a given sentence might have to be revised in virtue of what turns out to be *derivable* from other sentences; suppose, for instance that "if  $A$  then it is obligatory that  $B$ " is in the initial set, and is assumed not to be a CTD; if it then transpires that "it is obligatory that not  $A$ " is derivable from other members of the initial set, then the conditional becomes a CTD and its logical form has to be changed accordingly. This change may, in turn, have further repercussions regarding what can be derived ... and so on. Now with a *small* initial set, such as Chisholm's, it will of course be relatively easy to see where changes need to be made; but with a large corpus of norms it is not difficult to imagine that the problem could become intractable. The disadvantages which accrue from this kind of context-dependence of logical form are so great, in our opinion, that any approach to the analysis of CTDs which manages to avoid it is - other things being equal - to be preferred.

Thus, we have so far identified the following requirements that an adequate formalisation of the Chisholm set should meet:

- (i) consistency;
- (ii) logical independence of the members;
- (iii) applicability to (at least apparently) timeless and actionless CTD-examples;
- (iv) analogous logical structures for the two conditional sentences, (b) and (c).

One important group of deontic logics that satisfy these requirements employs a primitive dyadic conditional obligation operator  $\text{O}(-/-)$ , where

$O(B/A)$  is read “it is obligatory that  $B$ , given that  $A$ ”. These logics usually take the unconditional obligation  $O B$  to be equivalent to  $O(B/\top)$ , and they represent the Chisholm set as follows:

- (a)  $O(\text{help} / \top)$
- (b)  $O(\text{tell} / \text{help})$
- (c)  $O(\neg\text{tell} / \neg\text{help})$
- (d)  $\neg\text{help}$

Following [Lower and Belzer, 1983] we can distinguish between two main “families” of dyadic deontic logics, according to the kind of detachment principles they support: one supports the “factual detachment” principle (FD), and we call it the “FD-family”;<sup>4</sup> the other supports the “deontic detachment” principle (DD), and we call it the “DD-family”.<sup>5</sup>

Returning again to the Chisholm set, it is clear that (as for its proposed representation within SDL) acceptance of *both* (FD) and (DD) would permit the derivation of  $O \neg\text{tell}$  (by (FD) on lines (c) and (d)) and  $O \text{tell}$  (by (DD) on lines (a) and (b)). If the (D)-schema is accepted, then the situation arising from adoption of both (FD) and (DD) would of course be one of *logical* inconsistency. But even if the (D)-schema is not accepted, so that the conjunction  $O \text{tell} \wedge O \neg\text{tell}$  is not deemed to be *logically* inconsistent, the derivation from the Chisholm set of a *conflict of obligations* of the type expressed by this conjunction is surely unacceptable from the intuitive point of view. The situation described by the Chisholm set does not present the agent concerned with a *moral dilemma*, on our view. Requirement (i), above, should be understood as one to the effect that a conjunction of the form  $O A \wedge O \neg A$  should not be derivable from the formal representation of the set, regardless of whether that conjunction is deemed *logically* inconsistent.

Of course, neither the FD-family nor the DD-family accepts *both* (FD) and (DD). Nevertheless, it might be suggested that a fully adequate representation of the Chisholm set should be able to capture, in a way which generates neither inconsistency nor a moral dilemma, *both* the fact that — given the circumstances, and particularly the occurrence of the violation of the obligation expressed by line (a) — the agent’s actual obligation is

<sup>4</sup>In general the logics in this family have a semantics based on minimal models (proposed, independently, by Dana Scott and Montague, and popularised by [Chellas, 1980]). As representatives of this family [Lower and Belzer, 1983] mention [Mott, 1973; al-Hibri, 1978; Chellas, 1974]; however, as regards [Chellas, 1974] it is not entirely clear whether Chellas commits himself to acceptance of (FD).

<sup>5</sup>[Lewis, 1974] presents an overview of several members of the DD-family. These logics introduce, in the semantics, a preference relation between the worlds, that orders the worlds according to their ideality; then  $O(B/A)$  is true at a world iff there is some world where  $A \wedge B$  is true and that is more ideal than any world where  $A \wedge \neg B$  is true. See also, below, Sections 7.1 and 7.3.



not to tell his neighbours he is coming, *and* the fact that — under ideal circumstances, in the absence of violation of the obligation expressed by line (a) — the agent’s obligation would be to help his neighbours and to tell them he is coming. Accepting these suggestions, we offer three further requirements which we believe an adequate representation of the Chisholm set should meet:

- (v) capacity to derive *actual* obligations;
- (vi) capacity to derive *ideal* obligations;<sup>6</sup>
- (vii) capacity to represent the fact that a *violation* of an obligation has occurred.

Neither the FD-family nor the DD-family meet both (v) and (vi).<sup>7</sup> We will return later to the issues raised by (vii).

### 3.3 The “pragmatic oddity”

One of the logics that fulfills all these requirements is the one proposed in [Jones and Pörn, 1985]. Jones and Pörn adopt a completely different approach from that of the dyadic deontic logics, and define a deontic logic where non-normal obligation operators are obtained as Boolean combinations of normal modal operators, following a strategy that had already been used in the field of action logic by Kanger and by Pörn.

Taking as its point of departure the observation that SDL fails in its attempt to capture CTDs because — from the semantical point of view - SDL considers only the *ideal* versions of each world, Jones and Pörn propose, in addition to SDL’s accessibility relation, a second accessibility relation which picks out the *sub-ideal* versions of a given world (and they further require that each world is either an ideal or a sub-ideal version of itself).

Then they introduce into the logical language two modal necessity operators,<sup>8</sup>Note that the notation employed here for the operators differs in most cases from that used in [Jones and Pörn, 1985]. here denoted by  $\boxed{\rightarrow}$  and  $\boxed{\rightarrow\rightarrow}$ . The first of these is just the obligation operator of SDL, so that an

<sup>6</sup>Some might call them *prima facie obligations*. However, we avoid using this term here since its meaning in the literature seems to us to be far from clear. Furthermore, [Prakken and Sergot, 1997] provide good reasons for supposing that the term is most at home in the discussion of defeasibility, rather than CTDs.

<sup>7</sup>In [Jones, 1993] it is argued that a further problem of the (FD)-family is that they reject the “principle of strengthening of the antecedent” (SA) whilst at the same time accepting unrestricted factual detachment. The problem is that one of the reasons for rejecting (SA) is that one wants to be able to represent conjunctions of the form  $\bigcirc (B/A) \wedge \bigcirc (\neg B/A \wedge C)$ , without getting logical inconsistency or moral dilemma of the form  $\bigcirc B \wedge \bigcirc \neg B$ , even in circumstances in which both  $A$  and  $C$  are true.

expression of the form  $\boxed{i_+}A$  is true at a given world  $w$  iff  $A$  is true at all of the ideal versions of  $w$ . By contrast,  $\boxed{s_+}A$  is true at a given world  $w$  iff  $A$  is true at all of the sub-ideal versions of  $w$ . (A sub-ideal version  $w_1$ , of  $w$ , is informally seen as a version of  $w$  in which at least one of the obligations in force at  $w$  is violated.) The duals of  $\boxed{i_+}$  and  $\boxed{s_+}$  are, respectively,  $\blacklozenge$  and  $\blacklozenge_{s_+}$ .

Finally they introduce both a deontic necessity operator  $\boxed{\rightarrow}$ , defined as follows:

$$\boxed{\rightarrow}A =_{\text{df}} \boxed{i_+}A \wedge \boxed{s_+}A$$

and an actual-obligation operator  $\text{Ought}$ , defined by:

$$\text{Ought } A =_{\text{df}} \boxed{i_+}A \wedge \blacklozenge_{s_+} \neg A$$

(the second conjunct guarantees that  $\vdash \neg \text{Ought } \top$ ).

The Chisholm set is then represented in [Jones and Pörn, 1985] as follows:

- (a)  $\text{Ought help}$
- (b)  $\boxed{\rightarrow}(\text{help} \rightarrow \text{Ought tell})$
- (c)  $\boxed{\rightarrow}(\neg \text{help} \rightarrow \text{Ought } \neg \text{tell})$
- (d)  $\neg \text{help}$

The set, on this representation, is consistent and its members are logically independent of each other. Lines (c) and (d) imply  $\text{Ought } \neg \text{tell}$  (note that  $\boxed{\rightarrow}$  is a “success” operator, i.e. it satisfies the (T)-schema), and lines (a) and (b) imply  $\boxed{i_+}\text{Ought tell}$ . Furthermore, the conjunction of (a) and (d) may be taken as expressing the fact that the unconditional obligation to help the neighbours has been violated; and, had it been the case that “(d’) help”, rather than (d), were true, then from (b) one could have deduced the actual obligation to tell,  $\text{Ought tell}$ . Apparently, all is well!

However, [Prakken and Sergot, 1994; Prakken and Sergot, 1996] point out that the Jones and Pörn treatment of Chisholm, in common with a number of others, generates what they call the “pragmatic oddity”: line (a), together with the derived actual obligation  $\text{Ought } \neg \text{tell}$ , require that, in all ideal versions of the given world, the agent concerned goes to help his neighbours but does not tell them he is coming — a result which appears highly counterintuitive.

Prakken and Sergot correctly point out that, *for a number of cases*, a reasonable temporal interpretation is available which enables the pragmatic oddity to be avoided. For instance, perhaps the obligation expressed in line (a) would ordinarily be understood as an obligation to go to help the

neighbours *no later than a particular time, t*. Then, if line (d) were to be true after time *t*, the accessible deontically ideal worlds would be characterised in such a way that, after time *t*, these worlds would require that the agent does not tell his neighbours he is coming (but they would not, of course, also require that he goes to help, since it would then be too late).

However, as we indicated above, Prakken and Sergot also point out that there are instances of the Chisholm set which may be interpreted in such a way that the temporal dimension is completely absent ([Jones, 1993, pp. 153-4] makes a similar point). Example 2, above, is one such case: a very ordinary way of understanding that set takes each sentence to be true at one and the same moment of time, and — without any insertion of *temporal* qualifications concerning *when* there ought to be no dog, or *when* there ought to be a warning sign — allows the conclusion to be drawn that there ought, in the circumstances, to be a warning sign, *without* thereby generating the pragmatic oddity, i.e., *without* forcing the further conclusion that, in all ideal versions of the given situation, there is no dog but there is a sign warning of one.

Unfortunately, Prakken and Sergot offer little by way of explanation of the pragmatic oddity: they say little about what it is that creates the sense of oddity. In [Carmo and Jones, 1997] we suggest an explanation which exploits a parallel between examples of type Example 2, which on the [Jones and Pörn, 1985] analysis exhibit the pragmatic oddity *simpliciter*, and examples like Example 3 above (also due to Prakken and Sergot) which, by virtue of some assumed logical truth, are inconsistent when formalised in the style of [Jones and Pörn, 1985] (according to which, in all ideal versions of the given world, there is a white fence and no fence at all!). The suggested parallel is as follows: as represented in the language of [Jones and Pörn, 1985], Example 2 exhibits the pragmatic oddity because an inconsistency *would* be generated were one to add to the example the further constraint that it ought not to be the case that there is both no dog and a sign warning of one. The sense of oddity arises because there is an interpretation of Example 2 according to which it remains consistent even if supplemented with that further constraint; and the problem with the [Jones and Pörn, 1985] approach is that it fails to capture *that* interpretation.

Thus we add another requirement which an adequate representation of CTDs should satisfy:

- (viii) capacity to avoid the pragmatic oddity (interpreted according to the previous diagnosis).

### 3.4 Two attempts to resolve the “pragmatic oddity”

In [Prakken and Sergot, 1994; Prakken and Sergot, 1996], Prakken and Sergot argue that the proper response to the problems raised by Example 2 —

and in particular the problem of pragmatic oddity — is to assign distinct logical forms to primary obligations, on the one hand, and CTD obligations, on the other. For CTD obligations, they relativise an obligation operator to a specific “context of violation”; more precisely, an expression of the form  $O_A B$  is intended to be read as “there is a secondary obligation that  $B$  given that, or presupposing, the sub-ideal context  $A$ ”, or “given that  $A$ , which is a violation of some primary obligation, there is a secondary, compromise obligation that  $B$ ” [Prakken and Sergot, 1996, section 5]. They emphasise that expressions of the form  $O_A B$  are not to be read as conditional primary obligations. “The expression  $O_A B \dots$  represents a particular kind of obligation. There is no meaningful sense  $\dots$  in which the obligation  $O B$  can be detached from the expression  $O_A B$ ” [*loc.cit.*]. Their representation of Example 2 takes the form:

- (a)  $O \neg \text{dog}$
- (b)  $\neg \text{dog} \rightarrow O \neg \text{sign}$
- (c)  $\text{dog} \rightarrow O_{\text{dog}} \text{sign}$
- (d)  $\text{dog}$

We shall not pursue the Prakken and Sergot 94 treatment of CTDs here (although we return to their work briefly in Section 7). Suffice it to say that their approach (in [Prakken and Sergot, 1994; Prakken and Sergot, 1996]) rejects the fourth of our requirements for a satisfactory theory in this area. For them, the choice of logical form for an apparently conditional deontic sentence will itself be dependent on which other norms are contained in, or derivable from, the set of norms being formalised.<sup>9</sup>

In [Carmo and Jones, 1995]<sup>10</sup> we attempted a different kind of approach to the problem of the pragmatic oddity, distinguishing between “ideal obligations” (line (a) in Examples 1, 2 and 3, for instance) and “actual obligations”, which indicate what is to be done given the (perhaps less-than-ideal) circumstances. The operator  $O_a$ , for representing actual obligations, was defined in the same way as the Ought-operator of [Jones and Pörn, 1985], described above. As regards ideal obligations, the basic model-theoretic

<sup>9</sup>There are also some difficulties in understanding how  $O_A B$  should be interpreted, particularly since Prakken and Sergot insist that  $A$  (in  $O_A B$ ) necessarily represents a context of violation. For instance, the formula  $(P A \wedge O_A B) \rightarrow O B$  is valid, on their account (where  $P$  is the permission operator), but not trivially so. As we see it, intuitively the conjunction in the antecedent of this conditional (given their reading of  $O_A B$ ) could only be false, so it should imply anything. Furthermore, what can they possibly mean by the claim that  $O B$  is an abbreviation of  $O_{\top} B$ ? Are we to suppose that it is obligatory that  $B$  only if the tautology represents a context of violation?

<sup>10</sup>We there adapt the logic proposed in [Carmo and Jones, 1994; Carmo and Jones, 1996] for the analysis of deontic integrity constraints.

idea was to distinguish between ideal *versions* of a given world (the fundamental feature of SDL), and *ideal worlds themselves*. Accordingly, we divided the set of possible worlds  $W$  into two mutually exclusive sub-sets, the set of ideal worlds and the set of sub-ideal worlds; importantly for our purposes, we allowed that a world  $w_1$  could be an ideal *version* of a given (*sub-ideal*) world  $w$  *without* also itself being an ideal *world*. And we fixed truth conditions for expressions of the form  $O_i B$  (“it ought ideally to be the case that  $B$ ”) in terms of the truth of  $B$  *in all ideal worlds* and falsity *in some sub-ideal world*.<sup>11</sup>

We represented Example 2 in the following way:

- (a)  $O_i \neg \text{dog}$
- (b)  $\Box (\neg \text{dog} \rightarrow O_a \neg \text{sign})$
- (c)  $\Box (\text{dog} \rightarrow O_a \text{sign})$
- (d)  $\text{dog}$

All the requirements (i)–(viii) are met by this analysis. In particular, the pragmatic oddity disappears because the conjunction  $O_i \neg \text{dog} \wedge O_a \text{sign}$ , which is clearly derivable from (a)–(d), does not imply that, in all ideal versions of the given world, there is no dog but a sign warning of one. What the conjunction *does* say, essentially, is that in all ideal *worlds* there is no dog, but in all ideal *versions* of the given (clearly sub-ideal) world there is a warning sign. The proposal worked well for this and a number of other examples, and we have defined a complete axiomatization for the logic.

However, as we now see things, this approach suffered from a defect similar to the one we have criticised in relation to [Prakken and Sergot, 1994; Prakken and Sergot, 1996]: the assignment of logical form for some of the norms in the set is dependent on the other norms in it. In [Prakken and Sergot, 1994; Prakken and Sergot, 1996] this was reflected in the use of different obligation operators for representing the deontic conditionals expressed by lines (b) and (c); in [Carmo and Jones, 1995] it is reflected in the use of different obligation operators for representing line (a) and lines (b) and (c). So, in order to capture the *general* issue motivating the adoption of adequacy requirement (iv), it should be reformulated as follows:

- (iv) the assignment of logical form to each of the norms in the set should be independent of the other norms in it.

---

<sup>11</sup>The second conjunct simply guarantees the violability of ideal obligations (i.e.  $\models \neg O_i \top$ ). [Carmo and Jones, 1995] contains discussion of possible connections between the notions of ideal/sub-ideal world and ideal/sub-ideal versions of a world, but we omit those details here.

A related observation is that problems appear within the [Carmo and Jones, 1995] approach if we add to the Chisholm set other norms that interact in some significant way with the norms in the original set. In particular, serious difficulties arise as soon as a “second-level” of CTDs is considered. Suppose, for instance, that lines (e) and (f), below, are added to Example 2:

- (e) If there is a dog and no warning sign, there ought to be a high fence.
- (f) There is no warning sign.

The [Carmo and Jones, 1995] representation of this extended set is:

- (a)  $O_i \neg \text{dog}$
- (b)  $\Box (\neg \text{dog} \rightarrow O_a \neg \text{sign})$
- (c)  $\Box (\text{dog} \rightarrow O_a \text{sign})$
- (d)  $\text{dog}$
- (e)  $\Box (\text{dog} \wedge \neg \text{sign} \rightarrow O_a \text{fence})$
- (f)  $\neg \text{sign}$

And the pragmatic oddity now re-appears, since the conjunction  $O_a \text{sign} \wedge O_a \text{fence}$  is derivable. So, in all ideal versions of the given world, there is a sign and a fence. (If this does not seem “odd”, imagine that the sign says “Beware of the unfenced dog”: it may well be forbidden to have both a sign of that kind and a fence. Thus the pragmatic oddity, in the sense of our proposed diagnosis, re-emerges.)

The problem of “further levels” of CTDs would force the [Carmo and Jones, 1995] approach to allow the possibility of an infinity of obligation operators: the need to associate (in some way or other) a context to each obligation operator seems to re-appear.

#### 4 CONTRARY-TO-DUTIES: A NEW APPROACH

On one very common interpretation of the set (a)–(f) above, the actual obligation which *applies* in the circumstances is the obligation to put up a fence, and it applies because the other two obligations (not to have a dog, and to put up a sign if there is a dog) have been violated. As we have emphasised above, it would be incorrect to say that the obligations not to have a dog, and to put up a sign if there is a dog, have been *defeated*, or *overturned*; they have been *violated*, and any proper representation of the situation must register the fact that, because of these violations, the obligation which becomes actual is the obligation to erect a fence. But how are these points to be articulated in a formal theory? To that question we now turn.

#### 4.1 Motivation

Consider again Example 2, particularly lines (a), (c) and (d). The norms governing, or in force in, the situation are that there ought to be no dog, and that if there is a dog there ought to be a warning sign; and the relevant fact is that there is a dog. So what is the actual obligation, of the agent concerned, in these circumstances? To erect a warning sign? But why not insist on getting rid of the dog, rather than on erecting a warning sign<sup>12</sup>? We wish to suggest that the answer to such questions turns on the *status* assigned to the fact that there is a dog — in the following sense: so long as there is a dog, but this, for one reason or another, is not deemed to be a *fixed, unalterable* feature of the situation, then the actual obligation which applies is that there ought to be no dog. However, as soon as, for one reason or another, the fact that there is a dog is deemed *fixed*, i.e., it is seen as a *necessary*, unavoidable feature of the situation, so that — in consequence — the practical possibility of satisfying the obligation that there ought to be no dog has to all intents and purposes been eliminated, then the actual obligation which applies is that there ought to be a warning sign.<sup>13</sup>

What do we mean when we say that *for some reason or another* a fact of the situation — in this case that there is a dog — may be deemed a fixed, necessary, unalterable feature of that situation? Well, there are various ways in which this “fixity” might arise; those who proposed temporal solutions to the problems associated with CTDs focussed on *one* of these ways. If books shall be returned by date due, then if you still have the books after the date due there is no way *that* obligation can be met. It is too late! It is unalterably the case that the books are not returned by the date due, and consequently the possibility of satisfying the obligation to return the books by date due has been eliminated.

But temporal reasons, although very common, are not the only reasons why things become fixed, in the sense of necessity or unalterability we here seek to explicate; for instance, it is not for temporal reasons that the deed of killing, once done, cannot be undone. What explains fixity in this case is not temporal necessity, but rather causal necessity. Nor need temporal considerations have any role to play in explaining why the presence of a dog may be, to all intents and purposes, an unalterable feature of the situation; it may, for some reason, be practically impossible in the situation to remove the dog; perhaps, for instance, its owner stubbornly refuses to remove it, and nobody else dares attempt the feat. The presence of the dog is a fixed fact: the dog remains unless the intervention of some agent leads to its

---

<sup>12</sup>Remember that, in keeping with our analysis of the pragmatic oddity, we seek an answer to these questions which is compatible with a further assumption to the effect that there ought not to be *both* no dog *and* a warning sign.

<sup>13</sup>Some remarks in a similar spirit are to be found in [Hansson, 1971, section XIII: “on the interpretation of circumstances”].

removal, and no agent is prepared to perform the required action. From the practical point of view — from the point of view of deciding which obligation actually applies to the situation — the key feature is that the possibility of satisfaction of the requirement that there be no dog is effectively eliminated.

As a further illustration, consider next the example of the “considerate assassin”.<sup>14</sup>

EXAMPLE 4.

- (a) You should not kill Mr. X.
- (b) -
- (c) But if you kill Mr. X, you should offer him a cigarette.

When does the assassin have an actual obligation to offer Mr. X a cigarette? After killing him? But then it is too late! One intuitively acceptable interpretation is that the assassin’s actual obligation to offer Mr. X a cigarette arises when he firmly decides that he is going to kill Mr. X. It is then that it becomes a settled or fixed fact that Mr. X will be killed, and then that the assassin’s actual obligation is to offer a cigarette.

Notice that the examples indicate that two different notions of necessity — and their associated notions of possibility — need to be considered. Mr. X, once killed, cannot be offered a cigarette because nobody has the ability or the opportunity to make offers to the dead, just as nobody has the ability or opportunity to return a book by date due if the date due has passed. On the other hand, the dog-owner may have both the ability and the opportunity to remove the dog, and the assassin may have both the ability and the opportunity to refrain from killing Mr. X; but once each has made a firm and definite decision (to keep the dog and to kill Mr. X, respectively), then to all intents and purposes the persistent presence of the dog and the future performance of the assassination become fixed features of the respective situations; so questions about which actual obligations arise in these situations have to be answered in the light of the fact that alternatives in which there is no dog, or no assassination, are not actually available.

Now it may well be that the judge, at the assassin’s trial, insists that the assassin should never have decided to commit the murder, just as it may be that the manager of the housing estate refuses to accept that the dog owner was entitled to decide that he would keep his dog. Furthermore, it is a well known feature of, for instance, disputes in legal cases, that the parties to the dispute may disagree about what an agent was able to do, or

<sup>14</sup>This example can also be found in [Prakken and Sergot, 1996] (using “the witness” instead of “Mr. X”). In fact [Prakken and Sergot, 1996] provides an excellent survey of the principal examples of CTDs, and we use them in Section 5 to test the adequacy of our proposal.



what he had the opportunity to do. But the existence of disagreements of these kinds is perfectly compatible with the approach to CTD scenarios we develop below. For it will not be the task of our logical system to determine the reasons which justify the classification of some fact as settled. Rather, what the system will do is this: first, it will specify the role of assumptions about two types of fixity in reasoning about actual and ideal obligations; and, second, it will show which actual/ideal obligations can be derived from a given set of norms *when some facts are taken to be fixed in the one sense or the other*.<sup>15</sup>

#### 4.2 *The new theory and its fundamental semantic features*

We now present the basic features of a modal-logical language designed to capture the approach to CTDs described above in a way that conforms to the constraints, or requirements, (i)–(viii). We shall then show how the new language may be applied to the Chisholm set, and to the analysis of some other problematic CTD “scenarios”.

We adopt the following approach to the formal representation of these scenarios: their *deontic component* (the obligation norms which they explicitly contain) will be represented throughout in terms of a dyadic, conditional obligation operator  $O(\_/\_)$ ; their *factual component* will be represented by means of either unmodalised sentences, or modalised sentences in two categories. These two categories correspond to the two notions of necessity (and their associated dual notions of possibility) which we shall employ to articulate the ideas regarding fixity, or unalterability, of facts alluded to in the previous subsection.

From the deontic and factual components taken together, some further obligation sentences may be derivable. The *derived obligation sentences* are of two types, pertaining to *actual* obligations and *ideal* obligations, respectively. There is an intimate conceptual connection between these two notions of derived obligation, on the one hand, and the two notions of necessity/possibility used in characterising the factual component, on the other.

Consider first the dyadic conditional obligation operator. How do we wish to interpret a sentence of the kind “if there is a dog then there shall be a warning sign”? On our view, this sentence is to be understood as saying that in any context in which the presence of a dog is a fixed, unalterable fact, it is obligatory to have a warning sign, if this is possible. We think of a *context* as a set of worlds — the set of relevant worlds for the situation at hand. So the above sentence is to be understood as saying that, for any context in which there is a dog (i.e., for any context in which there is a dog in each world of that context), if it is possible to have a warning sign then

---

<sup>15</sup>Our thanks to Layman Allen for raising a question at the Sesimbra  $\Delta$ EON96 Workshop related to this point.

it is obligatory to have a warning sign.<sup>16</sup> In order to capture this idea we introduce in our models a function  $ob:\wp(W) \rightarrow \wp(\wp(W))$  which picks out, for each context, the propositions which represent that which is obligatory in that context. That is,  $\|B\| \in ob(X)$  (where  $\|B\|$  denotes the truth set of  $B$  in the model in question) if and only if the proposition expressed by  $B$  represents something obligatory in context  $X$ . Accordingly, we say that a sentence  $O(B/A)$  is true in a model if and only if, in any context  $X$  where  $A$  is true and  $B$  is possible (i.e. in any context having  $A$  true in each of its worlds and  $B$  true in at least one of its worlds), it is obligatory that  $B$ .<sup>17</sup>

On the basis of this operator we could now derive the obligations that were applicable in each context, *assuming* that our language contained a means of representing contexts. *The question is:* what are the types of contexts that we need to be able to talk about in our formal language, given that we want to be able to derive sentences of two kinds, describing *actual* obligations and *ideal* obligations, respectively? We answer this question in terms of the two notions of necessity.

The first of these will be denoted by  $\boxed{\rightarrow}$ , and its dual possibility notion denoted by  $\diamond$ . Intuitively,  $\boxed{\rightarrow}$  is intended to capture that which — in a particular situation — is *actually* fixed, or unalterable, given (among other factors) what the agents concerned have decided to do and not to do. In, for instance, the “dog scenario” (Example 2), if the agents concerned have firmly decided that the dog is not going to be removed, then the sentence  $\boxed{\rightarrow}$ dog is true of that situation (the presence of the dog is *actually* an unalterable fact). On the other hand, if some actual possibility existed for getting rid of the dog, then the situation would be appropriately described by  $\diamond\neg$ dog (i.e.  $\neg\boxed{\rightarrow}$ dog). Which *actual obligations* arise in the dog scenario will depend, in particular, on whether or not  $\diamond\neg$ dog is true.

We must emphasise one important difference between this notion of necessity/possibility and the second one we employ. For the reasons discussed in the previous subsection, we do not exclude, *a priori*, that a sentence of the form  $\boxed{\rightarrow}A$  might be true even though the agents concerned have the ability and the opportunity to see to it that  $\neg A$ . That is, we shall want to consider scenarios where, despite their abilities and opportunities for action, the agents have firmly resolved not to see to it that  $\neg A$ , and where — given that this is what the agents have decided — there is (for all intents and purposes) no way that  $A$  could be false.

<sup>16</sup>We are here using the term “obligatory” in a *weak sense*; in a *strict sense*, for a sentence  $B$  to be obligatory in a context  $X$  we would also claim that there must exist at least one world in  $X$  where  $B$  is false (i.e., we would insist, for the strict sense, that obligations must be violable). However, our actual and ideal obligations, to be defined next, will be considered in this *strict sense*.

<sup>17</sup>We also add the further requirement that the conjunction of  $A$  and  $B$  is not contradictory, in order to avoid some “absurd” vacuous conditional obligations, and as one of the conditions needed to secure the result that if  $O(B/A)$  is true then  $\|B\| \in ob(\|A\|)$ .

In order to capture the semantics of the necessity operator  $\boxed{\rightarrow}$ , our semantical models will contain a function,  $av$ , which picks out (for any given world  $w$ ) a set of worlds  $av(w)$  — the set of worlds which are the *actual versions* of  $w$  (the *open alternatives* for the current world  $w$ ), those which constitute the context that it is actually relevant to take into account in determining which obligations are actually in force, or actually apply, at  $w$ . Accordingly, a sentence of the form  $\boxed{\rightarrow}A$  will be said to be true at a given world  $w$  if and only if  $A$  is true at *all* of the worlds contained in  $av(w)$ .

Given the way the function  $ob$  is understood, the set of propositions  $ob(av(w))$  will be the set of propositions which represent that which is obligatory in the context  $av(w)$  (that is to say, in the context of the alternatives that are actually open at  $w$ ). In line with this, we shall say that a sentence of the form  $O_a A$  (read as “it is actually obligatory that  $A$ ”, or “it actually ought to be the case that  $A$ ”) is true at a world  $w$  only if the proposition expressed by  $A$  is one of those propositions picked out by  $ob$  for the argument  $av(w)$ . In addition, the truth of  $O_a A$  at  $w$  will require that there is at least one world in  $av(w)$  where the sentence  $A$  is false; the reason for this second requirement is that that which is actually obligatory might actually fail to obtain.

The second of the two notions of necessity will be denoted by  $\boxed{\square}$ , and its dual possibility notion denoted by  $\diamond$ . Intuitively,  $\boxed{\square}$  is intended to capture that which — in a particular situation — is not only actually fixed, but would still be fixed even if different decisions had been made, by the agents concerned, regarding how they were going to behave. For instance, certain features of the situation will be such that it is beyond the power of the agents to change them — they may lack the ability, or the opportunity, or both. Of such features it is appropriate to say that they are fixed in the sense that they *could not have been* avoided by the agents concerned, no matter what they had done. It is not even *potentially* possible for the agents to alter them. In the original Chisholm scenario, for example, if the bridge that leads to the man’s neighbours’ house has been destroyed by a storm, and the man is unable to repair it, then clearly it is a necessary feature of the situation, in this second sense of necessity, that he does not help his neighbours. This is to be understood in contrast to the situation in which it is potentially possible for the agent to go to his neighbours’ house to help them, but the agent has made a definite decision, from which he will not budge, not to go. His will is firm, and thus in all actual relevant alternatives open to the agent he does not go to help them (and so actually he ought not to tell them he is coming). But given that he has the ability and opportunity to go, it is *potentially* possible — in the sense expressed by the  $\diamond$  operator — that he does so<sup>18</sup>, and so we want then be able to derive

---

<sup>18</sup>So the best short readings for these two pairs of operators we can offer are the following:

that his *ideal* obligation was to go and to tell them he was coming.

To articulate the semantics of the second pair of notions of necessity and possibility, we introduce into the models a function,  $pv$ , which picks out (for any given world  $w$ ) a set of worlds  $pv(w)$  — the set of worlds which are the *potential versions* of  $w$ . These worlds will constitute the context it is relevant to take into account in determining which *ideal* obligations hold at  $w$  (“*what should have been done*”). A sentence of the form  $\Box A$  will be said to be true at a given world  $w$  if and only if  $A$  is true at all of the worlds contained in  $pv(w)$ . Furthermore, given the way the function  $ob$  is understood, the set of propositions  $ob(pv(w))$  will be the set of propositions which represent that which is obligatory in the context  $pv(w)$ . Thus we shall say that a sentence of the form  $O_i A$  (read as “it is ideally obligatory that  $A$ ”, or “it ideally ought to be the case that  $A$ ”) is true at a world  $w$  only if the proposition expressed by  $A$  is one of those propositions picked out by  $ob$  for the argument  $pv(w)$ . In addition, the truth of  $O_i A$  at  $w$  will require that there is at least one world in  $pv(w)$  where the sentence  $A$  is false — since that which is ideally obligatory might potentially fail to obtain.

Finally, we define the notion of violation in terms of the notion of ideal obligation, as follows:

$$\text{viol}(A) =_{\text{df}} O_i A \wedge \neg A$$

This choice is in accordance with the intuitive idea that ideal obligations express what should have been done, and fits in well with our treatment of the pragmatic oddity and other features of CTD scenarios, as will become clearer when we analyse a number of examples in some detail. Briefly, the main points may already be explained as follows: in, for instance, the dog scenario, if it is a fixed fact that there is a dog (i.e., if  $\Box \text{dog}$  is true), but it is *actually* possible that a sign may be erected and *potentially* possible that there is no dog, then we shall be able to derive that it is actually obligatory that a sign is erected and ideally obligatory that there is no dog. The pragmatic oddity will be avoided because it will not, in *these* circumstances, be possible to derive an actual obligation that there be no dog. Nevertheless, we still of course want to say that an obligation (that there be no dog) has been violated, and this result is secured if violation is characterised as above. As the formal analysis of this example will show,

- 
- $\Diamond A$  : it is actually possible that  $A$
  - $\blacklozenge A$  : it is potentially possible that  $A$
  - $\Box A$  : it is not actually possible that  $\neg A$
  - $\blacklozenge A$  : it is not potentially possible that  $\neg A$

(In a number of cases, the natural reading of statements about potential possibility will be in the past tense.)

we shall also be able to derive a second violation in this situation, if no sign has been erected.

The semantic models described next will be subject to various constraints, designed to achieve a particular pattern of relationships between the dyadic obligation operator, the two types of necessity/possibility operators, and the operators for actual and ideal obligations.

### 4.3 Syntax and semantics of the formal language

#### Syntax

*Alphabet:*

- a set of (natural language) terms (dog, fence, sign, ...) for atomic sentences
- $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$  (sentential connectives)
- $(, )$  (parentheses)
- $\boxed{\rightarrow}$  (dual:  $\diamond \stackrel{\text{df}}{=} \neg \boxed{\rightarrow} \neg$ )
- $\boxed{\square}$  (dual:  $\diamond \stackrel{\text{df}}{=} \neg \boxed{\square} \neg$ )
- $O (/)$  (dyadic deontic operator)
- $O_a$  (monadic deontic operator - for actual obligation)
- $O_i$  (monadic deontic operator - for ideal obligation)

*Rules for construction of well-formed sentences:* as usual

- $\text{viol}(A) \stackrel{\text{df}}{=} O_i A \wedge \neg A$

#### Semantics

*Models:*

$\mathcal{M} = \langle W, av, pv, ob, V \rangle$ , where:

- 1)  $W \neq \emptyset$
- 2)  $V$  - a function assigning a truth set to each atomic sentence
- 3)  $av : W \rightarrow \wp(W)$   
(alternatively:  $Ra \subseteq W \times W$  and  $av(w) = \{w_1 : wR_a w_1\}$ )  
such that:

- 3-a)  $av(w) \neq \emptyset$
- 4)  $pv : W \rightarrow \wp(W)$   
 (alternatively:  $Rp \subseteq W \times W$  and  $pv(w) = \{w_1 : wR_p w_1\}$ )  
 such that:
- 4-a)  $av(w) \subseteq pv(w)$
- 4-b)  $w \in pv(w)$
- 5)  $ob : \wp(W) \rightarrow \wp(\wp(W))$   
 such that (where  $X, Y, Z$  designate arbitrary sets of members of  $W$ ):
- 5-a)  $\emptyset \notin ob(X)$
- 5-b) if  $Y \cap X = Z \cap X$ , then  $(Y \in ob(X) \text{ iff } Z \in ob(X))$
- 5-c) if  $Y, Z \in ob(X)$ , then  $Y \cap Z \in ob(X)$
- 5-d) if  $Y \subseteq X$  and  $Y \in ob(X)$  and  $X \subseteq Z$ , then  $((Z - X) \cup Y) \in ob(Z)$

*Truth in a world  $w$  in a model  $\mathcal{M} = \langle W, av, pv, ob, V \rangle$  is characterised as follows (where  $\|A\| = \|A\|^{\mathcal{M}} = \{w \in W : \mathcal{M} \models_w A\}$ ):*

$$\mathcal{M} \models_w p \quad \text{iff} \quad w \in V(p)$$

... (the usual truth conditions for the connectives  $\neg, \wedge, \vee, \rightarrow$  and  $\leftrightarrow$ )

$$\mathcal{M} \models_w \boxed{\rightarrow} A \quad \text{iff} \quad av(w) \subseteq \|A\|$$

$$\mathcal{M} \models_w \boxed{\square} A \quad \text{iff} \quad pv(w) \subseteq \|A\|$$

$$\mathcal{M} \models_w \bigcirc (B/A)^{19} \quad \text{iff}$$

$$\|A\| \cap \|B\| \neq \emptyset \text{ and}$$

$$(\forall X)(\text{if } X \subseteq \|A\| \text{ and } X \cap \|B\| \neq \emptyset, \text{ then } \|B\| \in ob(X))$$

$$\mathcal{M} \models_w \bigcirc_a A \quad \text{iff} \quad \|A\| \in ob(av(w)) \text{ and } av(w) \cap \|\neg A\| \neq \emptyset$$

$$\text{(i.e. iff } \|A\| \in ob(av(w)) \text{ and } av(w) \cap (W - \|A\|) \neq \emptyset)$$

$$\mathcal{M} \models_w \bigcirc_i A \quad \text{iff} \quad \|A\| \in ob(pv(w)) \text{ and } pv(w) \cap \|\neg A\| \neq \emptyset$$

<sup>19</sup>An alternative would be to define the dyadic obligation operator in the *strict sense* referred to in footnote 16, in which case  $\mathcal{M} \models_w \bigcirc (B/A)$  iff  $\|A\| \cap \|B\| \neq \emptyset$  and  $\|A\| \cap \|\neg B\| \neq \emptyset$  and  $(\forall X)(\text{if } X \subseteq \|A\| \text{ and } X \cap \|B\| \neq \emptyset \text{ and } X \cap \|\neg B\| \neq \emptyset, \text{ then } \|B\| \in ob(X))$ . In that case we would require that if  $Y \in ob(X)$  then  $X \cap (W - Y) \neq \emptyset$ ; and the truth in a world  $w$  of  $\bigcirc_a A$  (respectively  $\bigcirc_i A$ ) would be defined as follows:  $\mathcal{M} \models_w \bigcirc_a A$  iff  $\|A\| \in ob(av(w))$  (resp.  $\mathcal{M} \models_w \bigcirc_i A$  iff  $\|A\| \in ob(pv(w))$ ). With both approaches we get exactly the same semantics for  $\bigcirc_a A$  and  $\bigcirc_i A$ .

(Note that the definition of  $\mathcal{M} \models_w O(B/A)$  entails that if  $\mathcal{M} \models_w O(B/A)$ , then  $\|B\| \in ob(\|A\|)$ .)

A sentence  $A$  is said to be *true in a model*  $\mathcal{M} = \langle W, av, pv, ob, V \rangle$ , written  $\mathcal{M} \models A$ , iff  $\|A\|^{\mathcal{M}} = W$ ; and  $A$  is said to be *valid*, written  $\models A$ , iff  $\mathcal{M} \models A$  in all models  $\mathcal{M}$ .

Some comments about the conditions:

- i) As would be expected, the set  $av(w)$  is required to be a subset of  $pv(w)$ , for any  $w$  (condition 4-a), so that actual possibility entails potential possibility. Conditions 3-a) and 4-b) are also obvious.

In [Carmo and Jones, 1997]<sup>20</sup> we required that  $w \in av(w)$  (which implies 3-a) and, together with 4-a), also implies 4-b)). Although in most scenarios it makes sense to say that the actual world is always an actual alternative to itself, we sometimes need to be able to describe situations where  $A$  is not yet the case, but nevertheless in all relevant future alternatives open to the agent,  $A$  is the case. Consider, for instance, the scenario of the “considerate assassin” (Example 4): there may be situations where, although the assassin has not yet killed Mr. X, in all the relevant future alternatives open to the assassin Mr. X is going to be killed by him (because the assassin has so decided); the natural way to represent this situation in our logic is:  $\neg\text{kill} \wedge \boxed{\rightarrow}\text{kill}$ . But of course this can only be expressed consistently if we do not require that, for all  $w, w \in av(w)$ .

There exist other conditions that it may seem natural to impose on  $av$  and  $pv$ , such as the transitivity of both the actual and potential relations. But, for simplicity, we consider here only those conditions which appear to have a direct bearing on the analysis of the key examples of CTD scenarios in Section 5.

- ii) Condition 5-a) means that we do not accept that a contradiction might be obligatory.
- iii) Condition 5-b) means that if, from the point of view of a context  $X$ , two propositions  $Y$  and  $Z$  are indistinguishable, then one of them is obligatory iff the other is (this corresponds to a kind of “contextual” RE-rule).

It is also appropriate at this point to state some of the main consequences of condition 5-b), and to attach numbered labels to them, in order to facilitate later discussion of a possible weakening of 5-b).

---

<sup>20</sup>Where we use  $va$ ,  $vp$  and  $pi$ , instead of the more suggestive names,  $av$ ,  $pv$  and  $ob$ , used in this chapter.

Since  $Y \cap X = (Y \cap X) \cap X$ , we get as particular cases of 5-b):

5-b1) if  $Y \in ob(X)$ , then  $Y \cap X \in ob(X)$

5-b2) if  $Y \cap X \in ob(X)$ , then  $Y \in ob(X)$

On the other hand, using 5-a) and 5-b1) we get the condition (which in turn implies 5-a)):

5-ab) if  $Y \in ob(X)$ , then  $Y \cap X \neq \emptyset$

- iv) Condition 5-c) requires that the conjunction of two obligatory propositions within a context  $X$  is also obligatory in that context. A natural extension of condition 5-c) would be to require the closure of  $ob$  under arbitrary intersections (and not only under finite intersections):

5-c+) if  $\beta \subseteq ob(X)$  and  $\beta \neq \emptyset$ , then  $(\bigcap \beta) \in ob(X)$

(where  $\beta$  is any set of subsets of  $W$ ,  $(\bigcap \beta)$  is defined as:

$$(\bigcap \beta) = \{w \in W : (\forall X \in \beta)(w \in X)\}$$

If we impose this stronger condition then we would get  $(\bigcap ob(X)) \in ob(X)$ , if  $ob(X)$  is non-empty; note also that, by 5-b1), if  $ob(X) \neq \emptyset$  then  $(\bigcap ob(X)) \subseteq X$ ; if  $ob(X) = \emptyset$  then, by definition,  $(\bigcap ob(X)) = W$ . However, for reasons to be explained, what we shall in fact propose later is a *weakening* of condition 5-c).

- v) Condition 5-d) states that if a subset  $Y$  of  $X$  is an obligatory proposition in a context  $X$ , then in a bigger context  $Z$  it is obligatory to be either in  $Y$  or else in that part of  $Z$  which is not in  $X$ .

Taking into account condition 5-b), it may be shown that each of the following conditions is equivalent to 5-d) - they can sometimes be used to simplify proofs:

5-bd1) if  $Y \subseteq X$  and  $Y \in ob(X)$  and  $X \subseteq Z$ , then  $((W - X) \cup Y) \in ob(Z)$

5-bd2) if  $Y \in ob(X)$  and  $X \subseteq Z$ , then  $((Z - X) \cup Y) \in ob(Z)$

5-bd3) if  $Y \in ob(X)$  and  $X \subseteq Z$ , then  $((W - X) \cup Y) \in ob(Z)$

5-bd4) if  $Y \in ob(X)$  and  $X \subseteq Z$ , then  $((W - X) \cup (X \cap Y)) \in ob(Z)$

Using conditions 5-b), 5-c) and 5-d), we can also prove that

$$\text{if } Z \in ob(X) \text{ and } Z \in ob(Y), \text{ then } Z \in ob(X \cup Y).$$

#### 4.4 *Syntactic/axiomatic characterisation of the modal operators*

In what follows we introduce the axioms and rules for the various modal operators. For some of the axioms and theorems we introduce special labels



— if there is no standard label — in order to facilitate reference to them later on.

*Characterisation of  $\Box$ :*

1.  $\Box$  is a normal modal operator of type KT.

*Characterisation of  $\Box\rightarrow$ :*

2.  $\Box\rightarrow$  is a normal modal operator of type KD.

*Relationship between  $\Box$  and  $\Box\rightarrow$ :*

3.  $\Box A \rightarrow \Box\rightarrow A$  (axiom schema ( $\Box \rightarrow \Box\rightarrow$ ))

*Characterisation of  $O$ :*

4.  $\neg O(\perp/A)$  (the schema  $\neg N$  for  $O : (O - \neg N)$ )
5.  $O(B/A) \wedge O(C/A) \rightarrow O(B \wedge C/A)$   
(the schema  $C$  for  $O : (O - C)$ )
6. Restricted principle of strengthening of the antecedent - 1:  
 $O(B/A) \rightarrow O(B/A \wedge B)$  (SA1)
7. the *RE-rule* with respect to (w.r.t.) the *antecedent*:  
if  $\vdash (A \leftrightarrow B)$  then  $\vdash O(C/A) \leftrightarrow O(C/B)$
8. the “*contextual RE-rule*” w.r.t. the *consequent*:  
if  $\vdash C \rightarrow (A \leftrightarrow B)$  then  $\vdash O(A/C) \leftrightarrow O(B/C)$

*Relationship between  $O$  and  $\Box$ :*

9.  $\Diamond O(B/A) \rightarrow \Box O(B/A)$  ( $\Diamond O \rightarrow \Box O$ )
10. Restricted principle of strengthening of the antecedent - 2:  
 $\Diamond(A \wedge B \wedge C) \wedge O(C/B) \rightarrow O(C/A \wedge B)$  (SA2)

*Characterisation of  $O_a / O_i$ :*

11.  $O_a A \wedge O_a B \rightarrow O_a (A \wedge B)$  ( $O_a - C$ )  
 $O_i A \wedge O_i B \rightarrow O_i (A \wedge B)$  ( $O_i - C$ )

*Relationships between  $O_a$  (respectively:  $O_i$ ) and  $\Box\rightarrow$  (resp.:  $\Box$ ):*

12.  $\boxed{\rightarrow} A \rightarrow (\neg O_a A \wedge \neg O_a \neg A) \quad (\neg O_a)$   
 $\boxed{\square} A \rightarrow (\neg O_i A \wedge \neg O_i \neg A) \quad (\neg O_i)$
13.  $\boxed{\rightarrow} (A \leftrightarrow B) \rightarrow (O_a A \leftrightarrow O_a B) \quad (\leftrightarrow O_a)$   
 $\boxed{\square} (A \leftrightarrow B) \rightarrow (O_i A \leftrightarrow O_i B) \quad (\leftrightarrow O_i)$

*Relationships between  $O, O_a$  (resp.:  $O_i$ ) and  $\boxed{\rightarrow}$  (resp.:  $\boxed{\square}$ ):*

14. Restricted factual detachment:  
 $O(B/A) \wedge \boxed{\rightarrow} A \wedge \diamond B \wedge \diamond \neg B \rightarrow O_a B \quad (O_a - FD)$   
 $O(B/A) \wedge \boxed{\square} A \wedge \diamond B \wedge \diamond \neg B \rightarrow O_i B \quad (O_i - FD)$
15.  $O(B/A) \wedge \diamond(A \wedge B) \wedge \diamond(A \wedge \neg B) \rightarrow O_a(A \rightarrow B) \quad (O \rightarrow O_a \rightarrow)$   
 $O(B/A) \wedge \diamond(A \wedge B) \wedge \diamond(A \wedge \neg B) \rightarrow O_i(A \rightarrow B) \quad (O \rightarrow O_i \rightarrow)$

Notes:

- i) Simply using Propositional Calculus, we can deduce from 3. ( $\boxed{\square} \rightarrow \boxed{\rightarrow}$ ) the following useful theorem:

$$\vdash \diamond \diamond A \rightarrow \diamond A \quad (\diamond \rightarrow \diamond)$$

- ii) From the contextual RE-rule (8.) it follows that if  $\vdash A \leftrightarrow B$  then  $\vdash O(A/C) \leftrightarrow O(B/C)$ ; thus  $O$  is classical w.r.t. to both of its arguments.
- iii) Some comments are in order regarding axiom 9. It reflects the fact that norms which comprise the deontic component of a CTD scenario are themselves taken to be fixed, in the sense that they are features of the scenario which the agents concerned have neither the ability nor the opportunity to change. We maintain that this is a reasonable assumption to make, given that our concern is *not* with the dynamics of normative systems, but with the determination of which ideal and actual obligations may be derived from a *fixed* set of norms, given the facts of the case.

But it might be asked whether our analysis of the given norms (i.e., of the dyadic deontic conditionals) is *compatible* with the obvious fact that norms are man-made and may be subject to change. Our answer is that there is no incompatibility here: were we to switch our interest from normative statics to normative dynamics, one natural move — from the semantic point of view — would be to treat the *models* of our current semantics as the *worlds* of a semantical framework for the investigation of normative dynamics.

- iv) Axiom 10. embodies only a weak restriction on strengthening of the antecedent and it is clear that our dyadic deontic conditionals are not “exception-allowing” default conditionals. Any attempt to re-define them as default conditionals — and thus to eliminate 10. from the class of theorems — would have to be accompanied by elimination of the factual detachment principles expressed in 14. (cf. footnote 7, above). The factual detachment principles would then need to be replaced by a theory of default reasoning, defining the conditions under which sentences about actual and ideal obligations could be drawn as default conclusions from any given deontic and factual scenario. In our view, changes of *this* kind in the underlying logic of deontic conditionals would have no direct bearing on how the CTD problems themselves are to be handled.
- v) The counterpart to axiom 12. (first part) also held in the system DL of [Jones and Pörn, 1985], where their modality Ought represented actual obligation. The point captured by this axiom is as follows: if it is not actually possible that  $A$  is false, then  $A$  is not actually obligatory (for if it is guaranteed that  $A$  is true, any such obligation has been discharged); in addition,  $\neg A$  is not actually obligatory (for that which it is not actually possible to realise cannot be actually obligatory). However, the fact that it is not *actually* possible that  $\neg A$  does not rule out the possibility that, *ideally*, it ought not to be the case that  $A$ .

- vi) Using the axiom schemas  $(\leftrightarrow O_a)$  and  $(\leftrightarrow O_i)$ , we can prove the *RE-rules*:

if  $\vdash (A \leftrightarrow B)$  then  $\vdash O_a A \leftrightarrow O_a B$

and

if  $\vdash (A \leftrightarrow B)$  then  $\vdash O_i A \leftrightarrow O_i B$

Thus both  $O_a$  and  $O_i$  are classical operators.

- vii) Using the schemas  $(\neg O_a)$  and  $(\neg O_i)$ , we can deduce the following theorems:

$$\begin{array}{llll} \vdash \neg O_a \top & (O_a - \neg N) & \vdash \neg O_i \top & (O_i - \neg N) \\ \vdash \neg O_a \perp & (O_a - OD) & \vdash \neg O_i \perp & (O_i - OD) \end{array}$$

And any classical operator with  $(OD)$  and  $(C)$  also has the schema  $(D)$  as a theorem; a revision of the logic which avoids  $(D)$  will be discussed in Section 6.

It is important that neither  $O_a$  nor  $O_i$  validates the schema  $(M)$  — the converse of  $(C)$  — otherwise we would get the  $(RM)$ -rule, to be discussed further in Section 6.1.

- viii) We write just  $(FD)$ , instead of  $(O_a - FD)$  or  $(O_i - FD)$ , whenever it is clear from the context to which of the two obligation operators we are referring.

**RESULT 1.**

The previous axiomatisation is *sound*.<sup>21</sup>

Hint to the proof:

- 1.-4. Besides the relevant truth conditions, simply use (respectively) the condition: 4-b), 3-a), 4-a), and 5-a) (for this last case recall that  $\mathcal{M} \models_w O(B/A)$  implies  $\|B\| \in ob(\|A\|)$ ).
5. Use 5-c) and 5-ab) (plus the relevant truth condition); condition 5-ab) is needed only to prove that  $\|A \wedge B \wedge C\| \neq \emptyset$ .
6. Trivial (simply use the relevant truth condition).
- 7.-8. It is trivial to prove that these rules preserve truth in a model (and so also preserve validity, as desired). For the contextual RE-rule use condition 5-b).
- 9.-10. Simply use the relevant truth condition.  
(w.r.t. 10., note that  $\mathcal{M} \models_w \Diamond(A \wedge B \wedge C)$  implies  $\|A \wedge B \wedge C\| \neq \emptyset$ .)
- 11.-13. Besides the relevant truth conditions, use (respectively): 5-c), 5-ab), and 5-b).
14. Simply use the relevant truth conditions.
15. Using conditions 5-b) and 5-d), prove that if  $\mathcal{M} \models_w O(B/A)$  and  $Z \cap \|A\| \cap \|B\| \neq \emptyset$ , then  $((W - \|A\|) \cup \|B\|) \in ob(Z)$ ; and then apply that result with  $Z = av(w)$  (resp.  $pv(w)$ ).

End-proof

The next result provides a list of some other useful theorems and rules concerning the obligation operators.

**RESULT 2.**

- i)  $\vdash O(B/A) \leftrightarrow O(A \wedge B/A)$
- ii)  $\vdash \neg O(A/\perp) \quad \vdash \neg O(\neg A/A)$
- iii)  $\vdash \Diamond \neg O(B/A) \rightarrow \Box \neg O(B/A) \quad (\Diamond \neg O \rightarrow \Box \neg O)$
- iv)  $\vdash \Diamond(A \wedge B) \wedge O(A \rightarrow B/\top) \rightarrow O(B/A)$
- v) if  $\vdash A \rightarrow B$  then  $\vdash \Diamond(A \wedge C) \wedge O(C/B) \rightarrow O(C/A)$

<sup>21</sup>Some completeness results are also available, but are not reproduced in this chapter.

- vi)  $\vdash \boxed{\rightarrow} A \rightarrow (O_a B \rightarrow O_a (A \wedge B)) \quad (O_a \rightarrow O_a \wedge)$   
 $\vdash \boxed{\square} A \rightarrow (O_i B \rightarrow O_i (A \wedge B)) \quad (O_i \rightarrow O_i \wedge)$
- vii) Restricted deontic detachment:<sup>22</sup>  
 $\vdash O_a A \wedge O (B/A) \wedge \diamond(A \wedge B) \rightarrow O_a (A \wedge B) \quad (O_a - DD)$   
 $\vdash O_i A \wedge O (B/A) \wedge \diamond(A \wedge B) \rightarrow O_i (A \wedge B) \quad (O_i - DD)$

Hint to the proof:

- i) From the contextual RE-rule, since  $\vdash A \rightarrow (B \leftrightarrow A \wedge B)$ .
- ii) From the contextual RE-rule and  $(O - \neg N)$  (since  $\vdash \perp \rightarrow (\perp \leftrightarrow A)$  and  $\vdash A \rightarrow (\perp \leftrightarrow \neg A)$ ).
- iii) From  $(\diamond O \rightarrow \square O)$ .
- iv) Since  $\vdash \diamond(A \wedge B) \rightarrow \diamond(A \wedge \top \wedge (A \rightarrow B))$ , using (SA2) we get  $\vdash \diamond(A \wedge B) \wedge O (A \rightarrow B/\top) \rightarrow O (A \rightarrow B/A \wedge \top)$ , and the desired theorem follows by using the RE-rule (w.r.t. the antecedent) and the contextual RE-rule (since  $\vdash A \rightarrow ((A \rightarrow B) \leftrightarrow B)$ ).
- v) If  $\vdash A \rightarrow B$  then  $\vdash \diamond(A \wedge C) \rightarrow \diamond(A \wedge B \wedge C)$ ; by (SA2),  $\vdash \diamond(A \wedge C) \wedge O (C/B) \rightarrow O (C/A \wedge B)$ ; the theorem follows by the RE-rule w.r.t. the antecedent (since  $\vdash A \rightarrow B$  implies  $\vdash A \wedge B \leftrightarrow A$ ).
- vi) From  $(\leftrightarrow O_a)$ , since  $\vdash \boxed{\rightarrow} A \rightarrow \boxed{\square} (B \leftrightarrow A \wedge B)$ . (Analogously for  $O_i$ .)
- vii) Using axioms  $(O \rightarrow O_a \rightarrow)$  and  $(O_a - C)$ , we deduce  $\vdash \diamond(A \wedge \neg B) \rightarrow (O_a A \wedge O (B/A) \wedge \diamond(A \wedge B) \rightarrow O_a (A \wedge B))$ ; by axiom  $(\leftrightarrow O_a)$ , we deduce  $\vdash \neg \diamond(A \wedge \neg B) \rightarrow (O_a A \rightarrow O_a (A \wedge B))$  (since  $\vdash \neg \diamond(A \wedge \neg B) \rightarrow \boxed{\rightarrow} (A \leftrightarrow A \wedge B)$ ); thus  $\vdash O_a A \wedge O (B/A) \wedge \diamond(A \wedge B) \rightarrow O_a (A \wedge B)$ . (Analogously for  $O_i$ .)

End-proof

As can be seen in the next section, the theorems that play the dominant

<sup>22</sup>We note, however, that we cannot detach  $O_a B(O_i B)$  from the antecedent of these theorems. In fact, even the “weaker” formulas below are *not valid*:

$$O_a A \wedge O (B/A) \wedge \diamond(A \wedge B) \wedge \diamond(A \wedge \neg B) \rightarrow O_a B$$

$$O_i A \wedge O (B/A) \wedge \diamond(A \wedge B) \wedge \diamond(A \wedge \neg B) \rightarrow O_i B.$$

Although it may seem, at first sight, that the failure of these implications represents a weakness of the logic, the analysis of some examples, to follow, indicates that it is — on the contrary — an advantage.

role in determining what may, or may not, be derived from a chosen representation of a CTD scenario are:

- the factual detachment axioms (FD);
- axioms  $(O \rightarrow O_a \rightarrow)$  and  $(O \rightarrow O_i \rightarrow)$ ;
- the deontic detachment theorems (DD);
- axioms  $(\neg O_a)$  and  $(\neg O_i)$ ;
- axioms  $(\leftrightarrow O_a)$  and  $(\leftrightarrow O_i)$ ;
- and the theorems  $(O_a \rightarrow O_a \wedge)$  and  $(O_i \rightarrow O_i \wedge)$ .

Besides these theorems, we will also make extensive use of the  $T$ -normality of  $\square$ ; the  $D$ -normality of  $\boxrightarrow$ ; axiom  $(\square \rightarrow \boxrightarrow)$  and theorem  $(\diamond \rightarrow \diamond)$ .

## 5 THE ANALYSIS OF SOME CTD SCENARIOS

We shall focus on six scenarios which exhibit CTD structures: Scenario 1 is the Chisholm set — Example 1 above; Scenario 2 is the extended version of Example 2 — “the dog scenario” involving “contrary-to-contrary-to-duties”; Scenario 3 is “the white fence” - Example 3 above; Scenario 4 is “the gentle murderer”; Scenario 5 is “the considerate assassin” — Example 4 above; Scenario 6 is the so-called “Reykjavic scenario”<sup>23</sup>.

The scenarios illustrate the range of problems which a theory of CTD should be able to handle; and they enable us to exhibit the expressive and deductive capabilities of our logical system.

---

<sup>23</sup>This is also discussed in [Belzer, 1987] and [McCarty, 1994], and is one of the many examples considered in [Prakken and Sergot, 1994; Prakken and Sergot, 1996].

SCENARIO 1. *The Chisholm Set**The deontic component*

- (d1) It ought to be that a certain man go to help his neighbours
- (d2) It ought to be that if he goes he tell them he is coming
- (d3) If he does not go, he ought not to tell them he is coming

*Logical representation*

- (d1)  $O(\text{help} / \top)$
- (d2)  $O(\text{tell} / \text{help})$
- (d3)  $O(\neg \text{tell} / \neg \text{help})$

*Assumptions regarding the representation of the facts*

In this example we assume the following obvious hypotheses regarding the representation of the facts:

- (a)  $\text{help} \rightarrow \boxed{\rightarrow} \text{help}$  (thus, by  $(\boxed{\rightarrow} - D)$ ,  $\boxed{\rightarrow} \neg \text{help} \rightarrow \neg \text{help}$ )
- (b)  $\text{tell} \rightarrow \boxed{\rightarrow} \text{tell}$  (analogously,  $\boxed{\rightarrow} \neg \text{tell} \rightarrow \neg \text{tell}$ )

Moreover, we also assume that

- (c)  $(\text{help} \wedge \neg \text{tell}) \rightarrow \boxed{\rightarrow} \neg \text{tell}$  (but not  $\neg \text{tell} \rightarrow \boxed{\rightarrow} \neg \text{tell}$ )

since when the agent concerned has helped his neighbours but did not tell them he was coming, it makes no sense to consider any actual alternative where he tells them he is coming. On the other hand, although in some of the cases it might be reasonable to accept both  $\neg \text{help} \rightarrow \boxed{\rightarrow} \neg \text{help}$  and  $\neg \text{tell} \rightarrow \boxed{\rightarrow} \neg \text{tell}$ , we shall not adopt either of these assumptions. (Nevertheless, by  $(\boxed{\square} - T)$ , we have, for any sentence  $A$ ,  $\vdash A \rightarrow \boxed{\rightarrow} A$ .)

## CASE 1.1.

*Factual component*

- (f1) X (the agent concerned) decides not to go to help his neighbour (and, of course, he has not yet gone to help them).

- (f2) But it is potentially possible for X to help and to tell and potentially possible for X to help and not to tell.
- (f3) X has not in fact told that he is coming, although it is still actually possible that he does tell and actually possible that he does not tell.

*Logical representation*

- (f1)  $\boxed{\rightarrow} \neg \text{help}$
- (f2)  $\diamond(\text{help} \wedge \text{tell}) \wedge \diamond(\text{help} \wedge \neg \text{tell})$
- (f3)  $\neg \text{tell} \wedge \diamond \text{tell} \wedge \diamond \neg \text{tell}$

(Note that, without assumption (a), we would represent (f1) as  $\neg \text{help} \wedge \boxed{\rightarrow} \neg \text{help}$ , since we have not adopted the  $T$ -schema for  $\boxed{\rightarrow}$ .)

*Conclusions*

In virtue of the factual detachment axioms (FD), we may derive the following:

$$* \quad \text{viol}(\text{help}) \wedge O_a \neg \text{tell}$$

that is to say, X violates his obligation to help his neighbours, and is actually (given the circumstances) under an obligation not to tell them he is coming. It is also possible to derive the conclusion that X violates his obligation to “help and tell”, since it is also possible to deduce, in virtue of the deontic detachment theorem (DD), that  $O_i(\text{help} \wedge \text{tell})$ . We remark here (cf. footnote 22 regarding Result 2-vii)) that it is not also possible to conclude that X has violated an obligation to tell *simpliciter*, since that particular obligation would not come into effect until X’s helping was a fixed fact. This result seems to accord well with intuition.

Note also that the pragmatic oddity, as we have diagnosed it, is avoided, for we could consistently add to the deontic component above an obligation to the effect that X ought not both go to help and not say that he is coming, i.e., an obligation of the form  $O(\neg \text{help} \vee \text{tell} / \top)$ .

CASE 1.2.

*Factual component*

- (f1) X has helped his neighbours and told them he was coming.
- (f2) But it was potentially possible that X did not help his neighbours.



*Logical representation*

- (f1)  $\text{help} \wedge \text{tell}$   
 (by assumptions (a) and (b), this implies  $\boxed{\rightarrow}(\text{help} \wedge \text{tell})$ )
- (f2)  $\diamond\neg\text{help}$

*Conclusions*

We may derive the following (using (FD) and (DD)):

- \*  $O_i \text{help} \wedge O_i (\text{help} \wedge \text{tell}) \wedge \text{help} \wedge \text{tell}$

So, X has met his ideal obligations, and no actual obligation arises (as can be seen by taking into account  $(\neg O_a)$ ).

## CASE 1.3.

*Factual component*

- (f1) X has helped his neighbours and he did not tell them he was coming.
- (f2) It was potentially possible that he both helped and told, as well as that he did not help them.

*Logical representation*

- (f1)  $\text{help} \wedge \neg\text{tell}$   
 (by assumptions (a) and (c), this implies  $\boxed{\rightarrow}(\text{help} \wedge \neg\text{tell})$ )
- (f2)  $\diamond(\text{help} \wedge \text{tell}) \wedge \diamond\neg\text{help}$

*Conclusions*

We may derive the following (using (FD) and (DD)):

- \*  $O_i \text{help} \wedge \text{help} \wedge \text{viol}(\text{help} \wedge \text{tell})$

So, X meets his ideal obligation to help, but violates his obligation to help and tell; and no actual obligations are derivable.

## CASE 1.4.

*Factual component*

- (f1) X has helped his neighbours, although it was potentially possible that he did not help them.
- (f2) X did not tell his neighbours he was coming, since that was potentially impossible for X to do (imagine that there were no available means of communication).

*Logical representation*

(f1)  $\text{help} \wedge \Diamond \neg \text{help}$  (recall that, by (a),  $\text{help} \rightarrow \Box \text{help}$ )

(f2)  $\Box \neg \text{tell}$

*Conclusions*

We may derive the following (using (FD)):

\*  $O_i \text{ help} \wedge \text{help}$

So, X has not violated any obligation: the obligation sentence  $O_i$  (help  $\wedge$  tell) cannot be derived, because it would be impossible to satisfy such an obligation; furthermore, X has met his ideal obligation to help. No actual obligations are derivable.

## CASE 1.5.

*Factual component*

(f1) It is not potentially possible for X to help his neighbours (for some reason or other — perhaps, for instance, there are no available means for X to travel to his neighbours' house); however, X tells his neighbours he is coming, but he might not have told them so.

*Logical representation*

(f1)  $\Box \neg \text{help} \wedge \text{tell} \wedge \Diamond \neg \text{tell}$

*Conclusions*

We may derive the following (again using (FD)):

\*  $\text{viol}(\neg \text{tell})$

SCENARIO 2. *The dog example – extended with a second-level CTD*

*The deontic component*

- (d1) There ought to be no dog
- (d2) If there is no dog, there ought not to be a warning sign
- (d3) If there is a dog, there ought to be a warning sign
- (d4) If there is a dog and no warning sign, there ought to be a high fence

*Logical representation*

- (d1)  $O(\neg\text{dog} / \top)$
- (d2)  $O(\neg\text{sign} / \neg\text{dog})$
- (d3)  $O(\text{sign} / \text{dog})$
- (d4)  $O(\text{fence} / \text{dog} \wedge \neg\text{sign})$

*Assumptions regarding the representation of the facts*

We here adopt no specific hypotheses regarding the representation of the facts.

## CASE 2.1.

*Factual component*

- (f1) There is a dog, and it is actually possible to keep it or to get rid of it (and there is no information regarding the possibility of having, or not, a sign or a fence).

*Logical representation*

- (f1)  $\text{dog} \wedge \Diamond\text{dog} \wedge \Diamond\neg\text{dog}$

*Conclusions*

We may derive the following regarding violation and actual obligation:

$$* \quad \text{viol}(\neg\text{dog}) \wedge O_a \neg\text{dog}$$

So, there is a violation of the obligation not to have a dog, and there is an actual obligation to get rid of it.

## CASE 2.2.

*Factual component*

- (f1) There is a dog and it is not actually possible that there is not a dog (i.e., the presence of the dog is actually fixed), but it was potentially possible that there was no dog.

*Logical representation*

$$(f1) \text{ dog} \wedge \boxed{\rightarrow} \text{dog} \wedge \diamond \neg \text{dog}$$

*Conclusions*

We may derive the following regarding violation and actual obligation:

$$* \quad \text{viol}(\neg \text{dog})$$

There is a violation of the prohibition against having a dog, and no actual obligations may be derived - although it is possible to derive

$$\diamond \text{sign} \wedge \diamond \neg \text{sign} \rightarrow O_a \text{sign}$$

## CASE 2.3.

*Factual component*

- (f1) There is a dog and that fact is actually fixed (i.e., it is not actually possible that there is not a dog), but there might potentially have been no dog.
- (f2) There is no sign and it is not actually possible that there is a sign, but there might potentially have been both a dog and a sign.
- (f3) There is not a fence, and both the presence and the absence of a fence are actual possibilities.

*Logical representation*

$$(f1) \text{ dog} \wedge \boxed{\rightarrow} \text{dog} \wedge \diamond \neg \text{dog}$$

$$(f2) \neg \text{sign} \wedge \boxed{\rightarrow} \neg \text{sign} \wedge \diamond (\text{dog} \wedge \text{sign})$$

$$(f3) \neg \text{fence} \wedge \diamond \neg \text{fence} \wedge \diamond \text{fence}$$

*Conclusions*

We may derive, in particular, the following violations and actual obligation:

$$* \quad \text{viol}(\neg \text{dog}) \wedge \text{viol}(\text{dog} \rightarrow \text{sign}) \wedge \text{viol}(\text{dog} \wedge \neg \text{sign} \rightarrow \text{fence}) \wedge O_a \text{fence}$$

The ideal obligation that there be no dog is obtained by applying (FD). By application of the axiom schema  $(O \rightarrow O_i \rightarrow)$ , we derive that:

- (i) it is ideally obligatory that if there is a dog then there is a sign

(note that, from (f1) and (f2) we get  $\text{dog} \wedge \neg \text{sign}$ , and so  $\Diamond(\text{dog} \wedge \neg \text{sign})$ ; the application of  $(\text{O} \rightarrow \text{O}_i \rightarrow)$  is then trivial, taking into account (d3) and (f2));

and

(ii) it is ideally obligatory that if there is a dog and no sign then there is a fence

(for instance, from (f1), (f2) and (f3), plus the normality of  $\Box$ , deduce  $\Diamond(\text{dog} \wedge \neg \text{sign} \wedge \text{fence})$  and  $\Diamond(\text{dog} \wedge \neg \text{sign} \wedge \neg \text{fence})$ , which imply (by  $(\Diamond \rightarrow \Diamond)$ )  $\Diamond(\text{dog} \wedge \neg \text{sign} \wedge \text{fence})$  and  $\Diamond(\text{dog} \wedge \neg \text{sign} \wedge \neg \text{fence})$ , and then use  $(\text{O} \rightarrow \text{O}_i \rightarrow)$  and (d4)).

Axiom (FD) enables us to derive that, given the circumstances, the actual obligation is to put up a fence.

### SCENARIO 3. *The white fence case - Example 3*

*The deontic component*

- (d1) There must be no fence
- (d2) But, if there is a fence it must be white

*Logical representation*

- (d1)  $\text{O}(\neg \text{fence} / \top)$
- (d2)  $\text{O}(\text{white-fence} / \text{fence})$

*Assumptions regarding the representation of the facts*

$\Box(\text{white-fence} \rightarrow \text{fence})$

#### CASE 3.1.

*Factual component*

- (f1) There is no fence, and it is still actually possible not to erect a fence and actually possible to erect a fence, white or not.

*Logical representation*

- (f1)  $\neg \text{fence} \wedge \Diamond \neg \text{fence} \wedge \Diamond \text{white-fence} \wedge \Diamond(\text{fence} \wedge \neg \text{white-fence})$
- (Note that, by the assumption made, from  $\Diamond \text{white-fence}$  we can derive  $\Diamond(\text{fence} \wedge \text{white-fence})$ .)

*Conclusions*

We may derive the following regarding ideal/actual obligations:

$$* \quad O_i \neg\text{fence} \wedge O_i (\text{fence} \rightarrow \text{white-fence}) \wedge \neg\text{fence} \wedge \\ O_a \neg\text{fence} \wedge O_a (\text{fence} \rightarrow \text{white-fence})$$

There is an ideal obligation not to have a fence and an ideal obligation that if there is a fence then it must be white; neither of these ideal obligations has been violated; and both persist as actual obligations.

## CASE 3.2

*Factual component*

- (f1) There is a white fence, and it is actually fixed that there will be a fence, possibly white or of another colour.  
 (f2) But it was potentially possible not to have a fence.

*Logical representation*

- (f1)  $\text{white-fence} \wedge \boxed{\rightarrow} \text{fence} \wedge \blacklozenge \text{white-fence} \wedge \blacklozenge (\text{fence} \wedge \neg\text{white-fence})$   
 (f2)  $\blacklozenge \neg\text{fence}$

*Conclusions*

We may derive, in particular, the following violation, and ideal and actual obligations:

$$* \quad \text{viol}(\neg\text{fence}) \wedge O_i (\text{fence} \rightarrow \text{white-fence}) \wedge \\ O_a (\text{fence} \rightarrow \text{white-fence}) \wedge O_a \text{white-fence}$$

The ideal obligation not to have a fence has been violated and does not persist as an actual obligation, since it can no longer actually be fulfilled; the obligation to the effect that if there is a fence then it must be white has not been violated, and it persists as an actual obligation. We may also derive the actual obligation to have a white fence.

SCENARIO 4. *The gentle murderer**The deontic component*

(d1) You should not kill Mr. X

(d2) But, if you kill Mr. X, you should do it gently

*Logical representation*(d1)  $O(\neg\text{kill} / \top)$ (d2)  $O(\text{kill-gently} / \text{kill})$ *Assumptions regarding the representation of the facts*

We make the following assumptions:

- (a)  $\Box(\text{kill-gently} \rightarrow \text{kill})$   
(thus:  $(\text{kill-gently} \rightarrow \text{kill})$  and  $\Box(\text{kill-gently} \rightarrow \text{kill})$ )
- (b)  $\text{kill} \rightarrow \Box\text{kill}$
- (c)  $\text{kill-gently} \rightarrow \Box\text{kill-gently}$
- (d)  $\text{kill} \wedge \neg\text{kill-gently} \rightarrow \Box\neg\text{kill-gently}$

In the cases discussed below, assumptions (b), (c) and (d) are relevant only in regard to determining that no actual obligations arise. Note that the difference between Scenarios 4 and 3 is that the counterparts to assumptions (b), (c) and (d) cannot be adopted for Scenario 3.

## CASE 4.1

*Factual component*

- (f1) The assassin has killed Mr. X, but gently.
- (f2) It was potentially possible for the assassin not to kill and potentially possible that he killed “non-gently”.

*Logical representation*

- (f1)  $\text{kill-gently}$
  - (f2)  $\Diamond\neg\text{kill} \wedge \Diamond(\text{kill} \wedge \neg\text{kill-gently})$
- (Note that  $\vdash \text{kill-gently} \rightarrow \Diamond\text{kill-gently}$ .)

*Conclusions*

We may derive the following (using (FD) and  $(O \rightarrow O_i \rightarrow)$ ):

$$* \quad \text{viol}(\neg\text{kill}) \wedge O_i(\text{kill} \rightarrow \text{kill-gently}) \wedge \text{kill-gently}$$

The assassin has violated his obligation not to kill and has fulfilled his obligation to kill Mr. X gently if he was going to kill him; no actual obligation arises; (using, for instance, (f1), (a) and (b) we can derive  $\Box$  kill: so, it is actually impossible to fulfil the obligation not to kill Mr. X; also we cannot derive an actual obligation to kill gently, since that cannot actually be violated, taking into account (c)).

## CASE 4.2.

*Factual component*

- (f1) The assassin has killed Mr. X and not gently.
- (f2) It was potentially possible for the assassin not to kill and potentially possible for him to kill gently.

*Logical representation*

- (f1)  $\text{kill} \wedge \neg\text{kill-gently}$
- (f2)  $\Diamond\neg\text{kill} \wedge \Diamond\text{kill-gently}$

*Conclusions*

We may derive the following:

$$* \quad \text{viol}(\neg\text{kill}) \wedge \text{viol}(\text{kill} \rightarrow \text{kill-gently})$$

The assassin has violated both his obligation not to kill and his obligation to kill gently if he does kill; no actual obligation arises; (use (f1), (b) and (d)).

## CASE 4.3.

*Factual component*

- (f1) It has been proved (in court) that the “assassin” has killed Mr. X, but gently (and this is admitted by the “assassin”).
- (f2) The “assassin” argues in court that he had no other choice, i.e. that it was potentially impossible for him not to kill Mr. X, because a real assassin had told him that he would kill his son if he (the “assassin”) did not kill Mr. X. The prosecution argues that it was potentially possible for the “assassin” not to kill Mr. X (for instance because he could ask the police for protection for his son).



*Logical representation of the facts according to the point of view of the defence*

- (f1) kill-gently  
 (f2)  $\square$  kill  $\wedge$   $\diamond$   $\neg$ kill-gently

*Conclusions*

Since it was impossible not to kill Mr. X the ideal obligation not to kill cannot be derived; using (FD) we can derive the ideal obligation to kill gently, and this obligation was fulfilled.

*Logical representation of the facts according to the point of view of the prosecution*

- (f1) kill-gently  
 (f2)  $\diamond$   $\neg$ kill  $\wedge$   $\diamond$ (kill  $\wedge$   $\neg$ kill-gently)

*Conclusions*

As in case 4.1, we can derive  $\text{viol}(\neg\text{kill})$ . So, the prosecution argues: the “assassin” should be considered guilty.

*Conclusions*

Of course it may well be suggested that the tactic of the defence here is most unwise. Perhaps they should first accept the prosecution’s claim that it was potentially possible for the “assassin” to refrain from killing, and thus that the “assassin” was ideally obliged not to kill; but then the defence should point out that obviously the “assassin” acted under duress, because of the threat to his son, and thus that the option of not killing was not one which the “assassin” could reasonably be expected to choose.

Again we note here that the logic does not determine the status of the facts; but its language is capable of representing the opposing views concerning their status - in this case, concerning what is taken to be potentially possible. The logic’s task is to show which conclusions regarding obligations and violations follow from a given set of norms, once a particular proposal has been made as to the status of the facts.

SCENARIO 5. *The considerate assassin**The deontic component*

(d1) You should not kill Mr. X

(d2) But, if you kill Mr. X, you should offer him a cigarette

*Logical representation*(d1)  $O(\neg\text{kill} / \top)$ (d2)  $O(\text{offer} / \text{kill})$ *Assumptions regarding the representation of the facts*(a)  $\text{kill} \rightarrow \boxed{\rightarrow} \text{kill}$ (b)  $\text{offer} \rightarrow \boxed{\rightarrow} \text{offer}$ (c)  $\text{kill} \wedge \neg\text{offer} \rightarrow \boxed{\rightarrow} \neg\text{offer}$ 

## CASE 5.1.

*Factual component*

(f1) The assassin has not yet killed Mr. X and has not offered him a cigarette.

(f2) It is still actually possible for the assassin to kill Mr. X and to offer him a cigarette or to kill and not offer a cigarette or not to kill and offer him a cigarette or not to kill and not offer him a cigarette.

*Logical representation*(f1)  $\neg\text{kill} \wedge \neg\text{offer}$ (f2)  $\diamond(\text{kill} \wedge \text{offer}) \wedge \diamond(\text{kill} \wedge \neg\text{offer}) \wedge \diamond(\neg\text{kill} \wedge \text{offer}) \wedge \diamond(\neg\text{kill} \wedge \neg\text{offer})$ *Conclusions*We may derive the following (using (FD),  $(O \rightarrow O_a \rightarrow)$ , and  $(O \rightarrow O_i \rightarrow)$ ):

$$* \quad O_i \neg\text{kill} \wedge O_i (\text{kill} \rightarrow \text{offer}) \wedge \neg\text{kill} \wedge \neg\text{offer} \wedge O_a \neg\text{kill} \wedge O_a (\text{kill} \rightarrow \text{offer})$$

The assassin has not violated his ideal obligations not to kill Mr. X and to offer Mr. X a cigarette if he was going to kill him, and these obligations persist as actual obligations.

## CASE 5.2

*Factual component*

- (f1) The assassin has not yet killed Mr. X but he has firmly decided to kill him.  
 (f2) It is actually possible for the assassin to offer or not a cigarette.

*Logical representation*

- (f1)  $\neg\text{kill} \wedge \boxed{\rightarrow}\text{kill}$   
 (f2)  $\diamond\text{offer} \wedge \diamond\neg\text{offer}$

*Conclusions*

We may derive, in particular:

$$* \quad O_i \neg\text{kill} \wedge O_a (\text{kill} \wedge \text{offer}) \wedge O_a \text{offer}$$

Given that it is actually a fixed fact that the assassin is going to kill Mr. X, his actual obligation is to offer Mr. X a cigarette. Of course, his ideal obligation is not to kill. Note that from  $O_a \text{offer}$  (by  $(O \rightarrow O_a \rightarrow)$ ) we can derive  $O_a (\text{kill} \rightarrow \text{offer})$ ; and the obligation  $O_a (\text{kill} \wedge \text{offer})$  can then be obtained by (f1) and  $(O_a \rightarrow O_a \wedge)$ . We shall comment on this conclusion in Section 6, below, but note at this point that  $O_a \text{kill}$  is *not* derivable.

SCENARIO 6. *The Reykjavik scenario**The deontic component*

Consider the following instructions given to officials accompanying Reagan and Gorbachov at the Reykjavik meeting:

- (d1) The secret shall be told neither to Reagan nor to Gorbachov  
 (d2) But if the secret is told to Reagan it shall also be told to Gorbachov  
 (d3) And if the secret is told to Gorbachov it shall also be told to Reagan

*Logical representation*

Suppose that 'r' represents 'Reagan knows the secret' and that 'g' represents 'Gorbachov knows the secret'

- (d1)  $O (\neg r \wedge \neg g / \top)$   
 (d2)  $O (g/r)$   
 (d3)  $O (r/g)$

*Comments on the logical representation of the deontic component*

- (a) Sentence (d1) ought to be represented in the form indicated and not as  $O(\neg r/\top) \wedge O(\neg g/\top)$ . The reason is that there is no “absolute” obligation according to which, e.g., Reagan is not to know the secret (in just any context in which that would be possible); what is “absolutely” obligatory is that neither of them knows the secret. In fact, if the above alternative formulation of (d1) were to be employed, then in the situation where, for instance, the secret has been told to Gorbachov but not to Reagan, it would be possible to derive conflicting actual obligations, and that derivation would be intuitively correct. It is a merit of the logic proposed that it would detect such a conflict. (A similar point is made in [Prakken and Sergot, 1994; Prakken and Sergot, 1996] in their discussion of this scenario.)
- (b) A possible alternative representation of (d2)+(d3) would be (d2+3):  $O(g \wedge r/g \vee r)$ . (Although it would then be possible to generate the same conclusions as from (d2)+(d3), different patterns of derivation would be involved.)

*Assumptions regarding the representation of the facts*

We make the following assumptions:

- (a)  $r \rightarrow \boxed{\rightarrow} r$   
 (b)  $g \rightarrow \boxed{\rightarrow} g$   
 (c)  $\diamond(r \wedge g)$  (which gives:  $\diamond r$  and  $\diamond g$ , as well as  $\diamond(r \wedge g)$ )  
 (d)  $\neg r \rightarrow \diamond \neg r$   
 (e)  $\neg g \rightarrow \diamond \neg g$

## CASE 6.1.

*Factual component*

- (f1) The secret has not yet been told to either of them

*Logical representation*

- (f1)  $\neg r \wedge \neg g$

*Conclusions*

$$* \quad O_i(\neg r \wedge \neg g) \wedge \neg r \wedge \neg g \wedge O_a(\neg r \wedge \neg g)$$

The obligation to tell the secret to neither of them has not been violated, and persists as an actual obligation. Application of (FD) is instrumental in the derivation of these conclusions.

## CASE 6.2

*Factual component*

- (f1) Reagan knows the secret but Gorbachov does not
- (f2) But it was potentially possible that neither of them knew

*Logical representation*

- (f1)  $r \wedge \neg g$
- (f2)  $\Diamond(\neg r \wedge \neg g)$

*Conclusions*

\*  $\text{viol}(\neg r \wedge \neg g) \wedge O_a g$  (it suffices to use (FD))

We may also derive  $O_a(r \wedge g)$ , by direct application of  $(O_a \rightarrow O_a \wedge)$ , on the basis of the prior derivation of  $O_a g$ . On the other hand, if we employ (d2+3) instead of (d2) and (d3), we get  $O_a(r \wedge g)$  by direct application of (FD), and then on the basis of that result we can use  $(\leftrightarrow O_a)$  to obtain  $O_a g$ .

## CASE 6.3

*Factual component*

- (f1) The secret has been told at the same time to Reagan and Gorbachov

*Logical representation*

- (f1)  $r \wedge g \wedge \Diamond(\neg r \wedge \neg g)$

*Conclusions*

\*  $\text{viol}(\neg r \wedge \neg g)$

There is violation of the obligation to tell neither of them, and no actual obligations are now derivable.

## CASE 6.4

*Factual component*

- (f1) One, and only one, of Reagan and Gorbachev has been told the secret
- (f2) It might potentially have been the case that neither of them was told

*Logical representation*

(f1)  $(r \vee g) \wedge \neg(r \wedge g)$

(f2)  $\Diamond(\neg r \wedge \neg g)$

*Conclusions*

\*  $\text{viol}(\neg r \wedge \neg g) \wedge O_a(r \wedge g)$

Note that the derivation of  $O_a(r \wedge g)$  requires the use of (FD), and then the application of  $(O_a \rightarrow O_a \wedge)$ ; the derivation could be made directly by application of (FD) if (d2+3) were used to represent the second and third lines of the deontic component.

## 6 EVALUATING THE PROPOSED APPROACH

## 6.1 Closure under implication

Our ideal/actual obligation operators are not closed under the (RM)-rule. Since we have  $(\neg N)$  (for both  $O_i$  and  $O_a$ ), and since the operators are classical, closure under the (RM)-rule would yield the result that any ideal or actual obligation implies a contradiction. But even if  $O_i$  and  $O_a$  had been defined in such a way that  $(\neg N)$  were not valid, there are still reasons for not wanting the (RM)-rule; as we remarked in Section 2.2, from the point of view of *violation* the Ross problem — which acceptance of the (RM)-rule would generate — does seem to be genuine. So, given the focus on violation in our approach, the Ross problem is to be avoided.

This was a main reason for not supplementing our models with respect to ‘ob’. If we had imposed that “if  $Y \subseteq Z$  and  $Y \in \text{ob}(X)$ , then  $Z \in \text{ob}(X)$ ”, then although we would not get the (RM)-rule in its full generality (because of the second conjunct in the truth conditions for actual and ideal obligation sentences), we would still get weaker versions of it, such as: “if  $\models A \rightarrow B$  then  $\models \Diamond \neg B \rightarrow (O_i A \rightarrow O_i B)$ ”,<sup>24</sup> that are strong enough to generate the Ross problem.

Nevertheless, some would maintain<sup>25</sup> that at least weaker versions of the (RM)-rule are needed, since if, for instance, it is forbidden to kill, it seems strange that we cannot also derive that it is forbidden to strangle.

The claim seems to be this: for a particular class of pairs of sentences  $(A, B)$ , the conditional relation between  $A$  and  $B$  is such that it should license the derivation of “if  $B$  is forbidden then  $A$  is also forbidden”. Suppose that we introduce a connective  $\Rightarrow$  to express the kind of conditional

<sup>24</sup>Using condition 5-b), we would even get the stronger result “ $\models (\Box(A \rightarrow B) \wedge \Diamond \neg B) \rightarrow (O_i A \rightarrow O_i B)$ ” (and “ $\models (\Box(A \rightarrow B) \wedge \Diamond \neg B) \rightarrow (O_a A \rightarrow O_a B)$ ” with respect to actual obligations).

<sup>25</sup>We are grateful to Henry Prakken for this criticism.

relation concerned (leaving its semantics open for the moment). Instances of this conditional relationship (taken from the scenarios analysed in Section 5) would be (white-fence  $\Rightarrow$  fence) and (kill-gently  $\Rightarrow$  kill), besides the previously mentioned (strangle  $\Rightarrow$  kill). Using this connective, one possible way of interpreting the claim, in terms of our approach, might be to require the validity of the following sentences:

$$(\$i) (A \Rightarrow B) \wedge \Diamond A \rightarrow (O_i \neg B \rightarrow O_i \neg A)$$

$$(\$a) (A \Rightarrow B) \wedge \Diamond A \rightarrow (O_a \neg B \rightarrow O_a \neg A)$$

It would not be a difficult matter to define appropriate truth conditions for  $\Rightarrow$  and to relate them to the semantics of ‘*ob*’ in such a way as to secure the validity of these two sentences. But we do not pursue this line here, since we think that, in terms of our approach, there are reasons for supposing that they should not be deemed valid.

To see why, recall again the gentle murderer scenario, supposing that “*A*” is “kill-gently” and “*B*” is “kill”. (In what follows we will concentrate on (\$i) although a similar argument could be raised against (\$a).) Consider the case where it was potentially possible to kill or not to kill, and potentially possible to kill gently or not to kill gently. According to (\$i) we would derive that there is an ideal obligation not to kill gently. Is this an acceptable derivation? We think not. In our opinion what we should derive — as we do in our logic — is that  $O_i \neg \text{kill} \wedge O_i (\text{kill} \rightarrow \text{kill-gently})$ .

Of course, if the assassin kills gently he violates the prohibition to kill, and that is indeed secured by our logic: assuming the obvious hypothesis that  $\Box (\text{kill-gently} \rightarrow \text{kill})$  then we may derive

$$(*) O_i \neg \text{kill} \rightarrow (\text{kill-gently} \rightarrow \text{viol}(\text{kill}))$$

The fact that we can derive (\*) goes a long way towards accommodating the point behind Prakken’s criticism, we believe.

## 6.2 Axiom ( $\leftrightarrow O_a$ ) and condition 5-b)

First possible counter-example:

Suppose that Mr. X has an actual obligation to help his friend move on Saturday, and suppose that Mr. X (firmly) decides that he will help his friend move on Saturday if and only if he will borrow his brother’s convertible on Saturday. Thus we will have  $O_a \text{help} \wedge \Box (\text{help} \leftrightarrow \text{borrow})$  and, using axiom ( $\leftrightarrow O_a$ ), we derive  $O_a \text{borrow}$ . However, it has been suggested<sup>26</sup> that Mr. X might have an actual obligation to help his friend, but not an actual obligation to borrow the convertible. We disagree: the

<sup>26</sup>We are grateful to Donald Nute for this criticism.

central point, as we see it, is that the truth of  $\boxed{\leftrightarrow}$  (help $\leftrightarrow$ borrow) means that it is not actually possible that either Mr. X helps his friend and does not borrow the convertible or Mr. X borrows the convertible and does not help his friend. Given that actual impossibility, there can be an actual obligation to help the friend if and only if there is an actual obligation to borrow the convertible. The fact that the decision to help a friend involves a commitment to another person, whereas the decision to borrow the car perhaps does not, is irrelevant to the conception of actual obligation we wish to explicate.

Second possible counter-example:

Suppose that Mr. X has an actual obligation to go to the cinema on Saturday, and suppose that Mr. X decides that he will go to the cinema on Saturday if and only if his friend, Mr. Y, also goes to the cinema on Saturday. Thus we will have  $O_a \text{ go-X} \wedge \boxed{\leftrightarrow} (\text{go-X} \leftrightarrow \text{go-Y})$  and we derive  $O_a \text{ go-Y}$ . Does this mean that Mr. Y has an actual obligation to go to the cinema on Saturday? Clearly not! Examples of this sort indicate that there are cases where the formal language needs to be extended to include means of indexing decisions to particular agents, and means of relativising obligations to particular agents. Then it could be made explicit that X has taken the decision, that X bears the obligation to go to the cinema, and that X also bears the obligation to secure Y's presence.

Third possible counter-example:

Recall Case 5.2, from “the considerate assassin” scenario, in Section 5. The facts were: (f1) “The assassin has not yet killed Mr. X but he has firmly decided to kill him” and (f2) “It is actually open for the assassin to offer or not a cigarette”; from these facts, and the norms, and some background assumptions, we concluded not only that the assassin has an actual obligation to offer Mr. X a cigarette ( $O_a \text{ offer}$ ), but also (using the theorem  $(O_a \rightarrow O_a \wedge)$ , which follows from the axiom  $(\leftrightarrow O_a)$ ) that the assassin has an actual obligation to kill Mr. X and to offer him a cigarette ( $O_a (\text{kill} \wedge \text{offer})$ ). This result may seem odd, just because it seems odd to say that someone has an actual obligation to kill (and whatever). But there is a good reason why this result should indeed be forthcoming. For remember that it is a fixed fact that the assassin will kill Mr. X — it is assumed that it is actually impossible not to kill; thus, whatever actual obligations now come into force do so in the context of that assumption. Importantly,  $O_a \text{ kill}$  cannot be derived from  $O_a (\text{kill} \wedge \text{offer})$ , and clearly there cannot be an actual obligation to kill, since it must be actually possible that that which is actually obligatory fails to obtain.

There is a connection between the last point and the Good Samaritan paradox of SDL mentioned in Section 2.2. In our view, the proper response to the Good Samaritan scenario is as follows: it is a fixed fact of the situation that a man X has been robbed. The actual obligation to help him thus arises



in the context of that fixed fact, and so it is appropriate to represent the content of the actual obligation in terms of a conjunction: “X is robbed and X is helped”. In our system, in contrast to SDL, it does *not* follow that there is an obligation to rob; and we have a good explanation for why it should not follow, for that which is actually fixed cannot actually be otherwise, and thus cannot be the object of an actual obligation.

### 6.3 Violations, and ideal and actual obligations

#### Violations of conditional obligations

A sentence of form  $\text{viol}(B)$  is true if we can deduce from a set of deontic norms, and a set of facts, both that ideally it ought to be the case that  $B$  ( $O_i B$ ) and that  $B$  is not the case ( $\neg B$ ). However, it might be suggested that a normgiver often wants to express explicitly in the object language sentences of the form “violation(deontic norm)  $\rightarrow$  Sanction”. The question is how “violation( $O(B/A)$ )” could be represented in our system?<sup>27</sup>

Our answer is that “violation( $O(B/A)$ )” can be characterized as:

$$O(B/A) \wedge \Diamond(A \wedge B) \wedge A \wedge \neg B$$

The explanation is as follows:

Suppose  $O(B/A)$ . Two factual situations are then possible:

- First possibility:  $\Box A$  is the case.

Then (by (FD) and ( $\neg O_i$ )) we derive an ideal obligation  $O_i B$  iff  $\Diamond B \wedge \Diamond \neg B$ . And in order to get  $\text{viol}(B)$  we must have  $\neg B$  (which implies  $\Diamond \neg B$ ). Thus a violation occurs if  $\Diamond B \wedge \neg B$ , which is equivalent, given that  $\Box A$ , to  $\Diamond(A \wedge B) \wedge A \wedge \neg B$ .

- Second possibility:  $\neg \Box A$  is the case.

In this case (using axiom ( $O \rightarrow O_i \rightarrow$ )) we derive from  $O(B/A)$  an ideal obligation of the form  $O_i(A \rightarrow B)$  if  $\Diamond(A \wedge B) \wedge \Diamond(A \wedge \neg B)$ . And in order to get  $\text{viol}(A \rightarrow B)$  we must have  $A \wedge \neg B$  (which implies  $\Diamond(A \wedge \neg B)$ ). Thus a violation occurs if  $\Diamond(A \wedge B) \wedge A \wedge \neg B$ .

In the absence of other conditional obligations, no other ideal obligations can be deduced from  $O(B/A)$ . The only other way we could deduce an ideal obligation from  $O(B/A)$ , would be if we already had  $O_i A$  and  $\Diamond(A \wedge B)$ , in which case we could deduce,

---

<sup>27</sup>We are grateful to Henry Prakken for posing us this question.

by (DD), the further obligation  $O_i(A \wedge B)$ . We omit further details here, but it may be shown that this case too provides no difficulties for the proposed characterisation of “violation( $O(B/A)$ )”.

### Violation vs non-fulfillment

Consider the following new case of the Reykjavik scenario (Scenario 6 in Section 5):

#### *Factual component*

Suppose that the following facts hold at three different points in time ( $t_1 < t_2 < t_3$ ):

- ( $t_1$ -f1) Neither Gorbachov nor Reagan knew the secret (which official 007 knows), and official 007 tells the secret to Reagan
- ( $t_2$ -f1) Official 007 has not yet told the secret to Gorbachov
- ( $t_3$ -f1) Official 007 tells the secret to Gorbachov

#### *Logical representation of the facts*

- ( $t_1$ -f1)  $r \wedge \neg g \wedge \Diamond(\neg r \wedge \neg g)$
- ( $t_2$ -f1)  $\neg g \wedge \Box r$
- ( $t_3$ -f1)  $g \wedge \Diamond\neg g \wedge \Box r$

#### *Conclusions*

- At time  $t_1$ :  $\text{viol}(\neg r \wedge \neg g) \wedge O_a g$
- At time  $t_2$ :  $O_i g \wedge \neg g \wedge O_a g$
- At time  $t_3$ :  $O_i g \wedge g$

The natural reading of these conclusions is as follows: at time  $t_1$ , Official 007 has violated his obligation to tell neither of them the secret, and gets an actual obligation to tell the secret to Gorbachov; at time  $t_2$ , Official 007 has an ideal obligation to tell the secret to Gorbachov, which he has not yet fulfilled, and which persists as an actual obligation; at time  $t_3$ , Official 007 has fulfilled his ideal obligation to tell the secret to Gorbachov, and has no actual obligation.

However, according to our definition of violation, we derive that at time  $t_2$  007 has violated his ideal obligation  $O_i g$ . Is this an intuitively correct interpretation of the situation at  $t_2$ ? Surely, we can say that at time  $t_2$  there exists a violation of the deontic norm ( $d_2$ ) ( $O(g/r)$ ), in the sense described above. But should we conclude that at  $t_2$  the obligation  $O_i g$  has been violated, or rather that it has not yet been fulfilled?

This suggests that in some cases where there is a clear temporal dimension involved it may be natural to distinguish between the violation of

an obligation, on the one hand, and the situation in which an obligation has not been fulfilled, on the other. The way we have defined violation ( $O_i A \wedge \neg A$ ) would perhaps more appropriately be said to correspond to the latter, weaker notion, since that definition does not rule out the actual possibility of fulfilling the (ideal) obligation concerned ( $O_i A \wedge \neg A \wedge \Diamond A$ ); whereas violation in the full sense of the obligation  $O_i A$  implies that falsity of  $A$  is an actual necessity. Analogously, we may wish to distinguish between different degrees of fulfillment of an ideal obligation. We do not pursue this point here, but it seems clear that our formal language as it stands is expressive enough to capture some of the distinctions concerned.

#### Relationship between ideal and actual obligations

In our approach, there is no direct logical connection between the notions of actual and ideal obligation. However, a reasonable question to raise is this: should it not be supposed that an ideal obligation which it is still actually possible to fulfill and actually possible to violate entails an actual obligation to the same effect? An affirmative answer would require the validity of the following sentence:

$$O_i A \wedge \Diamond \neg A \wedge \Diamond A \rightarrow O_a A \quad (O_i \rightarrow O_a)$$

This result would be secured were the following model condition to be adopted:

$$5\text{-e) if } Y \subseteq X \text{ and } Z \in ob(X) \text{ and } Y \cap Z \neq \emptyset, \text{ then } Z \in ob(Y)$$

In [Carmo and Jones, 1997], we conjecture that a condition of this kind could be adopted without untoward consequences. However, the situation is not quite so simple; 5-e) may conflict with the other conditions on our models.<sup>28</sup> A deeper analysis shows that the problem depends fundamentally on the combination of 5-d) and 5-e), together with 5-c), for reasons we now explain.

Consider the graphical description in Figure 1 and suppose  $Y \in ob(X)$ . Then, by condition 5-d),  $((Z - X) \cup Y) \in ob(Z)$ : if a subset  $Y$  of  $X$  is an obligatory proposition in a context  $X$ , then in a bigger context  $Z$  it should be the case that either we are not in  $X$  or we are in  $Y$ . With respect to the set  $S$ , 5-d) does not require  $((S - X) \cup Y) \in ob(S)$ , contrary to what we would obtain if we also adopted condition 5-e) — taking also into account 5-b).

<sup>28</sup>The following counter-example is due to Bjørn Kjos-Hanssen: Suppose  $O(A/\top)$  is true in a model where  $W = \{\alpha_1, \alpha_2, \alpha_3\}$  and  $\|A\| = \{\alpha_1\}$ ; then it can be shown that the conditions 5-b) and 5-d), together with the truth of  $O(A/\top)$ , imply that if  $\alpha_1 \in X$ , then  $ob(X) = \{Z : \alpha_1 \in Z\}$ . Thus, supposing  $Y = \{\alpha_2, \alpha_3\}$ , if 5-e) is also assumed, then we would get (from  $ob(W)$ ) that  $\{\alpha_1, \alpha_2\} \in ob(Y)$  and  $\{\alpha_1, \alpha_3\} \in ob(Y)$ ; and, from 5-b), we would get  $\{\alpha_2\} \in ob(Y)$  and  $\{\alpha_3\} \in ob(Y)$ , and a contradiction would follow from 5-c) and 5-a).

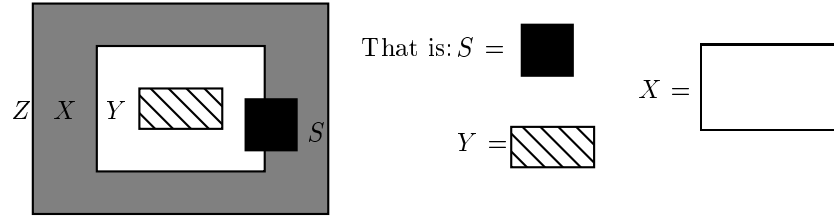


Figure 1.

But  $((S - X) \cup Y) \in ob(S)$  means that  $(S - X) \in ob(S)$ ! Is this acceptable? And, if the deontic norms also require that  $X \in ob(S)$ , we would derive a conflict of obligations. The following case of the dog scenario illustrates the point.

*Factual component*

Suppose that the facts are as follows:

- (f1) there is no dog and there is a sign warning of one; and it was also potentially possibly to have no dog and no sign, or to have a dog and no sign, or to have a dog and a sign
- (f2) the agent is firmly decided to have a sign (for instance, because he wants to frighten possible robbers), and it is still actually possibly to have, or not, a dog

*Logical representation*

- (f1)  $\neg \text{dog} \wedge \text{sign} \wedge \Diamond(\neg \text{dog} \wedge \neg \text{sign}) \wedge \Diamond(\text{dog} \wedge \neg \text{sign}) \wedge \Diamond(\text{dog} \wedge \text{sign})$
- (f2)  $\Box \text{sign} \wedge \Diamond \text{dog} \wedge \Diamond \neg \text{dog}$

From these facts, as our logic stands, we can derive that the ideal obligation not to have a dog has not been violated, and that there was a violation of the ideal obligation not to have a sign if there is no dog; with respect to the actual obligations, we derive the actual obligation not to have a dog. If we now adopt condition 5-e), and introduce  $(O_i \rightarrow O_a)$  as a new axiom schema, then, since we have  $O_i(\neg \text{dog} \rightarrow \neg \text{sign})$ ,  $\Diamond(\neg \text{dog} \rightarrow \neg \text{sign})$  and  $\Diamond \neg(\neg \text{dog} \rightarrow \neg \text{sign})$ , we would derive also  $O_a(\neg \text{dog} \rightarrow \neg \text{sign})$ ; and, since  $\Box \text{sign} \rightarrow \Box((\neg \text{dog} \rightarrow \neg \text{sign}) \leftrightarrow \text{dog})$ , we derive  $O_a \text{dog}$ , conflicting with  $O_a \neg \text{dog}$ . So the question is, in this situation which conclusion should follow, from the intuitive point of view: simply the actual obligation not to have a dog, or a conflict of obligations?

If intuition indicates that a conflict of obligations should not be derivable from this scenario, then 5-e) and  $(O_i \rightarrow O_a)$  must not be adopted. Then

there will be no direct logical connection between ideal and actual obligations - each of them will be derived, independently, directly from the deontic norms, taking into account the relevant context; nevertheless, whenever an ideal obligation  $O_i A$  is deduced from a deontic norm  $O(A/B)$  by factual detachment (FD), we may still also deduce  $\Diamond \neg A \wedge \Diamond A \rightarrow O_a A$ .

If, on the contrary, intuition indicates there does indeed exist a conflict of obligations in the previous situation, then we should adopt condition 5-e), and we should weaken condition 5-c) so that conflicting obligations can be expressed without logical contradiction. For reasons to be explained in the next sub-section, we think it is necessary to weaken 5-c). Although we do not here commit ourselves to acceptance of 5-e), we note in passing that its adoption would have some further interesting consequences, besides providing a direct link between ideal and actual obligations.

First consequence: relationships between  $O(B/A)$  and  $O(A \rightarrow B/\top)$ , and the “pragmatic oddity”

Recall Result 2-iv) in subsection 4.4):  $\vdash \Diamond(A \wedge B) \rightarrow (O(A \rightarrow B/\top) \rightarrow O(B/A))$ . Adopting condition 5-e) (plus 5-b1) and 5-d), the formula below would also become valid:<sup>29</sup>

$$O(B/A) \rightarrow O(A \rightarrow B/\top) \quad (O \rightarrow O \rightarrow)$$

Thus, to represent a deontic conditional by  $O(B/A)$  rather than by  $O(A \rightarrow B/\top)$  would become almost a question of taste (the former is only slightly more general). Moreover, were  $(O \rightarrow O \rightarrow)$  to be adopted as a new axiom, we could conclude that the “pragmatic oddity”, as we have diagnosed it, does not introduce anything really new to the original Chisholm set, since  $O(B/A) \rightarrow O(\neg(A \wedge \neg B)/\top)$  — and so, for instance,  $O(\neg \text{sign} / \neg \text{dog}) \rightarrow O(\neg(\neg \text{dog} \wedge \text{sign}) / \top)$ .

Second consequence: redefinition of  $\mathcal{M} \models_w O(B/A)$

With the condition 5-e) (plus 5-ab)), the condition for  $\mathcal{M} \models_w O(B/A)$  becomes equivalent to the following, simpler, condition:

$$\mathcal{M} \models_w O(B/A) \quad \text{iff} \quad \|B\| \in ob(\|A\|)$$

---

<sup>29</sup>It is trivial to see that  $\mathcal{M} \models_w O(B/A)$  implies  $\|A \rightarrow B\| \neq \emptyset$ . On the other hand, let  $Z$  be such that  $Z \cap \|A \rightarrow B\| \neq \emptyset$ . We have  $\mathcal{M} \models_w O(B/A)$  implies  $\|B\| \in ob(\|A\|)$ ; by condition 5-bd3), this implies  $\|A \rightarrow B\| \in ob(\|A\| \cup Z)$ ; thus, by condition 5-e),  $\|A \rightarrow B\| \in ob(Z)$ .

#### 6.4 *Unrelated deontic norms, contrary-to-duties and conflicting obligations*

Consider again the Case 5.2 of the “considerate assassin” example (Scenario 5 of Section 5):

*Factual situation*

- (f1) The assassin has not yet killed Mr. X but has firmly decided to kill him
- (f2) It is actually possible for the assassin to offer, or not, to Mr. X a cigarette

and suppose now that there is also *another deontic norm applicable to the situation* saying that it is forbidden to offer cigarettes. Thus we have:

- (d1)  $O(\neg\text{kill} / \top)$
- (d2)  $O(\text{offer} / \text{kill})$
- (d3)  $O(\neg\text{offer} / \top)$
- (f1)  $\neg\text{kill} \wedge \boxed{\rightarrow}\text{kill}$
- (f2)  $\diamond\text{offer} \wedge \diamond\neg\text{offer}$

In this case we get, using (FD):  $O_a$  offer (from (f1), (f2) and (d2)) and  $O_a \neg\text{offer}$  (from (f2) and (d3)), and thus a contradiction (since  $O_a$  verifies schemas (C) and (OD)). Is this a problem? We think not. There is a reasonable interpretation of this situation according to which it does contain a conflict of obligations; the logical system we have described was not designed to solve conflicts of obligations, but it should certainly be able to detect them when — as here — they arise. The key feature of this scenario is as follows: (d2) was designed to be “a CTD w.r.t. (d1)”, i.e. to describe the obligations in force in a context of violation of the obligation specified by (d1); but (d1) and the new sentence (d3) express unrelated, or independent, deontic norms; (d2) cannot be seen to be “a CTD w.r.t. (d3)”, and so the obligation not to offer a cigarette “transports down”<sup>30</sup> to the context ( $\boxed{\rightarrow}\text{kill}$ ) of violation of (d1), and a conflict is obtained according to (d2).

“Now one may ask how this conflict should be resolved and, of course, one plausible option is to regard (d2) as an exception to (d3) and to formalize this with a suitable nonmonotonic defeat mechanisms. However, it is important to note that this is a separate issue, which has nothing to do with the CTD aspects of the example”. This remark is quoted from [Prakken and Sergot, 1994, pp.310-311] (replacing their sentences (2) and (3) by our

<sup>30</sup>Using the terminology of [Prakken and Sergot, 1994; Prakken and Sergot, 1996], where a similar result is obtained for this case.

(d2) and (d3)); and we fully agree with it, although we leave entirely open the question of how conflicts of obligations might be *resolved*, since that is a matter falling outside the logical analysis of CTDs themselves.

The above example indicates that semantical condition 5-c) needs to be weakened; consider also the following observation,<sup>31</sup> that in our logic any two “unrelated deontic norms” on the same “antecedent” will give rise to a contradiction, in the following sense:

Suppose that both  $O(B/A)$  and  $O(C/A)$  are true in a model  $\mathcal{M}$ , and suppose that  $\|A\| \cap \|B\| \cap \|\neg C\| \neq \emptyset$  and  $\|A\| \cap \|C\| \cap \|\neg B\| \neq \emptyset$ . Then, defining  $X = (\|A\| \cap \|B\| \cap \|\neg C\|) \cup (\|A\| \cap \|C\| \cap \|\neg B\|)$ , we get  $\|B\| \in ob(X)$  and  $\|C\| \in ob(X)$ ; and, from condition 5-c), it follows that  $(\|B\| \cap \|C\|) \in ob(X)$ , contradicting 5-ab).

Our response is to replace 5-c) by 5-c<sup>-</sup>):

5-c<sup>-</sup>) if  $Y, Z \in ob(X)$  and  $Y \cap Z \cap X \neq \emptyset$ , then  $Y \cap Z \in ob(X)$

It is easy to see that a weakening of this kind does not affect the theorems used in our analysis of the CTD scenarios. This move is in keeping with our belief (a) that sets of norms, as human artifacts, may indeed be imperfectly designed and thus contain the possibility for generating conflicting obligations; and (b) that it is the task of the logic to identify such conflicts when they arise, and to supply conflict-free representations of those CTD scenarios which are, from the intuitive point of view, normatively consistent.

### 6.5 Axiomatisation revisited

Given the weakening of condition 5-c), in order to obtain a *sound* axiomatisation we need to replace the axioms  $(O - C)$ ,  $(O_a - C)$  and  $(O_i - C)$  by the weaker schemas:

$$5'. \quad \Diamond(A \wedge B \wedge C) \wedge O(B/A) \wedge O(C/A) \rightarrow O(B \wedge C/A) \quad (O - C^-)$$

$$11'. \quad \Diamond(A \wedge B) \wedge O_a A \wedge O_a B \rightarrow O_a(A \wedge B) \quad (O_a - C^-)$$

$$\Diamond(A \wedge B) \wedge O_i A \wedge O_i B \rightarrow O_i(A \wedge B) \quad (O_i - C^-)$$

All the other axioms remain unchanged. Moreover, it is easy to see that all the theorems stated in Result 2 (of Subsection 4.4) are retained.

On the other hand, if we were also to adopt condition 5-e), then the axiom schemas  $(O \rightarrow O_a \rightarrow)$  and  $(O \rightarrow O_i \rightarrow)$  would be replaced by the new axiom  $(O \rightarrow O \rightarrow)$  (since the latter, in conjunction with (FD), allows

<sup>31</sup>We are grateful to Henry Prakken for this criticism.

the derivation of the former pair), and we would also add  $(O_i \rightarrow O_a)$  as a new axiom. Obviously, it is then possible to deduce some new theorems<sup>32</sup>.

## 7 A FURTHER LOOK AT SOME OTHER APPROACHES

### 7.1 Temporal approaches

A number of researchers have maintained that the problems raised by the Chisholm set essentially involve a temporal dimension, and that previously proposed solutions fail in as much as they do not capture this dimension. In this section we review some aspects of temporal deontic approaches and compare them with ours.

Temporal approaches to the semantics of deontic notions are generally based on tree-structures, representing branching time with the same past and open to the future. Following [Chellas, 1980, section 6.3], we can think of *time* as an ordered set  $T$  of moments (or *instants*), and — in order to regard the possible worlds as time-stretched — define them as functions from  $T$  into an otherwise unspecified set of momentary world-states  $S$ ; we use the term *history* to denote this specific interpretation of a possible world, and, in the models, we denote by  $H$  the set of the possible histories. Thus we can define the tree-like structures as a tuple  $\langle T, <, S, H, V \rangle$ , where:

- $T$  and  $S$  are non-empty sets;
  - $<$  is a strict linear order on  $T$  (i.e.  $<$  is irreflexive, transitive and  $(\forall i_1, i_2 \in T) (i_1 < i_2 \text{ or } i_2 < i_1 \text{ or } i_1 = i_2)$ );
  - $H$  is a non-empty subset of the set of all functions from  $T$  into  $S$ , satisfying the following restriction (where  $h \equiv_i h'$  iff  $h(i_1) = h'(i_1)$  at every instant  $i_1 \leq i$ ):
- (\*)  $(\forall h, h' \in H) (\forall i \in T) (\text{if } h(i) = h'(i) \text{ then } h \equiv_i h')$ .

Constraint (\*) is intended to give the tree-like form to the temporal structures, allowing branching to the future, but not to the past.

On top of these structures, temporal deontic logics typically define one necessity modal operator plus obligation operators (either monadic or dyadic<sup>33</sup>). However, a main difference appears in the way the temporal dimension is syntactically reflected in the formal language. One family of logics indexes

<sup>32</sup>For instance, we can prove that  $\vdash \Diamond((A \vee C) \wedge B) \wedge O(B/A) \wedge O(B/C) \rightarrow O(B/A \vee C)$  as follows: by  $(O \rightarrow O \rightarrow)$  we deduce  $\vdash O(B/A) \wedge O(B/C) \rightarrow O(A \rightarrow B/\top) \wedge O(C \rightarrow B/\top)$ ; and the new theorem then follows by  $(O - C^-)$  and Result 2-iv).

<sup>33</sup>We are excluding from this comparison the analysis of other deontic operators, such as permission operators, or the obligation-related operators proposed in [Brown, 1996].



the modal and deontic operators with temporal terms; as representatives of the “indexed” family, we may mention [van Eck, 1981] and [Lower and Belzer, 1983].<sup>34</sup> Another family introduces temporal operators that can be iterated with the modal and deontic operators; as representatives of the “non-indexed” family we may mention [Chellas, 1980, sections 6.3 and 6.4], [Åqvist and Hoepelman, 1981], [Thomason, 1981; Thomason, 2001] and [Brown, 1996]. Although the indexed family is more expressive and less abstract than the other family, this difference is not essential in regard to deontic aspects. So we shall introduce a uniform setting where we can abstract from such differences and concentrate on the main distinctions they provide regarding the deontic notions.

In order to see how the truth-value of a sentence can be evaluated within the temporal framework, we first need to extend the tree structures with a valuation of the atomic sentences. Tree-like models are then tuples  $\langle T, <, S, H, V \rangle$ , where:

-  $V$  assigns to each atomic sentence  $p$  a subset of  $H \times T$ , satisfying:

$$(**) (\forall i \in T) (\forall h, h' \in H) \\ (\text{if } h(i) = h'(i) \text{ then } (\langle h, i \rangle \in V(p) \text{ iff } \langle h', i \rangle \in V(p)))$$

Informally,  $\langle h, i \rangle \in V(p)$  iff  $p$  is true at the time instant  $i$  in the world-state  $h(i)$ .

Constraint  $(**)$  is intended to capture the idea that the truth-value of an atomic sentence is a function solely of the actual current world state, and independent of its past and future. We are here assuming that the atomic sentences take the simple form of propositional variables, thus avoiding most of the complications introduced by a first order component, such as the one considered in [van Eck, 1981].

With respect to truth in a model, we can follow Prior’s “Ockhamist” approach to indeterminist time [Prior, 1967], and say that a sentence  $A$  is true in  $\mathcal{M}(= \langle \mathcal{T}, <, \mathcal{S}, \mathcal{H}, \mathcal{V} \rangle)$  iff  $A$  is true at all pairs  $\langle h, i \rangle$  (belonging to  $H \times T$ ). Thus the basic semantic unit for the truth analysis is the pair  $\langle h, i \rangle$ : intuitively  $i$  denotes the current instant; although the intuition behind  $h$  is less obvious, we may see  $h$  as fixing the current and past states and pointing out a possible future (that we may call *actual*, or *prima facie*, following Prior). So  $\|A\|^{\mathcal{M}} = \{\langle h, i \rangle \in H \times T : \mathcal{M} \models_{\langle h, i \rangle} A\}$  and  $\|p\|^{\mathcal{M}} = V(p)$ ; as usual, we write  $\|A\|$ , leaving  $\mathcal{M}$  implicit.

Many temporal operators can be defined in terms of this semantical framework, and the various “non-indexed” temporal deontic logics vary a good deal with respect to the kinds of temporal operators they consider.

<sup>34</sup>In [Lower and Belzer, 1983] the dyadic  $\text{O}$ -operator is not time-indexed, but all the other modal/ deontic operators are.

Just to give one example, we can define a “sometime in the (actual) future” operator as follows (corresponding past operators are defined similarly):

$$\mathcal{M} \models_{\langle h, i \rangle} \text{F}A \quad \text{iff} \quad (\exists i < i_1) \mathcal{M} \models_{\langle h, i_1 \rangle} A$$

(dual: G; for past: P and H)

On the other hand, as we have mentioned, the “indexed” temporal deontic logics do not adopt any temporal operator; instead, they allow a direct reference to the underlying time structure in the proper formal (object) language. More precisely, they include time variables and possibly other time terms<sup>35</sup>  $t, t_1, \dots$  which refer to the time instants through a function  $v$ , added to the temporal models, that applies each term into an element of  $T$ . Their languages may be seen as a Boolean combination of two sublanguages: the sublanguage of time and the modal/deontic sublanguage, where the latter is built from atomic sentences of the form  $p_t$  (for  $t$  a time term), with time-indexed modal and deontic operators, and sentential connectives. Interestingly, if we ignore this time indexing, we can see that the modal/deontic sublanguages of the indexed and the non-indexed temporal deontic logics are analogous. So, if we could “separate” the time index from the modal/deontic operators, we would get a uniform semantical and logical setting for analysing the modal/deontic component of both types of temporal deontic logics, with obvious advantages for the purposes of comparison. This can be achieved by means of the temporal realisation operator of [Rescher and Urquhart, 1971]:  $(t)$ , for  $t$  a time term ( $\text{R}_t$  in [Rescher and Urquhart, 1971]), semantically defined as follows:

$$\mathcal{M} \models_{\langle h, i \rangle} (t)A \quad \text{iff} \quad \mathcal{M} \models_{\langle h, v(t) \rangle} A$$

Intuitively,  $(t)A$  signifies that  $A$  is true at time  $t$ , and using this operator we can translate each modal/deontic indexed sentence into a non-indexed sentence, with the same meaning, by replacing each  $p_t$  by  $(t)p$ , and each modal/deontic operator  $\#_t$  by  $(t)\#$ . For instance, the sentence  $t < t_1 \rightarrow \text{O}_{t_1} p_t$ , valid in [van Eck, 1981]’s logic, is translated to  $t < t_1 \rightarrow (t_1)\text{O} (t)p$ .

Turning now to the modal/deontic component of temporal deontic logics, we start by noting that they all introduce a necessity modality, expressing some kind of “inevitability”. However, a main division appears here between the temporal deontic logics — independently of whether or not they are indexed — regarding which of the following two types of necessity operator they adopt; for instance [Åqvist and Hoepelman, 1981], [Thomason, 1981; Thomason, 2001], [Lower and Belzer, 1983] and [Brown, 1996] adopt the first, whilst [Chellas, 1980] and [van Eck, 1981] opt for the second:

<sup>35</sup>For particular known numerical time structures  $T$ , it is usually considered that  $T$  is directly “included” in the formal language, allowing the existence of constants and functions, and supposing that terms (of sort  $T$ ) can be written in the formal language as in the semantics (e.g. in the form  $t + 1$ ). The same assumption will be made here whenever appropriate.

$$\mathcal{M} \models_{\langle h, i \rangle} \boxed{A} \quad \text{iff} \quad (\forall h' \in H) (\text{if } h \equiv_i h' \text{ then } \mathcal{M} \models_{\langle h', i \rangle} A)$$

(dual:  $\blacklozenge$ )

$$\mathcal{M} \models_{\langle h, i \rangle} \boxed{\bullet} A \quad \text{iff} \quad (\forall h' \in H) (\text{if } h \approx_i h' \text{ then } \mathcal{M} \models_{\langle h', i \rangle} A)$$

where  $h \approx_i h'$  iff  $h(i_1) = h'(i_1)$ , at every  $i_1 < i$

(dual:  $\blacklozenge$ )

Informally,  $\boxed{A}$  means “for all histories, with the same past and present”, and  $\boxed{\bullet} A$  means “for all histories, with the same past”. Obviously,  $\boxed{\bullet} A \rightarrow \boxed{A} A$  is a valid sentence within this semantics.<sup>36</sup>

A similar division also appears in the way the deontic components are defined, as we explain below.

Starting with the analysis of the unary obligation operator  $\text{O}$ , with the exception of [Brown, 1996], all the mentioned works employ a *normal* logic for  $\text{O}$ , which can be obtained as follows: enrich the temporal models  $\mathcal{M} = \langle T, <, S, H, V, v \rangle$  with a new component (“best” histories)  $bh : H \times T \rightarrow \wp(H)$ , and define

$$\mathcal{M} \models_{\langle h, i \rangle} \text{O} A \quad \text{iff} \quad (\forall h' \in bh(\langle h, i \rangle)) \mathcal{M} \models_{\langle h', i \rangle} A$$

Obviously, the function  $bh$  must satisfy some conditions, besides the non-emptiness of each  $bh(\langle h, i \rangle)$  required by all of these researchers. Importantly, we can distinguish between (a) those approaches — the ones that employ  $\boxed{A}$  — that can be said to interpret  $bh$  intuitively as giving the “best” histories that are still open, given the past and present (which are fixed); they adopt the following conditions (or conditions that imply them):

- $bh(\langle h, i \rangle) \subseteq \{h' : h \equiv_i h'\}$
- if  $h \equiv_i h'$  and  $i_1 \leq i$  then  $bh(\langle h, i_1 \rangle) = bh(\langle h', i_1 \rangle)$

and (b) those approaches — the ones that opt for  $\boxed{\bullet}$  — that can be said to interpret  $bh$  intuitively as giving the “best” histories, that were open, given the (fixed) past; they adopt:

- $bh(\langle h, i \rangle) \subseteq \{h' : h \approx_i h'\}$
- if  $h \approx_i h'$  and  $i_1 < i$  then  $bh(\langle h, i_1 \rangle) = bh(\langle h', i_1 \rangle)$

---

<sup>36</sup>The symbol  $\boxed{\phantom{A}}$  is normally used for both of them, but since we want to distinguish them - and since the symbol  $\boxed{\phantom{A}}$  has already been used for potential necessity — we use  $\boxed{A}$  and  $\boxed{\bullet}$ . We also note that, for an integer-like time, we can express  $\boxed{\bullet}$  as  $Y \boxed{A} X$ , where  $Y$  and  $X$  denote the previous and next time operators, semantically defined as follows:  $\mathcal{M} \models_{\langle h, i \rangle} Y A$  iff  $\mathcal{M} \models_{\langle h, i-1 \rangle} A$  and  $\mathcal{M} \models_{\langle h, i \rangle} X A$  iff  $\mathcal{M} \models_{\langle h, i+1 \rangle} A$ .

(In what follows we shall use  $\odot$  whenever we want to stress that we are referring to the unary obligation operator under this latter class of approaches.)

Some comparison can already be made between these temporal deontic approaches and our own. In fact, it is not difficult to see that the operators  $\blacksquare$  and  $\boxed{A}$  have some similarities with the operators for, respectively, potential and actual necessity,  $\square$  and  $\boxrightarrow$ , *in those contexts where they are assigned a specifically temporal reading*. More generally, it is easy to find connections between each pair of operators  $(\blacksquare, \odot)$  and  $(\boxed{A}, O)$ , and our pairs  $(\square, O_i)$  and  $(\boxrightarrow, O_a)$ . However, there are also some relevant differences. We compare them both at a technical level and from the point of view of their motivation, starting with the former.

As regards the relationships between each necessity operator and its associated obligation operator, we note that — as for our corresponding operators — what is obligatory must be possible to fulfil; that is, the following sentences are valid:

$$\begin{aligned} O A &\rightarrow \blacklozenge A \\ \odot A &\rightarrow \blacklozenge A \end{aligned}$$

However, in contrast to ours, these logics do not require that what is obligatory must also be violable. On the contrary, all necessary truths (in their sense of necessary) are (vacuously) obligatory; that is, the following sentences are valid:

$$\begin{aligned} \boxed{A} A &\rightarrow O A \\ \blacksquare A &\rightarrow \odot A \end{aligned}$$

Thus, since  $A \leftrightarrow \boxed{A} A$  is valid if  $A$  does not include reference to the future, if one wanted to require that an obligation must be both possible to fulfil and violable, then in order for  $O A$  to be true the sentence  $A$  would have to be characterised in such a way that it refers to the future; in the non-indexed logics this would imply that  $A$  can never be a propositional variable and must include future operators (such as  $F$  or  $X$ ); in the indexed case,  $O_t p_{t_1}$  could be true only if  $t_1 > t$ ; (for a general sentence  $O_t A$  one would need to require that the “temporality of”  $A$  is greater than  $t$ ; this notion is from trivial: see e.g. [van Eck, 1981]). On our approach, by contrast, we can simply write  $O_a p$ ,<sup>37</sup> abstracting both from the exact time

<sup>37</sup>The possibility of violation can be described in our logic by the valid sentence  $O_a A \rightarrow \blacklozenge \neg A$ , where for cases where there is a meaningful temporal interpretation, we can see  $\blacklozenge \neg A$  as saying that there is a future instant within some of the alternatives open to the agent where  $A$  is not the case. Note that, in contrast to what happens with  $\boxed{A}$ , our operator  $\boxrightarrow$  does not verify the  $(T)$ -schema.

instant  $t$  where this obligation appears (as in the non-indexed logics), and from the description of the other aspects related with the time of occurrence of the “events” encoded in  $p$ .

Consider now  $(\boxed{\bullet}, \odot)$  and  $(\boxed{\phantom{\bullet}}, O_i)$ . There seems to be a main technical difference between  $\boxed{\phantom{\bullet}}$  and  $\boxed{\bullet}$ . For the latter, the potential alternatives (using our terminology) to the actual current world-state correspond to the alternatives that were open immediately before the current instant, i.e., at  $t_c - 1$ , supposing that  $t_c$  denotes the current instant and assuming here, and in what follows, an integer-like time. On the other hand, the potential alternatives to the actual current world-state, as described by our operator  $\boxed{\phantom{\bullet}}$ , include the alternatives that were open at some relevant time before the current time instant, but not necessarily the one immediately before.

Even in cases where it is appropriate to give a temporal interpretation to our operators, it is necessary to consider only two time instants, on our view: the current instant  $t_c$  (the one relevant to determining the actual obligations), and some previous time instant  $t^*$ , relative to which an assessment is made of the potential possibilities which were open to the agent — an assessment which, in turn, will determine the agent’s ideal obligations at  $t_c$  and his violations at  $t_c$ .

As regards  $\odot$ , it appears that for the analysis of violations only the instant  $t_c - 1$ , immediately preceding  $t_c$ , is relevant. The informal idea seems to be that one can violate only those obligations that require that something be done immediately. But then there is a problem if the obligation was in fact incurred at some time  $t^*$  prior to  $t_c - 1$ : how is the persistence of the obligation to be represented? As is noted in [Prakken and Sergot, 1997], the issue of the persistence of obligations through time has been poorly studied within the temporal approaches to deontic logic, despite its apparently close connections with CTD-problems. One of the main exceptions is [Brown, 1996], where there is some discussion of persistence conditions according to which an obligation persists until it is (definitely) fulfilled or violated. In [Thomason, 2001] we also find a condition related with the persistence of obligations, a condition that more or less corresponds (in our temporal semantics) to: if  $i \leq i_1$  and  $h_1 \in bh(\langle h, i \rangle)$ , then  $bh(\langle h_1, i_1 \rangle) = bh(\langle h, i \rangle) \cap \{h' : h \equiv_{i_1} h_1\}$ . However, on the one hand, this condition seems too strong, since it seems not to allow the appearance of other obligations between  $i$  and  $i_1$ ; and, on the other hand, since it only refers to  $h_1 \in bh(\langle h, i \rangle)$ , it says nothing about how an obligation persists over time when some *other* obligation is violated.

Besides these technical similarities and differences, there is one *fundamental* difference between the motivation underlying our approach and the one underlying the temporal approaches. This difference has to do with the reasons why a history  $h_1$ , which was considered an alternative to the actual history  $h$  at some time instant  $i - 1$  (so that  $h_1 \approx_i h$ ), is no longer taken to be an alternative to  $h$  at  $i$  (i.e., not  $h_1 \equiv_i h$ ). For the temporal

approaches, the essential reason why, at  $i$ ,  $h_1$  is no longer an alternative to  $h$ , is that a state  $h(i)$  makes  $h_1$  impossible (that is,  $h_1(i) \neq h(i)$ ). On the other hand, in our approach, in cases for which a temporal interpretation can meaningfully be assigned to our necessity operators, we would say that our worlds roughly correspond to the pairs  $\langle h, i \rangle$ ; and although  $av(\langle h, i \rangle)$  would be seen as a subset of  $\{\langle h_1, j \rangle : h_1 \equiv_i h \text{ and } j > i\}$  (satisfying the condition that if  $j, k > i$  then  $\langle h_1, j \rangle \in av(\langle h, i \rangle)$  iff  $\langle h_1, k \rangle \in av(\langle h, i \rangle)$ ), it need not necessarily be equal to the set  $\{\langle h_1, j \rangle : h_1 \equiv_i h \text{ and } j > i\}$ ; thus a pair  $\langle h_1, j \rangle$  may belong to  $av(\langle h, i - 1 \rangle)$ , but may fail to belong to  $av(\langle h, i \rangle)$ , even if  $h_1 \equiv_i h$  and  $j > i$ . The essential reason for this is that the agent may have decided to exclude from the worlds that are actual alternatives to  $\langle h, i \rangle$  all the worlds provided by the history  $h_1$ , even if this history is still potentially possible to follow. This fundamental difference between our theory and the temporal approaches is the key to understanding why we are able to accommodate such a CTD-scenario as the gentle murderer, which remains problematic for the temporal approach.

We next note that there is a significant difference between our dyadic obligation operator and those proposed in, e.g., [Åqvist and Hoepelman, 1981] and [van Eck, 1981], both of which see the unary  $O$  as a limiting case of the dyadic  $O$ , in the sense that  $OA$  is an abbreviation of  $O(A/\top)$ . As we have indicated above, we reject a move of this kind; we view unary obligations (of types  $O_i$  and  $O_a$ ) as derivable from the dyadic, in ways that depend on the context, and on matters pertaining to the satisfiability and violability of obligation. The rejection of the idea that  $OA$  is an abbreviation of  $O(A/\top)$  is also a feature of the temporal-based approach in [Lower and Belzer, 1983], to which we now turn.

Loewer and Belzer's 3D-logic involves a combination of a temporal approach with a preference-based approach. Briefly, the basic idea of [Lower and Belzer, 1983] appears to be the following: there is a hypothetical time instant zero, at which some deontic norms are assumed to come into force; these are expressed by the use of a dyadic  $O$ -operator, which is not time-indexed. Actual obligations, expressed by means of a time-indexed unary operator  $O_t$ , evolve from this moment on in ways determined by the deontic norms and by the facts which are taken to be settled at each time  $t$  (settledness being represented by means of an operator of type  $\boxed{A}_t$ ). They also use a notion of "ethical sufficiency", and introduce two versions of sentences of the form  $R(A, B)$ , one time-indexed and one not, intended to express the idea that  $B$  is ethically sufficient for  $A$ . (We omit here the details of their semantical analysis of  $R$ .)

Although they have little to say regarding ideal obligations and their violations, they do offer a suggestion to the effect that ideal obligation of  $A$  can be defined as an abbreviation of  $O(A/\top)$ ; in our opinion this would lead to difficulties in cases where at the instant zero some of the deontic norms cannot be potentially satisfied. Consider, for instance, a case of Scenario

3 in which it is potentially impossible not to have a fence - suppose that the fence is installed in such a way that the agent concerned has neither the ability nor the opportunity to remove it; were  $O_i A$  to be defined as an abbreviation of  $O(A/\top)$ , then  $O_i \neg$ fence would be derived, instead of — as in our logic —  $O_i$  white-fence.

The semantics of the dyadic  $O$  is defined *à la* Lewis, imposing a ranking  $\leq$  on the members of  $H$ . This ranking  $\leq$  is then connected with the model component (“best” histories)  $bh : H \times T \rightarrow \wp(H)$ . In their own words, the intuitive idea is the following [Lower and Belzer, 1983, pp. 308]: “Our basic idea for connecting the ranking  $\leq$  with  $F$  (here denoted by “ $bh$ ”), that is connecting conditional obligations with actual obligations is this: At the first instant of time  $t_0$ , we will assume that some of the histories which are ideal according to  $\leq$  can be achieved. But as time proceeds, events and the actions of men may render the ideal histories unattainable. Still at every moment the actual obligation is to bring about one among those best histories that remain”.

The factual and deontic detachment principles that are validated by their semantics are as follows:

$$O(B/A) \wedge R(B/A) \wedge \boxed{A} \wedge \Diamond_t B \wedge \Diamond_t B \rightarrow O_t B$$

$$O(B/A) \wedge R(B/A) \wedge O_t A \wedge \Diamond_t B \wedge \Diamond_t B \rightarrow O_t B$$

Ignoring the “ethical sufficiency” operator  $R$ , their factual detachment principle strongly resembles a “temporal version” of ours (although, as in the previous logics, they do not require that an obligation must be violable). On the other hand, their deontic detachment principle is stronger than ours, and, for the reasons that we have previously mentioned, we suspect that it is too strong, particularly if one also wants to address the problem of violations of ideal obligations: recall our conclusions in Case 1.1, Section 5 above.

Finally, some further comments about the representation of CTD-scenarios within the temporal deontic approaches, as compared to our own.

Of the temporal approaches here analysed that explicitly discuss the representation of CTD-scenarios ([Åqvist and Hoepelman, 1981], [van Eck, 1981] and [Lower and Belzer, 1983]), the one most similar to ours is [Lower and Belzer, 1983]. In fact, it is the only one of the three that respects our requirement (iv) on the representation of CTD scenarios, and — like us — Lower and Belzer make a distinction between the “deontic norms” applicable to a situation (expressed through the dyadic  $O$ ) and the specific obligations that can be deduced from them given the facts of the case. However, it appears that they opt to represent the first sentence of the Chisholm set by a time-indexed actual obligation rather than as a deontic norm expressed in terms of the dyadic  $O$ -operator.

An important difference remains: in our approach, given the “deontic norms”, in order to derive which violations may have occurred and what

the actual obligations are, we need only consider what it was potentially and actually possible to do. No specific reference to any time instant is needed, in contrast to what happens in [Lower and Belzer, 1983] and other indexed temporal approaches, where sometimes one is forced to introduce time instants in the representation of cases in a way that seems more or less artificial, since their specific values are irrelevant. Moreover, as we have emphasised, our potential and actual necessitation operators are not tied to *temporal* settledness, and so can be used to represent cases where a temporal dimension is absent. In short, we think that our theory offers patterns of representation which are more abstract and simpler than those supplied by the typical temporal deontic approaches.

Of course, whenever it *is* essential to state the exact time of realisation of some obligatory state of affairs — for instance because there is an obligation to do something by a specific deadline — then an indexed temporal logic seems to be necessary. The simplicity of the logic will always depend on the degree of abstraction that we want to achieve, and when we choose, for instance, to use a deontic propositional language we abstract from many features — the temporality of some state of affairs being just one of these. But this abstraction is justifiable precisely on the grounds of the simplicity that it provides for illustrating the *essential* features underlying CTD-reasoning.

## 7.2 Action-based approaches

In the work of Castañeda, and of those influenced by him, it has frequently been maintained that the problems that beset deontic logic can be solved if proper recognition is given to the role of the concept of action. It is perhaps fair to say that Castañeda's own work is not always easy to penetrate, but fortunately in this case two papers by James Tomberlin [1983; 1986] supply both an outline introduction to Castañeda's system of deontic logic, and an appraisal of (respectively) his treatments of the Chisholm set and the Good Samaritan. And, judging by his replies — published in the same sources — Castañeda accepted that Tomberlin provided a faithful account of his position. We focus here on just those aspects which seem essential in regard to the analysis of the Chisholm set.

“Central to Castañeda's enterprise, we encounter his pivotal distinction between *practitions* and *propositions*” [Tomberlin, 1983, pp. 204]. Propositions are understood in the usual way, whereas practitions are understood, roughly, as the semantical content of such sentences as commands, orders, requests, entreaties. Within the scope of deontic operators, such as “it is obligatory that ...”, both types of components may occur: propositions indicate the circumstance, or condition, of a deontic judgment, whereas practitions indicate “. . . the target or focus of the deontic operator; components of this sort are *actions practically considered*. . .” [Tomberlin, 1983, pp. 235]. However, deontic operators are assumed to apply to practitions



only, and thus a mixed scope formula, within the scope of a deontic operator, will have a practition as its content. Deontic judgments are said to belong to different families, where each family determines a particular sense or type of obligation, permission, and so on. This type is indicated in the formal language by an index on the deontic operator.

Let us turn immediately to the representation of the Chisholm set, stating in due course those principles of Castañeda's logic which are relevant to an assessment of that representation. Suppose that the type of obligation concerned is denoted by "s", and that we use  $p, q, \dots$  to stand for propositions, and  $p^*, q^*, \dots$  to stand for practitions. Then line 1 of the Chisholm set is assigned the form:

$$1. O_s p^*$$

whereas line 4 expresses a proposition and is accordingly represented by:

$$4. \neg p$$

As regards line 3 (the CTD), Castañeda insists that it must be understood as specifying what is obligatory in those circumstances in which the obligation expressed by line 1 is violated. He insists, then, on factual detachment, requiring that 1 and 3 must together imply

$$5. O_s \neg q^*$$

Accordingly, 3 is assigned the following form, where  $\rightarrow$  is the material conditional:

$$3. \neg p \rightarrow O_s \neg q^*$$

Line 2 is interpreted as an obligation sentence with a mixed component within the scope of the deontic operator:

$$2. O_s (p \rightarrow q^*)$$

The crucial point to note now is that lines 1 and 2 do *not* imply

$$6. O_s q^*$$

because the embedded antecedent in line 2 is a proposition, whereas the scope formula in line 1 is a practition. So, the claim is, in essence: distinguish properly action components and propositional components, and incorporate that distinction in the representation of the set, and the inconsistency which Chisholm's formalisation contained will disappear. Although Castañeda's deontic logic contains the theorem

$$(CasK) \quad O_s (p^* \rightarrow q^*) \rightarrow (O_s p^* \rightarrow O_s q^*)$$

its application is of course restricted to instances where both the component sentences in the antecedent express propositions. *This* representation of the Chisholm set, then, is consistent.

Before moving on to Tomberlin's criticisms, note that it is already clear that Castañeda's solution generates the pragmatic oddity, in virtue of lines 1 and 5. Outlining the basic feature of Castañeda's semantics of obligation sentences, Tomberlin says that a sentence of the form  $O_s p^*$  "... is true at a world  $w$  if and only if the proposition  $p^*$  belongs to every world  $v$  such that  $v$  is deontically compossible with  $w$ ..." [Tomberlin, 1983, pp. 237]. Then in all the worlds which are deontically compossible with the given world in which the four sentences of the Chisholm set are true, the agent concerned helps his neighbours (by line 1) but does not tell them he is coming (by line 5).

Note that

$$(TC1) \quad (p \rightarrow O_s q^*) \leftrightarrow O_s (p \rightarrow q^*)$$

is a theorem of Castañeda's system. It follows immediately that line 4 implies line 2, and thus that the requirement of logical independence is not met. Tomberlin and Castañeda are among the very few deontic logicians who do not accept this requirement; we do not discuss their reasons for this rejection here — the reader is referred to [Tomberlin, 1983] and Castañeda's reply — but choose rather to focus on Tomberlin's criticism of Castañeda's treatment of the Chisholm set, which also begins from the observation that (TC1) is a theorem. What troubles Tomberlin is that, in virtue of (TC1), line 4 also implies

$$7. O_s (p \rightarrow \neg q^*)$$

Putting lines 2 and 7 together we have: it is obligatory that if X goes to help his neighbours then he tells them he is coming, and it is obligatory that if X goes to help his neighbours then he does not tell them he is coming. This is clearly an intuitively unacceptable consequence — it would not be implied by a *proper* representation of the Chisholm set!

So what has gone wrong? Tomberlin's view, in brief, is that line 2 is not the correct way to represent the second line of the Chisholm set. By virtue of (TC1), which he seems inclined to accept, line 2 is equivalent to a conditional obligation and thus, ignoring the negation signs, has the same logical structure as line 3. But Tomberlin is of the opinion that the second sentence of the Chisholm set should be understood as expressing an obligation *simpliciter*, of the form:

$$2'. O_s (p^* \rightarrow q^*)$$

But then, of course, as he immediately notes, the original Chisholm “paradox” re-emerges, since line 1 and line 2' together imply line 6, by (CasK).

Tomberlin advocates an alternative line of solution to the Chisholm problem, within the framework of Castañeda's deontic logic. “Biting the bullet”, as he puts it, one should accept that *another* aspect of Castañeda's deontic logic — that deontic operators can be relativised to different senses of obligation — has a key role to play in the representation of the Chisholm set. As Tomberlin sees it, the four sentences of the Chisholm set can be assumed to be both true and mutually consistent if and only if the sense or type of obligation pertaining to the third line is *different from* the sense or type of obligation expressed by the first and second lines. So, the first two lines are obligation sentences whose truth conditions refer to deontically perfect worlds (he calls these absolute obligations), whereas the third line is an obligation sentence whose truth conditions refer to worlds which — like the actual world in which line 4 is true — are deontically imperfect.

At this point, naturally, we begin to experience a gentle sense of *déjà vu*. Tomberlin is here treading one of the paths we investigated in Section 3, above, and which — as we there tried to indicate — leads only to further troubles. How would he cope with second-level CTDs of the kind exhibited by Scenario 2 (the dog-sign-fence example)? By introducing a third sense of obligation, i.e., a second sense of imperfection or sub-ideality? Clearly, this proliferation of senses of the fundamental deontic notions is to be avoided. To embrace it is to admit defeat.

Despite the difficulties that arise for Tomberlin's own positive proposal, it does seem clear that his criticism of Castañeda casts considerable doubt on whether the proposition/practition distinction has any role to play in solving Chisholm's puzzle. Our conclusion is that Castañeda failed to supply an analysis of deontic conditionals that both copes with the pragmatic oddity and generates suitable deontic and factual detachment principles. And we should add, with Prakken and Sergot, the observation that such CTD examples as Scenario 3 (the white fence) are apparently devoid of any action component whatsoever.

Nevertheless, we think that there is at least the following to be said for the action approach: the two notions of necessity to which we assigned a crucial role in the analysis of CTD scenarios are intimately connected to praxiological concepts, in particular *decision*, *ability* and *opportunity*. A more elaborate development of our logical system should make these connections explicit. But, even when that is done, we very much doubt that the resulting picture will provide confirmation of *Castañeda's* account of the role of action concepts.

It is suggested in [Hilpinen, 1993, pp. 89] that one interesting way to understand Castañeda's distinction is “. . . to consider it from the standpoint of dynamic deontic logic. In (propositional) dynamic logic, the non-logical

expressions are divided into action terms and propositions, and the deontic operators behave in the same way as in Castañeda's deontic logic . . . : they transform action terms into deontic propositions or statements . . . the way in which the distinction between action terms and propositions is made in the semantics of dynamic logic corresponds nicely to Castañeda's conception of propositions as descriptions of the circumstances under which an action is considered or performed."

Existing work (e.g., [Meyer, 1988]) falls short of providing convincing evidence that an approach to deontic logic based on dynamic logic holds the key to solving CTD problems. One difficulty, as Hilpinen himself notes [Hilpinen, 1993, pp. 94, footnote 4], is that the fourth line of the Chisholm set appears to have "... no plausible representation in Meyer's dynamic deontic logic". Hilpinen's suspicion is confirmed in a later paper by Meyer, Wieringa and Dignum [1997], in which they offer a representation of what they call the 'ought-to-do' version of the Chisholm set in a deontic logic (named PDeL) based on dynamic logic. Concerning their representation they make the following remark: "... the fourth premise of the set ... cannot be represented in PDeL. In some sense, statements of actions in PDeL and the underlying dynamic logic are of a hypothetical nature: 'if one (would) perform the action, the following holds'. The implication implicit in a formula  $[\alpha]\varphi$  is therefore more like a conditional in conditional logic. As such, it is not really important what actually happens. Here and in the sequel we shall just ignore the fourth assertion in the formal representation" [Meyer, Wieringa and Dignum, 1997, § 1.6.3]. (The formula  $[\alpha]\varphi$  says that execution of action  $\alpha$  leads to some state(s) where  $\varphi$  holds.) However, in the context of analysing CTD scenarios, what actually happens is — as we have seen — of paramount importance.

Meyer, Wieringa and Dignum further maintain that it is necessary to distinguish explicitly between the logic of 'ought-to-do' and the logic of 'ought-to-be', and accordingly they also offer an analysis of the 'ought-to-be' version of the Chisholm set. We refer the reader to their paper for the details [Meyer, Wieringa and Dignum, 1997, § 1.5], but note that their 'ought-to-be' logic relies essentially on the availability of an indefinite number of distinct obligation operators, each of which is relativised to a particular 'frame of reference', as they put it. They do not offer precise criteria for determining when one has moved from one frame of reference to another, but they take it for granted that the first three lines of the Chisholm set involve reference to three distinct frames of reference, one for each of the obligations contained. Not surprisingly, problems of inconsistency are thereby avoided, but at the cost of introducing an indefinite number of obligation operators, and in the absence of clear guidelines determining how many operators the representation of a given scenario will need.

### 7.3 Preference-based approaches

“SDL cannot distinguish between various grades or levels of non-ideality; in the semantics of SDL worlds are either ideal or non-ideal. Yet the expression ‘if you kill, kill gently’ says that some non-ideal worlds are more ideal than other non-ideal worlds; it says: presupposing that one kills, then in those non-ideal worlds that best measure up to the deontically perfect worlds, one kills gently. In formalising CTD reasoning the key problem is formalisation of what is meant by ‘best measure up’ ” [Prakken and Sergot, 1997, p. 244]. This passage is quoted from the most detailed study currently available of the treatment of CTD-phenomena within a preference-based approach to the semantics of obligation sentences. The passage captures very succinctly the reason why a number of researchers in deontic logic have accepted the idea that an appropriate semantics for obligation sentences calls for an ordering on possible worlds, in terms of preference or relative goodness. As the authors note, the idea stems from Bengt Hansson [1971], with the later work of David Lewis [1974] providing a more comprehensive investigation of the semantical framework involved.

Prakken and Sergot’s paper, although in parts rather dense — it is best read as a sequel to [Prakken and Sergot, 1996] — explores in considerable depth the Hansson–Lewis analysis of dyadic deontic logic, assesses its shortcomings in regard to the treatment of CTDs, and offers a remedy for them which draws on techniques deriving from the study of default reasoning. We shall not attempt to give a summary of their final position, but confine ourselves to a few observations.

A significant feature of the Prakken and Sergot account of the Hansson–Lewis approach is that they add a notion of alethic necessity, the function of which is in part to clarify the nature of detachment properties. As was pointed out in Section 3.2, above, the characteristic feature of the DD-family of dyadic deontic logics (to which the Hansson–Lewis approach belongs) is that it validates the deontic detachment principle, but does not validate the factual detachment property. Expressed in terms of the notation used by Prakken and Sergot in [Prakken and Sergot, 1997], this means that

$$(DD) \quad O [B]A \rightarrow (O B \rightarrow O A)$$

is valid, whereas

$$(FD) \quad O [B]A \rightarrow (B \rightarrow O A)$$

is not. (Formulas of the form  $O [B]A$  express what Prakken and Sergot call *contextual obligations*, and they correspond to formulas of the form  $O (A/B)$  in the Lewis notation. The formula  $O A$  is an abbreviation of  $O [\top]A$ , where  $\top$  is any tautology.) However, since

$$(\Box O) \quad \Box B \rightarrow O B$$

is valid in the Prakken and Sergot extended version of Hansson-Lewis, (DD) immediately yields

$$(SFD) \quad O[B]A \rightarrow (\Box B \rightarrow O A)$$

Prakken and Sergot call this principle ‘strong factual detachment’, and they say of the alethic necessity operator that it expresses the notion ‘objectively settled’. They do not say much by way of further characterisation of ‘objectively’, except that they want to distinguish it from necessity in a subjective sense, “. . . such as when an agent decides to regard it as settled *for him* that there will be a fence” [Prakken and Sergot, 1997, p. 241]. They have the further interesting comment to make about the import of (SFD) for CTD contexts: “For CTD obligations this form of strong factual detachment seems very appropriate, but it must be read with extreme care. As long as it is possible to avoid violation of a primary obligation  $O \neg B$  a CTD obligation  $O[B]A$  remains restricted to the context; it is only if the violation of  $O \neg B$  is unavoidable, if  $\Box B$  holds, that the CTD obligation comes into full effect, pertains to the context  $\top$ ” [*loc.cit.*]. Note, however, that if  $\Box B$  holds, then  $O B$  holds, and thus — since the (D)-schema is valid in their system —  $\neg O \neg B$  holds. That is, going back to the dog-and-sign scenario, one could detach the obligation to put up a sign only in circumstances in which there was no longer an obligation not to have a dog!

These observations bring to the fore one of the most basic differences between our approach and that adopted by Prakken and Sergot: there is a quite fundamental disagreement between us regarding what an adequate theory of CTD scenarios should be expected to achieve. As we see matters, it is of paramount importance that the theory can show which actual obligations are derivable in circumstances of violation of some primary obligation, and can do so without also requiring that the sentence expressing the primary obligation must be false. A deontic logic which cannot show what actually ought to be done *in circumstances of violation* is, in our view, of limited interest. Prakken and Sergot’s theory, by contrast, belongs to a tradition in deontic logic which, it seems, takes the detachment of actual obligations to be a matter of no particular importance.

The notion of ‘settledness’ Prakken and Sergot employ is peculiar, at least with respect to its relation to the concept of obligation. How can that which is settled, unalterable, be obligatory? Surely that which is genuinely obligatory must be violable! Here again is a basic point of contrast between our approach and that of Prakken and Sergot. The principles  $(\neg O_a)$  and  $(\neg O_i)$  — Section 4.4, above — express, we believe, the correct connection between settledness and obligation concepts, and they also play a key role, as we saw in Section 5, in determining the consequences which can be drawn from various CTD scenarios.

Returning to Prakken and Sergot’s discussion of the Hansson–Lewis framework, it should be noted that a prime reason for their dissatisfaction with

that framework resides in the fact that it fails to capture a reading of the ‘extended’ considerate assassin scenario according to which that scenario is inconsistent. (This scenario was discussed above, Section 6.4, in response to some critical questions from Prakken. It extends the considerate assassin scenario by adding the further requirement that it is forbidden to offer cigarettes. See [Prakken and Sergot, 1997, § 6.1].) The underlying problem, as they diagnose it, is that the Hansson–Lewis semantics “. . . allows for the possibility of sub-ideal worlds but has very little to say about what they are like and nothing to say about how they compare with ideal worlds” [Prakken and Sergot, 1997, p. 250]. Their attempted solution involves a rather complex extension of the Hansson–Lewis framework, adapting techniques from the study of default reasoning in order to provide a means of ranking sub-ideal worlds with respect to the degree to which they measure up to ideal worlds.

We shall not here attempt to summarise these complexities. But from the point of view of comparison with our own theory, we observe in particular that Prakken and Sergot chose to impose on their investigation two constraints which we feel — for reasons discussed, in particular, in Section 6, above — are best rejected: they insist on retaining the (D)-schema for obligation sentences, and they refuse to abandon the consequential closure principle expressed by

$$\Box (A \rightarrow C) \rightarrow (O [B]A \rightarrow O [B]C)$$

(However, one of the factors which further complicates their account is that they also introduce a notion they call ‘explicit obligation’, which is *not* closed under consequence.)

Our rejection of the (D)-schema and consequential closure, our insistence on the importance of (a restricted form of) factual detachment, our exploitation of a distinction between actual and ideal obligations, and our characterisation of the relationship between settledness and obligation — all of these factors mark fundamental differences between our theory and that of Prakken and Sergot. But perhaps none of these constitutes the *most* fundamental difference. For we have not found it necessary at all to resort to the use of an explicit preference ordering in the semantics, in order to capture an adequate representation of the various CTD scenarios. Recall Lewis’ claim in [Lewis, 1974, pp. 3]: “A mere division of worlds into the ideal and the less-than-ideal will not meet our needs. We must use more complicated value structures that somehow bear information about comparisons or gradations of value.” The treatment of the Chisholm scenario in [Jones and Pörn, 1985] deliberately attempted to indicate that Lewis was wrong: the semantics used a “mere division” into two types of worlds, defined two accessibility relations pertaining to them, and defined some simple relations between these relations. But it imposed no ranking, no ordering, on the possible worlds. Yet it supplied the basis for an analysis of the Chisholm

set that met all the standard adequacy criteria . . . until Prakken and Sergot identified the “pragmatic oddity”. The question then became: is a preference ordering in the semantics essential in order to cope with the pragmatic oddity, and to cope with a broader class of CTD-phenomena, including but not limited to those exposed by the Chisholm set? In contrast to Prakken and Sergot, our answer to the last question is negative; the alternative strategy is to focus on another of the notions mentioned by Hansson which, as we have seen, also plays a role for both Loewer and Belzer and Prakken and Sergot: the notion of settledness or fixity of the facts. Our basic conjecture is that, properly characterised, and appropriately connected to the notions of actual and ideal obligation, the concept(s) of settledness provide the fundamental key to unravelling the tangled knot of CTD-problems.

Future research, we trust, will facilitate comparison of the preference-based approach and the approach which has formed the core of this chapter, with a view to furthering our understanding of the Contrary-to-Duty, and thus of normative reasoning itself.

#### POSTSCRIPT (SEPTEMBER, 2001)

The writing of this chapter was completed in 1999. Of relevant material that has been published since that time, we would like in particular to mention Makinson and van der Torre’s work on “input/output” logics, which the authors claim to be applicable to the treatment of CTD-problems; see Makinson and van der Torre [2000; 2001].

#### ACKNOWLEDGEMENTS

Some of the research here reported was carried out within the Portuguese research projects no. PCSH/C/OGE/1038/95-MAGO, and PCEX/P/MAT/46/96-ACL, and the ESPRIT Basic Research Working Group 8319 MODELAGE (“A Common Formal Model of Cooperating Intelligent Agents”). Andrew Jones also wishes to thank the PRAXIS XXI Programme of the Portuguese Research Council for its support in the Spring of 1996. Finally, the authors are very grateful to Henry Prakken, Donald Nute, Bjørn Kjos-Hanssen and Marek Sergot for their criticisms and comments.

At the time of completion of the writing of this chapter, José Carmo was at the Department of Mathematics, Instituto Superior Technico, Lisbon, Portugal; and Andrew Jones was at the Department of Philosophy and Norwegian Research Centre for Computers and Law, University of Oslo, Norway.

José Carmo

*Department of Mathematics, University of Madeira, Portugal.*

Andrew Jones

*Department of Computer Science, King’s College, London, UK.*



## BIBLIOGRAPHY

- [Alchourrón, 1993] C. E. Alchourrón. Philosophical foundations of deontic logic and the logic of defeasible conditionals. In *Deontic Logic in Computer Science: Normative System Specification*, J.-J. Ch. Meyer and R. J. Wieringa, eds. pp. 43–84. John Wiley and Sons, 1993.
- [al-Hibri, 1978] A. al-Hibri. *Deontic Logic: A Comprehensive Appraisal and a New Proposal*. University Press of America, Washington, DC, 1978.
- [Åqvist and Hoepelman, 1981] L. Åqvist and J. Hoepelman. Some theorems about a ‘tree’ system of deontic tense logic. In *New Studies in Deontic Logic*. R. Hilpinen, ed. pp. 187–221. D. Reidel, Dordrecht, 1981.
- [Belzer, 1987] M. Belzer. Legal reasoning in 3–D. In *Proceedings of the First International Conference in Artificial Intelligence and Law*. pp. 155–163. ACM Press, Boston, 1987.
- [Bratman, 1987] M. E. Bratman. *Intention, Plans and Practical Reason*. Harvard University Press, Cambridge, MA, 1987.
- [Brown, 1996] M. Brown. Doing as we ought: towards a logic of simply dischargeable obligations. In *Deontic Logic, Agency and Normative Systems (Proceedings of the Third International Workshop on Deontic Logic in Computer Science)*, M. Brown and J. Carmo, eds. pp. 47–65. Workshops in Computing Series, Springer, 1996.
- [Bull and Segerberg, 2001] R. A. Bull and K. Segerberg. Basic modal logic. In *Handbook of Philosophical Logic, Volume 3*, 2nd edition, D. M. Gabbay and F. Guenther, eds. pp. 1–81. D. Reidel, Dordrecht, 2001.
- [Carmo and Jones, 1994] J. Carmo and A. J. I. Jones. Deontic database constraints and the characterisation of recovery. In *Proceedings of the Second International Workshop on Deontic Logic in Computer Science (ΔEON’94)*, A. J. I. Jones and M. J. Sergot, eds. pp. 56–85. Complex 1/94 NRCCCL, Tano A. S., Oslo, 1994.
- [Carmo and Jones, 1995] J. Carmo and A. J. I. Jones. Deontic Logic and Different Levels of Ideality. RRD MIST 1/95, 1995.
- [Carmo and Jones, 1996] J. Carmo and A. J. I. Jones. Deontic database constraints, violation and recovery. *Studia Logica*, **57**, 139–165, 1996.
- [Carmo and Jones, 1997] J. Carmo and A. J. I. Jones. A new approach to contrary-to-duty obligations. In *Defeasible Deontic Logic*, D. Nute, ed. Synthese Library, pp. 317–344, 1997.
- [Chellas, 1974] B. J. Chellas. Conditional obligation. In *Logical Theory and Semantic Analysis*, Stenlund, ed. pp. 23–33. D. Reidel, Dordrecht, 1974.
- [Chellas, 1980] B. J. Chellas. *Modal Logic - An Introduction*. Cambridge University Press, Cambridge, 1980.
- [Chisholm, 1963] R. M. Chisholm. Contrary-to-duty imperatives and deontic logic. *Analysis*, **24**, 33–36, 1963.
- [ΔEON91, ] J.-J. Meyer and R. Wieringa, eds. *Proceedings of the First International Workshop on Deontic Logic in Computer Science (ΔEON’91)*, Amsterdam, 1991. Revised copies of selected papers appear in *Deontic Logic in Computer Science: Normative System Specification*, John Wiley and Sons, 1993, and in *Annals of Mathematics and Artificial Intelligence*, **9**, 1993.
- [ΔEON94, ] A. J. I. Jones and M. J. Sergot, eds. *Proceedings of the Second International Workshop on Deontic Logic in Computer Science (ΔEON’94)*, Complex 1/94 NRCCCL, Oslo, Tano A.S., 1994. Revised copies of selected papers appear in *Studia Logica*, **57**, 1996.
- [ΔEON96, ] M. Brown and J. Carmo, eds. *Deontic Logic, Agency and Normative Systems (Proceedings of the Third International Workshop on Deontic Logic in Computer Science)*, Springer, Workshops in Computing Series, 1996.
- [Elgesem, 1993] D. Elgesem. *Action Theory and Modal Logic*. PhD thesis, Dept. of Philosophy, University of Oslo, 1993.
- [Hamilton, 1978] A. G. Hamilton. *Logic for Mathematicians*. Cambridge University Press, 1978.

- [Hansson, 1971] B. Hansson. An analysis of some deontic logics. In *Deontic Logic: Introductory and Systematic Readings*, R. Hilpinen, ed. pp. 121–147. D. Reidel, Dordrecht, 1971. (2nd edn. 1981).
- [Hilpinen, 1993] R. Hilpinen. Actions in deontic logic. In *Deontic Logic in Computer Science - Normative System Specification*, J.-J. Ch. Meyer and R. J. Wieringa, eds. pp. 85–100. John Wiley and Sons, Chichester, UK, 1993.
- [Hintikka, 1975] J. Hintikka. Impossible possible worlds vindicated. *Journal of Philosophical Logic*, **4**, 475–484, 1975.
- [Hughes and Cresswell, 1968] G. E. Hughes and M. J. Cresswell. *A Companion to Modal Logic*. Methuen, London, 1968.
- [Jones, 1991] A. J. I. Jones. Intentions and the Logic of Norms. *First collection of papers from the ESPRIT Basic Research Action MEDLAR (“Mechanizing Deduction in the Logics of Practical Reasoning”)*, Dept. of Computing, Imperial College, London, 1991.
- [Jones, 1993] A. J. I. Jones. Towards a formal theory of defeasible deontic conditionals. *Annals of Mathematics and Artificial Intelligence*, **9**, 151–166, 1993.
- [Jones and Pörn, 1985] A. J. I. Jones and I. Pörn. Ideality, sub-ideality and deontic logic. *Synthese*, **65**, 275–290, 1985.
- [Jones and Sergot, 1992] A. J. I. Jones and M. J. Sergot. Deontic logic in the representation of law: towards a methodology. *Artificial Intelligence and Law*, **1**, 45–64, 1992.
- [Jones and Sergot, 1993] A. J. I. Jones and M. J. Sergot. On the characterisation of law and computer systems: the normative systems perspective. In *Deontic Logic in Computer Science - Normative System Specification*, J.-J. Ch. Meyer and R. J. Wieringa, eds. pp. 275–307. John Wiley and Sons, Chichester, UK, 1993.
- [Konolige and Pollack, 1993] K. Konolige and M. E. Pollack. A Representationalist theory of intention. In *Proceedings of IJCAI’93*, pp. 390–395, 1993.
- [Lewis, 1973] D. Lewis. *Counterfactuals*. Blackwell, Oxford, 1973.
- [Lewis, 1974] D. Lewis. Semantic Analysis for Dyadic Deontic Logic. In *Logical Theory and Semantic Analysis*. Stenlund, ed. pp. 1–14. D. Reidel, Dordrecht, 1974.
- [Lower and Belzer, 1983] B. Lower and M. Belzer. Dyadic deontic detachment. *Synthese*, **54**, 295–318, 1983.
- [Makinson and van der Torre, 2000] D. Makinson and L. van der Torre. Input/output logics. *Journal of Philosophical Logic*, **29**, 383–408, 2000.
- [Makinson and van der Torre, 2001] D. Makinson and L. van der Torre. Constraints for input/output logics. *Journal of Philosophical Logic*, **30**, 155–185, 2001.
- [McCarty, 1994] L. T. McCarty. Defeasible deontic reasoning. *Fundamenta Informaticae*, **21**, 125–148, 1994.
- [Mendelson, 1979] E. Mendelson. *Introduction to Mathematical Logic*. Van Nostrand, 2nd edn. 1979.
- [Meyer, 1988] J.-J. Ch. Meyer. A different approach to deontic logic: deontic logic viewed as a variant of dynamic logic. *Notre Dame Journal of Formal Logic*, **29**, 109–136, 1988.
- [Meyer, Wieringa and Dignum, 1997] J.-J. Ch. Meyer, R. J. Wieringa and F. P. M. Dignum. The role of deontic logic in the specification of information systems. In *Logics for Databases and Information Systems*, J. Chomicki and G. Saake, eds. Ch. 1, Kluwer Academic Publishers, Boston/Dordrecht/London, 1997.
- [Mott, 1973] P. L. Mott. On Chisholm’s Paradox. *Journal of Philosophical Logic*, **2**, 197–211, 1973.
- [Pörn, 1977] I. Pörn. *Action Theory and Social Science: Some Formal Models*. Synthese Library 120, D. Reidel, Dordrecht, Holland, 1977.
- [Prior, 1967] A. Prior. *Past, Present and Future*. Clarendon Press, Oxford, 1967.
- [Prakken and Sergot, 1994] H. Prakken and M. J. Sergot. Contrary-to-duty imperatives, defeasibility and violability. In *Proceedings of the Second International Workshop on Deontic Logic in Computer Science (ΔEON’94)*, A. J. I. Jones and M. J. Sergot, eds. pp. 296–318. Complex 1/94 NRCCCL, Tano A. S., Oslo, 1994.
- [Prakken and Sergot, 1996] H. Prakken and M. J. Sergot. Contrary-to-duty obligations and defeasible deontic reasoning. *Studia Logica*, **57**, 91–115, 1996.

- [Prakken and Sergot, 1997] H. Prakken and M. J. Sergot. Dyadic deontic logic and contrary-to-duty obligations. In *Defeasible Deontic Logic*. D. Nute, ed. Synthese Library, pp. 223–262, 1997.
- [Rescher and Urquhart, 1971] N. Rescher and A. Urquhart. *Temporal Logic*. Springer-Verlag, 1971.
- [Santos and Carmo, 1996] F. Santos and J. Carmo. Indirect action, influence and responsibility. In *Deontic Logic, Agency and Normative Systems (Proceedings of the Third International Workshop on Deontic Logic in Computer Science)*, M. Brown and J. Carmo, eds. pp. 194–215. Springer, Workshops in Computing Series, 1996.
- [Thomason, 1981] R. H. Thomason. Deontic logic as founded on tense logic. In *New Studies in Deontic Logic*. R. Hilpinen, ed. pp. 165–176. D. Reidel, Dordrecht, 1981.
- [Thomason, 2001] R. H. Thomason. Combinations of tense and modality. In *Handbook of Philosophical Logic*, 2nd edition, Volume 6, D. M. Gabbay and F. Guenther, eds. D. Reidel, Dordrecht, 2001.
- [Tomberlin, 1983] J. E. Tomberlin. Contrary-to-duty imperatives and Castañeda's system of deontic logic. In *Agent, Language, and the Structure of the World*, J. E. Tomberlin, ed. pp. 231–249, Hackett Publishing Co., Indianapolis, USA, 1983. (A reply by Castañeda appears at pp. 441–448 of the same volume.)
- [Tomberlin, 1986] J. E. Tomberlin. Good samaritans and Castañeda's system of deontic logic. In *Hector-Neri Castañeda*, J. E. Tomberlin, ed. pp. 255–272. D. Reidel, Dordrecht, Holland, 1986. (A reply by Castañeda appears at pp. 373–375 of the same volume.)
- [van Eck, 1981] J. van Eck. *A System of Temporally Relative Modal and Deontic Predicate Logic and its Philosophical Applications*, Ph. D. thesis, Rijksuniversiteit te Groningen, 1981.



# INDEX

- absolute *vs.* relative, *see* normative concept
- absurdity, *see* Falsum
- accessibility relation, 231
- Ackermann, W., 170
- act-utilitarianism, 198, 201
- action  
    logic of, 154, 197, 204
- action dimension of CTDs, 274, 332–336
- actual/potential possibility/necessity  
     $\diamond$ ,  $\diamond$ ,  $\boxplus$ ,  $\square$ , 285–288, 290, 291, 293–295, 297, 298, 317, 319–323, 326, 328–332, 337–340
- adequacy, *see* translation
- adequacy criteria (for a theory of commitment), 186
- adjunction, 177, 181, 229
- al-Hibri, A., 190
- alethic modal logic, 149, 151, 152, 156, 188  
    with a monadic  $Q$  connective, 152, 235, 236, 241  
    with a propositional constant, 149, 151, 179, 185, 217
- alethic models, *see* models
- almost reflexive relation, 209
- almost symmetric relation, 209, 210
- alphabet, 166, *see* also vocabulary, 205, 217
- alternative relation, *see* accessibility relation to an action
- ancestral, *see* chain (of a relation)
- Anderson, A. R., 148–150, 152, 153, 156–158, 160, 179, 184–187, 189, 204, 259
- Andersonian  
    definitions of deontic operators, 151  
    reduction of deontic logic to alethic logic, 179
- answer, 1
- Åqvist, L., 147, 149, 152, 159, 178, 189, 190, 202, 203, 216, 230, 236, 238–240, 243, 249, 250, 252, 254–257
- Åqvist, L., 5
- Åqvist, L., 147
- Åqvist, L., 15, 24, 31, 33, 49, 50
- Aristotle, 3, 150
- assignment (function), 214, 221, 251
- autonomy, 164
- auxiliary symbols, 161
- axiom, single *vs.* schema, 155
- axiomatic  
    point of view, 150
- axiomatic systems, *see* theories, proof theories
- Bailhache, P., 154
- Belnap's Analysis, 10
- Belnap's Hauptsatz, 22
- Belnap, N. D., 2, 4, 5, 7–10, 34, 37, 40, 41, 49–51
- Bentham, J., 150, 154
- Bergström, L., 199, 202
- Berkemann, J., 259
- bestness, *see* optimality
- betterness, 193
- bi-intuitionistic logic, 118
- biconditionality, 164
- Bolinger, D., 7

- Bolzano criterion, 153, 173–175, 181, 182
- Bolzano, B., 153, 173–175
- brackets, *see* auxiliary symbols
- Bromberger, S., 41
- ca-complete, 55
- ca-derivation, 55
- canonical model, 221
- Carnap, R., 3
- Castañeda theorem, 202
- Castañeda, H.-N., 200–203
- ceteris paribus proviso, 198, 199
- chain (proper ancestral) of a relation, 231
- Chellas, B. F., 185, 188, 200, 203, 232, 259
- Chisholm's Paradox, 151, 153, 190, 273–277, 279, 281, 282, 285, 287, 298, 299, 301, 302, 321, 324, 332–336, 339, 340
- theorem on, 192, 197
- Chisholm's paradox, 272, 282, 300, 332, 334
- Chisholm, R. M., 148, 151, 153, 190, 192, 197, 198, 204
- choice-offering *vs.* alternative-presenting (sense of disjunctive obligations and permissions), *see* free choice permission
- circumstance, 148, 201
- closure, 158, 159
- co-permissibility, *see* accessibility relation
- Cohen, F., 3
- coherence principal, 30
- coincidence lemma, 214, 216, 222
- commanding *vs.* commands, 148
- commitment, 148, 149, 181, 183–187, 189, 190, 197, 198
- paradox of, 148, 151, 153, 179, 180, 197
- completeness, 8, 16, 89
- completeness theorem, 152, 211, 212, 214, 221, 223, 249
- strong, 221, 222, 256
- weak, 211, 221, 255
- completeness-claim, 14, 15
- compounding, 18
- compounds, 52
- conclusive answer, 31
- conclusiveness, 32
- conditional obligation, 148
- conditional proof (rule of), *see* deduction theorem, implication introduction, 215
- conflict of obligations, 266, 270, 276, 320, 322, 323
- connective
- degree of, 162, 166, 168, 206
- dyadic (binary), 148
- main, 168
- monadic, 148–152
- zero-place, 232, 253
- consequential closure, 266–270, 314, 315, 339
- considerate assassin example, 284, 291, 310, 311, 316, 322, 339
- consistency, 166, 172, 190, 191, 193, 195, 198, 207, 218, 234, 253
- constant, *see* variable *vs.* constant, 152, 155, 161, 165, 166, 186, 187, 203, 219, 232
- Conte, A. G., 259
- contingency
- alethic, 157
- deontic, 157
- contradiction, *see* falsum
- correction, 28
- corrections-accumulating sequence, 49
- correctness, *see* truth, soundness theorem
- Cresswell, M., 44
- Cresswell, M. J., 218

- crucial lemma, 229  
 $\Delta$ EON workshops, 285  
 $\Delta$ EON workshops, 265  
 Dacey, R., 6  
 Dahlquist, T., 189  
 Danielsson, S., 152, 185, 188, 198, 200  
 decision theory, 179  
 deducibility, *see* derivability  
 deduction theorem, 196  
 deductive completeness, 10  
 deductive equivalence, 257, 258  
 defeasible obligations, 271, 282  
 definite descriptions, 205  
 definitional  
   enrichment of a language, 164, 165, 168, 204  
   extensions of a theory, 153, 157  
 definitions  
   theory of, 165  
 deliberative question, 43  
 denotational semantics of proofs, 110  
 deontic contingency, 157  
 deontic detachment, 273, 276, 297, 298, 331, 335, 337  
 deontic logic, 147  
   dyadic, 148, 151, 153  
   monadic, 148–152, 155, 169  
   natural *vs.* formal, 152, 153, 172, 205  
   reduction to alethic modal logic, *see* Andersonian  
   relationship of act-utilitarianism to, 198, 201  
   von Wright-type, 148, 149, 154, 155, 159, 179, 241, *see* Hansson dyadic systems, Smiley–Hanson monadic systems  
 deontic models, *see* models  
 deontic paraadoxex, *see* paradox
- deontic/epistemic paradox, 269, 270  
*Deontik* (Mally's), 150  
 depreivation of counterintuitive force, 183  
 derivability, 172, 191, 210, 218, 234, 253  
 derived obligation, *see* commitment  
 detachment, principle of, 156, *see* *Modus ponens*  
 di Bernardo, G., 259  
 dilemma, *see* paradox  
   Jephta, 198, 200, 201, 203  
   of commitment and detachment, 150  
***DIntKt***, 103  
 direct answer, 15  
 direct answers, 1, 29  
 discharging of hypotheses, *see* implication introduction, negation introduction, disjunction elimination, existential instantiation  
 disjunction elimination, 215  
 disjunction introduction, 163, 205  
 display logic, 81  
 distinctness-claims, 15  
***DKt***, 89  
 dog example, 274, 279–286, 288, 302, 304, 320, 335, 338  
 doxastic logic, 270  
 duty, conflict of, 151  
 dyadic deontic logic, 276, 285, 286, 289, 290, 292, 294–298, 317–323, 330, 331, 337–340  
 dynamic logic, 335, 336
- easy lemma, 228, 230, 231, 236  
 effectiveness, 8  
 effectivity, 16  
 Egli, U., 4, 10  
 elementary questions, 12  
 elementary-like questions, 17  
 epistemic analysis of questions, 24

- equivalence (material), 163  
 equivalence relation, 235, 236  
 erotetic, 2  
 erotetic logic, 2, 51  
 ethics (ethical theory), 150, 201, 204, 259  
 Euathlus, 147  
 Euclidean relation, 209–211, 220  
 existential generalisation, 177, 181, 215, 229  
 existential instantiation, 215  
 expressive completeness, 9  
 expressive incompleteness, 9  
 expressive resources of deontic languages, 150  
 extension, *see* definitional, truth-set  
  
 factual detachment, 273, 276, 295, 321, 331, 333, 335–339  
 faithfulness, *see* representation  
 fallacy of many questions, 35  
*falsum*, 155, 163, 192, 205, 247  
 Ficht, H., 4  
 Finn, V., 38  
 Fisher, M., 154  
 fixed/unalterable/settled facts, 283–288, 294, 316, 330, 338–340  
 forbiddance, *see* prohibition  
 formalization, 153, *see* translation, 161, 169, 170, 183, 184, 190, 191, 197, 204  
 formula(s), *see* sentences  
 formulas-as-types for temporal logics, 101  
 fragment (generated by definitions), *see* representability, representation  
     deontic, 150, 151, 223–225, 233, 236, 241  
     of normative English, 153, 167, 168, 170, 171, 175, 179, 183  
  
 free choice permission, 178, 269  
 full answer, 31  
 function from sentences into binary relations, 168, 193, 208, 219, 232  
 Føllesdal, D., 150, 158, 174, 175, 178, 185, 190, 197, 259  
  
 Gärdenfors, P., 156  
 gaggle theory, 129  
 general erotetic logic, 51, 52  
 gentle murderer example, 307–309, 315, 330  
 Gentzen system, 62  
 Gentzen systems for normal modal logics, 63  
 Gentzen terms, 62  
 genuine *vs.* spurious deontic sentences, 154  
*Gestalt*, 162  
 GMA ('Give Me an Answer'), 34  
 good samaritan, 269, 316  
 Gornstein, I., 4  
 Grewendorf, G., 41  
 Groenendijk, J., 2, 7, 32, 37, 44  
 grouping, ambiguity of, 169  
**GS5**, 80  
 guarding, 28  
 guarding a risky question, 28  
  
 Halonen, I., 42  
 Hamblin's Postulate, 37, 43, 44  
 Hamblin's Postulates, 5  
 Hamblin, C., 2, 3, 5, 34  
 Hanson, W. H., 150, 151, 154, 207  
 Hansson systems of  
     dyadic deontic logic, 152, 185, 193, 236, 241, 244, 247  
     partial syntactic identification, 248  
     semantic identification, 248  
 Harrah, D., 1, 9, 21, 35, 52, 53, 57  
 Harris, S., 32  
 Hedenius, I., 154



- Henkin technique in modal logic,  
151, 225
- Henkin, L. A., 151, 225
- Hiž, H., 4, 23, 38
- Higginbotham, J., 4, 36, 37, 44,  
51
- higher-arity sequent systems, 75
- higher-dimensional sequent systems,  
74
- higher-level sequent systems, 73
- Hilbert, D., 170
- Hilpinen, R., 150, 158, 174, 175,  
178, 185, 190, 197, 259
- HIIntKt**, 135
- Hintikka deontic consequence, 159
- Hintikka's development, 31
- Hintikka, J., 5, 25, 31, 32, 40, 42,  
150, 159, 199, 200, 203
- historical possibility (necessity), 149,  
217
- Hoepelman, J., 203
- Hohfeld, W. N., 204
- how questions, 43
- Hughes, G. E., 218
- Hull, R., 37
- hypersequents, 80
- hypothesis, *see* discharging
- ideal/actual obligations,  $O_i A/O_a A$ ,  
271, 273, 278, 283–286,  
288–297, 314–317, 319–  
323, 328–331, 338–340
- identity, 171, 203
- imperative *vs.* descriptive inter-  
pretation of deontic sen-  
tences, 154
- imperatives  
contrary-to-duty, 148, 153, 190,  
192, 195, 197, 203  
logic of, 148, 259
- implication introduction, 196, *see*  
conditional proof (rule of)
- implying, 44
- improved formalization, 183, 184
- incompleteness, 249
- independence, 195
- individual acts, 203
- inductive definition, *see* recursive  
definition
- inferential equivalence, *see* deduc-  
tive equivalence
- information-retrieval, 51
- internal *vs.* external deontic sen-  
tences, 154
- interpretation, 148, 149, 154, 163,  
183, *see* assignment, mod-  
els, truth, imperative *vs.*  
descriptive, internal *vs.*  
external, modal *vs.* fac-  
tual, genuine *vs.* spuri-  
ous
- interrogative, 1
- interrogative model of inquiry, 32
- interrogative operators, 26
- interrogatives, 26
- introduction rules, 86
- introduction schemata, 68
- IQW ('It Is the Question Whether'),  
38
- iteration, 247
- Jephta, *see* dilemma
- Kalinowski, G., 159
- Kalish, D., 176, 181
- Kamp, J. A. W., 178
- Kampé de Fériet, J., 50
- Kanger, H., 204
- Kanger, S., 150, 151, 175, 186,  
201, 203, 204, 217, 235,  
236
- Karttunen, L., 7
- Keenan, E., 37
- Kiefer, F., 7, 32
- Knuuttila, S., 150, 200
- Koura, A., 41
- Kracht's algorithm, 96

- Kripke, S., 34, 150, 151, 218, 241, 247
- Kt**, 97
- Kubiński, T., 3, 10, 15, 17, 23, 24, 44, 46
- LS5**, 66
- language, *see* alphabet, vocabulary
- of alethic modal logic with **Q** connective, 151, 152
  - of dyadic deontic logic, 151, 152, 159, 186, 187
  - of monadic deontic logic, 152, 161, 169, 173, 175
  - of normative English, 153, 167
- legal theory, 149, 259
- Lehnert, W., 4, 44
- Leibniz, G. W., 150
- lemma
  - on relations, 230, 231, 236
  - on **G**, 255
- Lemmon, E. J., 151, 218, 222
- Lenzen, W., 147, 203
- Leszko, R., 48
- Lewis, D., 33, 152, 178, 185
- Lewis, S., 33
- limit assumption (limitedness), 248
- Lindahl, L., 150, 204
- Lindenbaum lemma, 213, 221, 251
- Lindenbaum, A., 213, 221, 251
- logic, *see* alethic modal logic, deontic logic, predicate logic, preference theory, action
- logic of action, 270, 332, 334–336
- logic of intention, 270
- logical
  - analysis, 149
  - connectives, 206
  - consequence, 159, 172, 176–178
  - dictionary, 163
  - validity (truth), 153, 163, 171–174, 176, 181
- logics, *see* theories
- LS4**, 69
- main connective, *see* principal sign
- Makinson lemma, 213, 215, 221
- Makinson, D., 151, 156, 203, 213, 215, 218, 222
- Mally, E., 150, 158
- Materna, P., 37
- matrix method, 225
- maximal (complete) consistent set, *see* saturated sets
- maximality under a preference relation, *see* optimality
- Mayo, B., 43
- McGuinness, F., 200
- mixed sentences, 51
- MMB (‘Make Me Believe’), 33
- MMK (or ‘Make Me Know’) approach, 24
- modal logic, *see* alethic modal logic
- modal *vs.* factual interpretation of normative sentences, 154
- modality, *see* predicate *vs* modality
- model-theoretical semantics, 150, 183, 203
- models
  - $\succ$ -supplemented deontic, 244
  - $\succ$ -supplemented dyadic deontic, 245
  - classification of, 209
  - classification of alethic, 219
  - deontic, 242
- modus ponens*, 156, 196, 206, 212, 215, 217, 233, 237, 249, 253, 256
- modus tollens*, 182, 189
- monadic *vs.* dyadic, *see* deontic logic, normative concept
- Montague, R., 150, 176, 181, 200, 203
- Moritz, M., 156

- morphology, *see* alphabet, vocabulary
- Mott, P. L., 190
- multiple-sequent systems, 78
- natural deontic logic, 153, 171
- natural *vs.* restricted sentence-set (over a vocabulary), 153, 171, 172
- necessitation (inference rule of), 159
- N**, 159, 249, 253
- O**, 155, 156, 206, 213
- OB**, 233
- negation introduction, 215
- negative requirements, *see* unprovable
- nominally true, 20
- non-normal operators, 277
- normal deontic logic, 155, 157
- normative concept
- absolute *vs.* relative, 148
  - monadic *vs.* dyadic, 148
  - unconditional *vs.* conditional, 148
- notation (for dyadic normative concepts)
- probabilistic, 186
  - sententially indexed modality, 232
  - standard binary connective, 232
- Nute, D., 178
- obedience *vs.* disobedience, 178
- obligation
- prima facie vs.* actual, 198, 199
  - conditional, *see* commitment
  - expressions of, 148, 178
- Oppenheim, F., 149
- opt-serial relation, 219, 220, 231
- optimal obligation and permission, 217
- optimality (bestness), 217
- or, 18
- other-things-being-equal, *see ceteris paribus* proviso
- ought-implies-can, 198, 199
- overrule, 199
- Pörn, I., 204
- Paradox
- Chisholm, 151
  - Good Samaritan, 173, 198
  - Morning Star, 201, 203
  - Protagoras, 147
  - Ross, 153
- partial answer, 44
- partial answers, 1
- penalty, 148, 185
- permission, 148, *see* commitment, deontic logic (dyadic), 152, 185
- expressions of, 151, 154, 163, 178, 186
- Picard, C., 50
- Polish notation, 169
- positive law, 148
- possible worlds semantics, *see* model-theoretical semantics
- power-relation, *see* chain (proper ancestral) of a relation
- Powers, L., 185, 188, 190
- pragmatic oddity, 278–279, 288, 300, 321, 334, 335
- Prawitz, D., 202
- predicament, *see* duty, conflict of
- predicate logic, 126
- predicate logic, temporally relative, 203
- predicate *vs.* modality, 159
- preference
- logic of, 217, 256
  - relation, 152, 202, 217, 241, 242, 254
  - theory, 152, 217
- preference-based semantics, 276, 331, 337–340

- prescription, 148  
 presumptions, 56  
 presupposes, 20  
 presupposition, 20, 28  
 Price, R., 199  
 principal sign, 168  
 Prior, A. N., 2, 3, 148, 151, 153,  
     156, 157, 159, 179, 184,  
     200, 203  
 Prior, M., 2, 3  
 prohairetic, 217, 218, 232  
 prohibition  
     expressions of, 148, 151, 154,  
     157  
 proof theory, 205, 206, 217, 237,  
     253  
 proof-theoretical point of view, *see*  
     axiomatic point of view  
 propositional  
     constant, 148, 149, 151, 152,  
     162, 165, 179, 217, 242  
     letter, 152, 155, 156, 161, 164,  
     167, 169, 216  
     parameter, 167  
     variable, 162, 165  
     von Wright-type deontic logic,  
     148, 154, 155, 159  
 Protagoras, 147, 203  
 provability, 172, 173, 186, 207, 218,  
     234, 253  
 prudence, safety, risk in embouli-  
     atic logic, 179  
 Purtil, R. L., 199  
  
**Q**-connective  
     monadic, 151, 235  
     zero-place, 232, 253  
 QIE (quantified imperative-epistemic)  
     logic, 25  
 quantification, 203  
 queriables, 12  
 question trees, 48  
 question-types, 1  
 questions, 1  
  
 questions as context descriptions,  
     34  
 questions as hyper-complete enti-  
     ties, 38  
 questions as incomplete entities,  
     37  
 questions as intensional entities,  
     35  
 questions as their presuppositions,  
     34  
 Quine, W. V. O., 174  
 quotation marks, 168  
  
 raising, 44  
 rational reconstruction, 149  
 reading, 162  
 real answer, 16  
 really true, 20  
 recursive definition, 162, 179, 208,  
     257  
*reductio ad absurdum*, 178  
 reduction of questions, 46  
 relative product, 231  
 relevant replies, 53, 54  
 reply, 1  
 representability, 217, 256  
 representation, 217, 238  
     faithful, 174  
     left-to-right faithful, 174  
     right-to-left faithful, 174  
 requirement of consistency andnon-  
     redundancy, 190  
 Rescher, N., 148, 150, 178, 185,  
     259  
 residuation, 81  
 Reykjavik example, 311–314, 318  
 riskiness, 28  
 risky, 22, 28  
 Robison, J., 204  
 Rosetus, R., 200  
 Ross paradox, 268, 269, 314, 315  
 Ross, A., 153, 165, 175  
 Ross, Sir D., 199  
 rule of proof (inference), 156, 161

- Russell, B., 147
- S5**, 65
- SA reduction of questions, 5
- safe, 22, 28
- sanction, 148, 151, 185
- satisfiability, 210, 219, 235, 238
- saturated sets, 151, 212  
lemma on, 212
- saturation Lemma for canonical models, 214, 216, 222, 223
- Schleichert, H., 4
- Schliechert, H., 10
- Schmidt-Radefeldt, J., 44, 48
- Scott, D., 151, 159, 218, 222, 236
- Segeberg, K., 156, 185
- Sellars, w., 190
- semantic entailment, 238
- semantic tableaux, 151
- semantics, *see* model theoretical semantics
- sentences  
basic, 154, 161, 163, 170  
set of all well formed, 154, 157, 161, 231, 232, 247, 253
- sequences of questions, 48
- sequencing, 48
- serial relation, 219, 231
- set-of-answers methodology, 4, 5
- Smiley, T. J., 150, 151
- Smiley–Hanson monadic systems, 150–153, 155, 159, 161, 163, 169, 173, 176, 179, 180, 204–206, 210, 216
- Smullyan, R. M., 147, 203
- soundness theorem, 194, 195, 210, 211, 249
- speech act theory, 150
- Spohn, W., 185, 197
- square of opposition, 149
- Stahl, G., 3, 6, 7, 19, 49
- standard deontic logic, 266–270, 272–274, 277, 317
- Steel, T., 2, 4, 5, 9, 10, 40, 41, 49, 51
- Stenius, E., 154, 156, 157
- Stokhof, M., 2, 7, 32, 37, 44
- Strasser, P., 259
- strict implication, 185
- strict preference, *see* betterness
- strong cut-elimination, 90
- strongfactual detachment, 337, 338
- strong normality (of deontic logics), 152, 156–159
- strong *vs.* weak permission, 179
- structural rules, 87
- stylistic elegance, 167
- stylistic variant, 176, 181
- substitution, 155, 156, 165, 167
- substructure property, 70
- sufficient reply, 56, 57
- Suppes, P., 164, 166
- suppressing, 44
- symbolization, *see* formalization
- symmetric relation, 219, 220, 235
- Szaniawski, K., 6
- temporal  
deontic logic, 203  
resources of expression, 150, 153, 157
- temporal dimension of CTDs, 274, 279, 283, 324–331
- temporally dependent (relative) possibility (necessity), 199, 200
- theories, 148, 153, 162
- thesishood, *see* provability
- Thomason, R. H., 203
- Tichý, P., 5, 7, 9, 35, 37, 51
- TMT (‘Tell Me Truly’), 33
- Todt, G., 44, 48
- Tomberlin, J. E., 190, 200
- transitive relation, 209, 219, 220, 235, 243
- translation, 153, 161, 169, 173, 223, 236, 258

- from natural language into formal, 153
- fully adequate, 174, 175, 177, 181
- left-to-right adequacy, 173, 174, 197, 204, 205
- right-to-left adequacy, 173, 174, 177, 229
- translation of hypersequents, 125
- translation of multiple-sequent systems, 124
- translation theorem
  - for monadic deontic logic (Smiley), 225
  - for dyadic deontic logic (Åqvist), 152, 236
  - for monadic deontic logic (Smiley), 151, 152, 225
- true, 20
- truth, 20
  - condition (definition), 183, 193, 208, 222, 235, 237, 248
  - function, 155, 162
  - set, 164, 194, 234, 239, 253
- truth value analysis of interrogatives, 39
- truth-or-falsity of deontic sentences, 154
- turn rules, 82
- typed  $\lambda$ -calculus  $\lambda_t$ , 105
- types of answer, 21
- types of elementary-like question, 23
- types of question, 40
- types of reply, 43
- unsoundness, 249
- ungetrenntes ganzes*, 170
- universal generalization, 223
- universal instantiation, 215
- universal necessity and possibility, 159, 236, 237, 247
- univocity, 16
- unprovables, 158
- validity, 149, 153, 163, 171–174, 188, 210, 219, 235, 238, 250
- valuation (function), *see* assignment
- van Eck, J., 150, 153, 159, 188, 190, 197–200, 202, 203, 232
- van Fraassen, B. C., 152, 185, 188, 200
- variable *vs.* constant, 149
- vectored sentences, 52
- verification lemma, 214, 216, 222, 223, 252
- verum*, 155, 163, 187, 205, 247
- violation, 265, 269, 271, 277, 282, 288, 314, 315, 317–319, 322, 329–331, 333, 338
- vocabulary, 252, 256
- von Kutschera, F., 152, 185, 260
- von Wright, G. H., 148, 149, 153, 154, 156, 157, 161, 178, 179, 184–187, 190, 194, 200, 203, 204, 259
- Wachowicz, K., 24
- Wedberg, A., 149
- what, 40
- Wheatley, J., 43
- whether-questions, 12
- which-questions, 12
- white fence example, 274, 305, 306, 315, 331, 335
- who, 40
- why, 40
- why questions, 41, 42
- Wiśniewski, A., 2, 4, 10, 24, 44, 45, 47

# Handbook of Philosophical Logic

2nd Edition

Volume 9

edited by Dov M. Gabbay and F. Guentner





## CONTENTS

Editorial Preface	vii
<b>Dov M. Gabbay</b>	
Rewriting Logic as a Logical and Semantic Framework	1
<b>N. Martí-Oliet and J. Meseguer</b>	
Logical Frameworks	89
<b>D. Basin and S. Matthews</b>	
Proof Theory and Meaning	165
<b>Goran Sundholm</b>	
Goal Directed Deductions	199
<b>Dov M. Gabbay and Nicola Olivetti</b>	
On Negation, Completeness and Consistency	287
<b>Arnon Avron</b>	
Logic as General Rationality: A Survey	321
<b>Ton Sales</b>	
Index	367



## PREFACE TO THE SECOND EDITION

It is with great pleasure that we are presenting to the community the second edition of this extraordinary handbook. It has been over 15 years since the publication of the first edition and there have been great changes in the landscape of philosophical logic since then.

The first edition has proved invaluable to generations of students and researchers in formal philosophy and language, as well as to consumers of logic in many applied areas. The main logic article in the Encyclopaedia Britannica 1999 has described the first edition as ‘the best starting point for exploring any of the topics in logic’. We are confident that the second edition will prove to be just as good!

The first edition was the second handbook published for the logic community. It followed the North Holland one volume *Handbook of Mathematical Logic*, published in 1977, edited by the late Jon Barwise. The four volume *Handbook of Philosophical Logic*, published 1983–1989 came at a fortunate temporal junction at the evolution of logic. This was the time when logic was gaining ground in computer science and artificial intelligence circles.

These areas were under increasing commercial pressure to provide devices which help and/or replace the human in his daily activity. This pressure required the use of logic in the modelling of human activity and organisation on the one hand and to provide the theoretical basis for the computer program constructs on the other. The result was that the *Handbook of Philosophical Logic*, which covered most of the areas needed from logic for these active communities, became their bible.

The increased demand for philosophical logic from computer science and artificial intelligence and computational linguistics accelerated the development of the subject directly and indirectly. It directly pushed research forward, stimulated by the needs of applications. New logic areas became established and old areas were enriched and expanded. At the same time, it socially provided employment for generations of logicians residing in computer science, linguistics and electrical engineering departments which of course helped keep the logic community thriving. In addition to that, it so happens (perhaps not by accident) that many of the Handbook contributors became active in these application areas and took their place as time passed on, among the most famous leading figures of applied philosophical logic of our times. Today we have a handbook with a most extraordinary collection of famous people as authors!

The table below will give our readers an idea of the landscape of logic and its relation to computer science and formal language and artificial intelligence. It shows that the first edition is very close to the mark of what was needed. Two topics were not included in the first edition, even though

they were extensively discussed by all authors in a 3-day Handbook meeting. These are:

- a chapter on non-monotonic logic
- a chapter on combinatory logic and  $\lambda$ -calculus

We felt at the time (1979) that non-monotonic logic was not ready for a chapter yet and that combinatory logic and  $\lambda$ -calculus was too far removed.<sup>1</sup> Non-monotonic logic is now a very major area of philosophical logic, alongside default logics, labelled deductive systems, fibring logics, multi-dimensional, multimodal and substructural logics. Intensive re-examinations of fragments of classical logic have produced fresh insights, including at time decision procedures and equivalence with non-classical systems.

Perhaps the most impressive achievement of philosophical logic as arising in the past decade has been the effective negotiation of research partnerships with fallacy theory, informal logic and argumentation theory, attested to by the Amsterdam Conference in Logic and Argumentation in 1995, and the two Bonn Conferences in Practical Reasoning in 1996 and 1997.

These subjects are becoming more and more useful in agent theory and intelligent and reactive databases.

Finally, fifteen years after the start of the Handbook project, I would like to take this opportunity to put forward my current views about logic in computer science, computational linguistics and artificial intelligence. In the early 1980s the perception of the role of logic in computer science was that of a specification and reasoning tool and that of a basis for possibly neat computer languages. The computer scientist was manipulating data structures and the use of logic was one of his options.

My own view at the time was that there was an opportunity for logic to play a key role in computer science and to exchange benefits with this rich and important application area and thus enhance its own evolution. The relationship between logic and computer science was perceived as very much like the relationship of applied mathematics to physics and engineering. Applied mathematics evolves through its use as an essential tool, and so we hoped for logic. Today my view has changed. As computer science and artificial intelligence deal more and more with distributed and interactive systems, processes, concurrency, agents, causes, transitions, communication and control (to name a few), the researcher in this area is having more and more in common with the traditional philosopher who has been analysing

---

<sup>1</sup>I am really sorry, in hindsight, about the omission of the non-monotonic logic chapter. I wonder how the subject would have developed, if the AI research community had had a theoretical model, in the form of a chapter, to look at. Perhaps the area would have developed in a more streamlined way!

such questions for centuries (unrestricted by the capabilities of any hardware).

The principles governing the interaction of several processes, for example, are abstract and similar to principles governing the cooperation of two large organisations. A detailed rule based effective but rigid bureaucracy is very much similar to a complex computer program handling and manipulating data. My guess is that the principles underlying one are very much the same as those underlying the other.

I believe the day is not far away in the future when the computer scientist will wake up one morning with the realisation that he is actually a kind of formal philosopher!

The projected number of volumes for this Handbook is about 18. The subject has evolved and its areas have become interrelated to such an extent that it no longer makes sense to dedicate volumes to topics. However, the volumes do follow some natural groupings of chapters.

I would like to thank our authors and readers for their contributions and their commitment in making this Handbook a success. Thanks also to our publication administrator Mrs J. Spurr for her usual dedication and excellence and to Kluwer Academic Publishers for their continuing support for the Handbook.

Dov Gabbay  
King's College London

Logic	IT			
	Natural language processing	Program control specification, verification, concurrency	Artificial intelligence	Logic programming
<b>Temporal logic</b>	Expressive power of tense operators. Temporal indices. Separation of past from future	Expressive power for recurrent events. Specification of temporal control. Decision problems. Model checking.	Planning. Time dependent data. Event calculus. Persistence through time—the Frame Problem. Temporal query language. temporal transactions.	Extension of Horn clause with time capability. Event calculus. Temporal logic programming.
<b>Modal logic. Multi-modal logics</b>	generalised quantifiers	Action logic	Belief revision. Inferential databases	Negation by failure and modality
<b>Algorithmic proof</b>	Discourse representation. Direct computation on linguistic input	New logics. Generic theorem provers	General theory of reasoning. Non-monotonic systems	Procedural approach to logic
<b>Non-monotonic reasoning</b>	Resolving ambiguities. Machine translation. Document classification. Relevance theory	Loop checking. Non-monotonic decisions about loops. Faults in systems.	Intrinsic logical discipline for AI. Evolving and communicating databases	Negation by failure. Deductive databases
<b>Probabilistic and fuzzy logic</b>	logical analysis of language	Real time systems	Expert systems. Machine learning	Semantics for logic programs
<b>Intuitionistic logic</b>	Quantifiers in logic	Constructive reasoning and proof theory about specification design	Intuitionistic logic is a better logical basis than classical logic	Horn clause logic is really intuitionistic. Extension of logic programming languages
<b>Set theory, higher-order logic, <math>\lambda</math>-calculus, types</b>	Montague semantics. Situation semantics	Non-well-founded sets	Hereditary finite predicates	$\lambda$ -calculus extension to logic programs

<b>Imperative vs. declarative languages</b>	<b>Database theory</b>	<b>Complexity theory</b>	<b>Agent theory</b>	<b>Special comments: A look to the future</b>
Temporal logic as a declarative programming language. The changing past in databases. The imperative future	Temporal databases and temporal transactions	Complexity questions of decision procedures of the logics involved	An essential component	Temporal systems are becoming more and more sophisticated and extensively applied
Dynamic logic	Database updates and action logic	Ditto	Possible actions	Multimodal logics are on the rise. Quantification and context becoming very active
Types. Term rewrite systems. Abstract interpretation	Abduction, relevance	Ditto	Agent's implementation rely on proof theory.	
	Inferential databases. Non-monotonic coding of databases	Ditto	Agent's reasoning is non-monotonic	A major area now. Important for formalising practical reasoning
	Fuzzy and probabilistic data	Ditto	Connection with decision theory	Major area now
Semantics for programming languages. Martin-Löf theories	Database transactions. Inductive learning	Ditto	Agents constructive reasoning	Still a major central alternative to classical logic
Semantics for programming languages. Abstract interpretation. Domain recursion theory.		Ditto		More central than ever!

<b>Classical logic. Classical frag- ments</b>	Basic back- ground lan- guage	Program syn- thesis	A basic tool	
<b>Labelled deductive systems</b>	Extremely use- ful in modelling		A unifying framework. Context theory.	Annotated logic programs
<b>Resource and substructural logics</b>	Lambek calcu- lus		Truth maintenance systems	
<b>Fibring and combining logics</b>	Dynamic syn- tax	Modules. Combining languages	Logics of space and time	Combining fea- tures
<b>Fallacy theory</b>				
<b>Logical Dynamics</b>	Widely applied here			
<b>Argumentation theory games</b>		Game seman- tics gaining ground		
<b>Object level/ metalevel</b>			Extensively used in AI	
<b>Mechanisms: Abduction, default relevance</b>			ditto	
<b>Connection with neural nets</b>				
<b>Time-action- revision mod- els</b>			ditto	



	Relational databases	Logical complexity classes	The workhorse of logic	The study of fragments is very active and promising.
	Labelling allows for context and control.		Essential tool.	The new unifying framework for logics
Linear logic			Agents have limited resources	
	Linked databases. Reactive databases		Agents are built up of various fibred mechanisms	The notion of self-fibring allows for self-reference
				Fallacies are really valid modes of reasoning in the right context.
			Potentially applicable	A dynamic view of logic
				On the rise in all areas of applied logic. Promises a great future
			Important feature of agents	Always central in all areas
			Very important for agents	Becoming part of the notion of a logic
				Of great importance to the future. Just starting
			A new theory of logical agent	A new kind of model



## REWRITING LOGIC AS A LOGICAL AND SEMANTIC FRAMEWORK

### 1 INTRODUCTION

The relationships between logic and computation, and the mutual interactions between both fields, are becoming stronger and more pervasive than they have ever been. In fact, our way of thinking about both logic and computation is being altered quite strongly. For example, there is such an increasingly strong connection—in some cases to the point of complete identification—between computation and deduction, and such impressive progress in compilation techniques and computing power, that the frontiers between logical systems, theorem provers, and declarative programming languages are shifting and becoming more and more tenuous, with each area influencing and being influenced by the others.

Similarly, in the specification of languages and systems there is an increasing shift from mathematically precise but somewhat restricted formalisms towards specifications that are not only mathematical, but actually logical in nature, as exemplified, for example, by specification formalisms such as algebraic specifications and structural operational semantics. In this way, languages and systems that in principle may not seem to bear any resemblance to logical systems and may be completely “conventional” in nature, end up being conceptualized primarily as *formal* systems.

However, any important development brings with it new challenges and questions. Two such questions, that we wish to address in this paper are:

- *How can the proliferation of logics be handled?*
- *Can flexible logics allowing the specification and prototyping of a wide variety of languages and systems with naturalness and ease be found?*

Much fruitful research has already been done with the aim of providing adequate answers to these questions. Our aim here is to contribute in some measure to their ongoing discussion by suggesting that rewriting logic [Meseguer, 1992] seems to have particularly good properties recommending its use as both a *logical framework* in which many other logics can be represented, and as a general *semantic framework* in which many languages and systems can be naturally specified and prototyped.

### 1.1 *Rewriting logic as a logical framework*

In our view, the main need in handling the proliferation of logics is primarily conceptual. What is most needed is a *metatheory* of logics helping us to better understand and explore the boundaries of the “space” of all logics, present and future, and to relate in precise and general ways many of the logics that we know or wish to develop.

Following ideas that go back to the original work of Goguen and Burstall [1984] on *institutions*, we find very useful understanding the space of all logics as a *category*, with appropriate translations between logics as the arrows or morphisms between them. The work on institutions has been further developed by their original proponents and by others [Goguen and Burstall, 1986; Goguen and Burstall, 1992; Tarlecki, 1984; Tarlecki, 1985], and has influenced other notions proposed by different authors [Mayoh, 1985; Poigné, 1989; Fiadeiro and Sernadas, 1988; Meseguer, 1989; Harper *et al.*, 1989a; Salibra and Scollo, 1993; Ehrig *et al.*, 1991; Astesiano and Cerioli, 1993]. Some of the notions proposed are closely related to institutions; however, in other cases the main intent is to substantially expand the primarily model-theoretic viewpoint provided by institutions to give an adequate treatment of proof-theoretic aspects such as entailment and proof structures. The theory of general logics [Meseguer, 1989] that we present in summary form in Section 2 is one such attempt to encompass also proof-theoretic aspects, and suggests not just one space or category of logics, but several, depending on the proof-theoretic or model-theoretic aspects that we wish to focus on.

In our view, the quest for a *logical framework*, understood as a logic in which many other logics can be represented, is important but is not the primary issue. Viewed from the perspective of a general space of logics, such a quest can in principle—although perhaps not in all approaches—be understood as the search within such a space for a logic  $\mathcal{F}$  such that many other logics  $\mathcal{L}$  can be represented in  $\mathcal{F}$  by means of mappings  $\mathcal{L} \rightarrow \mathcal{F}$  that have particularly nice properties such as being conservative translations.

Considered in this way, and assuming a very general axiomatic notion of logic and ambitious enough requirements for a framework, there is in principle no guarantee that such an  $\mathcal{F}$  will necessarily be found. However, somewhat more restricted successes such as finding an  $\mathcal{F}$  in which all the logics of “practical interest,” having finitary presentations of their syntax and their rules, can be represented can be very valuable and can provide a great economy of effort. This is because, if an implementation for such a framework logic exists, it becomes possible to implement through it all the other “object logics” that can be adequately represented in the framework logic.

Much work has already been done in this area, including the Edinburgh logical framework LF [Harper *et al.*, 1993; Harper *et al.*, 1989; Gardner, 1992] and meta-theorem provers such as Isabelle [Paulson, 1989],  $\lambda$ Prolog

[Nadathur and Miller, 1988; Felty and Miller, 1990], and Elf [Pfenning, 1989], all of which adopt as framework logics different variants of higher-order logics or type theories. There has also been important work on what Basin and Constable [1993] call *metallogical* frameworks. These are frameworks supporting reasoning about the metalogical aspects of the logics being represented. Typically, this is accomplished by reifying as “data” the proof theory of the logic being represented in a process that is described in [Basin and Constable, 1993] as *externalizing* the logic in question. This is in contrast to the more *internalized* form in which logics are represented in LF and in meta-theorem provers, so that deduction in the object logic is mirrored by deduction—for example, type inference—in the framework logic. Work on metalogical frameworks includes the already mentioned paper by Basin and Constable [1993], who advocate constructive type theory as the framework logic, work of Matthews, Smaill, and Basin [1993], who use Feferman’s  $FS_0$  [Feferman, 1989], a logic designed with the explicit purpose of being a metalogical framework, earlier work by Smullyan [1961], and work by Goguen, Stevens, Hopley, and Hilberdink [1992] on the 2OBJ meta-theorem prover, which uses order-sorted equational logic [Goguen and Meseguer, 1992; Goguen *et al.*, 2000].

A difficulty with systems based on higher-order type theory such as LF is that it may be quite awkward and of little practical use to represent logics whose structural properties differ considerably from those of the type theory. For example, linear and relevance logics do not have adequate representations in LF, in a precise technical sense of “adequate” [Gardner, 1992, Corollary 5.1.8]. Since in metalogical frameworks a direct connection between deduction in the object and framework logics does not have to be maintained, they seem in principle much more flexible in their representational capabilities. However, this comes at a price, since the possibility of directly using an implementation of the framework logic to implement an object logic is compromised.

In relation to this previous work, rewriting logic seems to have great flexibility to represent in a natural way many other logics, widely different in nature, including equational, Horn, and linear logics, and any sequent calculus presentation of a logic under extremely general assumptions about such a logic. Moreover, quantifiers can also be treated without problems. More experience in representing other logics is certainly needed, but we are encouraged by the naturalness and directness—often preserving the original syntax and rules—with which the logics that we have studied can be represented. This is due to the great simplicity and generality of rewriting logic, since in it all syntax and structural axioms are user-definable, so that the abstract syntax of an object logic can be represented as an algebraic data type, and is also due to the existence of only a few general “meta” rules of deduction relative to the rewrite rules given by a specification, where such a specification can be used to describe with rewrite rules the rules of deduc-

tion of the object logic in question. In addition, the direct correspondence between proofs in object logics and proofs in the framework logic can often be maintained in a *conservative* way by means of maps of logics, so that an implementation of rewriting logic can directly support an implementation of an object logic. Furthermore, given the directness with which logics can be represented, the task of proving conservativity is in many cases straightforward. Finally, although we do not discuss this aspect which is left for a subsequent paper, externalization of logics to support metalogical reasoning is also possible in rewriting logic.

Another important difference is that most approaches to logical frameworks are proof-theoretic in nature, and thus they do not address the model theories of the logics being represented. By contrast, several of the representations into rewriting logic that we consider—such as those for equational logic, Horn logic, and linear logic—involve both models and proofs and are therefore considerably more informative than purely proof-theoretic representations.

The fact that rewriting logic is *reflective* [Clavel and Meseguer, 1996; Clavel and Meseguer, 1996a] has very important practical consequences for its use as a logical framework. Note that a representation map  $\Psi : \mathcal{L} \rightarrow RWLogic$  for a logic  $\mathcal{L}$  is by its very nature a *metatheoretic* construction above the object levels of both  $\mathcal{L}$  and  $RWLogic$ . In particular,  $\Psi$  includes as one of its key components a function  $\Psi_{\mathbf{Th}} : \mathbf{Th}_{\mathcal{L}} \rightarrow \mathbf{Th}_{RWLogic}$  translating theories in  $\mathcal{L}$  into rewrite theories. However, thanks to the fact that the finitely presentable rewrite theories can be reified as an abstract data type  $RWL\text{-}ADT$ , for  $\mathcal{L}$  a logic having a finitary presentation of its syntax and its deduction rules, and such that  $\Psi$  maps finitely presented theories in  $\mathcal{L}$  to finitely presented rewrite theories, we can often *reify* a metatheoretic construction such as  $\Psi$  inside rewriting logic by first defining an abstract data type  $\mathcal{L}\text{-}ADT$  representing the finitely presentable theories of  $\mathcal{L}$ , and then reifying  $\Psi$  itself as an equationally defined function  $\overline{\Psi} : \mathcal{L}\text{-}ADT \rightarrow RWL\text{-}ADT$ . In this way, the translation  $\Psi$  becomes itself expressible and executable inside rewriting logic.

## 1.2 Rewriting logic as a semantic framework

As we have already mentioned, the distinction between a logical system and a language or a model of computation is more and more in the eyes of the beholder, although of course efficiency considerations and the practical uses intended may indeed strongly influence the design choices. A good case in point is the isomorphism between the Petri net model of concurrent computation [Reisig, 1995] and the tensor fragment of linear logic [Girard, 1987] (see [Martí-Oliet and Meseguer, 1991] and references therein). Therefore, even though at the most basic mathematical level there may be little distinction between the general way in which a logic, a programming language,

a system, or a model of computation are represented in rewriting logic, the criteria and case studies to be used in order to judge the merits of rewriting logic as a semantic framework are different from those relevant for its use as a logical framework.

One important consideration is that, from a computational point of view, rewriting logic deduction is intrinsically *concurrent*. In fact, it was the search for a general concurrency model that would help unify the somewhat bewildering heterogeneity of existing models that provided the original impetus for the first investigations on rewriting logic [Meseguer, 1992]. Since the generality and naturalness with which many concurrency models can be expressed in rewriting logic has already been illustrated at length in [Meseguer, 1992], only a brief summary is given in this paper. However, the CCS [Milner, 1989] and the concurrent object-oriented programming models are discussed in some detail to provide relevant examples.

Concurrent object-oriented programming is of particular interest. Given that the semantics of object-oriented programs is still poorly understood, and that the semantics of concurrent object-oriented systems is even less well understood, the ease with which rewriting logic can be used to give a precise semantics to concurrent object-oriented programs and to make such programs declarative is quite encouraging. In this paper, only the basic ideas of such a semantics are sketched; a much more detailed account can be found in [Meseguer, 1993].

The similarities between rewriting logic and structural operational semantics [Plotkin, 1981; Kahn, 1987] already noted in [Meseguer, 1992] are further explored in this paper. We give examples showing that different styles of structural operational semantics can be regarded as special cases of rewriting logic. The two main differences are the greater expressive power of rewriting logic due to the ability for rewriting modulo user-definable axioms, and the fact that rewriting logic is a full-fledged logic with both a proof and a model theory, whereas structural operational semantics accounts are only proof-theoretic.

Deduction with constraints can greatly increase the efficiency of theorem provers and logic programming languages. The most classical constraint solving algorithm is syntactic unification, which corresponds to solving equations in a free algebra, the so-called Herbrand model, and is used in resolution. However, much more efficient deduction techniques than those afforded by resolution can be obtained by building in additional knowledge of special theories in the form of constraint solving algorithms such as, for example, semantic unification, or equalities and inequalities in a numerical domain. In the past few years many authors have become aware that many constraint solving algorithms can be specified declaratively using rewrite rules. However, since constraint solving is usually nondeterministic, the usual equational logic interpretation of rewrite rules is clearly inadequate as a mathematical semantics. By contrast, rewriting logic completely avoids

such inadequacies and can serve as a semantic framework for logical systems and languages using constraints, including parallel ones.

The frame problem in artificial intelligence is caused by the need, typical of classical logic representations, to specify changes of state by stating not only what changes, but also what does not change. This is basically due to the essentially Platonic character of classical logic. Since rewriting logic is by design a logic of change that allows sound and complete deductions about the transitions of a system whose basic changes are axiomatized by rewrite rules, the difficulties associated with the frame problem disappear [Martí-Oliet and Meseguer, 1999]. In addition, the conservative mappings of Horn logic with equality and of linear logic studied in Sections 4.2 and 4.3, respectively, directly show how other logics of change recently proposed [Hölldobler and Schneeberger, 1990; Große *et al.*, 1996; Große *et al.*, 1992; Masseron *et al.*, 1990; Masseron *et al.*, 1993] can be subsumed as special cases. Added benefits include the straightforward support for concurrent change and the logical support for object-oriented representation.

The paper begins with a summary of the theory of general logics proposed in [Meseguer, 1989] that provides the conceptual basis for our discussion of logical frameworks. Then the rules of deduction and the model theory of rewriting logic are introduced, and the Maude and MaudeLog languages based on rewriting logic are briefly discussed. This is followed by a section presenting examples of logics representable in the rewriting logic framework. The role of rewriting logic as a semantic framework is then discussed and illustrated with examples. The paper ends with some concluding remarks.

## 2 GENERAL LOGICS

A general axiomatic theory of logics should adequately cover all the key ingredients of a logic. These include: a *syntax*, a notion of *entailment* of a sentence from a set of axioms, a notion of *model*, and a notion of *satisfaction* of a sentence by a model. A flexible axiomatic notion of a *proof calculus*, in which proofs of entailments, not just the entailments themselves, are first class citizens should also be included. This section gives a brief review of the required notions and axioms that will be later used in our treatment of rewriting logic as a logical framework; a more detailed account with many examples can be found in [Meseguer, 1989].

### 2.1 *Syntax*

Syntax can typically be given by a *signature*  $\Sigma$  providing a grammar on which to build *sentences*. For first-order logic, a typical signature consists of a list of function symbols and a list of predicate symbols, each with a prescribed number of arguments, which are used to build up sentences by



means of the usual logical connectives. For our purposes, it is enough to assume that for each logic there is a category **Sign** of possible signatures for it, and a functor  $sen$  assigning to each signature  $\Sigma$  the set  $sen(\Sigma)$  of all its sentences.

## 2.2 Entailment systems

For a given signature  $\Sigma$  in **Sign**, *entailment* (also called *provability*) of a sentence  $\varphi \in sen(\Sigma)$  from a set of axioms  $\Gamma \subseteq sen(\Sigma)$  is a relation  $\Gamma \vdash \varphi$  which holds if and only if we can prove  $\varphi$  from the axioms  $\Gamma$  using the rules of the logic. We make this relation relative to a signature.

In what follows,  $|\mathcal{C}|$  denotes the collection of objects of a category  $\mathcal{C}$ .

**DEFINITION 1.** [Meseguer, 1989] An *entailment system* is a triple  $\mathcal{E} = (\mathbf{Sign}, sen, \vdash)$  such that

- **Sign** is a category whose objects are called *signatures*,
- $sen : \mathbf{Sign} \rightarrow \mathbf{Set}$  is a functor associating to each signature  $\Sigma$  a corresponding set of  $\Sigma$ -sentences, and
- $\vdash$  is a function associating to each  $\Sigma \in |\mathbf{Sign}|$  a binary relation  $\vdash_\Sigma \subseteq \mathcal{P}(sen(\Sigma)) \times sen(\Sigma)$  called  $\Sigma$ -entailment such that the following properties are satisfied:
  1. *reflexivity*: for any  $\varphi \in sen(\Sigma)$ ,  $\{\varphi\} \vdash_\Sigma \varphi$ ,
  2. *monotonicity*: if  $\Gamma \vdash_\Sigma \varphi$  and  $\Gamma' \supseteq \Gamma$  then  $\Gamma' \vdash_\Sigma \varphi$ ,
  3. *transitivity*: if  $\Gamma \vdash_\Sigma \varphi_i$ , for all  $i \in I$ , and  $\Gamma \cup \{\varphi_i \mid i \in I\} \vdash_\Sigma \psi$ , then  $\Gamma \vdash_\Sigma \psi$ ,
  4.  *$\vdash$ -translation*: if  $\Gamma \vdash_\Sigma \varphi$ , then for any  $H : \Sigma \rightarrow \Sigma'$  in **Sign**,  $sen(H)(\Gamma) \vdash_{\Sigma'} sen(H)(\varphi)$ .

Except for the explicit treatment of syntax translations, the axioms are very similar to Scott's axioms for a consequence relation [Scott, 1974].

**DEFINITION 2.** [Meseguer, 1989] Given an entailment system  $\mathcal{E}$ , its category **Th** of *theories* has as objects pairs  $T = (\Sigma, \Gamma)$  with  $\Sigma$  a signature and  $\Gamma \subseteq sen(\Sigma)$ . A *theory morphism*  $H : (\Sigma, \Gamma) \rightarrow (\Sigma', \Gamma')$  is a signature morphism  $H : \Sigma \rightarrow \Sigma'$  such that if  $\varphi \in \Gamma$ , then  $\Gamma' \vdash_{\Sigma'} sen(H)(\varphi)$ .

A theory morphism  $H : (\Sigma, \Gamma) \rightarrow (\Sigma', \Gamma')$  is called *axiom-preserving* if it satisfies the condition that  $sen(H)(\Gamma) \subseteq \Gamma'$ . This defines a subcategory **Th**<sub>0</sub> with the same objects as **Th** but with morphisms restricted to be axiom-preserving theory morphisms. Notice that the category **Th**<sub>0</sub> does not depend at all on the entailment relation  $\vdash$ .

### 2.3 Institutions

The axiomatization of a model theory is due to the seminal work on *institutions* by Goguen and Burstall [1984; 1992].

DEFINITION 3. [Goguen and Burstall, 1984] An *institution* is a 4-tuple  $\mathcal{I} = (\mathbf{Sign}, sen, \mathbf{Mod}, \models)$  such that

- $\mathbf{Sign}$  is a category whose objects are called *signatures*,
- $sen : \mathbf{Sign} \rightarrow \mathbf{Set}$  is a functor associating to each signature  $\Sigma$  a set of  $\Sigma$ -sentences,
- $\mathbf{Mod} : \mathbf{Sign} \rightarrow \mathbf{Cat}^{op}$  is a functor that gives for each signature  $\Sigma$  a category whose objects are called  $\Sigma$ -models, and
- $\models$  is a function associating to each  $\Sigma \in |\mathbf{Sign}|$  a binary relation  $\models_{\Sigma} \subseteq |\mathbf{Mod}(\Sigma)| \times sen(\Sigma)$  called  $\Sigma$ -satisfaction satisfying the following *satisfaction condition* for each  $H : \Sigma \rightarrow \Sigma'$  in  $\mathbf{Sign}$ : for all  $M' \in |\mathbf{Mod}(\Sigma')|$  and all  $\varphi \in sen(\Sigma)$ ,

$$M' \models_{\Sigma'} sen(H)(\varphi) \iff \mathbf{Mod}(H)(M') \models_{\Sigma} \varphi.$$

The satisfaction condition just requires that, for any syntax translation between two signatures, a model of the second signature satisfies a translated sentence if and only if the translation of this model satisfies the original sentence. Note that  $\mathbf{Mod}$  is a contravariant functor, that is, translations of models go backwards.

Given a set of  $\Sigma$ -sentences  $\Gamma$ , we define the category  $\mathbf{Mod}(\Sigma, \Gamma)$  as the full subcategory of  $\mathbf{Mod}(\Sigma)$  determined by those models  $M \in |\mathbf{Mod}(\Sigma)|$  that satisfy all the sentences in  $\Gamma$ , i.e.,  $M \models_{\Sigma} \varphi$  for each  $\varphi \in \Gamma$ .

Since the definition above of the category of theories  $\mathbf{Th}_0$  only depends on signatures and sentences, it also makes sense for an institution.

### 2.4 Logics

Defining a *logic* is now almost trivial.

DEFINITION 4. [Meseguer, 1989] A *logic* is a 5-tuple  $\mathcal{L} = (\mathbf{Sign}, sen, \mathbf{Mod}, \vdash, \models)$  such that:

- $(\mathbf{Sign}, sen, \vdash)$  is an entailment system,
- $(\mathbf{Sign}, sen, \mathbf{Mod}, \models)$  is an institution,

and the following *soundness condition* is satisfied: for any  $\Sigma \in |\mathbf{Sign}|$ ,  $\Gamma \subseteq sen(\Sigma)$ , and  $\varphi \in sen(\Sigma)$ ,

$$\Gamma \vdash_{\Sigma} \varphi \implies \Gamma \models_{\Sigma} \varphi,$$

where, by definition, the relation  $\Gamma \models_{\Sigma} \varphi$  holds if and only if  $M \models_{\Sigma} \varphi$  holds for any model  $M$  that satisfies all the sentences in  $\Gamma$ .

The logic is called *complete* if the above implication is in fact an equivalence.

## 2.5 Proof calculi

A given logic may admit many different proof calculi. For example, in first-order logic we have Hilbert style, natural deduction, and sequent calculi among others, and the way in which proofs are represented and generated by rules of deduction is different for each of these calculi. It is useful to make proofs relative to a given theory  $T$  whose axioms we are allowed to use in order to prove theorems.

A proof calculus associates to each theory  $T$  a *structure*  $P(T)$  of proofs that use axioms of  $T$  as hypotheses. The structure  $P(T)$  typically has an *algebraic structure* of some kind so that we can obtain new proofs out of previously given proofs by operations that mirror the rules of deduction of the calculus in question. We need not make a choice about the particular types of algebraic structures that should be allowed for different proof calculi; we can abstract from such choices by simply saying that for a given proof calculus there is a category  $\mathbf{Str}$  of such structures and a functor  $P : \mathbf{Th}_0 \rightarrow \mathbf{Str}$  assigning to each theory  $T$  its structure of proofs  $P(T)$ . Of course, it should be possible to extract from  $P(T)$  the underlying set *proofs*( $T$ ) of all the proofs of theorems of the theory  $T$ , and this extraction should be functorial. Also, each proof, whatever it is, should contain information about what theorem it is a proof of; this can be formalized by postulating a “projection function”  $\pi_T$  (parameterized by  $T$  in a natural way) that maps each proof  $p \in \text{proofs}(T)$  to the sentence  $\varphi$  that it proves. Of course, each theorem of  $T$  must have at least one proof, and sentences that are not theorems should have no proof. To summarize, a *proof calculus* [Meseguer, 1989] consists of an entailment system together with:

- A functorial assignment  $P$  of a structure  $P(T)$  to each theory  $T$ .
- An additional functorial assignment of a set *proofs*( $T$ ) to each structure  $P(T)$ .
- A natural function  $\pi_T$  assigning a sentence to each proof  $p \in \text{proofs}(T)$  and such that, for  $\Gamma$  the axioms of  $T$ , a sentence  $\varphi$  is in the image of  $\pi_T$  if and only if  $\Gamma \vdash \varphi$ .

It is quite common to encounter proof systems of a specialized nature. In these calculi, only certain signatures are admissible as syntax—e.g., finite signatures—, only certain sentences are allowed as axioms, and only certain sentences—possibly different from the axioms—are allowed as conclusions. The obvious reason for introducing such specialized calculi is that

proofs are simpler under the given restrictions. In computer science the choice between an efficient and an inefficient calculus may have dramatic practical consequences. For logic programming languages, such calculi do (or should) coincide with what is called their *operational semantics*, and mark the difference between a hopelessly inefficient theorem prover and an efficient programming language. In practice, of course, we are primarily interested in proof calculi and proof subcalculi that are computationally effective. This is axiomatized by the notion of an (*effective*) *proof subcalculus* which can be found in [Meseguer, 1989].

## 2.6 Mapping logics

The advantage of having an axiomatic theory of logics is that the “space” of all logics (or that of all entailment systems, institutions, proof calculi, etc.) becomes well understood. This space is not just a collection of objects bearing no relationship to each other. In fact, the most interesting fruit of the theory of general logics outlined in this section is that it gives us a method for *relating* logics in a general and systematic way, and to exploit such relations in many applications. The simplest kind of relation is a *sublogic* (subentailment system, etc.) relation. Thus, first-order equational logic and Horn logic are both sublogics of first-order logic with equality. However, more subtle and general ways of relating logics are possible. For example, we may want to represent the universal fragment of first-order logic in a purely functional way by taking all the predicates and formulas to be *functions* whose value is either *true* or *false* so that a universal formula then becomes an equation equating a given term to *true*. The general way of relating logics (entailment systems, etc.) is to consider *maps* that interpret one logic into another. A detailed treatment of such maps is given in [Meseguer, 1989]; here we summarize some of the key ideas.

Let us first discuss in some detail *maps of entailment systems*. Basically, a map of entailment systems  $\mathcal{E} \rightarrow \mathcal{E}'$  maps the language of  $\mathcal{E}$  to that of  $\mathcal{E}'$  in a way that respects the entailment relation. This means that signatures of  $\mathcal{E}$  are functorially mapped to signatures of  $\mathcal{E}'$ , and that sentences of  $\mathcal{E}$  are mapped to sentences of  $\mathcal{E}'$  in a way that is coherent with the mapping of their corresponding signatures. In addition, such a mapping  $\alpha$  must respect the entailment relations  $\vdash$  of  $\mathcal{E}$  and  $\vdash'$  of  $\mathcal{E}'$ , i.e., we must have  $\Gamma \vdash \varphi \Rightarrow \alpha(\Gamma) \vdash' \alpha(\varphi)$ . It turns out that for many interesting applications, including the functional representation of first-order logic sketched above, one wants to be more general and allow maps that send a signature of  $\mathcal{E}$  to a *theory* of  $\mathcal{E}'$ . These maps extend to maps between theories, and in this context the coherence with the mapping at the level of signatures is expressed by the notion of *sensible functor* defined in [Meseguer, 1989].

DEFINITION 5. [Meseguer, 1989] Given entailment systems  $\mathcal{E} = (\mathbf{Sign}, sen, \vdash)$  and  $\mathcal{E}' = (\mathbf{Sign}', sen', \vdash')$ , a *map of entailment systems*  $(\Phi, \alpha) : \mathcal{E} \rightarrow \mathcal{E}'$  consists of a natural transformation  $\alpha : sen \Rightarrow \Phi; sen'$  and an  $\alpha$ -sensible functor<sup>1</sup>  $\Phi : \mathbf{Th}_0 \rightarrow \mathbf{Th}'_0$  satisfying the following property:

$$\Gamma \vdash_{\Sigma} \varphi \implies \Gamma' \cup \alpha_{\Sigma}(\Gamma) \vdash'_{\Sigma'} \alpha_{\Sigma}(\varphi),$$

where, by convention,  $(\Sigma', \Gamma') = \Phi(\Sigma, \Gamma)$ .

We say that  $(\Phi, \alpha)$  is a *conservative map of entailment systems* when the above implication is an equivalence.

The property of being conservative may be essential for many applications. For example, since proof calculi are in a sense computational engines on which the design and implementation of theorem provers and logic programming languages can be based, we can view the establishment of a map of proof calculi having nice properties, such as conservativity, as a proof of correctness for a *compiler* that permits implementing a system based on the first calculus in terms of another system based on the second. Besides establishing correctness, the map itself specifies the compilation function.

A *map of institutions*<sup>2</sup>  $\mathcal{I} \rightarrow \mathcal{I}'$  is similar in its syntax part to a map of entailment systems. In addition, for models we have a natural functor  $\beta : \mathbf{Mod}'(\Phi(\Sigma)) \rightarrow \mathbf{Mod}(\Sigma)$  “backwards” from the models in  $\mathcal{I}'$  of a translated signature  $\Phi(\Sigma)$  to the models in  $\mathcal{I}$  of the original signature  $\Sigma$ , and such a mapping respects the satisfaction relations  $\models$  of  $\mathcal{I}$  and  $\models'$  of  $\mathcal{I}'$ , in the sense that  $M' \models' \alpha(\varphi) \iff \beta(M') \models \varphi$ .

DEFINITION 6. [Meseguer, 1989] Given institutions  $\mathcal{I} = (\mathbf{Sign}, sen, \mathbf{Mod}, \models)$  and  $\mathcal{I}' = (\mathbf{Sign}', sen', \mathbf{Mod}', \models')$ , a *map of institutions*  $(\Phi, \alpha, \beta) : \mathcal{I} \rightarrow \mathcal{I}'$  consists of a natural transformation  $\alpha : sen \Rightarrow \Phi; sen'$ , an  $\alpha$ -sensible functor  $\Phi : \mathbf{Th}_0 \rightarrow \mathbf{Th}'_0$ , and a natural transformation  $\beta : \Phi^{op}; \mathbf{Mod}' \Rightarrow \mathbf{Mod}$  such that for each  $\Sigma \in |\mathbf{Sign}|$ ,  $\varphi \in sen(\Sigma)$ , and  $M' \in |\mathbf{Mod}'(\Phi(\Sigma, \emptyset))|$  the following property is satisfied:

$$M' \models'_{\Sigma'} \alpha_{\Sigma}(\varphi) \iff \beta_{(\Sigma, \emptyset)}(M') \models_{\Sigma} \varphi,$$

where  $\Sigma'$  is the signature of the theory  $\Phi(\Sigma, \emptyset)$ .

A *map of logics* has now a very simple definition. It consists of a pair of maps: one for the underlying entailment systems, and another for the underlying institutions, such that both maps agree on how they translate signatures and sentences.

<sup>1</sup>We refer to [Meseguer, 1989] for the detailed definition of  $\alpha$ -sensible functor. Basically, what is required is that the provable consequences of the theory  $\Phi(\Sigma, \Gamma)$  are entirely determined by  $\Phi(\Sigma, \emptyset)$  and by  $\alpha(\Gamma)$ . Note that  $\alpha$  depends only on signatures, not theories.

<sup>2</sup>Such maps are different from the “institution morphisms” considered by Goguen and Burstall in [1984; 1992].

DEFINITION 7. [Meseguer, 1989] Given logics  $\mathcal{L} = (\mathbf{Sign}, sen, \mathbf{Mod}, \vdash, \models)$  and  $\mathcal{L}' = (\mathbf{Sign}', sen', \mathbf{Mod}', \vdash', \models')$ , a *map of logics*  $(\Phi, \alpha, \beta) : \mathcal{L} \rightarrow \mathcal{L}'$  consists of a functor  $\Phi : \mathbf{Th}_0 \rightarrow \mathbf{Th}'_0$ , and natural transformations  $\alpha : sen \Rightarrow \Phi; sen'$  and  $\beta : \Phi^{op}; \mathbf{Mod}' \Rightarrow \mathbf{Mod}$  such that:

- $(\Phi, \alpha) : (\mathbf{Sign}, sen, \vdash) \rightarrow (\mathbf{Sign}', sen', \vdash')$  is a map of entailment systems, and
- $(\Phi, \alpha, \beta) : (\mathbf{Sign}, sen, \mathbf{Mod}, \models) \rightarrow (\mathbf{Sign}', sen', \mathbf{Mod}', \models')$  is a map of institutions.

We say that  $(\Phi, \alpha, \beta)$  is *conservative* when if  $(\Phi, \alpha)$  is so as a map of entailment systems.

There is also a notion of map of proof calculi, for which we refer the reader to [Meseguer, 1989].

### 2.7 The idea of a logical framework

As we have already explained in the introduction, viewed from the perspective of a general space of logics that can be related to each other by means of mappings, the quest for a *logical framework* can be understood as the search within such a space for a logic  $\mathcal{F}$  (the *framework* logic) such that many other logics (the *object* logics) such as, say,  $\mathcal{L}$  can be represented in  $\mathcal{F}$  by means of mappings  $\mathcal{L} \rightarrow \mathcal{F}$  that have good enough properties. The minimum requirement that seems reasonable to make on a representation map  $\mathcal{L} \rightarrow \mathcal{F}$  is that it should be a *conservative* map of entailment systems. Under such circumstances, we can reduce issues of provability in  $\mathcal{L}$  to issues of provability in  $\mathcal{F}$ , by mapping the theories and sentences of  $\mathcal{L}$  into  $\mathcal{F}$  using the conservative representation map. Given a computer implementation of deduction in  $\mathcal{F}$ , we can use the conservative map to prove theorems in  $\mathcal{L}$  by proving the corresponding translations in  $\mathcal{F}$ . In this way, the implementation for  $\mathcal{F}$  can be used as a generic theorem prover for many logics.

However, since maps between logics can, as we have seen, respect additional logical structure such as the model theory or the proofs, in some cases a representation map into a logical framework may be particularly informative because, in addition to being a conservative map of entailment systems, it is also a map of institutions, or a map of proof calculi. For example, when rewriting logic is chosen as a logical framework, appropriate representation maps for equational logic, Horn logic, and propositional linear logic can be shown to be maps of institutions also (see Section 4). In general, however, since the model theories of different logics can be very different from each other, it is not reasonable to expect or require that the representation maps into a logical framework will always be maps of institutions. Nevertheless,

what it can always be done is to “borrow” the additional logical structure that  $\mathcal{F}$  may have (institution, proof calculus) to endow  $\mathcal{L}$  with such a structure, so that the representation map does indeed preserve the extra structure [Cerioli and Meseguer, 1996].

Having criteria for the adequacy of maps representing logics in a logical framework is not enough. An equally important issue is having criteria for the *generality* of a logical framework, so that it is in fact justified to call it by that name. That is, given a candidate logical framework  $\mathcal{F}$ , how many logics can be adequately represented in  $\mathcal{F}$ ? We can make this question precise by defining the *scope* of a logical framework  $\mathcal{F}$  as the class of entailment systems  $\mathcal{E}$  having conservative maps of entailment systems  $\mathcal{E} \rightarrow \mathcal{F}$ . In this regard, the axioms of the theory of general logics that we have presented are probably too general; without adding further assumptions it is not reasonable to expect that we can find a logical framework  $\mathcal{F}$  whose scope is the class of *all* entailment systems. A much more reasonable goal is finding an  $\mathcal{F}$  whose scope includes all entailment systems of “practical interest,” having finitary presentations of their syntax and their rules of deduction. Axiomatizing such finitely presentable entailment systems and proof calculi so as to capture—in the spirit of the more general axioms that we have presented, but with stronger requirements—all logics of “practical interest” (at least for computational purposes) is a very important research task.

Another important property that can help measuring the suitability of a logic  $\mathcal{F}$  as a logical framework is its *representational adequacy*, understood as the naturalness and ease with which entailment systems can be represented, so that the representation  $\mathcal{E} \rightarrow \mathcal{F}$  mirrors  $\mathcal{E}$  as closely as possible. That is, a framework requiring very complicated encodings for many object logics of interest is less representationally adequate than one for which most logics can be represented in a straightforward way, so that there is in fact little or no “distance” between an object logic and its corresponding representation. Although at present we lack a precise definition of this property, it is quite easy to observe its absence in particular examples. We view representational adequacy as a very important practical criterion for judging the relative merits of different logical frameworks.

In this paper, we present rewriting logic as a logic that seems to have particularly good properties as a logical framework. We conjecture that the scope of rewriting logic contains all entailment systems of “practical interest” for a reasonable axiomatization of such systems.

## 2.8 Reflection

We give here a brief summary of the notion of a universal theory in a logic and of a reflective entailment system introduced in [Clavel and Meseguer, 1996]. These notions axiomatize reflective logics within the theory of general

logics [Meseguer, 1989]. We focus here on the simplest case, namely entailment systems. However, reflection at the proof calculus level—where not only sentences, but also proofs are reflected—is also very useful; definitions for that case are also in [Clavel and Meseguer, 1996].

A reflective logic is a logic in which important aspects of its metatheory can be represented at the object level in a consistent way, so that the object-level representation correctly simulates the relevant metatheoretic aspects. Two obvious metatheoretic notions that can be so reflected are theories and the entailment relation  $\vdash$ . This leads us to the notion of a universal theory. However, universality may not be absolute, but only relative to a class  $\mathcal{C}$  of *representable* theories. Typically, for a theory to be representable at the object level, it must have a finitary description in some way—say, being recursively enumerable—so that it can be represented as a piece of language.

Given an entailment system  $\mathcal{E}$  and a set of theories  $\mathcal{C}$ , a theory  $U$  is  $\mathcal{C}$ -*universal* if there is a recursive injective function, called a *representation function*,

$$\overline{(\_ \vdash \_)} : \bigcup_{T \in \mathcal{C}} \{T\} \times \text{sen}(T) \longrightarrow \text{sen}(U)$$

such that for each  $T \in \mathcal{C}$ ,  $\varphi \in \text{sen}(T)$ ,

$$T \vdash \varphi \iff U \vdash \overline{T \vdash \varphi}.$$

If, in addition,  $U \in \mathcal{C}$ , then the entailment system  $\mathcal{E}$  is called  $\mathcal{C}$ -*reflective*.

Note that in a reflective entailment system, since  $U$  itself is representable, representation can be iterated, so that we immediately have a “reflective tower”

$$T \vdash \varphi \iff U \vdash \overline{T \vdash \varphi} \iff U \vdash \overline{U \vdash \overline{T \vdash \varphi}} \dots$$

### 3 REWRITING LOGIC

This section gives the rules of deduction and semantics of rewriting logic, and explains its computational meaning. The Maude and MaudeLog languages, based on rewriting logic, are also briefly discussed.

#### 3.1 Basic universal algebra

Rewriting logic is parameterized with respect to the version of the underlying equational logic, which can be unsorted, many-sorted, order-sorted, or the recently developed membership equational logic [Bouhoula *et al.*, 2000; Meseguer, 1998]. For the sake of simplifying the exposition, we treat here the *unsorted* case.

A set  $\Sigma$  of function symbols is a ranked alphabet  $\Sigma = \{\Sigma_n \mid n \in \mathbb{N}\}$ . A  $\Sigma$ -algebra is then a set  $A$  together with an assignment of a function



$f_A : A^n \rightarrow A$  for each  $f \in \Sigma_n$  with  $n \in \mathbb{N}$ . We denote by  $T_\Sigma$  the  $\Sigma$ -algebra of ground  $\Sigma$ -terms, and by  $T_\Sigma(X)$  the  $\Sigma$ -algebra of  $\Sigma$ -terms with variables in a set  $X$ . Similarly, given a set  $E$  of  $\Sigma$ -equations,  $T_{\Sigma,E}$  denotes the  $\Sigma$ -algebra of equivalence classes of ground  $\Sigma$ -terms modulo the equations  $E$ ; in the same way,  $T_{\Sigma,E}(X)$  denotes the  $\Sigma$ -algebra of equivalence classes of  $\Sigma$ -terms with variables in  $X$  modulo the equations  $E$ . Let  $[t]_E$  or just  $[t]$  denote the  $E$ -equivalence class of  $t$ .

Given a term  $t \in T_\Sigma(\{x_1, \dots, x_n\})$  and terms  $u_1, \dots, u_n \in T_\Sigma(X)$ , we denote  $t(u_1/x_1, \dots, u_n/x_n)$  the term in  $T_\Sigma(X)$  obtained from  $t$  by *simultaneously substituting*  $u_i$  for  $x_i$ ,  $i = 1, \dots, n$ . To simplify notation, we denote a sequence of objects  $a_1, \dots, a_n$  by  $\bar{a}$ ; with this notation,  $t(u_1/x_1, \dots, u_n/x_n)$  can be abbreviated to  $t(\bar{u}/\bar{x})$ .

### 3.2 The rules of rewriting logic

A *signature* in rewriting logic is a pair  $(\Sigma, E)$  with  $\Sigma$  a ranked alphabet of function symbols and  $E$  a set of  $\Sigma$ -equations. Rewriting will operate on equivalence classes of terms modulo the set of equations  $E$ . In this way, we free rewriting from the syntactic constraints of a term representation and gain a much greater flexibility in deciding what counts as a *data structure*; for example, string rewriting is obtained by imposing an associativity axiom, and multiset rewriting by imposing associativity and commutativity. Of course, standard term rewriting is obtained as the particular case in which the set  $E$  of equations is empty. Techniques for rewriting modulo equations have been studied extensively [Dershowitz and Jouannaud, 1990] and can be used to implement rewriting modulo many equational theories of interest.

Given a signature  $(\Sigma, E)$ , *sentences* of rewriting logic are “sequents” of the form  $[t]_E \rightarrow [t']_E$ , where  $t$  and  $t'$  are  $\Sigma$ -terms possibly involving some variables from the countably infinite set  $X = \{x_1, \dots, x_n, \dots\}$ . A *theory* in this logic, called a *rewrite theory*, is a slight generalization of the usual notion of theory as in Definition 2 in that, in addition, we allow the axioms—in this case the sequents  $[t]_E \rightarrow [t']_E$ —to be labelled. This is very natural for many applications, and customary for automata—viewed as labelled transition systems—and for Petri nets, which are both particular instances of our definition.

**DEFINITION 8.** A *rewrite theory*  $\mathcal{R}$  is a 4-tuple  $\mathcal{R} = (\Sigma, E, L, R)$  where  $\Sigma$  is a ranked alphabet of function symbols,  $E$  is a set of  $\Sigma$ -equations,  $L$  is a set of *labels*, and  $R$  is a set of pairs  $R \subseteq L \times T_{\Sigma,E}(X)^2$  whose first component is a label and whose second component is a pair of  $E$ -equivalence classes of terms, with  $X = \{x_1, \dots, x_n, \dots\}$  a countably infinite set of variables. Elements of  $R$  are called *rewrite rules*<sup>3</sup>. We understand a

<sup>3</sup>To simplify the exposition the rules of the logic are given for the case of *unconditional* rewrite rules. However, all the ideas presented here have been extended to conditional

rule  $(r, ([t], [t']))$  as a labelled sequent and use for it the notation  $r : [t] \longrightarrow [t']$ . To indicate that  $\{x_1, \dots, x_n\}$  is the set of variables occurring in either  $t$  or  $t'$ , we write  $r : [t(x_1, \dots, x_n)] \longrightarrow [t'(x_1, \dots, x_n)]$ , or in abbreviated notation  $r : [t(\bar{x})] \longrightarrow [t'(\bar{x})]$ .

Given a rewrite theory  $\mathcal{R}$ , we say that  $\mathcal{R}$  *entails* a sequent  $[t] \longrightarrow [t']$  and write  $\mathcal{R} \vdash [t] \longrightarrow [t']$  if and only if  $[t] \longrightarrow [t']$  can be obtained by finite application of the following *rules of deduction*:

1. **Reflexivity.** For each  $[t] \in T_{\Sigma, E}(X)$ ,

$$\frac{}{[t] \longrightarrow [t]}.$$

2. **Congruence.** For each  $f \in \Sigma_n$ ,  $n \in \mathbb{N}$ ,

$$\frac{[t_1] \longrightarrow [t'_1] \quad \dots \quad [t_n] \longrightarrow [t'_n]}{[f(t_1, \dots, t_n)] \longrightarrow [f(t'_1, \dots, t'_n)]}.$$

3. **Replacement.** For each rewrite rule in the theory  $R$  of the form  $r : [t(x_1, \dots, x_n)] \longrightarrow [t'(x_1, \dots, x_n)]$ ,

$$\frac{[w_1] \longrightarrow [w'_1] \quad \dots \quad [w_n] \longrightarrow [w'_n]}{[t(\bar{w}/\bar{x})] \longrightarrow [t'(\bar{w}'/\bar{x})]}.$$

4. **Transitivity.**

$$\frac{[t_1] \longrightarrow [t_2] \quad [t_2] \longrightarrow [t_3]}{[t_1] \longrightarrow [t_3]}.$$

*Equational logic* (modulo a set of axioms  $E$ ) is obtained from rewriting logic by adding the following rule:

5. **Symmetry.**

$$\frac{[t_1] \longrightarrow [t_2]}{[t_2] \longrightarrow [t_1]}.$$

With this new rule, sequents derivable in equational logic are *bidirectional*; therefore, in this case we can adopt the notation  $[t] \leftrightarrow [t']$  throughout and call such bidirectional sequents *equations*.

A nice consequence of having defined rewriting logic is that concurrent rewriting, rather than emerging as an operational notion, actually coincides with deduction in such a logic.

**DEFINITION 9.** Given a rewrite theory  $\mathcal{R} = (\Sigma, E, L, R)$ , a  $(\Sigma, E)$ -sequent  $[t] \longrightarrow [t']$  is called a *concurrent  $\mathcal{R}$ -rewrite* (or just a *rewrite*) if and only if it can be derived from  $\mathcal{R}$  by means of the rules 1-4, i.e.,  $\mathcal{R} \vdash [t] \longrightarrow [t']$ .

rules in [Meseguer, 1992] with very general rules of the form

$$r : [t] \longrightarrow [t'] \text{ if } [u_1] \longrightarrow [v_1] \wedge \dots \wedge [u_k] \longrightarrow [v_k].$$

This increases considerably the expressive power of rewrite theories, as illustrated by several of the examples presented in this paper.

### 3.3 *The meaning of rewriting logic*

A logic worth its salt should be understood as a method of correct reasoning about some class of entities, not as an empty formal game. For equational logic, the entities in question are sets, functions between them, and the relation of identity between elements. For rewriting logic, the entities in question are *concurrent systems* having *states*, and evolving by means of *transitions*. The *signature* of a rewrite theory describes a particular structure for the states of a system—e.g., multiset, binary tree, etc.—so that its states can be distributed according to such a structure. The *rewrite rules* in the theory describe which *elementary local transitions* are possible in the distributed state by concurrent local transformations. The rules of rewriting logic allow us to reason correctly about which *general* concurrent transitions are possible in a system satisfying such a description. Clearly, concurrent systems should be the *models* giving a semantic interpretation to rewriting logic, in the same way that algebras are the models giving a semantic interpretation to equational logic. A precise account of the model theory of rewriting logic, giving rise to an initial model semantics for Maude modules and fully consistent with the above system-oriented interpretation, is sketched in Section 3.5 and developed in full detail for the more general conditional case in [Meseguer, 1992].

Therefore, in rewriting logic a sequent  $[t] \longrightarrow [t']$  should not be read as “[ $t$ ] equals [ $t'$ ],” but as “[ $t$ ] becomes [ $t'$ ].” Clearly, rewriting logic is a logic of *becoming* or *change*, not a logic of equality in a static sense. The apparently innocent step of adding the symmetry rule is in fact a *very strong* restriction, namely assuming that *all change is reversible*, thus bringing us into a timeless Platonic realm in which “before” and “after” have been identified.

A related observation, which is particularly important for the use of rewriting logic as a logical framework, is that  $[t]$  should not be understood as a *term* in the usual first-order logic sense, but as a *proposition* or *formula*—built up using the *connectives* in  $\Sigma$ —that asserts being in a certain *state* having a certain *structure*. However, unlike most other logics, the logical connectives  $\Sigma$  and their structural properties  $E$  are entirely *user-definable*. This provides great flexibility for considering many different state structures and makes rewriting logic very general in its capacity to deal with many different types of concurrent systems, and also in its capacity to represent many different logics. For the case of concurrent systems, this generality is discussed at length in [Meseguer, 1992] (see also [Martí-Oliet and Meseguer, 1999] for the advantages of this generality in the context of unifying AI logics of action). In a similar vein, but with a broader focus, Section 5 discusses the advantages of rewriting logic as a general semantic framework in which to specify and prototype languages and systems. Finally, Section 4 explores the generality of rewriting logic as a logical framework in which logics can

be represented and prototyped.

In summary, the rules of rewriting logic are rules to reason about *change in a concurrent system*, or, alternatively, metarules for reasoning about *deduction in a logical system*. They allow us to draw valid conclusions about the evolution of the system from certain basic types of change known to be possible, or, in the alternative viewpoint, about the correct deductions possible in a logical system. Our present discussion is summarized as follows:

<i>State</i>	$\leftrightarrow$	<i>Term</i>	$\leftrightarrow$	<i>Proposition</i>
<i>Transition</i>	$\leftrightarrow$	<i>Rewriting</i>	$\leftrightarrow$	<i>Deduction</i>
<i>Distributed</i>	$\leftrightarrow$	<i>Algebraic</i>	$\leftrightarrow$	<i>Propositional</i>
<i>Structure</i>		<i>Structure</i>		<i>Structure</i>

Section 4 will further clarify and illustrate each of the correspondences in the last two columns of the diagram, and Section 5 will do the same for the first two columns.

### 3.4 The Maude and MaudeLog languages

Rewriting logic can be used directly as a wide spectrum language supporting specification, rapid prototyping, and programming of concurrent systems. As explained later in this paper, rewriting logic can also be used as a logical framework in which other logics can be naturally represented, and as a semantic framework for specifying languages and systems. The Maude language [Meseguer, 1993; Clavel *et al.*, 1996] supports all these uses of rewriting logic in a particularly modular way in which modules are rewrite theories and in which functional modules with equationally defined data types can also be declared in a functional sublanguage. The examples given later in this paper illustrate the syntax of Maude. Details about the language design, its semantics, its parallel programming and wide spectrum capabilities, and its support of object-oriented programming can be found in [Meseguer, 1992; Meseguer and Winkler, 1992; Meseguer, 1993; Meseguer, 1993b]. Here we provide a very brief sketch that should be sufficient for understanding the examples presented later.

In Maude there are three kinds of *modules*:

1. *Functional modules*, introduced by the keyword `fmod`,
2. *System modules*, introduced by the keyword `mod`, and
3. *Object-oriented modules*, introduced by the keyword `omod`.

Object-oriented modules can be reduced to a special case of system modules for which a special syntax is used; therefore, in essence we only have functional and system modules. Maude's functional and system modules are respectively of the form

- `fmod  $\mathcal{E}$  endfm`, and
- `mod  $\mathcal{R}$  endm`,

for  $\mathcal{E}$  an equational theory and  $\mathcal{R}$  a rewrite theory<sup>4</sup>. In functional modules, equations are declared with the keywords `eq` or `ceq` (for conditional equations), and in system or object-oriented modules with the keywords `ax` or `cax`. In addition, certain equations, such as any combination of associativity, commutativity, or identity, for which rewriting modulo is provided, can be declared together with the corresponding operator using the keywords `assoc`, `comm`, `id`. Rules can only appear in system or object-oriented modules, and are declared with the keywords `r1` or `cr1`.

In Maude a module can have *submodules*, which can be imported with either `protecting` or `including` qualifications stating the degree of integrity enjoyed by the submodule when imported by the supermodule.

The version of rewriting logic used for Maude in this paper is *order-sorted*<sup>5</sup>. This means that rewrite theories are typed (types are called *sorts*) and can have subtypes (subsorts), and that function symbols can be overloaded. In particular, functional modules are order-sorted equational theories [Goguen and Meseguer, 1992] and they form a sublanguage similar to OBJ [Goguen *et al.*, 2000].

Like OBJ, Maude has also *theories* to specify semantic requirements for interfaces and to make high level assertions about modules. They are of the three kinds:

1. *Functional theories*, introduced by the keyword `fth`,
2. *System theories*, introduced by the keyword `th`, and
3. *Object-oriented theories*, introduced by the keyword `oth`.

Also as OBJ, Maude has *parameterized modules* and *theories*, again of the three kinds, and *views* that are theory interpretations relating theories to modules or to other theories.

Maude can be further extended to a language called MaudeLog that unifies the paradigms of functional programming, Horn logic programming, and concurrent object-oriented programming. In fact, Maude’s design is based on a general axiomatic notion of “logic programming language” based on the general axiomatic theory of logic sketched in Section 2 [Meseguer, 1989; Meseguer, 1992b]. Technically, a unification of paradigms is achieved by mapping the logics of each paradigm into a richer logic in which the

---

<sup>4</sup>This is somewhat inaccurate in the case of system modules having functional submodules because we have to “remember” that the submodule is functional.

<sup>5</sup>The latest version of Maude [Clavel *et al.*, 1996] is based on the recently developed membership equational logic, which extends order-sorted equational logic and at the same time has a simpler and more general model theory [Bouhoula *et al.*, 2000; Meseguer, 1998].

paradigms are unified. In the case of Maude and MaudeLog, what is done is to define a new logic (rewriting logic) in which concurrent computations, and in particular concurrent object-oriented computations, can be expressed in a natural way, and then to formally relate this logic to the logics of the functional and relational paradigms, i.e., to equational logic and to Horn logic, by means of maps of logics that provide a simple and rigorous unification of paradigms. As it has already been mentioned, we actually assume an order-sorted structure throughout, and therefore the logics in question are: order-sorted rewriting logic, denoted *OSRWLogic*, order-sorted equational logic, denoted *OSEqtl*, and order-sorted Horn logic, denoted *OSHorn*.

The logic of equational programming can be embedded within (order-sorted) rewriting logic by means of a map of logics

$$OSEqtl \longrightarrow OSRWLogic.$$

The details of this map of logics are discussed in Section 4.1. At the programming language level, such a map corresponds to the inclusion of Maude's functional modules (essentially identical to OBJ modules) within the language.

Since the power and the range of applications of a multiparadigm logic programming language can be substantially increased if it is possible to solve queries involving *logical variables* in the sense of relational programming, as in the Prolog language, we are naturally led to seek a unification of the three paradigms of functional, relational and concurrent object-oriented programming into a single multiparadigm logic programming language. This unification can be attained in a language extension of Maude called MaudeLog. The integration of Horn logic is achieved by a map of logics

$$OSHorn \longrightarrow OSRWLogic$$

that systematically relates order-sorted Horn logic to order-sorted rewriting logic. The details of this map are discussed in Section 4.2.

The difference between Maude and MaudeLog does not consist of any change in the underlying logic; indeed, both languages are based on rewriting logic, and both have rewrite theories as programs. It resides, rather, in an enlargement of the set of *queries* that can be presented, so that, while keeping the same syntax and models, in MaudeLog we also consider queries involving existential formulas of the form

$$\exists \bar{x} \ [u_1(\bar{x})] \longrightarrow [v_1(\bar{x})] \wedge \dots \wedge [u_k(\bar{x})] \longrightarrow [v_k(\bar{x})].$$

Therefore, the sentences and the deductive rules and mechanisms that are now needed require further extensions of rewriting logic deduction. In particular, solving such existential queries requires performing *unification*, specifically, given Maude's typing structure, order-sorted *E*-unification for a set *E* of structural axioms [Meseguer *et al.*, 1989].

### 3.5 The models of rewriting logic

We first sketch the construction of initial and free models for a rewrite theory  $\mathcal{R} = (\Sigma, E, L, R)$ . Such models capture nicely the intuitive idea of a “rewrite system” in the sense that they are systems whose states are  $E$ -equivalence classes of terms, and whose transitions are concurrent rewrites using the rules in  $R$ . By adopting a logical instead of a computational perspective, we can alternatively view such models as “logical systems” in which formulas are validly rewritten to other formulas by concurrent rewrites which correspond to proofs for the logic in question. Such models have a natural *category* structure, with states (or formulas) as objects, transitions (or proofs) as morphisms, and sequential composition as morphism composition, and in them dynamic behavior exactly corresponds to deduction.

Given a rewrite theory  $\mathcal{R} = (\Sigma, E, L, R)$ , the model that we are seeking is a category  $\mathcal{T}_{\mathcal{R}}(X)$  whose objects are equivalence classes of terms  $[t] \in T_{\Sigma, E}(X)$  and whose morphisms are equivalence classes of “proof terms” representing proofs in rewriting deduction, i.e., concurrent  $\mathcal{R}$ -rewrites. The rules for generating such proof terms, with the specification of their respective domain and codomain, are given below; they just “decorate” with proof terms the rules 1-4 of rewriting logic. Note that we always use “diagrammatic” notation for morphism composition, i.e.,  $\alpha; \beta$  always means the composition of  $\alpha$  followed by  $\beta$ .

1. **Identities.** For each  $[t] \in T_{\Sigma, E}(X)$ ,

$$\overline{[t] : [t] \longrightarrow [t]}.$$

2.  **$\Sigma$ -structure.** For each  $f \in \Sigma_n$ ,  $n \in \mathbb{N}$ ,

$$\frac{\alpha_1 : [t_1] \longrightarrow [t'_1] \quad \dots \quad \alpha_n : [t_n] \longrightarrow [t'_n]}{f(\alpha_1, \dots, \alpha_n) : [f(t_1, \dots, t_n)] \longrightarrow [f(t'_1, \dots, t'_n)]}.$$

3. **Replacement.** For each rewrite rule  $r : [t(\bar{x}^n)] \longrightarrow [t'(\bar{x}^n)]$  in  $R$ ,

$$\frac{\alpha_1 : [w_1] \longrightarrow [w'_1] \quad \dots \quad \alpha_n : [w_n] \longrightarrow [w'_n]}{r(\alpha_1, \dots, \alpha_n) : [t(\bar{w}/\bar{x})] \longrightarrow [t'(\bar{w}'/\bar{x})]}.$$

4. **Composition.**

$$\frac{\alpha : [t_1] \longrightarrow [t_2] \quad \beta : [t_2] \longrightarrow [t_3]}{\alpha; \beta : [t_1] \longrightarrow [t_3]}.$$

**Convention.** In the case when the same label  $r$  appears in two different rules of  $R$ , the “proof terms”  $r(\bar{\alpha})$  can sometimes be ambiguous. We assume that such ambiguity problems have been resolved by disambiguating the

label  $r$  in the proof terms  $r(\bar{\alpha})$  if necessary; with this understanding, we adopt the simpler notation  $r(\bar{\alpha})$  to ease the exposition.

Each of the above rules of generation defines a different proof term taking certain proof terms as arguments and returning a resulting proof term. In other words, proof terms form an algebraic structure  $\mathcal{P}_{\mathcal{R}}(X)$  consisting of a graph with nodes  $T_{\Sigma, E}(X)$ , with identity arrows, and with operations  $f$  (for each  $f \in \Sigma$ ),  $r$  (for each rewrite rule), and  $_;$  (for composing arrows). Our desired model  $\mathcal{T}_{\mathcal{R}}(X)$  is the quotient of  $\mathcal{P}_{\mathcal{R}}(X)$  modulo the following equations<sup>6</sup>:

1. **Category.**

(a) *Associativity.* For all  $\alpha, \beta, \gamma$ ,

$$(\alpha; \beta); \gamma = \alpha; (\beta; \gamma).$$

(b) *Identities.* For each  $\alpha : [t] \longrightarrow [t']$ ,

$$\alpha; [t'] = \alpha \quad \text{and} \quad [t]; \alpha = \alpha.$$

2. **Functoriality of the  $\Sigma$ -algebraic structure.** For each  $f \in \Sigma_n$ ,  $n \in \mathbb{N}$ ,

(a) *Preservation of composition.* For all  $\alpha_1, \dots, \alpha_n, \beta_1, \dots, \beta_n$ ,

$$f(\alpha_1; \beta_1, \dots, \alpha_n; \beta_n) = f(\alpha_1, \dots, \alpha_n); f(\beta_1, \dots, \beta_n).$$

(b) *Preservation of identities.*

$$f([t_1], \dots, [t_n]) = [f(t_1, \dots, t_n)].$$

3. **Axioms in  $E$ .** For each axiom  $t(x_1, \dots, x_n) = t'(x_1, \dots, x_n)$  in  $E$ , for all  $\alpha_1, \dots, \alpha_n$ ,

$$t(\alpha_1, \dots, \alpha_n) = t'(\alpha_1, \dots, \alpha_n).$$

4. **Exchange.** For each rule  $r : [t(x_1, \dots, x_n)] \longrightarrow [t'(x_1, \dots, x_n)]$  in  $R$ ,

$$\frac{\alpha_1 : [w_1] \longrightarrow [w'_1] \quad \dots \quad \alpha_n : [w_n] \longrightarrow [w'_n]}{r(\bar{\alpha}) = r(\bar{[w]}); t'(\bar{\alpha}) = t(\bar{\alpha}); r(\bar{[w']})}.$$

---

<sup>6</sup>In the expressions appearing in the equations, when compositions of morphisms are involved, we always implicitly assume that the corresponding domains and codomains match.



Note that the set  $X$  of variables is actually a parameter of these constructions, and we need not assume  $X$  to be fixed and countable. In particular, for  $X = \emptyset$ , we adopt the notation  $\mathcal{T}_{\mathcal{R}}$ . The equations in 1 make  $\mathcal{T}_{\mathcal{R}}(X)$  a category, the equations in 2 make each  $f \in \Sigma$  a functor, and 3 forces the axioms  $E$ . The exchange law states that any rewrite of the form  $r(\bar{\alpha})$ —which represents the *simultaneous* rewriting of the term at the top using rule  $r$  and “below,” i.e., in the subterms matched by the variables, using the rewrites  $\bar{\alpha}$ —is equivalent to the sequential composition  $r(\overline{[w]}); t'(\bar{\alpha})$ , corresponding to first rewriting on top with  $r$  and then below on the subterms matched by the variables with  $\bar{\alpha}$ , and is also equivalent to the sequential composition  $t(\bar{\alpha}); r(\overline{[w']})$  corresponding to first rewriting below with  $\bar{\alpha}$  and then on top with  $r$ . Therefore, the exchange law states that rewriting at the top by means of rule  $r$  and rewriting “below” using  $\bar{\alpha}$  are processes that are independent of each other and can be done either simultaneously or in any order. Since  $[t(x_1, \dots, x_n)]$  and  $[t'(x_1, \dots, x_n)]$  can be regarded as functors  $\mathcal{T}_{\mathcal{R}}(X)^n \rightarrow \mathcal{T}_{\mathcal{R}}(X)$ , from the mathematical point of view the exchange law just asserts that  $r$  is a *natural transformation*, i.e.,

LEMMA 10. [Meseguer, 1992] For each rewrite rule  $r : [t(x_1, \dots, x_n)] \rightarrow [t'(x_1, \dots, x_n)]$  in  $R$ , the family of morphisms

$$\{r(\overline{[w]}) : [t(\bar{w}/\bar{x})] \rightarrow [t'(\bar{w}/\bar{x})] \mid \overline{[w]} \in T_{\Sigma, E}(X)^n\}$$

is a natural transformation  $r : [t(x_1, \dots, x_n)] \Rightarrow [t'(x_1, \dots, x_n)]$  between the functors  $[t(x_1, \dots, x_n)], [t'(x_1, \dots, x_n)] : \mathcal{T}_{\mathcal{R}}(X)^n \rightarrow \mathcal{T}_{\mathcal{R}}(X)$ .

The exchange law provides a way of *abstracting* a rewriting computation by considering immaterial the order in which rewrites are performed “above” and “below” in the term; further abstraction among proof terms is obtained from the functoriality equations. The equations 1-4 provide in a sense the *most abstract* “true concurrency” view of the computations of the rewrite theory  $\mathcal{R}$  that can reasonably be given.

The category  $\mathcal{T}_{\mathcal{R}}(X)$  is just one among many *models* that can be assigned to the rewrite theory  $\mathcal{R}$ . The general notion of model, called an  *$\mathcal{R}$ -system*, is defined as follows:

DEFINITION 11. Given a rewrite theory  $\mathcal{R} = (\Sigma, E, L, R)$ , an  *$\mathcal{R}$ -system*  $\mathcal{S}$  is a category  $\mathcal{S}$  together with:

- a  $(\Sigma, E)$ -algebra structure given by a family of functors

$$\{f_{\mathcal{S}} : \mathcal{S}^n \rightarrow \mathcal{S} \mid f \in \Sigma_n, n \in \mathbb{N}\}$$

satisfying the equations  $E$ , i.e., for any  $t(x_1, \dots, x_n) = t'(x_1, \dots, x_n)$  in  $E$  we have an identity of functors  $t_{\mathcal{S}} = t'_{\mathcal{S}}$ , where the functor  $t_{\mathcal{S}}$  is defined inductively from the functors  $f_{\mathcal{S}}$  in the obvious way.

- for each rewrite rule  $r : [t(\bar{x})] \rightarrow [t'(\bar{x})]$  in  $R$  a natural transformation  $r_{\mathcal{S}} : t_{\mathcal{S}} \Rightarrow t'_{\mathcal{S}}$ .

An  $\mathcal{R}$ -homomorphism  $F : \mathcal{S} \rightarrow \mathcal{S}'$  between two  $\mathcal{R}$ -systems is then a functor  $F : \mathcal{S} \rightarrow \mathcal{S}'$  such that it is a  $\Sigma$ -algebra homomorphism—i.e.,  $f_{\mathcal{S}} * F = F^n * f_{\mathcal{S}'}$ , for each  $f$  in  $\Sigma_n$ ,  $n \in \mathbb{N}$ —and such that “ $F$  preserves  $R$ ,” i.e., for each rewrite rule  $r : [t(\bar{x})] \rightarrow [t'(\bar{x})]$  in  $R$  we have the identity of natural transformations<sup>7</sup>  $r_{\mathcal{S}} * F = F^n * r_{\mathcal{S}'}$ , where  $n$  is the number of variables appearing in the rule. This defines a category  $\mathcal{R}\text{-Sys}$  in the obvious way.

The above definition captures formally the idea that the models of a rewrite theory are *systems*. By a “system” we mean a machine-like entity that can be in a variety of *states*, and that can change its state by performing certain *transitions*. Such transitions are transitive, and it is natural and convenient to view states as “idle” transitions that do not change the state. In other words, a system can be naturally regarded as a *category*, whose objects are the states of the system and whose morphisms are the system’s transitions.

For *sequential* systems such as labelled transition systems this is in a sense the end of the story; such systems exhibit *nondeterminism*, but do not have the required algebraic structure in their states and transitions to exhibit true concurrency. Indeed, what makes a system *concurrent* is precisely the existence of an additional *algebraic structure* [Meseguer, 1992]. First, the states themselves are distributed according to such a structure; for example, for Petri nets [Reisig, 1995] the distribution takes the form of a multiset. Second, concurrent transitions are themselves distributed according to the same algebraic structure; this is what the notion of  $\mathcal{R}$ -system captures, and is for example manifested in the concurrent firing of Petri nets, the evolution of concurrent object-oriented systems [Meseguer, 1993] and, more generally, in any type of concurrent rewriting.

The expressive power of rewrite theories to specify concurrent transition systems is greatly increased by the possibility of having not only transitions, but also *parameterized transitions*, i.e., *procedures*. This is what rewrite rules with variables provide. The family of states to which the procedure applies is given by those states where a component of the (distributed) state is a substitution instance of the lefthand side of the rule in question. The rewrite rule is then a *procedure* which transforms the state *locally*, by replacing such a substitution instance by the corresponding substitution instance of the righthand side. The fact that this can take place concurrently with other transitions “below” is precisely what the concept of a *natural transformation* formalizes. The following table summarizes our present discussion:

---

<sup>7</sup>Note that we use diagrammatic order for the *horizontal*,  $\alpha * \beta$ , and *vertical*,  $\gamma; \delta$ , composition of natural transformations [Mac Lane, 1971].

<i>System</i>	$\longleftrightarrow$	<i>Category</i>
<i>State</i>	$\longleftrightarrow$	<i>Object</i>
<i>Transition</i>	$\longleftrightarrow$	<i>Morphism</i>
<i>Procedure</i>	$\longleftrightarrow$	<i>Natural Transformation</i>
<i>Distributed Structure</i>	$\longleftrightarrow$	<i>Algebraic Structure</i>

A detailed proof of the following theorem on the existence of initial and free  $\mathcal{R}$ -systems for the more general case of conditional rewrite theories is given in [Meseguer, 1992], where the soundness and completeness of rewriting logic for  $\mathcal{R}$ -system models is also proved.

**THEOREM 12.**  $\mathcal{T}_{\mathcal{R}}$  is an initial object in the category  $\mathcal{R}\text{-Sys}$ . More generally,  $\mathcal{T}_{\mathcal{R}}(X)$  has the following universal property: Given an  $\mathcal{R}$ -system  $\mathcal{S}$ , each function  $F : X \rightarrow |\mathcal{S}|$  extends uniquely to an  $\mathcal{R}$ -homomorphism  $F^{\natural} : \mathcal{T}_{\mathcal{R}}(X) \rightarrow \mathcal{S}$ .

#### *Preorder, poset, and algebra models*

Since  $\mathcal{R}$ -systems are an “essentially algebraic” concept<sup>8</sup>, we can consider classes  $\Theta$  of  $\mathcal{R}$ -systems defined by the satisfaction of additional equations. Such classes give rise to full subcategory inclusions  $\Theta \hookrightarrow \mathcal{R}\text{-Sys}$ , and by general universal algebra results about essentially algebraic theories [Barr and Wells, 1985] such inclusions are *reflective* [Mac Lane, 1971], i.e., for each  $\mathcal{R}$ -system  $\mathcal{S}$  there is an  $\mathcal{R}$ -system  $R_{\Theta}(\mathcal{S}) \in \Theta$  and an  $\mathcal{R}$ -homomorphism  $\rho_{\Theta}(\mathcal{S}) : \mathcal{S} \rightarrow R_{\Theta}(\mathcal{S})$  such that for any  $\mathcal{R}$ -homomorphism  $F : \mathcal{S} \rightarrow \mathcal{D}$  with  $\mathcal{D} \in \Theta$  there is a unique  $\mathcal{R}$ -homomorphism  $F^{\diamond} : R_{\Theta}(\mathcal{S}) \rightarrow \mathcal{D}$  such that  $F = \rho_{\Theta}(\mathcal{S}); F^{\diamond}$ . The assignment  $\mathcal{S} \mapsto R_{\Theta}(\mathcal{S})$  extends to a functor  $\mathcal{R}\text{-Sys} \rightarrow \Theta$ , called the *reflection functor*.

Therefore, we can consider subcategories of  $\mathcal{R}\text{-Sys}$  that are defined by certain equations and be guaranteed that they have initial and free objects, that they are closed by subobjects and products, etc. Consider for example the following equations:

$$\begin{aligned} \forall f, g \in \text{Arrows}, f = g \text{ if } \partial_0(f) = \partial_0(g) \wedge \partial_1(f) = \partial_1(g) \\ \forall f, g \in \text{Arrows}, f = g \text{ if } \partial_0(f) = \partial_1(g) \wedge \partial_1(f) = \partial_0(g) \\ \forall f \in \text{Arrows}, \partial_0(f) = \partial_1(f), \end{aligned}$$

where  $\partial_0(f)$  and  $\partial_1(f)$  denote the source and target of an arrow  $f$  respectively. The first equation forces a category to be a preorder, the addition of the second requires this preorder to be a poset, and the three equations

---

<sup>8</sup>In the precise sense of being specifiable by an “essentially algebraic theory” or a “sketch” [Barr and Wells, 1985]; see [Meseguer, 1992] for more details.

together force the poset to be *discrete*, i.e., just a set. By imposing the first one, the first two, or all three, we get full subcategories

$$\mathcal{R}\text{-Alg} \subseteq \mathcal{R}\text{-Pos} \subseteq \mathcal{R}\text{-Preord} \subseteq \mathcal{R}\text{-Sys}.$$

A routine inspection of  $\mathcal{R}\text{-Preord}$  for  $\mathcal{R} = (\Sigma, E, L, R)$  reveals that its objects are preordered  $\Sigma$ -algebras  $(A, \leq)$ —i.e., preordered sets with a  $\Sigma$ -algebra structure such that all the operations in  $\Sigma$  are monotonic—that satisfy the equations  $E$  and such that for each rewrite rule  $r : [t(\bar{x})] \rightarrow [t'(\bar{x})]$  in  $R$  and for each  $\bar{a} \in A^n$  we have  $t_A(\bar{a}) \leq t'_A(\bar{a})$ . The poset case is entirely analogous, except that the relation  $\leq$  is a partial order instead of being a preorder. Finally,  $\mathcal{R}\text{-Alg}$  is the category of ordinary  $\Sigma$ -algebras that satisfy the equations  $E \cup eq(R)$ , where  $eq(r : [t] \rightarrow [t']) = \{t_1 = t_2 \mid t_1 \in [t] \text{ and } t_2 \in [t']\}$ , and  $eq(R) = \bigcup \{eq(r : [t] \rightarrow [t']) \mid [t] \rightarrow [t'] \in R\}$ .

The reflection functor associated with the inclusion  $\mathcal{R}\text{-Preord} \subseteq \mathcal{R}\text{-Sys}$  sends  $\mathcal{T}_{\mathcal{R}}(X)$  to the familiar  $\mathcal{R}$ -rewriting relation<sup>9</sup>  $\rightarrow_{\mathcal{R}(X)}$  on  $E$ -equivalence classes of terms with variables in  $X$ . Similarly, the reflection associated to the inclusion  $\mathcal{R}\text{-Pos} \subseteq \mathcal{R}\text{-Sys}$  maps  $\mathcal{T}_{\mathcal{R}}(X)$  to the partial order  $\leq_{\mathcal{R}(X)}$  obtained from the preorder  $\rightarrow_{\mathcal{R}(X)}$  by identifying any two  $[t], [t']$  such that  $[t] \rightarrow_{\mathcal{R}(X)} [t']$  and  $[t'] \rightarrow_{\mathcal{R}(X)} [t]$ . Finally, the reflection functor into  $\mathcal{R}\text{-Alg}$  maps  $\mathcal{T}_{\mathcal{R}}(X)$  to  $T_{\mathcal{R}}(X)$ , the free  $\Sigma$ -algebra on  $X$  satisfying the equations  $E \cup eq(R)$ ; therefore, the classical *initial algebra semantics* of (functional) equational specifications reappears here associated with a very special class of models which—when viewed as systems—have only trivial identity transitions.

## 4 REWRITING LOGIC AS A LOGICAL FRAMEWORK

The adequacy of rewriting logic as a logical framework in which other logics can be represented by means of maps of logics or of entailment systems is explored by means of relevant examples, including equational, Horn, and linear logic, a general approach to the treatment of quantifiers, and a very general method for representing sequent presentations of a logic.

### 4.1 Mapping equational logic

As mentioned in Section 3.2, one can get equational logic from rewriting logic by adding the symmetry rule. Moreover, the syntax of rewriting logic includes equations in order to impose structural axioms on terms. Therefore, it should not be surprising to find out that there are many connections between both logics.

---

<sup>9</sup>It is perhaps more suggestive to call  $\rightarrow_{\mathcal{R}(X)}$  the *reachability relation* of the system  $\mathcal{T}_{\mathcal{R}}(X)$ .

Even in the case of equational logic it can be convenient to allow sometimes a distinction between structural axioms and equations, so that an equational theory can then be described as a triple  $(\Sigma, E, Q)$ , with  $Q$  a set of equations of the form  $[u]_E = [v]_E$ . This increases the expressiveness of equational theories, because we can allow more flexible description of equations—for example, omitting parentheses in the case when  $E$  contains an associativity axiom—and also supports a built-in treatment of the structural axioms in equational deduction. Indeed, this is fully consistent with the distinction made in OBJ3 and in Maude’s functional modules between the equational *attributes* of an operator—such as associativity, commutativity, etc.—which are declared together with the operator, and the equations given, which are used modulo such attributes.

In order to define a map of entailment systems

$$(\Phi, \alpha) : \text{ent}(OSEqtl) \longrightarrow \text{ent}(OSRWLogic)$$

in principle we need to map an equation  $[u]_E = [v]_E$  to a sequent, and the obvious choices are either  $[u]_E \longrightarrow [v]_E$  or  $[v]_E \longrightarrow [u]_E$ . However this choice involves giving a fixed orientation to an equation, with the well-known problems that this causes. To avoid this choice, we would like to give the equation *both* orientations. We can achieve this by slightly generalizing Definition 5 of map of entailment systems in such a way that a sentence is mapped to a set of sentences<sup>10</sup>. In our case,  $\alpha$  maps an equation  $[u]_E = [v]_E$  to the set of sequents  $\{[u]_E \longrightarrow [v]_E, [v]_E \longrightarrow [u]_E\}$ , and  $\Phi$  maps an equational theory  $T = (\Sigma, E, Q)$  to the rewrite theory  $\Phi(T) = (\Sigma, E, L, \alpha(Q))$ , where  $\alpha(Q) = \bigcup\{\alpha(e) \mid e \in Q\}$ , and  $L$  is a labelling of the rewrite rules such that, for example, each rule is labelled by itself. This map satisfies

$$(\Sigma, E, Q) \vdash_{EL} e \iff (\Sigma, E, L, \alpha(Q)) \vdash_{RL} \alpha(e).$$

This can be easily proved by induction on the deduction rules of equational logic, using the fact that all the rules of rewriting logic are also rules of equational logic and the following lemma.

LEMMA 13.

$$(\Sigma, E, L, \alpha(Q)) \vdash_{RL} [u] \rightarrow [v] \iff (\Sigma, E, L, \alpha(Q)) \vdash_{RL} [v] \rightarrow [u].$$

Therefore, we have a *conservative* map of entailment systems.

In order to extend this map to a map of logics, a simple idea concerning models is to send a  $\Phi(T)$ -system  $\mathcal{C}$  to  $R_{\mathbf{Alg}}(\mathcal{C})$ , where  $R_{\mathbf{Alg}}$  is the reflection functor associated with the inclusion  $\Phi(T)\text{-}\mathbf{Alg} \subseteq \Phi(T)\text{-}\mathbf{Sys}$ , as discussed

<sup>10</sup>This generalization is also very useful in relating other logics; see for example [Meseguer, 1998].

in Section 3.5. By definition,  $R_{\mathbf{Alg}}(\mathcal{C})$  is a model of the equational theory  $T$ . However, this map does *not* satisfy the condition in Definition 6 of map of institutions. The difficulty is that, in general, from an equation  $t = t'$  one can deduce that there is a chain  $t \rightarrow t_1 \leftarrow t_2 \cdots t_n \leftarrow t'$ , but not that  $t \rightarrow t'$ , as the reader familiar with term rewriting knows. To solve this problem, we consider a different quotient of the underlying  $(\Sigma, E)$ -algebra  $|\mathcal{C}|$  in which two objects  $A$  and  $B$  are identified if and only if there exist morphisms  $f : A \rightarrow B$  and  $g : B \rightarrow A$  in  $\mathcal{C}$ . In this way, we obtain a  $(\Sigma, E)$ -algebra  $\beta_T(\mathcal{C})$  that satisfies all the sentences in  $Q$ . Moreover, the condition in Definition 6 of map of institutions holds for this map. In short, we have obtained a conservative map of logics

$$(\Phi, \alpha, \beta) : OSEqtl \longrightarrow OSRWLogic.$$

There is also another map of logics

$$(\Phi', \alpha', \beta') : OSEqtl \longrightarrow OSRWLogic$$

that, instead of sending equations to sequents, sends equations to equations. This requires making explicit the fact, left implicit in Section 3, that equations can also be considered as sentences of rewriting logic, where, by definition,

$$(\Sigma, E, L, R) \vdash_{RL} t = t' \iff E \vdash_{EL} t = t'.$$

From this point of view,  $\Phi'$  maps an equational theory  $(\Sigma, E)$  to the rewrite theory  $(\Sigma, E, \emptyset, \emptyset)$ , and at the level of sentences  $\alpha'$  is just an inclusion, trivially satisfying the requirement for a map of entailment systems. Note that in this context the distinction between structural axioms and equations is not necessary.

With respect to the models,  $\beta'_T$  maps a  $(\Sigma, E, \emptyset, \emptyset)$ -system  $\mathcal{C}$  to the underlying  $(\Sigma, E)$ -algebra structure on  $|\mathcal{C}|$ , trivially satisfying also the condition in Definition 6 and being therefore a map of institutions. Notice that  $(\Phi', \alpha', \beta')$  is *conservative* in a straightforward way.

On the opposite direction there is also a map of logics

$$(\Psi, \gamma, \delta) : OSRWLogic \longrightarrow OSEqtl$$

mapping a rewrite theory  $(\Sigma, E, L, R)$  to the equational theory  $(\Sigma, E, \gamma(R))$  where  $\gamma$  removes the labels from the rules and turns the sequent signs “ $\longrightarrow$ ” into equality signs. For the models,  $\delta_{\mathcal{R}}$  is the inclusion  $\mathcal{R}\text{-Alg} \subseteq \mathcal{R}\text{-Sys}$  defined in Section 3.5.

Notice that the composition of maps of logics  $(\Phi, \alpha, \beta); (\Psi, \gamma, \delta)$  is the identity.

## 4.2 Mapping Horn logic

Horn logic signatures are of the form  $(F, P)$ , with  $F$  a set of function symbols and  $P$  a set of predicate symbols. In the order-sorted case such symbols have ranks  $f : s_1 \dots s_n \rightarrow s$ , and  $p : s_1 \dots s_n$ , specified by strings of sorts in the poset of sorts  $S$ . Models are  $F$ -algebras  $M$  together with, for each predicate symbol  $p : s_1 \dots s_n$ , a subset  $p_M \subseteq M_{s_1} \times \dots \times M_{s_n}$ , which can alternatively be viewed as a characteristic function  $p_M : M_{s_1} \times \dots \times M_{s_n} \rightarrow Bool$  to the two element Boolean algebra  $Bool$ . Satisfaction of a Horn clause

$$q_1(\bar{u}_1), \dots, q_n(\bar{u}_n) \Rightarrow p(\bar{t})$$

in a model  $M$  can be expressed as either the subset containment of the intersection of the interpretations of  $q_1(\bar{u}_1), \dots, q_n(\bar{u}_n)$  in  $M$  inside the corresponding interpretation of  $p(\bar{t})$ , or, in a characteristic function description, as the functional inequality

$$q_1(\bar{u}_1)_M \text{ and } \dots \text{ and } q_n(\bar{u}_n)_M \leq p(\bar{t})_M$$

between the corresponding interpretations in  $M$  of the conjunction of the premises and of the conclusion as characteristic functions, where the inequality between the functions means inequality of their values for each of the arguments in the Boolean algebra ordering. A *homomorphism*  $f : M \rightarrow M'$  between two such models is an  $F$ -homomorphism which in addition satisfies  $(f_{s_1} \times \dots \times f_{s_n})(p_M) \subseteq p_{M'}$  for each  $p : s_1 \dots s_n$ , or in characteristic function form the functional inequality

$$p_M \leq (f_{s_1} \times \dots \times f_{s_n}); p_{M'}.$$

Horn logic is a particularly simple logic that does not use the full power of classical first-order logic and is in fact compatible with a variety of other nonclassical interpretations such as for example intuitionistic logic. It is therefore reasonable to enlarge the class of models just described by keeping the  $F$ -algebra parts as before, but allowing instead interpretations of the predicate symbols  $p$  as “characteristic functions”

$$p_M : M_{s_1} \times \dots \times M_{s_n} \rightarrow M_{Prop}$$

into a partially ordered set  $M_{Prop}$  of “propositions” which is not required to be fixed, i.e., it can vary from model to model. We require of any such poset the “bare minimum” structure of having a top element *true*:  $Prop Prop \rightarrow Prop$  that is monotonic and has *true* as its neutral element. Of course,  $Bool$  is one such poset, where conjunction is interpreted as *and*. Satisfaction of Horn clauses can be defined by a functional inequality just as before, but changing  $Bool$  by the appropriate poset  $M_{Prop}$  being chosen for the model.

The natural generalization of the notion of homomorphism  $f : M \rightarrow M'$  is to again require an  $F$ -homomorphism for the operations in  $F$ , whereas for predicate symbols  $p : s_1 \dots s_n$  we require the functional inequality

$$(\dagger) \quad p_M; f_{Prop} \leq (f_{s_1} \times \dots \times f_{s_n}); p_{M'}$$

where  $f_{Prop} : M_{Prop} \rightarrow M'_{Prop}$  is an additional component of the homomorphism, namely, a monotonic function preserving *true* and conjunction “up to inequality” between the posets of propositions  $M_{Prop}$  and  $M'_{Prop}$  chosen for the models  $M$  and  $M'$ , in the sense that we have  $f_{Prop}(true_M) \leq true_{M'}$ , and  $f_{Prop}(x, y) \leq f_{Prop}(x), f_{Prop}(y)$ , for  $x, y \in M_{Prop}$ . This defines a category of models  $(F, P)$ -**Mod**.

In addition, we can consider the generalization to Horn theories of the form  $(F, P, E, H)$  where  $E$  is a set of  $F$ -equations, and  $H$  is a set of Horn clauses involving the predicates in  $P$  but not equations (again, equations in  $E$  can be viewed as structural axioms forming part of the signature). A model satisfies this theory when the underlying  $F$ -algebra satisfies all the equations in  $E$  and the model satisfies the Horn clauses in  $H$ , defining in this way a full subcategory  $(F, P, E, H)$ -**Mod** of  $(F, P)$ -**Mod**. We denote by  $OSHorn^=$  the logic whose theories are such generalized Horn theories  $(F, P, E, H)$  with equational axioms  $E$ , and whose models we have just described.

The map of logics

$$(\Phi, \alpha, \beta) : OSHorn^= \rightarrow OSRWLogic$$

that we define now is a considerable simplification and extension of the map described in [Meseguer, 1992b].

A Horn theory  $(F, P, E, H)$  is mapped to a rewrite theory

$$\Phi(F, P, E, H) = (F \cup P^\diamond, E \cup ACI, \{*\} \cup H, \{x_{Prop} \rightarrow true\} \cup H^\diamond),$$

where

- $F \cup P^\diamond$  is the order-sorted signature that extends  $F$  by adding the additional sort *Prop*, a constant *true* : *Prop*, a binary operator  $\_ , \_$  on *Prop*, and, for each predicate symbol  $p : s_1 \dots s_n$  in  $P$ , an operator  $p : s_1 \dots s_n \rightarrow Prop$ ;
- *ACI* is the set of associativity, commutativity, and identity (*true*) structural axioms for the conjunction operator  $\_ , \_$ ;
- “\*” is the label for the rewrite rule  $x_{Prop} \rightarrow true$ , where  $x_{Prop}$  is a variable of sort *Prop*;
- $H^\diamond$  is a set of rewrite rules labelled by the Horn clauses  $H$  themselves in such a way that a Horn clause of the form  $q_1(\bar{u}_1), \dots, q_n(\bar{u}_n) \Rightarrow p(\bar{t})$  labels the rewrite rule  $q_1(\bar{u}_1), \dots, q_n(\bar{u}_n) \rightarrow p(\bar{t})$ , whereas a Horn clause of the form  $p(\bar{t})$  labels the rewrite rule  $true \rightarrow p(\bar{t})$ .



At the level of sentences,  $\alpha$  maps each Horn clause to its corresponding labelled rewrite rule in the above manner.

As to models, given a Horn theory  $T$ , a  $\Phi(T)$ -system consists of a category  $\mathcal{C}_s$  for each sort  $s$  in the poset  $S$ , and a category  $\mathcal{P}$  for the sort *Prop*, together with a collection of functors satisfying the equations in  $\Phi(T)$  and natural transformations interpreting the rewrite rules in  $\Phi(T)$ . The functor  $\beta_T$  sends such a system to the  $T$ -model consisting of the underlying (order-sorted) algebra structure on the family of sets  $\{\mathcal{C}_s \mid s \in S\}$ , and the poset  $R_{\mathbf{Pos}}(\mathcal{P})$ , where  $R_{\mathbf{Pos}}$  is the reflection functor associated to the inclusion  $\Phi(T)\text{-Pos} \subseteq \Phi(T)\text{-Sys}$ , discussed in Section 3.5. By definition of this reflection functor,  $A \leq B$  in  $R_{\mathbf{Pos}}(\mathcal{P})$  if and only if there is a morphism  $A \rightarrow B$  in  $\mathcal{P}$ . Therefore, a Horn clause  $q_1(\bar{u}_1), \dots, q_n(\bar{u}_n) \Rightarrow p(\bar{t})$  is satisfied by this  $T$ -model if and only if there is a morphism in  $\mathcal{P}$  interpreting the rewrite sequent  $q_1(\bar{u}_1), \dots, q_n(\bar{u}_n) \longrightarrow p(\bar{t})$  if and only if this sequent is satisfied by the original  $\Phi(T)$ -system. Thus,  $(\Phi, \alpha, \beta)$  is indeed a map of institutions.

Notice that, by the conditions for  $\mathcal{R}$ -homomorphisms in Definition 11, for the homomorphisms in the image of  $\beta_T$  the functional inequality  $(\dagger)$  above becomes an equality. In addition,  $\beta_T$  maps free  $\Phi(T)$ -systems to (weakly) free Horn  $T$ -models; since the entailment relation coincides with satisfaction in free models (see the proof of Theorem 3.13 in [Meseguer, 1992]), this provides a short proof of the fact that  $(\Phi, \alpha)$  is indeed a map of entailment systems, and moreover, it is *conservative*.

The same discussion applies to the case of preorders instead of posets, by considering the reflection functor associated to the inclusion  $\Phi(T)\text{-Preord} \subseteq \Phi(T)\text{-Sys}$ , which would have given a slightly more general notion of model for a Horn theory in which propositions would form a preorder.

### 4.3 Mapping linear logic

In this section, we describe a map of logics  $LinLogic \longrightarrow OSRWLogic$  mapping theories in full quantifier-free first-order linear logic to rewrite theories. We do not provide much motivation for linear logic, referring the reader to [Girard, 1987; Troelstra, 1992; Martí-Oliet and Meseguer, 1991] for example. We need to point out, nonetheless, the way linear logic satisfies the conditions given in Definition 1 of entailment system. If one thinks of formulas as sentences and of the turnstile symbol “ $\vdash$ ” in a sequent as the entailment relation, then this relation is not monotonic, because in linear logic the structural rules of weakening and contraction are forbidden, so that, for example, we have the sequent  $A \vdash A$  as an axiom, but we cannot derive either  $A, B \vdash A$  or even  $A, A \vdash A$ . The point is that, for  $\Sigma$  a linear logic signature, the elements of  $sen(\Sigma)$  should not be identified with *formulas* but with *sequents*. Viewed as a way of generating sequents, i.e., identifying our entailment relation  $\vdash$  with the closure

of the horizontal bar relation among linear logic sequents, the entailment of linear logic is indeed reflexive, monotonic, and transitive. This idea is also supported by the categorical models for linear logic [Seely, 1989; Martí-Oliet and Meseguer, 1991], in which sequents are interpreted as morphisms, and leads to a very natural correspondence between the models of rewriting and linear logic.

### *Expressing linear logic in rewriting logic*

We use the syntax of the Maude language to write down the map of entailment systems from linear logic to rewriting logic. Note that any sequence of characters starting with either “---” or “\*\*\*” and ending with “end-of-line” is a comment. Moreover, we usually drop the equivalence class square brackets, adopting the convention that a term  $t$  denotes the equivalence class  $[t]_E$  for the appropriate set of structural axioms  $E$ .

We first define the *functional* theory  $\text{PROPO}[X]$  which introduces the syntax of propositions as a parameterized abstract data type. The parameterization permits having additional structure at the level of atoms if desired. In order to provide a proper treatment of negation, only equations are given, and no rewrite rules are introduced in this theory; they are introduced afterwards in the  $\text{LINLOG}[X]$  theory. The purpose of the equations in the  $\text{PROPO}[X]$  theory is to push negation to the atom level, by using the dualities of linear logic; this is a well-known process in classical and linear logic.

```
fth ATOM is
  sort Atom .
endfth

--- linear logic syntax
fth PROPO[X :: ATOM] is
  sort Prop0 .
  subsort Atom < Prop0 .
  ops 1 0 ⊥ ⊤ : -> Prop0 .
  op _⊥_ : Prop0 -> Prop0 .
  op _⊗_ : Prop0 Prop0 -> Prop0 [assoc comm id: 1] .
  op _⌘_ : Prop0 Prop0 -> Prop0 [assoc comm id: ⊥] .
  op _⊕_ : Prop0 Prop0 -> Prop0 [assoc comm id: 0] .
  op _&_ : Prop0 Prop0 -> Prop0 [assoc comm id: ⊤] .
  op !_ : Prop0 -> Prop0 .
  op ?_ : Prop0 -> Prop0 .

  vars A B : Prop0 .
  eq (A ⊗ B)⊥ = A⊥ ⌘ B⊥ .
  eq (A ⌘ B)⊥ = A⊥ ⊗ B⊥ .
  eq (A & B)⊥ = A⊥ ⊕ B⊥ .
```

```

eq (A ⊕ B)⊥ = A⊥ & B⊥ .
eq (!A)⊥ = ?(A⊥) .
eq (?A)⊥ = !(A⊥) .
eq A⊥⊥ = A .
eq 1⊥ = ⊥ .
eq ⊥⊥ = 1 .
eq ⊤⊥ = 0 .
eq 0⊥ = ⊤ .
endft

```

Note that the equations can be used as oriented rules from left to right at the implementation level in order to obtain a canonical form for expressions in `Prop0`.

The `LINLOG[X]` theory introduces linear logic propositions and the rules of the logic. Propositions are of the form `[A]` for `A` an expression in `Prop0`. All logical connectives work similarly for `Prop0` expressions and for propositions, except negation, which is defined only for `Prop0` expressions.

Some presentations of linear logic are given in the form of one-sided sequents  $\vdash \Gamma$  where negation has been pushed to the atom level, and there are no rules for negation in the sequent calculus [Girard, 1987]. In this section, in order to make the connections with category theory and with rewriting logic more direct, we prefer to use standard sequents of the more general form  $\Gamma \vdash \Delta$ . In a later section, we will also use one-sided sequents just in order to reduce the number of rules.

The style of our formulation adopts a categorical viewpoint for the proof theory and semantics of linear logic [Seely, 1989; Martí-Oliet and Meseguer, 1991]. This style exploits the close connection between the models of linear logic and those of rewriting logic which are also categories, as we have explained in Section 3.5. Without going into details that the reader can find for example in [Martí-Oliet and Meseguer, 1991] and the references therein, the tensor and linear implication connectives are interpreted in a closed symmetric monoidal category  $(\mathcal{C}, \otimes, \multimap)$ . Negation is interpreted by means of a dualizing object  $\perp$  and the definition  $A^\perp = A \multimap \perp$  (with this definition of negation,  $\mathcal{C}$  becomes a  $*$ -autonomous category [Barr, 1979]). The categorical product  $\&$  interprets additive conjunction. The interpretation of the exponential  $!$  is given by a comonad  $(!A, !A \rightarrow A, !A \rightarrow !!A)$  that maps the comonoid structure  $\top \leftarrow A \rightarrow A \& A$  into a comonoid structure  $1 \leftarrow !A \rightarrow !A \otimes !A$  via isomorphisms  $!\top \cong 1$  and  $!(A \& A) \cong !A \otimes !A$ .

The dual connectives  $\wp$ ,  $\oplus$ , and  $?$  can be defined using negation:  $A \wp B = (A^\perp \otimes B^\perp)^\perp = A^\perp \multimap B$ ,  $A \oplus B = (A^\perp \& B^\perp)^\perp$ ,  $?A = (!A^\perp)^\perp$ . Without negation,  $\oplus$  needs the presence of coproducts and  $?$  is interpreted by means of a monad with a monoid structure.

When seeking the minimal categorical structure required for interpreting linear logic, an important question is how to interpret the connective  $\wp$  without using negation, and how to axiomatize its relationship with the

tensor  $\otimes$ . Cockett and Seely have answered this question with the notion of a *weakly distributive category* [Cockett and Seely, 1992]. A weakly distributive category consists of a category  $\mathcal{C}$  with two symmetric tensor products  $\otimes, \wp : \mathcal{C} \times \mathcal{C} \rightarrow \mathcal{C}$ , and a natural transformation  $A \otimes (B \wp C) \rightarrow (A \otimes B) \wp C$  (weak distributivity) satisfying some coherence equations<sup>11</sup>. Negation is added to a weakly distributive category by means of a function  $(\_)^\perp : |\mathcal{C}| \rightarrow |\mathcal{C}|$  on the objects of  $\mathcal{C}$ , and natural transformations  $1 \rightarrow A \wp A^\perp$  and  $A \otimes A^\perp \rightarrow \perp$  satisfying some coherence equations. Cockett and Seely then prove that the concepts of weakly distributive category with negation and of  $*$ -autonomous category are equivalent, providing in this way a categorical semantics for linear logic in which the *par* connective  $\wp$  is primitive and is not defined in terms of tensor and negation.

In the following theory, the rewrite rules for  $\otimes, \wp$ , and negation correspond to the natural transformations in the definition of a weakly distributive category, as explained above. The rules for  $\&$  ( $\oplus$ , respectively) mirror the usual definition of final object and product (initial object and coproduct, respectively). Finally, the axioms and rules for the exponential  $!$  ( $?$ , respectively) correspond to the comonad with a comonoid structure (monad with monoid structure, respectively). Note that some rules are redundant, but we have decided to include them in order to make the connectives less interdependent, so that, for example, if the connective  $\&$  is omitted we do not need to add new rules for the modality  $!$ .

```

--- linear logic rules
th LINLOG[X :: ATOM] is
  protecting PROPO[X] .
  sort Prop .
  ops 1 0  $\perp$   $\top$  : -> Prop .
  op  $\_ \otimes \_$  : Prop Prop -> Prop [assoc comm id: 1] .
  op  $\_ \wp \_$  : Prop Prop -> Prop [assoc comm id:  $\perp$ ] .
  op  $\_ \oplus \_$  : Prop Prop -> Prop [assoc comm id: 0] .
  op  $\_ \& \_$  : Prop Prop -> Prop [assoc comm id:  $\top$ ] .
  op  $! \_$  : Prop -> Prop .
  op  $? \_$  : Prop -> Prop .

  op  $[\_]$  : Prop0 -> Prop .

  vars A B : Prop0 .
  ax  $[A \otimes B] = [A] \otimes [B]$  .
  ax  $[A \wp B] = [A] \wp [B]$  .
  ax  $[A \& B] = [A] \& [B]$  .
  ax  $[A \oplus B] = [A] \oplus [B]$  .
  ax  $[!A] = ![A]$  .

```

<sup>11</sup>Cockett and Seely develop in [1992] the more general case in which the tensor products are not assumed to be symmetric.

```

ax [?A] = ?[A] .
ax [1] = 1 .
ax [ $\perp$ ] =  $\perp$  .
ax [ $\top$ ] =  $\top$  .
ax [0] = 0 .

*** [_] is injective
cax A = B if [A] = [B] .

*** Rules for negation
rl 1 => [A]  $\wp$  [A⊥] .
rl [A]  $\otimes$  [A⊥] =>  $\perp$  .

vars P Q R : Prop .
*** Rules for  $\otimes$  and  $\wp$ 
rl P  $\otimes$  (Q  $\wp$  R) => (P  $\otimes$  Q)  $\wp$  R .

*** Rules for &
rl P =>  $\top$  . *** (1)
rl P & Q => P .
crl R => P & Q if R => P and R => Q . *** (2)

*** Rules for  $\oplus$ 
rl 0 => P . *** (3)
rl P => P  $\oplus$  Q .
crl P  $\oplus$  Q => R if P => R and Q => R . *** (4)

*** Structural axioms and rules for !
ax !(P & Q) = !P  $\otimes$  !Q . *** (5)
ax ! $\top$  = 1 . *** (6)
rl !P => P .
rl !P => !!P .
rl !P => 1 . *** redundant from (1) and (6) above
rl !P => !P  $\otimes$  !P . *** redundant from (2) and (5) above

*** Structural axioms and rules for ?
ax ?(P  $\oplus$  Q) = ?P  $\wp$  ?Q . *** (7)
ax ?0 =  $\perp$  . *** (8)
rl P => ?P .
rl ??P => ?P .
rl  $\perp$  => ?P . *** redundant from (3) and (8) above
rl ?P  $\wp$  ?P => ?P . *** redundant from (4) and (7) above
endt

```

A linear logic formula is built from a set of propositional constants using the logical constants and connectives of linear logic. Notice that linear implication  $A \multimap B$  is not necessary because it can be defined as  $A^\perp \wp B$ .

*Representing a linear logic theory in rewriting logic*

A *linear theory*  $T$  in propositional linear logic consists of a finite set  $C$  of propositional constants and a finite set  $S$  of double-sided sequents of the form  $A_1, \dots, A_n \vdash B_1, \dots, B_m$ , where each  $A_i$  and  $B_j$  is a linear logic formula built from the constants in  $C$ . Given such a theory  $T$ , it is interpreted in rewriting logic as follows.

First, we define a functional theory to interpret the propositional constants in  $C$ . For example, if  $C = \{a, b, c\}$  we would define

```
fth C is
  sort Atom .
  ops a b c : -> Atom .
endft
```

Then, we can instantiate the parameterized theory  $\text{LINLOG}[X]$  using this functional theory, with the default view  $\text{ATOM} \rightarrow C$ :

```
make LINLOG0 is LINLOG[C] endmk
```

A linear logic formula  $A$  (with constants in  $C$ ) is interpreted in  $\text{LINLOG0}$  as the term  $[A]$  of sort  $\text{Prop}$ . For example, the formula  $(a \otimes b)^\perp \oplus (! (a \& c^\perp))^\perp$  is interpreted as the term

$$[(a \otimes b)^\perp \oplus (! (a \& c^\perp))^\perp]$$

which, using the equations for negation in  $\text{PROPO}[X]$  and the structural axioms in  $\text{LINLOG}[X]$ , is equal to the term

$$([a^\perp] \wp [b^\perp]) \oplus ?([a^\perp] \oplus [c]).$$

Finally, we extend the theory  $\text{LINLOG0}$  by adding a rule

```
r1 [A1] \otimes ... \otimes [An] => [B1] \wp ... \wp [Bm] .
```

for each sequent  $A_1, \dots, A_n \vdash B_1, \dots, B_m$  in the linear theory  $T$ . For example, if

$$T = \{a \otimes b, !c \oplus a \vdash a, (c \oplus b)^\perp, \\ a \wp b, ?(c^\perp) \vdash (?b \wp !c)^\perp, a \oplus b\},$$

the corresponding rewrite theory is

```
th LINLOG(T) is
  including LINLOG0 .
  r1 [a] \otimes [b] \otimes (![c] \oplus [a]) => [a] \wp ([c^\perp] \& [b^\perp]) .
  r1 ([a] \wp [b]) \otimes ?[c^\perp] => (![b^\perp] \otimes ?[c^\perp]) \wp ([a] \oplus [b]) .
endt
```

Note that this technique can also be used to interpret quantifier-free first-order linear logic formulas, where, instead of propositional constants, we have literals built using functions and predicates. In general, we can allow any abstract data type ADT defining constants, functions and predicates. Then, we define the instantiation

```
make LINLOGO is LINLOG[ADT] endmk
```

which is finally extended with the corresponding rules to a theory  $\text{LINLOG}(T)$  corresponding to the desired theory  $T$ .

The main result is the following conservativity theorem.

**THEOREM 14.** Given a linear theory  $T$ , a sequent  $A_1, \dots, A_n \vdash B_1, \dots, B_m$  is provable in linear logic from the axioms in  $T$  if and only if the sequent

$$[A1] \otimes \dots \otimes [An] \longrightarrow [B1] \wp \dots \wp [Bm]$$

is a  $\text{LINLOG}(T)$ -rewrite, i.e., it is provable in rewriting logic from the rewrite theory  $\text{LINLOG}(T)$ .

To show that a linear logic proof can be translated into a rewriting logic proof, the idea is similar to the proof of the soundness theorem for the categorical semantics of linear logic, where a sequent is interpreted as a morphism (see [Martí-Oliet and Meseguer, 1991, Theorem 40]). What is important to realize is that the categorical constructions of these morphisms can be seen as rewriting logic proofs; for example, functoriality corresponds to the *Congruence* rule of rewriting logic, something made completely explicit in the categorical semantics of rewriting logic, as outlined in Section 3.5 and developed in detail in [Meseguer, 1992].

### *The map of logics*

The fully detailed development in the previous sections provides a map of entailment systems between linear logic and rewriting logic, which is conservative because of Theorem 14. We have already discussed briefly the models of linear logic in Section 4.3 by way of motivation to the rules in the theory  $\text{LINLOG}[X]$ . Now, in order to complete the construction of the map of logics  $\text{LinLogic} \longrightarrow \text{OSRWLogic}$ , we need a way of getting a (categorical) model of a linear theory  $T$  from a rewrite system that is a model of the rewrite theory  $\text{LINLOG}(T)$ .

The first thing to note, recalling the definition of  $\mathcal{R}$ -system in Section 3.5, is that for each rewrite rule in  $R$  we require just a natural transformation in the system, but we do not impose any coherence or uniqueness conditions on these natural transformations. For this reason, a  $\text{LINLOG}(T)$ -system interprets  $A \& B$  as a weak product instead of a product, for example. A way of obtaining uniqueness would be considering the generalized rewrite

theories defined in [Meseguer, 1992b], but we do not need that for our purposes here. On the other hand, the attributes of the operations, like associativity or commutativity, are interpreted as identities, instead of the more general natural isomorphisms, thus satisfying all coherence conditions automatically.

In general, given a linear theory  $T = (C, S)$ , a  $\text{LINLOG}(T)$ -system consists of an algebra  $\mathcal{A}$  interpreting all the structure of the functional theory  $\text{PROPO}[C]$ , a category  $\mathcal{C}$  with all the morphisms necessary to interpret the rewrite rules in the theory  $\text{LINLOG}[C]$  and the rules corresponding to all the sequents in  $S$ , and an injective homomorphism  $\mathcal{A} \rightarrow |\mathcal{C}|$  that, without loss of generality, we can consider to be an inclusion. Note that, as  $\mathcal{A}$  is closed under all the operations in the theory  $\text{LINLOG}[C]$ , the full subcategory of  $\mathcal{C}$  generated by  $\mathcal{A}$  has the same structure as  $\mathcal{C}$ , and, in addition, there is a function  $(\_)^\perp : \mathcal{A} \rightarrow \mathcal{A}$  interpreting negation. Therefore, this full subcategory is *almost* a weakly distributive category with negation, products, coproducts, a comonad with a comonoid structure, and a monad with a monoid structure. What is possibly missing is the satisfaction of a set of equations between morphisms which ensure that all this structure is really what we want.

Thus, in order to get a Girard category  $\mathcal{L}$  from the original  $\text{LINLOG}(T)$ -system, we do the quotient of the full subcategory of  $\mathcal{C}$  generated by  $\mathcal{A}$  by this set of equations. Clearly, there is a morphism  $A \rightarrow B$  in  $\mathcal{L}$  if and only if there is a morphism  $A \rightarrow B$  in  $\mathcal{C}$ , i.e.,  $\mathcal{L}$  satisfies a linear sequent if and only if  $\mathcal{C}$  satisfies the rewriting logic version of that sequent. This is true because the constants in  $\mathcal{C}$  are interpreted always as the corresponding constants in  $\mathcal{A}$ , and variables in a sequent are also interpreted as elements of  $\mathcal{A}$  (note that variables appear in a theory  $\text{ADT}$  that is used to instantiate  $\text{PROPO}[X]$ ). In summary, we have a *conservative* map of logics  $\text{LinLogic} \longrightarrow \text{OSRWLogic}$ .

#### 4.4 Quantifiers

In Section 4.3 we have defined a map of logics between quantifier-free linear logic and rewriting logic. In this section, we show how to extend that map at the level of entailment systems to quantifiers. The choice of linear logic to illustrate the treatment of quantifiers is irrelevant; we could have chosen any other logic. It has only the expository advantage of building upon an example already introduced in this paper. In fact, our equational treatment of quantification, inspired by ideas of Laneve and Montanari on the definition of the lambda calculus as a theory in rewriting logic [Laneve and Montanari, 1992; Laneve and Montanari, 1996], is very general and encompasses not only existential and universal quantification, but also lambda abstraction and other such binding mechanisms.

The main idea is to internalize as operations in the theory the notions



of free variables and substitution that are usually defined at the metalevel. Then, the typical definitions of such notions by structural induction on terms can be easily written down as equations in the theory, but, more importantly, we can consider terms modulo these axioms and we can also use the operation of substitution explicitly in the rules introducing or eliminating quantifiers. This is similar to the lambda calculus with explicit substitutions defined by Abadi, Cardelli, Curien, and Lévy in [1991], and to the work on binding structures by Talcott [1993].

We begin by presenting the example of the lambda abstraction binding mechanism in the lambda calculus, as defined by Laneve and Montanari in [1992] (see also [Laneve and Montanari, 1996], where this technique is generalized to combinatory reduction systems). Since in this case the syntax is much simpler, the main ideas can become more explicit and clearer to the reader.

We assume a parameterized functional module `SET[X]` that provides finite sets over a parameter set `X` with operations `_U_` for union, `_-_` for set difference, `{_}` for singleton, `emptyset` for the empty set, and a predicate `_is-in_` for membership.

```

--- variable names
fth VAR is
  sort Var .
  protecting SET[Var] .
  op new : Set -> Var .
  var S : Set .
  eq new(S) is-in S = false . *** new variable
endft

--- lambda calculus syntax with substitution
fmod LAMBDA[X :: VAR] is
  including SET[X] .
  sort Lambda .
  subsort Var < Lambda .
  op λ_ : Var Lambda -> Lambda . *** variables
  op _ : Lambda Lambda -> Lambda . *** lambda abstraction
  op _[_/_] : Lambda Lambda Var -> Lambda . *** application
  op fv : Lambda -> Set . *** substitution
  *** free variables

  vars X Y : Var .
  vars M N P : Lambda .

  *** Free variables
  eq fv(X) = {X} .
  eq fv(λX.M) = fv(M) - {X} .
  eq fv(MN) = fv(M) U fv(N) .
  eq fv(M[N/X]) = (fv(M) - {X}) U fv(N) .
    
```

```

*** Substitution equations
eq X[N/X] = N .
ceq Y[N/X] = Y if not(X == Y) .
eq (MN)[P/X] = (M[P/X])(N[P/X]) .
eq ( $\lambda$ X.M)[N/X] =  $\lambda$ X.M .
ceq ( $\lambda$ Y.M)[N/X] =  $\lambda$ Y.(M[N/X])
    if not(X == Y) and (not(Y is-in fv(N)) or not(X is-in fv(M))) .
ceq ( $\lambda$ Y.M)[N/X] =  $\lambda$ (new(fv(MN))).(M[new(fv(MN))/Y])[N/X]
    if not(X == Y) and Y is-in fv(N) and X is-in fv(M) .
endfm

```

Note that substitution is here another term constructor instead of a meta-syntactic operation. Of course, using the above equations, all occurrences of the substitution constructor can be eliminated. After having defined in the previous functional module the class of lambda terms with substitution, we just need to add the equational axiom of alpha conversion and the beta rule in the following module:

```

--- lambda calculus rules
mod ALPHA-BETA[X :: VAR] is
    including LAMBDA[X] .
    vars X Y : Var .
    vars M N : Lambda .

*** Alpha conversion
cax  $\lambda$ X.M =  $\lambda$ Y.(M[Y/X]) if not(Y is-in fv(M)) .

*** Beta reduction
r1 ( $\lambda$ X.M)N => M[N/X] .
endm

```

In order to introduce quantifiers, we can develop a similar approach, by first introducing substitution in the syntax together with the quantifiers, and then adding rewrite rules for the new connectives. In the same way that we had to duplicate the logical connectives in both theories `PROPO[X]` and `LINLOG[X]` in Section 4.3 in order to have a correct treatment of negation, we also have to duplicate the operations and equations for substitution in the two modules `FO-PROPO[X]` and `FO-LINLOG[X]` below. This technicality, due to the treatment of negation, makes the exposition somewhat longer, but should not obscure the main ideas about the treatment of quantification that have been illustrated more concisely before with the lambda calculus example.

We assume an abstract data type ADT defining constants, functions and predicates over a set `Var` of variable names. Substitution must also be defined in this module. For example, we can have something like the following module:

```

fmod ADT[X :: VAR] is
  including SET[X] .
  sort Term .                *** terms
  subsort Var < Term .       *** variables are terms
  op c : -> Term .           *** constant symbol
  op f : Term Term -> Term . *** function symbol
  sort Atom .                *** atomic formulas
  op p : Term Term -> Atom . *** predicate symbol

  op va : Term -> Set .      *** set of variables
  op va : Atom -> Set .      *** set of variables
  op _[_/_] : Term Term Var -> Term . *** substitution
  op _[_/_] : Atom Term Var -> Atom . *** substitution

  vars X Y : Var .          vars T U V : Term .
  var P : Atom .

  *** Set of variables
  eq va(X) = {X} .
  eq va(c) = emptyset .
  eq va(f(T,V)) = va(T) U va(V) .
  eq va(p(T,V)) = va(T) U va(V) .
  eq va(V[T/X]) = (va(V) - {X}) U va(T) .
  eq va(P[T/X]) = (va(P) - {X}) U va(T) .

  *** Substitution equations
  eq X[T/X] = T .
  ceq Y[T/X] = Y if not(X == Y) .
  eq c[T/X] = c .
  eq f(U,V)[T/X] = f(U[T/X],V[T/X]) .
  eq p(U,V)[T/X] = p(U[T/X],V[T/X]) .
endfm

--- linear logic syntax with quantifiers
fmod FO-PROPO[X :: VAR] is
  including PROPO[ADT[X]] .
  op _[_/_] : Prop0 Term Var -> Prop0 . *** substitution
  op fv : Prop0 -> Set .                 *** free variables
  op ∀_.. : Var Prop0 -> Prop0 .        *** universal quantifier
  op ∃_.. : Var Prop0 -> Prop0 .        *** existential quantifier

  vars A B : Prop0 .          vars X Y : Var .
  var P : Atom .              var T : Term .

  *** Negation and quantifiers
  eq (∀X.A)⊥ = ∃X.A⊥ .
  eq (∃X.A)⊥ = ∀X.A⊥ .

```

```

*** Free variables
eq fv(P) = va(P) .
eq fv(1) = emptyset .
eq ... *** similar equations for the other logical constants
eq fv(A⊥) = fv(A) .
eq fv(A ⊗ B) = fv(A) U fv(B) .
eq ... *** similar equations for the other logical connectives
eq fv(∀X.A) = fv(A) - {X} .
eq fv(∃X.A) = fv(A) - {X} .
eq fv(A[T/X]) = (fv(A) - {X}) U va(T) .

*** Substitution equations
eq 1[T/X] = 1 .
eq ... *** similar equations for the other logical constants
eq A⊥[T/X] = A[T/X]⊥ .
eq (A ⊗ B)[T/X] = A[T/X] ⊗ B[T/X] .
eq ... *** similar equations for the other logical connectives
eq (∀X.A)[T/X] = ∀X.A .
ceq (∀Y.A)[T/X] = ∀Y.(A[T/X])
    if not(X == Y) and (not(Y is-in va(T)) or not(X is-in fv(A))) .
ceq (∀Y.A)[T/X] =
    ∀(new(va(T) U fv(A)).((A[new(va(T) U fv(A))/Y])[T/X]))
    if not(X == Y) and Y is-in fv(T) and X is-in fv(A) .
eq ... *** similar equations for the existential quantifier
endfm

mod FO-LINLOG[X :: VAR] is
  including LINLOG[ADT[X]] . ***
  protecting FO-PROPO[X] . *** Note PROPO[ADT[X]] is shared

op _[_/_] : Prop Term Var -> Prop0 . *** substitution
op fv : Prop -> Set . *** free variables
op ∀_.. : Var Prop -> Prop . *** universal quantifier
op ∃_.. : Var Prop -> Prop . *** existential quantifier

var P Q : Prop . var A : Prop0 .
var X : Var . var T : Term .

ax [∀X.A] = ∀X.[A] .
ax [∃X.A] = ∃X.[A] .

*** Free variables
ax fv(1) = emptyset .
ax ... *** similar axioms for the other logical constants
ax fv(P ⊗ Q) = fv(P) U fv(Q) .
ax ... *** similar axioms for the other logical connectives

```

```

ax fv( $\forall X.P$ ) = fv( $P$ ) -  $\{X\}$  .
ax fv( $\exists X.P$ ) = fv( $P$ ) -  $\{X\}$  .
ax fv( $P[T/X]$ ) = (fv( $P$ ) -  $\{X\}$ )  $\cup$  va( $T$ ) .
ax fv( $[A]$ ) = fv( $A$ ) .

*** Substitution axioms
ax 1 $[T/X]$  = 1 .
ax ... *** similar axioms for the other logical constants
ax ( $P \otimes Q$ ) $[T/X]$  =  $P[T/X] \otimes Q[T/X]$  .
ax ... *** similar axioms for the other logical connectives
ax ( $\forall X.P$ ) $[T/X]$  =  $\forall X.P$  .
cax ( $\forall Y.P$ ) $[T/X]$  =  $\forall Y.(P[T/X])$ 
    if not( $X == Y$ ) and (not( $Y$  is-in va( $T$ )) or not( $X$  is-in fv( $P$ ))) .
cax ( $\forall Y.P$ ) $[Q/X]$  =
     $\forall(\text{new}(\text{va}(\text{T}) \cup \text{fv}(\text{P}))).((\text{P}[\text{new}(\text{va}(\text{T}) \cup \text{fv}(\text{P})/Y)])[\text{T}/X])$ 
    if not( $X == Y$ ) and  $Y$  is-in va( $T$ ) and  $X$  is-in fv( $P$ ) .
ax ... *** similar axioms for the existential quantifier
ax  $[A][T/X]$  =  $[A[T/X]]$  .

*** Rules for quantifiers
rl  $\forall X.P \Rightarrow P[T/X]$  .
rl  $P[T/X] \Rightarrow \exists X.P$  .
crl  $P \Rightarrow \forall X.A \wp Q$ 
    if  $P \Rightarrow A \wp Q$  and not( $X$  is-in fv( $P \otimes Q$ )) .
crl  $P \otimes \exists X.A \Rightarrow Q$ 
    if  $P \otimes A \Rightarrow Q$  and not( $X$  is-in fv( $P \otimes Q$ )) .
endm

```

In this way, we have defined a map of entailment systems

$$\text{ent}(\text{FOLinLogic}) \longrightarrow \text{ent}(\text{OSRWLogic})$$

which is also *conservative*.

#### 4.5 Mapping sequent systems

In Section 4.3, we have mapped linear logic formulas to terms, and linear logic sequents to rewrite rules in rewriting logic. There is another map of entailment systems between linear logic and rewriting logic in which linear sequents become also terms, and rewrite rules correspond to rules in a Gentzen sequent calculus for linear logic. In order to reduce the number of rules of this calculus, we consider one-sided linear sequents in this section, but a completely similar treatment can be given for two-sided sequents. Thus, a linear logic sequent will be a turnstile symbol “ $\vdash$ ” followed by a multiset  $M$  of linear logic formulas, that in our translation to rewriting logic will be represented by the term  $\vdash M$ . Using the duality of linear logic

negation, a two-sided sequent  $A_1, \dots, A_n \vdash B_1, \dots, B_m$  can in this notation be expressed as the one-sided sequent  $\vdash A_1^\perp, \dots, A_n^\perp, B_1, \dots, B_m$ .

First, we define a parameterized module for multisets. The elements in the parameter are considered singleton multisets via a subsort declaration `Elem < Mset`, and there is a multiset union operator `_,_` which is associative, commutative, and has the empty multiset `null` as neutral element. Note that what makes the elements of `Mset` multisets instead of lists is the attribute `comm` of commutativity of the union operator `_,_`.

```
fth ELEM is
  sort Elem .
endft

fmod MSET[X :: ELEM] is
  sort Mset .
  subsort Elem < Mset .
  op null : -> Mset .
  op _,_ : Mset Mset -> Mset [assoc comm id: null] .
endfm
```

Now we can use this parameterized module to define the main module for sequents<sup>12</sup> and give the corresponding rules. A sequent calculus rule of the form

$$\frac{\vdash M_1, \dots, \vdash M_n}{\vdash M}$$

becomes the rewrite rule

```
r1 \vdash M1 ... \vdash Mn => \vdash M .
```

on the sort `Configuration`. Recalling that “`---`” introduces a comment, this rule can be written as

```
r1  \vdash M1 ... \vdash Mn
    => -----
        \vdash M .
```

This displaying trick that makes it possible to write a sequent calculus rule in a similar way to the usual presentation in logical textbooks is due to K. Futatsugi.

---

<sup>12</sup>The multiset structure is one particular way of building in certain *structural rules*, in this case *exchange*. Many other such data structuring mechanisms are as well possible to build in, or to drop, desired structural properties. Appropriate parameterized data types can similarly be used for this purpose. For example, we use later a data type of lists to define 2-sequents in which exchange is not assumed.

```

--- one-sided sequent calculus for linear logic
mod LL-SEQUENT[X :: VAR] is
  protecting FO-PROPO[X] .
  including MSET[FO-PROPO[X]] .

--- a configuration is a multiset of sequents
sort Configuration .
op ⊢_ : Mset -> Configuration .
op empty : -> Configuration .
op -- : Configuration Configuration -> Configuration
                                             [assoc comm id: empty] .

op ?_ : Mset -> Mset .
vars M N : Mset .
ax ?null = null .
ax ?(M,N) = (?M,?N) .
op fv : Mset -> Set .
ax fv(null) = emptyset .
ax fv(M,N) = fv(M) U fv(N) .

var P : Atom .      vars A B : Prop0 .
var T : Term .      var X : Var .

*** Identity
rl      empty
=> -----
    ⊢ P, P⊥ .

*** Cut
rl      (⊢ M,A) (⊢ N,A⊥)
=> -----
    ⊢ M,N .

*** Tensor
rl      (⊢ M,A) (⊢ B,N)
=> -----
    ⊢ M,A ⊗ B,N .

*** Par
rl      ⊢ M,A,B
=> -----
    ⊢ M,A ⋈ B .

*** Plus
rl      ⊢ M,A
=> -----
    ⊢ M,A ⊕ B .

```

```

*** With
rl      (⊢ M,A) (⊢ M,B)
=> -----
      ⊢ M,A & B .

*** Weakening
rl      ⊢ M
=> -----
      ⊢ M,?A .

*** Contraction
rl      ⊢ M,?A,?A
=> -----
      ⊢ M,?A .

*** Dereliction
rl      ⊢ M,A
=> -----
      ⊢ M,?A .

*** Storage
rl      ⊢ ?M,A
=> -----
      ⊢ ?M,!A .

*** Bottom
rl      ⊢ M
=> -----
      ⊢ M,⊥ .

*** One
rl      empty
=> -----
      ⊢ 1 .

*** Top
rl      empty
=> -----
      ⊢ M,⊤ .

*** Universal
crl     ⊢ M,A
=> -----
      ⊢ M,∀X.A
if not(X is-in fv(M)) .

```



```

*** Existential
r1  ⊢ M,A[T/X]
    => -----
        ⊢ M.∃X.A .
endm

```

Note that in the module `FO-PROPO[X]` (via the reused theory `PROPO[X]`) we have imposed associativity and commutativity attributes for some connectives, making syntax a bit more abstract than usual. However, in this case, this has no significance at all, except for the convenient fact that we only need a rule for  $\oplus$  instead of two; of course, these attributes can be removed if a less abstract presentation is preferred.

Given a linear theory  $T = (C, S)$  (where we can assume that all the sequents in  $S$  are of the form  $\vdash A_1, \dots, A_n$ ), we instantiate the parameterized module `LL-SEQUENT[X]` using a functional module `C` that interprets the propositional constants in  $C$ , as in Section 4.3, and then extend it by adding a rule

```
r1 empty => ⊢ A1, ..., An .
```

for each sequent  $\vdash A_1, \dots, A_n$  in  $S$ , obtaining in this way a rewrite theory `LL-SEQUENT(T)`.

With this map we have also an immediate conservativity result:

**THEOREM 15.** Given a linear theory  $T$ , a linear logic sequent  $\vdash A_1, \dots, A_n$  is provable in linear logic from the axioms in  $T$  if and only if the sequent

```
empty → ⊢ A1, ..., An
```

is provable in rewriting logic from the rewrite theory `LL-SEQUENT(T)`.

It is very important to realize that the technique used in this conservative map of entailment systems is very general and it is in no way restricted to linear logic. Indeed, it can be applied to any sequent calculus, be it for intuitionistic, classical or any other logic. In general, we need an operation

```
op _|_ : FormList FormList -> Sequent .
```

that turns two lists of formulas (multisets, or sets in some cases) into a term representing a sequent. Then we have a sort `Configuration` representing multisets of sequents, with a union operator written using empty syntax. A sequent calculus rule

$$\frac{G_1 \vdash D_1, \dots, G_n \vdash D_n}{G \vdash D}$$

becomes a rewrite rule

```
r1 (G1 ⊢ D1) ... (Gn ⊢ Dn) => (G ⊢ D) .
```

on the sort `Configuration`, that we have displayed above also as

$$\text{r1} \quad \frac{(G_1 \vdash D_1) \dots (G_n \vdash D_n)}{(G \vdash D) .}$$

in order to make even clearer that the rewrite rule and the sequent notations in fact capture the same idea. In the particular case of linear logic the situation is somewhat simplified by the use of one-sided sequents. Notice also that sometimes the rewrite rule can be conditional to the satisfaction of some auxiliary side conditions like, for example, in the rule for the universal quantifier in the module above.

As another example illustrating the generality of this approach, we sketch a presentation in rewriting logic of the *2-sequent calculus* defined by Masini and Martini in order to develop a proof theory for modal logics [Masini, 1993; Martini and Masini, 1993]. In their approach, a *2-sequent* is an expression of the form  $\Gamma \vdash \Delta$ , where  $\Gamma$  and  $\Delta$  are not lists of formulas as usual, but they are lists of lists of formulas, so that sequents are endowed with a vertical structure. For example,

$$\frac{A, B \quad D}{C \vdash \frac{E, F}{G}}$$

is a 2-sequent, which will be represented in rewriting logic as

$$A, B; C \vdash D; E, F; G.$$

In order to define 2-sequents, we first need a parameterized module for lists, assuming a module `NAT` defining a sort `Nat` of natural numbers with zero `0`, a successor function `s_`, an addition operation `_+_`, and an order relation `_<=_`, as well as a module `BOOL` defining a sort `Bool` of truth values `true`, `false` and corresponding Boolean operations.

```
fmod LIST[X :: ELEM] is
  protecting NAT BOOL .
  sort List .
  subsort Elem < List .
  op nil : -> List .
  op _;_ : List List -> List [assoc id: nil] .
  op length : List -> Nat .
  op _in_ : Elem List -> Bool .

  vars E E' : Elem .
  vars L L' : List .
  eq length(nil) = 0 .
  eq length(E) = s0 .
```

```

eq length(L;L') = length(L) + length(L') .
eq E in nil = false .
eq E in E' = if E == E' then true else false .
eq E in (L;L') = (E in L) or (E in L') .
endfm

```

This module is instantiated twice in order to get the module of 2-sequents, using a sort of formulas `Form` whose definition is not presented here, and that should have an operation

```
op []_ : Form -> Form .
```

corresponding to the modality  $\Box$ .

```

make 2-LIST is
  LIST[LIST[Form]*(op _;_ to _,_)]*(sort List to 2-List,
                                   op length to depth)
endmk

```

Note that in the 2-LIST module the concatenation operation `_;` is renamed to `_,_` in the case of lists of formulas, whereas in the case of lists of lists of formulas, called 2-lists, the notation `_;` is kept. Also, to emphasize the vertical structure of 2-sequents, the operation `length` for 2-lists is renamed to `depth`.

Now we can define 2-sequents as follows:

```

fmod 2-SEQUENT is
  protecting 2-LIST .
  sort 2-Sequent .
  op _\_ : 2-List 2-List -> 2-Sequent .
endfm

```

The basic rules for the modality  $\Box$  are

$$\frac{\begin{array}{c} \Gamma \\ \alpha \\ \beta, A \\ \Gamma' \end{array} \vdash \Delta}{\begin{array}{c} \Gamma \\ \alpha, \Box A \\ \beta \\ \Gamma' \end{array} \vdash \Delta} (\Box-L) \qquad \frac{\begin{array}{c} \Delta \\ \Gamma \vdash \alpha \\ A \end{array}}{\Gamma \vdash \begin{array}{c} \Delta \\ \alpha, \Box A \end{array}} (\Box-R)$$

where  $\Gamma, \Gamma', \Delta$  denote 2-lists,  $\alpha, \beta$  denote lists of formulas, and the rule  $\Box-R$  has the side condition that  $\text{depth}(\Gamma) \leq \text{depth}(\Delta) + 1$ , i.e., the formula  $A$  is the only formula in the last level of the 2-sequent.

These rules are represented in rewriting logic as follows.

```

mod 2-SEQUENT-RULES is
  protecting 2-SEQUENT .
  sort Configuration .
  subsort 2-Sequent < Configuration .
  op empty : -> Configuration .
  op _ : Configuration Configuration -> Configuration
                                     [assoc comm id: empty] .

  vars R R' S : 2-List .
  vars L L' : List .
  var A : Form .
  rl   R ; L ; L',A ; R' ⊢ S
      => -----
          R ; L,[]A ; L' ; R' ⊢ S .

  crl  R ⊢ S ; L ; A
      => -----
          R ⊢ S ; L,[]A
      if depth(R) <= s(depth(S)) .
endm

```

The dual rules for the modality  $\diamond$  are treated similarly.

This general method of viewing sequents as rewrite rules can even be applied to systems more general than traditional sequent calculi. Thus, besides the possibilities of being one-sided or two-sided, one-dimensional or two-dimensional, etc., a “sequent” can for example be a sequent presentation of natural deduction, a term assignment system, or even any predicate defined by structural induction in some way such that the proof is a kind of tree, as for example the operational semantics of CCS given later in Section 5.3 and any other use of the so-called structural operational semantics (see [Hennessy, 1990] and Section 5.4 later), including type-checking systems. The general idea is to map a rule in the “sequent” system to a rewrite rule over a “configuration” of sequents or predicates, in such a way that the rewriting relation corresponds to provability of such a predicate.

#### 4.6 Reflection in rewriting logic

Clavel and Meseguer have shown in [1996; 1996a] that rewriting logic is reflective in the sense of Section 2.8. That is, there is a rewrite theory  $\mathcal{U}$  with a finite number of operations and rules that can simulate any other finitely presentable rewrite theory  $\mathcal{R}$  in the following sense: given any two terms  $t, t'$  in  $\mathcal{R}$ , there are corresponding terms  $\langle \overline{\mathcal{R}}, \overline{t} \rangle$  and  $\langle \overline{\mathcal{R}}, \overline{t'} \rangle$  in  $\mathcal{U}$  such that we have

$$\mathcal{R} \vdash t \longrightarrow t' \iff \mathcal{U} \vdash \langle \overline{\mathcal{R}}, \overline{t} \rangle \longrightarrow \langle \overline{\mathcal{R}}, \overline{t'} \rangle.$$

Moreover, it is often possible to reify inside rewriting logic itself a representation map  $\mathcal{L} \rightarrow \text{OSRWLogic}$  for the finitely presentable theories of  $\mathcal{L}$ .

Such a reification takes the form of a map between the abstract data types representing the finitary theories of  $\mathcal{L}$  and of *OSRWLogic*. In this section we illustrate this powerful idea with the linear logic mapping defined in Section 4.3.

We have defined a linear theory  $T$  as a finite set  $C$  of propositional constants together with a finite set  $S$  of sequents of the form  $A_1, \dots, A_n \vdash B_1, \dots, B_m$ , where each  $A_i$  and  $B_j$  is a linear logic formula built from the constants in  $C$ . Note that with this definition, all linear theories are finitely presentable. First, we define an abstract data type LL-ADT to represent linear theories. A linear theory is represented as a term  $\langle C \mid G \rangle$ , where  $C$  is a list of propositional constants (that is, identifiers), and  $G$  is a list of sequents written in the usual way. Moreover, all the propositional constants in  $G$  must be included in  $C$ . To enforce this condition, we use a sort constraint [Meseguer and Goguen, 1993], which is introduced with the keyword `set` and defines a subsort `LLTheory?` of a sort `LLTheory?` by means of the given condition. In the functional module below, we do not give the equations defining the auxiliary functions `const` that extracts the constants of a list of sequents, and the list containment predicate `_=<_`. These functions are needed to write down the sort constraint for theories.

```
fmod LL-ADT is
  protecting QID .
  sorts Ids Formula Formulas Sequent .
  sorts Sequents LLTheory? LLTheory .

  subsort Id < Formula .
  ops 1 0  $\perp$   $\top$  : -> Formula .
  op  $\otimes$  : Formula Formula -> Formula .
  op  $\wp$  : Formula Formula -> Formula .
  op  $\oplus$  : Formula Formula -> Formula .
  op  $\&$  : Formula Formula -> Formula .
  op !_ : Formula -> Formula .
  op ?_ : Formula -> Formula .
  op  $\perp$  : Formula -> Formula .

  subsort Formula < Formulas .
  op null : -> Formulas .
  op  $\_,\_$  : Formulas Formulas -> Formulas [assoc comm id: null] .
  op ( $\_ \vdash \_$ ) : Formulas Formulas -> Sequent .

  subsort Id < Ids .
  op nil : -> Ids .
  op  $\_,\_$  : Ids Ids -> Ids [assoc id: nil] .

  subsort Sequent < Sequents .
  op nil : -> Sequents .
```

```

op _,_ : Sequents Sequents -> Sequents [assoc id: nil] .
op <_|_> : Ids Sequents -> LLTheory? .

var C : Ids .
var G : Sequents .
sct <C | G> : LLTheory if const(G) =< C .
eq ...
*** several equations defining the auxiliary operations
*** "const" and "_=<" used in the sort constraint condition
eq ...
endfm

```

An order-sorted rewrite theory has much more structure, and therefore the corresponding RWL-ADT is more complex, but the basic ideas are completely similar as we sketch here. First we have an order-sorted signature, declaring sorts, subsorts, constants, operations, and variables. Then, in addition, we have equations and rules. Thus, a finitely presentable rewrite theory is represented as a term  $\langle S \mid E \mid R \rangle$ , where  $S$  is a term representing a signature,  $E$  is a list of equations, and  $R$  is a list of rules. In turn, the term  $S$  has the form  $\langle T ; B ; C ; O ; V \rangle$  where each subterm corresponds to a component of a signature as mentioned before. In addition, several sort constraints are necessary to ensure for example that the variables used in equations and rules are included in the list of variables. Just to give the flavor of the construction, here is a small fragment of the module RWL-ADT, where we have omitted most of the list constructors, operations to handle conditional equations and rules, and sort constraints.

```

sorts Sort Subsort Constant Op Var .
sorts Term Equation Rule Signature RWLTheory .

op sort{ _ } : Id -> Sort .
subsort Sort < Sorts .
op nil : -> Sorts .
op __ : Sorts Sorts -> Sorts [assoc id: nil] .
op ( _<_ ) : Id Id -> Subsort .
subsort Subsort < Subsorts .

op ( cons{ _ } : sort{ _ } ) : Id Id -> Constant .
subsort Constant < Constants .
op nil : -> Constants .
op _,_ : Constants Constants -> Sorts [assoc id: nil] .

op ( op{ _ } : _ -> sort{ _ } ) : Id Sorts Id -> Op .
subsort Op < Ops .
op ( var{ _ } : sort{ _ } ) : Id Id -> Var .
subsort Var < Vars .

```

```

op <_:_;_:_;_:_> : Sorts Subsorts Constants Ops Vars -> Signature .
subsort Var < Term .
subsort Constant < Term .
subsort Term < Terms .
op nil : -> Terms .
op op{_[_]}[_] : Id Terms -> Term .
op _,_ : Terms Terms -> Terms [assoc id: nil] .

op (=__) : Term Term -> Equation .
subsort Equation < Equations .
op (=>_) : Term Term -> Rule .
subsort Rule < Rules .
op <_|_|_> : Signature Equations Rules -> RWLTheory .
    
```

Having defined the abstract data types to represent both linear and rewrite theories, we define a function  $\overline{\Phi}$  mapping a term in `LLTheory` representing a linear theory  $T$  to a term in `RWLTheory` representing the corresponding rewrite theory `LINLOG(T)` as defined in Section 4.3. First note that the rewrite theory `LINLOG` presented in Section 4.3 gives rise to a term in `RWLTheory` that we denote

$$\langle\langle T_{LL} ; B_{LL} ; C_{LL} ; O_{LL} ; V_{LL} \rangle \mid E_{LL} \mid R_{LL} \rangle.$$

The representation  $\langle C \mid F_1 \vdash G_1, \dots, F_n \vdash G_n \rangle$  of a linear logic theory is then mapped by  $\overline{\Phi}$  to the following term

$$\langle\langle T_{LL} ; B_{LL} ; \text{cons}(C), C_{LL} ; O_{LL} ; V_{LL} \rangle \mid E_{LL} \mid R_{LL}, ([\text{tensor}(F_1)] \Rightarrow [\text{par}(G_1)]), \dots, ([\text{tensor}(F_n)] \Rightarrow [\text{par}(G_n)]) \rangle$$

where the auxiliary operations `cons`, `tensor` and `par` are defined as follows, and correspond exactly to the description in Section 4.3.

```

op tensor : Formulas -> Formula .
op par : Formulas -> Formula .
op cons : Ids -> Constants .

var F : Formula .    vars F1 F2 : Formulas .
var I : Id .         var L : Ids .

eq tensor(null) = 1 .
eq tensor(F) = F .
eq tensor(F1,F2) = tensor(F1) ⊗ tensor(F2) .
eq par(null) = ⊥ .
eq par(F) = F .
eq par(F1,F2) = par(F1) ⋈ par(F2) .
eq cons(nil) = nil .
eq cons(I,L) = (cons{I}:sort{Atom}),cons(L) .
    
```

We can summarize the reification  $\overline{\Phi} : \text{LL-ADT} \rightarrow \text{RWL-ADT}$  of the map of logics  $\Phi : \text{LinLogic} \rightarrow \text{OSRWLogic}$  we have just defined by means of the following commutative diagram:

$$\begin{array}{ccc}
 \text{LL-ADT} & \xrightarrow{\overline{\Phi}} & \text{RWL-ADT} \\
 \downarrow & & \downarrow \\
 \text{LinLogicTh} & \xrightarrow{\Phi} & \text{OSRWLogicTh}
 \end{array}$$

This method is completely general, in that it should apply to any effectively presented map of logics  $\Psi : \mathcal{L} \rightarrow \text{RWLogic}$  that maps finitely presentable theories in  $\mathcal{L}$  to finitely presentable theories in rewriting logic. Indeed, the effectiveness of  $\Psi$  should exactly mean that the corresponding  $\overline{\Psi} : \mathcal{L}\text{-ADT} \rightarrow \text{RWL-ADT}$  is a computable function and therefore, by the metatheorem of Bergstra and Tucker [1980], that it is specifiable by a finite set of Church-Rosser and terminating equations inside rewriting logic.

## 5 REWRITING LOGIC AS A SEMANTIC FRAMEWORK

After an overview of rewriting logic as a general model of computation that unifies many other existing models, the cases of concurrent object-oriented programming and of Milner's CCS are treated in greater detail. Structural operational semantics is discussed as a specification formalism similar in some ways to rewriting logic, but more limited in its expressive capabilities. Rewriting logic can also be very useful as a semantic framework for many varieties of constraint solving in logic programming and in automated deduction. Finally, the representation of action and change in rewriting logic and the consequent solution of the “frame problem” difficulties associated with standard logics are also discussed.

### 5.1 Generality of rewriting logic as a model of computation

Concurrent rewriting is a very general model of concurrency from which many other models can be obtained by specialization. Except for concurrent object-oriented programming and CCS that are further discussed in Sections 5.2 and 5.3, respectively, we refer the reader to [Meseguer, 1992; Meseguer, 1996] for a detailed discussion of the remaining models, and summarize here such specializations using Figure 1, where RWL stands for rewriting logic, the arrows indicate specializations, and the subscripts  $\emptyset$ ,  $AI$ , and  $ACI$  stand for syntactic rewriting, rewriting modulo associativity and identity, and rewriting modulo associativity, commutativity, and identity, respectively.



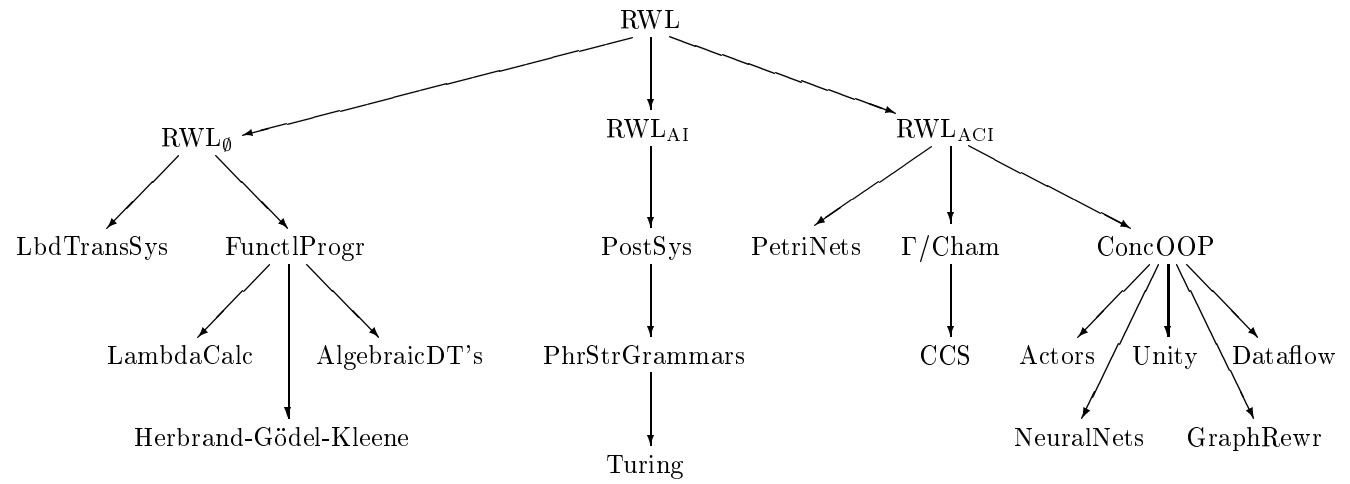


Figure 1. Unification of models of computation.

Within syntactic rewriting we have labelled transition systems, which are used in interleaving approaches to concurrency; functional programming (in particular Maude’s functional modules) corresponds to the case of *confluent*<sup>13</sup> rules, and includes the lambda calculus and the Herbrand-Gödel-Kleene theory of recursive functions. Rewriting modulo *AI* yields Post systems and related grammar formalisms, including Turing machines. Besides the general treatment by *ACI*-rewriting of concurrent object-oriented programming, briefly described in Section 5.2, that contains Actors [Agha, 1986], neural networks, graph rewriting, and the dataflow model as a special case [Meseguer, 1996], rewriting modulo *ACI* includes Petri nets [Reisig, 1995], the Gamma language of Banâtre and Le Métayer [1990], and Berry and Boudol’s *chemical abstract machine* [1992] (which itself specializes to CCS [Milner, 1989]; see [Berry and Boudol, 1992] and also the treatment in Section 5.3), as well as Unity’s model of computation [Chandy and Misra, 1988].

The *ACI* case is quite important, since it contains as special subcases a good number of concurrency models that have already been studied. In fact, the associativity and commutativity of the axioms appear in some of those models as “fundamental laws of concurrency.” However, from the perspective of this work the *ACI* case, while being important and useful, does not have a monopoly on the concurrency business. Indeed, “fundamental laws of concurrency” expressing associativity and commutativity are only valid in this particular case. They are for example meaningless for the tree-structured case of functional programming. The point is that the laws satisfied by a concurrent system cannot be determined *a priori*. They essentially depend on the actual distributed structure of the system, which is its algebraic structure.

## 5.2 Concurrent object-oriented programming

Concurrent object-oriented programming is a very active area of research. An important reason for this interest is the naturalness with which this style of programming can model concurrent interactions between objects in the real world. However, the field of concurrent object-oriented programming seems at present to lack a clear, agreed-upon semantic basis.

Rewriting logic supports a logical theory of concurrent objects that addresses these conceptual needs in a very direct way. We summarize here the key ideas regarding Maude’s object-oriented modules; a full discussion of Maude’s object-oriented aspects can be found in [Meseguer, 1993; Meseguer, 1993b].

An *object* in a given state can be represented as a term

---

<sup>13</sup>Although not reflected in the picture, rules confluent *modulo* equations *E* are also functional.

```
< 0 : C | a1 : v1, ... , an : vn >
```

where  $0$  is the object's name, belonging to a set  $OID$  of object identifiers,  $C$  is its class, the  $a_i$ 's are the names of the object's *attributes*, and the  $v_i$ 's are their corresponding values, which typically are required to be in a sort appropriate for their corresponding attribute. The *configuration* is the distributed state of the concurrent object-oriented system and is represented as a multiset of objects and messages according to the following syntax:

```
subsorts Object Message < Configuration .
op __ : Configuration Configuration -> Configuration
      [assoc comm id: null] .
```

where the operator  $__$  is associative and commutative with identity `null` and is interpreted as multiset union, and the sorts `Object` and `Message` are subsorts of `Configuration` and generate data of that sort by multiset union. The system evolves by concurrent *ACI*-rewriting of the configuration by means of rewrite rules specific to each particular system, whose lefthand and righthand sides may in general involve patterns for several objects and messages. By specializing to patterns involving only one object and one message, we can obtain an abstract, declarative, and truly concurrent version of the Actor model [Agha, 1986] (see [Meseguer, 1993, Section 4.7]).

Maude's syntax for object-oriented modules is illustrated by the object-oriented module `ACCNT` below which specifies the concurrent behavior of objects in a very simple class `Accnt` of bank accounts, each having a `bal(ance)` attribute, which may receive messages for crediting or debiting the account, or for transferring funds between two accounts. We assume an already defined functional module `INT` for integers with a subsort relation `Nat < Int` and an ordering predicate `_>=_`.

After the keyword `class`, the name of the class (`Accnt` in this case) is given, followed by a “|” and by a list of pairs of the form `a : S` separated by commas, where `a` is an attribute identifier and `S` is the sort inside which the values of such an attribute identifier must range in the given class. In this example, the only attribute of an account is its `bal(ance)`, which is declared to be a value in `Nat`. The three kinds of messages involving accounts are `credit`, `debit`, and `transfer` messages, whose user-definable syntax is introduced by the keyword `msg`. The rewrite rules specify in a declarative way the behavior associated to the `credit`, `debit`, and `transfer` messages.

```
omod ACCNT is
  protecting INT .
  class Accnt | bal : Nat .
  msgs credit debit : OId Nat -> Msg .
  msg transfer_from_to_ : Nat OId OId -> Msg .

  vars A B : OId .
```

```

vars M N N' : Nat .
rl credit(A,M) < A : Accnt | bal: N >
=> < A : Accnt | bal: N + M > .
crl debit(A,M) < A : Accnt | bal: N >
=> < A : Accnt | bal: N - M > if N >= M .
crl transfer M from A to B
< A : Accnt | bal: N > < B : Accnt | bal: N' >
=> < A : Accnt | bal: N - M > < B : Accnt | bal: N' + M >
if N >= M .
endom

```

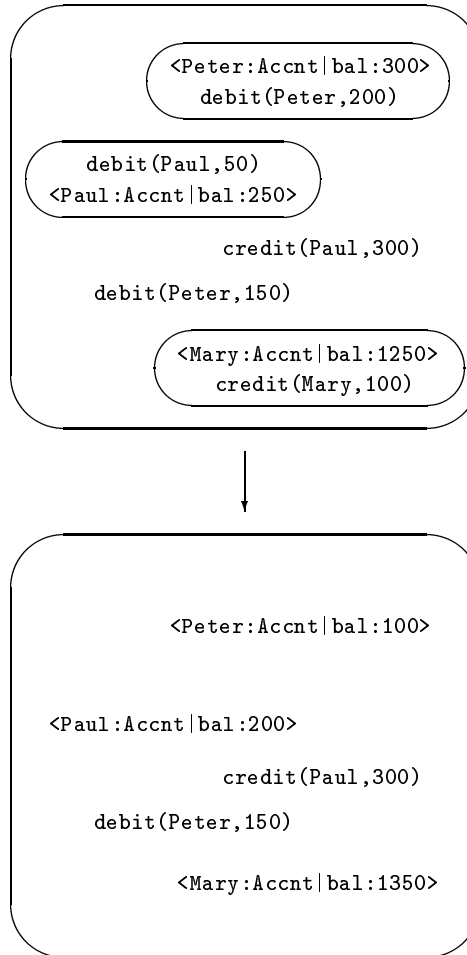


Figure 2. Concurrent rewriting of bank accounts.

The multiset structure of the configuration provides the top level distributed structure of the system and allows concurrent application of the rules. For example, Figure 2 provides a snapshot in the evolution by concurrent rewriting of a simple configuration of bank accounts. To simplify the picture, the arithmetic operations required to update balances have already been performed. However, the reader should bear in mind that the values in the attributes of an object can also be computed by means of rewrite rules, and this adds yet another important level of concurrency to a concurrent object-oriented system, which might be called *intra-object concurrency*.

Intuitively, we can think of messages as “traveling” to come into contact with the objects to which they are sent and then causing “communication events” by application of rewrite rules. In rewriting logic, this traveling is accounted for in a very abstract way by the *ACI* structural axioms. This abstract level supports both synchronous and asynchronous communication and provides great freedom and flexibility to consider a variety of alternative implementations at lower levels.

Although Maude provides convenient syntax for object-oriented modules, the syntax and semantics of such modules can be reduced to those of system modules, i.e., we can systematically translate an object-oriented module `omod  $\mathcal{O}$  endom` into a corresponding system module `mod  $\mathcal{O}\#$  endm`, where  $\mathcal{O}\#$  is a theory in rewriting logic. A detailed account of this translation process can be found in [Meseguer, 1993].

### 5.3 CCS

Milner’s *Calculus of Communicating Systems* (CCS) [Milner, 1980; Milner, 1989; Milner, 1990] is among the best well-known and studied concurrency models, and has become the paradigmatic example of an entire approach to “process algebras.” We just give a very brief introduction to CCS, referring the reader to Milner’s book [1989] for motivation and a comprehensive treatment, before giving two alternative formulations of CCS in rewriting logic and showing the conservativity of these formulations.

We assume a set  $A$  of *names*; the elements of the set  $\bar{A} = \{\bar{a} \mid a \in A\}$  are called *co-names*, and the members of the (disjoint) union  $\mathcal{L} = A \cup \bar{A}$  are *labels* naming ordinary actions. The function  $a \mapsto \bar{a}$  is extended to  $\mathcal{L}$  by defining  $\bar{\bar{a}} = a$ . There is a special action called *silent action* and denoted  $\tau$ , intended to represent internal behaviour of a system, and in particular the synchronization of two processes by means of actions  $a$  and  $\bar{a}$ . Then the set of *actions* is  $\mathcal{L} \cup \{\tau\}$ . The set of processes is intuitively defined as follows:

- 0 is an inactive process that does nothing.
- If  $\alpha$  is an action and  $P$  is a process,  $\alpha.P$  is the process that performs  $\alpha$  and subsequently behaves as  $P$ .

- If  $P$  and  $Q$  are processes,  $P + Q$  is the process that may behave as either  $P$  or  $Q$ .
- If  $P$  and  $Q$  are processes,  $P|Q$  represents  $P$  and  $Q$  running concurrently with possible communication via synchronization of the pair of ordinary actions  $a$  and  $\bar{a}$ .
- If  $P$  is a process and  $f : \mathcal{L} \rightarrow \mathcal{L}$  is a relabelling function such that  $f(\bar{a}) = \overline{f(a)}$ ,  $P[f]$  is the process that behaves as  $P$  but with the actions relabelled according to  $f$ , assuming  $f(\tau) = \tau$ .
- If  $P$  is a process and  $L \subseteq \mathcal{L}$  is a set of ordinary actions,  $P \setminus L$  is the process that behaves as  $P$  but with the actions in  $L \cup \bar{L}$  prohibited.
- If  $P$  is a process,  $I$  is a process identifier, and  $I =_{def} P$  is a defining equation where  $P$  may recursively involve  $I$ , then  $I$  is a process that behaves as  $P$ .

This intuitive explanation can be made precise in terms of the following structural operational semantics that defines a labelled transition system for CCS processes.

*Action:*

$$\frac{}{\alpha.P \xrightarrow{\alpha} P}$$

*Summation:*

$$\frac{P \xrightarrow{\alpha} P'}{P + Q \xrightarrow{\alpha} P'} \qquad \frac{Q \xrightarrow{\alpha} Q'}{P + Q \xrightarrow{\alpha} Q'}$$

*Composition:*

$$\frac{P \xrightarrow{\alpha} P'}{P|Q \xrightarrow{\alpha} P'|Q} \qquad \frac{Q \xrightarrow{\alpha} Q'}{P|Q \xrightarrow{\alpha} P|Q'}$$

$$\frac{P \xrightarrow{a} P' \quad Q \xrightarrow{\bar{a}} Q'}{P|Q \xrightarrow{\tau} P'|Q'}$$

*Relabelling:*

$$\frac{P \xrightarrow{\alpha} P'}{P[f] \xrightarrow{f(\alpha)} P'[f]}$$

*Restriction:*

$$\frac{P \xrightarrow{\alpha} P'}{P \setminus L \xrightarrow{\alpha} P' \setminus L} \quad \alpha \notin L \cup \bar{L}$$

*Definition:*

$$\frac{P \xrightarrow{\alpha} P'}{I \xrightarrow{\alpha} P'} \quad I =_{def} P$$

We now show how CCS can be described and given semantics in rewriting logic. The following modules have been motivated by, but are considerably different from, the corresponding examples in [Meseguer *et al.*, 1992].

```
fth LABEL is
  sort Label .      *** ordinary actions
  op ~_ : Label -> Label .
  var N : Label .
  eq ~N = N .
endft

--- an action is the silent action or a label
fmod ACTION[X :: LABEL] is
  sort Act .
  subsort Label < Act .
  op tau : -> Act .      *** silent action
endfm

fth PROCESSID is
  sort ProcessId .   *** process identifiers
endft

--- CCS syntax
fmod PROCESS[X :: LABEL, Y :: PROCESSID] is
  protecting ACTION[X] .
  sort Process .
  subsort ProcessId < Process .
  op 0 : -> Process .      *** inaction
  op _._ : Act Process -> Process .      *** prefix
  op _+_ : Process Process -> Process [assoc comm idem id: 0] .
      *** summation
  op _|_ : Process Process -> Process [assoc comm id: 0] .
      *** composition
  op _[_/_] : Process Label Label -> Process .
      *** relabelling: [b/a] relabels "a" to "b"
  op _\_ : Process Label -> Process .      *** restriction
endfm
```

Before defining the operational semantics of CCS processes, we need an auxiliary module in order to build contexts in which process identifiers can be associated with processes, providing in this way recursive definitions of processes. A sort constraint [Meseguer and Goguen, 1993], which is introduced with the keyword `sct` and defines a subsort `Context` by means

of a condition, is used to enforce the requirement that the same process identifier cannot be associated with two different processes in a context.

```

--- defining equations and contexts
fmod CCS-CONTEXT[X :: LABEL, Y :: PROCESSID] is
  protecting PROCESS[X,Y] .
  sorts Def Context Context? .
  subsorts Def < Context < Context? .
  op (_ =def _) : ProcessId Process -> Def .
  protecting LIST[ProcessId]*(op _;_ to __) .
  protecting LIST[Def]*(sort List to Context?) .
  op nil : -> Context .
  op pid : Context? -> List .

  var X : ProcessId .      var P : Process .
  var C : Context .        vars D D' : Context? .

  eq pid(nil) = nil .
  eq pid((X =def P)) = X .
  eq pid(D;D') = pid(D) pid(D') .
  sct (X =def P);C : Context if not(X in pid(C)) .
endfm

```

The semantics of CCS processes is usually defined relative to a given context that provides defining equations for all the necessary process identifiers [Milner, 1989, Section 2.4]. The previous module defines the data type of all contexts. We now need to parameterize the module defining the CCS semantics by the choice of a context. This is accomplished by means of the following theory that picks up a context in the sort `Context`.

```

fth CCS-CONTEXT*[X :: LABEL, Y :: PROCESSID] is
  protecting CCS-CONTEXT[X,Y] .
  op context : -> Context .
endft

```

As in the case of linear logic, we have two possibilities in order to write the operational semantics for CCS by means of rewrite rules. On the one hand, we can interpret a transition  $P \xrightarrow{\alpha} P'$  as a rewrite, so that the above operational semantics rules become conditional rewrite rules. On the other hand, the transition  $P \xrightarrow{\alpha} P'$  can be seen as a term, forming part of a configuration, in such a way that the semantics rules correspond to rewrite rules, as a particular case of the general mapping of sequent systems into rewriting logic that we have presented in Section 4.5.

```

--- CCS transitions
mod CCS1[X :: LABEL, Y :: PROCESSID, C :: CCS-CONTEXT*[X,Y]] is
  sort ActProcess .

```



```

subsort Process < ActProcess .
op {_}_ : Act ActProcess -> ActProcess .
*** {A}P means action A has been performed thus becoming process P

vars P P' Q Q' : Process .    vars L M : Label .
var X : ProcessId .          var A : Act .

*** Prefix
rl A . P => {A}P .

*** Summation
crl P + Q => {A}P' if P => {A}P' .

*** Composition
crl P | Q => {A}(P' | Q) if P => {A}P' .
crl P | Q => {tau}(P' | Q') if P => {L}P' and Q => {~L}Q' .

*** Restriction
crl P \ L => {A}(P' \ L)
    if P => {A}P' and not(A == L) and not(A == ~L) .

*** Relabelling
crl P [M / L] => {M}(P' [M / L]) if P => {L}P' .
crl P [M / L] => {~M}(P' [M / L]) if P => {~L}P' .
crl P [M / L] => {A}(P' [M / L])
    if P => {A}P' and not(A == L) and not(A == ~L) .

*** Definition
crl X => {A}P' if (X =def P) in context and P => {A}P' .
endm

```

In the above module, the rewrite rules have the property of being sort-increasing, i.e., in a rule  $[t] \longrightarrow [t']$  the least sort of  $[t']$  is bigger than the least sort of  $[t]$ . Thus, one rule cannot be applied unless the resulting term is well-formed. This prevents, for example, rewrites of the following form:

$$\{A\}(P \mid Q) \longrightarrow \{A\}(\{B\}P' \mid \{C\}Q')$$

because the term on the righthand side is not well formed according to the order-sorted signature of the module  $\text{CCS1}[X, Y, C[X, Y]]$ . More precisely, the *Congruence* rule of order-sorted rewriting logic, like the corresponding rule of order-sorted algebra [Goguen and Meseguer, 1992], cannot be applied unless the resulting term  $f(t_1, \dots, t_n)$  is well formed according to the given order-sorted signature. To illustrate this point further, although  $A.P \longrightarrow \{A\}P$  is a correct instance of the *Prefix* rewrite rule, we cannot use the *Congruence* rule to derive

$$(A.P) \mid Q \longrightarrow (\{A\}P) \mid Q$$

because the second term  $(\{A\}P) \mid Q$  is not well formed.

The net effect of this restriction is that an `ActProcess` term of the form  $\{A1\} \dots \{Ak\}P$  can only be rewritten into another term of the same form  $\{A1\} \dots \{Ak\}\{B\}P'$ , assuming that  $P \longrightarrow \{B\}P'$  is a  $\text{CCS1}[X, Y, C[X, Y]]$ -rewrite. As another example, a process of the form  $A.B.P$  can be rewritten first into  $\{A\}B.P$  and then into  $\{A\}\{B\}P$ , but cannot be rewritten into  $A.\{B\}P$ , because this last term is not well formed. After this discussion, it is easy to see that we have the following conservativity result.

**THEOREM 16.** Given a CCS process  $P$ , there are processes  $P_1, \dots, P_{k-1}$  such that

$$P \xrightarrow{a_1} P_1 \xrightarrow{a_2} \dots \xrightarrow{a_{k-1}} P_{k-1} \xrightarrow{a_k} P'$$

if and only if  $P$  can be rewritten into  $\{a1\} \dots \{ak\}P'$  using the rules in the module  $\text{CCS1}[X, Y, C[X, Y]]$ .

Note also that, since the operators  $\_+\_$  and  $\_|\_$  are declared commutative, one rule is enough for each one, instead of the two rules in the original presentation. On the other hand, we need three rules for relabelling, due to the representation of the relabelling function.

Let us consider now the second possibility, using the same idea described in Section 4.5 for the linear logic sequent calculus, that, as we have already mentioned there, is applicable to many more cases, with a very broad understanding of the term “sequent.”

```

--- CCS operational semantics
mod CCS2[X :: LABEL, Y :: PROCESSID, C :: CCS-CONTEXT*[X,Y]] is
  sort Configuration .

  op (_:-->_) : Act Process Process -> Configuration .
  op empty : -> Configuration .
  op -- : Configuration Configuration -> Configuration
          [assoc comm id: empty] .
  *** a configuration is a multiset of transitions

  vars P P' Q Q' : Process .    vars L M : Label .
  var X : ProcessId .          var A : Act .

  *** Prefix
  rl
    empty
  => -----
    (A : (A . P) --> P) .

  *** Summation
  rl
    (A : P --> P')
  => -----
    (A : P + Q --> P') .

```

```

*** Composition
rl      (A : P --> P')
=> -----
      (A : P | Q --> P' | Q) .

rl      (L : P --> P')(~L : Q --> Q')
=> -----
      (tau : P | Q --> P' | Q') .

*** Restriction
crl     (A : P --> P')
=> -----
      (A : P \ L --> P' \ L)
if not(A == L) and not(A == ~L) .

*** Relabelling
rl      (L : P --> P')
=> -----
      (M : P[M / L] --> P'[M / L]) .

rl      (~L : P --> P')
=> -----
      (~M : P[M / L] --> P'[M / L]) .

crl     (A : P --> P')
=> -----
      (A : P[M / L] --> P'[M / L])
if not(A == L) and not(A == ~L) .

*** Definition
crl     (A : P --> P')
=> -----
      (A : X --> P')
if (X =def P) in context .

endm

```

Except for the difference in the number of rules for some operators, as already pointed out above for the module  $\text{CCS1}[X, Y, C[X, Y]]$ , this presentation is closer to the original one, and therefore the following conservativity result is immediate.

**THEOREM 17.** For CCS processes  $P$  and  $P'$ , a transition  $P \xrightarrow{A} P'$  is possible according to the structural operational semantics of CCS if and only if

$$\text{empty} \longrightarrow (A : P \longrightarrow P')$$

is provable in rewriting logic from the rewrite theory  $\text{CCS2}[X, Y, C[X, Y]]$ .

#### 5.4 Structural operational semantics

Structural operational semantics is an approach originally introduced by Plotkin [1981] in which the operational semantics of a programming language is specified in a logical way, independent of machine architecture or implementation details, by means of rules that provide an inductive definition based on the structure of the expressions in the language. We refer the reader to Hennessy’s book [1990] for a clear introduction to this subject.

Within “structural operational semantics,” two main approaches coexist:

- *Big-step semantics* (also called *natural semantics* by Kahn [1987], Gunter [1991], and Nielson and Nielson [1992], and *evaluation semantics* by Hennessy [1990]). In this approach, the main inductive predicate describes the overall result or value of executing a computation until its termination. For this reason, it is not well suited for languages like CCS where most programs are not intended to be terminating.
- *Small-step semantics* (also called *structural operational semantics* by Plotkin [1981], and Nielson and Nielson [1992], *computation semantics* by Hennessy [1990], and *transition semantics* by Gunter [1991]). In this approach, the main inductive predicate describes in more detail the execution of individual steps in a computation, with the overall computation roughly corresponding to the transitive closure of such small steps. The structural operational semantics of CCS presented at the beginning of Section 5.3 is an example.

Both big-step and small-step approaches to structural operational semantics can be naturally expressed in rewriting logic:

- Big-step semantics can be seen as a particular case of the mapping of sequent systems described in Section 4.5, where semantics rules are mapped to rewrite rules over a “configuration” of sequents or predicates, and the rewriting relation means provability of such a predicate.
- Small-step semantics corresponds to the use of conditional rewrite rules, where a rewrite  $t \longrightarrow t'$  means a transition or computation step from a state  $t$  to a new state  $t'$  as in the explanation of rewriting logic given in Section 3.3. This is illustrated by the  $\text{CCS1}[X, Y, C[X, Y]]$  example in Section 5.3. However, as the  $\text{CCS2}[X, Y, C[X, Y]]$  example shows, the technique of sequent systems of Section 4.5 can also be used in this case.

Since the CCS example has already been discussed in detail in Section 5.3, we give here another example, describing the operational semantics of the functional language Mini-ML taken with slight modifications from Kahn’s

paper [1987]. The first thing to point out about this example is that the specification of a language's syntax is outside of the structural operational semantics formalism. By contrast, thanks to the order-sorted type structure of rewriting logic, such specification is now given by a functional module in Maude, as follows:

```
fmod NAT-TRUTH-VAL is
  sort Nat .
  op 0 : -> Nat .
  op s : Nat -> Nat .
  sort TruthVal .
  ops true false : -> TruthVal .
endfm

--- syntax: values, patterns and expressions
fmod ML-SYNTAX[X :: VAR] is
  protecting NAT-TRUTH-VAL .
  sorts Exp Value Pat NullPat Lambda .

  subsorts NullPat Var < Pat .
  op () : -> NullPat .
  op (_,_) : Pat Pat -> Pat .
  subsorts TruthVal Nat NullPat < Value .
  op (_,_) : Value Value -> Value .
  subsorts Value Var Lambda < Exp .
  op s : Exp -> Exp .
  op _+_ : Exp Exp -> Exp [comm] .
  op not : Exp -> Exp .
  op _and_ : Exp Exp -> Exp .
  op if_then_else_ : Exp Exp Exp -> Exp .
  op (_,_) : Exp Exp -> Exp .
  op __ : Exp Exp -> Exp .
  op λ_._ : Pat Exp -> Lambda .
  op let=_in_ : Pat Exp Exp -> Exp .
  op letrec=_in_ : Pat Exp Exp -> Exp .
endfm

--- environments are lists of pairs pattern-value
fmod AUX[X :: VAR] is
  protecting ML-SYNTAX[X] .
  sort Pair .
  op <_,_> : Pat Value -> Pair .
  protecting LIST[Pair]*(sort List to Env, op _;_ to __) .
  op Clos : Lambda Env -> Value .
endfm
```

The following module constitutes a direct translation of the natural semantics specification for Mini-ML given by Kahn in [1987], using the general

technique for sequent systems introduced in Section 4.5. Note that the natural semantics rules are particularly well suited for Prolog search, and indeed they are so executed in the system described in [Kahn, 1987].

```

--- natural semantics a la Kahn
mod ML-NAT-SEMANT[X :: VAR] is
  including AUX[X] .
  sort Config .
  op (_|-_-->_) : Env Exp Value -> Config .
  op empty : -> Config .
  op -- : Config Config -> Config [assoc comm id: empty] .

vars V W : Env .      vars E F G : Exp .
vars X Y : Var .      vars P Q : Pat .
vars A B C : Value .  vars N M : Nat .
var T : TruthVal .

*** Variables
rl      empty
=> -----
      ((V <X,A>) |- X --> A) .

crl      (V |- X --> A)
=> -----
      ((V <Y,B>) |- X --> A)
if not(X == Y) .

rl      (V <P,A> <Q,B> |- X --> C)
=> -----
      (V <(P,Q), (A,B)> |- X --> C) .

*** Arithmetic expressions
rl      empty
=> -----
      (V |- 0 --> 0) .

rl      (V |- E --> A)
=> -----
      (V |- s(E) --> s(A)) .

crl      (V |- E --> A)(V |- F --> B)
=> -----
      (V |- E + F --> C)
if A + B => C .

rl 0 + N => N .
rl s(N) + s(M) => s(s(N + M)) .

```

```

*** Boolean expressions
rl      empty
=> -----
    (V |- true --> true) .

rl      empty
=> -----
    (V |- false --> false) .

rl      (V |- E --> true)
=> -----
    (V |- not(E) --> false) .

rl      (V |- E --> false)
=> -----
    (V |- not(E) --> true) .

crl    (V |- E --> A)(V |- F --> B)
=> -----
    (V |- E and F --> C)
if (A and B) => C .

rl T and true => T .
rl T and false => false .

*** Conditional expressions
rl      (V |- E --> true)(V |- F --> A)
=> -----
    (V |- if E then F else G --> A) .

rl      (V |- E --> false)(V |- G --> A)
=> -----
    (V |- if E then F else G --> A) .

*** Pair expressions
rl      empty
=> -----
    (V |- () --> ()) .

rl      (V |- E --> A)(V |- F --> B)
=> -----
    (V |- (E,F) --> (A,B)) .

*** Lambda expressions
rl      empty
=> -----
    (V |-  $\lambda P.E$  --> Clos( $\lambda P.E,V$ )) .
    
```

```

r1 (V |- E --> Clos( $\lambda$ P.G,W))(V |- F --> A)(W <P,A> |- G --> B)
=> -----
(V |- E F --> B) .

*** Let and letrec expressions
r1 (V |- F --> A)(V <P,A> |- E --> B)
=> -----
(V |- let P = F in E --> B) .

r1 (V <P,A> |- F --> A)(V <P,A> |- E --> B)
=> -----
(V |- letrec P = F in E --> B) .

endm

```

The following module gives an alternative description of the semantics of the Mini-ML language in terms of the small-step approach. The rules can be directly used to perform reduction on Mini-ML expressions, and therefore constitute a very natural functional interpreter for the language.

```

--- sos semantics
mod ML-SOS-SEMANT[X :: VAR] is
  including AUX[X] .
  op [[_]]_ : Exp Env -> Value .

  vars V W : Env .    vars E F G : Exp .
  vars X Y : Var .    vars P Q : Pat .
  vars A B : Value .

  *** Variables
  r1 [[X]](V <X,A>) => A .
  cr1 [[X]](V <Y,B>) => [[X]]V if not(X == Y) .
  r1 [[X]](V <(P,Q),(A,B)>) => [[X]](V <P,A> <Q,B>) .

  *** Arithmetic expressions
  r1 0 + E => E .
  r1 s(E) + s(F) => s(s(E + F)) .
  r1 [[0]]V => 0 .
  r1 [[s(E)]]V => s([[E]]V) .
  r1 [[E + F]]V => [[E]]V + [[F]]V .

  *** Boolean expressions
  r1 not(false) => true .
  r1 not(true) => false .
  r1 E and true => E .
  r1 E and false => false .
  r1 [[true]]V => true .
  r1 [[false]]V => false .

```



```

r1 [[not(E)]V => not([[E]]V) .
r1 [[E and F]]V => [[E]]V and [[F]]V .

*** Conditional expressions
r1 if true then E else F => E .
r1 if false then E else F => F .
r1 [[if E then F else G]]V => if [[E]]V then [[F]]V else [[G]]V .

*** Pair expressions
r1 [[()]V => () .
r1 [[(E,F)]V => ([[E]]V,[[F]]V) .

*** Lambda expressions
r1 [[λP.E]]V => Clos(λP.E,V) .
r1 [[E F]]V => [[E]]V [[F]]V .
r1 Clos(λP.E,W) [[F]]V => [[E]](W <P,[[F]]V>) .

*** Let and letrec expressions
r1 [[let P = E in F]]V => [[F]](V <P,[[E]]V>) .
cr1 [[letrec P = E in F]]V => [[F]](V <P,A>)
    if [[E]]((V <P,A>) => A) .
endm

```

This concludes our discussion of structural operational semantics. Compared with rewriting logic, one of its limitations is the lack of support for structural axioms yielding more abstract data representations. Therefore, the rules must follow a purely syntactic structure, and more rules may in some cases be necessary than if an abstract representation had been chosen. In the case of multiset representations (corresponding to associativity, commutativity, and identity axioms), this has led Milner to favor multiset rewriting presentations [Milner, 1992] in the style of the chemical abstract machine of Berry and Boudol [1992] over the traditional syntactic presentation of structural operational semantics.

### 5.5 Constraint solving

Deduction can in many cases be made much more efficient by making use of *constraints* that can drastically reduce the search space, and for which special purpose constraint solving algorithms can be much faster than the alternative of expressing everything in a unique deduction mechanism such as some form of resolution.

Typically, constraints are symbolic expressions associated with a particular *theory*, and a constraint solving algorithm uses intimate knowledge about the truths of the theory in question to find solutions for those expressions by transforming them into expressions in *solved form*. One of the simplest examples is provided by standard syntactic unification—the

constraint solver for resolution in first-order logic without equality and in particular for Prolog—where the constraints in question are equalities between terms in a free algebra, i.e., in the so-called Herbrand universe. There are however many other constraints and constraint solving algorithms that can be used to advantage in order to make the representation of problems more expressive and logical deduction more efficient. For example,

- *Semantic unification* (see for example the survey by Jouannaud and Kirchner [1991]), which corresponds to solving equations in a given equational theory.
- *Sorted unification*, either many-sorted or order-sorted [Walther, 1985; Walther, 1986; Schmidt-Schauss, 1989; Meseguer *et al.*, 1989; Smolka *et al.*, 1989; Jouannaud and Kirchner, 1991], where type constraints are added to variables in equations.
- *Higher-order unification* [Huet, 1973; Miller, 1991], which corresponds to solving equations between  $\lambda$ -expressions.
- *Disunification* [Comon, 1991], which corresponds to solving not only equalities but also negated equalities.
- *Solution of equalities and inequalities in a theory*, as for example the solution of numerical constraints built into the constraint logic programming language  $CLP(\mathcal{R})$  [Jaffar and Lassez, 1987] and in other languages.

A remarkable property shared by most constraint-solving processes, and already implicit in the approach to syntactic unification problems proposed by Martelli and Montanari [1982], is that the process of solving constraints can be naturally understood as one of applying transformations to a set or multiset of constraints. Furthermore, many authors have realized that the most elegant and simple way to specify, prove correct, or even implement many constraint solving problems is by expressing those transformations as rewrite rules (see for example [Goguen and Meseguer, 1988; Jouannaud and Kirchner, 1991; Comon, 1990; Comon, 1991; Nipkow, 1993]). In particular, the survey by Jouannaud and Kirchner [1991] makes this viewpoint the cornerstone of a unified conceptual approach to unification.

For example, the so-called *decomposition* transformation present in syntactic unification and in a number of other unification algorithms can be expressed by a rewrite rule of the form

$$\begin{aligned} \text{r1 } f(t_1, \dots, t_n) &=?= f(t'_1, \dots, t'_n) \\ &=> (t_1 =?= t'_1) \dots (t_n =?= t'_n) . \end{aligned}$$

where in the righthand side multiset union has been expressed by juxtaposition.

Although the operational semantics of such rewrite rules is very obvious and intuitive, their logical or mathematical semantics has remained ambiguous. Although appeal is sometimes made to equational logic as the framework in which such rules exist, the fact that many of these rules are nondeterministic, so that, except for a few exceptions such as syntactic unification, there is in general not a unique solution but rather a, sometimes infinite, set of solutions, makes an interpretation of the rewrite rules as equations highly implausible and potentially contradictory.

We would like to suggest that rewriting logic provides a very natural framework in which to interpret rewrite rules of this nature and, more generally, deduction processes that are nondeterministic in nature and involve the exploration of an entire space of solutions. Since in rewriting logic rewrite rules go only in one direction and its models do not assume either the identification of the two sides of a rewrite step, or even the possible reversal of such a step, all the difficulties involved in an equational interpretation disappear.

Such a proposed use of rewriting logic for constraint solving and constraint programming seems very much in the spirit of recent rewrite rule approaches to constrained deduction such as those of C. Kirchner, H. Kirchner, and M. Rusinovitch [1990] (who use a general notion of constraint language proposed by Smolka [1989]), Bachmair, Ganzinger, Lynch, and Snyder [1992], Nieuwenhuis and Rubio [1992], and Giunchiglia, Pecchiari, and Talcott [1996]. In particular, the ELAN language of C. Kirchner, H. Kirchner, and M. Vittek [1995] (see also [Borovanský *et al.*, 1996]) proposes an approach to the prototyping of constraint solving languages similar in some ways to the one that would be natural using a Maude interpreter.

Exploring the use of rewriting logic as a semantic framework for languages and theorem-proving systems using constraints seems a worthwhile research direction not only for systems used in automated deduction, but also for parallel logic programming languages such as those surveyed in [Shapiro, 1989], the Andorra language [Janson and Haridi, 1991], concurrent constraint programming [Saraswat, 1992], and the Oz language [Henz *et al.*, 1995].

### 5.6 Action and change in rewriting logic

In the previous sections, we have shown the advantages of rewriting logic as a logical framework in which other logics can be represented, and as a semantic framework for the specification of languages and systems. We would like the class of systems that can be represented to be as wide as possible, and their representation to be as natural and direct as possible. In particular, an important point that has to be considered is the representation of action and change in rewriting logic. In our paper [Martí-Oliet and Meseguer, 1999], we show that rewriting logic overcomes the frame problem, and subsumes and

unifies a number of previously proposed logics of change. In this section, we illustrate this claim by means of an example, referring the reader to the cited paper for more examples and discussion.

The frame problem [McCarthy and Hayes, 1969; Hayes, 1987; Janlert, 1987] consists in formalizing the assumption that facts are preserved by an action unless the action explicitly says that a certain fact becomes true or false. In the words of Patrick Hayes [1987],

“There should be some economical and principled way of succinctly saying what changes an action makes, without having to explicitly list all the things it doesn’t change as well [...]. *That* is the frame problem.”

Recently, some new logics of action and change have been proposed, among which we can point out the approach of Hölldobler and Schneeberger [1990] (see also [Große *et al.*, 1996; Große *et al.*, 1992]), based on Horn logic with equations, and the approach of Masseron, Tollu, and Vauzeilles [1990; 1993], based on linear logic. The main interest of these formalisms is that they need not explicitly state frame axioms, because they treat facts as resources which are produced and consumed. Having proved in Sections 4.2 and 4.3, respectively, that Horn logic with equations and linear logic can be conservatively mapped into rewriting logic, it is not surprising that the advantages of the two previously mentioned approaches are also shared by rewriting logic. In particular, the rewriting logic rules automatically take care of the task of preserving context, making unnecessary the use of any frame axioms stating the properties that do not change when a rule is applied to a certain state.

We illustrate this point by means of a blocksworld example, borrowed from [Hölldobler and Schneeberger, 1990; Masseron *et al.*, 1990].

```
fth BLOCKS is
  sort BlockId .
endft

mod BLOCKWORLD[X :: BLOCKS] is
  sort Prop .
  op table : BlockId -> Prop .          *** block is on the table
  op on : BlockId BlockId -> Prop .     *** block A is on block B
  op clear : BlockId -> Prop .          *** block is clear
  op hold : BlockId -> Prop .           *** robot hand is holding block
  op empty : -> Prop .                  *** robot hand is empty

  sort State .
  subsort Prop < State .
  op 1 : -> State .
  op _⊗_ : State State -> State [assoc comm id: 1] .
```

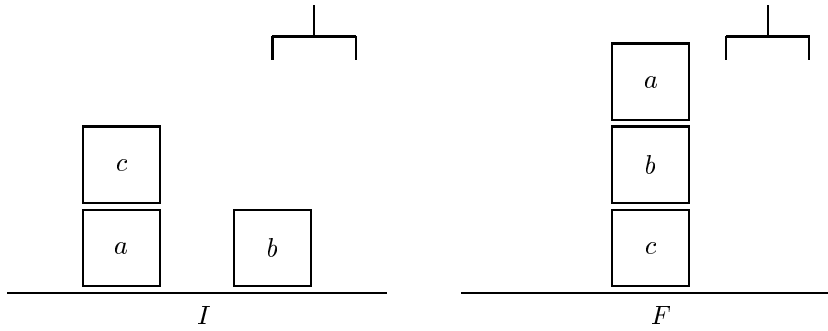


Figure 3. Two states of a blockworld.

```

vars X Y : BlockId .
rl pickup(X) : empty ⊗ clear(X) ⊗ table(X) => hold(X) .
rl putdown(X) : hold(X) => empty ⊗ clear(X) ⊗ table(X) .
rl unstack(X,Y) : empty ⊗ clear(X) ⊗ on(X,Y)
=> hold(X) ⊗ clear(Y) .
rl stack(X,Y) : hold(X) ⊗ clear(Y)
=> empty ⊗ clear(X) ⊗ on(X,Y) .
endm
    
```

In order to create a world with three blocks  $\{a,b,c\}$ , we consider the following instantiation of the previous parameterized module.

```

fmod BLOCKS3 is
  sort BlockId .
  ops a b c : -> BlockId .
endfm

make WORLD is BLOCKWORLD[BLOCKS3] endmk
    
```

Consider the states described in Figure 3; the state *I* on the left is the initial one, described by the following term of sort *State* in the rewrite theory (Maude program) *WORLD*

```
empty ⊗ clear(c) ⊗ clear(b) ⊗ table(a) ⊗ table(b) ⊗ on(c,a) .
```

Analogously, the final state *F* on the right is described by the term

```
empty ⊗ clear(a) ⊗ table(c) ⊗ on(a,b) ⊗ on(b,c) .
```

The fact that the plan

```
unstack(c,a);putdown(c);pickup(b);stack(b,c);pickup(a);stack(a,b)
```

moves the blocks from state  $I$  to state  $F$  corresponds directly to the following WORLD-rewrite (proof in rewriting logic), where we also show the use of the structural axioms of associativity and commutativity:

$$\begin{aligned}
& \text{empty} \otimes \text{clear}(c) \otimes \text{clear}(b) \otimes \text{table}(a) \otimes \text{table}(b) \otimes \text{on}(c,a) \\
& = \\
& \text{empty} \otimes \text{clear}(c) \otimes \text{on}(c,a) \otimes \text{clear}(b) \otimes \text{table}(a) \otimes \text{table}(b) \\
& \longrightarrow \text{Cong}[\text{Repl}[\text{unstack}(c,a)], \text{Refl}] \\
& \text{hold}(c) \otimes \text{clear}(a) \otimes \text{clear}(b) \otimes \text{table}(a) \otimes \text{table}(b) \\
& \longrightarrow \text{Cong}[\text{Repl}[\text{putdown}(c)], \text{Refl}] \\
& \text{empty} \otimes \text{clear}(c) \otimes \text{table}(c) \otimes \text{clear}(a) \otimes \text{clear}(b) \otimes \\
& \hspace{15em} \text{table}(a) \otimes \text{table}(b) \\
& = \\
& \text{empty} \otimes \text{clear}(b) \otimes \text{table}(b) \otimes \text{clear}(c) \otimes \text{table}(c) \otimes \\
& \hspace{15em} \text{clear}(a) \otimes \text{table}(a) \\
& \longrightarrow \text{Cong}[\text{Repl}[\text{pickup}(b)], \text{Refl}] \\
& \text{hold}(b) \otimes \text{clear}(c) \otimes \text{table}(c) \otimes \text{clear}(a) \otimes \text{table}(a) \\
& \longrightarrow \text{Cong}[\text{Repl}[\text{stack}(b,c)], \text{Refl}] \\
& \text{empty} \otimes \text{clear}(b) \otimes \text{on}(b,c) \otimes \text{table}(c) \otimes \text{clear}(a) \otimes \text{table}(a) \\
& = \\
& \text{empty} \otimes \text{clear}(a) \otimes \text{table}(a) \otimes \text{clear}(b) \otimes \text{on}(b,c) \otimes \text{table}(c) \\
& \longrightarrow \text{Cong}[\text{Repl}[\text{pickup}(a)], \text{Refl}] \\
& \text{hold}(a) \otimes \text{clear}(b) \otimes \text{on}(b,c) \otimes \text{table}(c) \\
& \longrightarrow \text{Cong}[\text{Repl}[\text{stack}(a,b)], \text{Refl}] \\
& \text{empty} \otimes \text{clear}(a) \otimes \text{on}(a,b) \otimes \text{on}(b,c) \otimes \text{table}(c) \\
& = \\
& \text{empty} \otimes \text{clear}(a) \otimes \text{table}(c) \otimes \text{on}(a,b) \otimes \text{on}(b,c)
\end{aligned}$$

Hopefully this notation is self-explanatory. For example, the expression  $\text{Cong}[\text{Repl}[\text{pickup}(b)], \text{Refl}]$  means the application of the *Congruence* rule of rewriting logic to the two WORLD-rewrites obtained by using *Replacement* with the rewrite rule  $\text{pickup}(b)$  and *Reflexivity*. The *Transitivity* rule is used several times to go from the initial state  $I$  to the final state  $F$ .

Große, Hölldobler, and Schneeberger prove in [1996] (see also [Große *et al.*, 1992; Hölldobler, 1992]) that, in the framework of conjunctive planning, there is an equivalence between plans generated by linear logic proofs as used by Masseron *et al.* [1990; 1993], and the equational Horn logic approach of Hölldobler and Schneeberger [1990]. In the light of the example above, it is not surprising that we can add to the above equivalence the plans generated by proofs in rewriting logic [Martí-Oliet and Meseguer, 1999]. Moreover, this result extends to the case of disjunctive planning [Brüning *et al.*, 1993; Martí-Oliet and Meseguer, 1999]. In our opinion, rewriting logic compares favorably with these formalisms, not only because it subsumes them, but also because it is intrinsically concurrent, and it is more flexible and general, supporting user-definable logical connectives, which can be chosen to fit the problem at hand. In the words of Reichwein, Fiadeiro, and Maibaum [1992],

“It is not enough to have a convenient formalism in which to represent action and change: the representation has to reflect the structure of the represented system.”

In this respect, we show in [Martí-Oliet and Meseguer, 1999] that the object-oriented point of view supported by rewriting logic becomes very helpful in order to represent action and change.

## 6 CONCLUDING REMARKS

Rewriting logic has been proposed as a logical framework that seems particularly promising for representing logics, and its use for this purpose has been illustrated in detail by a number of examples. The general way in which such representations are achieved is by:

- Representing formulas or, more generally, proof-theoretic structures such as sequents, as *terms* in an order-sorted equational data type whose equations express structural axioms natural to the logic in question.
- Representing the rules of deduction of a logic as rewrite rules that transform certain patterns of formulas into other patterns modulo the given structural axioms.

Besides, the theory of general logics [Meseguer, 1989] has been used as both a method and a criterion of adequacy for defining these representations as conservative maps of logics or of entailment systems. From this point of view, our tentative conclusion is that, at the level of entailment systems, rewriting logic should in fact be able to represent any finitely presented logic via a conservative map, for any reasonable notion of “finitely presented logic.” Making this tentative conclusion definite will require proposing an intuitively reasonable formal version of such a notion in a way similar to previous proposals of this kind by Smullyan [1961] and Feferman [1989].

In some cases, such as for equational logic, Horn logic with equality, and linear logic, we have in fact been able to represent logics in a much stronger sense, namely by conservative maps of logics that also map the models. Of course, such maps are much more informative, and may afford easier proofs, for example for conservativity. However, one should not expect to find representations of this kind for logics whose model theory is very different from that of rewriting logic.

Although this paper has studied the use of rewriting logic as a logical framework, and not as a metalogical one in which metalevel reasoning about an object logic is performed, this second use is not excluded and is indeed one of the most interesting research directions that we plan to study. For this purpose, as stressed by Constable [1995], we regard *reflection* as a key

technique to be employed. Some concrete evidence for the usefulness of reflection has been given in Section 4.6.

The uses of rewriting logic as a semantic framework for the specification of languages, systems, and models of computation have also been discussed and illustrated with examples. Such uses include the specification and prototyping of concurrent models of computation and concurrent object-oriented systems, of general programming languages, of automated deduction systems and logic programming languages that use constraints, and of logical representation of action and change in AI.

From a pragmatic point of view, the main goal of this study is to serve as a guide for the design and implementation of a theoretically-based high-level system in which it can be easy to define logics and to perform deductions in them, and in which a very wide variety of systems, languages, and models of computation can similarly be specified and prototyped. Having this goal in mind, the following features seem particularly useful:

- *Executability*, which is not only very useful for prototyping purposes, but is in practice a must for debugging specifications of any realistic size.
- *Abstract user-definable syntax*, which can be specified as an order-sorted equational data type with the desired structural axioms.
- *Modularity and parameterization*<sup>14</sup>, which can make specifications very readable and reusable by decomposing them in small understandable pieces that are as general as possible.
- *Simple and general logical semantics*, which can naturally express both logical deductions and concurrent computations.

These features are supported by the Maude interpreter [Clavel *et al.*, 1996]. A very important additional feature that the Maude interpreter has is good support for flexible and expressive strategies of evaluation [Clavel *et al.*, 1996; Clavel and Meseguer, 1996a], so that the user can explore the space of rewritings in intelligent ways.

#### POSTSCRIPT (2001)

During the five years that have passed since this paper was last revised until its final publication, the ideas put forward here have been greatly developed by several researchers all over the world. The survey paper [Martí-Oliet and Meseguer, 2001] provides a recent snapshot of the state of the art in the theory and applications of rewriting logic with a bibliography including

---

<sup>14</sup>Parameterization is based on the existence of relatively free algebras in rewriting logic, which generalizes the existence of initial algebras.



more than three hundred papers in this area. Here we provide some pointers to work closely related to the main points developed in this paper, and refer the reader to [Martí-Oliet and Meseguer, 2001] for many more references.

The paper [Clavel *et al.*, 2001] explains and illustrates the main concepts behind Maude's language design. The Maude system, a tutorial and a manual, a collection of examples and case studies, and a list of related papers are available at <http://maude.csl.sri.com>.

The reflective properties of rewriting logic and its applications have been developed in detail in [Clavel, 2000; Clavel and Meseguer, 2001]. A full reflective implementation developed by Clavel and Martí-Oliet of the map from linear logic to rewriting logic described in Section 4.6 appears in [Clavel, 2000]. Reflection has been used to endow Maude with a powerful module algebra of parameterized modules and module composition operations implemented in the Full Maude tool [Durán, 1999]. Moreover, reflection allows Maude to become a powerful *metatool* that has itself been used to build formal tools such as an inductive theorem prover; a tool to check the Church-Rosser property, coherence, and termination, and to perform Knuth-Bendix completion; and a tool to specify, analyze and model check real-time specifications [Clavel *et al.*, 2000; Clavel *et al.*, 1999; Olveczky, 2000].

A good number of examples of representations of logics in rewriting logic have been given by different authors, often in the form of executable specifications, including a map  $HOL \rightarrow Nuprl$  between the logics of the HOL and Nuprl theorem provers, and a natural representation map  $PTS \rightarrow RWLogic$  of pure type systems (a parametric family of higher-order logics) in rewriting logic [Stehr, 2002].

Thanks to reflection and to the existence of initial models, rewriting logic can not only be used as a logical framework in which the deduction of a logic  $\mathcal{L}$  can be faithfully simulated, but also as a *metalogical framework* in which we can reason about the metalogical properties of a logic  $\mathcal{L}$ . Basin, Clavel, and Meseguer [2000] have begun studying the use of reflection, induction, and Maude's inductive theorem prover enriched with reflective reasoning principles to prove such metalogical properties.

Similarly, the use of rewriting logic and Maude as semantic framework has been greatly advanced. A number of encouraging case studies giving rewriting logic definitions of programming languages have already been carried out by different authors. Since those specifications usually can be executed in a rewriting logic language, they in fact become *interpreters* for the languages in question. In addition, such formal specifications allow both formal reasoning and a variety of formal analyses for the languages so specified. See [Martí-Oliet and Meseguer, 2001] for a considerable number of references related to these topics.

The close connections between rewriting logic and structural operational semantics have been further developed by Mosses [1998] and Braga [2001] in

the context of Mosses’s modular structural operational semantics (MSOS) [Mosses, 1999]. In particular, Braga [2001] proves the correctness of a mapping translating MSOS specifications into rewrite theories. Based on these ideas, an interpreter for MSOS specifications [Braga, 2001] and a Maude Action Tool [Braga *et al.*, 2000; Braga, 2001] to execute Action Semantics specifications have been built using Maude.

The implementation of CCS in Maude has been refined and considerably extended to take into account the Hennessy-Milner modal logic by Verdejo and Martí-Oliet [2000]. The semantic properties of this map from CCS to rewriting logic have been studied in detail in [Carabetta *et al.*, 1998; Degano *et al.*, 2000].

### ACKNOWLEDGEMENTS

We would like to thank Manuel G. Clavel, Robert Constable, Harmut Ehrig, Fabio Gadducci, Claude Kirchner, H el ene Kirchner, Patrick Lincoln, Ugo Montanari, Natarajan Shankar, Sam Owre, Miguel Palomino, Gordon Plotkin, Axel Poign e, and Carolyn Talcott for their comments and suggestions that have helped us improve the final version of this paper. We are also grateful to the participants of the May 1993 Dagstuhl Seminar on Specification and Semantics, where this work was first presented, for their encouragement of, and constructive comments on, these ideas.

The work reported in this paper has been supported by Office of Naval Research Contracts N00014-90-C-0086, N00014-92-C-0518, N00014-95-C-0225 and N00014-96-C-0114, National Science Foundation Grant CCR-9224005, and by the Information Technology Promotion Agency, Japan, as a part of the Industrial Science and Technology Frontier Program “New Models for Software Architecture” sponsored by NEDO (New Energy and Industrial Technology Development Organization). The first author was also supported by a Postdoctoral Research Fellowship of the Spanish Ministry for Education and Science and by CICYT, TIC 95-0433-C03-01.

*SRI International, Menlo Park, CA 94025, USA and  
Center for the Study of Language and Information  
Stanford University, Stanford, CA 94305, USA*

*Current affiliations:*

*Narciso Mart ı-Oliet  
Depto. de Sistemas Inform aticos y Programaci on  
Facultad de Ciencias Matem aticas  
Universidad Complutense de Madrid  
28040 Madrid, Spain  
E-mail: narciso@sip.ucm.es*

José Meseguer  
 Department of Computer Science  
 University of Illinois at Urbana-Champaign  
 1304 W. Springfield Ave  
 Urbana IL 61801, USA  
 E-mail: meseguer@cs.uiuc.edu

## BIBLIOGRAPHY

- [Abadi *et al.*, 1991] M. Abadi, L. Cardelli, P.-L. Curien, and J.-J. Lévy. Explicit substitutions, *Journal of Functional Programming*, **1**, 375–416, 1991.
- [Agha, 1986] G. Agha. *Actors*, The MIT Press, 1986.
- [Aitken *et al.*, 1995] W. E. Aitken, R. L. Constable, and J. L. Underwood. *Metalogical frameworks II: Using reflected decision procedures*, Technical report, Computer Science Department, Cornell University, 1995. Also, lecture by R. L. Constable at the Max Planck Institut für Informatik, Saarbrücken, Germany, July 21, 1993.
- [Astesiano and Cerioli, 1993] E. Astesiano and M. Cerioli. Relationships between logical frameworks. In *Recent Trends in Data Type Specification*, M. Bidoit and C. Choppy, eds. pp. 126–143. LNCS 655, Springer-Verlag, 1993.
- [Bachmair *et al.*, 1992] L. Bachmair, H. Ganzinger, C. Lynch, and W. Snyder. Basic paramodulation and superposition. In *Proc. 11th. Int. Conf. on Automated Deduction*, Saratoga Springs, NY, June 1992, D. Kapur, ed. pp. 462–476. LNAI 607, Springer-Verlag, 1992.
- [Banâtre and Le Métayer, 1990] J.-P. Banâtre and D. Le Métayer. The Gamma model and its discipline of programming, *Science of Computer Programming*, **15**, 55–77, 1990.
- [Barr, 1979] M. Barr. *\*-Autonomous Categories*, Lecture Notes in Mathematics 752, Springer-Verlag, 1979.
- [Barr and Wells, 1985] M. Barr and C. Wells. *Toposes, Triples and Theories*, Springer-Verlag, 1985.
- [Basin *et al.*, 2000] D. Basin, M. Clavel, and J. Meseguer. Rewriting logic as a meta-logical framework. In *Foundations of Software Technology and Theoretical Computer Science*, New Delhi, India, December 2000, S. Kapoor and S. Prasad, eds. pp. 55–80. LNCS 1974, Springer-Verlag, 2000.
- [Basin and Constable, 1993] D. A. Basin and R. L. Constable. Metalogical frameworks. In *Logical Environments*, G. Huet and G. Plotkin, eds. pp. 1–29. Cambridge University Press, 1993.
- [Bergstra and Tucker, 1980] J. Bergstra and J. Tucker. Characterization of computable data types by means of a finite equational specification method. In *Proc. 7th. Int. Colloquium on Automata, Languages and Programming*, J. W. de Bakker and J. van Leeuwen, eds. pp. 76–90. LNCS 81, Springer-Verlag, 1980.
- [Berry and Boudol, 1992] G. Berry and G. Boudol. The chemical abstract machine, *Theoretical Computer Science*, **96**, 217–248, 1992.
- [Borovanský *et al.*, 1996] P. Borovanský, C. Kirchner, H. Kirchner, P.-E. Moreau, and M. Vittek. ELAN: A logical framework based on computational systems. In *Proc. First Int. Workshop on Rewriting Logic and its Applications*, Asilomar, California, September 1996, J. Meseguer, ed. pp. 35–50. ENTCS 4, Elsevier, 1996. <http://www.elsevier.nl/locate/entcs/volume4.html>
- [Bouhoula *et al.*, 2000] A. Bouhoula, J.-P. Jouannaud, and J. Meseguer. Specification and proof in membership equational logic, *Theoretical Computer Science*, **236**, 35–132, 2000.
- [Braga, 2001] C. Braga. *Rewriting Logic as a Semantic Framework for Modular Structural Operational Semantics*, Ph.D. thesis, Departamento de Informática, Pontifícia Universidade Católica do Rio de Janeiro, Brazil, 2001.

- [Braga *et al.*, 2000] C. Braga, H. Haeusler, J. Meseguer, and P. Mosses. Maude Action Tool: Using reflection to map action semantics to rewriting logic. In *Algebraic Methodology and Software Technology* Iowa City, Iowa, USA, May 2000, T. Rus, ed. pp. 407–421. LNCS 1816, Springer-Verlag, 2000.
- [Brüning *et al.*, 1993] S. Brüning, G. Große, S. Hölldobler, J. Schneeberger, U. Sigmund, and M. Thielscher. Disjunction in plan generation by equational logic programming. In *Beiträge zum 7. Workshop Planen und Konfigurieren*, A. Horz, ed. pp. 18–26. Arbeitspapiere der GMD 723, 1993.
- [Carabetta *et al.*, 1998] G. Carabetta, P. Degano, and F. Gadducci. CCS semantics via proved transition systems and rewriting logic. In *Proc. Second Int. Workshop on Rewriting Logic and its Applications*, Pont-à-Mousson, France, September 1998, C. Kirchner and H. Kirchner, eds. pp. 253–272. ENTCS 15, Elsevier, 1998. <http://www.elsevier.nl/locate/entcs/volume15.html>
- [Cerioli and Meseguer, 1996] M. Cerioli and J. Meseguer. May I borrow your logic? (Transporting logical structure along maps), *Theoretical Computer Science*, **173**, 311–347, 1997.
- [Clavel, 2000] M. Clavel. *Reflection in Rewriting Logic: Metalogical Foundations and Metaprogramming Applications*, CSLI Publications, 2000.
- [Clavel *et al.*, 2001] M. Clavel, F. Durán, S. Eker, P. Lincoln, N. Martí-Oliet, J. Meseguer, and J. F. Quesada. Maude: Specification and programming in rewriting logic, *Theoretical Computer Science*, 2001. To appear.
- [Clavel *et al.*, 2000] M. Clavel, F. Durán, S. Eker, and J. Meseguer. Building equational proving tools by reflection in rewriting logic. In *Cafe: An Industrial-Strength Algebraic Formal Method*, K. Futatsugi, A. T. Nakagawa, and T. Tamai, eds. pp. 1–31. Elsevier, 2000.
- [Clavel *et al.*, 1999] M. Clavel, F. Durán, S. Eker, J. Meseguer, and M.-O. Stehr. Maude as a formal meta-tool. In *FM'99 — Formal Methods, World Congress on Formal Methods in the Development of Computing Systems, Volume II*, Toulouse, France, September 1999, J. M. Wing, J. Woodcock, and J. Davies, eds. pp. 1684–1703. LNCS 1709, Springer-Verlag, 1999.
- [Clavel *et al.*, 1996] M. Clavel, S. Eker, P. Lincoln, and J. Meseguer. Principles of Maude. In *Proc. First Int. Workshop on Rewriting Logic and its Applications*, Asilomar, California, September 1996, J. Meseguer, ed. pp. 65–89. ENTCS 4, Elsevier, 1996. <http://www.elsevier.nl/locate/entcs/volume4.html>
- [Clavel and Meseguer, 1996] M. Clavel and J. Meseguer. Axiomatizing reflective logics and languages. In *Proc. Reflection'96*, San Francisco, USA, G. Kiczales, ed. pp. 263–288. April 1996.
- [Clavel and Meseguer, 1996a] M. Clavel and J. Meseguer. Reflection and strategies in rewriting logic. In *Proc. First Int. Workshop on Rewriting Logic and its Applications*, Asilomar, California, September 1996, J. Meseguer, ed. pp. 125–147. ENTCS 4, Elsevier, 1996. <http://www.elsevier.nl/locate/entcs/volume4.html>
- [Clavel and Meseguer, 2001] M. Clavel and J. Meseguer. Reflection in conditional rewriting logic, *Theoretical Computer Science*, 2001. To appear.
- [Chandy and Misra, 1988] K. M. Chandy and J. Misra. *Parallel Program Design: A Foundation*, Addison-Wesley, 1988.
- [Cockett and Seely, 1992] J. R. B. Cockett and R. A. G. Seely. Weakly distributive categories. In *Applications of Categories in Computer Science*, M. P. Fourman, P. T. Johnstone, and A. M. Pitts, eds. pp. 45–65. Cambridge University Press, 1992.
- [Comon, 1990] H. Comon. Equational formulas in order-sorted algebras. In *Proc. 17th. Int. Colloquium on Automata, Languages and Programming*, Warwick, England, July 1990, M. S. Paterson, ed. pp. 674–688. LNCS 443, Springer-Verlag, 1990.
- [Comon, 1991] H. Comon. Disunification: A survey. In *Computational Logic: Essays in Honor of Alan Robinson*, J.-L. Lassez and G. Plotkin, eds. pp. 322–359. The MIT Press, 1991.
- [Degano *et al.*, 2000] Pierpaolo Degano, Fabio Gadducci, and Corrado Priami. *A causal semantics for CCS via rewriting logic*. Manuscript, Dipartimento di Informatica, Università di Pisa, submitted for publication, 2000.

- [Dershowitz and Jouannaud, 1990] N. Dershowitz and J.-P. Jouannaud. Rewrite systems. In *Handbook of Theoretical Computer Science, Vol. B: Formal Models and Semantics*, J. van Leeuwen *et al.*, eds. pp. 243–320. The MIT Press/Elsevier, 1990.
- [Durán, 1999] F. Durán. *A Reflective Module Algebra with Applications to the Maude Language*, Ph.D. thesis, Universidad de Málaga, Spain, 1999.
- [Ehrig *et al.*, 1991] H. Ehrig, M. Baldamus, and F. Cornelius. Theory of algebraic module specification including behavioural semantics, constraints and aspects of generalized morphisms. In *Proc. Second Int. Conf. on Algebraic Methodology and Software Technology*, Iowa City, Iowa, pp. 101–125. 1991.
- [Feferman, 1989] S. Feferman. Finitary inductively presented logics. In *Logic Colloquium '88*, R. Ferro *et al.*, eds. pp. 191–220. North-Holland, 1989.
- [Felyt and Miller, 1990] A. Felyt and D. Miller. Encoding a dependent-type  $\lambda$ -calculus in a logic programming language. In *Proc. 10th. Int. Conf. on Automated Deduction*, Kaiserslautern, Germany, July 1990, M. E. Stickel, ed. pp. 221–235. LNAI 449, Springer-Verlag, 1990.
- [Fiadeiro and Sernadas, 1988] J. Fiadeiro and A. Sernadas. Structuring theories on consequence. In *Recent Trends in Data Type Specification*, D. Sannella and A. Tarlecki, eds. pp. 44–72. LNCS 332, Springer-Verlag, 1988.
- [Gardner, 1992] P. Gardner. *Representing Logics in Type Theory*, Ph.D. Thesis, Department of Computer Science, University of Edinburgh, 1992.
- [Girard, 1987] J.-Y. Girard. Linear logic, *Theoretical Computer Science*, **50**, 1–102, 1987.
- [Giunchiglia *et al.*, 1996] F. Giunchiglia, P. Pecchiari, and C. Talcott. Reasoning Theories - Towards an Architecture for Open Mechanized Reasoning Systems. In *Proc. First Int. Workshop on Frontiers of Combining Systems*, F. Baader and K. U. Schulz, eds. pp. 157–174, Kluwer Academic Publishers, 1996.
- [Goguen and Burstall, 1984] J. A. Goguen and R. M. Burstall. Introducing institutions. In *Proc. Logics of Programming Workshop*, E. Clarke and D. Kozen, eds. pp. 221–256. LNCS 164, Springer-Verlag, 1984.
- [Goguen and Burstall, 1986] J. A. Goguen and R. M. Burstall. A study in the foundations of programming methodology: Specifications, institutions, charters and parchments. In *Proc. Workshop on Category Theory and Computer Programming*, Guildford, UK, September 1985, D. Pitt *et al.*, eds. pp. 313–333. LNCS 240, Springer-Verlag, 1986.
- [Goguen and Burstall, 1992] J. A. Goguen and R. M. Burstall. Institutions: Abstract model theory for specification and programming, *Journal of the Association for Computing Machinery*, **39**, 95–146, 1992.
- [Goguen and Meseguer, 1988] J. A. Goguen and J. Meseguer. Software for the Rewrite Rule Machine. In *Proc. of the Int. Conf. on Fifth Generation Computer Systems*, Tokyo, Japan, pp. 628–637. ICOT, 1988.
- [Goguen and Meseguer, 1992] J. A. Goguen and J. Meseguer. Order-sorted algebra I: Equational deduction for multiple inheritance, overloading, exceptions, and partial operations, *Theoretical Computer Science*, **105**, 217–273, 1992.
- [Goguen *et al.*, 1992] J. A. Goguen, A. Stevens, K. Hobley, and H. Hilberdink. 2OBJ: A meta-logical framework based on equational logic, *Philosophical Transactions of the Royal Society, Series A*, **339**, 69–86, 1992.
- [Goguen *et al.*, 2000] J. A. Goguen, T. Winkler, J. Meseguer, K. Futatsugi, and J.-P. Jouannaud. Introducing OBJ. In *Software Engineering with OBJ: Algebraic Specification in Action*, J. A. Goguen and G. Malcolm, eds. pp. 3–167. Kluwer Academic Publishers, 2000.
- [Große *et al.*, 1996] G. Große, S. Hölldobler, and J. Schneeberger. Linear deductive planning, *Journal of Logic and Computation*, **6**, 233–262, 1996.
- [Große *et al.*, 1992] G. Große, S. Hölldobler, J. Schneeberger, U. Sigmund, and M. Thielscher. Equational logic programming, actions, and change. In *Proc. Int. Joint Conf. and Symp. on Logic Programming*, K. Apt, ed. pp. 177–191. The MIT Press, 1992.
- [Gunter, 1991] C. Gunter. Forms of semantic specification, *Bulletin of the EATCS*, **45**, 98–113, 1991.

- [Harper *et al.*, 1993] R. Harper, F. Honsell, and G. Plotkin. A framework for defining logics, *Journal of the Association for Computing Machinery*, **40**, 143–184, 1993.
- [Harper *et al.*, 1989] R. Harper, D. Sannella, and A. Tarlecki. Structure and representation in LF. In *Proc. Fourth Annual IEEE Symp. on Logic in Computer Science*, Asilomar, California, pp. 226–237. 1989.
- [Harper *et al.*, 1989a] R. Harper, D. Sannella, and A. Tarlecki. Logic representation in LF. In *Category Theory and Computer Science*, Manchester, UK, September 1989, D. H. Pitt *et al.*, eds. pp. 250–272. LNCS 389, Springer-Verlag, 1989.
- [Hayes, 1987] P. J. Hayes. What the frame problem is and isn't. In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Z. W. Pylyshyn, ed. pp. 123–137. Ablex Publishing Corp., 1987.
- [Hennessy, 1990] M. Hennessy. *The Semantics of Programming Languages: An Elementary Introduction Using Structural Operational Semantics*, John Wiley and Sons, 1990.
- [Henz *et al.*, 1995] M. Henz, G. Smolka, and J. Würtz. Object-oriented concurrent constraint programming in Oz. In *Principles and Practice of Constraint Systems: The Newport Papers*, V. Saraswat and P. van Hentenryck, eds. pp. 29–48. The MIT Press, 1995.
- [Hölldobler, 1992] S. Hölldobler. On deductive planning and the frame problem. In *Logic Programming and Automated Reasoning*, St. Petersburg, Russia, July 1992, A. Voronkov, ed. pp. 13–29. LNAI 624, Springer-Verlag, 1992.
- [Hölldobler and Schneeberger, 1990] S. Hölldobler and J. Schneeberger. A new deductive approach to planning, *New Generation Computing*, **8**, 225–244, 1990.
- [Huet, 1973] G. Huet. A unification algorithm for typed lambda calculus, *Theoretical Computer Science*, **1**, 27–57, 1973.
- [Jaffar and Lassez, 1987] J. Jaffar and J. Lassez. Constraint logic programming. In *Proc. 14th. ACM Symp. on Principles of Programming Languages*, Munich, Germany, pp. 111–119. 1987.
- [Janlert, 1987] L.-E. Janlert. Modeling change—The frame problem. In *The Robot's Dilemma: The Frame Problem in Artificial Intelligence*, Z. W. Pylyshyn, ed. pp. 1–40. Ablex Publishing Corp., 1987.
- [Janson and Haridi, 1991] S. Janson and S. Haridi. Programming paradigms of the Andorra kernel language. In *Proc. 1991 Int. Symp. on Logic Programming*, V. Saraswat and K. Ueda, eds. pp. 167–186. The MIT Press, 1991.
- [Jouannaud and Kirchner, 1991] J.-P. Jouannaud and C. Kirchner. Solving equations in abstract algebras: A rule-based survey of unification. In *Computational Logic: Essays in Honor of Alan Robinson*, J.-L. Lassez and G. Plotkin, eds. pp. 257–321. The MIT Press, 1991.
- [Kahn, 1987] G. Kahn. *Natural semantics*, Technical report 601, INRIA Sophia Antipolis, February 1987.
- [Kirchner *et al.*, 1990] C. Kirchner, H. Kirchner, and M. Rusinowitch. Deduction with symbolic constraints, *Revue Française d'Intelligence Artificielle*, **4**, 9–52, 1990.
- [Kirchner *et al.*, 1995] C. Kirchner, H. Kirchner, and M. Vittek. Designing constraint logic programming languages using computational systems. In *Principles and Practice of Constraint Systems: The Newport Papers*, V. Saraswat and P. van Hentenryck, eds. pp. 133–160. The MIT Press, 1995.
- [Laneve and Montanari, 1992] C. Laneve and U. Montanari. Axiomatizing permutation equivalence in the  $\lambda$ -calculus. In *Proc. Third Int. Conf. on Algebraic and Logic Programming*, Volterra, Italy, September 1992, H. Kirchner and G. Levi, eds. pp. 350–363. LNCS 632, Springer-Verlag, 1992.
- [Laneve and Montanari, 1996] C. Laneve and U. Montanari. Axiomatizing permutation equivalence, *Mathematical Structures in Computer Science*, **6**, 219–249, 1996.
- [Mac Lane, 1971] S. Mac Lane. *Categories for the Working Mathematician*, Springer-Verlag, 1971.
- [Martelli and Montanari, 1982] A. Martelli and U. Montanari. An efficient unification algorithm, *ACM Transactions on Programming Languages and Systems*, **4**, 258–282, 1982.

- [Martini and Masini, 1993] S. Martini and A. Masini. *A computational interpretation of modal proofs*, Technical report TR-27/93, Dipartimento di Informatica, Università di Pisa, November 1993.
- [Martí-Oliet and Meseguer, 1991] N. Martí-Oliet and J. Meseguer. From Petri nets to linear logic through categories: A survey, *International Journal of Foundations of Computer Science*, **2**, 297–399, 1991.
- [Martí-Oliet and Meseguer, 1999] N. Martí-Oliet and J. Meseguer. Action and change in rewriting logic. In *Dynamic Worlds: From the Frame Problem to Knowledge Management*, R. Pareschi and B. Fronhöfer, eds. pp. 1–53. Kluwer Academic Publishers, 1999.
- [Martí-Oliet and Meseguer, 2001] N. Martí-Oliet and J. Meseguer. Rewriting logic: Roadmap and bibliography, *Theoretical Computer Science*, 2001. To appear.
- [Masini, 1993] A. Masini. *A Proof Theory of Modalities for Computer Science*, Ph.D. Thesis, Dipartimento di Informatica, Università di Pisa, 1993.
- [Masseron *et al.*, 1990] M. Masseron, C. Tollu, and J. Vauzeilles. Generating plans in linear logic. In *Foundations of Software Technology and Theoretical Computer Science*, Bangalore, India, December 1990, K. V. Nori and C. E. Veni Madhavan, eds. pp. 63–75. LNCS 472, Springer-Verlag, 1990.
- [Masseron *et al.*, 1993] M. Masseron, C. Tollu, and J. Vauzeilles. Generating plans in linear logic I: Actions as proofs, *Theoretical Computer Science*, **113**, 349–370, 1993.
- [Matthews *et al.*, 1993] S. Matthews, A. Smaill, and D. Basin. Experience with  $FS_0$  as a framework theory. In *Logical Environments*, G. Huet and G. Plotkin, eds. pp. 61–82. Cambridge University Press, 1993.
- [Mayoh, 1985] B. Mayoh. *Galleries and institutions*, Technical report DAIMI PB-191, Computer Science Department, Aarhus University, 1985.
- [McCarthy and Hayes, 1969] J. McCarthy and P. J. Hayes. Some philosophical problems from the standpoint of artificial intelligence. In *Machine Intelligence 4*, B. Meltzer and D. Michie, eds. pp. 463–502. Edinburgh University Press, 1969.
- [Meseguer, 1989] J. Meseguer. General logics. In *Logic Colloquium '87*, H.-D. Ebbinghaus *et al.*, eds. pp. 275–329. North-Holland, 1989.
- [Meseguer, 1990] J. Meseguer. A logical theory of concurrent objects. In *Proc. OOPSLA-ECOOP'90*, N. Meyrowitz, ed. pp. 101–115. ACM Press, 1990.
- [Meseguer, 1992] J. Meseguer. Conditional rewriting logic as a unified model of concurrency, *Theoretical Computer Science*, **96**, 73–155, 1992.
- [Meseguer, 1992b] J. Meseguer. Multiparadigm logic programming. In *Proc. Third Int. Conf. on Algebraic and Logic Programming*, Volterra, Italy, September 1992, H. Kirchner and G. Levi, eds. pp. 158–200. LNCS 632, Springer-Verlag, 1992.
- [Meseguer, 1993] J. Meseguer. A logical theory of concurrent objects and its realization in the Maude language. In *Research Directions in Object-Based Concurrency*, G. Agha, P. Wegner, and A. Yonezawa, eds. pp. 314–390. The MIT Press, 1993.
- [Meseguer, 1993b] J. Meseguer. Solving the inheritance anomaly in concurrent object-oriented programming. In *Proc. ECOOP'93, 7th. European Conf.*, Kaiserslautern, Germany, July 1993, O. M. Nierstrasz, ed. pp. 220–246. LNCS 707, Springer-Verlag, 1993.
- [Meseguer, 1996] J. Meseguer. Rewriting logic as a semantic framework for concurrency: A progress report. In *CONCUR'96: Concurrency Theory*, Pisa, Italy, August 1996, U. Montanari and V. Sassone, eds. pp. 331–372. LNCS 1119, Springer-Verlag, 1996.
- [Meseguer, 1998] J. Meseguer. Membership algebra as a logical framework for equational specification. In *Recent Trends in Algebraic Development Techniques*, Tarquinia, Italy, June 1997, F. Parisi-Presicce, ed. pp. 18–61. LNCS 1376, Springer-Verlag, 1998.
- [Meseguer *et al.*, 1992] J. Meseguer, K. Futatsugi, and T. Winkler. Using rewriting logic to specify, program, integrate, and reuse open concurrent systems of cooperating agents. In *Proc. IMSA'92, Int. Symp. on New Models for Software Architecture*, Tokyo, 1992.
- [Meseguer and Goguen, 1993] J. Meseguer and J. A. Goguen. Order-sorted algebra solves the constructor-selector, multiple representation and coercion problems, *Information and Computation*, **104**, 114–158, 1993.

- [Meseguer *et al.*, 1989] J. Meseguer, J. A. Goguen, and G. Smolka. Order-sorted unification, *Journal of Symbolic Computation*, **8**, 383–413, 1989.
- [Meseguer and Winkler, 1992] J. Meseguer and T. Winkler. Parallel programming in Maude. In *Research Directions in High-Level Parallel Programming Languages*, J. P. Banâtre and D. Le Métayer, eds. pp. 253–293. LNCS 574, Springer-Verlag, 1992.
- [Miller, 1991] D. Miller. A logic programming language with lambda-abstraction, function variables, and simple unification, *Journal of Logic and Computation*, **1**, 497–536, 1991.
- [Milner, 1980] R. Milner. *A Calculus of Communicating Systems*, LNCS 92, Springer-Verlag, 1980.
- [Milner, 1989] R. Milner. *Communication and Concurrency*, Prentice Hall, 1989.
- [Milner, 1990] R. Milner. Operational and algebraic semantics of concurrent processes. In *Handbook of Theoretical Computer Science, Vol. B: Formal Models and Semantics*, J. van Leeuwen *et al.*, eds. pp. 1201–1242. The MIT Press/Elsevier, 1990.
- [Milner, 1992] R. Milner. Functions as processes, *Mathematical Structures in Computer Science*, **2**, 119–141, 1992.
- [Mosses, 1998] P. D. Mosses. Semantics, modularity, and rewriting logic. In *Proc. Second Int. Workshop on Rewriting Logic and its Applications*, Pont-à-Mousson, France, September 1998, C. Kirchner and H. Kirchner, eds. ENTCS 15, Elsevier, 1998. <http://www.elsevier.nl/locate/entcs/volume15.html>
- [Mosses, 1999] P. D. Mosses. *Foundations of modular SOS*, Technical report, Research Series RS-99-54, BRICS, Department of Computer Science, University of Aarhus, 1999.
- [Nadathur and Miller, 1988] G. Nadathur and D. Miller. An overview of  $\lambda$ Prolog. In *Fifth Int. Joint Conf. and Symp. on Logic Programming*, K. Bowen and R. Kowalski, eds. pp. 810–827. The MIT Press, 1988.
- [Nielson and Nielson, 1992] H. R. Nielson and F. Nielson. *Semantics with Applications: A Formal Introduction*, John Wiley and Sons, 1992.
- [Nieuwenhuis and Rubio, 1992] R. Nieuwenhuis and A. Rubio. Theorem proving with ordering constrained clauses. In *Proc. 11th. Int. Conf. on Automated Deduction*, Saratoga Springs, NY, June 1992, D. Kapur, ed. pp. 477–491. LNAI 607, Springer-Verlag, 1992.
- [Nipkow, 1993] T. Nipkow. Functional unification of higher-order patterns. In *Proc. Eighth Annual IEEE Symp. on Logic in Computer Science*, Montreal, Canada, pp. 64–74. 1993.
- [Olveczky, 2000] P. C. Ölveczky. *Specification and Analysis of Real-Time and Hybrid Systems in Rewriting Logic*, Ph.D. thesis, University of Bergen, Norway, 2000.
- [Paulson, 1989] L. Paulson. The foundation of a generic theorem prover, *Journal of Automated Reasoning*, **5**, 363–397, 1989.
- [Pfenning, 1989] F. Pfenning. Elf: A language for logic definition and verified metaprogramming. In *Proc. Fourth Annual IEEE Symp. on Logic in Computer Science*, Asilomar, California, pp. 313–322. 1989.
- [Plotkin, 1981] G. D. Plotkin. *A structural approach to operational semantics*, Technical report DAIMI FN-19, Computer Science Department, Aarhus University, September 1981.
- [Poigné, 1989] A. Poigné. Foundations are rich institutions, but institutions are poor foundations. In *Categorical Methods in Computer Science with Aspects from Topology*, H. Ehrig *et al.* eds. pp. 82–101. LNCS 393, Springer-Verlag, 1989.
- [Reichwein *et al.*, 1992] G. Reichwein, J. L. Fiadeiro, and T. Maibaum. Modular reasoning about change in an object-oriented framework, Abstract presented at the *Workshop Logic & Change at GWAI'92*, September 1992.
- [Reisig, 1995] W. Reisig. *Petri Nets: An Introduction*, Springer-Verlag, 1985.
- [Salibra and Scollo, 1993] A. Salibra and G. Scollo. A soft stairway to institutions. In *Recent Trends in Data Type Specification*, M. Bidoit and C. Choppy, eds. pp. 310–329. LNCS 655, Springer-Verlag, 1993.
- [Saraswat, 1992] V. J. Saraswat. *Concurrent Constraint Programming*, The MIT Press, 1992.



- [Schmidt-Schauss, 1989] M. Schmidt-Schauss. *Computational Aspects of Order-Sorted Logic with Term Declarations*, LNCS 395, Springer-Verlag, 1989.
- [Scott, 1974] D. Scott. Completeness and axiomatizability in many-valued logic. In *Proceedings of the Tarski Symposium*, L. Henkin *et al.* eds. pp. 411–135. American Mathematical Society, 1974.
- [Seely, 1989] R. A. G. Seely. Linear logic, \*-autonomous categories and cofree coalgebras. In *Categories in Computer Science and Logic*, Boulder, Colorado, June 1987, J. W. Gray and A. Scedrov, eds. pp. 371–382. Contemporary Mathematics 92, American Mathematical Society, 1989.
- [Shapiro, 1989] E. Shapiro. The family of concurrent logic programming languages, *ACM Computing Surveys*, **21**, 412–510, 1989.
- [Smolka, 1989] G. Smolka. *Logic Programming over Polymorphically Order-Sorted Types*, Ph.D. Thesis, Fachbereich Informatik, Universität Kaiserslautern, 1989.
- [Smolka *et al.*, 1989] G. Smolka, W. Nutt, J. A. Goguen, and J. Meseguer. Order-sorted equational computation. In *Resolution of Equations in Algebraic Structures, Volume 2*, M. Nivat and H. Ait-Kaci, eds. pp. 297–367. Academic Press, 1989.
- [Smullyan, 1961] R. M. Smullyan. *Theory of Formal Systems*, Annals of Mathematics Studies 47, Princeton University Press, 1961.
- [Stehr, 2002] M.-O. Stehr. *Rewriting Logic and Type Theory — From Applications to Unification*, Ph.D. thesis, Computer Science Department, University of Hamburg, Germany, 2002. In preparation.
- [Talcott, 1993] C. Talcott. A theory of binding structures and applications to rewriting, *Theoretical Computer Science*, **112**, 99–143, 1993.
- [Tarlecki, 1984] A. Tarlecki. Free constructions in algebraic institutions. In *Proc. Mathematical Foundations of Computer Science '84*, M. P. Chytil and V. Koubek, eds. pp. 526–534. LNCS 176, Springer-Verlag, 1984.
- [Tarlecki, 1985] A. Tarlecki. On the existence of free models in abstract algebraic institutions, *Theoretical Computer Science*, **37**, 269–304, 1985.
- [Troelstra, 1992] A. S. Troelstra. *Lectures on Linear Logic*, CSLI Lecture Notes 29, Center for the Study of Language and Information, Stanford University, 1992.
- [Verdejo and N. Martí-Oliet, 2000] A. Verdejo and N. Martí-Oliet. Implementing CCS in Maude. In *Formal Methods for Distributed System Development*, Pisa, Italy, October 2000, T. Bolognesi and D. Latella, eds. pp. 351–366. Kluwer Academic Publishers, 2000.
- [Walther, 1985] C. Walther. A mechanical solution to Schubert's steamroller by many-sorted resolution, *Artificial Intelligence*, **26**, 217–224, 1985.
- [Walther, 1986] C. Walther. A classification of many-sorted unification theories. In *Proc. 8th. Int. Conf. on Automated Deduction*, Oxford, England, 1986, J. H. Siekmann, ed. pp. 525–537. LNCS 230, Springer-Verlag, 1986,



DAVID BASIN, SEÁN MATTHEWS

## LOGICAL FRAMEWORKS

### 1 INTRODUCTION

One way to define a logic is to specify a language and a deductive system. For example, the language of first-order logic consists of the syntactic categories of terms and formulae, and its deductive system establishes which formulae are theorems. Typically we have a specific language in mind for a logic, but some flexibility about the kind of deductive system we use; we are able to select from, e.g., a Hilbert calculus, a sequent calculus, or a natural deduction calculus. A *logical framework* is an abstract characterization of one of these kinds of deductive system that we can use to formalize particular examples. Thus a logical framework for natural deduction should allow us to formalize natural deduction for a wide range of logics from, e.g., propositional logic to intuitionistic type-theories or classical higher-order logic.

Exactly how a logical framework abstractly characterizes a kind of deductive system is difficult to pin-down formally. From a high enough level of abstraction, we can see a deductive system as defining sets; i.e. we have a recursive set corresponding to well-formed syntax, a recursive set corresponding to proofs, and a recursively enumerable set of provable formulae.<sup>1</sup> But this view is really too coarse: we expect a logical framework to be able to capture more than just the sets of well-formed and provable formulae associated with a logic. If this were all that we wanted, then any Turing complete programming language would constitute a logical framework, in so far as it can implement a proof-checker for any logic.

In this chapter we present and examine two different kinds of frameworks, each representing a different view of what a deductive system is:  $\rightarrow$ -frameworks (deduction interpreted as reasoning in a weak logic of implication) and ID-frameworks (deduction interpreted as showing that a formula is a member of an inductively defined set). Either of these can be used to formalize any recursively enumerable relation. However, before calling a system a logical framework we will demand that it preserves additional structure. Thus we first consider what are the important and distinguishing characteristics of the different kinds of deductive systems, then we examine frameworks based on different sorts of possible *metallogic* (or *metatheory*) and we show that these are well-suited to representing the deductive systems for certain classes of *object logics* (or *object theories*). By providing procedures by which the deductive system of any object logic in a class can be naturally encoded in some metallogic, we show how effective frameworks are for formalizing particular logics.

---

<sup>1</sup>We make the assumption in this chapter that the property of being a well-formed syntactic entity or a proof is recursive; i.e. we know one when we see it.

When we say that an encoding is *natural* we shall mean not only that it is high-level and declarative but also that there is an appropriate bijection between derivations in the object logic and derivations manipulating the encoding in the metalogic. For example, a Hilbert system can be naturally encoded using an inductive definition where each rule in the object logic corresponds to a rule in the metalogic. Similarly, there are natural encodings of many natural deduction systems in a fragment of the language of higher-order logic, which make use of a simple, uniform, method for writing down rules as formulae of the metalogic, so that it is possible to translate between natural deduction proofs of the object logic and derivations in the metalogic.

The term ‘logical framework’ came into use in the 1980s; however the study of metalogics and methods of representing one logic in another has a longer history in the work of logicians interested in languages for the metatheoretic analysis of logics, and computer scientists seeking conceptual and practical foundations for implementing logic-based systems. Although these motivations differ and are, at least in part, application oriented, behind them we find something common and more general: logical frameworks clarify our ideas of what we mean by a ‘deductive system’ by reducing it to its abstract principles. Thus work on logical frameworks contributes to the work of Dummett, Gentzen, Hacking, Prawitz, and others on the larger question of what is a logic.

### 1.1 *Some historical background*

Historically, the different kinds of deductive systems have resulted from different views of logics and their applications. Thus the idea of a deductive system as an inductive definition developed out of the work of Frege, Russell, Hilbert, Post, Gödel and Gentzen attempting to place mathematics on firm foundational ground. In particular, Hilbert’s program (which Gödel famously showed to be impossible) was to use the theory of proofs to establish the consistency of all of classical mathematics, using only finitary methods. From this perspective, of *metatheory as proof theory*, a deductive system defines a set of objects in terms of an initial set and a set of rules that generate new objects from old, and a proof is the tree of rule applications used to show that an object is in the set. For Frege and Hilbert these objects were simply theorems, but later Gentzen took them to be sequents, i.e. pairs of collections of formulae. But, either way, a deductive system defines a recursively enumerable set, which is suitable for analysis using inductive arguments.

How should the metatheory used to define these inductive definitions be characterized? Hilbert required it to be finitary, so it has traditionally been taken to be the primitive recursive fragment of arithmetic (which is essentially, e.g., what Gödel [1931] used for his incompleteness theorems). However, despite the endorsement by Hilbert and Gödel, arithmetic is remarkably unsuitable, in the primitives it offers, as a general theory of inductive definitions. Thus, more recent investigations, such as those of Smullyan [1961] and Feferman [1990], have

proposed theories of inductive definitions based on sets of strings or S-expressions, structures more tractable in actual use.

A different view of deductive systems is found in work in computer science and artificial intelligence, where logics have been formalized for concrete applications like machine checked proof. The work of, e.g., de Bruijn [Nederpelt *et al.*, 1994] does not attempt to analyze the meta-theoretic properties of deductive systems, but concentrates rather on common operations, such as binding and substitution. The goal is to provide an abstract characterization of such operations so that deductive systems can be easily implemented and used to prove theorems; i.e. metatheory is seen as providing a *unifying language* for implementing, and ultimately using, deductive systems. A result of this concern with ease in use (i.e. building proofs) rather than ease of metatheoretic analysis (i.e. reasoning about proofs) is that work has emphasized technically more complex, but more usable, notations such as natural deduction, and resulted in frameworks based on higher-order logics and intuitionistic type-theories instead of the inductive definitions of the older proof theory tradition.

## 1.2 Focus and organization

This chapter presents the concepts underlying the various logical frameworks that have been proposed, examining the relationship between different kinds of deductive systems and metatheory. Many issues thus fall outside our scope. For example, we do not investigate semantically based approaches to formalizing and reasoning about logics, such as the *institutions* of Goguen and Burstall [1992] or the *general logics* of Meseguer [1989]. Even within our focus, we do not attempt a comprehensive survey of all the formalisms that have been proposed. Neither do we directly consider implementation questions, even though computer implementations now play an important role in the field. Further references are given in the bibliography; the reader is referred in particular to the work of Avron *et al.*, Paulson, Pfenning, Pollack, and ourselves, where descriptions of first-hand experience with logical frameworks can be found.

The remainder of this chapter is organized as follows. In §2 we briefly survey three kinds of deductive systems, highlighting details relevant for formalizing abstractions of them. In §3 we consider a logic based on minimal implication as an abstraction of natural deduction. It turns out that this abstraction is closely related to a generalization of natural deduction due to Schroeder-Heister. In §4 we consider in detail a particular metatheory that formalizes this abstraction. We also consider quantification and so-called ‘higher-order syntax’. In §5 we present a case study: the problem of formalizing modal logics. In §6 we examine sequent calculi and their more abstract formalization as consequence relations. In §7 we investigate the relationship between sequent systems and inductive definitions, and present a particular logic for inductive definitions. Finally, we draw conclusions and point to some current and possible future research directions.

## 2 KINDS OF DEDUCTIVE SYSTEMS

Many different kinds of formalization of deduction have been proposed, for a range of purposes. However in this chapter we are interested in deductive systems designed for the traditional purposes of logicians and philosophers, of investigating the foundations of language and reasoning; in fact we restrict our interest even further, to three kinds of system, which are commonly called *Hilbert calculi*, *sequent calculi* and *natural deduction*. But this restriction is more apparent than real since, between them, these three cover the vast majority of deductive systems that have been proposed. We describe only the details that are important here, i.e. the basic mechanics; for deeper and more general discussion, the reader is referred to Sundholm's articles on systems of deduction elsewhere in this handbook [1983; 1986].

For the purposes of comparison, we shall present, as an example, the same simple logics in each style: propositional logics of minimal and classical implication. The language we will work with then is as follows.

**DEFINITION 1.** Given a set of atomic propositions  $P$ , the language of propositions,  $L$ , is defined as the smallest set containing  $P$ , where if  $A$  and  $B$  are in  $L$ , then  $(A \supset B)$  is in  $L$ .

For the sake of readability we assume that  $\supset$  associates to the right and omit unnecessary parentheses; for example, we abbreviate  $(A \supset (B \supset C))$  as  $A \supset B \supset C$ . We now proceed to the different presentations.

### 2.1 Hilbert calculi

Historically, Hilbert calculi are the oldest kind of presentation we consider. They are also, in some technical sense, the simplest. A Hilbert calculus defines a set of *theorems* in terms of a set of *axioms*  $Ax$  and a set of *rules of proof*  $R$ . A *rule of proof* is a relation between formulae  $A_1, \dots, A_n$  and a formula  $A$ . A rule is usually written in a two-dimensional format and sometimes decorated with its name. For example

$$\frac{A_1 \dots A_n}{A} \textit{ name}$$

says that by rule *name*, given  $A_1, \dots, A_n$ , it follows that  $A$ . The set of formulae defined by a Hilbert calculus is the smallest set containing  $Ax$  and closed under  $R$ ; we call this set a *theory*, and the formulae in it *theorems*.

The set of theorems in a logic of minimal implication can be defined as a Hilbert calculus where the set of axioms contains all instances of the axiom schemata  $K$

$$A \supset B \supset A$$

and  $S$

$$(A \supset B) \supset (A \supset B \supset C) \supset A \supset C.$$

We call these schemata because  $A$ ,  $B$  and  $C$  stand for arbitrary formulae in  $L$ . In addition, we have one rule of proof, *detachment*, which takes the form

$$\frac{A \supset B \quad A}{B} \text{Det.}$$

The Hilbert calculus defined by  $K$ ,  $S$  and  $Det$  is the set of theorems of the minimal (or intuitionistic) logic of implication. We call this theory  $HJ^{\supset}$ .<sup>2</sup>

The set of theorems in a logic of classical implication,  $HK^{\supset}$ , is slightly larger than  $HJ^{\supset}$ . We can formalize  $HK^{\supset}$  by adding to  $HJ^{\supset}$  a third axiom schema, for *Peirce's law*

$$((A \supset B) \supset A) \supset A. \tag{1}$$

A proof in a Hilbert calculus consists of a demonstration that a formula is in the (inductively defined) set of theorems. We can think of a proof as either a tree, where the leaves are axioms, and the branches represent rule applications, or as a list of formulae, ending in the theorem, where each entry is either an axiom, or follows from previous entries by a rule. Following common practice, we use the list notation here.

The following is an example of a proof of  $A \supset A$  in  $HJ^{\supset}$  and thus also in  $HK^{\supset}$ :

- |  |            |
|--|------------|
| 1. $A \supset A \supset A$   | $K$        |
| 2. $(A \supset A \supset A) \supset (A \supset (A \supset A) \supset A) \supset A \supset A$ | $S$        |
| 3. $A \supset (A \supset A) \supset A$   | $K$        |
| 4. $(A \supset (A \supset A) \supset A) \supset A \supset A$                                 | $Det\ 2,1$ |
| 5. $A \supset A$   | $Det\ 4,3$ |

It turns out that proving theorems in a Hilbert calculus by hand, or even on a machine, is not practical: proofs can quickly grow to be enormous in size, and it is often necessary (e.g. in the proof we have just presented) to invent instances of axiom schemata that have no intuitive relationship to the formula being proven. However Hilbert calculi were never really intended to be used to build proofs, but rather as a tool for the metatheoretic analysis of logical concepts such as deduction. And from this point of view, the most important fact about Hilbert calculi is that they are essentially inductive definitions;<sup>3</sup> i.e. well-suited for arguments (about provability) by induction.

---

<sup>2</sup>For the proof systems presented in this section, we follow the tradition where the first letter indicates the kind of deduction system (H for Hilbert, N for natural deduction, and L for sequent calculus), the second letter indicates the logic (J for minimal, or intuitionistic logic, and K for classical logic), and superscripts name the connectives.

<sup>3</sup>The two are so closely related that, e.g., Aczel [1977] identifies them.

## 2.2 Sequent calculi

Useful though Hilbert calculi are, Gentzen [1934] found them unsatisfactory for his particular purposes, and so developed a very different style that has since become the standard notation for much of proof theory, and which is known as *sequent calculus*.

With a sequent calculus we do not define directly the set of theorems; instead we define a binary ‘sequent’ relation between collections of formulae,  $\Gamma$  and  $\Delta$ , and identify a subset of instances of this relation with theorems. We shall write this relation as  $\Gamma \vdash \Delta$ , where  $\Gamma$  is called the *antecedent* and  $\Delta$  the *succedent*. This is often read as ‘if all the formulae in  $\Gamma$  are true, then at least one of the formulae in  $\Delta$  is true’.

The rules of a sequent calculus are traditionally divided into two subsets consisting of the *logical* rules, which define logical connectives, and the *structural* rules, which define the abstract properties of the sequent relation itself. The basic properties of any sequent system are given by the following rules which state that  $\vdash$  is reflexive for singleton collections (*Basic*) and satisfies a form of transitivity (*Cut*):

$$\frac{}{A \vdash A} \textit{Basic} \quad \frac{\Gamma \vdash A, \Delta \quad \Gamma', A \vdash \Delta'}{\Gamma, \Gamma' \vdash \Delta, \Delta'} \textit{Cut} \quad (2)$$

A typical set of structural rules is then

$$\begin{array}{cc} \frac{\Gamma \vdash \Delta}{\Gamma, A \vdash \Delta} \textit{WL} & \frac{\Gamma \vdash \Delta}{\Gamma \vdash A, \Delta} \textit{WR} \\ \frac{\Gamma, A, A \vdash \Delta}{\Gamma, A \vdash \Delta} \textit{CL} & \frac{\Gamma \vdash A, A, \Delta}{\Gamma \vdash A, \Delta} \textit{CR} \\ \frac{\Gamma, A, B, \Gamma' \vdash \Delta}{\Gamma, B, A, \Gamma' \vdash \Delta} \textit{PL} & \frac{\Gamma \vdash \Delta, A, B, \Delta'}{\Gamma \vdash \Delta, B, A, \Delta'} \textit{PR} \end{array} \quad (3)$$

which define  $\vdash$  to be a relation on finite sets that is also monotonic in both arguments (what we will later call *ordinary*). The names *WL*, *CL* and *PL* stand for *Weakening*, *Contraction* and *Permutation Left* of the sequent, while *WR*, *CR* and *PR* name the same operations on the right.

We can give a deductive system for classical implication, which we call  $\text{LK}^{\supset}$ , in terms of this sequent calculus by adding logical rules for  $\supset$ :

$$\frac{\Gamma \vdash A, \Delta \quad \Gamma', B \vdash \Delta'}{\Gamma, \Gamma', A \supset B \vdash \Delta, \Delta'} \supset\text{-L} \quad \frac{\Gamma, A \vdash B, \Delta}{\Gamma \vdash A \supset B, \Delta} \supset\text{-R}$$

The names  $\supset\text{-L}$  and  $\supset\text{-R}$  indicate that these are rules for introducing the implication connective on the left and the right of  $\vdash$ . We get a sequent calculus for



minimal implication,  $LJ^{\supset}$ , by adding the restriction on  $\vdash$  that its succedent is a sequence containing exactly one formula.<sup>4</sup>

Like Hilbert proofs, sequent calculus proofs can be equivalently documented as lists or trees. Since proofs are less unwieldy than for Hilbert calculi, trees are a practical notation. A sequent calculus proof of  $A \supset A$  is trivial, so instead we take as an example a proof of Peirce's law (1):

$$\begin{array}{c}
 \frac{}{A \vdash A} \textit{Basic} \\
 \frac{}{A \vdash B, A} \textit{WR} \\
 \frac{}{\vdash (A \supset B), A} \supset\textit{-R} \quad \frac{}{A \vdash A} \textit{Basic} \\
 \frac{}{((A \supset B) \supset A) \vdash A, A} \supset\textit{-L} \\
 \frac{}{((A \supset B) \supset A) \vdash A} \textit{CR} \\
 \frac{}{\vdash ((A \supset B) \supset A) \supset A} \supset\textit{-R}
 \end{array}$$

Notice that it is critical in this proof that the succedent can consist of more than one formula, and that *CR* can be applied. Neither this proof (since *CR* is not available), nor any other, of Peirce's law is possible in  $LJ^{\supset}$ .

Technically we can regard a sequent calculus as a Hilbert calculus for a binary connective  $\vdash$ ; however the theorems of this system are in the language of sequents over  $L$ , not formulae in the language of  $L$  itself. The set of theorems a sequent calculus defines is taken to be the set of formulae  $A$  in  $L$  such that  $\vdash A$  is provable, i.e., there is a proof of the sequent  $\Gamma \vdash \Delta$ , where  $\Gamma$  is empty, and  $\Delta$  is the singleton  $A$ .

### 2.3 Natural deduction

A second important kind of deductive system that Gentzen [1934] (and subsequently Prawitz [1965]) developed was *natural deduction*. In contrast to the sequent calculus, which is intended as a formalism that supports metatheoretic analysis, natural deduction, as its name implies, is intended to reflect the way people 'naturally' work out logical arguments. Thus Gentzen suggested that, e.g., in order to convince ourselves that the *S* axiom schema

$$(A \supset B) \supset (A \supset B \supset C) \supset A \supset C$$

is true, we would, reading  $\supset$  as 'implies', informally reason as follows: 'Assuming  $A \supset B$  and then  $A \supset B \supset C$  we have to show that  $A \supset C$ , and to do this, it is enough to show that  $C$  is true assuming  $A$ . But if  $A$  is true then, from the first assumption,  $B$  is true, and, given that  $A$  and  $B$  are true, by the second assumption  $C$  is true. Thus the *S* schema is true (under the intended interpretation)'.

<sup>4</sup>As a result, in  $LJ^{\supset}$  the structural rules (*WR*, *CR* and *PR*) become inapplicable, and  $\Delta = \emptyset$  in the logical rule  $\supset\textit{-L}$ .

To represent this style of argument *under assumption* we need a new kind of rule, called a *rule of inference*, that allows temporary hypotheses to be made and discharged. This is best explained by example. Consider implication; we can informally describe some of its properties as: (i) if, assuming  $A$ , it follows that  $B$ , then it follows that  $A \supset B$ , and (ii) if  $A \supset B$  and  $A$ , then it follows that  $B$ . We might represent this reasoning diagrammatically as follows:

$$\frac{\begin{array}{c} [A] \\ \vdots \\ B \end{array} \supset\text{-}I}{A \supset B} \quad \frac{A \supset B \quad A}{B} \supset\text{-}E \quad (4)$$

where  $\supset\text{-}I$  and  $\supset\text{-}E$  are to be pronounced as ‘Implication Introduction’ and ‘Implication Elimination’, since they explain how to introduce and to eliminate (i.e. to create and to use) a formula with  $\supset$  as the main connective.

We call  $A \supset B$  in  $\supset\text{-}E$  the *major premise*, and  $A$  the *minor premise*. In general, the major premise of an elimination rule is the premise in which the eliminated connective is exhibited and all other premises are minor premises. Square brackets around assumptions indicate that they are considered to be temporary, and made only for the course of the derivation; when applying the rule we can *discharge* these occurrences. Thus, when applying  $\supset\text{-}I$ , we can discharge (zero or more) occurrences of the assumption  $A$  which has been made for the purposes of building a derivation of  $B$ , which shows that  $A \supset B$ . Of course, when applying the  $\supset\text{-}I$  rule, there may be other assumptions (so called open assumptions) that are not discharged by the rule. Similarly, the two premises of  $\supset\text{-}E$  may each have open assumptions, and the conclusion follows under the union of these.

To finish our account of  $\supset$ , we must explain how these rules can be used to build formal proofs. Gentzen formalized natural deduction derivations just like in sequent and Hilbert calculi, as trees, explaining how formulae are derived from formulae. With natural deduction though, there is an added complication: we also have to track the temporary assumptions that are made and discharged. Thus along with the tree of formulae, a natural deduction derivation has a *discharge function*, which associates with each node  $N$  in the tree leaf nodes above  $N$  that the rule application at node  $N$  discharges. Moreover, there is the proviso that each leaf node can be discharged by at most one node. A proof, then, is a derivation where all the assumptions are discharged (i.e. all assumptions are temporary); the formula proven is a theorem of the logic.

A natural deduction proof, with its discharge function, is a complex object in comparison with a Hilbert or sequent proof. However, it has a simple two-dimensional representation: we just decorate each node with the name of the rule to which it corresponds, and a unique number, and decorate with the same number each leaf node of the tree that it discharges. In this form we can document the previously given informal argument of the truth of the  $S$  axiom schema as a formal

proof (we only number the nodes that discharge some formula):<sup>5</sup>

$$\begin{array}{c}
 \frac{[A \supset B \supset C]_2 \quad [A]_3 \supset\text{-}E}{B \supset C} \supset\text{-}E \quad \frac{[A \supset B]_1 \quad [A]_3 \supset\text{-}E}{B} \supset\text{-}E \\
 \frac{\quad}{C} \supset\text{-}E \\
 \frac{C}{A \supset C} \supset\text{-}I_3 \\
 \frac{A \supset C}{(A \supset B \supset C) \supset A \supset C} \supset\text{-}I_2 \\
 \frac{(A \supset B \supset C) \supset A \supset C}{(A \supset B) \supset (A \supset B \supset C) \supset A \supset C} \supset\text{-}I_1
 \end{array}$$

Once we have committed ourselves to this formalization of natural deduction, we find that the two rules (4) together define exactly minimal implication, in a deductive system which we call  $\text{NJ}^\supset$ .

While the claims of natural deduction to be the most intuitive way to reason are plausible, the style does have some problems. First, there is the question of how to encode classical implication. We can equally easily give a direct presentation of either classical or minimal implication, using a Hilbert or sequent calculus, but not using natural deduction. While  $\text{NJ}^\supset$  is standard for minimal implication, there is, unfortunately, no simple equivalent for classical implication (what we might imagine calling  $\text{NK}^\supset$ ): the standard presentation of classical implication in natural deduction is as part of a larger, functionally complete, set of connectives, including, e.g., negation and disjunction, through which we can appeal to the law of excluded middle. Alternatively we can simply accept Peirce's law as an axiom. We cannot, however, using the language of natural deduction, define classical implication simply in terms of introduction and elimination rules for the  $\supset$  connective.

A second problem concerns proofs themselves, which are complex in comparison to the formal representations of proofs in Hilbert or sequent calculi. It is worth noting that a Hilbert calculus can be seen as a special simpler case of natural deduction, since axioms of a Hilbert calculus can be treated as rules with no premises, and the rules of a Hilbert calculus correspond to natural deduction rules where no assumptions are discharged.

### 3 NATURAL DEDUCTION AND THE LOGIC OF IMPLICATION

Given the previous remarks about the complexity of formalizations of natural deduction, it might seem unlikely that a satisfactory logical framework for it is possible. In fact, this is not the case. In this section we show that there is an abstraction of natural deduction that is the basis of an effective logical framework. Preliminary to this, however, we consider another way of formalizing natural deduction using sequents.

<sup>5</sup>It may help the reader trying to compare this formal proof diagram with the previous informal argument, to read it 'bottom up'; i.e. upwards from the conclusion at the root.

### 3.1 Natural deduction and sequents

We can describe natural deduction informally as ‘proof under assumption’, so the basic facility needed to formalize such a calculus is a mechanism for managing assumptions. Sequents provide this: assumptions can be stored in the antecedent of the sequent and deleted when they are discharged. Thus, if we postulate a relation  $\vdash$  satisfying *Basic* and the structural rules (3), where the succedent is a singleton set, then we can encode  $\text{NJ}^\supset$  using the rules:<sup>6</sup>

$$\frac{\Gamma, A \vdash B}{\Gamma \vdash A \supset B} \supset\text{-I} \quad \frac{\Gamma \vdash A \supset B \quad \Gamma \vdash A}{\Gamma \vdash B} \supset\text{-E} \quad (5)$$

This view of natural deduction is often technically convenient (we shall use it ourselves later) but it is unsatisfactory in some respects. Gentzen, and later Prawitz, had a particular idea in mind of how natural deduction proofs should look, and they made a distinction between it and the sequent calculus, using proof trees with discharge functions for natural deduction, and sequent notation for sequent calculus. We also later consider ‘natural’ generalizations of natural deduction that cannot easily be described using a sequent notation.

### 3.2 Encoding rules using implication

The standard notation for natural deduction, based on assumptions and their discharge, while intuitive, is formally complex. Reducing it to sequents allows us to formalize presentations in a simpler ‘Hilbert’ style; however we have also said that this is not altogether satisfactory. We now consider another notation that can encode not only natural deduction but also generalizations that have been independently proposed.

A natural ‘horizontal’ notation for rules can be based on lists  $\langle A_1, \dots, A_n \rangle$  and an arrow  $\rightarrow$ . Consider the rules (4): we can write the elimination rule in horizontal form simply as

$$\frac{A \supset B \quad A}{B} \equiv \langle A \supset B, A \rangle \rightarrow B$$

and the introduction rule as

$$\frac{\begin{array}{c} [A] \\ \vdots \\ B \end{array}}{A \supset B} \equiv \langle A^\dagger \rightarrow B \rangle \rightarrow A \supset B.$$

In the introduction rule we have marked the assumption  $A$  with the symbol  $\dagger$  to indicate that it is discharged. But this is actually unnecessary; using this linear

<sup>6</sup>Importantly, we can prove that the relation defined by this encoding satisfies the *Cut* rule above; it thus defines a consequence relation (see §6).

notation, precisely the formulae (here there is only one) occurring on the left-hand side of some  $\rightarrow$ -connective in the premise list are discharged.

This new notation is suggestive: Gentzen explicitly motivated natural deduction as a formal notation for intuitive reasoning, and we have now mapped natural language connectives such as ‘follows from’ onto the symbol  $\rightarrow$ . Why not make this explicit and read  $\rightarrow$  as formal implication, and ‘ $\langle \dots \rangle$ ’ as a conjunction of its components? The result, if we translate into the conventional language of (quantifier free) predicate logic, is just:

$$\begin{aligned} (T(A \supset B) \ \& \ T(A)) \rightarrow T(B) \\ (T(A) \rightarrow T(B)) \rightarrow T(A \supset B) \end{aligned} \tag{6}$$

which, reading the unary predicate symbol  $T$  as ‘True’, is practically identical with Gentzen’s natural language formulation.

What is the significance of this logical reading of rules? The casual relationship observed above is not sufficient justification for an identification, and we will see that we must be careful. However, after working out the details, this interpretation provides exactly what we need for an effective logical framework, allowing us to trade the complex machinery of trees and discharge functions for a pure ‘logical’ abstraction.

We separate the investigation of such interpretations into several parts. First, we present a metalogic based on (minimal) implication and conjunction and give a uniform way of translating a set of natural deduction rules  $\mathcal{R}$  into a set of formulae  $\mathcal{R}^*$  in the metalogic. For an object logic  $\mathcal{L}$  presented by a set of natural deduction rules  $\mathcal{R}$ , this yields a metatheory  $\mathcal{L}^*$  given by the metalogic extended with  $\mathcal{R}^*$ . Second, we demonstrate that for any such  $\mathcal{L}$ , the translation is *adequate*. This means that  $\mathcal{L}^*$  is ‘strong enough’ to derive (the representative of) any formula derivable in  $\mathcal{L}$ . Third, we demonstrate that the translation is *faithful*. That is that  $\mathcal{L}^*$  is not too strong; we can only derive in  $\mathcal{L}^*$  representatives of formulae that are derivable in  $\mathcal{L}$ , and further, given such a derivation, we can recover a proof in  $\mathcal{L}$ .<sup>7</sup>

### 3.3 The metatheory and translation

We have previously defined natural deduction in terms of formula-trees and discharge functions. We now describe the logic with which we propose to replace it.

Thus we assume that  $\rightarrow$  and  $\&$  have at least the properties they have in minimal logic. Above we have provided three separate formalizations of (the implicational fragment of) minimal logic: as a Hilbert calculus, as a sequent calculus, and as natural deduction. Since these all formalize the same set of theorems, we could base our analysis on any of them. However, it will be convenient for our development

---

<sup>7</sup>Notice that this is a stronger requirement than the model-theoretic requirement of soundness, which requires that we should not be able to prove anything false in the model, but not that we can recover a proof of anything that we have shown to be true.

to use natural deduction.<sup>8</sup> Even though this may seem circular, it isn't: we simply need some calculus to fix the meaning of these connectives and later to formalize the correctness of our embeddings.

For the rest of this section, we will formalize deductive systems associated with *propositional* logics. We leave the treatment of quantifiers in the object logic, for which we need quantification in the metalogic (here we need only, as we will see, quantifier-free schemata), and a general theory of syntax, to §4. For now, we simply assume a suitable term algebra formalizing the language of the object logic. We will build formulae in the metalogic from the unary predicate  $T$  and the connectives  $\rightarrow$  and  $\&$ . To aid readability we assume that  $\&$  binds tighter than  $\rightarrow$ , which (as usual) associates to the right. Now, let  $\text{NJ}^{\rightarrow, \&}$  be the natural deduction calculus based on the following rules:

$$\frac{A \quad B}{A \& B} \&-I \qquad \frac{\begin{array}{c} [A, B] \\ \vdots \\ C \end{array}}{A \& B} \&-E$$

$$\frac{\begin{array}{c} [A] \\ \vdots \\ B \end{array}}{A \rightarrow B} \rightarrow-I \qquad \frac{A \rightarrow B \quad A}{B} \rightarrow-E$$

We also formally define the translation, given by the mapping ‘\*’, from rules of natural deduction to the above language. Here  $A_i$  varies over formulae in the logic, and  $\Phi_i$  over the premises (along with the discharged hypotheses) of a rule  $\Theta$ .

$$\begin{aligned} \left( \frac{\Phi_1 \dots \Phi_n}{A} \right)^* &\rightsquigarrow \Phi_1^* \& \dots \& \Phi_n^* \rightarrow T(A) \\ \left( \frac{[A_1, \dots, A_n]}{A} \right)^* &\rightsquigarrow T(A_1) \& \dots \& T(A_n) \rightarrow T(A) \end{aligned} \quad (7)$$

Note that axioms and rules that discharge no hypotheses constitute degenerate cases, i.e.  $A^* \rightsquigarrow T(A)$ . We extend this mapping to sets of rules and formulae, i.e.  $\mathcal{R}^* \equiv \{\Theta^* \mid \Theta \in \mathcal{R}\}$ .

<sup>8</sup>One reason is that we can then directly relate derivability in the metalogic with derivability in the object logic. As is standard [Prawitz, 1965], derivability refers to natural deduction derivations with possibly open assumptions. Provability is a special case where there are no open assumptions. This generalization is also relevant to showing that we can correctly represent particular consequence relations (see §6).

### 3.4 Adequacy

If  $A$  is a theorem of an object logic  $\mathcal{L}$  defined by rules  $\mathcal{R}$ , then we would like  $T(A)$  to be provable in the metatheory  $\mathcal{L}^* = \text{NJ}^{\rightarrow, \&} + \mathcal{R}^*$  (i.e.  $\text{NJ}^{\rightarrow, \&}$  extended by the rules  $\mathcal{R}^*$ ). More generally, if  $A$  is derivable in  $\mathcal{L}$  under the assumptions  $\Delta = \{A_1, \dots, A_n\}$ , then we would like for  $T(A)$  to be derivable in  $\mathcal{L}^*$  under the assumptions  $\Delta^*$ .

Consider, for example, the object logic  $\mathcal{L}$  over the language  $\{\oplus, \otimes, \times, +\}$  defined by the following rules  $\mathcal{R}$ :

$$\frac{\frac{+}{\otimes} \alpha \quad \frac{+}{\times} \beta \quad \frac{\otimes \times}{\oplus} \gamma}{\oplus} \delta \quad \begin{array}{c} [+ \\ \vdots \\ \oplus \\ \hline \oplus \end{array} \quad (8)$$

We can prove, for example,  $\oplus$  by:

$$\frac{\frac{[+]_1}{\otimes} \alpha \quad \frac{[+]_1}{\times} \beta}{\oplus} \gamma \quad \begin{array}{c} \oplus \\ \hline \oplus \\ \delta_1 \end{array} \quad (9)$$

Under our proposed encoding, the rules are translated to the following set  $\mathcal{R}^*$ :

$$\begin{aligned} T(+) &\rightarrow T(\otimes) & (\alpha^*) \\ T(+) &\rightarrow T(\times) & (\beta^*) \\ T(\otimes) \ \& \ T(\times) &\rightarrow T(\oplus) & (\gamma^*) \\ (T(+) &\rightarrow T(\oplus)) &\rightarrow T(\oplus) & (\delta^*) \end{aligned}$$

and we can prove  $T(\oplus)$  in the metatheory  $\mathcal{L}^* = \text{NJ}^{\rightarrow, \&} + \mathcal{R}^*$  by:

$$\frac{\frac{\frac{[T(+)]_1}{\vdots \Sigma} T(\oplus)}{T(+) \rightarrow T(\oplus)} \rightarrow -I_1}{(T(+) \rightarrow T(\oplus)) \rightarrow T(\oplus)} \delta^* \quad \frac{}{T(+) \rightarrow T(\oplus)} \rightarrow -E}{T(\oplus)} \rightarrow -E$$

where  $\Sigma$  is:

$$\frac{\frac{\frac{\frac{}{T(+) \rightarrow T(\otimes)} \alpha^*}{T(+) \rightarrow T(\otimes)} \rightarrow -E \quad \frac{\frac{\frac{}{T(+) \rightarrow T(\times)} \beta^*}{T(+) \rightarrow T(\times)} \rightarrow -E \quad T(+)}{T(\otimes) \ \& \ T(\times)} \rightarrow -E}{T(\otimes) \ \& \ T(\times) \rightarrow T(\oplus)} \gamma^* \quad \frac{T(\otimes) \quad T(\times)}{T(\otimes) \ \& \ T(\times)} \&-I}{T(\oplus)} \rightarrow -E$$

Notice how assumptions in  $\mathcal{L}$  are modeled directly by assumptions in  $\mathcal{L}^*$ , and how, in general, a rule application in  $\mathcal{L}$  corresponds to a fragment of the derivation in  $\mathcal{L}^*$ . We have, for instance, the equivalence

$$\begin{array}{c} [ + ] \\ \vdots \\ \oplus \\ \frac{}{\oplus} \delta \\ \oplus \end{array} \equiv \frac{\frac{}{\quad} \delta^* \quad \frac{[T(+)]_1 \quad \vdots \quad T(\oplus)}{T(+)} \rightarrow -I_1}{\frac{}{\quad} \rightarrow -E} T(\oplus)$$

Intuitively, then, we use  $\rightarrow -E$  to unpack the rule, and  $\rightarrow -I$  to gather together the subproofs and discharged hypotheses.

We call the right-hand side of this the corresponding *characteristic fragment* for the rule  $\delta$ . In the same way there are characteristic fragments for each of the other rules, and out of these we can build a meta-derivation corresponding to any derivation in our original logic. Moreover, these characteristic fragments can be restricted to have a special form. Given a natural deduction rule  $\Theta$  then  $\Theta^*$  (in the general case) has the form

$$\begin{array}{c} (T(A_{1_1}) \& \dots \& T(A_{1_{m_1}}) \rightarrow T(A_1)) \& \dots \\ \& (T(A_{n_1}) \& \dots \& T(A_{n_{m_n}}) \rightarrow T(A_n)) \rightarrow T(A), \end{array}$$

and we can show:

**LEMMA 2.** *Given a rule  $\Theta$ , there is a characteristic fragment  $\Sigma$  where for  $1 \leq i \leq n$ , atomic assumptions  $T(A_{i_1}), \dots, T(A_{i_{m_i}})$  are discharged in subderivations of  $T(A_i)$ , and then these subderivations are combined together with  $\Theta^*$  to prove  $T(A)$ . Further, if we insist also that the major premise of an elimination rule is never the result of the application of an introduction rule (i.e.  $\rightarrow -I$  or  $\& -I$ ), while the minor premise is always either atomic or the result of the application of an introduction rule, then  $\Sigma$  is unique.*

**Proof.** By an analysis of the possible structure of  $\Theta$ . ■

For a rule  $\Theta$  that does not discharge assumptions, i.e., of the form

$$\frac{A_1 \dots A_n}{A},$$

the characteristic fragment is just a demonstration that, given  $\Theta^*$ , the rule

$$\frac{T(A_1) \dots T(A_n)}{T(A)}$$

is derivable in  $\mathcal{L}^*$ . That is, there is a derivation of  $T(A)$  where the only open assumptions are  $T(A_1), \dots, T(A_n)$ . Note that this standard notion of derivability (see, e.g., Troelstra [1982] or Hindley and Seldin [1986]), can be extended



(see Schroeder-Heister [1984b] and §3.6 below) to account for rules that discharge assumptions; in this extended sense, each characteristic fragment justifies a derived rule that allows us to simulate  $\mathcal{L}$ -derivations in  $\mathcal{L}^*$ .

Given the existence of characteristic fragments, it is a simple matter to model  $\mathcal{L}$ -derivations in  $\mathcal{L}^*$ . To begin with, since our metalogic is a natural deduction calculus, we can model assumptions in the encoded logic as assumptions in the metalogic. Each such assumption  $B_i$  is modeled by an assumption  $T(B_i)$ . Now, given a derivation in the object logic, we can inductively replace rules by corresponding characteristic fragments to produce a derivation in  $\mathcal{L}^*$ . For example, in (9) we proved  $\oplus$  using four rule applications; after we gave a proof in  $\mathcal{L}^*$  of  $T(\oplus)$  built from the four characteristic fragments that correspond to these applications.

In this form, however, this observation assumes not just a metalogic ( $\text{NJ}^{\rightarrow, \&}$ ) but also a particular proof calculus for the metalogic (natural deduction), which, since we want a purely logical characterization, we want to avoid. It is easy to remove this assumption by observing that  $A$  follows from the assumptions  $A_1$  to  $A_n$  in  $\text{NJ}^{\rightarrow, \&}$  iff  $A_1 \& \cdots \& A_n \rightarrow A$  follows without the assumptions. Thus we have a theorem that states the adequacy of encodings of natural deduction calculi in  $\text{NJ}^{\rightarrow, \&}$ .<sup>9</sup>

**THEOREM 3 (Adequacy).** *For any natural deduction calculus  $\mathcal{L}$  defined by a set of rules  $\mathcal{R}$ , if the formula  $A$  is derivable under assumptions  $A_1, \dots, A_n$ , then  $T(A_1) \& \cdots \& T(A_n) \rightarrow T(A)$  is provable in  $\mathcal{L}^* = \text{NJ}^{\rightarrow, \&} + \mathcal{R}^*$ .*

Notice that the proof is based on translation, and hence is constructive: given a derivation in the object logic we can construct one in the metalogic.

### 3.5 Faithfulness

In proving adequacy, we only used the metatheory  $\mathcal{L}^*$  in a limited way where derivations had a simple structure built by pasting characteristic fragments together. Of course, there are other ways to build proofs in  $\mathcal{L}^*$  and it could be that this freedom allows us to derive representations of formulae that are not derivable in  $\mathcal{L}$ . We now show that our translation is faithful, i.e. this is not the case.

To see why we have to be careful about faithfulness, consider the following: in the deductive system  $\text{NJ}^{\rightarrow, \&}$ , the  $\rightarrow$  connective is minimal implication (the logic is a conservative extension of  $\text{NJ}^{\rightarrow}$ ). But what happens if we strengthen the metalogic to be classical, e.g. by adding classical negation or assuming Peirce's law? We can still show adequacy of our encodings, since derivations in  $\text{NJ}^{\rightarrow, \&}$  remain valid when additional rules are present, but we can also use the encoding to derive formulae that are not derivable in the original system. Consider what happens if we try to prove something that is classically but not minimally valid,

<sup>9</sup>However notice that the translation is only defined on rules without side conditions, i.e. for 'pure' natural deduction; for more idiosyncratic logics, see the discussion in §6.

like, for instance, Peirce's law itself:

$$T(((A \supset B) \supset A) \supset A). \quad (10)$$

Using the axioms in (6), we can reduce this to the problem of proving

$$((T(A) \rightarrow T(B)) \rightarrow T(A)) \rightarrow T(A). \quad (11)$$

But this is itself an instance of Peirce's law, and is provable if  $\rightarrow$  is classical implication. Thus  $T(\cdot)$  does not here define the set of minimal logic theorems.

If  $\rightarrow$  is read as minimal implication, then the same trick does not work. We can still reduce (10) to (11), but we are not able to take the final step of appealing to Peirce's law in the metalogic.

We now show that assuming  $\rightarrow$  to be minimal really does lead to a faithful presentation of object logics, by demonstrating a direct relationship between derivations of formulae  $T(A)$  in the  $\mathcal{L}^*$  and derivations of  $A$  in  $\mathcal{L}$ .

The desired relationship is not a simple isomorphism: we pointed out in discussing adequacy above, that for each natural deduction rule translated into our logical language, it is possible to find a characteristic fragment, and using these fragments we can translate derivations in  $\mathcal{L}$  into derivations in  $\mathcal{L}^*$ . However an arbitrary derivation in  $\mathcal{L}^*$  may not have a corresponding derivation in  $\mathcal{L}$  (it is a simple exercise to construct a proof of  $T(\oplus)$  that does not use characteristic fragments). But if we look more carefully at the derivation of  $T(\oplus)$  we have given, and the fragments from which it is constructed, we can see that it has a particularly simple structure, and this structure has a technical characterization, similar to that we used for the characteristic fragments themselves. The derivations we build to show adequacy are in what is called *expanded normal form* (ENF): the major premise of an elimination rule (i.e.  $\rightarrow$ -E or  $\&$ -E) is never the result of the application of an introduction rule (i.e.  $\rightarrow$ -I or  $\&$ -I), and all minor premises are either atomic or the result of introduction rules. Not all derivations are in ENF, but any derivation can be transformed into one that is. We have the following:

**FACT 4** (Prawitz [1971]). There is an algorithm transforming any derivation in  $\text{NJ}^{\rightarrow, \&}$  into an equivalent derivation in ENF, and this equivalent derivation is unique.

From this we get the theorem we need; again, like for the statement of adequacy, we abstract away from the deductive system to get a pure logical characterization:

**THEOREM 5** (Faithfulness). *For any natural deduction system  $\mathcal{L}$  defined by a set of rules  $\mathcal{R}$ , if we can prove  $T(A_1) \& \dots \& T(A_n) \rightarrow T(A)$  in  $\mathcal{L}^*$ , then  $A$  is derivable in  $\mathcal{L}$  from the assumptions  $A_1, \dots, A_n$ .*

**Proof.** The theorem follows immediately from the existence of ENF derivations and Lemma 6 below. ■

LEMMA 6. *There is an effective transformation from ENF derivations in  $\mathcal{L}^*$  of  $T(A)$  from atomic assumptions  $\Delta$ , to natural deduction derivations in  $\mathcal{L}$  of  $A$  from assumptions  $\{B \mid T(B) \in \Delta\}$ .*

**Proof.** We prove this by induction on the structure of ENF derivations. Intuitively, we show that an ENF derivation is built out of characteristic fragments, and thus can easily be translated back into the original deductive system. Consider a derivation  $\Sigma$  of  $T(A)$ .

Base case:  $T(A)$  corresponds to either an assumption in  $\Delta$  or a (premissless) rule of the encoded theory. The translation then consists either of the assumption  $A$  or of the corresponding premissless rule with conclusion  $A$ .

Step case: Since the derivation is in ENF and the conclusion is atomic, the last rule applied is an elimination rule; more specifically, the derivation must have the form

$$\frac{\frac{\Phi_1^* \& \cdots \& \Phi_n^* \rightarrow T(A)}{\Phi_1^* \& \cdots \& \Phi_n^* \rightarrow T(A)} \Theta^* \quad \frac{\begin{array}{c} \vdots \Sigma_1 \quad \vdots \Sigma_n \\ \Phi_1^* \quad \cdots \quad \Phi_n^* \\ \hline \Phi_1^* \& \cdots \& \Phi_n^* \# \end{array}}{\Phi_1^* \& \cdots \& \Phi_n^* \rightarrow T(A)} \#}{T(A)} \rightarrow\text{-}E \quad (12)$$

where  $\#$  consists only of applications of  $\&$ -I and  $\Theta^*$  is

$$\Phi_1^* \& \cdots \& \Phi_n^* \rightarrow T(A)$$

for some rule  $\Theta$  of the encoded system.

By the definition of the encoding, each  $\Phi_i^*$ , derived by a proof  $\Sigma_i$ , is then of the form

$$T(A_{i_1}) \& \cdots \& T(A_{i_{m_i}}) \rightarrow T(A_i).$$

Since the conclusion of  $\Sigma_i$  proves an implication, and is in ENF, the last rule applied must be  $\rightarrow$ -I; thus  $\Sigma_i$  must have the form

$$\frac{\frac{\begin{array}{c} [T(A_{i_1})] \cdots [T(A_{i_{m_i}})] \\ \vdots \Sigma'_i \\ [T(A_{i_1}) \& \cdots \& T(A_{i_{m_i}})]_1 \end{array}}{T(A_i)} \# \quad T(A_i)}{T(A_{i_1}) \& \cdots \& T(A_{i_{m_i}}) \rightarrow T(A_i)} \rightarrow\text{-}I_1$$

with open assumptions  $\Delta$ , where  $\#$  consists only of applications of  $\&$ -E, which ‘unpack’ the assumption  $T(A_{i_1}) \& \cdots \& T(A_{i_{m_i}})$  into its component propositions. In other words, (12) corresponds to the characteristic fragment for  $\Theta$ .

By induction we can apply the same analysis to each  $\Sigma'_i$ , taking account not only of the undischarged assumptions  $\Delta$ , but also the new atomic assumptions  $\Delta + \{T(A_{i_1}), \dots, T(A_{i_{m_i}})\}$  which have been introduced, and we are finished.  $\blacksquare$

### 3.6 Generalizations of natural deduction

The distinction we mentioned above (in §3.1) that Gentzen and Prawitz make between natural deduction and sequent calculi has recently been emphasized by Schroeder-Heister [1984a; 1984b], who has proposed a ‘generalized natural deduction’. This extension, which follows directly from Gentzen’s appeal to intuition to justify natural deduction, is not formalizable using sequents in the way suggested in §3.<sup>10</sup>

Consider again Gentzen’s proposed ‘natural language’ analysis of logical connectives. In these terms we can characterize Implication Elimination as

If  $A \supset B$  and  $A$ , then it follows that  $B$

which we formalize as:

$$\frac{A \supset B \quad A}{B}$$

Now consider the slightly more convoluted

If  $A \supset B$  and assuming that if, given that from  $A$  we could derive  $B$ , we could derive  $C$ , then we can derive  $C$ .

which is also true. There is a natural generalization of rules that allows us to express this in the spirit of natural deduction; namely,

$$\frac{\begin{array}{c} \left[ \begin{array}{c} A \\ \hline B \end{array} \right] \\ \vdots \\ A \supset B \quad C \end{array}}{C} \quad (13)$$

where we can assume rules as hypotheses (and by generalization, have rules as hypotheses of already hypothetical rules, and so on). Schroeder-Heister shows that it is quite a simple matter to extend the formula-tree/discharge-function formalization to cope with this extension: we simply allow discharge functions to point to subtrees rather than just leaves.

Why is such a generalization interesting? Schroeder-Heister uses it to analyze systematically the intuitionistic logical connectives. However it has more general use, and at least one immediate application: we can use it to encode Peirce’s law without introducing new connectives, as

$$\frac{\begin{array}{c} \left[ \begin{array}{c} A \\ \hline B \end{array} \right] \\ \vdots \\ A \end{array}}{A}$$

<sup>10</sup>Though see Avron [1990] for relevant discussion.

which, in English, is equivalent to

if, by assuming that some  $B$  follows from  $A$ , it follows that  $A$ , then it follows that  $A$ .

This provides, finally, a self-contained generalized natural deduction formulation of classical implication. (At least in the sense that the meaning of implication can be defined by rules involving no auxiliary connectives.)

The process of generalization can obviously be continued beyond allowing rules as hypotheses, though. There is no reason why we cannot also have them as conclusions. Thus, for instance, it is clearly true that

$$\frac{A \supset B}{\left( \frac{A}{B} \right)} \rightarrow\text{-}E_C$$

(where round brackets do not, in the manner of square brackets, represent the discharge of an assumption, but simply grouping), which might reasonably be read as:

If  $A \supset B$  then from  $A$  it follows that  $B$ .

By now, however, while our intuition still seems to be functioning soundly, the formula-tree/discharge-function formalization is beginning to break down. It can cope with rules as assumptions, but the more general treatment of both rules and formulae in derivations that we are now proposing is more difficult. With our proposed alternative of the metalogic  $\text{NJ}^{\rightarrow, \&}$ , on the other hand, the same problems do not arise. In fact one way of looking at the faithfulness argument developed above is as essentially a demonstration that this sort of ‘confusion’ can be unwound for the purpose of recovering a proof in a system that does not in the end allow it. The question then is, how closely does our logic match the traditional formulation?

In the most general case, allowing rules as both hypotheses and conclusions, such a question is not quite meaningful, since we do not have a traditional formulation against which we can compare. However if we limit ourselves to the restricted case where we have rules as assumptions, it is still possible to be properly formal, since we can still compare our encoding to Schroeder-Heister’s more traditional formalization in terms of formula-trees and discharge-functions.

Such a formulation is enough to allow us, by a ‘natural’ generalization of the encoding in (7), to formalize, for instance, (13) as

$$(T(A \supset B) \& ((T(A) \rightarrow T(B)) \rightarrow T(C))) \rightarrow T(C)$$

and  $\rightarrow\text{-}E_C$  as

$$T(A \supset B) \rightarrow T(A) \rightarrow T(B).$$

It is now possible, by a corresponding generalization of the notion of deduction (see Schroeder-Heister [1984b]) to show adequacy and faithfulness for this larger class of deductive systems.

*Generalized natural deduction and curried rules*

Until now we have been using  $\text{NJ}^{\rightarrow, \&}$  to encode and reason with deductive systems. If we want to show a direct relationship between the traditional notation of natural deduction and our encoding, then we seem to need both the  $\rightarrow$  and  $\&$  connectives in the metalogic. However, compare  $\rightarrow\text{-}E$  and  $\rightarrow\text{-}E_C$ ; while there seems to be large differences between the two as rules, there is very little either in natural language or in logical framework terms: one is simply the ‘curried’ form of the other; in this case the conjunction seems almost redundant.

On the other hand, one generalization that we have not made so far, and one that is suggested by  $\text{NJ}^{\rightarrow, \&}$ , is to allow a natural deduction rule with multiple conclusions; e.g. assuming an ordinary conjunction  $\wedge$ , have a (tableau-like) rule

$$\frac{A \wedge B}{A, B} \&\text{-}E_{MC}$$

corresponding to the natural language

If  $A \wedge B$  then it follows that  $A$  and  $B$ .

Both of which correspond to the framework formalization

$$T(A \wedge B) \rightarrow T(A) \& T(B).$$

However, there seems to be less of a need for rules of this form: a single encoded rule  $C \rightarrow A \& B$  can be replaced by a pair  $C \rightarrow A$  and  $C \rightarrow B$ .

If we are willing to accept this restriction to the language of  $\rightarrow$  alone, then we can simplify the metalogic we are using: if we have conjunctions only in the antecedents of implications  $\rightarrow$ , they can always be eliminated by ‘currying’ the formulae conjoined in the antecedent, allowing us to dispose of conjunction completely. For example, the curried form of the axiom for  $\rightarrow\text{-}E$  in (6) is then

$$T(A \supset B) \rightarrow T(A) \rightarrow T(B).$$

Notice that this form of  $\rightarrow\text{-}E$  is indistinguishable from the encoding we give above for  $\rightarrow\text{-}E_C$ : we are no longer able to distinguish some different rules in generalized natural deduction, and thus we lose the faithfulness/adequacy bijection that we have previously demonstrated. However this problem is not serious: we are only identifying proofs that differ in simple ways, e.g., by the application of curried versus uncurried rules. Furthermore, this possible confusion arises in the case of *generalized* natural deduction, but not in the traditional form.

In the next section we describe a full logical framework based on the ideas we have developed here. In fact it turns out that we can adopt the same notation to formalize not only deductive systems, but also languages, in a uniform manner.

## 4 FRAMEWORKS BASED ON TYPE-THEORIES

In the last section we established a relationship between natural deduction and the logic of implication. However we considered only reasoning about fragments of propositional logic, and when we turn to predicate logics, we find that the mechanisms of binding and substitution introduce some entirely new problems for us to solve.

The first problem is simply how to encode languages where operators bind variables. Such variable binding operators include standard logical quantifiers, the  $\lambda$  of the  $\lambda$ -calculus, fixedpoint operators like  $\mu$  in fixedpoint logics, etc. Until now we have been using a simple term algebra to represent syntax, where, e.g. a binary connective like implication is represented by a binary function. However, with the introduction of binding and substitution this approach is less satisfactory. For instance  $\forall x. \phi(x)$  and  $\forall y. \phi(y)$  are distinct syntactically, but not in terms of the deductive system (any proof of one proves the other). Binding operators also complicate operations performed on syntax; e.g. substitution. The second problem is that proof-rules become more complex: the rules for quantifiers place conditions on the contexts (e.g. insisting that certain variables do not appear free) in which they can be applied.

Now we extend our investigation to deal with these problems, and complete the development of a practical  $\rightarrow$ -framework. We tackle the two problems of language encoding and rule encoding together, by introducing the  $\lambda$ -calculus as a representation language into our system. This provides us with a way to encode quantifiers using *higher-order* syntax and then to encode rules for these quantifiers.

There are different ways that we can combine the  $\lambda$ -calculus with a metalogic. One possibility is simply to add it to the term language, extending, e.g., the theory  $\text{NJ}^{\rightarrow}$  to a fragment of higher-order logic.<sup>11</sup> Another, similar, possibility, and one which offers some theoretical advantages, is to use a *type-theory*. We investigate the type-theoretic approach in this section. Type-theories based on the  $\lambda$ -calculus are well-known to be closely related to intuitionistic logics like  $\text{NJ}^{\rightarrow}$  via ‘Curry-Howard’ (see Howard [1980]), or *propositions-as-types*, isomorphisms, a fact which allows us to carry across much of what we already know about encoding deductive systems from  $\text{NJ}^{\rightarrow}$ . Moreover, an expressive enough type-theory provides a unified language for representing not just syntax, but also proof-rules and proofs. Thus a type-theoretic logical framework can provide a single solution to the apparently distinct problems of encoding languages and deductive systems: The encoding problems are reduced to declaration of appropriate (higher-order) signatures and the checking problems (e.g. well-formedness of syntax and the correctness of derivations) to the problem of type checking against these signatures.

---

<sup>11</sup> See, e.g., Felty [1989], Paulson [1994] or Simpson [1992], for examples of this approach.

#### 4.1 Some initial observations

We begin by considering the problem of formalizing a binding operator, taking  $\forall$  as our example. Consider what happens if we follow through the analysis given above for natural deduction in propositional logic. First we state, in Gentzen style natural language, some of the properties of  $\forall$ ; e.g.

if, for arbitrary  $t$ , it follows that  $\phi(t)$ , then it follows that  $\forall x. \phi(x)$

and

if it follows that  $\forall x. \phi(x)$ , then for any  $t$  it follows that  $\phi(t)$ .

These are traditionally represented, in the notation of natural deduction, as

$$\frac{\phi}{\forall x. \phi} \forall\text{-I} \quad \text{and} \quad \frac{\forall x. \phi}{\phi[x \leftarrow t]} \forall\text{-E}$$

where, in  $\forall\text{-I}$ , the variable  $x$  does not appear free in any undischarged assumption, and the notation  $\phi[x \leftarrow t]$  denotes the formula  $\phi$  where the term  $t$  has been substituted through for  $x$  (care being taken to avoid capturing variables). The relationship between these rules and their informal characterizations is less direct here than in the propositional case; for example, we model the statement ‘If, for arbitrary  $t$ , it follows that  $\phi(t)$ ’ indirectly, by assuming that an arbitrary variable  $x$  can stand for an arbitrary term, then ensuring that  $x$  really is arbitrary by requiring that it does not occur free in the current assumptions.

Consider how we might use a ‘logical’ language instead of rules by extending our language based on minimal implication. To start with, we need a way of saying that a term is arbitrary, which we can accomplish with universal quantification. Furthermore, unlike in the propositional case, we have two syntactic categories. As well as formulae ( $fm$ ), we now have terms ( $tm$ ), and we will use types to formally distinguish between them. If we combine quantification with typing, by writing  $(x:y)$  to mean ‘for all  $x$  of type  $y$ ’, then a first attempt at translating natural language into (semi)formal language results in the following:

$$\begin{aligned} T(\forall x. \phi(x)) &\rightarrow (t:tm)T(\phi(t)) \\ (t:tm)T(\phi(t)) &\rightarrow T(\forall x. \phi(x)) \end{aligned} \tag{14}$$

However while this appears to capture our intuitions, it is not clear that it is formally meaningful. If nothing else, one might question the cavalier way we have treated the distinction between object-level and metalevel languages. In the following sections we show that this translation does in fact properly correspond to the intuitive reading we have suggested.

#### 4.2 Syntax as typed terms

We begin by considering how languages that include variable-binding operators can be represented. We want a method of representing syntax where terms in the



object logic are represented by terms in the metalogic. This representation should be computable and we want too that it is compositional; i.e. that the representative of a formula is built directly from the representatives of its subformulae.

Types provide a starting point for solving these problems. A type system can be used to classify terms into types where well-typed terms correspond to well-formed syntax. Consider the example of first-order logic; this has two syntactic categories, terms and formulae, and, independent of whether we view syntax as strings, trees, or something more abstract, we must keep track of the categories to which subexpressions belong. One way to do this is to tag expressions with their categories and have rules for propagating these tags to ensure that entire expressions are ‘well-tagged’. Even if there is only one syntactic category we still need some notion of well-formed syntax; in minimal logic, for instance, some strings built from implication and variables, such as  $\phi \supset \psi$ , are well-formed formulae, while others, like  $\phi \supset \supset \psi$ , are not.

In a typed setting, we can reduce the problem of syntactic well-formedness to the problem of well-typedness by viewing syntax as a typed term algebra: we associate a type of data with each syntactic category and regard operators over syntax as typed functions. For instance,  $\supset$  corresponds to a function that builds a formula given two formulae, i.e., a function of type  $fm \times fm \rightarrow fm$ . Under this reading, and using infix notation,  $\phi \supset \psi$  is a well-typed formula provided that  $\phi$  and  $\psi$  are both well-typed formulae of type  $fm$ , whereas  $\phi \supset \supset \psi$  is ill-typed.

This view of syntax as typed terms provides a formalization of the treatment of propositional languages as term algebras that we have informally adopted up to now. We will see that it also provides a basis for formalizing languages with quantification. In fact the mechanism by which this is done is so flexible that it is possible to claim:

THEESIS 7. The typed  $\lambda$ -calculus provides a logical basis for representing many common kinds of logical syntax.

Note that we do not claim that the typed  $\lambda$ -calculus, as we shall use it, is a universal solution applicable to formalizing any language. We cannot, for instance, formalize a syntax based on a non-context-free grammar. Neither can we give a finite signature for languages that require infinitely many (or parameterized) sets of productions. The notation is remarkably flexible nevertheless, and, especially if we refine the type system a little further, can deal with a great many subtle problems — a good example of this can be found in the analysis of the syntax of higher-order logic developed by Harper *et al.* [1993].

### *The simply typed $\lambda$ -calculus*

We assume that the reader is familiar with the basics of the  $\lambda$ -calculus, e.g. reduction (we denote one-step reduction by  $\rightarrow_\beta$  and its reflexive-transitive closure by  $\rightarrow_\beta^*$ ) and conversion ( $=_\beta$ ), and review only syntax and typing here; further details can be found in Barendregt [1984; 1991] or Hindley and Seldin [1986]. We now

define a type theory based on this, called the *simply typed*  $\lambda$ -calculus, or more succinctly  $\lambda^\rightarrow$ . Our development is based loosely on that of Barendregt and anticipates extensions we develop later.

We start by defining the syntax we use:

**DEFINITION 8.** Fix a denumerable set of variables  $\mathcal{V}$ . The terms and types of  $\lambda^\rightarrow$  are defined relative to a signature consisting of a non-empty set of *base types*  $\mathcal{B}$ , a set of *constants*  $\mathcal{K}$ , and a *constant signature*  $\Sigma$ , which is a set  $\{c_i:A_i \mid 1 \leq i \leq n, c_i \in \mathcal{K}, A_i \in \text{Ty}\}$ , where the  $c_i$  are distinct, and:

- The set of *types*  $\text{Ty}$  is given by

$$\text{Ty} ::= \mathcal{B} \mid \text{Ty} \rightarrow \text{Ty}.$$

- The set of *terms*  $\mathcal{T}$  is given by

$$\mathcal{T} ::= \mathcal{V} \mid \mathcal{K} \mid \mathcal{T}\mathcal{T} \mid \lambda\mathcal{V}^{\text{Ty}}.\mathcal{T}.$$

A variable  $x$  occurs *bound* in a term if it is in the scope of a  $\lambda x$  and  $x$  occurs *free* otherwise. We implicitly identify terms that are identical under a renaming of bound variables.

- A *typing context* is a sequence  $x_1:A_1, \dots, x_n:A_n$  of *bindings* where the  $x_i$  are distinct variables and  $A_i \in \text{Ty}$  for  $(1 \leq i \leq n)$ .

For convenience, we shall overload the  $\in$  relation, extending it from sets to sequences in the obvious manner (i.e. so that  $x_i:A_i \in x_1:A_1, \dots, x_n:A_n$  iff  $1 \leq i \leq n$ ).

- A *type assignment relation*  $\vdash_\Sigma$  is a binary relation, indexed by  $\Sigma$ , defined between typing contexts  $\Gamma$  and typing assertions  $M:A$  where  $M \in \mathcal{T}$  and  $A \in \text{Ty}$ , by the inference rules:

$$\begin{array}{c} \frac{c:A \in \Sigma}{\Gamma \vdash_\Sigma c:A} \text{assum} \qquad \frac{\Gamma, x:A \vdash_\Sigma M:B}{\Gamma \vdash_\Sigma (\lambda x^A. M):(A \rightarrow B)} \text{abst} \\ \frac{x:A \in \Gamma}{\Gamma \vdash_\Sigma x:A} \text{hyp} \qquad \frac{\Gamma \vdash_\Sigma M:A \rightarrow B \quad \Gamma \vdash_\Sigma N:A}{\Gamma \vdash_\Sigma (MN):B} \text{appl} \end{array}$$

This definition states that types consist of the closure of a set of base types under the connective  $\rightarrow$  and that terms are either variables, constants, applications or (typed)  $\lambda$ -abstractions. The rules constitute a deductive system that defines when a term is well-typed relative to a context and a signature, which in turn assign types to free variables and constants. Notice that we do not have to make explicit the standard side condition for *abst* that the variable  $x$  does not appear free in  $\Gamma$ , since this is already implicitly enforced by the requirement that variables in a type context must be distinct.

As an example, if  $\{A, B\} \subseteq \text{Ty}$  then

$$\frac{\frac{\frac{x:A \in x:A, y:B}{x:A, y:B \vdash_{\Sigma} x:A} \text{assum}}{x:A \vdash_{\Sigma} \lambda y^B. x:B \rightarrow A} \text{abst}}{\vdash_{\Sigma} \lambda x^A. \lambda y^B. x:A \rightarrow (B \rightarrow A)} \text{abst} \quad (15)$$

is a proof (for any  $\Sigma$ ). Further, it is easy to see that:

FACT 9. Provability for this system (and thus well-typing) is decidable.

*A Curry-Howard isomorphism for  $\lambda^{\rightarrow}$*

The rules for  $\lambda^{\rightarrow}$  suggest the natural deduction presentation for  $\text{NJ}^{\rightarrow}$  (see §3.1): if we rewrite all instances of  $\Gamma \vdash_{\Sigma} M:A$  in a proof in  $\lambda^{\rightarrow}$  as  $\Gamma, \Sigma \vdash M:A$  and uniformly replace each typing assertion  $c:A$ ,  $x:A$ , and  $M:A$  by the type  $A$ , then *abst* and *appl* correspond to  $\rightarrow\text{-I}$  and  $\rightarrow\text{-E}$ , while *hyp* and *assum* together correspond to *Basic*. The following makes this correspondence more precise.

FACT 10.

1. There is a bijection between types in  $\lambda^{\rightarrow}$  and propositions in  $\text{NJ}^{\rightarrow}$  where a proposition in  $\text{NJ}^{\rightarrow}$  is provable precisely when the corresponding  $\lambda^{\rightarrow}$  type is inhabited (has a member).
2. There is a bijection between members of types in  $\lambda^{\rightarrow}$  and proofs of the corresponding propositions in  $\text{NJ}^{\rightarrow}$ .

Part 1 of this characterizes the direct syntactic bijection between types and propositions where base types correspond to atomic propositions<sup>12</sup> and the function space constructor corresponds to the implication connective. The correspondence between provability and inhabitation then follows from the correspondence between the proof-rules of the two systems. Then part 2 refines the bijection to the level of inhabiting terms on the one hand, and proofs on the other.

Fact 10 states an isomorphism between types and propositions that we can characterize as ‘truth is inhabitation’: if  $M:A$  then the proposition corresponding to  $A$  is provable and, moreover, there is a corresponding notion of reduction in the two settings where  $\beta$ -reduction of  $M$  in the  $\lambda$ -calculus corresponds to reduction of the proof  $A$  (in the sense of Prawitz).

We exploit this isomorphism in reasoning about encodings in this chapter. In this section we show the correctness of encodings by reasoning about normal forms of terms in type-theory and in §5 we will reason about normal forms of proofs.

<sup>12</sup>This fact holds for the pure version of  $\lambda^{\rightarrow}$  without constants. We also implicitly assume a bijection between base types and propositional variables.

REMARK 11. For reasoning about the correctness of encodings it is sometimes necessary to use a more complex isomorphism based on both  $\beta$  and  $\eta$ -conversion. It is possible to establish a correspondence between so called long  $\beta\eta$ -normal forms<sup>13</sup> and corresponding classes of derivations in  $\text{NJ}^{\rightarrow}$ . More details can be found in [Harper *et al.*, 1993].

### Representation

In order to represent syntax in  $\lambda^{\rightarrow}$  we define a suitable signature  $\Sigma$  and establish a correspondence between terms in the metalogic and syntax in the object logic. Our signature will consist of a base type for each syntactic category of the object logic and a constant for each constructor in the object logic. We will see that we do not need a type *variable* to formalize variables in the object logic because we can use instead variables of the metalogic. Syntax belonging to a given category in the object logic then corresponds to terms in the metalogic belonging to the type of that category.

This is best illustrated with a simple example: we represent the syntax of minimal logic itself. We follow Church's [1940] convention, where  $o$  is the type of propositions, so the signature is the (singleton) set of base types  $\{o\}$ , and the constant signature is

$$\Sigma \equiv \{imp: o \rightarrow o \rightarrow o\}. \quad (16)$$

The constructor  $imp$  builds a proposition from two propositions. With respect to this signature,  $imp\ x\ (imp\ y\ z)$  is a term that belongs to  $o$  when  $x$ ,  $y$ , and  $z$  belong to  $o$ ; i.e.  $x:o, y:o, z:o \vdash imp\ x\ (imp\ y\ z):o$ . This corresponds to the fact that  $x \supset (y \supset z)$  is a proposition of minimal logic provided that  $x$ ,  $y$ , and  $z$  are propositions.

Thus formulae of minimal logic correspond to well-typed terms of type  $o$ . Formulating this correspondence requires some care though: we have to check adequacy and faithfulness for the represented syntax; this amounts to showing that (i) every formula in minimal logic is represented by a term of type  $o$ , and (ii) that every term of type  $o$  represents a formula in minimal logic.

As a first attempt to establish such a correspondence, consider the following mapping  $\ulcorner \cdot \urcorner$ , from terms in minimal logic to  $\lambda$ -calculus terms:

$$\begin{aligned} \ulcorner x \urcorner &= x \\ \ulcorner t_1 \supset t_2 \urcorner &= imp\ \ulcorner t_1 \urcorner\ \ulcorner t_2 \urcorner \end{aligned}$$

As an example, under this representation the formula  $x \supset (y \supset z)$  corresponds to the term  $imp\ x\ (imp\ y\ z)$ . The representation function is an injection from propositions to terms of type  $o$ , provided that variables are declared to be of type

<sup>13</sup>A term  $M = \lambda x_1 \dots x_n. x M_1 \dots M_m$  is in long  $\beta\eta$ -normal form when 1)  $x$  is an  $x_i$ , a constant, or a free variable of  $M$ , 2)  $x M_1 \dots M_n$  is of base type (i.e.  $x$  is applied fully to all possible arguments) and 3) each  $M_i$  is in long  $\beta\eta$ -normal form.

$o$  in the context. However, it is not surjective: there are too many terms of type  $o$ . For example in a context where  $x$  is of type  $o \rightarrow o$  and  $y$  is of type  $o$ , then  $x y$  is of type  $o$ , and so is  $\text{imp } ((\lambda z^o. z) y) y$ ; but neither of these is in the image of  $\ulcorner \cdot \urcorner$ . The problem with the first example is that the variable  $x$  is of higher type (i.e. the function type  $o \rightarrow o$ ) and not a base type in  $\mathcal{B}$ . The problem with the second example is that it is not in normal form. Any such term, however, has a unique equivalent  $\beta$ -normal form, which we can compute. In the above example the term has the normal form  $\text{imp } y y$ , which is  $\ulcorner y \supset y \urcorner$ . Our correspondence is thus a bijection when we exclude the cases above: we consider only  $\beta$ -normal form terms where free variables are of type  $o$ .

**THEOREM 12.**  $\ulcorner \cdot \urcorner$  is a bijection between propositions in minimal logic with propositional variables  $x_1, \dots, x_n$ , and  $\beta$ -normal form terms of type  $o$  containing only free variables  $x_1, \dots, x_n$ , all of type  $o$ .

### First-order syntax

The syntax of minimal logic is very simple, and we do not need all of  $\lambda^\rightarrow$  to encode it. We can be precise about what we mean by ‘not all’ if we associate types with orders: observe that any type  $\gamma$  has a unique representation as  $\alpha_1 \rightarrow \dots \rightarrow \alpha_n \rightarrow \beta$ , where  $\rightarrow$  associates to the right and  $\beta$  is a base type. Now define the *order* of  $\gamma$  to be 0 if  $n = 0$ , and  $1 + \max(\text{Ord}(\alpha_1), \dots, \text{Ord}(\alpha_n))$  otherwise.<sup>14</sup> In our encoding of minimal propositional logic we have used only the first-order fragment of  $\lambda^\rightarrow$ ; i.e. variables are restricted to the base type  $o$  and the only function constant  $\text{imp}$  is first-order (since its two arguments are of base type); this is another way of saying that an encoding using a simple term algebra is enough.

This raises an obvious question: why adopt a full higher-order notation if a simple first-order notation is enough? Indeed, in a first-order setting, results like Theorem 12 are much easier to prove because there are no complications introduced by reduction and normal forms. The answer is that the situation changes when we introduce quantifiers and other variable binding operators. A ‘naive’ encoding of syntax is still possible but is much more complicated.

Consider, as an example, the syntax of first-order arithmetic. This is defined in terms of two syntactic categories, terms and formulae, which are usually specified as:

$$\begin{aligned} \text{terms } T &::= x \mid 0 \mid sT \mid T + T \mid T \times T \\ \text{formulae } F &::= T = T \mid \neg F \mid F \wedge F \mid F \vee F \mid F \supset F \mid \forall x. F \mid \exists x. F \end{aligned}$$

How should we represent this? A possible first-order notation is as follows: we define a base type  $v$  of variables, in addition to types  $i$  for terms (i.e. ‘individuals’) and  $o$  for formulae. Since variables are also terms, our signature requires a ‘coercion’ function mapping elements of type  $v$  to elements of type  $i$ . The

<sup>14</sup>Note that there is not complete agreement in the literature about the order of base types, which is sometimes defined to be 1, not 0.

rest of the signature is formalized as we would expect; e.g. *plus* is a constant of type  $i \rightarrow i \rightarrow i$ , atomic formulae are built from the equality relation *eq* of type  $i \rightarrow i \rightarrow o$ , connectives are defined as propositional functions over  $o$ , and a quantifier like  $\forall$  is formalized by declaring a function constant *all* of type  $v \rightarrow o \rightarrow o$  taking a variable and a formula to a formula.

This part of the encoding presents no difficulties. However the problems with first-order representations of a language with binding are not directly in the representation, but rather in the way we will use that representation. It is in the formalization of proof-rules where we encounter the problems, in particular with substitution. Consider, for instance,  $\forall$ -*E* in (4.1), where we use the notation  $\phi[x \leftarrow t]$ ; we need to formalize an analogue for our encoding, which we can do by introducing a ternary function, *sub* of type  $o \rightarrow i \rightarrow v \rightarrow o$ , where  $sub(\phi, t, x) = \psi$  is provable precisely when  $\phi[x \leftarrow t] = \psi$ . With this addition  $\forall$ -*E* is axiomatizable as

$$\forall \phi, \psi: o. \forall t: i. \forall x: v. (sub(\phi, t, x) = \psi) \rightarrow T(all\ x\ \phi) \rightarrow T(\psi). \quad (17)$$

There are several ways we might axiomatize the details of *sub*. The most direct approach is simply to formalize the standard textbook account of basic concepts such as free and bound variables, capture, and equivalence under bound variable renaming, which we can easily do by structural recursion on terms and formulae. The definitions are well-known, although complex enough that some care is required, e.g. bound variables must sometimes be renamed to avoid capture. Note too that in (17) we have used the equality predicate over formulae (not to be confused with equality over terms in the object logic, i.e. *eq*), which must either be provided by the metalogic or additionally formalized. Examples of such equational encodings of logics are given by Martí-Oliet and Meseguer [2002].

Other encodings have also been explored: for instance we can use a representation of terms that finesses problems involving bound variable names by eliminating them entirely. Such an approach was originally suggested by de Bruijn [1972] who represents terms with bound variables by replacing occurrences of such variables with ‘pointers’ to the operators that bind them. A related approach has been proposed by Talcott [1993], who has axiomatized a general theory of binding structure and substitution, which can be used as the basis for encoding logics that require such facilities. Another recent development, which is gaining some popularity, is to provide extensions of the  $\lambda$ -calculus that formalize operators for ‘explicit substitutions’ [Abadi *et al.*, 1991].

In all of these approaches, direct or indirect, there is considerable overhead, and not only at the formalization stage itself: when we are building proofs, we have to construct subproofs using the axiomatization of *sub* at every application of  $\forall$ -*E*. Being forced to take such micro-steps in the metatheory simply to apply a rule in the object logic is both awkward and tedious. And there is another serious problem that appears when we consider a rule like  $\forall$ -*I* in (4.1), where we have a side condition that *the variable x does not appear free in any undischarged*

*assumption.* There is no direct way, in the language we have been developing, to complete the formalization of this: we cannot reify the semiformal

$$\forall\phi:o.\forall x:v. x \text{ not free in context} \rightarrow T(\phi) \rightarrow T(\text{all } x \phi) \quad (18)$$

into a formal statement (we cannot, in general, refer to contexts in this way).

Again there are ways to get around the problem by complicating the formalization. For instance while we cannot tell which variables are free in the context, we can, under certain circumstances, keep track of those that might be. We can define a new type  $sv$ , of sets of variables, with associated algebra and predicates, where, e.g.,  $\text{notin}(x, c)$  is a predicate on types  $v$  and  $sv$  that is true iff  $x$  does not occur in the set  $c$ , and  $\text{union}$  is a function of type  $sv \rightarrow sv \rightarrow sv$  that returns the union of its arguments. In this setting we can then expand  $T(\cdot)$  so that rather than being a predicate on  $o$ , it is a predicate on  $o$  and  $sv$ ; this yields the formalization

$$\forall\phi:o. \forall x:v. \forall c:sv. \text{notin}(x, c) \rightarrow T(\phi, c) \rightarrow T(\text{all } x \phi, c).$$

In addition, we have, of course, to modify all the other rules so they keep track of the variables that might be added to the context; for instance  $\rightarrow-I$  now has the formalization

$$\forall\phi, \psi:o. \forall c:sv. (T(\phi, c) \rightarrow T(\psi, \text{union}(c, \text{fv}(\phi)))) \rightarrow T(\text{imp } \phi \psi, c),$$

where  $\text{fv}$  returns the set of variables free in a formula (i.e. for  $\psi$  we have added all the free variables of  $\phi$  to the free variables in the context).

Clearly, first-order syntax in combination with  $\rightarrow$  is becoming unacceptably complicated at this point, and is far removed from the ‘sketched’ characterization in (14). If part of the motivation of a logical framework is to provide a high-level abstraction of a deductive system, which we can then use to implement particular systems, then, since substitution and binding are standard requirements, we might expect them to be built into the framework, rather than encoded from scratch each time we need them. We thus now examine a very different sort of encoding in terms of  $\lambda^\rightarrow$  that does precisely this.

### *Higher-order syntax*

When using a first-order encoding of syntax, each time we apply a proof-rule like  $\forall-E$ , we must construct a subproof about substitution. But in  $\lambda^\rightarrow$  we already have a substitution mechanism available that we can exploit if we formalize variable binding operators a bit differently. The formalization is based on what is now commonly called *higher-order* syntax.

We start by observing that in  $\lambda^\rightarrow$  the  $\lambda$  operator binds variables,  $\beta$ -reduction provides the sort of substitution we want, and we have a built-in equivalence that accounts for renaming of bound variables. Higher-order syntax is a way of exploiting this, using higher-order functions to formalize the variable binding operators of

an encoded logic directly, avoiding the complications associated with a first-order encoding.

The idea is best illustrated with an example. If we return to the problem of how to encode the language of arithmetic, then, using higher-order syntax our signature need contain only the two sorts, for terms and formulae; i.e.  $\mathcal{B} = \{i, o\}$ .

We do not need a sort corresponding to a syntactic category of variables, because now we will represent them directly using variables of the metalogic itself, which are either declared in the context or bound by  $\lambda$  in the metalogic. The signature  $\Sigma$  is then

$$\begin{aligned} \{0:i, s:i \rightarrow i, plus:i \rightarrow i \rightarrow i, times:i \rightarrow i \rightarrow i, eq:i \rightarrow i \rightarrow o, \\ falsum:o, neg:o \rightarrow o, or:o \rightarrow o \rightarrow o, and:o \rightarrow o \rightarrow o, \\ imp:o \rightarrow o \rightarrow o, all:(i \rightarrow o) \rightarrow o, exists:(i \rightarrow o) \rightarrow o\}. \end{aligned} \quad (19)$$

In this signature *all* and *exists* no longer have the (first-order) type  $v \rightarrow o \rightarrow o$ ; instead they are second order, taking as their arguments predicate valued functions (which have first-order types).

Using higher-order syntax, an operator that binds a variable can be conceptually decomposed into two parts. First, if  $\phi$  is an encoded formula, i.e. of type  $o$ , possibly including a free metavariable  $x$  of type  $i$ , then  $\lambda x^i. \phi$  is the abstraction of  $\phi$  over  $x$ , where  $x$  is now bound. The result is, however, of type  $i \rightarrow o$ , not of type  $o$ . Second, we convert  $\lambda x^i. \phi$  back into an object of type  $o$ , and indicate the variable binding operator we want, by applying that operator to it. For example, applying *all* to  $\lambda x^i. \phi$  yields the term *all* ( $\lambda x^i. \phi$ ) of type  $o$ . Similarly, for a substitution we reverse the procedure. Given *all* ( $\lambda x^i. \phi$ ), for which we want to generate the substitution instance  $\phi[x \leftarrow t]$ , we first strip off the operator *all* and apply (in  $\lambda^\rightarrow$ ) the result to  $t$ , to get  $(\lambda x^i. \phi)t$ . But in  $\lambda^\rightarrow$  this reduces to  $\phi[x \leftarrow t]$ . Hence, we needn't formalize explicitly any substitution mechanism for the object logic since we can exploit the substitution that is (already) formalized for the metalogic.

Of course we must check that all of this works. But it is easy to extend adequacy (Theorem 12) for this signature to show that the terms and formulae of first-order arithmetic are correctly represented by normal form members of types  $i$  and  $o$  respectively.

Now we can formalize  $\forall$ -*E* and  $\forall$ -*I* in a way that directly reflects the sketch in (14):

$$\begin{aligned} \forall \phi^{i \rightarrow o}. (T(all \phi) \rightarrow \forall x^i. T(\phi x)) \\ \forall \phi^{i \rightarrow o}. ((\forall x^i. T(\phi x)) \rightarrow T(all \phi)) \end{aligned} \quad (20)$$

If we compare this with (18) above, we can see how, by using metavariables as object variables, we are able to formalize the side condition ‘ $x$  not free in context’ in (20) by having  $x$  bound directly by a universal quantifier at the metalevel.

In conclusion, higher-order syntax provides strong supporting evidence for Thesis 7 by providing a mechanism for using the  $\lambda$  of the metatheory to provide directly the machinery needed for variable binding, substitution, variable renaming,



and the like, which are typically needed for representing and using object logics that contain variable binding operators.

### 4.3 Rules of proof and dependent types

We have shown how we can use a type-theory to represent syntax, reducing the problem of syntactic well-formedness to the problem of decidable well-typing. We now extend the language we are developing so that we can do the same with proof-rules.

We can use  $\lambda^\rightarrow$  as a representation language for proofs too, but it is too weak to reduce proof checking to type checking alone. To see why, consider the two function symbols *times* and *plus* in the signature defined in (19). Both are of type  $i \rightarrow i \rightarrow i$ , which means that as far as the typing enforced by  $\lambda^\rightarrow$  is concerned, they are interchangeable; i.e., if  $t$  is a well-typed term in  $\lambda^\rightarrow$ , and we replace every occurrence of the constant *times* with the constant *plus*, we get a well-typed term  $t'$ . If  $t$  is supposed to represent a piece of syntax, this is what we want; for instance if we have used the type-checking of  $\lambda^\rightarrow$  to show that  $t \equiv eq(times\ 0\ (s0))\ 0$  is a well-formed formula, i.e. that  $t:o$ , then we immediately know that  $t'$  is a well-formed formula too. Unfortunately, what is useful for encoding syntax makes it impossible to define a type of proofs: in arithmetic we want  $t$ , but not  $t'$ , to be provable, but we cannot make this distinction in  $\lambda^\rightarrow$ : we cannot define a type  $pr$  such that  $a:pr$  iff  $a$  is the encoding of a proof, since we would not be able to tell whether a ‘proof’ is of  $t$  or of  $t'$ .

This observation may seem at odds with the relationship between  $NJ^\rightarrow$  and  $\lambda^\rightarrow$  established in §3, since we have already used  $NJ^\rightarrow$  to encode (propositional) proofs. But in our discussion of the Curry-Howard isomorphism, we were careful to talk about the *propositional fragment* of  $NJ^\rightarrow$ , but in order to encode even propositional proofs in  $NJ^\rightarrow$  we have used the language of *quantifier-free predicate* logic, not propositional logic, and in order to encode the rules for quantifiers, we needed explicit quantification.

To represent proofs we proceed by extending  $\lambda^\rightarrow$  with *dependent* types, that is, with types that can depend on terms. Specifically, we introduce a new operator,  $\Pi$ , where, if  $A$  is a type, and for every  $t \in A$ ,  $B[x \leftarrow t]$  is a type, then so is  $\Pi x^A. B$ . In other words, we use  $\Pi$  to build families of types,  $B[x \leftarrow t]$ , indexed by  $A$ .  $\Pi$  is sometimes called a dependent function space constructor because its members are functions  $f$  where, for every  $t \in A$ ,  $f(t)$  belongs to the type  $B[x \leftarrow t]$ . The addition of dependent types generalizes  $\lambda^\rightarrow$  since when  $x$  does not occur free in  $B$ , the type  $\Pi x^A. B$  is simply  $A \rightarrow B$  because its members are just the functions from  $A$  to  $B$  that we have in  $\lambda^\rightarrow$ .

Given dependent function types, we can define the provability relation for a logic as a type-valued function: instead of having  $pr$  be a single type, we index it over the formulae it might prove, i.e. we define it to be a function from objects  $\phi$  of type  $o$  (i.e. formulae) to the type of proofs of  $\phi$ . Using this, we define typed function constants that correspond to rules of proof. For example, we can now

formalize implication elimination as

$$\text{impe}:\Pi x^o. \Pi y^o. \text{pr}(\text{imp } x \ y) \rightarrow \text{pr}(x) \rightarrow \text{pr}(y)$$

i.e.,  $\text{impe}$  is a function which, given formulae  $x$  and  $y$  (objects of type  $o$ ), and terms proving  $x \supset y$  and  $x$ , returns a term proving  $y$ .

We now provide the formal details of an extension of  $\lambda^\top$  in which we can build dependent types, and show that the approach to representing deductive systems using type systems actually works the way we want.

### The metalogic $\lambda^p$

The particular theory we present, which we call  $\lambda^p$ , is closely related to the Edinburgh LF type-theory [Harper *et al.*, 1993] and the  $\Pi$  fragment of the AUTOMATH language AUT-PI [de Bruijn, 1980]. Our presentation is based on a similar presentation by Barendregt [1991; 1992], which we have chosen for its relative simplicity.

We define the expressions and types of  $\lambda^p$  together as follows:

**DEFINITION 13 (Pseudo-Terms).** Let  $\mathcal{V}$  be an infinite set of variables and  $\mathcal{K}$  be a set of constants that contains at least two elements,  $*$  and  $\square$ , which are called *sorts*. A set of *pseudo-terms*  $\mathcal{T}$  is described by the following grammar

$$\mathcal{T} ::= \mathcal{V} \mid \mathcal{K} \mid \mathcal{T}\mathcal{T} \mid \lambda\mathcal{V}^{\mathcal{T}}. \mathcal{T} \mid \Pi\mathcal{V}^{\mathcal{T}}. \mathcal{T}$$

$\Pi$  binds variables exactly like  $\lambda$ . Substitution (respecting bound variables) and bound variable renaming are defined in the standard manner.

**DEFINITION 14 (A deductive system for  $\lambda^p$ ).** We define, together, judgments for a *valid signature*, a *valid context* and a *valid typing*. In the following,  $s$  ranges over  $\{*, \square\}$ :

- A *signature* is a sequence given by the grammar

$$\Sigma ::= \langle \rangle \mid \Sigma, c:A$$

where  $c$  ranges over  $\mathcal{K}$ . A *signature*  $\Sigma$  is *valid* when it satisfies the relation  $\vdash_s$  defined by:

$$\frac{}{\vdash_s \langle \rangle} \quad \frac{\vdash_s \Sigma \quad \vdash_s A:s \quad c \notin \text{dom}(\Sigma)}{\vdash_s \Sigma, c:A}$$

- A *context* is a sequence given by the grammar

$$\Gamma ::= \langle \rangle \mid \Gamma, x:A$$

where  $x$  ranges over  $\mathcal{V}$ . A *context* is *valid* with respect to a valid signature  $\Sigma$  when it satisfies the relation  $\vdash_c$  defined by:

$$\frac{}{\vdash_c \langle \rangle} \quad \frac{\vdash_c \Gamma \quad \Gamma \vdash_s A:s \quad x \notin \text{dom}(\Gamma)}{\vdash_c \Gamma, x:A}$$

- A *type assignment relation*  $\vdash_{\Sigma}$ , indexed by a valid signature  $\Sigma$ , is defined between valid typing contexts  $\Gamma$  and typing assertions  $a:B$  where  $a, B \in \mathcal{T}$  is given by the rules:

$$\begin{array}{c}
\frac{}{\Gamma \vdash_{\Sigma} *:\square} \textit{axiom} \qquad \frac{\Gamma \vdash_{\Sigma} A:* \quad \Gamma, x:A \vdash_{\Sigma} B:s}{\Gamma \vdash_{\Sigma} \Pi x^A. B:s} \textit{form} \\
\frac{c:A \in \Sigma}{\Gamma \vdash_{\Sigma} c:A} \textit{assum} \qquad \frac{\Gamma, x:A \vdash_{\Sigma} b:B \quad \Gamma \vdash_{\Sigma} \Pi x^A. B:s}{\Gamma \vdash_{\Sigma} \lambda x^A. b:\Pi x^A. B} \textit{abst} \\
\frac{x:A \in \Gamma}{\Gamma \vdash_{\Sigma} x:A} \textit{hyp} \qquad \frac{\Gamma \vdash_{\Sigma} f:\Pi x^A. B \quad \Gamma \vdash_{\Sigma} a:A}{\Gamma \vdash_{\Sigma} f(a):B[x \leftarrow a]} \textit{appl} \\
\frac{\Gamma \vdash_{\Sigma} a:B \quad \Gamma \vdash_{\Sigma} B':s \quad B =_{\beta} B'}{\Gamma \vdash_{\Sigma} a:B'} \textit{conv}
\end{array}$$

We use the two sorts  $*$  and  $\square$  to classify entities in  $\mathcal{T}$  into levels. We say that  $*$  is the set of types and  $\square$  is the set of kinds. As in  $\lambda^{\rightarrow}$ , terms, which are here a subset of the pseudo-terms, belong to types; unlike in  $\lambda^{\rightarrow}$ , types, which are here also pseudo-terms, belong to  $*$ . For example, if  $o$  is a type (i.e.  $o:*$ ), then  $\lambda x^o. x$  is a term of type  $\Pi x^o. o$ , which we can abbreviate as  $o \rightarrow o$ , since  $x$  does not occur free in  $o$ . It is possible to build kinds, in limited ways, using the constant  $*$ ; in particular, the rules we give allow the formation of kinds with range  $*$ , e.g.  $o \rightarrow *$  but exclude kinds with domain  $*$ , e.g.  $* \rightarrow o$ . Hence we can form kinds like  $o \rightarrow *$  that have type-valued functions as members, but we cannot form kinds by quantifying over the set of types.

We state without proof a number of facts about this system. They have been proven in the more general setting of the so-called  $\lambda$ -cube (a family of eight related type systems) and generalized type systems examined by Barendregt [1991; 1992], but see also Harper *et al.* [1993] who show that the closely related LF type-theory has similar properties.

FACT 15. Term reduction in  $\mathcal{L}^p$  is Church-Rosser: given  $A, B, B' \in \mathcal{T}$ , then if  $A \xrightarrow{*}_{\beta} B$  and  $A \xrightarrow{*}_{\beta} B'$  there exists  $C \in \mathcal{T}$  where both  $B \xrightarrow{*}_{\beta} C$  and  $B' \xrightarrow{*}_{\beta} C$ .

FACT 16. Terms in  $\mathcal{L}^p$  are strongly normalizing: if  $\Gamma \vdash_{\Sigma} A:B$ , then  $A$  and  $B$  are strongly normalizing (all  $\beta$ -reductions starting with  $A$  or  $B$  terminate).

FACT 17.  $\mathcal{L}^p$  satisfies unicity of types: if  $\Gamma \vdash_{\Sigma} A:B$  and  $\Gamma \vdash_{\Sigma} A:B'$ , then  $B =_{\beta} B'$ .

From the decidability of these operations it follows that:

FACT 18. All judgments of  $\mathcal{L}^p$  are decidable.

#### *Relationship to other metalogics*

The proof-rules for  $\mathcal{L}^p$  extend the rules given for  $\lambda^{\rightarrow}$  in Definition 8. The rules for  $\lambda^{\rightarrow}$  essentially correspond to the identically named rules for  $\mathcal{L}^p$  restricted so

that in every typing assertion  $a:B$ ,  $a$  is a term of  $\lambda^\rightarrow$  and  $B$  is a simple type. The correspondence is not quite exact since in  $\lambda^p$  we have to prove that a signature, context, and type are well-formed (i.e., the first three parts of Definition 8), whereas this is assumed to hold in  $\lambda^\rightarrow$ . The need for this explicit demonstration of well-formedness is also reflected in the second premise of the  $\lambda^p$  rule for abstraction.

An example should clarify the connection between the two systems. In §4.2 we gave a signature  $\Sigma$  for minimal logic

$$\{ \text{imp}: o \rightarrow o \rightarrow o \}.$$

In  $\lambda^p$  we have

$$\Sigma = o:*, \text{imp}: o \rightarrow o \rightarrow o.$$

According to this, if  $x$  is of type  $o$ , then we can show in  $\lambda^p$  that  $\text{imp } x \ x$  is a well-formed proposition, i.e.,  $\Gamma \vdash_{\Sigma} \text{imp } x \ x:o$  where  $\Gamma = x:o$ , as follows:

$$\frac{\frac{\text{imp}: o \rightarrow o \rightarrow o \in \Sigma}{\Gamma \vdash_{\Sigma} \text{imp}: o \rightarrow o \rightarrow o} \text{assum} \quad \frac{x:o \in \Gamma}{\Gamma \vdash_{\Sigma} x:o} \text{hyp}}{\Gamma \vdash_{\Sigma} \text{imp } x:o \rightarrow o} \text{appl} \quad \frac{x:o \in \Gamma}{\Gamma \vdash_{\Sigma} x:o} \text{hyp}}{\Gamma \vdash_{\Sigma} \text{imp } x \ x:o} \text{appl}$$

However, the rules of  $\lambda^p$  formalize a strictly more expressive type-theory than  $\lambda^\rightarrow$ , and correspond, via a Curry-Howard isomorphism, to a more expressive logic. Terms are built, as we have already seen, by declaring function constants that form typed objects from other typed objects, e.g.,  $\text{imp } x \ x$  above corresponds to a term of type  $o$ . An  $n$ -ary predicate symbol  $P$ , which takes arguments of types  $s_1, \dots, s_n$ , has the kind  $s_1 \rightarrow \dots \rightarrow s_n \rightarrow *$ . The  $\Pi$ -type constructor corresponds either to universal quantification or (in its non-dependent form) implication. For example, given the signature

$$\Sigma = s_1:*, s_2:*, p:s_1 \rightarrow s_2 \rightarrow *$$

we can show that there is a  $t$  such that

$$\vdash_{\Sigma} t: (\Pi x^{s_1}. \Pi y^{s_2}. p(x, y)) \rightarrow \Pi y^{s_2}. \Pi x^{s_1}. p(x, y),$$

which corresponds to demonstrating the provability of the formula

$$\vdash_{\Sigma} (\forall x. \forall y. p(x, y)) \rightarrow \forall y. \forall x. p(x, y)$$

in a traditional sorted first-order setting.

In  $\lambda^p$  we generalize  $\lambda^\rightarrow$  so that types can depend on terms. We have not carried through this generalization to allow, e.g., types depending on types, which would allow impredicative higher-order quantification. As a result, and given the above

discussion, logics like  $\lambda^p$  and the LF are often described as first-order. Alternatively, since we can also quantify over functions (as opposed to predicates) at all types, some authors prefer to talk about minimal implicational predicate logic with quantification over all higher types [Simpson, 1992], or  $\omega$ -order logic ( $hh^\omega$ ) [Felty, 1991], to emphasize that these logics are more than first-order, but are not fully higher-order.

#### 4.4 Representation in $\lambda^p$

We have reached the point where the intuitions we have formulated about the relationship between natural deduction calculi and the logic of implication are reduced to a single formal system, the type-theory  $\lambda^p$ . In this system, the problems of encoding the syntax and proof rules of a deductive system are reduced to the single problem of providing a signature  $\Sigma$  and the problems of checking well-formedness of syntax and proof checking are reduced to (decidable) type-checking. We will expand on these points with two examples.

##### *A simple example: minimal logic*

A deductive system is encoded in  $\lambda^p$  by a signature that encodes

1. The language of the object logic and
2. The deductive system.

In §4.3 we gave a signature suitable for encoding the language of minimal logic. As we have seen, this consists first of an extension of the signature with types corresponding to syntactic categories and then with function constants over these types. The encoding of the deductive system also proceeds in two stages. First, we represent the basic judgments of the object logic.<sup>15</sup> To do this, for each judgment we augment the signature with a function from the relevant syntactic categories to a type. For minimal logic we have one judgment, that a formula is provable, so we add to the signature a function  $pr$ , of kind  $o \rightarrow *$ , where for any proposition  $p \in o$ ,  $pr(p)$  should be read as saying that the formula represented by  $p$  is provable. Second, we add constants to the signature that build (representatives of) proofs. Each constant is associated with a type that encodes (under the propositions-as-types correspondence) a proof rule of the object logic. For minimal logic we add constants with types that encode the formulae given in (6) from §3.2, which axiomatize the rules for minimal logic.

---

<sup>15</sup>Recall that *judgments* are assertions such as, e.g., that a proposition is provable. Typically, a logic only has a single judgment, but not always; for instance  $\lambda^p$  itself, in our presentation, has three judgments: a signature is well-formed, a context is well-formed, and a typing assertion is provable relative to a well-formed signature and context. The reader should be aware of the following possible source of confusion. By using a metalogic we have judgments at two levels: we use the judgment in  $\lambda^p$  that a typing assertion is provable relative to a signature and a context to demonstrate the truth of judgments in some object logic.

Putting the above pieces together, minimal logic is formalized by the following signature.

$$\begin{aligned} \Sigma \equiv & o:*, pr:o \rightarrow *, imp:o \rightarrow o \rightarrow o, \\ & impi:\Pi A^\circ B^\circ. (pr(A) \rightarrow pr(B)) \rightarrow pr(imp A B), \\ & imp:e:\Pi A^\circ B^\circ. pr(imp A B) \rightarrow pr(A) \rightarrow pr(B) \end{aligned}$$

It is an easy exercise to prove in  $\lambda^p$  that this is a well-formed signature.

Now consider how we use this signature to prove the proposition  $A \supset A$ . We encode this as  $imp A A$  and prove it by showing that the judgment  $pr(imp A A)$  has a member. We require, of course, that  $A$  is a proposition and we formalize this by the context  $A:o$ , which is well-formed relative to  $\Sigma$ . (In the following proof we have omitted rule names, but these can be easily reconstructed.)

Part I:

$$\frac{\frac{\frac{impi:\Pi A^\circ B^\circ. (pr(A) \rightarrow pr(B)) \rightarrow pr(imp A B) \in \Sigma \quad A:o \in A:o}{A:o \boxvDash impi:\Pi A^\circ B^\circ. (pr(A) \rightarrow pr(B)) \rightarrow pr(imp A B)} \quad A:o \boxvDash A:o}{A:o \boxvDash impi A:\Pi B^\circ. (pr(A) \rightarrow pr(B)) \rightarrow pr(imp A B)} \quad A:o \boxvDash A:o}{A:o \boxvDash impi A A:(pr(A) \rightarrow pr(A)) \rightarrow pr(imp A A)}}$$

Part II:

$$\frac{\frac{y:pr(A) \in A:o, y:pr(A)}{A:o, y:pr(A) \boxvDash y:pr(A)} \quad \dots}{A:o \boxvDash pr(A) \rightarrow pr(A):*} \quad \frac{}{A:o \boxvDash \lambda y^{pr(A)}. y:pr(A) \rightarrow pr(A)}}$$

(We have elided the subproof showing that  $pr(A) \rightarrow pr(A)$  is well-formed, which is straightforward using the formation rule *form*.) Putting the two parts together gives:

$$\frac{\text{Part I} \quad \text{Part II}}{A:o \boxvDash impi A A (\lambda y^{pr(A)}. y):pr(imp A A)}$$

Note that the reader interested in actually using a metalogic for machine supported proof construction should not be frightened away by the substantial ‘meta-level overhead’ that is associated with carrying out a proof of even very simple propositions like  $A \supset A$ . Real implementations of logical frameworks can hide much of this detail by partially automating the work of proof construction. Because all the judgments of  $\lambda^p$  are decidable, the well-formedness of signatures and contexts can be checked automatically, as can the typing of the terms that encode proofs.<sup>16</sup>

<sup>16</sup>This second point is not so important: Although the decidability of syntactic well-formedness is important, in practice, a framework is not used to decide if a given proof is valid, but as an interactive tool for building proofs.

This example shows how both well-formedness and proof checking are uniformly reduced to type checking. With respect to well-formedness of syntax, a proof that the type  $pr(\text{imp } A \ A)$  is inhabited is only possible if  $\text{imp } A \ A$  is of type  $o$ , i.e. it represents a well-formed proposition. That members of type  $o$  really represent well-formed propositions follows from adequacy and faithfulness of the representation of syntax, which for this example was argued (for  $\lambda^+$ ) in §4.2. With respect to proof checking, we have proven that the term  $\text{impi } A \ A \ (\lambda y^{pr(A)}. y)$  inhabits the type  $pr(\text{imp } A \ A)$ . In the same way that a term of type  $o$  represents a proposition, a term of type  $pr(p)$  represents a proof of  $p$ . In this example, the term represents the following natural deduction proof of  $A \supset A$ .

$$\frac{[A]_y}{A \supset A} \supset\text{-I}_y$$

The exact correspondence (adequacy and faithfulness) between terms and the proofs that they encode can be formalized (see Harper *et al.* [1993] for details), though we do not do this here since it requires first formalizing natural deduction proof trees and the representation of discharge functions. The idea is simple enough though: the proof rule  $\supset\text{-I}$  is encoded using a constant  $\text{impi}$ , and a well-typed application of this constructs a proof-term formalizing the operation of discharging an assumption. Specifically,  $\text{impi}$  builds an object (proof representative) of the type (proposition)  $pr(\text{imp } A \ B)$  given an object of type  $pr(A) \rightarrow pr(B)$ , i.e. a proof that can take any object of proof  $pr(A)$  (the hypothesis), and from it produce an object of type  $pr(B)$ .

In the example above the function must construct a proof of  $pr(A)$  from  $pr(A)$ , and  $\lambda y^{pr(A)}. y$  does this. In general, the question of which occurrences of  $pr(A)$  are discharged and how the proof of  $B$  is built is considerably more complex. Consider for example

$$\text{impi } A \ (\text{imp } B \ A) \ (\lambda x^{pr(A)}. \text{impi } B \ A \ (\lambda y^{pr(B)}. x)), \quad (21)$$

which is a member of the type  $pr(\text{imp } A \ (\text{imp } B \ A))$  in a context where  $A:o$  and  $B:o$ . This term represents a proof where Implication Introduction has been applied twice and the first (reading left to right) application discharges an assumption  $x$  and the second discharge (of  $y$ ) is vacuous. This proof-term corresponds to the following natural deduction proof.

$$\frac{\frac{[A]_x}{B \supset A} \supset\text{-I}_y}{A \supset (B \supset A)} \supset\text{-I}_x \quad (22)$$

*A larger example: first-order arithmetic*

A more complex example of a theory that we can easily formalize in  $\lambda^p$  is first-order arithmetic, and in fact we can define this as a direct extension of the system

$$\begin{aligned}
\text{oril} &: \Pi A^\circ B^\circ. pr(A) \rightarrow pr(\text{or } A B) \\
\text{orir} &: \Pi A^\circ B^\circ. pr(B) \rightarrow pr(\text{or } A B) \\
\text{ore} &: \Pi A^\circ B^\circ C^\circ. pr(\text{or } A B) \rightarrow (pr(A) \rightarrow pr(C)) \\
&\quad \rightarrow (pr(B) \rightarrow pr(C)) \rightarrow pr(C) \\
\text{raa} &: \Pi A^\circ. (pr(\text{imp } A \text{ falsum}) \rightarrow pr(\text{falsum})) \rightarrow pr(A) \\
\text{alli} &: \Pi A^{i \rightarrow o}. (\Pi x^i. pr(A(x))) \rightarrow pr(\text{all}(A)) \\
\text{alle} &: \Pi A^{i \rightarrow o} x^i. pr(\text{all}(A)) \rightarrow pr(A(x)) \\
\text{existsi} &: \Pi A^{i \rightarrow o}. \Pi x^i. pr(A(x)) \rightarrow pr(\text{exists}(A)) \\
\text{existse} &: \Pi A^{i \rightarrow o}. \Pi C^\circ. pr(\text{exists}(A)) \rightarrow (\Pi x^i. pr(A(x)) \rightarrow pr(C)) \rightarrow pr(C) \\
\text{ind} &: \Pi A^{i \rightarrow o}. pr(A(0)) \rightarrow (\Pi x^i. pr(A(x)) \rightarrow pr(A(sx))) \rightarrow pr(\text{all}(A))
\end{aligned}$$

Figure 1. Some proof-rules for arithmetic

we have already formalized. We extend the signature with the formalization of the syntax of arithmetic that we developed in §4.2 then we formalize the new rules, axioms and axiom-schemas that we need.

We have formalized some of the proof-rules in Figure 1 and most are self-explanatory. The first five extend minimal logic to propositional logic by adding rules for disjunction and falsum. We use the constant *falsum* to encode  $\perp$  (from which we can define negation as *not*  $A \equiv \text{imp } A \text{ falsum}$ ).

In our rules we assume a ‘classical’ falsum, i.e. the rule:

$$\begin{array}{c}
[A \supset \perp] \\
\vdots \\
\perp \\
\hline
A \quad \perp_c
\end{array}$$

encoded as *raa*. If we wanted an ‘intuitionistic’ falsum, we would replace this with the simpler rule encoded by

$$\Pi A^\circ. pr(\text{falsum}) \rightarrow pr(A).$$

For the quantifier rules we have not only given the rules for universal quantification (*alli* and *alle*) but also the rules for existential quantification, given by *existsi* and *existse*.

$$\frac{A[x \leftarrow t]}{\exists x. A} \exists\text{-I} \quad \frac{\begin{array}{c} [A] \\ \vdots \\ \exists x. A \quad C \end{array}}{C} \exists\text{-E}$$

These come with the usual side conditions: in  $\forall\text{-I}$ ,  $x$  cannot be free in any undischarged assumptions on which  $A$  depends and, for  $\exists\text{-E}$ ,  $x$  cannot be free in  $C$  or any assumptions other than  $A$  upon which (in the subderivation)  $C$  depends.



Object Logic		Metalogic
Syntactic Categories terms, individuals	$\rightsquigarrow$	Base Types $\{i:*, o:*\}$
Connectives & Constructors $\vee$	$\rightsquigarrow$	First-Order Constants $or:o \rightarrow o \rightarrow o$
Variable Binding Operators $\forall$	$\rightsquigarrow$	Higher-Order Constants $all:(i \rightarrow o) \rightarrow o$
Judgment $\vdash p$	$\rightsquigarrow$	Type Valued Functions $pr:o \rightarrow *$
Inference Rule $\frac{A}{A \vee B} \vee\text{-IL}$	$\rightsquigarrow$	Constant Declaration $oril:\Pi A^o B^o. pr(A) \rightarrow pr(or A B)$
Deductive System		Signature Declaration
Deduction		Typing Proof

Figure 2. Correspondence between object logics and their encodings

If we stop with the quantifier rules, the result is an encoding of first-order logic over the language of arithmetic. We have to add more rules to formalize the theory of equality and arithmetic. Thus, for example, *ind* formalizes the induction rule

$$\frac{A[x \leftarrow 0] \quad \begin{array}{c} [A] \\ \vdots \\ A[x \leftarrow sx] \end{array}}{\forall x. A}$$

and enforces the side condition that  $x$  does not occur free in any assumptions other than those discharged by the application of the rule. The other rules of arithmetic are formalized in a similar fashion.

#### 4.5 Summary

Figure 2 contains a summary. It is worth emphasizing that there is a relationship between the metalogic and the way that it is used, and an  $\rightarrow$ -framework like  $\lambda^p$  is well-suited to particular kinds of encodings. The idea behind higher-order syntax and the formalization of judgments using types is to internalize within the metalogic as much of the structure of terms and proofs as possible. By this we mean that syntactic notions and operations are subsumed by operations provided by the framework logic. In the case of syntax, we have seen how variable binding in the object logic is implemented by  $\lambda$ -abstraction in the framework logic and how substitution is implemented by  $\beta$ -reduction. Similarly, when representing proof-

rules and proof-terms, rather than our having to formalize, and then reason explicitly about, assumptions and their discharging, this is also captured directly in the metalogic. Support for this sort of internalization is one of the principles behind the design of these framework logics. The alternative (also possible in  $\lambda^P$ ) is to *externalize*, i.e., explicitly represent, such entities. The external approach is taken when using frameworks based on inductive definitions, which we will consider in §7.

## 5 ENCODING LESS WELL BEHAVED LOGICS

So far, we have restricted our attention to fairly standard, e.g. intuitionistic or classical, logics. We now consider how an  $\rightarrow$ -framework can treat the more ‘unconventional’ logics that we encounter in, for example, philosophy or artificial intelligence, for which such simple calculi are not available. As previously observed, most metalogics (and all the examples examined in this chapter) are ‘universal’ in the sense that they can represent any recursively enumerable relation, and thus any logic expressible in terms of such relations. However, there is still the question of how effective and natural the resulting encodings are.

We take as our example one of the more common kinds of philosophical logics: modal logic, i.e. propositional logic extended with the unary  $\Box$  connective and the *necessitation* rule (see Bull and Segerberg [1984]). Modal logics, as a group, have common features; for example, ‘canonical’ presentations use Hilbert calculi and, when natural deduction presentations are known, the proof-rules typically are not encodable in terms of the straightforward translation presented in §3.2. In §7 we will see how Hilbert presentations of these logics can be directly encoded as inductive definitions. Here we consider the problem of developing natural deduction presentations in an  $\rightarrow$ -framework. We explore two different possibilities, *labelled deductive systems* and *multiple judgment systems*, consider how practical they are, and how they compare.

### 5.1 Modal logic

We consider two modal logics in this section: K and an important extension, S4. A standard presentation of K is as a Hilbert system given by the axiom schemata

$$\begin{aligned} (A \supset B) \supset (A \supset B \supset C) \supset A \supset C \\ A \supset B \supset A \\ ((A \supset \perp) \supset \perp) \supset A \\ \Box(A \supset B) \supset \Box A \supset \Box B \end{aligned}$$

and the rules

$$\frac{A \supset B \quad A}{B} \textit{Det} \quad \text{and} \quad \frac{A}{\Box A} \textit{Nec}$$

The first of these rules is just the rule of detachment from §2.1, and the second is called necessitation. We get S4 from K by adding the additional axiom schemata:

$$\begin{aligned} \Box A \supset \Box \Box A \\ \Box A \supset A \end{aligned}$$

As noted above, we can see a Hilbert calculus as a special case of a natural deduction calculus, where the rules discharge no assumptions and axioms are premiseless rules.

There is an important difference between Hilbert and natural deduction calculi however, which is in the nature of what they reason about: Hilbert calculi manipulate formulae that are true in all contexts, i.e. valid (theorems), in contrast to natural deduction calculi, which typically manipulate formulae that are true under assumption. This difference causes problems when we try to give natural deduction-like presentations of modal logics, i.e. presentations that allow reasoning under temporary assumptions. The problem can be easily summarized:

PROPOSITION 19. *The deduction theorem (see §7.3) fails for K and S4.*

**Proof.** First, observe (e.g., semantically using Kripke structures; see §5.2) that  $A \supset \Box A$  is not provable in K or S4. However, if the deduction theorem held, we could derive this formula as follows: assume  $A$ , then, by necessitation, we have  $\Box A$ , and by the deduction theorem we would have that  $A \supset \Box A$  is a theorem. This is a contradiction. ■

The deduction theorem is a justification for the natural deduction rule  $\supset$ -I, but this in turn is precisely the rule that distinguishes natural deduction-like from Hilbert calculi: without it, one collapses into the other.

The problem of natural deduction encodings of modal logics is well known, and various fixes have been proposed. In some of these, the rules  $\supset$ -I and  $\supset$ -E are kept intact by extending the language of natural deduction itself. For instance if we allow global side conditions on rules then (following Prawitz [1965]) for S4 we have the rules

$$\frac{\begin{array}{c} \vdots * \\ A \end{array}}{\Box A} \supset\text{-I} \quad \text{and} \quad \frac{\Box A}{A} \supset\text{-E}$$

where  $*$  means that *all undischarged assumptions are boxed; i.e. of the form  $\Box B$* . Notice that given this side condition, the argument we have used to illustrate the failure of the deduction theorem no longer works. But the language of  $\rightarrow$  does not provide the vocabulary to express this side condition on  $\Box$ -I, so we cannot encode such a proof rule in the same fashion as proof-rules were encoded in §3.2.

## 5.2 A Kripke semantics for modal logics

A common way of understanding the meaning of formulae in a modal logic is in terms of the *Kripke*, or *possible worlds* semantics (see Kripke [1963] or van Benthem [1984] for details). We shall use this style of interpretation in developing our encodings.

A Kripke model  $(W, R, V)$  for a modal logic consists of a nonempty set of *worlds*  $W$ , a binary *accessibility* relation  $R$  defined over  $W$ , and a *valuation* predicate  $V$  over  $W$  and the propositional variables. We then define a *forcing* relation  $\Vdash$  between worlds and formulae as follows:  $a \Vdash A$  iff  $V(a, A)$  for  $A$  atomic;  $a \Vdash A \supset B$  iff  $a \Vdash A$  implies  $a \Vdash B$ ; and  $a \Vdash \Box A$  iff for all  $b \in W$  if  $a R b$  then  $b \Vdash A$ .

Using the Kripke semantics, we can classify modal logics by the behavior of  $R$  alone. For instance we have

FACT 20. Let  $R$  be the accessibility relation of a Kripke model.

- A formula  $A$  is a theorem of K iff  $A$  is forced at all worlds of all Kripke models.
- A formula  $A$  is a theorem of S4 iff  $A$  is forced at all worlds of all Kripke models where  $R$  is reflexive ( $x R x$ ) and transitive (if  $x R y$  and  $y R z$ , then  $x R z$ ).

It is now possible to see why the deduction theorem fails. Consider a Kripke model  $(W, R, V)$  and a formula  $A \supset B$ . In the deduction theorem we assume, for the sake of argument,  $A$  as a new axiom, and show that  $B$  is then a theorem; i.e. assuming  $\forall a \in W. a \Vdash A$ , we show that  $\forall a \in W. a \Vdash B$ . But it does not follow from this that  $A \supset B$  is a theorem; i.e. that  $\forall a \in W. a \Vdash A \supset B$ .

It is however easy to find a correct ‘semantic’ analogue of the deduction theorem:

FACT 21. For any Kripke model  $(W, R, V)$

$$\forall a \in W. (a \Vdash A \rightarrow a \Vdash B) \rightarrow a \Vdash A \supset B.$$

The problem of providing a natural deduction encoding of a modal logic can be reduced to the problem of capturing this semantic property of  $\supset$  in rules that can be directly encoded in the language of implication. We will consider two ways of doing this, which differ in the extent to which they make the Kripke semantics explicit.

## 5.3 Labelled deductive systems<sup>17</sup>

The above analysis suggests one possible solution to our problem: we can internalize the semantics into the deductive calculus. Hence, instead of reasoning with

<sup>17</sup>The work described in this section was done in collaboration with Luca Viganò.

formulae, we reason about formulae in worlds; i.e. we work with pairs  $a:A$  where  $a$  is a world and  $A$  is a formula.

Taking this approach, the rules

$$\begin{array}{c}
 \begin{array}{c} [x:A] \\ \vdots \\ x:B \\ \hline x:A \supset B \end{array} \supset-I \quad \begin{array}{c} [x R y] \\ \vdots \\ y:A \\ \hline x:\Box A \end{array} \Box-I \\
 \\
 \begin{array}{c} [x:A \supset \perp] \\ \vdots \\ y:\perp \\ \hline x:A \end{array} \perp-E \quad \begin{array}{c} x:A \supset B \quad x:A \\ \hline x:B \end{array} \supset-E \quad \begin{array}{c} x:\Box A \quad x R y \\ \hline y:A \end{array} \Box-E
 \end{array}$$

define a natural deduction calculus, which we call  $K_L$ . (We also require the side conditions that in  $\Box-I$   $y$  is different from  $x$  and does not occur in the assumptions on which  $y:A$  depends, except those of the form  $x R y$  that are discharged by the inference.) These rules formalize the meaning of both  $\supset$  and  $\Box$  in terms of the Kripke semantics; i.e., we locate applications of  $\supset-I$  in some particular world, and take account of the other worlds in defining the behavior of  $\Box$  and  $\perp$  (where it suffices to derive a contradiction in any world). We can show:

FACT 22 (Basin *et al.* [1997a]).  $a:A$  is provable in  $K_L$  iff  $A$  is true in all Kripke models, and therefore, by the completeness of  $K$  with respect to the set of all Kripke models, iff  $A$  is a theorem of  $K$ .

As an example of a proof in  $K_L$  of a  $K$  theorem, we show that  $\Box$  distributes over  $\supset$ .

$$\begin{array}{c}
 \frac{\frac{[a:\Box(A \supset B)]_1 \quad [a R b]_3}{b:A \supset B} \Box-E \quad \frac{[a:\Box A]_2 \quad [a R b]_3}{b:A} \Box-E}{\frac{b:B}{a:\Box B} \Box-I_3} \supset-E \\
 \frac{\frac{a:\Box B}{a:\Box A \supset \Box B} \supset-I_2}{a:\Box(A \supset B) \supset \Box A \supset \Box B} \supset-I_1
 \end{array}$$

Further, and essential for our purpose, there are no new kinds of side conditions on the rules of  $K_L$ , so we have no difficulty in formalizing these in an  $\rightarrow$ -framework.<sup>18</sup> The following is a signature for  $K_L$  in  $\lambda^p$  (note that for the sake of

<sup>18</sup>There is of course the side condition on  $\Box-I$ . But this can be formalized in the same way that eigenvariable conditions are formalized in logics with quantifiers, by using universal quantification in the metalogic.

readability we write ‘:’ and  $R$  in infix form):

$$\begin{aligned} \Sigma_{K_L} \equiv & w:*, o:*, ‘:’:w \rightarrow o \rightarrow *, R:w \rightarrow w \rightarrow *, \\ & falsum:o, imp:o \rightarrow o \rightarrow o, box:o \rightarrow o, \\ & FalseE:\Pi A^o x^w y^w. (x:imp A falsum \rightarrow y:falsum) \rightarrow x:A, \\ & impI:\Pi A^o B^o x^w. (x:A \rightarrow x:B) \rightarrow x:imp A B, \\ & impE:\Pi A^o B^o x^w. x:A \rightarrow x:imp A B \rightarrow x:B, \\ & boxI:\Pi A^o x^w. (\Pi y^w. x R y \rightarrow y:A) \rightarrow x:box A, \\ & boxE:\Pi A^o x^w y^w. x:box A \rightarrow x R y \rightarrow y:A \end{aligned}$$

The signature reflects that there are two types, a type of worlds  $w$  and type of formulae  $o$ , and two judgments; one about the relationship between worlds and formulae, asserting that a formula is true at that world, and a second, between two worlds, asserting that the first accesses the second. Adequacy and faithfulness follow by the style of analysis given in §3.

$K_L$  as a base for other modal logics

We can now take  $K_L$  as a base upon which to formalize other modal logics. Since modal logics are characterized, in terms of Kripke models, purely in terms of their accessibility relations, to get other modal logics we must simply modify the behavior of  $R$  in our encoding. Thus, since S4 corresponds to the class of Kripke models with transitive and reflexive accessibility relations, we can enrich our signature with:

$$\begin{aligned} Ref:& \Pi x^w. x R x \\ Trans:& \Pi x^w y^w z^w. x R y \rightarrow y R z \rightarrow x R z \end{aligned}$$

Again, we can show [Basin *et al.*, 1997a] that this really formalizes S4.

*The limits of  $K_L$*

It might appear from the discussion above that we can implement any modal logic we want, simply by adding the axioms for the appropriate accessibility relation to  $K_L$ . That is, we represent a logic by embedding its semantics in the metalogic, a formalization technique that is sometimes called *semantic embedding* (see van Benthem [1984] or Ohlbach [1993] for details on this approach). We must be careful though; not every embedding based on labelling accurately captures the semantics and different kinds of embeddings capture more structure than others.

Consider  $K_L$  again: the rules for  $\supset$  and  $\Box$  reflect the meaning that the Kripke semantics gives the connectives. On the other hand, the semantics does not explicitly state how the rules for  $\perp$  should function using labels. Following from the

rules for  $\supset$ , a plausible formalization is

$$\frac{\begin{array}{c} [x:A \supset \perp] \\ \vdots \\ x:\perp \end{array}}{x:A} \perp\text{-}E^-$$

which, like the rule for implication, stays in one world and ignores the existence of different possible worlds. But if we investigate the logic that results from using this rule (instead of  $\perp\text{-}E$ ), we find that it is not complete with respect to the Kripke semantics.

An examination of the Gentzen-style natural language characterization of  $\perp$  shows where the problem lies:

If the assumption that  $A \supset \perp$  in world  $x$  is inconsistent with the interpretation, then  $A$  is true in world  $x$ .

This says nothing about where the inconsistency might be and specifically does not say that it should be in the world  $x$  itself. The role of negation in encoding the semantics of logics is subtle and we lack space to develop this topic here. We therefore restrict ourselves to a few comments; much more detail can be found in [Basin *et al.*, 1997a]. In  $K_L$  we assumed that it is enough to be able to show that the inconsistency is in some world. It turns out that this is sufficient for a large class of logics; but again this does not reflect the complete behavior of  $\perp$ . Some accessibility relations require a richer metalogic than one based on minimal implication and this may in turn require formalizing all of first or even higher-order logic. In such formalizations, we must take account of the possibility that the inconsistency of an assumption about a world might manifest itself as a contradiction in the theory of the accessibility relation, or *vice versa*. It is possible then to use classical first (or higher-order) logic as a metatheory to formalize the Kripke semantics in a complete way, however the result also has drawbacks. In particular, we lose structure in the proofs available in  $K_L$ . In  $K_L$  we reason in two separate systems. We can reason in just the theory of the accessibility relation and then use the results of this in the theory of the labelled propositions; however, we cannot go in the other direction, i.e. we cannot use reasoning in the theory of labelled propositions as part of an argument about the relations. This enforced separation provides extra structure that we can exploit, e.g., to bound proof search (see, e.g., [Basin *et al.*, 1997b]). And in spite of enforcing this separation,  $K_L$  is a sufficient foundation for a very large class of standard logics.<sup>19</sup>

---

<sup>19</sup>In [Basin *et al.*, 1997a] we show that it is sufficient to define almost all the modal logics of the so-called Geach hierarchy, which includes most of those usually of interest, i.e. K, T, S4, S5, etc., though not, e.g., the modal logic of provability  $G$  [Boolos, 1993].

### 5.4 Alternative multiple judgment systems

Using a labelled deductive system, we built a deductive calculus for a modal logic based on two judgments: a formula is true in a world and one world accesses another. But this is only one of many possible presentations. We now consider another possibility, using multiple judgments that distinguish between truth (in a world) and validity that is due originally to Avron *et al.* [1992] (and further developed in [Avron *et al.*, 1998]).

#### Validity

Starting with K, we can proceed in a Gentzen-like manner, by writing down the behavior of the logical connectives as given by the Kripke semantics. If we abbreviate ‘ $A$  is true in all worlds’ to  $V(A)$  ( $A$  is valid; i.e.  $A$  is a theorem), then we have

$$\frac{V(A \supset B) \quad V(A)}{V(B)} \supset\text{-}E_V \quad \text{and} \quad \frac{V(A)}{V(\Box A)} \Box\text{-}I_V, \quad (23)$$

which can be easily verified against the Kripke semantics (they directly reflect the two rules *Det* and *Nec*). Since the deduction theorem fails, we do not have an introduction rule for  $\supset$  in terms of  $V$ , neither do we have a rule for  $\perp$ .

Thus the first part of the signature for K simply records the rules (23):

$$\begin{aligned} \Sigma_1 \equiv & o:*, \quad V:o \rightarrow *, \\ & \text{False}:o, \quad \text{imp}:o \rightarrow o \rightarrow o, \quad \text{box}:o \rightarrow o, \\ & \text{imp}E_V:\Pi A^\circ B^\circ. V(A) \rightarrow V(\text{imp } A B) \rightarrow V(B), \\ & \text{box}I_V:\Pi A^\circ. V(A) \rightarrow V(\text{box } A) \end{aligned}$$

And, as observed above, we have

**LEMMA 23.** *The rules encoded in  $\Sigma_1$  are sound with respect to the standard Kripke semantics of K, if we interpret  $V(A)$  as  $\forall a. a \Vdash A$ .*

**The rest of  $V$ .** The rest of the details about  $V(\cdot)$  could be summarized simply by declaring all the axioms of K to be valid. In this case  $V(\cdot)$  would simply encode a Hilbert presentation of K. However there is a more interesting possibility, which supports proof under assumption.

#### Truth in a world

As previously observed we do have a kind of semantic version of the deduction theorem relativized to any given world. We can use this to formalize more about



the meaning of implication and  $\perp$ . If we abbreviate ‘ $A$  is true in some arbitrary but fixed world  $c$ ’ to  $T(A)$ , then we have:

$$\frac{T(A \supset B) \quad T(A)}{T(B)} \supset\text{-}E_T \quad \frac{\begin{array}{c} [T(A)] \\ \vdots \\ T(B) \end{array}}{T(A \supset B)} \supset\text{-}I_T \quad \frac{\begin{array}{c} [T(A \supset \perp)] \\ \vdots \\ T(\perp) \end{array}}{T(A)} \perp\text{-}E_T$$

When we talked above about validity we did not have a rule for introducing implication, or reflecting the behavior of  $\perp$ . Similarly, when we are talking about truth in some world, we do not have a rule reflecting the behavior of  $\Box$ , since that is not dependent on just the one world. Thus for instance, there is no introduction (or elimination) rule for this operator. All we can say is that

$$\frac{T(\Box(A \supset B)) \quad T(\Box A)}{T(\Box B)} \text{Norm}_T$$

And again we can verify these rules against the semantics.

Thus the second part of the signature is

$$\begin{aligned} \Sigma_2 \equiv & T:o \rightarrow *, \\ & \text{imp}E:\Pi A^\circ B^\circ. T(\text{imp } A B) \rightarrow T(A) \rightarrow T(B), \\ & \text{imp}I:\Pi A^\circ B^\circ. (T(A) \rightarrow T(B)) \rightarrow T(\text{imp } A B), \\ & \text{false}E:\Pi A^\circ. (T(\text{imp } A \text{ falsum}) \rightarrow T(\text{falsum})) \rightarrow T(A), \\ & \text{norm}:\Pi A^\circ B^\circ. T(\text{box } (\text{imp } A B)) \rightarrow T(\text{box } A) \rightarrow T(\text{box } B) \end{aligned}$$

and we have

LEMMA 24. *The rules encoded in  $\Sigma_1, \Sigma_2$  are sound with respect to the standard Kripke semantics of  $\mathbb{K}$ , if we interpret  $V(A)$  as in Lemma 23 and  $T(A)$  as  $c \Vdash A$  where  $c$  is a fixed constant.*

### Connecting $V$ and $T$

We have now formalized two separate judgments, defined by the predicates  $V$  and  $T$ , which we have to connect together. To this end, we introduce two rules,  $C$  and  $R$ , which intuitively allow us to introduce a validity judgment, given a truth judgment, and eliminate (or use) a validity judgment.

$C$  states that if  $A$  is true in an arbitrary world, then it is valid.

$$\frac{T(A)}{V(A)} C$$

For this rule to be sound, we require that the world where  $A$  is true really is arbitrary, and this will hold so long as there are no other assumptions  $T(A')$  current when we apply it. It is easy to see that this condition is ensured for any proof of the atomic proposition  $V(A)$ , so long as  $C$  is the only rule connecting  $T$  and  $V$  together, since the form of the rules then ensures that there can be no hypotheses  $T(A')$  at a place where  $C$  is applied.

The addition of  $C$  yields a complete inference system for reasoning about valid formulae. However, the resulting deductive system is awkward: once we end up in the  $V$  fragment of the system, which by itself is essentially a Hilbert calculus, we are forced to stay there.

We thus extend our system with an elimination rule for  $V$ , to allow us to return to the natural deduction-like  $T$  fragment. Important to our justification of the rule  $C$  was that the premise followed in an *arbitrary* world. Any further rule that we add must not invalidate this assumption. However we observe that given  $V(A)$ , then, since  $A$  is valid, it is true in an arbitrary world, so adding  $T(A)$  as an open assumption to an application of  $C$  does not harm the semantic justification of that rule application. We can encode this as the following rule  $R$ .

$$\frac{\begin{array}{c} [T(A)] \\ \vdots \\ V(A) \quad V(B) \end{array}}{V(B)} R$$

Taken together, these rules complete our proposed encoding of  $K$ .

$$\begin{aligned} \Sigma_{K_{MJ}} &\equiv \Sigma_1, \Sigma_2, \\ &C:\Pi A^o. T(A) \rightarrow V(A), \\ &R:\Pi A^o B^o. V(A) \rightarrow (T(A) \rightarrow V(B)) \rightarrow V(B) \end{aligned}$$

To establish correctness formally, we begin by proving that:

**PROPOSITION 25.** *If  $A$  is a theorem of  $K$ , then  $V(A)$  is a theorem of the proof calculus encoded as  $\Sigma_{K_{MJ}}$ .*

**Proof.** We observe that if  $A$  is one of the listed axioms of  $K$ , then we can show that  $T(A)$ , and thus, by  $C$ , that  $V(A)$ . Therefore we need not declare these to be ‘valid’. These, and the rules encoded in  $\Sigma_1$  allow us to reconstruct any proof in Hilbert  $K$  (see also the remarks after Lemma 23 above). ■

Next that:

**PROPOSITION 26.** *If  $V(A)$  is a theorem of the proof calculus encoded as  $\Sigma_{K_{MJ}}$ , then  $A$  is a theorem of  $K$ .*

We prove a slight generalization, for which we need

LEMMA 27. For an arbitrary set  $\Gamma$  of theorems of  $K$ , if  $T(A)$  is a theorem of  $\Sigma_{K_{MJ}}$  extended with the axiom set  $\{T(x) \mid x \in \Gamma\}$ , then  $A$  is a theorem of  $K$ .

**Proof.** First notice that only the rules encoded as *impI*, *impE*, *FalseE* and *norm* can occur in a proof of  $T(A)$ . By Lemma 24, reading  $T(A)$  as  $c \Vdash A$ , only theorems of  $K$  follow from this fragment of the proof calculus extended with  $\Gamma$ , where  $\Gamma$  consists of theorems of  $K$ . ■

The generalization of Proposition 26 is then

LEMMA 28. For an arbitrary set  $\Gamma$  of theorems of  $K$ , if  $V(A)$  is a theorem of the proof calculus encoded as  $\Sigma_{K_{MJ}}$  extended with  $\{T(x) \mid x \in \Gamma\}$ , then  $A$  is a theorem of  $K$ .

**Proof.** The proof is by induction on the size of a proof  $\Pi$  of  $V(A)$ . We need to consider three cases: (i) The last rule in the proof is an application of one of the rules encoded in  $\Sigma_1$  from theorems  $V(A_i)$ , in which case, by appeal to the induction hypothesis,  $A_i$  are theorems of  $K$  and thus  $A$  is a theorem of  $K$ . (ii) The last rule is an application of  $C$ , in which case the sub-proof is a proof of the theorem  $T(A)$  (there are no undischarged assumptions for the theorem  $V(A)$  and hence  $T(A)$ ) and by Lemma 27,  $A$  is a theorem of  $K$ . (iii) The last rule is an application of  $R$  to proofs  $\Pi_1$  of  $V(B)$  from  $T(A)$  and  $\Pi_2$  of the theorem  $V(A)$ . Since  $\Pi_2$  is smaller than  $\Pi$ , by the induction hypothesis  $A$  is a theorem of  $K$ . Then we can transform  $\Pi_1$  into a proof of the theorem  $V(B)$  in the proof calculus formalized as  $\Sigma_{K_{MJ}}$  extended with  $\{T(x) \mid x \in \Gamma \cup \{A\}\}$  by replacing any appeal to a hypothesis  $T(A)$  in  $\Pi_1$  by an appeal to the axiom  $T(A)$ . Since the result is a proof no bigger than  $\Pi_1$ , which in turn is smaller than  $\Pi$ , by appeal to the induction hypothesis,  $B$  is a theorem of  $K$ . ■

We can combine Propositions 25 and 26 as

THEOREM 29.  $A$  is a theorem of  $K$  iff  $V(A)$  follows from  $\Sigma_{K_{MJ}}$ .

### Encoding other modal logics

We can extend the encoding of  $K$  easily to deal with S4; there are in fact several ways we can do this: one possibility (see Avron *et al.* [1992] for more discussion) is to add the rules

$$\frac{\begin{array}{c} [V(\Box A)] \\ \vdots \\ V(B) \end{array}}{V(\Box A \supset B)} \supset\text{-}I_V \quad \text{and} \quad \frac{V(\Box A)}{V(A)} \Box\text{-}E_V$$

(given these rules we can show that the *Norm<sub>T</sub>* rule is redundant). This produces an encoding that is closely related to the version of S4 suggested by Prawitz.

### *An alternative view of $K_{MJ}$*

We have motivated and developed the  $K_{MJ}$  presentation of  $K$  using a Kripke semantics, as a parallel with  $K_L$ . Unlike  $K_L$ , however, the interpretation is implicit, not explicit (there is no mention of particular worlds in the final proof calculus) so we are not committed to it. In fact, and importantly, this presentation can be understood from an entirely different perspective that uses just the Hilbert axiomatization itself, as follows.

We observe in §7 that it is possible to prove a deduction theorem for classical propositional logic, justifying the  $\supset$ -I rule by proof-theoretic means in terms of the Hilbert presentation. If we examine that proof, then we can see that it is easily modified for the fragment of (our Hilbert presentation of)  $K$  *without* the rule *Nec*. But this is precisely the fragment of  $K$  that is defined by the *T* fragment of  $K_{MJ}$ . Equally, the system defined by *V*, can be seen as the full Hilbert calculus.

Thus we can alternatively view the two judgments of  $K_{MJ}$  not as indicating whether we are speaking of truth in some world, or truth in all worlds, but rather whether or not we are allowed to apply the deduction theorem. This perspective provides the possibility of an entirely different proof of the correctness of our encoding, based on the standard Hilbert encoding, and without any intervening semantic argument.

### *5.5 Some conclusions*

We have presented two different encodings of two well-known modal logics in this section as examples of approaches to representing nonstandard logics in an  $\rightarrow$ -framework. Which approach is preferable depends, in the end, on how the resulting encoding will be used.  $K_L$  and  $S4_L$  make the semantic foundation of the presentation more explicit. This is advantageous if we take for granted the view that modal logics are the logics of Kripke models since the user is able to exploit the associated intuitions in building proofs. On the other hand this may be a problem if we want to use our encoding in circumstances where our intuitions are different. The opposite holds for  $K_{MJ}$  and  $S4_{MJ}$ : it is more difficult to make direct use of any intuitions we might have from the Kripke semantics, but, since the proof systems involve no explicit, or even necessary, commitment to that interpretation, we have fewer problems in assuming another.<sup>20</sup>

The solutions we have examined, while tailored for modal logics, do have some generality. However, each different logic must be considered in turn and the approaches presented here may not always be applicable, or the amount of effort in modifying them may be considerable. For instance it is possible to interpret relevance logics in terms of a Kripke semantics that can be adopted as the basis of

<sup>20</sup>In fact it would be fairly easy to adapt a multiple judgment presentation of  $S4$  to the different circumstances of say relevance, or linear, propositional logic, which share many properties with traditional  $S4$ . Such a project would be considerably more difficult, not to mention questionable, starting from  $S4_L$ . This issue of commitment to a particular interpretation is discussed at length with regard to labelled deductive systems in general by Gabbay [1996].

a labelled deductive system, similar in style to (though considerably more complex than)  $K_L$  (see Basin *et al.* [1998b]). But such an implementation is tricky to use and we reach, eventually, the limits of encodings that are understandable and usable by humans.

## 6 CONSEQUENCE RELATIONS

In the previous sections we considered an abstraction of natural deduction calculi and in the following section we will consider an abstraction of Hilbert calculi. Here, we consider the third style of proof calculus we mention in the introduction: the sequent calculus. It turns out that, unlike the other two, little work has been done on systems that directly abstract away from sequent calculi in a way that we can use as a logical framework. This certainly does not mean that there has been no work on the principles of the sequent calculus, just that work has concentrated not on the concrete implementational aspects so much as on the abstract properties of the sequent relation,  $\vdash$ , which when investigated in isolation is called a *consequence relation*.

Consequence relations provide a powerful tool for systematically analyzing properties of a wide range of logics, from the traditional logics of mathematics to modal or substructural logics, in terms that we can then use as the starting point of an implementation. In fact it is often possible to encode the results of a sequent calculus analysis directly in an  $\rightarrow$ -framework.

What, then, is a consequence relation? There are several definitions in the literature (e.g. Avron [1991; 1992], Scott [1974] and Hacking [1979]); we adopt that of Avron, along with his vocabulary, where possible.

**DEFINITION 30.** A *consequence relation* is a binary relation between finite multisets of formulae  $\Gamma, \Delta$ , usually written  $\Gamma \vdash \Delta$ , and satisfying at least *Basic* and *Cut* in (2) and *PL* and *PR* in (3).<sup>21</sup>

This is, however, a very general definition. In fact most logics that we might be interested in encoding have natural presentations in terms of more constrained *ordinary* consequence relations:<sup>22</sup>

**DEFINITION 31.** A consequence relation is said to be *ordinary* if it satisfies the rules *WL*, *CL*, *WR* and *CR* of (3).

Examples of ordinary consequence relations are  $LJ^\supset$  and  $LK^\supset$  defined in §2.2,

<sup>21</sup>In the rules given in §2.2, the antecedent and succedent are *sequences* of formulae whereas here they are *multisets*. In practice, the permutation rules *PL* and *PR* are often omitted and multi-sets are taken as primitive, as here. This is not always possible though, e.g., in  $\lambda^p$ , where the ordering in the sequence matters. Note also that this definition does not take account of variables; for an extension to that case, see Avron [1992].

<sup>22</sup>We do not have space to consider the best known exception, the *substructural* logics, e.g. relevance and linear logic. However what we say in this section generalizes, given a suitably modified  $\rightarrow$ -framework, to these cases. Readers interested in the (non-trivial) technical details of this modification are referred to [Cervesato and Pfenning, 1996] and, especially, [Ishtiaq and Pym, 1998].

as is the encoding in terms of  $\vdash$  of  $\text{NJ}^\supset$  in §3.1 (even though *Cut* is not a basic property of this presentation, we can show that it is admissible; i.e. we do not change the set of provable sequents if we assume it). Most of the traditional logics of mathematics can be presented as ordinary (in fact, *pure* ordinary, see below) consequence relations.

### 6.1 The meaning of a consequence relation

While it is possible to treat a consequence relation purely in syntactic terms, often one can be understood, and may have been motivated, by some idea of the ‘meaning’ of the formulae relative to one another. For instance we can read a sequent  $\Gamma \vdash \Delta$  of  $\text{LK}^\supset$  as ‘if all the formulae in  $\Gamma$  are true, then at least one of the formulae in  $\Delta$  is true.’ Because they have this reading in terms of truth, we call the systems defined by  $\text{LJ}^\supset$ ,  $\text{LK}^\supset$  and  $\text{NJ}^\supset$  ‘truth’ consequence relations. Notice that the meaning of ‘truth’ here is not fixed: when we say that something is true we might mean that it is classically true, intuitionistically true, Kripke-semantically true, relevance true, or even something else more exotic.

We can derive a different sort of consequence relation from a Hilbert calculus: if we assume (1) *Basic*, i.e. that  $A \vdash A$ , (2) that if  $A$  is an axiom then  $\Gamma \vdash A$ , and (3) that for each rule of proof

$$\frac{A_1 \dots A_n}{A}$$

we have that

$$\frac{\Gamma_1 \vdash A_1 \dots \Gamma_n \vdash A_n}{\Gamma_1, \dots, \Gamma_n \vdash A}$$

then it is easy to show that the resulting system satisfies *Cut*, and that  $\vdash A$  iff  $A$  is a theorem. This is not a truth consequence relation: the natural reading of  $\Gamma \vdash A$  is ‘if  $\Gamma$  are *theorems*, then  $A$  is a *theorem*’. We thus call  $\vdash$  a *validity* consequence relation.

Of course truth and validity are not the only possibilities. We can define consequence relations any way we want,<sup>23</sup> the only restriction we might impose is that in order to be effectively mechanizable on a computer, the relation  $\vdash$  should be recursively enumerable.

### 6.2 Ordinary pure consequence relations and $\rightarrow$ -frameworks

Part of the problem with mechanizing consequence relations, if we formalize them directly, is their very generality. Many systems are based on ordinary consequence relations, and if an encoding forces us to deal explicitly with all the rules that

<sup>23</sup>For some examples of other consequence relations which can arise in the analysis of a modal logic, see Fagin *et al.* [1992].

formalize this, then proof construction will often require considerable and tedious structural reasoning. This may explain, in part, why there has been little practical work on logical frameworks based directly on consequence relations.<sup>24</sup> However another reason is that an  $\rightarrow$ -framework can very effectively encode many ordinary consequence relations directly. In the remainder of this section we explore this reduction, which clarifies the relationship between consequence relations and  $\rightarrow$ -frameworks as well as illuminating some of the strengths and weaknesses of  $\rightarrow$ -frameworks.

In order to use an  $\rightarrow$ -framework for representing consequence relations, it helps if we impose a restriction in addition to ordinaryness.

DEFINITION 32. We say that a consequence relation is *pure* if, given that

$$\frac{\Gamma_1 \vdash \Delta_1 \quad \dots \quad \Gamma_n \vdash \Delta_n}{\Gamma_0 \vdash \Delta_0}$$

holds, then there are  $\Gamma'_i, \Delta'_i$ , which are sub-multisets of  $\Gamma_i, \Delta_i$ , such that for arbitrary  $\Gamma''_i, \Delta''_i$

$$\frac{\Gamma'_1, \Gamma''_1 \vdash \Delta'_1, \Delta''_1 \quad \dots \quad \Gamma'_n, \Gamma''_n \vdash \Delta'_n, \Delta''_n}{\Gamma'_0, \Gamma''_0 \vdash \Delta'_0, \Delta''_0}.$$

Notice that the consequence relations discussed at the beginning of this section all satisfy the definition of purity; this is also the case for most of the logics we encounter in mathematics. In order to find counterexamples we must look to some of the systems arising in philosophical logic. For instance in modal logic (see §5, and the discussion by Avron [1991]) we get a natural truth consequence relation satisfying the rule

$$\frac{\vdash A}{\vdash \Box A}$$

but not, for arbitrary  $\Gamma$ ,

$$\frac{\Gamma \vdash A}{\Gamma \vdash \Box A}$$

We shall not, here, consider frameworks that can handle general impure consequence relations satisfactorily.

### *Single conclusioned consequence*

As we said earlier, most of the standard logics of mathematics have intuitive presentations as ordinary pure consequence relations. Avron [1991] adds one more restriction before claiming that

<sup>24</sup>There are, however, many computer implementations of particular sequent calculi, e.g. the PVS system for higher-order logic, by Owre *et al.* [1995], not to mention many tableau calculi, which are, essentially, sequent calculi.

Every ordinary, pure, single-conclusioned [consequence relation] system can, e.g., quite easily be implemented in the Edinburgh LF.

We begin by considering the single conclusioned case, and then follow it with multiple conclusioned consequence relations and systems based on multiple consequence relations.

The encoding is uniform and consists of two parts. We first explain how to encode sequents and then rules. We can encode

$$A_1, \dots, A_n \vdash A$$

as

$$T(A_1) \rightarrow \dots \rightarrow T(A_n) \rightarrow T(A)$$

and it is easy to show that this satisfies Definition 31 of an ordinary consequence relation (in this case, single conclusioned). Notice how the structural properties of the consequence relation are directly reflected by the logical properties of  $\rightarrow$ .

We can then represent basic and derived rules expressed in terms of such a consequence relation, *assuming it is pure* as follows. Consider a rule of such a consequence relation  $\vdash$ , where, for *arbitrary*  $\Gamma_i$ :

$$\frac{\Gamma_1, A_{(1,1)}, \dots, A_{(1,n_1)} \vdash A_1 \quad \dots \quad \Gamma_m, A_{(m,1)}, \dots, A_{(m,n_m)} \vdash A_m}{\Gamma_1, \dots, \Gamma_m, A_{(0,1)}, \dots, A_{(0,n_0)} \vdash A_0}$$

We can encode this as

$$\begin{aligned} (T(A_{(1,1)}) \rightarrow \dots \rightarrow T(A_{(1,n_1)}) \rightarrow T(A_1)) \rightarrow \dots \\ \rightarrow (T(A_{(m,1)}) \rightarrow \dots \rightarrow T(A_{(m,n_m)}) \rightarrow T(A_m)) \\ \rightarrow T(A_{(0,1)}) \rightarrow \dots \rightarrow T(A_{(0,n_0)}) \rightarrow T(A_0) \end{aligned}$$

leaving the  $\Gamma_i$  implicit (the condition of purity is important because it allows us to do this).

As an example, consider  $\text{LJ}^\supset$  from §2.2, for which we have the rules

$$\frac{\Gamma \vdash A \quad \Gamma', B \vdash C}{\Gamma, \Gamma', A \supset B \vdash C} \supset-L \quad \frac{\Gamma, A \vdash B}{\Gamma \vdash A \supset B} \supset-R$$

We can encode these rules (assuming a suitable formalization of the language) as

$$T(A) \rightarrow (T(B) \rightarrow T(C)) \rightarrow T(A \supset B) \rightarrow T(C) \quad (24)$$

and

$$(T(A) \rightarrow T(B)) \rightarrow T(A \supset B). \quad (25)$$



Then we can simulate the effect of applying  $\supset$ -L, i.e.

$$\frac{F_1, \dots, F_m \vdash A \quad G_1, \dots, G_n, B \vdash C}{F_1, \dots, F_m, G_1, \dots, G_n, A \supset B \vdash C},$$

encoded as (24), to the encoded assumptions

$$T(F_1) \rightarrow \dots \rightarrow T(F_m) \rightarrow T(A) \quad (26)$$

and

$$T(G_1) \rightarrow \dots \rightarrow T(G_n) \rightarrow T(B) \rightarrow T(C). \quad (27)$$

By (26) and (24) we have

$$T(F_1) \rightarrow \dots \rightarrow T(F_m) \rightarrow (T(B) \rightarrow T(C)) \rightarrow T(A \supset B) \rightarrow T(C) \quad (28)$$

which combines with (27) to yield

$$T(F_1) \rightarrow \dots \rightarrow T(F_m) \rightarrow T(G_1) \rightarrow \dots \rightarrow T(G_n) \rightarrow T(A \supset B) \rightarrow T(C),$$

which encodes the conclusion of  $\supset$ -L.

From this, it is easy to see that our encoding of  $\text{LJ}^\supset$  is adequate. We show faithfulness by a modification of the techniques discussed above.

#### *An observation about implementations*

Note that the last step, ‘combining’ (28) with (27), actually requires a number of steps in the metalogic; e.g. shuffling around formulae by using the fact that, in the metalogic of the framework itself,  $T(A) \rightarrow T(B) \rightarrow T(C)$  is equivalent to  $T(B) \rightarrow T(A) \rightarrow T(C)$ . But such reasoning must come out somewhere in any formalism of a system based on consequence relations. There is no getting around some equivalent of structural reasoning; the question is simply of how, and where, it is done, and how visible it is to the user.

In fact in actual implementations of  $\rightarrow$ -frameworks, e.g., Isabelle [Paulson, 1994], the cost of this structural reasoning is no greater than in a custom implementation since the framework theory itself will be implemented in terms of a pure, ordinary single conclusioned consequence relation; i.e. as a sequent or natural deduction calculus. If we write the consequence relation of the implementation as  $\Longrightarrow$ , then it is easy to see that an encoded sequent such as (26) can be quickly transformed into the logically equivalent

$$T(F_1), \dots, T(F_m) \Longrightarrow T(A)$$

at which point it is possible to ‘piggy-back’ the rest of the proof off of  $\Longrightarrow$ , and thus make use of the direct (and thus presumably efficient) implementation of its structural properties.

### *Multiple concluded consequence*

Having examined how we might encode ordinary pure single concluded consequence relations in an  $\rightarrow$ -framework, we now consider the general case of multiple concluded relations, which occur in standard presentations of many logics (e.g.  $\text{LK}^\supset$  described in §2.2).

There is no obvious correspondence between such relations and formulae in the language of  $\rightarrow$ . However, by refining our encoding a little, we can easily extend Avron's observation to the general case and thereby give a direct and effective encoding of multiple concluded consequence relations. We take as our example the system  $\text{LK}^\supset$ . For larger developments and applications of this style, the reader is referred to [Pfenning, 2000].

A multiple concluded consequence relation is a pair of multisets of formulae, which we can refer to as 'left' and 'right'; i.e.

$$\overbrace{A_1, \dots, A_n}^{\text{left}} \vdash \overbrace{B_1, \dots, B_m}^{\text{right}}. \quad (29)$$

To encode this we need not one judgment  $T$ , but *two* judgments which we call  $T_L$  and  $T_R$ ; we also define a new propositional constant  $E$ . We can encode (29) in an  $\rightarrow$ -framework as

$$T_L(A_1) \rightarrow \dots \rightarrow T_L(A_n) \rightarrow T_R(B_1) \rightarrow \dots \rightarrow T_R(B_m) \rightarrow E.$$

This is not quite enough to give us a consequence relation though; unlike in the single concluded case, we do not automatically get that *Basic* is true. However we can remedy this by declaring that

$$T_L(A) \rightarrow T_R(A) \rightarrow E$$

and then we can show that the encoding defines an ordinary consequence relation.

We can then extend the style of encoding described to define, e.g., the right and left rule for implication in  $\text{LK}^\supset$ , which are

$$(T_L(A) \rightarrow T_R(B) \rightarrow E) \rightarrow T_R(A \supset B) \rightarrow E$$

and

$$(T_R(A) \rightarrow E) \rightarrow (T_L(B) \rightarrow E) \rightarrow T_L(A \supset B) \rightarrow E.$$

As in the single concluded case, a necessary condition for this encoding is that the consequence relation is pure. It is also easy to show that the encoding is adequate and faithful with respect to the derivability of sequents in  $\text{LK}^\supset$ .

*Multiple consequence relation systems*

Finally, we come to the problem of how to formalize systems based on more than one consequence relation. Here we will briefly consider just the single conclusioned case; the same remarks, suitably modified, also apply to the multiple conclusioned case.

One approach to analyzing a formal system (and quite useful if we want the analysis to be in terms of pure consequence relations) is, rather than using a single relation, to decompose it into a set of relations  $\vdash_1$  to  $\vdash_n$ . We can encode each of these relations, and their rules, just like in the single relation case, using predicates  $T_1$  to  $T_n$ ; i.e.

$$A_1, \dots, A_n \vdash_i A$$

as

$$T_i(A_1) \rightarrow \dots \rightarrow T_i(A_n) \rightarrow T_i(A).$$

We encounter a new encoding problem with multiple consequence systems. These typically (always, if they are interesting) contain rules that relate consequence relations to each other, i.e., rules where the premises are built from different consequence relations than the conclusion. Consider the simplest example of this, which simply declares that one consequence relation is a subrelation of another (what we might call a *bridge* rule):

$$\frac{\Gamma \vdash_1 A}{\Gamma \vdash_2 A} \text{ bridge}$$

We can encode this as the pair of schemata

$$T_1(A) \rightarrow T_2(A) \tag{30}$$

and

$$(T_1(A) \rightarrow T_2(B)) \rightarrow T_2(A) \rightarrow T_2(B) \tag{31}$$

and show that the resulting encoding is properly closed. That is, given

$$T_1(A_1) \rightarrow \dots \rightarrow T_1(A_n) \rightarrow T_1(B)$$

by (30) we get

$$T_1(A_1) \rightarrow \dots \rightarrow T_1(A_n) \rightarrow T_2(B)$$

then by  $n$  applications of (31) we eventually get

$$T_2(A_1) \rightarrow \dots \rightarrow T_2(A_n) \rightarrow T_2(B).$$

As this example shows, working with a multiple consequence relation system in an  $\rightarrow$ -framework may require many metalevel steps to simulate a proof step in the object logic (here the number depends on the size of the sequent). More practical experience using such encodings is required to judge whether they are really usable in practice as opposed to just being a theoretically interesting way of encoding multiple consequence relation systems in the logic of  $\rightarrow$ .

In fact we have already encountered an example of this kind of an encoding: the multiple judgment encoding of modal logic developed in [Avron *et al.*, 1992; Avron *et al.*, 1998], described in §5.4, can be seen as the encoding of a two consequence relation system for truth and validity where  $T_1$  is  $T$ ,  $T_2$  is  $V$  and *bridge* is implemented by  $R$  and  $C$ .

## 7 DEDUCTIVE SYSTEMS AS INDUCTIVE DEFINITIONS

In the introduction we discussed two separate traditions of metatheory: metatheory as a unifying language and metatheory as proof theory. We have shown too how  $\rightarrow$ -frameworks fit into the unifying language tradition, and the way different logics can be encoded in them and used to carry out proofs. However,  $\rightarrow$ -frameworks are inadequate for proof theory: in exchange for ease of reasoning *within* a logic, reasoning *about* the logic becomes difficult or impossible.

In order better to understand this point, and some of the subtleties it involves, consider the following statements about the (minimal) logic of  $\supset$ .

1.  $A \supset A$  is a theorem.
2.  $A \supset B \supset C$  is true on the assumption that  $B \supset A \supset C$  is true.
3. The deduction theorem holds for  $HJ^{\supset}$ .
4. The deduction theorem holds for all extensions of  $HJ^{\supset}$  with additional axioms.

Statement 1 can be formalized in a metalogic as a statement about provability in any complete presentation of the logic of  $\supset$ ; e.g.  $NJ^{\supset}$ ,  $LJ^{\supset}$  or  $HJ^{\supset}$ . As a statement about provability we might regard it as, in some sense, a proof-theoretic statement. But as such, it is very weak since, by completeness, it must be provable irrespective of the deductive system used.

Statement 2 expresses a derived rule of the logic of  $\supset$ . Its formalization requires that we can reason about the truth of one formula relative to others. As explained in §6, representing this kind of a (truth) consequence relation can easily be reduced to a provability problem in an  $\rightarrow$ -framework. For example, in the  $\lambda^p$  encoding of  $NJ^{\supset}$  we prove that the type  $\Pi A^o B^o C^o. pr(B \supset A \supset C) \rightarrow pr(A \supset B \supset C)$  is inhabited.

As with statement 1, the metalogic must be able to build proofs in the object logic (in this case though under assumptions, encoded using  $\rightarrow$  in the metalogic), but proofs themselves need not be further analyzed.

Statement 3 (which we prove in this section) is an example of a metatheoretic statement that is more in the proof theory tradition. In order to prove it we must analyze the structure of arbitrary proofs in the deductive system of  $HJ^\supset$  using an inductive argument. The difference between this and the previous example is important: for a proof of statement 2 we need to know which axioms and rules are available in the formalized deductive system, while for a proof of statement 3 we also need to know that no other rules are present, since this is what justifies an induction principle over proofs (the rules of  $HJ^\supset$  can be taken as constituting an inductive definition). An  $\rightarrow$ -framework like  $\lambda^p$  contains no provisions for carrying out induction over the structure of an encoded deductive system.

Statement 4 is an extension of statement 3, which introduces the problem of *theory structuring*. Structuring object theories to allow metatheoretic results to be ‘imported’ and used in related theories is not a substantial problem in the kinds of metamathematical investigations undertaken by proof theorists, who are typically interested in proving particular results about particular systems. However, for computer scientists working on formal theorem proving, it is enormously important: it is good practice for a user reasoning about complex theories to formalize a collection of simpler theories (e.g. for numbers and arithmetic, sequences, relations, orders, etc.) that later are combined together as needed. So some kind of theory structuring facility is practically important and many systems provide support for it.<sup>25</sup>

Unfortunately, hierarchical structure in  $\rightarrow$ -frameworks is the result of an assumption (necessary anyway for other reasons) that the languages and deductive systems of encoded logics are ‘open to extensions’ (see §7.4 below), something that automatically rules out any arguments requiring induction on the structure of the language<sup>26</sup> or proofs of a theory, e.g. the deduction theorem. If we ‘close’ the language or deductive system by explicitly adding induction over the language or proofs, in order to prove metatheorems, it is unsound later to assume those theorems in extensions of the deductive system or language. In §7.4, we suggest that there is a way to avoid this problem if we formulate metatheorems in a metalogic based on inductive definitions.

These examples illustrate that there are different sorts of metatheoretic statements, which are distinguished by how conscious we have to be of the metatheoretic context in order to prove them. The central role of induction in carrying

---

<sup>25</sup> For example HOL [Gordon and Melham, 1993], Isabelle [Paulson, 1994] and their predecessor LCF [Gordon *et al.*, 1979] support simple theory hierarchies where theorems proven in a theory may be used in extensions.

<sup>26</sup> An example of a metatheorem for which we need induction over the language, not the derivations, is that *In classical logic, a propositional formula that contains only the  $\Leftrightarrow$  connective is valid if and only if each propositional variable occurs an even number of times.*

out many kinds of the more general metatheoretic arguments is the reason we now consider logical frameworks based on inductive definitions.

### 7.1 *Historical background: From Hilbert to Feferman*

What kind of a metalogic is suited for carrying out metatheoretic arguments that require induction on the language or deductive system of the object logic? To answer this question we draw on experience gained by proof theorists dating back to Hilbert, in particular Post and Gödel, and later Feferman, who have asked very similar questions. We can also draw on practical experience over the last 30 years in the use of computers in work with formal systems.

#### *The work of Post and Gödel*

In the early part of this century, Post [1943] (see Davis [1989] for a short survey of Post's work) investigated the decidability of logics like that of *Principia Mathematica*. He showed that such systems could be formalized as (what we now recognize as) recursively enumerable classes of strings and that this provided a basis for metatheoretic analysis. Although Post's work is a large step towards answering our question, one important aspect, from our point of view, is missing from his formalization. There is no consideration of *formal* metatheory. Post was interested in a formal characterization of deductive systems, not of their metatheories; he simply assumed, reasonably for his purposes, that arbitrary mathematical principles could be adopted as necessary for the metatheory.

We cannot make the same assumption with a logical framework: we must decide first which mathematical principles we need, and then formalize them. The work of Post is thus complemented, for our purposes, by Gödel's [1931] work on the incompleteness of systems like *Principia Mathematica*, which shows that a fragment of arithmetic is sufficient for complex and general metatheory. Logicians have since been able to narrow that fragment down to the theory of primitive recursive arithmetic, which seems sufficient for general syntactic metatheory.<sup>27</sup>

The natural numbers, although adequate for Gödel's purposes, are too unwieldy to use as a logical framework. Indeed, a large part of Gödel's paper is taken up with a complicated technical development showing that arithmetic really can represent syntax. The result is unusable in a practical framework: the relationship between syntax and its encoding is only indirectly given by (relatively) complicated arithmetic functions and the numbers generated in the encoding can be enormous. This is in contrast to Post's strings (further investigated in the early sixties by mathematicians such as Smullyan [1961]), which have a simple and direct correspondence with the structures we want to encode, and a compact representation.

---

<sup>27</sup> This is a thesis, of course, not a theorem; and exceptions have to be made for, e.g., proof normalization theorems, for which (as a corollary of Gödel's result itself) we know there can be no single general metatheory.

### *S-expressions and $FS_0$*

It is possible to build a formal metatheory based on strings, in the manner of Post. However, experience in computer science in formalizing and working with large symbolic systems has shown that there is an even more natural language for modeling formal (and other kinds of symbolic) systems. The consensus of this experience is found in Lisp [McCarthy, 1981; Steele Jr. and Gabriel, 1996], which for more than 30 years has remained the most popular language for building systems for symbolic computation.<sup>28</sup> Further, Lisp is not only effective, but its basic data-structure, which has, in large part, contributed to its effectiveness, is remarkably simple: the *S-expression*. This is the data-type freely generated from a base type by a binary function ‘Cons’. Experience with Lisp has shown that just about any syntactic structure can be mapped directly onto S-expressions in such a way that it is very easy to manipulate. In fact, we can even restrict the base type of S-expressions to be a single constant (often called ‘*nil*’) and still get this expressiveness. Further evidence for the effectiveness of Lisp and S-expressions is that one of the most successful theorem proving systems ever built, NQTHM [Boyer and Moore, 1981], verifies recursive functions written in a version of Lisp that uses precisely this class of S-expressions.

An example of a theory that integrates all the above observations is  $FS_0$ , due to Feferman. This provides us with a language in which we can define exactly all the recursively enumerable classes of S-expressions. Moreover, it permits inductive arguments over these inductively defined classes, and thus naturally subsumes both the theory of recursively enumerable classes and primitive recursive arithmetic. It has proved usable too in case-studies of computer supported metatheory, in the proof-theoretic tradition (see Matthews [1992; 1993; 1994; 1997b]). In the rest of this chapter we will use a slightly abstracted version of  $FS_0$  for our discussion.

## 7.2 *A theory of inductive definitions: $FS_0$*

$FS_0$  is a simple minimal theory of inductive definitions, which we present here in an abstract form (we elide details in order to emphasize the general, rather than the particular, features). A full description of the theory is provided by Feferman [1990], while an implementation is described by Matthews [1996].

$FS_0$  is a theory of inductively defined sets, embedded in first-order logic and based on Lisp-style S-expressions. The class of S-expressions is the least class containing *nil* and closed under the pairing (*cons*) operator, which we write as an infix comma  $(\cdot, \cdot)$ , such that  $nil \neq (a, b)$  for all S-expressions *a* and *b*; we also assume, for convenience, that comma associates to the right, so that  $(a, (b, c))$  can

---

<sup>28</sup>This does not mean that there have not been successful programming languages that use strings as their basic data-structure; the SNOBOL family [Griswold, 1981] of languages, for example, is based on the theory of Markoff string transformation algorithms. However it is significant that SNOBOL, in spite of its mathematical elegance in many ways, has never been seen as a general purpose symbolic programming language like Lisp, or been adopted so enthusiastically by such a large and influential programming community.

be abbreviated to  $(a, b, c)$ . We then have functions  $car$  and  $cdr$ , which return the first and second elements of a pair.

Comprehension over first-order predicates is available. We write

$$x \in S \Leftrightarrow P(x) \quad \text{or} \quad S \equiv \{ x \mid P(x) \}$$

to indicate a set  $S$  so defined. Such definitions can be parameterized and the parameters are treated in a simple way, thus, for example, we can write

$$x \in S(a, b) \Leftrightarrow (x, a) \in b.$$

We can also define sets explicitly as inductive definitions using the  $I(\cdot, \cdot)$  construction: if  $A$  and  $B$  are sets, then  $I(A, B)$  is the least set containing  $A$  and closed under the rule

$$\frac{t_1 \quad t_2}{t}$$

where  $(t, t_1, t_2) \in B$ . Note that we only have inductive definitions with exactly two ‘predecessors’, but this is sufficient for our needs here and, with a little more effort, in general.

Finally, we can reason about inductively defined sets using the induction principle

$$\begin{aligned} Base \subseteq S \rightarrow \forall a, b, c. (b \in S \rightarrow c \in S \rightarrow \\ (a, b, c) \in Step \rightarrow a \in S) \rightarrow I(Base, Step) \subseteq S. \end{aligned} \quad (32)$$

This says that a set  $S$  contains all the members of a set  $I(Base, Step)$  if it contains the members of  $Base$  and whenever it contains two elements  $b$  and  $c$ , and  $(a, b, c)$  is an instance of the rule  $Step$ , then it also contains  $a$ . This induction principle applies to sets, not predicates, but it is easy, by comprehension, to generate one from the other, so this is not a restriction.<sup>29</sup>

### 7.3 A Hilbert theory of minimal implication

Having sketched a theory of inductive definitions, we now consider how it might actually be used, both to encode an object logic, and to prove metatheorems. As an example, we encode the theory  $HJ^{\supset}$  (of §2.1) and prove a deduction theorem.

*The definition of HJ*

**The language  $L_{HJ}$**  We define an encoding  $\ulcorner \cdot \urcorner$  of the standard language of implicational logic  $L_{HJ}$  (as usual we distinguish metalevel  $(\rightarrow)$  from object level  $(\supset)$ )

<sup>29</sup>In  $FS_0$  comprehension is restricted to essentially  $\Sigma_1^0$  predicates with the result that the theory is recursion-theoretically equivalent to primitive recursive arithmetic.



implication) as follows. We define two distinct S-expression constants (e.g. *nil* and  $(nil, nil)$ ) which we call `atom` and `imp`, then we have

$$\begin{aligned}\ulcorner a \urcorner &= (\text{atom}, \ulcorner a \urcorner) && (a \text{ atomic}) \\ \ulcorner a \supset b \urcorner &= (\text{imp}, \ulcorner a \urcorner, \ulcorner b \urcorner)\end{aligned}$$

(assuming  $\ulcorner a \urcorner$  for atomic  $a$  to be already, separately, defined). It is easy to see that  $\ulcorner \cdot \urcorner$  is an injection from  $L_{\text{HJ}}$  into the S-expressions, on the assumption that  $\ulcorner \cdot \urcorner$  is. For the sake of readability, in the future we will abuse notation, and write simply  $a \supset b$  when we mean the schema  $(\text{imp}, a, b)$ ; i.e.  $a$  and  $b$  here are variables in the theory of inductive definitions, not propositional variables in the encoded language  $L_{\text{HJ}}$ .

**The theory HJ** We now define a minimal theory of implication, HJ, as follows. We have two classes of axioms

$$x \in K \Leftrightarrow \exists a, b. x = (a \supset b \supset a)$$

and

$$x \in S \Leftrightarrow \exists a, b, c. x = ((a \supset b) \supset (a \supset b \supset c) \supset a \supset c)$$

and a rule of detachment

$$x \in Det \Leftrightarrow \exists a, b. x = (b, (a \supset b), a)$$

from which we define the theory HJ to be

$$\text{HJ} \equiv I(K \cup S, Det).$$

*Using HJ*

We can now use HJ to prove theorems of the Hilbert calculus  $\text{HJ}^\supset$  in the same way that we would use an encoding to carry out natural deduction proofs in an  $\rightarrow$ -framework. One difference is that we do not get a direct correspondence between proof steps in the encoded theory and steps in the derivation. However, in this case it is enough to prove first the following lemmas:

$$a \supset b \supset a \in \text{HJ} \tag{33}$$

$$(a \supset b) \supset (a \supset b \supset c) \supset a \supset c \in \text{HJ} \tag{34}$$

$$a \in \text{HJ} \rightarrow (a \supset b) \in \text{HJ} \rightarrow b \in \text{HJ} \tag{35}$$

Here, and in future, we assume theorems are universally closed. From these, it follows that if  $A$  is a theorem of minimal implication, then  $\ulcorner A \urcorner \in \text{HJ}$ .

For example, we show

$$a \supset a \in \text{HJ} \quad (36)$$

with the derivation

1.  $a \supset a \supset a \in \text{HJ}$  by (33)
2.  $a \supset (a \supset a) \supset a \in \text{HJ}$  by (33)
3.  $(a \supset a \supset a) \supset (a \supset (a \supset a) \supset a) \supset a \supset a \in \text{HJ}$  by (34)
4.  $(a \supset (a \supset a) \supset a) \supset a \supset a \in \text{HJ}$  by (35),1,3
5.  $a \supset a \in \text{HJ}$  by (35),2,4

The steps of this proof correspond, one to one, with the steps of the proof that we gave for  $\text{HJ}^\supset$  in §2.1. As this example suggests, at least for propositional Hilbert calculi, inductive definitions support object theory where proofs in the metatheory closely mirror proofs in the object theory. (For logics with quantifiers the relationship is less direct, since we have to encode and explicitly reason about variable binding and substitution.)

#### *Proving a deduction theorem for HJ*

Let us now consider an example that requires induction over proofs themselves: the deduction theorem for HJ. We only sketch the proof; the reader is referred to Basin and Matthews [2000] for details. Informally, the deduction theorem says:

If  $B$  is provable in HJ with the additional axiom  $A$  then  $A \supset B$  is provable in HJ.

Note that this theorem relates different deductive systems: HJ and its extension with the axiom  $A$ . Moreover, as  $A$  and  $B$  are schematic, ranging over all formulae of HJ, the theorem actually relates provability in HJ with provability in infinitely many extensions, one for each proposition  $A$ .

To formalize this theorem we define

$$\text{HJ}[\Gamma] \equiv I(K \cup S \cup \Gamma, \text{Det}); \quad (37)$$

i.e.  $\text{HJ}[\Gamma]$  is the deductive system HJ where the axioms are extended by all formulae in the class  $\Gamma$ . Now we can formalize the deduction theorem as

$$b \in \text{HJ}[\{a\}] \rightarrow (a \supset b) \in \text{HJ}, \quad (38)$$

which in  $\text{FS}_0$  can be transformed into

$$I(K \cup S \cup \{a\}, \text{Det}) \subseteq \{x \mid (a \supset x) \in \text{HJ}\}.$$

This in turn can be proved by induction on the inductively defined set  $I(K \cup S \cup \{a\}, \text{Det})$  using (32). The proof proceeds as follows:

**Base case** We have to show  $K \cup S \cup \{a\} \subseteq \{x \mid (a \supset x) \in \text{HJ}\}$ . This reduces, via  $x \in K \cup S \cup \{a\} \rightarrow (a \supset x) \in \text{HJ}$ , to showing that  $(a \supset x) \in \text{HJ}$  given either (i)  $x \in K \cup S$  or (ii)  $x \in \{a\}$ . For (i) we have  $x \in \text{HJ}$  and  $(x \supset a \supset x) \in \text{HJ}$ , and thus, by *Det* that  $(a \supset x) \in \text{HJ}$ . For (ii) we have  $x = a$  and thus have to show that  $(a \supset a) \in \text{HJ}$ , which we do following the proof of (36) above.

**Step case** There is only one rule (*Det*) and thus one case: given  $b \in \{x \mid (a \supset x) \in \text{HJ}\}$  and  $(b \supset c) \in \{x \mid (a \supset x) \in \text{HJ}\}$ , prove that  $c \in \{x \mid (a \supset x) \in \text{HJ}\}$ . This reduces to proving, given  $(a \supset b) \in \text{HJ}$  and  $(a \supset b \supset c) \in \text{HJ}$  that  $(a \supset c) \in \text{HJ}$ . This in turn follows by observing that  $((a \supset b) \supset (a \supset b \supset c) \supset a \supset c) \in \text{HJ}$  by (34), from which, by (35) twice with the hypotheses,  $(a \supset c) \in \text{HJ}$ .

Once we have proved the deduction theorem we can use it to build proofs where we reason under assumption in the style of natural deduction. This is useful, indeed in practice essential, if we really wish to use Hilbert calculi to prove anything. However it is also limited since this metatheorem can be applied *only* to HJ. Thus, we next consider how this limitation can be partially remedied.

#### 7.4 Structured theory and metatheory

At the beginning of this section, we discussed two examples of the deduction theorem (statements 3 and 4) where the second stated that the deduction theorem holds not just in  $\text{HJ}^\supset$  but also in extensions. We return to this example, which illustrates an important difference between ID-frameworks and  $\rightarrow$ -frameworks.

##### *Structuring in $\rightarrow$ -frameworks*

Let us first examine how theories can be structured in  $\rightarrow$ -frameworks. Consider the following: we can easily encode  $\text{HJ}^\supset$  as (assuming the encoding of the syntax is given separately) the axiom set  $\Gamma_{\text{HJ}^\supset}$ .

$$\begin{aligned} & T(A \supset B \supset A) \\ & T((A \supset B) \supset (A \supset B \supset C) \supset A \supset C) \\ & T(A) \rightarrow T(A \supset B) \rightarrow T(B) \end{aligned}$$

Then  $A$  is a theorem of  $\text{HJ}^\supset$  iff  $\Gamma_{\text{HJ}^\supset} \vdash T(A)$ , i.e.,  $T(A)$  is provable in the metalogic under the assumptions  $\Gamma_{\text{HJ}^\supset}$ . Now consider the deduction theorem in this setting; we would have to show that

$$\Gamma_{\text{HJ}^\supset} \vdash (T(A) \rightarrow T(B)) \rightarrow T(A \supset B). \quad (39)$$

This is not possible, however.

In an  $\rightarrow$ -framework, a basic property of  $\vdash$  is weakening; i.e. if  $\Gamma \vdash \phi$  then  $\Gamma, \Delta \vdash \phi$ . This is very convenient for structuring theories: a theorem proven

under the assumptions  $\Gamma$  holds under extension with additional assumptions  $\Delta$ . For example,  $\Delta$  might extend our formalization of  $\text{HJ}^\supset$  to a classical theory of  $\supset$  or perhaps to full first-order logic.<sup>30</sup> By weakening, given  $\Gamma_{\text{HJ}^\supset} \vdash \phi$  we immediately have  $\Gamma_{\text{HJ}^\supset}, \Delta \vdash \phi$ . Thus we get a natural hierarchy on the object theories we define: theory  $T'$  is a subtheory of theory  $T$  when its axioms are a subset of those of  $T$ . This allows us to reuse proven metatheorems since anything proven for a subtheory automatically follows in any supertheory.

Consider, on the other hand, the extension  $\Delta_K$  consisting of the axioms

$$\begin{aligned} T(A) &\rightarrow T(\Box A) \\ T(\Box A \supset \Box(A \supset B) \supset \Box B) \end{aligned}$$

with which we can extend  $\Gamma_{\text{HJ}^\supset}$  to a fragment of the modal logic  $K$ . The deduction theorem *does not* follow in  $K$ ; therefore, since by faithfulness we have

$$\Gamma_{\text{HJ}^\supset}, \Delta_K \not\vdash (T(A) \rightarrow T(B)) \rightarrow T(A \supset B),$$

we must also have, by weakening and contraposition,

$$\Gamma_{\text{HJ}^\supset} \not\vdash (T(A) \rightarrow T(B)) \rightarrow T(A \supset B).$$

This suggests that there is an either/or situation: we can have either hierarchically structured theories, as in an  $\rightarrow$ -framework, or general inductive metatheorems (like the deduction theorem), as in an ID-framework, but not both. In fact, as we will see, in an ID-framework things are not quite so clear-cut: there is the possibility both to prove metatheorems by induction and to use them in certain classes of extensions.

#### *Structuring in an ID-framework*

Part of the explanation of why we can prove (38), but not (39), is that it is not possible to extend the deduction theorem for  $\text{HJ}$  to arbitrary supertheories: (38) is a statement about  $\text{HJ}$  and it tells us nothing about anything else. However a theorem about  $\text{HJ}$  alone is of limited use: in practice we are likely to be interested in  $\text{HJ}^\supset$  as a fragment of some larger theory. We know, for instance, that the deduction theorem follows for many extensions of  $\text{HJ}$  (e.g. extensions to larger fragments of intuitionistic or classical logic). The problem is that the induction principle we use to prove the theorem is equivalent to a closure assumption, and such an assumption means that we are not able to use the theorem with extensions.

We seem to have confirmed, from the other side, the trade-off we have documented above for  $\rightarrow$ -frameworks: either we can have induction and no theory structuring (as theories are ‘closed’), or vice versa. However, if we look harder,

<sup>30</sup> An extension of the deduction theorem to first-order logic, however, is not trivial—we have to treat a new rule, which introduces complex side conditions to the statement of the theorem (Kleene [1952] discusses one way to do this).

there is sometimes a middle way that is possible in ID-frameworks. The crucial point is that an inductive argument does not always rely on *all* the closure assumptions of the most general case. Consider the assumptions that are made in the proof of (38):

- The proof of the base case relies on the fact that the axioms  $a \supset b \supset a$  and  $a \supset a$  are available in HJ.
- The proof of the step case relies on the fact that the *only* rule that can be applied is *Det*, and that the axiom

$$(a \supset b) \supset (a \supset b \supset c) \supset a \supset c$$

is available.

What bears emphasizing is that we do not need to assume that no axioms other than those explicitly mentioned are available, only that no rules other than *Det* are available.

We can take account of this observation to produce a more general version of (38)

$$b \in \text{HJ}[\Gamma \cup \{a\}] \rightarrow (a \supset b) \in \text{HJ}[\Gamma], \quad (40)$$

which we can still prove in the same way as (38). We call this version *open-ended* since it can be used with any axiomatic extension  $\Gamma$  of HJ. In particular (38) is just (40) where we take  $\Gamma$  to be  $\emptyset$ .

**Structuring theories with the deduction theorem** Unlike (38), we can make effective use of (40) in a hierarchy of theories in a way similar to what is possible in an  $\rightarrow$ -framework. The metatheorem can be applied to any extension  $\text{HJ}[\Gamma]$  where  $\Gamma$  is a collection of axioms. The fact that in the  $\rightarrow$ -framework we can add new rules, not just new axioms, is not as significant as it at first appears, so long as we have that

$$T(A) \rightarrow T(B) \quad \text{iff} \quad T(A \supset B) \quad (41)$$

since we can use this to find an axiomatic equivalent of any rule schema built from  $\rightarrow$  and  $T$  in terms of  $\supset$ .

The above observation, of course, only holds for theories that can be defined in terms of a single predicate  $T$  and which include a connective  $\supset$  for which (41) is true.<sup>31</sup>

---

<sup>31</sup> And, of course, some encodings use more than one metalevel predicate; e.g. in §5.4 we introduce a second predicate  $V$  for which there is no equivalent of (41). For these systems we have rules for which no axiomatic equivalent is available. This does not, however, mean that ID-frameworks are necessarily less effective for structuring collections of theories; it just means that we have to be more sophisticated in the way we exploit (41). See, e.g., Matthews [1997b] for discussion of how we can do this by introducing an ‘extra’ layer between the ID-framework and the theory to be encoded.

*A further generalization of the deduction theorem*

We arrived at the open-ended (40) by observing that other axioms could be present. And as previously observed, no such generalization is possible with arbitrary rules, e.g., the deduction theorem does not hold in  $K$  (which requires extensions by rules, as opposed to axioms). However, a more refined analysis of the step case of the proof is possible, and this leads to a further generalization of our metatheorem.

In the step case we need precisely that the theory is closed under

$$\frac{A \supset B \quad A \supset C}{A \supset D}$$

for each instance of a basic rule

$$\frac{B \quad C}{D}.$$

In the case of *Det* (the only rule in  $HJ^\supset$ ) we can show this by a combination of *Det* and the *S* axiom.

Using our ID-framework we can explicitly formalize these statements as part of the deduction theorem itself, proving a further generalization. If we extend the notation of (37) with a parameter  $\Delta$  for rules, i.e.,

$$HJ[\Gamma, \Delta] \equiv I(K \cup S \cup \Gamma, Det \cup \Delta) \quad (42)$$

then for the base case we have

$$b \in HJ[\Gamma \cup \{a\}, \Delta] \rightarrow (a \supset b) \in HJ[\Gamma, \Delta] \quad (43)$$

and

$$(a \supset a) \in HJ[\Gamma, \Delta] \quad (44)$$

while for the step case

$$\begin{aligned} (d, b, c) \in \Delta &\rightarrow (a \supset b) \in HJ[\Gamma, \Delta] \\ &\rightarrow (a \supset c) \in HJ[\Gamma, \Delta] \rightarrow (a \supset d) \in HJ[\Gamma, \Delta]. \end{aligned} \quad (45)$$

The formulae (43) and (44) follow immediately for any  $HJ[\Gamma, \Delta]$ , but (45) isn't always true. Thus, our third deduction theorem has the form

$$(45) \rightarrow b \in HJ[\Gamma \cup \{a\}, \Delta] \rightarrow (a \supset b) \in HJ[\Gamma, \Delta], \quad (46)$$

which can be proved in the same way as (40). Note too that this metatheorem generalizes (40), since (40) is just (38) where  $\Delta$  is  $\emptyset$  and the antecedent, which is therefore true, has been removed.

The deduction theorem can even be further generalized, but doing so would take us too far afield. In [Basin and Matthews, 2000] we show how a further generalization of (46) can be specialized to modal logics that extend S4. This generalization allows us to prove

$$b \in S4[\Gamma \cup \{a\}] \rightarrow (\Box a \supset b) \in S4[\Gamma].$$

That is, in S4 we can prove a deduction theorem that allows us to reason under ‘boxed’ assumptions  $\Box a$ .

### 7.5 Admissible and derived rules

Our examples suggest that inductive definitions offer considerable power and simplicity in organizing metatheories. Each metatheorem states the conditions an extension has to satisfy for it to apply; so once proved, we need only check these conditions before making use of it. Most metatheorems require only that certain axioms and rules are available and therefore hold in all extensions with additional axioms and rules. Others depend on certain things being absent (e.g. rules that do not satisfy certain properties, in the case of the deduction theorem); in such cases, we can prove more restricted theorems that are still usable in appropriate extensions.

How does this kind of metatheory compare with what is possible in theorem provers supporting hierarchical theories? We begin by reviewing the two standard notions of proof-rules. Our definitions are those of Troelstra [1982, § 1.11.1] translated into our notation, where  $\mathcal{T}[\Gamma, \Delta]$  is a deductive system  $\mathcal{T}$  extended with sets of axioms  $\Gamma$  and rules  $\Delta$ , e.g. (42).

Fix a language of formulae. A *rule* is an  $n + 1$ -ary relation over formulae  $\langle F_1, \dots, F_n, F_{n+1} \rangle$  where the  $F_1, \dots, F_n$  are the *premises* and  $F_{n+1}$  the *conclusion*. A rule is *admissible* for  $\mathcal{T}$  iff

$$\vdash_{\mathcal{T}[\emptyset, \emptyset]} F_1 \rightarrow \dots \rightarrow \vdash_{\mathcal{T}[\emptyset, \emptyset]} F_n \rightarrow \vdash_{\mathcal{T}[\emptyset, \emptyset]} F_{n+1}, \quad (adm)$$

and *derivable* for  $\mathcal{T}$  iff

$$\forall \Gamma. \vdash_{\mathcal{T}[\Gamma, \emptyset]} F_1 \rightarrow \dots \rightarrow \vdash_{\mathcal{T}[\Gamma, \emptyset]} F_n \rightarrow \vdash_{\mathcal{T}[\Gamma, \emptyset]} F_{n+1}. \quad (der)$$

It follows immediately from the definitions that derivability implies admissibility; however, the converse does not always hold. It is easy to show that Troelstra’s definition of derivability is equivalent to that of Hindley and Seldin [1986]; i.e.  $\vdash_{\mathcal{T}[\{F_1, \dots, F_n\}, \emptyset]} F_{n+1}$ , and that if a rule is derivable it holds in all extensions of  $\mathcal{T}$  with new axioms and rules.

Whereas in  $\rightarrow$ -frameworks we can only prove derived rules, logical frameworks based on inductive definitions allow us to prove that rules are admissible, as well as reason about other kinds of rules not fitting the above categories. For example, the languages or deductive systems for the  $F_i$  can be different, like in the various

versions of the deduction theorem that we have formalized; our deduction theorems are neither derived nor admissible since their statements involve different deductive systems.

### 7.6 Problems with ID-frameworks

Our examples provide evidence that a framework based on inductive definitions can serve as an adequate foundation for carrying out metatheory in the proof theory tradition and can be used to structure metatheoretic development. However, some aspects of formal metatheory are more difficult than with an  $\rightarrow$ -framework. The most fundamental difficulty, and one that is probably already clear from our discussion in this section, is the way languages are encoded. This is quite primitive in comparison to what is possible in an  $\rightarrow$ -framework: for the propositional examples that we have treated here, the view of a language as a recursively enumerable class is direct and effective. But this breaks down for logics with quantifiers and other variable binding operators where the natural equivalence relation for syntax is no longer identity but equivalence under the renaming of bound variables ( $\alpha$ -congruence). We have shown (in §4.2) that language involving binding has a natural and direct treatment in an  $\rightarrow$ -framework as higher-order syntax. Nothing directly equivalent is available in an ID-framework; we are forced to build the necessary facilities ourselves.

Since the user must formalize many basic syntactic operations in an ID-framework, any treatment of languages involving variable binding operators will be more ‘primitive’ than what we get in an  $\rightarrow$ -framework, but how much more primitive is not clear. So far, most experience has been with *ad hoc* implementations of binding (e.g. [Matthews, 1992]) but approaches that are both more sophisticated and more modular are possible, such as the binding structures proposed by Talcott [1993], a generalization of de Bruijn indices as an algebra. As yet, we do not know how effective such notations are.

The other property of ID-frameworks that might be criticized is that they are biased towards Hilbert calculi, which are recognized to be difficult to use. But metatheorems, in particular the deduction theorem, can play an important role in making Hilbert calculi usable in practice. And, if a Hilbert style presentation is not suitable, it may be possible to exploit a combination of the deduction theorem and the intuitions of  $\rightarrow$ -frameworks to provide direct encodings of natural deduction [Matthews, 1997b]. The same provisos about lack of experience with effective notations for handling bindings apply here though, since this work only discusses the propositional case.

## 8 CONCLUSIONS

This chapter does not try to give a final answer to the question of what a logical framework might be. Rather it argues that the question is only meaningful in terms



of some particular set of requirements, and in terms of ‘goodness of fit’; i.e. the relationship between the properties of a proposed metalogic and logics we want to encode.

Our central theme has been the relationship between different kinds of deductive systems and their abstractions as metatheories or metalogics, which we can use to encode and work with instances of them. We have showed that a logic of minimal implication and universal quantification can be used to encode both the language and the proof-rules of a natural deduction or sequent calculus, and then described in detail a particular logic of this type,  $\lambda^p$ . As a contrast, we have also considered how (especially Hilbert) calculi can be abstracted as inductive definitions and we sketched a framework based on this view,  $FS_0$ . We then used the metatheoretic facilities that we get with an ID-framework like  $FS_0$  to explore the relationship between metatheory and object theory, especially in the context of structured collections of theories.

The simple binary distinction between ID and  $\rightarrow$ -frameworks, which we make for the sake of space and explication, of course does not describe the whole range of frameworks that have been proposed and investigated. It does, however, help to define the space of possibilities and current research into frameworks can mostly be categorized in its terms.

For instance, research into the problem of the ‘goodness of fit’ relation between the metalogic and the object logic, especially for  $\rightarrow$ -frameworks, can be separated into two parts. We have shown that natural deduction calculi for standard mathematical (i.e., classical or intuitionistic, first or higher-order) logics fit well into an  $\rightarrow$ -framework. But the further we diverge from standard, mathematical, logics into philosophical (i.e. modal, relevance, etc.) logic the more complex and artificial the encodings become. In order to encode modal logics, for instance, we might introduce either multiple-judgment encodings or take explicit account of a semantics via a labelling. The particular encodings that we have described here are only some among a range of possibilities that could be imagined. A more radical possibility, not discussed here, can be found in, e.g., [Matthews, 1997a], where the possibility of extending a framework directly with a modal ‘validity’ connective is explored. However we do not yet know what the practical limits of these approaches are.<sup>32</sup> A similar problem of goodness of fit is also encountered in ID-frameworks, where the ‘natural’ deductive systems are Hilbert calculi and the ‘obvious’ encodings of consequence style systems are impractically unwieldy. Matthews [1997b] suggests how we might encode pure ordinary consequence relation based systems in such a framework in a way that is more effective than, and at least as intuitive as, the ‘naive’ approach of using inductively defined classes.

The particular problems of substructural logics (e.g. linear or relevance logics) have been the subject of substantial research. The consequence relations associated with these logics are not ordinary and hence cannot be encoded using tech-

---

<sup>32</sup>This is essentially a practical, not a theoretical question, since, as we pointed out earlier, an  $\rightarrow$ -framework can be used as a Turing complete programming language, so with sufficient ingenuity any deductive system can be encoded.

niques such as those suggested in §6. While labelled or multiple judgment presentations of these logics in an  $\rightarrow$ -framework are possible, they seem to be unwieldy; i.e. they ‘fit’ particularly badly. An alternative approach has been explored where the framework itself is modified: minimal implication is replaced or augmented in the framework logic with a substructural or linear implication, which does permit a good fit. In Ishtiaq and Pym [1998] and Cervesato and Pfenning [1996], systems are presented that are similar to  $\lambda^p$  except that they are based on linear implication, which is used to encode variants of linear and other relevance logics. There is also work that, rather than employing either  $\rightarrow$  or ID-frameworks, attempts to combine features of both (e.g. McDowell and Miller [1997] and Despeyroux *et al.* [1996]).

In short, then, there are many possibilities and, in the end, no absolute solutions: the suitability of a particular logical framework to a particular circumstance depends on empirical as well as theoretical issues; i.e. before we can choose we have to decide on the range of object logics we envision formalizing, the nature of the metatheoretic facilities that we want, and the kinds of compromises that we are willing to accept.

David Basin

*University of Freiburg, Germany.*

Seán Matthews

*IBM Unternehmensberatung GmbH, Frankfurt, Germany.*

## BIBLIOGRAPHY

- [Abadi *et al.*, 1991] Martín Abadi, Luca Cardelli, Pierre-Louis Curien, and Jean-Jacques Lévy. Explicit substitutions. *J. Functional Programming*, 1:375–416, 1991.
- [Aczel, 1977] Peter Aczel. An introduction to inductive definitions. In Jon Barwise, editor, *Handbook of Mathematical Logic*. North-Holland, Amsterdam, 1977.
- [Avron *et al.*, 1992] Arnon Avron, Furio Honsell, Ian Mason, and Robert Pollack. Using typed lambda calculus to implement formal systems on a machine. *J. Auto. Reas.*, 9:309–352, 1992.
- [Avron *et al.*, 1998] Arnon Avron, Furio Honsell, Marino Miculan, and Cristian Paravano. Encoding modal logics in logical frameworks. *Studia Logica*, 60(1):161–208, 1998.
- [Avron, 1990] Arnon Avron. Gentzenizing Shroeder-Heister’s natural extension of natural deduction. *Notre Dame Journal of Formal Logic*, 31:127–135, 1990.
- [Avron, 1991] Arnon Avron. Simple consequence relations. *Inform. and Comput.*, 92:105–139, 1991.
- [Avron, 1992] Arnon Avron. Axiomatic systems, deduction and implication. *J. Logic Computat.*, 2:51–98, 1992.
- [Barendregt, 1984] Henk Barendregt. *The Lambda Calculus: Its Syntax and Semantics*. North-Holland, Amsterdam, 2nd (revised) edition, 1984.
- [Barendregt, 1991] Henk Barendregt. Introduction to generalized type systems. *J. Functional Programming*, 2:125–154, 1991.
- [Barendregt, 1992] Henk Barendregt. Lambda calculi with types. In Samson Abramsky, Dov Gabbay, and Tom S. E. Maibaum, editors, *Handbook of Logic in Computer Science*, volume 2. Oxford University Press, 1992.
- [Basin and Matthews, 2000] David Basin and Sean Matthews. Structuring metatheory on inductive definitions. *Information and Computation*, 162(1–2), October/November 2000.
- [Basin *et al.*, 1997a] David Basin, Seán Matthews, and Luca Viganò. Labelled propositional modal logics: theory and practice. *J. Logic Computat.*, 7:685–717, 1997.

- [Basin *et al.*, 1997b] David Basin, Seán Matthews, and Luca Viganò. A new method for bounding the complexity of modal logics. In Georg Gottlob, Alexander Leitsch, and Daniele Mundici, editors, *Proc. Kurt Gödel Colloquium, 1997*, pages 89–102. Springer, Berlin, 1997.
- [Basin *et al.*, 1998a] David Basin, Seán Matthews, and Luca Viganò. Labelled modal logics: Quantifiers. *J. Logic, Language and Information*, 7:237–263, 1998.
- [Basin *et al.*, 1998b] David Basin, Seán Matthews, and Luca Viganò. Natural deduction for non-classical logics. *Studia Logica*, 60(1):119–160, 1998.
- [Boolos, 1993] George Boolos. *The Logic of Provability*. Cambridge University Press, 1993.
- [Boyer and Moore, 1981] Robert Boyer and J. Strother Moore. *A Computational Logic*. Academic Press, New York, 1981.
- [Bull and Segerberg, 1984] Robert Bull and Krister Segerberg. Basic modal logic. In Gabbay and Guenther [1983–89], chapter II.1.
- [Cervesato and Pfenning, 1996] Iliano Cervesato and Frank Pfenning. A linear logical framework. In *11th Ann. Symp. Logic in Comp. Sci.* IEEE Computer Society Press, 1996.
- [Church, 1940] Alonzo Church. A formulation of the simple theory of types. *J. Symbolic Logic*, 5:56–68, 1940.
- [Davis, 1989] Martin Davis. Emil Post’s contributions to computer science. In *Proc. 4th IEEE Ann. Symp. Logic in Comp. Sci.* IEEE Computer Society Press, 1989.
- [de Bruijn, 1972] Nicolas G. de Bruijn. Lambda calculus notation with nameless dummies, a tool for automatic formula manipulation, with application to the Church-Rosser theorem. *Indagationes Mathematicae*, 34:381–392, 1972.
- [de Bruijn, 1980] Nicolas G. de Bruijn. A survey of the project Automath. In J. R. Hindley and J. P. Seldin, editors, *To H. B. Curry: Essays in Combinatory Logic, Lambda Calculus and Formalism*, pages 579–606. Academic Press, New York, 1980.
- [Despeyroux *et al.*, 1996] Joëlle Despeyroux, Frank Pfenning, and Carsten Schürmann. Primitive recursion for higher-order abstract syntax. Technical Report CMU-CS-96-172, Dept. of Computer Science, Carnegie Mellon University, September 1996.
- [Dummett, 1978] Michael Dummett. The philosophical basis of intuitionistic logic. In *Truth and other enigmas*, pages 215–247. Duckworth, London, 1978.
- [Fagin *et al.*, 1992] Ronald Fagin, Joseph Y. Halpern, and Moshe Y. Vardi. What is an inference rule? *J. Symbolic Logic*, 57:1018–1045, 1992.
- [Feferman, 1990] Solomon Feferman. Finitary inductive systems. In *Logic Colloquium ’88*, pages 191–220. North-Holland, Amsterdam, 1990.
- [Felty, 1989] Amy Felty. *Specifying and Implementing Theorem Provers in a Higher Order Programming Language*. PhD thesis, University of Pennsylvania, 1989.
- [Felty, 1991] Amy Felty. Encoding dependent types in an intuitionistic logic. In Huet and Plotkin [1991].
- [Gabbay and Guenther, 1983–89] Dov Gabbay and Franz Guenther, editors. *Handbook of Philosophical Logic, vol. I–IV*. Reidel, Dordrecht, 1983–89.
- [Gabbay, 1996] Dov Gabbay. *Labelled Deductive Systems, vol. 1*. Clarendon Press, Oxford, 1996.
- [Gardner, 1995] Philippa Gardner. Equivalences between logics and their representing type theories. *Math. Struct. in Comp. Science*, 5:323–349, 1995.
- [Gentzen, 1934] Gerhard Gentzen. Untersuchen über das logische Schließen. *Math. Z.*, 39:179–210, 405–431, 1934.
- [Gödel, 1931] Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatsh. Math.*, 38:173–198, 1931.
- [Goguen and Burstall, 1992] Joseph A. Goguen and Rod M. Burstall. Institutions: Abstract model theory for specifications and programming. *J. Assoc. Comput. Mach.*, pages 95–146, 1992.
- [Gordon and Melham, 1993] Michael J. Gordon and Tom Melham. *Introduction to HOL: A Theorem Proving Environment for Higher Order Logic*. Cambridge University Press, Cambridge, 1993.
- [Gordon *et al.*, 1979] Michael J. Gordon, Robin Milner, and Christopher P. Wadsworth. *Edinburgh LCF: A Mechanized Logic of Computation*. Springer, Berlin, 1979.
- [Griswold, 1981] Ralph E. Griswold. A history of the Snobol programming languages. In Wexelblat [1981], pages 601–660.
- [Hacking, 1979] Ian Hacking. What is logic? *J. Philosophy*, 76(6), 1979.
- [Harper *et al.*, 1993] Robert Harper, Furio Honsell, and Gordon Plotkin. A framework for defining logics. *J. Assoc. Comput. Mach.*, 40:143–184, 1993.

- [Hindley and Seldin, 1986] J. Roger Hindley and Jonathan P. Seldin. *Introduction to Combinators and  $\lambda$ -Calculus*. Cambridge University Press, Cambridge, 1986.
- [Howard, 1980] William Howard. The formulas-as-types notion of construction. In J. P. Seldin and J. R. Hindley, editors, *To H. B. Curry: Essays on Combinatory Logic, Lambda-Calculus, and Formalism*. Academic Press, New York, 1980.
- [Huet and Plotkin, 1991] Gérard Huet and Gordon Plotkin, editors. *Logical Frameworks*. Cambridge University Press, Cambridge, 1991.
- [Ishtiaq and Pym, 1998] Samin Ishtiaq and David Pym. A relevant analysis of natural deduction. *J. Logic Comput.*, 8:809–838, 1998.
- [Kleene, 1952] Stephen C. Kleene. *Introduction to Metamathematics*. North-Holland, Amsterdam, 1952.
- [Kripke, 1963] Saul A. Kripke. Semantical analysis of modal logic I: normal propositional modal logic. *Z. Math. Logik Grundlag. Math.*, 8:67–96, 1963.
- [Martí-Oliet and Meseguer, 2002] Narciso Martí-Oliet and José Meseguer. Rewriting logic as a logical and semantic framework. In *Handbook of Philosophical Logic*. second edition, 2002.
- [Matthews et al., 1993] Seán Matthews, Alan Smaill, and David Basin. Experience with  $FS_0$  as a framework theory. In Gérard Huet and Gordon Plotkin, editors, *Logical Environments*. Cambridge University Press, Cambridge, 1993.
- [Matthews, 1992] Seán Matthews. *Metatheoretic and Reflexive Reasoning in Mechanical Theorem Proving*. PhD thesis, University of Edinburgh, 1992.
- [Matthews, 1994] Seán Matthews. A theory and its metatheory in  $FS_0$ . In Dov Gabbay, editor, *What is a Logical System?* Clarendon Press, Oxford, 1994.
- [Matthews, 1996] Seán Matthews. Implementing  $FS_0$  in Isabelle: adding structure at the metalevel. In Jacques Calmet and Carla Limongelli, editors, *Proc. Disco'96*. Springer, Berlin, 1996.
- [Matthews, 1997a] Seán Matthews. Extending a logical framework with a modal connective for validity. In Martín Abadi and Takayasu Ito, editors, *Proc. TACS'97*. Springer, Berlin, 1997.
- [Matthews, 1997b] Seán Matthews. A practical implementation of simple consequence relations using inductive definitions. In William McCune, editor, *Proc. CADE-14*. Springer, Berlin, 1997.
- [McCarthy, 1981] John McCarthy. History of Lisp. In Wexelblat [1981], pages 173–197.
- [McDowell and Miller, 1997] Raymond McDowell and Dale Miller. A logic for reasoning with higher-order abstract syntax. In *Proc. 12th IEEE Ann. Symp. Logic in Comp. Sci.*, pages 434–446. IEEE Computer Society Press, 1997.
- [Meseguer, 1989] José Meseguer. General logics. In Heinz-Dieter Ebbinghaus, J. Fernandez-Prida, M. Garrido, D. Lascar, and M. Rodríguez Artalejo, editors, *Logic Colloquium, '87*, pages 275–329. North-Holland, 1989.
- [Nederpelt et al., 1994] Rob P. Nederpelt, Herman J. Geuvers, and Roel C. de Vrijer, editors. *Selected papers on Automath*. Elsevier, Amsterdam, 1994.
- [Ohlbach, 1993] Hans-Jürgen Ohlbach. Translation methods for non-classical logics: an overview. *Bulletin of the IGPL*, 1:69–89, 1993.
- [Owre et al., 1995] Sam Owre, John Rushby, Natarajan Shankar, and Friedrich von Henke. Formal verification for fault-tolerant architectures: Prolegomena to the design of PVS. *IEEE Trans. Software Eng.*, 21:107–125, 1995.
- [Paulson, 1994] Lawrence C. Paulson. *Isabelle: A Generic Theorem Prover*. Springer, Berlin, 1994.
- [Pfenning, 1996] Frank Pfenning. The practice of logical frameworks. In Helene Kirchner, editor, *Proc. CAAP'96*. Springer, Berlin, 1996.
- [Pfenning, 2000] Frank Pfenning. Structural cut elimination I. intuitionistic and classical logic. *Information and Computation*, 157(1–2), March 2000.
- [Pollack, 1994] Randy Pollack. *The Theory of LEGO: A Proof Checker for the Extended Calculus of Constructions*. PhD thesis, University of Edinburgh, 1994.
- [Post, 1943] Emil Post. Formal reductions of the general combinatorial decision problem. *Amer. J. Math.*, 65:197–214, 1943.
- [Prawitz, 1965] Dag Prawitz. *Natural Deduction*. Almqvist and Wiksell, Stockholm, 1965.
- [Prawitz, 1971] Dag Prawitz. Ideas and results in proof theory. In J. E. Fensted, editor, *Proc. Second Scandinavian Logic Symp.*, pages 235–307. North-Holland, Amsterdam, 1971.
- [Pym and Wallen, 1991] David J. Pym and Lincoln Wallen. Proof-search in the  $\lambda\Pi$ -calculus. In Huet and Plotkin [1991], pages 309–340.

- [Schroeder-Heister, 1984a] Peter Schroeder-Heister. Generalised rules for quantifiers and the completeness of the intuitionistic operators  $\&$ ,  $\vee$ ,  $\supset$ ,  $\perp$ ,  $\forall$ ,  $\exists$ . In M. M. Richter et al., editors, *Computation and proof theory*. Springer, Berlin, 1984.
- [Schroeder-Heister, 1984b] Peter Schroeder-Heister. A natural extension of natural deduction. *J. Symbolic Logic*, 49:1284–1300, 1984.
- [Scott, 1974] Dana Scott. Rules and derived rules. In S. Stenlund, editor, *Logical Theory and Semantical Analysis*, pages 147–161. Reidel, Dordrecht, 1974.
- [Simpson, 1992] Alex K. Simpson. Kripke semantics for a logical framework. In *Proc. Workshop on Types for Proofs and Programs*, Båstad, 1992.
- [Smullyan, 1961] Raymond Smullyan. *Theory of Formal Systems*. Princeton University Press, 1961.
- [Steele Jr. and Gabriel, 1996] Guy L. Steele Jr. and Richard P. Gabriel. The evolution of Lisp. In Thomas J. Bergin and Richard G. Gibson, editors, *History of Programming Languages*, pages 233–330. ACM Press, New York, 1996.
- [Sundholm, 1983] Göran Sundholm. Systems of deduction. In Gabbay and Guentner [1983–89], chapter I.2.
- [Sundholm, 1986] Göran Sundholm. Proof theory and meaning. In Gabbay and Guentner [1983–89], chapter III.8.
- [Talcott, 1993] Carolyn Talcott. A theory of binding structures, and applications to rewriting. *Theoret. Comp. Sci.*, 112:99–143, 1993.
- [Troelstra, 1982] A. S. Troelstra. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Springer, Berlin, 1982.
- [van Benthem, 1984] Johan van Benthem. Correspondence theory. In Gabbay and Guentner [1983–89], chapter II.4.
- [Wexelblat, 1981] Richard L. Wexelblat, editor. *History of Programming Languages*. Academic Press, New York, 1981.



GÖRAN SUNDHOLM

## PROOF THEORY AND MEANING

Dedicated to Stig Kanger on the occasion of his 60th birthday

The meaning of a sentence determines how the truth of the proposition expressed by the sentence may be proved and hence one would expect proof theory to be influenced by meaning-theoretical considerations. In the present chapter we consider a proposal that also reverses the above priorities and determines meaning in terms of proof. The proposal originates in the criticism that Michael Dummett has voiced against a realist, truth-theoretical, conception of meaning and has been developed largely by him and Dag Prawitz, whose normalisation procedures in technical proof theory constitute the main technical basis of the proposal.

In a subject not more than 20–30 years old, and where much work is currently being done, any survey is bound to be out of date when it appears. Accordingly I have attempted not to give a large amount of technicalities, but rather to present the basic underlying themes and guide the reader to the ever-growing literature. Thus the chapter starts with a general introduction to meaning-theoretical issues and proceeds with a fairly detailed presentation of Dummett's argument against a realist, truth-conditional, meaning theory. The main part of the chapter is devoted to a consideration of the alternative proposal using 'proof-conditions', instead of truth-conditions, as the key concept. Finally, the chapter concludes with an introduction to the type theory of Martin-Löf.

I am indebted to Professors Dummett, Martin-Löf and Prawitz, and to my colleague Mr. Jan Lemmens, for many helpful conversations on the topics covered herein and to the editors for their infinite patience. Dag Prawitz and Albert Visser read parts of the manuscript and suggested many improvements.

### 1 THEORIES OF MEANING, MEANING THEORIES AND TRUTH THEORIES

A *theory of meaning* gives, one might not unreasonably expect, a general account of, or view on, the very concept of meaning: what it is and how it functions. Such theories *about* meaning, however, do not hold undisputed rights to the *appellation*; in current philosophy of language one frequently encounters discussions of theories of meaning for particular languages. Their task is to specify the meaning of all the sentences of the language in question. Following Peacocke [1981] I shall use the term 'meaning theory' for the latter, language-relative, sort of theory and reserve 'theory of meaning' for

the former. Terminological confusion is, fortunately, not the only connection between meaning theories and theories of meaning. On the contrary, the main reason for the study and attempted construction of meaning theories is that one hopes to find a correct theory of meaning through reflection on the various desiderata and constraints that have to be imposed on a satisfactory meaning theory. The study of meaning theories, so to speak, provides the *data* for the theory of meaning. In the present chapter we shall mainly treat meaning theories and some of their connection with (technical) proof theory and, consequently, we shall only touch on the theory of meaning in passing. (On the other hand the whole chapter can be viewed as a contribution to the theory of meaning.)

There is, since Frege, a large consensus that the sentence, rather than the word, is the primary bearer of (linguistic) meaning. The sentence is the least unit of language that can be used to *say* anything. Thus the theory of meaning directs that sentence-meaning is to be central in meaning theories and that word-meaning is to be introduced derivatively: the meaning of a word is the way in which the word contributes to the meaning of the sentences in which it occurs. It is natural to classify the sentences of a language according to the sort of linguistic act a speaker would perform through an utterance of the sentence in question, be it an assertion, a question or a command. Thus, in general, the meaning of a sentence seems to comprise (at least) two elements, because to know the meaning of — in order to understand an utterance of — the sentence in question one would have to know, first to what category the sentence belongs, i.e. one would have to know what sort of linguistic act that would be performed through an utterance of the sentence, and secondly one would have to know the *content* of the act.

This diversity of sentence-meaning, together with the idea that word-meaning is to be introduced derivatively (as a way of contributing to sentence-meaning), poses a certain problem for the putative meaning-theorist. If sentences from different categories have different *kinds* of meaning, it appears that the meaning of a word will vary according to the category of the sentences in which it occurs: uniform word-meanings are ruled out. But this is unacceptable as anyone familiar with a dictionary knows. The word 'door', say, has the same meaning in the three sentences 'Is the door open?', 'The door is open.', and 'Open the door!'. This *prima facie* difficulty is turned into a tool for investigating what internal structure ought to be imposed on a satisfactory meaning theory.

A meaning theory will have to comprise at least two parts: the *theory of sense* and the *theory of force*. The task of the latter is to identify the sort of act performed through an utterance of a sentence and the former has to specify the content of the acts performed. In order to secure the uniformity of word meaning the theory of sense has to be formulated in terms of some *key concept*, in terms of which the content of all sentences is to be given,



and the theory of force has to provide uniform, general, principles relating speech act to content. The meaning of a word is then taken as the way in which the word contributes to the content of the sentences in which it occurs (as given by the key concept in the theory of sense).

The use of such a notion of key concept also allows the meaning theories to account for certain (iterative) unboundedness-phenomena in language, e.g. that whenever  $A$  and  $B$  are understood sentences, then also ' $A$  and  $B$ ' would appear to be meaningful. This is brought under control in the meaning theory by expressing the condition for the application of the key concept  $P$  to ' $A$  and  $B$ ' in terms of  $P$  applied to  $A$  and  $P$  applied to  $B$ .

The most popular candidate for a key concept has undeniably been *truth*: the content of a sentence is given by its 'truth-condition'. One can, indeed, find many philosophers who have subscribed to the idea that meaning is to be given in terms of truth. Examples would be Frege, Wittgenstein, Carnap, Quine and Montague. It is doubtful, however, if they would accept that the way in which truth serves to specify meaning is as a key concept in a meaning theory (that is articulated into sense and force components respectively). Such a conception of the relation between meaning and truth has been advocated by Donald Davidson, who, in an important series of papers, starting with [1967], and now conveniently collected in his [1984], has proposed and developed the idea that meaning is to be studied via meaning theories. Davidson is quite explicit on the role of truth. It is going to take its rightful place within the meaning theory in the shape of a truth theory in the sense of Tarski [1956, Ch. VIII]. Tarski showed, for a given formal language  $L$ , how to define a predicate ' $\text{True}_L(x)$ ' such that for every sentence  $S$  of  $L$  it is provable from the definition that

$$(1) \quad \text{True}_L(\bar{S}) \text{ iff } f(S).$$

Here ' $\bar{S}$ ' is a *name* of, and  $f(S)$  a *translation* of, the object-language sentence  $S$  in the language of the meta-theory (= the theory in which the truth definition is given and where all instances of (1) must hold). *Using* the concept of meaning (in the guise of 'translation' from object-language to meta-language) Tarski gave a precise definition of what it is for a sentence of  $L$  to be true. Davidson reverses the theoretical priorities. Starting with a *truth theory* for  $L$ , that is a theory the language of which contains  $\text{True}_L(x)$  as a primitive, and where for each sentence  $S$  of  $L$

$$(2) \quad \text{True}_L(\bar{S}) \text{ iff } p.$$

holds for some sentence  $p$  of the language of the truth theory, he wanted to extract meaning from truth. Simply to consider an arbitrary truth theory will not do not capture meaning, though. It is certainly true that

$$(3) \quad \overline{\text{Snow is white}} \text{ is true-in-English iff snow is white}$$

but, unquestionably and unfortunately, it is equally true that

(4)  $\overline{\text{Snow is white}}$  is true-in-English iff grass is green

and the r.h.s. of (4) could not possibly by any stretch of imagination be said to provide even a rough approximation of the meaning of the English sentence

Snow is white.

Furthermore, a theory that had all instances of (2) as axioms would be unsatisfactory also in that it used infinitely many unrelated axioms; the theory would, it is claimed, be ‘unlearnable’.

Thus one might attempt to improve on the above simple-minded (2) by considering truth theories that are formulated in a meta-language that contains the object-language and that give their ‘*T*-theories’ (the instances of (2)), not as axioms, but as derivable from homophonic recursion clauses, e.g.

(5) for all  $\bar{A}$  and  $\bar{B}$  of L,  
 $\text{True}_L(\overline{\bar{A} \text{ and } \bar{B}} \text{ iff } \text{True}_L(\bar{A} \text{ and } \text{True}_L(\bar{B}))$

and

(6) for all  $\bar{A}$  of L,  
 $\text{True}_L(\overline{\text{not-}\bar{A}} \text{ iff not-True}_L(\bar{A}))$ .

Here one uses the word mentioned in the sentence on the l.h.s. when giving the condition for its truth on the r.h.s.; cf. the above remarks on the iterative unboundedness phenomena.

The treatment of quantification originally used Tarski’s device of ‘satisfaction relative to assignment by sequences’, where, in fact, one does not primarily recur on truth, but on satisfaction, and where truth is defined as satisfaction by all sequences. The problem which Tarski solved by the use of the sequences and the auxiliary notion of satisfaction was how to capture the right truth condition for ‘everything is *A*’ even though the object language does not contain a name for everything to be considered in the relevant domain of quantification. Another satisfactory solution which goes back to Frege, would be to use quantification over finite extensions  $L^+$  of  $L$  by means of new names. The interested reader is referred to [Evans, 1977, Section 2] or to [Davies, 1981, Chapter VI] for the (not too difficult) technicalities. A very extensive and careful canvassing of various alternative approaches to quantificational truth-theories is given by Baldwin [1979]. If we bypass the problem solve by Tarski and consider, say, the language of arithmetic, where the problem does not arise as the language contains a numeral for each element of the intended domain of quantification the universal-quantifier clause would be

for all  $\bar{A}$  of L,

$$(7) \quad \text{True}_L(\overline{\text{for every number } x, A(x)}) \text{ iff for every numeral } \bar{k}, \\ \text{True}_L(\overline{A(\bar{k}/x)}).$$

(here ' $A(\bar{k}/x)$ ' indicates the result of substituting the numeral  $k$  for the variable  $x$ .)

Unfortunately it is still not enough to consider these homophonic, finitely axiomatised truth theories in order to capture meaning. The basic clauses of a homophonic truth theory will have the form, say,

$$(8) \quad \text{for any name } \bar{t} \text{ of L,} \\ \text{True}_L(\overline{t \text{ is red}}) \text{ iff whatever } t \text{ refers to is red.}$$

If we now change this clause to

$$(9) \quad \text{for any } \bar{t} \text{ in L,} \\ \text{True}'_L(\overline{t \text{ is red}}) \text{ iff whatever } t \text{ refers to is red and grass is green}$$

and keep homophonic clauses for  $\text{True}'_L$  with respect to 'and' 'not', etc., the result will still be a finitely axiomatised and correct ('true') truth theory for L. We could equally well have chosen any other true contingent sentence instead of 'grass is green'. Seen from the perspective of 'real meaning' the truth condition of the primed theory is best explained as

$$(10) \quad \text{True}'_L(\bar{S}) \text{ iff } S \text{ and grass is green.}$$

The fact that a true, finitely axiomatised, homophonic truth-theory does not necessarily provide truth conditions that capture meaning was first observed by Foster and Loar in 1976. Various remedies and refinements of the original Davidsonian programme have been explored. We shall briefly consider an influential proposal due to John McDowell [1976; 1977; 1978].

The above attempts to find a meaning theory via truth start with a (true) truth theory and go on to seek further constraints that have to be imposed in order to capture meaning. McDowell, on the other hand, reverses this strategy and starts by considering a satisfactory theory of sense. Such a theory has to give content-ascriptions to the sentences  $\bar{S}$  of the language L, say in the general form

$$(11) \quad \bar{S} \text{ is } Q \text{ iff } p,$$

where  $p$  is a sentence of the meta-language that gives the content of  $\bar{S}$ , and, furthermore, the theory has to interact with a theory of force in such a way that the interpreting descriptions, based on the contents as assigned in (11), do in fact make sense of what speakers say and do when they utter sentences containing  $\bar{S}$ . A meaning theory, and thus also its theory

of sense, is part of an overall theory of understanding, the task of which is to make sense of human behaviour (and not just these speech-acts). If the theory of sense can serve as a content-specifying core in such a general theory, then (11) guarantees that the predicate  $Q$  is (co-extensional with) truth. But not only that is true; the pathological truth-theories that were manufactured for use in the Foster–Loar counter-examples are ruled out from service as theories of sense because their use would make the meaning theory issue incomprehensible, or outright false, descriptions of what people do. A theory of sense which uses a pathological truth-theory does not make sense. Thus we see that while an adequate theory of sense will be a truth theory, the opposite is false: not every truth theory for a language will be a theory of sense for the language.

In conclusion of the present section let us note the important fact that the Tarski homophonic truth-theories are completely neutral with respect to the underlying logic. The  $T$ -theorems are derivable from the basic homophonic recursion clauses using intuitionistic logic only (in fact even minimal logic will do).

No attempt has been made in the present section to achieve either completeness or originality. The very substantial literature on the Davidsonian programme is conveniently surveyed in two texts, [Platts, 1979] and [Davies, 1981], where the latter pays more attention to the (not too difficult) technicalities. Many of the important original papers are included in [Evans and McDowell, 1976], with an illuminating introduction by the editors, and [Platts, 1980], while mention has already been made of [Davidson, 1984]’s collection of essays.

## 2 INTERMEZZO: CLASSICAL TRUTH AND SEQUENT CALCULI

(Intended for readers of the method ‘semantic tableaux’, cf. Section 6 of Hodges’ chapter or section 3 of Sundholm’s chapter, both in Volume 1 of this *Handbook*.)

It is by now well-known that perhaps the easiest way to prove the completeness of classical predicate logic is to search systematically for a counter-model (or, more precisely, a falsifying ‘semi-valuation’, or ‘model set’) to the formula, or sequent, in question. This systematic search proceeds according to certain rules which are directly read off as necessary conditions from the relevant semantics. For instance, in order to falsify  $\forall xA(x) \rightarrow B$ , one needs to verify  $\forall xA(x)$  and falsify  $B$ , and in order to verify  $\forall xA(x)$  one has to verify  $A(t)$  for every  $t$ , etc. Thus the rules for falsification, in fact, also concern rules for verification and *vice versa* (consider verification of, e.g.  $\neg B$ ), and for each logical operator there will be two rules regulating the systematic search for a counter-model, one for verification and one for falsification. These rules turn out to be identical with Gentzen’s [1934–1935] left and

right introduction rules for the same operators. In some cases the search needs to take alternatives into account, e.g.  $A \rightarrow B$  is verified by falsifying  $A$  or verifying  $B$ . Thus one has two possibilities. The failure of the search along a possibility is indicated by that the rules would force one to assign both truth and falsity to one and the same formula. This corresponds, of course, to the axioms of Gentzen's sequent calculi. This method, where failure of existence of counter-models is equivalent to existence of a sequent calculus proof-tree, was discovered independently by Beth, Hintikka, Kanger and Schütte in the 1950s and a brilliant exposition can be found in [Kleene, 1967, Chapter VI], whereas [Smullyan, 1968] is the canonical reference for the various ways of taking the basic insight into account. Prawitz [1975] is a streamlined development of the more technical aspects which provides an illuminating answer to the question as to why the rules that generate counter-models turn out to be identical with the sequent calculus rules. There one also finds a good introduction to the notion of semi-valuation which has begun to play a role in recent investigations into the semantics of natural language (cf. [van Benthem and van Eijck, 1982] for an interesting treatment of the connection between recent work on 'partial structures' in the semantics of natural language and the more proof-theoretical notions that derive from the 'backwards' completeness proofs).

These semantical methods for proving completeness also lend themselves to immediate proof-theoretical applications. The *Cut-free* sequent calculus is complete, but cut is a sound rule. Hence it is derivable. A connection with the topic of our chapter is forged by reversing these proof-theoretic uses of semantical methods. Instead of proving the completeness via semantics, one could start by postulating the completeness of a cut-free formalism, and *read off a semantic* from the left and right introduction rules. (Proof theory determines meaning.) Such an approach was suggested by Hacking [1979] in an attempt towards a criterion for logical constanhood. Unfortunately, his presentation is marred by diverse technical infelicities (cf. [Sundholm, 1981]), and the problem still remains open how to find a workable proposal along these lines.

### 3 DUMMETT'S ARGUMENT AGAINST A TRUTH-CONDITIONAL VIEW ON MEANING

In the present section I attempt to set out one version of an argument due to Michael Dummett to the effect that truth cannot adequately serve as a key concept in a satisfactory meaning theory. Dummett has presented his argument in many places (cf. the note at the end of the section) and the presentation I offer is not to be attributed to him. In particular, the emphasis on manifestation that can be found in the present version of Dummett's argument I have come to appreciate through the writings of Colin McGinn

[1980] and Crispin Wright [1976]. Dummett's most forceful exposition is still his [1976], which will be referred to as "WTM2".

Dummett's views on the role and function of meaning theories are only in partial agreement with those presented in section 1. The essential difference consists mainly in the strong emphasis on what it is to know a language that can be found in Dummett's writings, and as a consequence his meaning theories are firmly cast in an epistemological mould: "questions about meaning are best interpreted as questions of understanding: a dictum about what the meaning of a sentence consists in must be construed as a thesis about what it is to *know* its meaning" (WTM2, p. 69). The task of the meaning theorist is to give a theoretical (propositional) representation of the complex practical capacity one has when one knows how to speak a language. The knowledge that a speaker will have of the propositions that constitute the theoretical representation in question will, in the end, have to be *implicit* knowledge. Indeed, one cannot demand that a speaker should be able to articulate explicitly; those very principles that constitute the theoretical representation of his practical mastery. Thus a meaning theory that gives such a theoretical representation must also comprise a part that would state what it is to know the other parts implicitly.

The inner structure of a meaning theory that could serve the aims of Dummett will have to be different from the simple bipartite version considered in section 1. Dummett's meaning theories are to be structured as follows. There is to be (ia) a core *theory of semantic value*, which states the condition for the application of the key concept to the sentences of the language, and, furthermore, there must be (ii) a *theory of force*, as before. In between these two, however, there must be (ib) a *theory of sense*, whose task it is to state what it is to know what is stated in the theory of semantic value, i.e. what it is to know the condition for the application of the key concept to a sentence. Thus the theory of sense in the proposals from section 1 does not correspond to the theory of sense<sub>D</sub> — 'D' for Dummett — but to the theory of semantic value. (The Fregean origin of Dummett's tripartite structure should be obvious. For further elaboration cf., his [1981].) The theory of sense<sub>D</sub> has no matching part in the theories from section 1. The corresponding match is as follows:

<i>Dummett</i>	<i>(Davidson-)McDowell</i>
(ia) Theory of semantic value (applies key concept to sentences)	(i) Theory of sense
(iv) Theory of sense <sub>D</sub> (states what it is to know the theory of semantic value)	
(ii) The theory of force	(ii) The theory of force

This difference is what lies at the heart of the matter in the discussion between Dummett and McDowell of whether a theory of meaning ought to be ‘modest’ or ‘fullblooded’ (cf. [McDowell, 1977, Section X]: should one demand that the meaning theory must give a link-up with practical capacities *independently of, and prior to, the theory of force?*

One should also note here that the right home for the theory of sense<sub>D</sub> is not quite clear. Here I have made it part of the meaning theory. It could perhaps be argued that a statement of wherein knowledge of meaning consists is something that had better be placed within a theory of meaning rather than in a meaning theory. Dummett himself does not draw the distinction between meaning theories and theories of meaning and one can, it seems to me, find traces of *both* notions in what Dummett calls a ‘theory of meaning’.

Dummett’s argument against the truth-theoretical conception of meaning makes essential use of the assumption that the meaning theories must contain a theory of sense<sub>D</sub>, which Dummett explicates in terms of how it can be manifested: since the knowledge is implicit, possession thereof can be construed only in terms of how one manifests that knowledge. Furthermore, this implicit knowledge of meaning, or more precisely, of the condition for applying the key concept to individual sentences, must be *fully* manifested in use. This is Dummett’s transformation of Wittgenstein’s dictum that meaning is use. Two reasons can be offered (cf. [McGinn, 1980, p. 20]). First, knowledge is one of many propositional attitudes and these are, in general, only attributed to agents on the basis of how they are manifested. Secondly, and more importantly, we are concerned with (implicit) knowledge of *meaning* and meaning is, *par excellence*, a vehicle of (linguistic) communication. If there were some components of the implicit knowledge that did not become fully manifest in use, they could not matter for communication and so they would be superfluous.

It was already noted above that the Tarskian truth-theories are completely neutral with respect to the logical properties of truth. What laws are obeyed is determined by the logic that is applied in the meta-theory, whereas the *T*-clauses themselves offer no information on this point. Dummett’s argument is brought to bear not so much against Tarskian truth as against the possibility that the key concept could be ‘recognition-transcendent’. Classical, *bivalent* truth is characterised by the law of bivalence that every sentence is either true or false independently of our capacity to decide, or find out, whichever is the case. Thus, in general, the truth-conditions will be such that they can obtain without us recognising that they do. There are a number of critical cases which produce such undecidable truth-conditions. (It should be noted that ‘undecidable’ is perhaps not the best choice here with its connotations from recursive function theory.) Foremost among these is undoubtedly quantification over infinite or unbounded domains. Fermat’s last theorem and the Reimann hypothesis are both famous examples from

mathematics and their form is purely universal  $\forall xA(x)$ , with decidable matrix  $A(x)$ . An existential example would be, say, ‘Somewhere in the universe there is a little green stone-eater’. Other sorts of examples are given by, respectively, counterfactual conditionals and claims about sentience in others, e.g. ‘Ronald Reagan is in pain’. A fourth class is given by statements about (remote) past and future time, e.g. ‘A city will be built here in a thousand years’, or ‘Two seconds before Brutus stabbed Casare thirty-nine geese cackled on the Capitol’.

The knowledge one has of how to apply the key concept cannot in its entirety be statable, explicit knowledge and so the theory of sense<sub>D</sub> will have to state, for at least some sentences, how one manifests knowledge of the condition for applying the key concept to them, in ways other than stating what one knows explicitly. Let us call the class of these ‘the non-statable fragment’. (Questions of the ‘division of linguistic labour’ may arise here. Is *the* fragment necessarily unique? Cf. [McGinn, 1980, p. 22].)

Assume now for a *reductio* that bivalent, possibly recognition-transcendent, truth-conditions can serve as key concept in a (Dummettian) meaning theory. Thus the theory of sense<sub>D</sub> has to state how one fully manifests knowledge of possibly recognition-transcendent truth-conditions. The ‘possibly’ can be removed: there are sentences in the non-statable fragment with undecidable truth-conditions. In order to see this, remember the four classes of undecidable sentences that were listed above. Demonstrably, undecidable sentences are present in the language and they must be present already in the non-statable fragment, because “the existence of such sentences cannot be due solely to the occurrence of sentences introduced by purely verbal explanations: a language all of whose sentences were decidable would continue to have this property when enriched by expressions so introduced” (WTM2, p. 81). An objection that may be (and has been) raised here is that one could start with a decidable fragment, e.g. the atomic sentences of arithmetic and get the undecidability through addition of new sentence-operators such as quantifiers. That is indeed so, but is not relevant here, where one *starts with a larger language* that, as a matter of fact contains undecidable sentences and then isolates a fragment within this language that also will have this property. Decidable sentences used for definitions could only provide decidable sentences and hence some of the sentences of the full language would be left out. Also it is not permissible to speak of adding, say, the quantifiers as their nature is *sub judice*: the meaning of a quantifier is not something independent of the rest of the language but, like any other word, its meaning is the way it contributes to the meaning of the sentences in which it occurs.

Now the argument is nearly at its end. The theory of sense<sub>D</sub> would be *incomplete* in that it could not state what it is to manifest fully implicit knowledge of the recognition-transcendent truth-condition of an undecidable sentence. If the theory attempted to do this, an observational void would



exist without observational warrant. We, as theorists, would be guilty of theoretical slack in our theory, because we could never see the agents manifest their implicit knowledge in response to the truth-conditions obtaining (or not), because *ex hypothesi*, they obtain unrecognisably. The agents, furthermore, could not see them obtain and so, independently of whether or not the theorist can see them response, they cannot manifest their knowledge in response to the truth-condition. (This is a point where the division of linguistic labour may play a role.)

Before we proceed, it might be useful to offer a short schematic summary of Dummett's argument as set out above. (Page references in brackets are to WTM2.)

1. To understand a language is to have knowledge of meaning. (p. 69)
2. Knowledge of meaning must in the end be implicit knowledge. (. 70)
3. Hence the meaning theory must contain a part, call it theory of sense<sub>D</sub>, that specifies 'in what having this knowledge consists, i.e. what counts as a manifestation of that knowledge. (pp. 70–71 and p. 127)
4. There are sentences in the language such that the speaker manifests his knowledge of their meaning in ways other than stating the meaning in other words. (The non-statable fragment is non-empty.) (p. 81)
5. Assume now that bivalent truth can serve as key concept. Bivalent truth-conditions are sometimes undecidable and hence recognition-transcendent. (p. 81)
6. Already in the non-statable fragment there must be sentences with recognition-transcendent truth-conditions. (p. 81)
7. Implicit knowledge of recognition-transcendent truth-conditions cannot be manifested, and so the theory of sense<sub>D</sub> is incomplete. (p. 82)

Supplementary notes concerning the argument:

- a. Dummett's argument is quite general and does not rest at all on any specific features of the language concerned. When it is applied to a particular area of discourse, or for a particular class of statements, it will lead to a metaphysical anti-realism for the area in question. Many examples of this can be found in Dummett's writings. Thus [1975] and [1977] both develop the argument within the philosophy of mathematics. The intuitionistic criticism of classical reasoning, and the ensuing explanations of the logical constants offered by Heyting, provided the main inspiration for Dummett's work on anti-realism. It should be stressed, however, and as is emphasised by Dummett himself

in [1975], that the semantical argument in favour of a constructivist philosophy of mathematics is very far from Brouwer's own position.

In Dummett [1968–1969] another one of the four critical classes of sentences is studied, viz. those concerning time, and in WTM2, Section 3, a discussion of counterfactual conditionals can be found, as well as a discussion of certain reductionist versions of anti-realism. They arise when the truth of statement  $A$  is reduced to the (simultaneous) truth of a certain possibly infinite class of reduction-sentences  $M_A$ . If it so happens that the falsity of the conjunction  $\bigwedge M_A$  does not entail the truth of the conjunction  $\bigwedge M_{\neg A}$ , then bivalence will fail for the statement  $A$ . Examples of such reductionist versions of anti-realism can be found in phenomenalist reductions of material objects or of sentences in others.

- b. It should be noted that Dummett's anti-realism, while verificationist in nature, must not be conflated with logico-empiricist verificationism. With a lot of simplification the matter can be crudely summarised by noting that for the logical empiricists classical logic was sacrosanct and certain sentences have non-verifiable classical truth-conditions. Hence they have no meaning. Dummett reverses this reasoning: obviously meaningful sentences have no good meaning if meaning is construed truth-conditionally. Hence classical meaning-theories are wrong.
- c. As one should expect, Dummett's anti-realist argument has not been allowed to remain uncontroverted. John McDowell has challenged the demand that the meaning theories should comprise a theory of sense $_D$ . In his [1977] and [1978] the criticism is mainly by implication as he is there more concerned with the development of the positive side of his own 'modest' version of a meaning theory, whereas in [1981] he explicitly questions the cogency of Dummett's full-blooded theories. McDowell's [1978a] is an answer to Dummett's [1969], and McDowell in his turn has found a critic in Wright [1980a].

Colin McGinn has been another persistent critic of Dummett's anti-realism and he has launched counter-arguments against most aspects of Dummett's position, cf. e.g. his [1980], [1979] and [1982].

Crispin Wright [1982] challenges Dummett by observing that a Strict Finitist can criticise a Dummettian constructivist in much the same way as a Platonist and so the uniquely privileged position that is claimed for constructivism (as the only viable alternative to classical semantics) is under pressure.

- d. Another sort of criticism is offered by Dag Prawitz [1977, 1978], who, like Wright, is in general sympathy with large parts of Dummett's meaning-theoretical position. Prawitz questions the demand for *full*

manifestation and suggests that the demand for a theory of sense<sub>D</sub> be replaced by an *adequacy condition* on meaning theories *T*:

if *T* is to be adequate, it must be possible  
to derive in *T* the implication  
if *P* knows the meaning of *A*, then *P* shows behaviour *B<sub>A</sub>*.  
Prawitz [1978, p. 27]

(Here “*B<sub>A</sub>*” is a kind of behaviour counted as a sign of grasping the meaning of *A*.)

The difference between this adequacy criterion and the constraints that McDowell imposes on his modest theories is not entirely clear to me. Only if the behaviour is to be shown before, and independently of, the theory of force (whose task it is to issue just the interpreting descriptions that tell what behaviour was exhibited by *P*) could something like a modification of Dummett’s argument be launched and even then it does not seem certain that the desired conclusion can be reached.

- e. In the presentation of Dummett’s argument I have relied solely on WTM2. The anti-realist argument can be found in many places though, e.g. [1973, chapter 14], [1969], [1975] and [1975a] as well as the more recent [1982]. It should be noted that Dummett often cf. e.g., [1969], lays equal or more stress on the acquisition of knowledge rather than its manifestation. Most of the articles mentioned are conveniently reprinted in TE.

Wright [1981], a review of TE, gives a good survey of Dummett’s work. Similarly, in his book [1980] Wright offers extensive discussion of anti-realist themes.

The already mentioned McGinn [1980] and Prawitz [1977], while not in entire agreement with Dummett, both give excellent expositions of the basic issues. It is a virtually impossible task to give a complete survey of the controversy around Dummett’s anti-realist position. In recent years almost every issue of the *Journal of Philosophy, Mind and Analysis*, as well as the *Proceedings of the Aristotelean Society*, contains material that directly, or indirectly, concerns itself with the Dummettian argument.

#### 4 PROOF AS A KEY CONCEPT IN MEANING THEORIES

As was mentioned above the traditional intuitionistic criticism of classical mathematical reasoning, cf. e.g., van Dalen (see Volume 5 of the second edition of this *Handbook*) was an important source of inspiration for Dummett’s anti-realist argument and it is also to intuitionism that he turns in

his search for an alternative key-concept to be used in the meaning theories in place of the bivalent, recognition-transcendent truth-conditions.

The simplest technical treatment of the truth-conditions approach to semantics is undoubtedly provided by the standard truth-tables (which, of course, are incorporated in the Tarski-treatment for, say, full predicate logic) and it is the corresponding constructive ‘proof-tables’ of Heyting that offer a possibility for Dummett’s positive proposal. Heyting’s explanations of the logical constants, cf. his [1956, Chapter 7] and [1960], can be set out roughly as follows:

A proof of the proposition	is given by
$A \wedge B$	a proof of $A$ and a proof of $B$
$A \vee B$	a proof of $A$ or a proof of $B$
$A \rightarrow B$	a method for obtaining proofs of $B$ from proof of $A$
$\perp$	nothing
$\forall x \in D A(x)$	a method which for every individual $d$ in $D$ provides a proof of $A(d)$
$\exists x \in D A(x)$	an individual $d$ in $D$ and a proof of $A(d)$ .

There are various versions of the above table of explanations, e.g. the one offered by Kreisel [1962], where ‘second clauses’ have been included in the explanations for implication and universal quantification to the effect that one has to include also a proof that the methods really have the properties required in the explanations above. The matter is dealt with at length in [Sundholm, 1983], where an attempt is made to sort out the various issues involved and where extensive bibliographical information can be found, cf. also Section 7 below on the type theory of Martin-Löf.

In the above explanations the meaning of a proposition is given by its ‘proof condition’ and, as was emphatically stressed by Kreisel [1962], in some sense, ‘we recognise a proof when we see one’. Thus it seems that the anti-realistic worries of Dummett can be alleviated with the use of proof as a key concept in meaning theories. (I will return to this question in the next section.) Independently of the desired immunity from anti-realist strictures, however, there are a number of other points that need to be taken into account here.

First among these is a logical gem invented by Prior [1960]. In the Heyting explanations the meaning of a proposition is given by its proof-condition. Conversely, does every proof-condition give a proposition? A positive answer to this question appears desirable, but the notion ‘proof-condition’

needs to be much more elucidated if any headway is to be made here. Prior noted that if by ‘proof-conditions’ one understands ‘rules that regulate deductive practice’ then a negative answer is called for. Let us introduce a proposition-forming operator, or connective ‘*tonk*’ by stipulating that its deductive practice is to be regulated by the following Natural Deduction rules (I here alter Prior’s rules inessentially):

$$\begin{array}{l} \text{tonk } I \quad \frac{A}{A \text{ tonk } B} \quad \frac{B}{A \text{ tonk } B} \\ \\ \text{tonk } E \quad \frac{A \text{ tonk } B}{A} \quad \frac{A \text{ tonk } B}{B} . \end{array}$$

As Prior observes one then readily proves false conclusion from true premises by means of first *tonk I* and then *tonk E*. In fact, given these two rule *any two propositions are logically equivalent* via the following derivation:

$$\frac{\frac{A^1}{A \text{ tonk } B} (\text{tonk } I) \quad \frac{B^2}{A \text{ tonk } B} (\text{tonk } I)}{\frac{B}{A} \quad \frac{A}{B}} 1, 2(\leftrightarrow)I$$

$$A \leftrightarrow B$$

Thus *tonk* leads to extreme egalitarianism in the underlying logic: from a logical point of view there is only one proposition. This is plainly absurd and something has gone badly wrong. Hence it is clear (and only what could be expected) that some constraints are needed for how the proof-conditions are to be understood; ‘rules regulating deductive practice’ is simply too broad. There is quite a literature dealing with *tonk* and the problems it causes: [Stevenson, 1961; Wagner, 1981; Hart, 1982] and, perhaps most importantly from our point of view, [Belnap, 1962], more about which below. The relevance of the *tonk*-problem for our present interests, was as far as I know, first noted by Dummett [1973, Chapter 13].

A second point to consider is the so-called *paradox of inference*, cf. Cohen and Nagel [1934, pp. 173–176]. This ‘paradox’ arises because of the tension between (a) the fact that the truth of the conclusion is already contained in the truth of the premises, and (b) the fact that logical inference is a way to gain ‘new’ knowledge. Cohen and Nagel formulate it thus:

*If in an inference the conclusion is not contained in the premise, it cannot be valid; and if the conclusion is not different from the premise, it is useless; but the conclusion cannot be contained in the premises and also possess novelty; hence inferences cannot be both valid and useful* [1934, p. 173]

So there is a tension between the legitimacy (the validity) and the utility of an inference, and one could perhaps reformulate the question posed by the ‘paradox’ as: How can logic function as a useful epistemological tool? For an inference to be legitimate, the process of recognising the premises as true must already have accomplished what is needed for the recognition of the truth of the conclusion, but if it is to be useful the recognition of the truth of the conclusion does not have to be present when the truth of the premises is ascertained. This is how Dummett poses the question in [1975a].

How does one use reasoning to gain new truths? By starting with known premises and drawing further conclusions. In most cases the use of valid inference has very little to do with how one would normally set about to verify the truth of something. For instance, the claim that I have seven coins in my pocket is best established by means of counting them. It would be possible, however, to deduce this fact from a number of diverse premises and some axioms of arithmetic. (The extra premises would be, say that I began the day with a £50 note, and I have then made such and such purchases for such and such sums, receiving such and such notes and coins in return, etc.) This would be a highly *indirect* way in comparison with the straightforward counting process. The utility of logical reasoning lies in that it provides indirect means of learning the truth of statements. Thus in order to account for this usefulness it seems that there must be a gap between the most direct ways of learning the truth and the indirect ways provided by logic. If we now explain meaning in terms of proof, it seems that we close this gap. The direct means, given directly by the meaning, would coincide, so to speak, with the indirect means of reasoning. The indirect means have then been made a part of the direct means of reasoning. (One should here compare the difference between direct and indirect means of recognising the truth with the solution to the ‘paradox’ offered by Cohen and Nagel [1934] that is formulated in terms of a concept called ‘conventional meaning’.)

The constraints we seek on our proof-explanations thus should take into account, on the one hand, that one must not be too liberal as witnessed by *tonk*, and, on the other hand, one must not make the identification between proof and meaning so tight that logic becomes useless.

Already Belnap [1962] noted what was wrong with *tonk* from our point of view. The (new) deductive practice that results from adding *tonk* with its stipulative rules, is not *conservative* over the old one. Using Dummett’s [1975] terminology, there is no *harmony* between the grounds for asserting, and the consequences that may be drawn from, a sentence of the form *A tonk B*. The introduction and elimination rules must, so to speak, match, not just in that each connective has introduction and elimination rules but also in that they must not interfere with the previous practice. Hence it seems natural to let one of the (two classes of) rules serve as meaning-giving and let the other one be chosen in such a way that it(s members) can be justified according to the meaning-explanation. Such a method of proceeding would

also take care of the ‘paradox’ of inference: one of the two types of rules would now serve as the direct, meaning-given (because meaning-giving!) way of learning the truth and the other would serve to provide the indirect means (in conjunction with other justified rules, of course).

The introduction rules are the natural choice for our purpose, since they are *synthesising* rules; they explain how a proof of, say  $A \& B$ , can be formed in terms of given proofs of  $A$  and of  $B$ , and thus some sort of compositionality is present (which is required for a key concept). Tentatively then, the meaning of a sentence is given by what counts as a *direct* (or *canonical*) *proof* of it. Other ways of formulating the same explanation would be to say that the meaning is given by the *direct grounds for asserting*, or by what counts as a *direct verification* of, the sentence in question. An (indirect) proof of a sentence would be a method, or program, for obtaining a direct proof.

In order to see that a sentence is true one does not in general have to produce the direct grounds for asserting it and so the desired gap between truth and truth-as-established-by-the-most-direct-means is open. Note that one could still say that the meaning of a sentence is given by its truth-condition, although the latter, of course, has to be understood in a way different from that of bivalent, and recognition-transcendent, truth: if a sentence is true it is possible to give a proof of it and this in turn can be used to produce a *direct proof*. Thus in order to explain what it is for a sentence to be true one has to explain what a direct proof of the sentence would be and, hence, one has to explain the meaning of the sentence in order to explain its truth-condition.

All of this is highly programmatic and it remains to be seen if, and how, the notion of *direct* (canonical) *proof* (verification, ground for asserting) can be made sense of also outside the confined subject-matter of mathematics. In the next section I shall attempt to spell out the Heyting explanations once again, but now in a modified form that closely links up with the discussion in the present section and with the so-called normalisation theorems in Natural Deduction style proof theory.

## 5 THE MEANING OF THE LOGICAL CONSTANTS AND THE SOUNDNESS OF PREDICATE LOGIC

In the present section, where knowledge of Natural Deduction rules is presupposed, we reconsider Heyting’s explanations and show that the introduction and elimination rules are sound for the intended meaning.

Thus we assume that  $A$  and  $B$  are *meaningful sentences*, or *propositions*, and, hence that we know what proofs (and direct proofs) are for them.

The *conjunction*  $A \wedge B$  is a proposition, such that a canonical proof of  $A \wedge B$  has the form:

$$\frac{\begin{array}{cc} D_1 & D_2 \\ A & B \end{array}}{A \wedge B}$$

where  $D_1$  and  $D_2$  are (not necessarily direct) proofs of  $A$  and  $B$ , respectively. On the basis of this meaning-explanation of the proposition  $A \wedge B$ , the rule ( $\wedge I$ ) is seen to be valid. We have to show that whenever the two premises  $A$  and  $B$  are true then so is  $A \wedge B$ . When  $A$  and  $B$  are true, they are so on the basis of proof and hence there can be found two proofs  $D_1$  and  $D_2$  respectively of  $A$  and  $B$ . These proofs can then be used to obtain a canonical proof of  $A \wedge B$ , which therefore is true.

Consider the elimination rule ( $\wedge E$ ), say,  $\frac{A \wedge B}{B}$ , and assume that  $A \wedge B$  is true. We have to see that  $B$  is true.  $A \wedge B$  is true on the basis of a proof  $D$ , which by the above meaning-explanation can be used to obtain a canonical proof  $D_3$  of the form specified above. Thus  $D_2$  is a proof of  $B$  and thus  $B$  is true.

Next we consider the *implication*  $A \rightarrow B$ , which is a proposition that is true if  $B$  is true on the assumption that  $A$  is true. Alternatively we may say that a canonical proof of  $A \rightarrow B$  has the form

$$\frac{\begin{array}{c} A^1 \\ D \\ B \end{array}}{A \rightarrow B^1}$$

where  $D$  is a proof of  $B$  using the assumption  $A$ . Again, the introduction rule ( $\rightarrow I$ ) is sound, since what has to be shown is that *if*  $B$  is true on the hypothesis that  $A$  is true, *then*  $A \rightarrow B$  is true. But this is directly granted by the meaning explanation above. For the elimination rule we consider

$$\frac{A \rightarrow B \quad A}{B}$$

and suppose that we have proofs  $D_1$  and  $D_2$  of respectively  $A \rightarrow B$  and  $A$ . As  $D_1$  is a proof it can be used to obtain a canonical proof  $D_3$  and thus we can find a hypothetical proof  $D$  of  $B$  from  $A$ . But then

$$\begin{array}{c} D_2 \\ A \\ D \\ B \end{array}$$



is a proof of  $B$  and thus  $B$  is true and  $(\rightarrow E)$  is a valid rule.

The *disjunction*  $A \vee B$  is a proposition, with canonical proofs of the forms

$$\frac{D_1}{A} \quad \text{and} \quad \frac{D_2}{B}$$

$$\frac{}{A \vee B}$$

where  $D_1$  and  $D_2$  are proofs of respectively  $A$  and  $B$ . The introduction rules are immediately seen to be valid, since they produce canonical proofs of their true premise. For the elimination rule, we assume that  $A \vee B$  is true, that  $C$  is true on assumption that  $A$  is true, and that  $C$  is true on assumption that  $B$  is true. Thus there are proofs  $D_1, D_2$  and  $D_3$  of, respectively  $A \vee B, C$  and  $C$ , where the latter two proofs are hypothetical, depending on respectively  $A$  and  $B$ . The proof  $D_1$  can be used to obtain a canonical proof  $D_4$  of  $A \vee \neg B$  in one of the two forms above, say the right, and so  $D_4$  contains a subproof  $D_5$ , that is a proof of  $B$ . Then we readily find a proof of  $C$  by combining  $D_5$  with the hypothetical  $D_3$  to get a proof of  $C$ , which thus is a true proposition.

The *absurdity*  $\perp$  is a proposition which has *no* canonical proof. We have to see that the rule  $\frac{}{\perp}$  is valid. Thus, we have to see that whenever the proposition  $\perp$  is true, then also  $A$  is true. But  $\perp$  is never true, since a proof of  $\perp$  could be used to obtain a *canonical* proof of  $\perp$  and by the explanation above there are no direct proofs of  $\perp$ .

The *universal quantification*  $(\forall x \in M)A(x)$  is a proposition such that its canonical proofs have the form

$$\frac{x \in M^1 \quad D \quad A(x)}{(\forall x \in M)A(x)^1}$$

that is, the proof of  $D$  of the premise is a hypothetical, free-variable, proof of  $A(x)$  from the assumption that  $x \in M$ . Again the introduction rule is valid, since if  $A(x)$  is true on the hypothesis that  $x \in M$ , there can be found a hypothetical proof of  $A(x)$  from assumption  $x \in M$ , and thus we immediately obtain a canonical proof of  $(\forall x \in M)A(x)$ . For the elimination rule  $(\forall E)$  consider

$$\frac{(\forall x \in M)A(x) \quad d \in M}{A(d)}$$

and suppose that the premises are true. Thus proofs  $D_1$  and  $D_2$  of, respectively,  $(\forall x \in M)A(x)$  and  $d \in M$ , can be found. As  $D_1$  is a proof it can be

used to obtain a direct proof of its conclusion, and hence we can extract a hypothetical proof of  $D_3$  of  $A(x)$  from assumption  $x \in M$ . Combining  $D_2$  with the free-variable proof  $D_3$  gives a proof

$$\begin{array}{ll} D_2 & ('D_3(d/x)' \text{ indicates the} \\ d \in M & \text{result of substituting} \\ D_3(d/x) & d \text{ for } x \text{ in } D_3.) \\ A(d) & \end{array}$$

of  $A(d)$ , so the rule  $(\forall E)$  is sound.

Finally the *existential quantification*  $(\exists x \in M)A(x)$  is a proposition such that its canonical proofs have the  $(\exists I)$  form

$$\frac{\begin{array}{ll} D_1 & D_2 \\ A(d) & d \in M \end{array}}{(\exists x \in M)A(x)}$$

Again the introduction rule is immediately seen to be valid as it produces canonical proofs of its conclusion from proofs of the premises. For the elimination rule  $(\exists E)$  consider the situation that  $(\exists x \in M)A(x)$  is true, and that  $C$  is true on the assumptions that  $x$  is in  $M$  and  $A(x)$  is true. Thus there can be found a proof  $D_3$  of  $(\exists x \in M)A(x)$  and a hypothetical free-variable proof  $D_4$  of  $C$  from hypotheses  $x \in M$  and  $A(x)$ . The proof  $D_3$  can be used to obtain a canonical proof of the form above, and combining the proofs  $D_1$  and  $D_2$  with the hypothetical free-variable proof  $D_4$  we obtain a proof of  $D$ :

$$\begin{array}{ll} D_2 & D_1 \\ d \in M & A(d) \\ D_4(d/x) & \\ C & \end{array}$$

Thus the rules of the intuitionistic predicate logic are all valid; no corresponding validation is known for, say, the classical law of Bivalence  $A \vee \neg A$  where  $\neg A$  is defined as  $A \rightarrow \perp$ .

The above treatment has been less precise and complete than would be desirable owing to limitations of space. First, questions of syntax have been left out especially where the quantifier rules are concerned, and secondly a whole complex of problems that arises from the fact that we need to know that  $A(x)$  is a proposition for any  $x$  in  $M$  in order to know that, say,  $(\forall x \in M)A(x)$  is a proposition has been ignored. The interested reader is referred to the type theory of Martin-Löf [1984] for detailed consideration and careful treatment of (analogues to) these and other lacunae, e.g. how to treat atomic sentences in our presentation.

The above explanations of why the rules of predicate logic are valid all follow the same pattern. The introduction rules are immediately seen to be valid, since canonical proofs are given introductory form. The elimination rules are then seen to be valid by noting that the introduction and elimination rules have the required harmony. The canonical grounds for asserting a sentence do contain sufficient grounds also for the consequences that may be drawn via the elimination rules for the sentence in question. Thus, in fact, we have here made use of the *reduction steps* first isolated and used by Prawitz [1965, 1971], in his proofs of the normalisation theorems for Natural Deduction-style formalisations.

Prawitz has in a long series of papers [1973, 1974, 1975, 1978 and 1980] been concerned to use this technical insight for meaning-theoretical purposes. His main concern, however, has been to give an explication of the notion of valid argument rather than to give direct meaning explanations in terms of proof. In the presentation here, which is inspired by Martin-Löf's meaning-explanations for his type theory, I have been more concerned with the task of giving constructivistic meaning-explanations while relying on the standard explication of validity as preservation of truth for a justification of the standard rules of inference.

One should, however, stop to consider the extent to which the above explanations constitute a meaning theory in the sense of section 1 above. In particular, in section 4 a promise was given to return to the question of decidability. Is it in fact true that the notion of proof is decidable? On our presentation at least this much is true: if we already have a proof it is decidable if it is in canonical form. As to the general question I would be inclined to think that the notion of proof is *semi-decidable*, in that we recognise a proof when we see one, but when we don't see one that does not necessarily mean that there is no proof there. One can compare the situation with understanding a meaningful sentence: we understand a meaningful sentence when we see (or hear!) one but if we don't understand that does not necessarily mean that there is nothing there to be understood. Failure to understand a meaningful sentence seems parallel to failure to follow, or grasp, a proof. Such a position, then, would not make the 'proof-condition' recognition-transcendent; when it obtains it can be seen to obtain, but when it is not seen to obtain no judgement is given (unless, of course it is seen not to obtain). Apart from the question of decidability, an important difference is that in explanations such as the above there is no mention of implicit knowledge and the like. It seems correct to speak of a theoretical representation of a (constructivistic) deductive practice, but it seems less natural to say that these explanations are known to everyone who draws logical inferences.

We used the notion of canonical proof as a key concept in order to provide the explanations, and in the literature one can find a number of alternatives as to how one ought to specify these, cf. the papers by Prawitz listed

above. In particular, one might wish to insist that all parts of a canonical proof should also be canonical (as is the case with the so-called normal derivations obtained by Prawitz in his normalization theorem [1971]). The choice I opted for here was motivated by, first, the success of the meaning-explanations of Martin-Löf in his type theory and, secondly, the fact that in Hallnäs [1983] a successful normalisation of strong set-theoretic systems is carried out using an analogous notion of normal derivation (Tennant's [1982] and his book [1978] are also interesting to the set-theoretically curious; in the former a treatment of the paradoxes is offered along Natural Deduction lines, and the latter contains a neat formulation of the rules of set theory.)

Finally, we should note that the explanations offered here have turned the formal system into an *interpreted formal system* (modulo not inconsiderable imprecision in the formulation of syntax and explanations). This is the main reason for the avoidance of Greek letters in the present Chapter.

## 6 QUESTIONS OF COMPLETENESS

In section 5 the meaning of the logical constants was explained and the standard deductive practice justified. In the case of classical, bivalent logic we know that the connectives  $\wedge, \vee$  and  $\neg$  are complete in that any truth-function can be generated from them. Does the corresponding property hold here? Clearly the answer is dependent on how the canonical proofs may be formed. It was shown by Prawitz [1978] and, independently of him, by Zucker and Tragesser [1978] that if we restrict ourselves to purely schematic means for obtaining canonical proofs (and for logical constants this does not seem unreasonable), then an affirmative answer is possible to the above question. As a typical example consider e.g. this Sheffer-stroke (which of course makes sense constructively as well). This is given the introduction rule ( $|I$ )

$$\frac{A^1 \dots B^2 \quad \vdots \quad \perp}{A|B^{1,2}}$$

A definition using  $\wedge, \rightarrow$  and  $\perp$  is found by putting  $A|B =_{\text{def}} A \wedge B \rightarrow \perp$ . If there are more premise-derivations in the introduction rule (= the rule for how canonical proofs may be obtained) for each of these one will get an implication of the above sort and they are all joined together by conjunctions. (Here it is presupposed that the rules have only finitely many premises. This does not seem unreasonable.) Finally, if there are more introduction

rules than one, the conjunctions are put together into a disjunction. (Here it is presupposed that there are only finitely many introduction rules. Again this does not seem unreasonable).

Only one case remains, namely that there are no introduction rules. Then there are no canonical proofs to be found and we have got the absurdity. Thus the fragment based on  $\rightarrow, \wedge, \vee$  and  $\perp$  is complete. For further details refer to the two original papers above as well as Schroeder-Heister [1982]. It should be noted that by Hendry [1981] we know that  $A \wedge B$  is equivalent also intuitionistically to  $(A \leftrightarrow B) \leftrightarrow (A \vee B)$  and that  $A \rightarrow B$  is equivalent to  $B \leftrightarrow A \vee B$ . Thus also  $\leftrightarrow, \vee$  and  $\perp$  are complete.

The standard elimination-rules ( $\wedge E$ ) can be replaced by the following rule:

$$\frac{A \wedge B \quad \begin{array}{c} A^1 \dots B^2 \\ \vdots \\ C \end{array}}{C} \quad 1,2$$

which rule seems quite well-motivated by the analogy with the introduction rule  $\frac{A \quad B}{A \wedge B}$ : everything which can be derived from the two premises  $A$  and  $B$  used as assumptions can also be derived from  $A \wedge B$  alone. The ( $\vee E$ ) rule has exactly this general pattern and the intuitionistic absurdity rule is a degenerate case without minor premise  $C$ :

$$\frac{\perp}{C}$$

Only implication does not obey the above pattern. Here the premise of the introduction rule is not just a sentence, but a hypothetical judgement that  $B$  is true whenever  $A$  is true. Thus, we have a sort of rule as premise: from  $A$  go to  $B$ , in symbols  $A \Rightarrow B$ . If we may use such rules as *dischargable* assumptions, one can keep the standard pattern also for implication, viz.

$$\frac{A \Rightarrow B \quad \begin{array}{c} A \Rightarrow B^1 \\ \vdots \\ C \end{array}}{C} \quad 1$$

whereas if we try to do the same using implication for the arrow  $\Rightarrow$ , we end up with the triviality

$$\frac{A \rightarrow B \quad \begin{array}{c} A \rightarrow B^1 \\ \vdots \\ C \end{array}}{C} \quad 1$$

which does not allow us to derive even *modus ponens*.

Using the rule with the higher level assumption  $A \Rightarrow B$  one can derive ( $\rightarrow E$ ) as follows:

$$\frac{A \rightarrow B \quad \frac{A}{B} (A \Rightarrow B)^1}{B} \quad 1$$

Given the use of the rule  $A \Rightarrow B$  as an assumption, from premise  $Q$  we can proceed to conclusion  $B$ , and the use of the major premise  $A \rightarrow B$  allows to *discharge the use of the rule*  $A \Rightarrow B$ .

This type of higher-level assumptions was introduced by Schroeder-Heister [1981] and it is a most interesting innovation in Natural Deduction-formulations of logic, cf. also his [1982] and [1983]. The elimination rule that the Prawitz method gives to the Sheffer-stroke would be

$$\frac{A \wedge B \rightarrow \perp^1 \quad \begin{array}{c} A \wedge B \rightarrow \perp^1 \\ \vdots \\ C \end{array}}{C} \quad 1$$

which follows the above pattern, but uses implication and conjunction. With the Schroeder-Heister conventions the rule can be given as

$$\frac{(A, B \rightarrow \perp)^1 \quad \begin{array}{c} (A, B \rightarrow \perp)^1 \\ \vdots \\ C \end{array}}{C} \quad 1$$

In words, if  $C$  is true under the assumptions that we may go from the premise  $A$  and  $B$  to conclusion  $\perp$ , then  $C$  is a consequence of  $A|B$  alone.

In Schroeder-Heister [1984] an extension of the above results is given and completeness is established also for the predicate calculus language.

The other question of completeness is also considered by Schroeder-Heister [1983]: is every valid inference derivable from the introduction and

elimination rules? This question gets a positive answer, but the concept of validity is extremely restrictive, i.e. the rule  $\frac{(A \wedge B) \wedge C}{A}$  is not a valid rule, cf. [1983, p. 374], which (given the concept of validity used in the present paper) it obviously must be. Thus I would consider the problem, first posed by Prawitz [1973], to establish the completeness of the predicate logic, for the present sort of meaning explanations, still to be open.

## 7 THE TYPE THEORY OF MARTIN-LÖF

Frege [1893], in the course of carrying out his logicist programme, designed a full-scale, completely formal language that was intended to suffice for mathematical practice. By today's standards, an almost unique feature of his attempt to secure a foundation of mathematics is that he uses an *interpreted* formal language for which he provides careful meaning explanations. The language proposed was, as we now know, not wholly successful, owing to the intervention of Russell's paradox. (The effects of the paradox on Frege's explanations of meaning are explored in Aczel [1980] and, from a different perspective, in Thiel [1975] and Martin [1982].) As the formal logic of Frege (and Whitehead–Russell) was transformed gradually into mathematical logic, notably by Tarski and Gödel, interest in the task of giving meaning explanations for interpreted formal languages faded out and after World War II the current distinction between syntax and (Tarskian, model-theoretic) semantics has become firmly entrenched.

The type theory of Martin-Löf [1975, 1982, 1984] represents a remarkable break with this tradition in that it returns to the original Fregean paradigm: interpreted formal language with careful explanations of meaning. Owing to limitations of space I shall not be able to give a detailed, precise description of the system here, (a task for which Martin-Löf [1984] uses close to a hundred ages), but will confine myself to trying to convey the basic flavour of the system.

A possible route to Martin-Löf's theory is through further examination of Heyting's explanations of the meaning of the logical constants. Our tentative semantics in section 5 above made tacit use of a refinement of the explanations: the proof-tables do not give just proofs but *canonical*, or *direct* proofs. A further refinement can be culled from Heyting's own writings. (In Sundholm [1983] a fairly detailed examination of Heyting's writings on this topic is offered.) According to Heyting, in order to prove a theorem one has to carry out certain constructions, 'die gewissen Bedingungen genügen', namely that it produces a mathematical object with certain specified properties, cf. e.g. his remarks on the proposition

"Euler's constant is rational"

in [1931, p. 113]. In Martin-Löf's system, the proof-tables are extended to

contain also the information about the objects that need to be constructed in order to establish the truth of the propositions in question. Thus, taking both refinements into account, the meaning of a proposition is explained by telling what a canonical object for the proposition would be. (A canonical object is not needed in order to assert the proposition; an object (method program) that can be evaluated to canonical form is enough. For more details here, see Martin-Löf [1984].) In fact, according to Martin-Löf, one also has to tell when two such objects are equal. On the other hand, when one defines a set constructively, one has to specify what the canonical elements are and what it is for two elements of the set to be equal elements. Thus, the explanations of what propositions are and of what sets are, are completely analogous and Martin-Löf's system does not differentiate between the two notions.

In ordinary formal theories, that are formulated in the predicate calculus, the derivable objects are *propositions* (or, rather, they are well-formed formulae, i.e. the formalistic counterparts of propositions). This leads to certain difficulties for the standard formulation where logical inference is a relation between proposition. As was already observed by Frege, the correct formulation of *modus ponens* is

$$\frac{A \rightarrow B \text{ is true} \quad A \text{ is true}}{B \text{ is true}};$$

It is simply not correct to say that the proposition  $B$  follows from the propositions  $A \rightarrow B$  and  $A$ . What is correct is that the *truth* of the proposition  $B$  follows from the *truth* of  $A \rightarrow B$  and the *truth* of  $A$ . Thus *the premises and conclusions of logical inferences are not propositions but judgements as to the truth of the propositions*. Furthermore, as Martin-Löf notes, that in order to keep the rules formal, one should also include the information that  $A$  and  $B$  are propositions in the premises of the rules, e.g.

$$\frac{A \text{ is a prop.} \quad B \text{ is a prop.} \quad A \text{ is true}}{A \vee B \text{ is true}}$$

is how  $\vee$ -introduction should be set out. Therefore, as the premises of inferences are judgements, and remembering the identification of propositions and sets, one finds two main sorts of judgements in the theory, namely

$$(a) \quad A \text{ set} \quad ('A \text{ is a set}')$$

and

$$(b) \quad a \in A \quad ('a \text{ is an element of the set } A')$$

(In fact, there are two further forms of judgement, namely ' $A$  is the same set as  $B$ ' and ' $a$  and  $b$  are equal elements of the set  $A$ '.)



In accordance with the above discussion, (a) also does duty for ‘ $A$  is a proposition’ and (b) can also be read as ‘the (proof-)object  $a$  is of the right sort for the proposition  $A$ , meets the condition specified by the proposition  $A$ ’. This reading of (b) is, constructively, a longhand for the judgement ‘(the proposition)  $A$  (is) *true*’, which is used whenever it is convenient to suppress the extra information contained in the proof-object. A third reading, deriving from Heyting and Kolmogorov, is possible, where (a) is taken in the sense ‘ $A$  is a *task* (or *problem*)’ and (b) in the sense ‘ $a$  is a method for carrying out the task  $A$  (solving the problem  $A$ )’. When the task-aspect is emphasised, another reading would be ‘ $a$  is a *program* that meets the *specification*  $A$ ’ and the type-theoretical language of Martin-Löf [1982] has, owing to this possibility, had considerable influence as a programming language.

Some feeling for the interaction between propositions and proof-objects may be obtained through consideration of the simple example of *conjunction*. The *proposition*  $A \wedge B$  (or *set*  $A \times B$ ) is explained, on the assumption that  $A$  and  $B$  are propositions, by laying down that a canonical element of  $A \times B$  is a pair  $(a, b)$  where  $a \in A$  and  $b \in B$ . Thus the  $\times$ -introduction rule is correct:

$$\frac{a \in A \quad b \in B}{(a, b) \in A \times B}.$$

Using the shorthand reading, when the proof-objects are left out, we also see that the rule of  $\wedge$ -introduction is correct:

$$\frac{A \text{ true} \quad B \text{ true}}{A \wedge B \text{ true}}.$$

For the  $\wedge$ -eliminations, we need the use of the *projection-functions*  $p$  and  $q$  that are associated with the pairing-function. Consider the rule

$$\frac{A \wedge B \text{ true}}{A \text{ true}}.$$

Restoring proof-objects, we see that from an element  $c \in A \wedge B$ , one has to find an element of  $A$ . But  $c$  is an element of  $A \wedge B$ , and so  $c$  is equal to (is a method for finding, can be evaluated, or executed, to) a *canonical* element  $(a, b) \in A \wedge B$ . Applying the projection  $p$ , we see that  $p(c) = p((a, b)) = a \in A$ , so the proper formulation will be

$$\frac{c \in A \wedge B}{p(c) \in A}.$$

It should be mentioned, however, that the conjunction is not a primitive set-formation operation in the language of Martin-Löf. On the contrary, a suitable candidate can be defined from other sets and the appropriate rules derived.

A slightly more complex example is provided by the *universal quantification*  $(\forall x \in A)B[x]$  and *implication*  $A \rightarrow B$ , both of which are treated as variants of the Cartesian product  $(\Pi x \in A)B[x]$  of a family of sets. This product may be formed only on the assumption that we have a family of sets over  $A$ , that is, provided that  $B[x]$  is a set, whenever  $x \in A$ . Thus the formation rule will take the form

$$\frac{\begin{array}{c} x \in A^1 \\ \vdots \\ A \text{ set } \quad B[x] \text{ set} \end{array}}{(\Pi x \in A)B[x] \text{ set}}.$$

(This serves to illustrate the important circumstance that the basic judgements may depend on assumptions. Better still, we should say that the right premise is a *hypothetical* judgement  $B[x]$  set (provided that  $x \in A$ .) In order to understand the  $\Pi$ -formation rule one needs to know what a canonical element of  $(\Pi x \in A)B[x]$  would be; this is told by the  $\Pi$ -introduction rule

$$\frac{\begin{array}{c} x \in A^1 \\ \vdots \\ b[x] \in B[x] \end{array}}{\lambda x.b[x] \in (\Pi x \in A)B[x]}$$

that is, the canonical elements are *functions*  $\lambda x b[x]$ , such that  $b[x] \in B[x]$  provided that  $x \in A$ . Just as in the case of conjunction, where the elimination rule was taken care of by matching the pairing function with a projection, one will obtain the elimination rule through a similar match between  $\lambda$ -abstraction and function-*application*, *ap*. Thus the rule take the form

$$\frac{f \in (\Pi x \in A)B[x] \quad a \in A}{ap(f, a) \in B[a/x]}.$$

(In order to understand this rule one makes use of an important connection between abstraction and application, namely the law

$$ap(\lambda x.b[x], a) = b[a/x].$$

For the details of the explanation, refer to Martin-Löf [1982] or [1984].)

If the set (proposition)  $B[x]$  does not depend on  $x$  the product is written as the set of functions  $B^A$  (as the proposition  $A \rightarrow B$ ). The rules are obvious, with the exception of  $\rightarrow$ -formation:

$$\frac{\begin{array}{c} A \text{ true}^1 \\ \vdots \\ A \text{ prop} \quad B \text{ prop} \end{array}}{A \rightarrow B \text{ prop}}^1.$$

Here the formation rule is stronger than the usual rule (where  $A$  and  $B$  both have to be propositions) because the right premise is weaker in that  $B$  has to be a proposition only when  $A$  is true. This concept of implication has been used by Stenlund in an elegant theory of definite descriptions, cf. his [1973] and [1975].

The other quantification is taken care of by means of the *disjoint union* of a family of sets. The  $\Sigma$ -formation rule takes the form

$$\frac{\begin{array}{c} x \in A^1 \\ \vdots \\ A \text{ set} \quad B[x] \text{ set} \end{array}}{(\Sigma x \in AB[x] \text{ set})}^1.$$

The canonical elements are given by the  $\Sigma$ -introduction rule

$$\frac{a \in A \quad b \in B[a/x]}{(a, b) \in (\Sigma x \in A)B[x]}.$$

On the propositional reading, where the disjoint union is written as the quantifier  $(\exists x \in A)B[x]$ , we see that in order to establish an existence claim one has to (i) exhibit a suitable witness  $a \in A$  and (ii) supply a suitable proof-object  $b$  that the witness  $a \in A$  does, in fact, satisfy the condition imposed by  $B[x]$ . The inclusion of the proof-object  $b$  allows yet a third use for the disjoint union, namely that of restricted comprehension-terms. What would, on a constructive reading, be meant by ‘an element of the set of  $x$ ’s in  $A$  such that  $B[x]$ ’? At least one would have to include a witness  $a \in A$  and information (=  $a$  proof-object) establishing that  $a$  satisfies the condition  $B[x]$ . Thus the canonical elements of the restricted comprehension-term  $\{x \in A : B[x]\}$  coincide with the canonical elements of the disjoint sum. This representation of ‘such that’ provides the key to the actual development of, say, the theory of real numbers given the set  $N$

of natural numbers. A real number will be an element of  $N^N$  such that it obeys a Cauchy-condition.

At this point I will refrain from further development of the language and instead I shall apply the type-theoretic abstractions that have been introduced so far to the notorious ‘donkey-sentence’

(\*) Every man who owns a donkey beats it.

The problem here is, of course, that formulations within ordinary predicate logic do not seem to provide any way to capture the back-reference of the pronoun ‘it’. A simple-minded formalisation yields

(\*\*)  $\forall x(\text{Man}(x) \wedge \exists y(\text{Donkey}(y) \wedge \text{Own}(x, y)) \rightarrow \text{Beats}(x, ?))$ .

There seems to be no way of filling the place indicated by ‘?’, as the donkey has been quantified away by ‘y’.

Using the disjoint-union manner of representation for restricted comprehension-terms one finds that ‘a man who owns a donkey’ is an element of the set

$$\{x \in \text{MAN} : (\exists y \in \text{DONKEY})\text{OWN}[x, y]\}.$$

Such an element, when in canonical form, is a pair  $(m, b)$ , where  $m \in \text{MAN}$  and  $b$  is a proof-object for  $(\exists y \in \text{DONKEY})\text{OWN}[m/x, y]$ . Thus  $b$ , in its turn, when brought to canonical form, will be a pair  $(d, c)$ , where  $d$  is a **DONKEY** and  $c$  a proof-object for  $\text{OWN}[m/x, d/y]$ . Thus for an element  $z$  of the comprehension-term ‘MAN who OWNS a DONKEY’ the left projection  $p(z)$  will be a man and the right projection  $q(z)$  will be a pair whose left projection  $p(q(z))$  will be the witnessing donkey. Putting it all together we get the formulation

(\*\*\*)  $(\forall z \in \{x \in \text{MAN} : (\exists y \in \text{DONKEY})\text{OWN}[x, y]\})\text{BEAT}[p(z), p(q(z))]$ .

In this manner, then, the type-theoretic abstractions suffice to solve the problem of the pronominal back-reference in (\*). It should be noted here that there is nothing *ad hoc* about the treatment, since all the notions used have been introduced for mathematical reasons in complete independence of the problem posed by (\*). On the other hand one should stress that it is not at all clear that one can export the ‘canonical proof-objects’ conception of meaning outside the confined area of constructive mathematics. In particular the treatment of atomic sentences such as ‘OWN[x, y]’ is left intolerably vague in the sketch above and it is an open problem how to remove that vagueness.

Martin-Löf’s type theory has attracted a measure of metamathematical attention. Peter Aczel [1977, 1987, 1980, 1982], in particular, has been a tireless explorer of the possibilities offered by the type theory. Other papers of interest are Diller [1980], Diller and Troelstra [1984] and Beeson [1982].

## NOTE ADDED IN PROOF (OCTOBER 1985)

Per Martin-Löf's 'On the meanings of the logical constants and the justifications of the logical laws' in *Atti degli incontri di logica matematica vol. 2*, Scuola di Specializzazione in Logica Matematica, Dipartimento di Matematica, Università di Siena, 1985, pp. 203–281, was not available during the writing of the present chapter. In these lectures, Martin-Löf deals with the topics covered in sections 4–6 above in great detail and carries the philosophical analysis considerably further.

*University of Nijmegen, The Netherlands.*

## EDITOR'S NOTE 2001

[For the most recent coverage of Martin-Löf's type theory, see the chapter by B. Nordström, K. Peterson and J. M. Smith in S. Abramsky, D. Gabbay and T. S. E. Maibaum, eds., *Handbook of Logic in Computer Science*, volume 5, pp. 1–37, Oxford University Press, 2000.]

## BIBLIOGRAPHY

- [Aczel, 1977] P. Aczel. The strength of Martin-Löf's type theory with one universe. In *Proceedings of the Symposium on Mathematical Logic (Oulo 1974)*, S. Miettinen and J. Väänänen, eds. pp. 1–32, 1977. Report No 2, Department of Philosophy, University of Helsinki.
- [Aczel, 1978] P. Aczel. The type theoretic interpretation of constructive set theory. In *Logic Colloquium 1977*, A. Macintyre et al., eds. pp. 55–66, 1978.
- [Aczel, 1980] P. Aczel. Frege structures and the notions of proposition, truth and set. In *The Kleene Symposium*, J. Barwise et al., eds. pp. 31–59. North-Holland, Amsterdam, 1980.
- [Aczel, 1982] P. Aczel. The type theoretic interpretation of constructive set theory: choice principles. In *The L. E. J. Brouwer Centenary Symposium*, A. S. Troelstra and D. van Dalen, eds. pp. 1–40. North-Holland, Amsterdam, 1982.
- [Baldwin, 1979] T. Baldwin. Interpretations of quantifiers. *Mind*, **88**, 215–240, 1979.
- [Beeson, 1982] M. Beeson. Recursive models for constructive set theories. *Annals Math. Logic*, **23**, 127–178, 1982.
- [Belnap, 1962] N. D. Belnap. Tonk, plonk and plink. *Analysis*, **22**, 130–134, 1962.
- [van Benthem and van Eijck, 1982] J. F. A. K. van Benthem and J. van Eijck. The dynamics of interpretation. *J. Semantics*, **1**, 3–20, 1982.
- [Cohen and Nagel, 1934] M. R. Cohen and E. Nagel. *An Introduction to Logic and Scientific Method*, Routledge and Kegan Paul, London, 1934.
- [Davidson, 1967] D. Davidson. Truth and meaning. *Synthese*, **17**, 304–323, 1967.
- [Davidson, 1984] D. Davidson. *Inquiries into Truth and Interpretation*, Oxford University Press, 1984.
- [Davies, 1981] M. K. Davies. *Meaning, Quantification, Necessity*. Routledge and Kegan Paul, London, 1981.
- [Diller, 1980] J. Diller. Modified realisation and the formulae-as-types notion. In *To H. B. Curry: Essays on Combinatory Logic, Lambda Calculus and Formalism*, J. P. Seldin and R. Hindley, eds. pp. 491–502. Academic Press, London, 1980.

- [Diller and Troelstra, 1984] J. Diller and A. S. Troelstra. Realisability and intuitionistic logic. *Synthese*, **60**, 153–282, 1984.
- [Dummett, 1968–1969] M. Dummett. The reality of the past. *Proc. Aristot. Soc.*, **69**, 239–258, 1968–69.
- [Dummett, 1973] M. Dummett. *Frege*. Duckworth, London, 1973.
- [Dummett, 1975] M. Dummett. The philosophical basis of intuitionistic logic. In *Logic Colloquium '73*, H. E. Rose and J. Shepherdson, eds. pp. 5–40. North-Holland, Amsterdam, 1975.
- [Dummett, 1975a] M. Dummett. The justification of deduction. *Proc. British Academy*, **LIX**, 201–321, 1975.
- [Dummett, 1976] M. Dummett. What is a theory of meaning? (II). [WTM2] In [Evans and McDowell, 1976], pp. 67–137, 1976.
- [Dummett, 2000] M. Dummett. *Elements of Intuitionism*, 2nd edition, Oxford University Press, 2000. (First edition 1977.)
- [Dummett, 1978] M. Dummett. *Truth and Other enigmas*, [TE], Duckworth, London, 1978.
- [Dummett, 1981] M. Dummett. Frege and Wittgenstein. In *Perspectives on the Philosophy of Wittgenstein*, I. Block, ed. pp. 31–42. Blackwell, Oxford, 1981.
- [Dummett, 1982] M. Dummett. Realism. *Synthese*, **52**, 55–112, 1982.
- [Evans, 1977] G. Evans. Pronouns, quantification and relative clauses (I). *Canadian J. Phil.*, reprinted in Platts [1980, pp. 255–317].
- [Evans and McDowell, 1976] G. Evans and J. McDowell, eds. *Truth and Meaning*. Oxford University Press, 1976.
- [Foster, 1976] J. A. Foster. Meaning and truth theory. In [Evans and McDowell, 1976, pp. 1–32].
- [Frege, 1893] G. Frege. *Grundgesetze der Arithmetik*, Jena, 1893.
- [Gentzen, 1934–1935] G. Gentzen. Untersuchungen über das logische Schliessen. *Mathematische Zeitschrift*, **39**, 176–210, 405–431, 1934–1935.
- [Hacking, 1979] I. Hacking. What is logic? *J. Philosophy*, **76**, 285–319, 1979. Reproduced in *What is a Logical System?*, D. Gabbay, ed. Oxford University Press, 1994.
- [Hallnäs, 1983] L. Hallnäs. *On Normalization of Proofs in Set Theory*, Dissertation, University of Stockholm, Preprint No. 1, Department of Philosophy, 1983.
- [Hart, 1982] W. D. Hart. Prior and Belnap. *Theoria*, **XLVII**, 127–138, 1982.
- [Hendry, 1981] H. E. Hendry. Does IPC have binary indigenous Sheffer function? *Notre Dame J. Formal Logic*, **22**, 183–186, 1981.
- [Heyting, 1931] A. Heyting. Die intuitionistische Grundlegung der Mathematik. *Erkenntnis*, **2**, 106–115, 1931.
- [Heyting, 1956] A. Heyting. *Intuitionism*, North-Holland, Amsterdam, 1956.
- [Heyting, 1960] A. Heyting. Remarques sur le constructivisme. *Logique et Analyse*, **3**, 177–182, 1960.
- [Kleene, 1967] S. C. Kleene. *Mathematical Logic*, John Wiley & Sons, New York, 1967.
- [Kreisel, 1962] G. Kreisel. Foundations of intuitionistic logic. In *Logic, Methodology and Philosophy of Science*, E. Nagel et al., eds. pp. 198–210. Stanford University Press, 1962.
- [Loar, 1976] B. Loar. Two theories of meaning. In [Evans and McDowell, 1976, pp. 138–161].
- [McDowell, 1976] J. McDowell. Truth conditions, bivalence and verificationism. In [Evans and McDowell, 1976, pp. 42–66].
- [McDowell, 1977] J. McDowell. On the sense and reference of a proper name. *Mind*, **86**, 159–185, 1977. Also in [Platts, 1980].
- [McDowell, 1978] J. McDowell. Physicalism and primitive denotation: Field on Tarski. *Erkenntnis*, **13**, 131–152, 1978. Also in [Platts, 1980].
- [McDowell, 1978a] J. McDowell. On 'The reality of the past'. In *Action and Interpretation*, C. Hookway and P. Pettit, eds. p. 127–144. Cambridge University Press, 1978.
- [McDowell, 1981] J. McDowell. Anti-realism and the epistemology of understanding. In *Meaning and Understanding*, H. Parrett and J. Bouveresse, eds. pp. 225–248. de Gruyter, Berlin, 1981.

- [McGinn, 1979] C. McGinn. An *a priori* argument for realism. *J. Philosophy*, **74**, 113–133, 1979.
- [McGinn, 1980] C. McGinn. Truth and use. In [Platts, 1980, pp. 19–40].
- [McGinn, 1982] C. McGinn. Realist semantics and content ascription. *Synthese*, **52**, 113–134, 1982.
- [Martin, 1982] E. Martin, Jr. Referentiality in Frege's *Grundgesetze*. *History and Philosophy of Logic*, **3**, 151–164, 1982.
- [Martin-Löf, 1975] P. Martin-Löf. An intuitionistic theory of types. In *Logic Colloquium '73*, H. E. Rose and J. Shepherdson, eds. pp. 73–118. North-Holland, Amsterdam, 1975.
- [Martin-Löf, 1982] P. Martin-Löf. Constructive mathematics and computer programming. In *Logic, Methodology and Philosophy of Science VI*, L. J. Cohen *et al.*, eds. pp. 153–175, North-Holland, Amsterdam, 1982.
- [Martin-Löf, 1984] P. Martin-Löf. *Intuitionistic Type Theory*. Notes by Giovanni Sambin of a series of lectures given in Padova, June 1980, Bibliopolis, Naples, 1984.
- [Peacocke, 1981] C. A. B. Peacocke. The theory of meaning in analytical philosophy. In *Contemporary Philosophy, vol. 1*, G. Flöistad, ed. pp. 35–36. M. Nijhoff, The Hague, 1981.
- [Platts, 1979] M. de B. Platts. *Ways of Meaning*. Routledge and Kegan Paul, London, 1979.
- [Platts, 1980] M. de B. Platts. *Reference, Truth and Reality*. Routledge and Kegan Paul, London, 1980.
- [Prawitz, 1965] D. Prawitz. *Natural Deduction*. Dissertation, University of Stockholm, 1965.
- [Prawitz, 1971] D. Prawitz. Ideas and results in proof theory. In *Proceedings of the Second Scandinavian Logic Symposium*, J.-E. Fenstad, ed. pp. 235–308. North-Holland, Amsterdam, 1971.
- [Prawitz, 1973] D. Prawitz. Towards a foundation of general proof theory. In *Logic, Methodology and Philosophy of Science IV*, P. Suppes *et al.*, eds. pp. 225–250. North-Holland, Amsterdam, 1973.
- [Prawitz, 1975] D. Prawitz. Comments on Gentzen-type procedures and the classical notion of truth. In *Proof Theory Symposium* J. Diller and G. H. Müller, eds. pp. 290–319. Lecture Notes in Mathematics 500, Springer, Berlin, 1975.
- [Prawitz, 1977] D. Prawitz. Meaning and Proofs. *Theoria*, **XLIII**, 2–40, 1977.
- [Prawitz, 1978] D. Prawitz. Proofs and the meaning and completeness of the logical constants. In *Essays on Mathematical and Philosophical Logic*, J. Hintikka *et al.*, eds. pp. 25–40. D. Reidel, Dordrecht, 1978.
- [Prawitz, 1980] D. Prawitz. Intuitionistic logic: a philosophical challenge. In *Logic and Philosophy*, G. H. von Wright, ed. pp. 1–10. M. Nijhoff, The Hague, 1980.
- [Prior, 1960] A. n. Prior. The runabout inference ticket. *Analysis*, **21**, 38–39, 1960.
- [Schroeder-Heister, 1981] P. Schroeder-Heister. *Untersuchungen zur regellogischen Deutung von Aussagenverknüpfungen*. Dissertation, University of Bonn, 1981.
- [Schroeder-Heister, 1982] P. Schroeder-Heister. Logische Konstanten und Regeln. *Concepts*, **16**, 45–60, 1982.
- [Schroeder-Heister, 1983] P. Schroeder-Heister. The completeness of intuitionistic logic with respect to a validity concept based on an inversion principle. *J. Philosophical Logic*, **12**, 359–377, 1983.
- [Schroeder-Heister, 1984] P. Schroeder-Heister. generalised rules for quantifiers and the completeness of the intuitionistic operators  $\&$ ,  $\vee$ ,  $\supset$ ,  $\perp$ ,  $\forall$ ,  $\exists$ . In *Computation and Proof Theory*, M. Richter, *et al.*, eds. pp. 399–426. Lecture notes in Mathematics, 1104, Springer, Berlin, 1984.
- [Smullyan, 1968] R. Smullyan. *First Order Logic*, Springer-Verlag, Berlin, 1968.
- [Stenlund, 1973] S. Stenlund. *The Logic of Description and Existence*. Department of Philosophy, University of Uppsala, 1973.
- [Stenlund, 1975] S. Stenlund. Descriptions in intuitionistic logic. In *Proceedings of the Third Scandinavian Logic Symposium*, S. Kanger, ed. pp. 197–212. North-Holland, Amsterdam, 1975.

- [Stevenson, 1961] J. T. Stevenson. Roundabout the runabout inference-ticket. *Analysis*, **21**, 124–128, 1961.
- [Sundholm, 1981] G. Sundholm. Hacking's logic. *J. Philosophy*, **78**, 160–168, 1981.
- [Sundholm, 1983] G. Sundholm. Constructions, proofs and the meaning of the logical constants. *J. Philosophical Logic*, **12**, 151–172, 1983.
- [Tarski, 1956] A. Tarski. *Logic, Semantics, Metamathematics*, Oxford University Press, 1956.
- [Tennant, 1978] N. Tennant. *Natural Logic*, Edinburgh University Press, 1978.
- [Tennant, 1982] N. Tennant. proof and paradox. *Dialectica*, **36**, 265–296, 1982.
- [Thiel, 1975] C. Thiel. Zur Inkonsistenz der Fregeschen Mengenlehre. In *Frege und die moderne Grundlagenforschung*, C. Thiel, ed. pp. 134–159. A. Hain, Meisenheim, 1975.
- [Wagner, 1981] S. Wagner. Tonk. *Notre Dame J. Formal Logic*, **22**, 289–300, 1981.
- [Wright, 1976] C. Wright. Truth conditions and criteria. *Proc. Arist. Soc.*, supp **50**, 217–245, 1976.
- [Wright, 1980] C. Wright. *Wittgenstein on the foundations of Mathematics*. Duckworth, London, 1980.
- [Wright, 1980a] C. Wright. Realism, truth-value links, other minds and the past. *Ratio*, **22**, 112–132, 1980.
- [Wright, 1981] C. Wright. Dummett and revisionism. *Philosophical Quarterly*, **31**, 47–67, 1981.
- [Wright, 1982] C. Wright. Strict finitism. *Synthese*, **51**, 203–282, 1982.
- [Zucker and Tragresser, 1978] J. Zucker and R. S. Tragresser. The adequacy problem for inferential logic. *J. Philosophical Logic*, **7**, 501–516, 1978.



## GOAL-ORIENTED DEDUCTIONS

## 1 INTRODUCTION

The topic of this chapter is to present a general methodology for automated deduction inspired by the logic programming paradigm. The methodology can and has been applied to both classical and non-classical logics. It comes without saying that the landscape of non-classical logics applications in computer science and artificial intelligence is now wide and varied, and this Handbook itself is a witness this fact. We will survey the application of goal-directed methods to classical, intuitionistic, modal, and substructural logics. For background information about these logical systems we refer to other chapters of this Handbook and to [Fitting, 1983; Anderson and Belnap, 1975; Anderson *et al.*, 1992; Gabbay, 1981; Troelstra, 1969; Dummett, 2001; Restall, 1999]. Our treatment will be confined to the propositional level.<sup>1</sup>

In the area of automated deduction and proof-theory there are several objectives which can be pursued. Methods suitable for one task are not necessarily the best ones for another. Consider propositional classical logic and the following tasks:

1. check if a randomly generated set of clauses is unsatisfiable;
2. given a formula  $A$  check whether  $A$  is valid;
3. given a set  $\Gamma$  containing say 5,000 formulas and a formula  $A$  check whether  $\Gamma \vdash A$ ;
4. (saturation) given a set of formulas  $\Gamma$  generate all atomic propositions which are entailed by  $\Gamma$ ;
5. (abduction) given a formula  $A$  and a set of formulas  $\Gamma$  such that  $\Gamma \not\vdash A$ , find a minimal set of atomic propositions  $S$  such that  $\Gamma \cup S \vdash A$  and satisfies some other constraints.

It is not difficult to see that all these problems can be reduced one to the other. However, it is quite likely that we need different methods to address each one of them efficiently. Consider task 3:  $\Gamma$  may represent a ‘deductive database’ and  $A$  a query. It might be that the formulas of  $\Gamma$  have a simple/uniform structure and only a small subset of the formula of  $\Gamma$  are relevant for getting a proof of  $A$  (if any): thus a general general

---

<sup>1</sup>The reader of the chapter of Basin and Matthews on logical frameworks can regard our chapter as a goal directed logical framework done in the *object level*.

SAT-algorithm applied  $\Gamma \cup \{\neg A\}$  might not be the most natural method, we would prefer a method capable of concentrating on the relevant data and ignoring the rest of  $\Gamma$ . Similar considerations applies to the other tasks. For instance in the case of abduction, we would likely calculate the set of abductive assumptions  $S$  from failed attempts to prove the data  $A$ , rather than guess an  $S$  arbitrarily and then check if it works. Moreover, in some applications we are not only interested to know whether a formula is valid or not, but also to find (and inspect) a proof of it in an understandable format. The goal-directed approach to deduction is useful to support deduction from large databases, abduction procedures, and proof search.

In a few words, the goal-directed paradigm is the same as the one underlying logic programming. The deduction process can be described as follows: we have a structured collection of formulas (called a database)  $\Delta$  and a goal formula  $A$ , and we want to know whether  $A$  follows from  $\Delta$  or not, in a specific logic. Let us denote by

$$\Delta \vdash^? A$$

the query ‘does  $A$  follows from  $\Delta$ ?’ (in a given logic). The deduction is *goal-directed* in the sense that the next step in a proof is determined by the form of the current goal: the goal is stepwise decomposed, according to its logical structure, until we reach its atomic constituents. An atomic goal  $q$  is then matched with the ‘head’ of a formula  $G' \rightarrow q$  (if any, otherwise we fail) in the database, and its ‘body’  $G'$  is asked in turn. This step can be understood as a resolution step, or as a generalized Modus Tollens. We will see that we can extend this backward reasoning, goal-directed paradigm to most non-classical logics. We can have a logic programming-like proof system presentation for classical, intuitionistic, modal, and substructural logics.

Here is a plan of the chapter: we start revising Horn classical logic as a motivating example, we then consider intuitionistic logic and full classical logic. Then we consider modal logics and substructural logics.

### *Notation and basic notions*

#### *Formulas*

By a propositional language  $\mathcal{L}$ , we denote the set of propositional formulas built from a denumerable set  $Var$  of propositional variables by applying the propositional connectives  $\neg, \wedge, \vee, \rightarrow, \perp$ .

Unless stated otherwise, we denote propositional variables (also called *atoms*) by lower case letters, and arbitrary formulas by upper case letters. We assign a *complexity*  $cp(A)$  to each formula  $A$  (as usual):

$$\begin{aligned} cp(q) &= 0 \text{ if } q \text{ is an atom,} \\ cp(\neg A) &= 1 + cp(A), \\ cp(A * B) &= cp(A) + cp(B) + 1, \text{ where } * \in \{\wedge, \vee, \rightarrow\}. \end{aligned}$$

(*Formula substitution*) We define the notion of substitution of an atom  $q$  by a subformula  $B$  within a formula  $A$ . This operation is denoted by  $A[q/B]$ .

$$p[q/B] = \begin{cases} p & \text{if } p \neq q \\ B & \text{if } p = q \end{cases}$$

$$(\neg A)[q/B] = \neg A[q/B]$$

$$(A * C)[q/B] = A[q/B] * C[q/B] \text{ where } * \in \{\wedge, \vee, \rightarrow\}.$$

### *Implicational formulas*

In much of the chapter we will be concerned with *implicational formulas*. These formulas are generated from a set of atoms by the only connective  $\rightarrow$ . We adopt some specific notations for them. We sometimes distinguish the *head* and the *body* of an implicational formula.<sup>2</sup> The head of a formula  $A$  is its rightmost nested atom, whereas the body is the *list* of the antecedents of its head. Given a formula  $A$ , we define  $Head(A)$  and  $Body(A)$  as follows:

$$\begin{aligned} Head(q) &= q, \text{ if } q \text{ is an atom,} \\ Head(A \rightarrow B) &= Head(B). \end{aligned}$$

$$\begin{aligned} Body(q) &= (), \text{ if } q \text{ is an atom,} \\ Body(A \rightarrow B) &= (A) * Body(B), \end{aligned}$$

where  $(A) * Body(B)$  denotes the list beginning with  $A$  followed by  $Body(B)$ .

Dealing with implicational formulas, we assume that implication associates on the right, i.e. we write

$$A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_{n-1} \rightarrow A_n,$$

instead of

$$A_1 \rightarrow (A_2 \rightarrow \dots \rightarrow (A_{n-1} \rightarrow A_n) \dots).$$

It turns out that every implicational formula  $A$  can be written as

$$A_1 \rightarrow A_2 \rightarrow \dots \rightarrow A_n \rightarrow q,$$

where we obviously have.

$$Head(A) = q \quad \text{and} \quad Body(A) = (A_1, \dots, A_n).$$

---

<sup>2</sup>This terminology is reminiscent of logic programming [Lloyd, 1984].

## 2 PROPOSITIONAL HORN DEDUCTION

To explain what we mean by goal-directed deduction style, we begin by recalling standard propositional Horn deductions. This type of deduction is usually interpreted in terms of classical resolution, but it is not the only possible interpretation.<sup>3</sup> The data are represented by a set of propositional Horn clauses, which we write as

$$a_1 \wedge \dots \wedge a_n \rightarrow b.$$

(or as  $a_1 \rightarrow \dots \rightarrow a_n \rightarrow b$ , according to the previous convention). The  $a_i$  are just propositional variables and  $n \geq 0$ . In case  $n = 0$ , the formula reduces to  $b$ . This formula is equivalent to:

$$\neg a_1 \vee \dots \vee \neg a_n \vee b.$$

Let  $\Delta$  be a set of such formulas, we can give a calculus to derive formulas, called ‘goals’ of the form  $b_1 \wedge \dots \wedge b_m$ . The rules are:

- $\Delta \vdash^? b$  succeeds if  $b \in \Delta$ ;
- $\Delta \vdash^? A \wedge B$  is reduced to  $\Delta \vdash^? A$  and  $\Delta \vdash^? B$ ;
- $\Delta \vdash^? q$  is reduced to
- $\Delta \vdash^? a_1 \wedge \dots \wedge a_n$ , if there is a clause in  $\Delta$  of the form

$$a_1 \wedge \dots \wedge a_n \rightarrow q.$$

The main difference from the traditional logic programming convention is that in the latter conjunction is eliminated and a goal is kept as a sequence of atoms  $b_1, \dots, b_m$ . The computation does not split because of conjunction, all the subgoals  $b_i$  are kept in parallel, and when some  $b_i$  succeeds (that is  $b_i \in \Delta$ ) it is deleted from the sequence. To obtain a real algorithm we should specify in which order we scan the database when we search for a clause whose head matches the goal. Let us see an example.

EXAMPLE 1. Let  $\Delta$  contain the following clauses

- (1)  $a \wedge b \rightarrow g$ ,
- (2)  $t \rightarrow g$ ,
- (3)  $p \wedge q \rightarrow t$ ,
- (4)  $h \rightarrow q$ ,
- (5)  $c \rightarrow d$ ,
- (6)  $c \wedge f \rightarrow a$ ,
- (7)  $d \wedge a \rightarrow b$ ,
- (8)  $a \rightarrow p$ ,
- (9)  $f \wedge t \rightarrow h$ ,
- (10)  $c$ ,
- (11)  $f$ .

---

<sup>3</sup>For a survey on foundations of logic programming, we refer to [Lloyd, 1984] and to [Gallier, 1987].

A derivation of  $g$  from  $\Delta$  can be displayed in the form of a tree and it is shown in Figure 1. The number in front of every non-leaf node indicates the clause of  $\Delta$  which is used to reduce the atomic goal in that node.

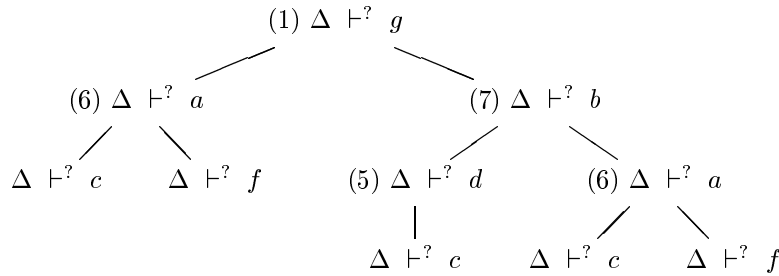


Figure 1. Derivation of Example 1.

We can make a few observations. First, we do not need to consider the whole database, it might even be infinite, and the derivation would be exactly the same; irrelevant clauses, that is those whose ‘head’ do not match with the current goal are ignored. The derivation is driven by the goal in the sense that each step in the proof simply replaces the current goal with the next one.

Notice also that in this specific case there is no other way to prove the goal, and the sequence of steps is entirely determined.

Two weak points of the method can also be noticed. Suppose that when asking for  $g$  we use the second formula, then we continue asking for  $t$ , then for  $h$ , and then we are lead to ask for  $t$  again. We are in a loop. An even simpler situation is the following

$$p \rightarrow p \vdash? p.$$

We can keep on asking for  $p$  without realizing that we are in a loop. To deal with this problem we should add a mechanism which ensures termination.

Another problem which has a bearing on the efficiency of the procedure is that a derivation may contain redundant subtrees. This occurs when the same goal is asked several times. In the previous example it happens with the subgoal  $a$ . In this case, the global derivation contains multiple subderivations of the same goal. It would be better to be able to remember whether a goal has already been asked (and succeeded) in order to avoid the duplication of its derivation. Whereas the problem of termination is crucial in the evaluation of the method (if we are interested in getting an answer eventually), the problem of redundancy will not be considered in this chapter. However, avoiding the type of redundancy we have described has a

dramatic effect on the efficiency of the procedure, for redundant derivations may grow exponentially with the size of the data.

Although the goal-directed procedure does not necessarily produce the shortest proofs, and does not always terminate, still it has the advantage that proofs, when they exist, are easily found. In the next sections we will see how to extend these type of proof systems to several families of non-classical logics.

### 3 INTUITIONISTIC AND CLASSICAL LOGICS

#### 3.1 Intuitionistic logic

Intuitionistic logic is the most known alternative to classical logic. For background motivation and information we refer to [Troelstra, 1969; Gabbay, 1981]. The reason why we initiate our tour from intuitionistic logic is *simplicity*. A proof procedure for the propositional implicational fragment of intuitionistic logic is just a minor extension of the Horn case. Moreover, the relation with the semantics, the role of cut, the problems, and the possible refinements are better understood for intuitionistic logic.

We recall a Hilbert style axiomatisation of the intuitionistic propositional calculus and the standard Kripke semantics for it. The axiomatization is obtained by considering the following set of axioms and rules we denote by **I**:<sup>4</sup>

1.  $(A \rightarrow B \rightarrow C) \rightarrow (A \rightarrow B) \rightarrow A \rightarrow C$
2.  $A \rightarrow B \rightarrow A$
3.  $A \rightarrow B \rightarrow (A \wedge B)$
4.  $A \wedge B \rightarrow A$
5.  $A \wedge B \rightarrow B$
6.  $(A \rightarrow C) \rightarrow (B \rightarrow C) \rightarrow (A \vee B \rightarrow C)$
7.  $A \rightarrow A \vee B$
8.  $A \rightarrow B \vee A$
9.  $\perp \rightarrow A$ .

In addition, it contains the *Modus Ponens* rule:

$$\frac{\vdash A \quad \vdash A \rightarrow B.}{\vdash B}$$

Negation is considered as a derived operator, by  $\neg A \stackrel{def}{=} A \rightarrow \perp$ .

Given a set of formulas  $\Delta$ , we can define  $A$  is derivable from  $\Delta$ ,  $\Delta \vdash A$  by the axiom systems above in the customary way.

---

<sup>4</sup>This axiom system is *separated*, that is to say, any theorem containing  $\rightarrow$  and a set of connectives  $S \subseteq \{\wedge, \vee, \neg, \perp\}$  can be proved by using the implicational axioms together with the axiom groups containing just the connectives in  $S$ .

If add to **I** any of the axioms below we get classical logic **C**:

**Alternative axioms for classical logic**

1. (Peirce's axiom)  $((A \rightarrow B) \rightarrow A) \rightarrow A$
2. (double negation)  $\neg\neg A \rightarrow A$ ,
3. (excluded middle)  $\neg A \vee A$ ,
4. ( $\vee, \rightarrow$ -distribution)  $[(A \rightarrow (B \vee C)) \rightarrow [(A \rightarrow B) \vee C]$ .

In particular, the addition of Peirce's axiom to the implicational axioms of intuitionistic logic give us an axiomatisation of classical implication.

We introduce a standard model-theoretic semantics of intuitionistic logic, called Kripke semantics.

**DEFINITION 2.** Given a propositional language  $\mathcal{L}$ , a Kripke model for  $\mathcal{L}$  is a structure of the form  $M = (W, \leq, V)$ , where  $W$  is a non-empty set,  $\leq$  is a reflexive and transitive relation on  $W$ ,  $V$  is a function of type:  $W \rightarrow Pow(Var_{\mathcal{L}})$ , that is  $V$  maps each element of  $W$  to a set of propositional variables of  $\mathcal{L}$ . We assume the following conditions:

- (1)  $w \leq w'$  implies  $V(w) \subseteq V(w')$ ;
- (2)  $\perp \notin V(w)$ , for all  $w \in W$ .

Given  $M = (W, \leq, V)$ ,  $w \in W$ , for any formula  $A$  of  $\mathcal{L}$ , we define ' $A$  is true at  $w$  in  $M$ ', denoted by  $M, w \models A$  by the following clauses:

- $M, w \models q$  iff  $q \in V(w)$ ;
- $M, w \models A \wedge B$  iff  $M, w \models A$  and  $M, w \models B$ ;
- $M, w \models A \vee B$  iff  $M, w \models A$  or  $M, w \models B$ ;
- $M, w \models A \rightarrow B$  iff for all  $w' \geq w$ , if  $M, w' \models A$  then  $M, w' \models B$ ;
- $M, w \models \neg A$  iff for all  $w' \geq w$   $M, w' \not\models A$ .

We say that  $A$  is *valid* in  $M$  if  $M, w \models A$ , for all  $w \in W$  and we denote this by  $M \models A$ . We say that  $A$  is *valid* if it is valid in every Kripke model  $M$ . We also define a notion of entailment between sets of formulas and formulas. Let  $\Gamma = \{A_1, \dots, A_n\}$  be a set of formulas and  $B$  be a formula, we say that  $\Gamma$  *entails*  $B$  denoted by  $\Gamma \models B$ <sup>5</sup> iff for every model  $M = (W, \leq, V)$ , for every  $w \in W$

if  $M, w \models A_i$  for all  $A_i \in \Gamma$ , then  $M, w \models B$ .

---

<sup>5</sup>To be precise, we should write  $\models_{\mathbf{I}}$  (and  $\vdash_{\mathbf{I}}$ ) to denote validity and entailment (respectively, provability and logical consequence) in intuitionistic logic **I**. To avoid burdening the notation, we usually omit the subscript unless there is a risk of confusion.

**THEOREM 3.** *For any set of formulas  $\Gamma$  and formula  $B$ ,  $\Gamma \vdash B$  iff  $\Gamma \models B$ . In particular  $\vdash_I B$  iff  $\models_I B$ .*

Classical interpretations can be thought as degenerated Kripke models  $M = (W, \leq, V)$ , where  $W = \{w\}$ .

### 3.2 Rules for intuitionistic implication

We start by presenting a goal-directed system for the implicational fragment of intuitionistic logic. We will then refine it and we will expand it with the other connectives later on. We give rules in this section to prove statements of the form  $\Delta \vdash A$ , where  $\Delta$  is a set of implicational formulas and  $A$  is an implicational formula. We use the usual conventions and we write  $\Gamma, A$  for  $\Gamma \cup \{A\}$  and  $\Gamma \cup \Delta$  for  $\Gamma \cup \Delta$ . Our rules hence manipulate queries  $Q$  of the form:

$$\Delta \vdash^? A.$$

We call  $\Delta$  the database and  $A$  the goal of the query  $Q$ . We use the symbol  $\vdash^?$  to indicate that we do not know whether the query succeeds or not. On the other hand the success of  $Q$  means that  $\Delta \vdash A$  according to intuitionistic logic.

**DEFINITION 4.**

- (success)  $\Delta \vdash^? q$  succeeds if  $q \in \Delta$ . We say that  $q$  is used in this query.
- (implication) from  $\Delta \vdash^? A \rightarrow B$  step to
 
$$\Delta, A \vdash^? B$$
- (reduction) from  $\Delta \vdash^? q$  if  $C \in \Delta$ , with  $C = D_1 \rightarrow D_2 \rightarrow \dots \rightarrow D_n \rightarrow q$  (that is  $Head(C) = q$  and  $Body(C) = (D_1, \dots, D_n)$ ) then step to

$$\Delta \vdash^? D_i, \text{ for } i = 1, \dots, n.$$

We say that  $C$  is used in this step.

A *derivation*  $\mathcal{D}$  of a query  $Q$  is a tree whose nodes are queries. The root of  $\mathcal{D}$  is  $Q$ , and the successors of every non-leaf query are determined by exactly one applicable rule (*implication or reduction*) as described above.

We say that  $\mathcal{D}$  is *successful* if the *success rule* may be applied to every leaf of  $\mathcal{D}$ . We finally say that a query  $Q$  *succeeds* if there is a successful derivation of  $Q$ .

By definition, a derivation  $\mathcal{D}$  might be an infinite tree. However if  $\mathcal{D}$  is successful then it must be finite. This is easily seen from the fact that,



in case of success, the height of  $\mathcal{D}$  is finite and every non-terminal node of  $\mathcal{D}$  has a finite number of successors, because of the form of the rules. Moreover, the databases involved in a deduction need not be finite. In a successful derivation only a finite number of formulas from the database will be used in the above sense.

Notice that the success of a query is defined in a non-deterministic way: a query succeeds if there is a successful derivation. To transform the proof rules into a deterministic algorithm one should give a method to search a successful derivation tree. In this respect we agree that when we come to an atomic goal we first try to apply the success rule and if it fails we try the reduction rule. Then the only choice we have is to indicate which formula, among those of the database whose head matches the current atomic goal, we use to perform a reduction step, if there are more than one. Thinking of the database as a list of formulas, we can choose the first one and remember the point up to which we have scanned the database as a backtracking point. This is exactly as in conventional logic programming [Lloyd, 1984].

EXAMPLE 5. We check

$$b \rightarrow d, a \rightarrow p, p \rightarrow b, (a \rightarrow b) \rightarrow c \rightarrow a, (p \rightarrow d) \rightarrow c \vdash b.$$

Let  $\Gamma = \{b \rightarrow d, a \rightarrow p, p \rightarrow b, (a \rightarrow b) \rightarrow c \rightarrow a, (p \rightarrow d) \rightarrow c\}$ , a successful derivation of  $\Gamma \vdash^? b$  is shown in Figure 2. A quick explanation: (2) is obtained by reduction wrt.  $p \rightarrow b$ , (3) by reduction wrt.  $a \rightarrow p$ , (4) and (8) by reduction wrt.  $(a \rightarrow b) \rightarrow c \rightarrow a$ , (6) by reduction wrt.  $p \rightarrow b$ , (7) by reduction wrt.  $a \rightarrow p$ , (9) by reduction wrt.  $(p \rightarrow d) \rightarrow c$ , (11) by reduction wrt.  $b \rightarrow d$ , (12) by reduction wrt.  $p \rightarrow b$ .

We state some simple, but important, properties of the deduction procedure defined above. The proof of them is left to the reader as an exercise.

PROPOSITION 6.

1. (Identity)  $\Delta \vdash^? G$  succeeds if  $G \in \Delta$ ;
2. (Monotony)  $\Delta \vdash^? G$  succeeds implies  $\Delta, \Gamma \vdash^? G$  succeeds;
3. (Deduction Theorem)  $\Delta \vdash^? A \rightarrow B$  succeeds iff  $\Delta, A \vdash^? B$  succeeds.

The soundness of the proof procedure with respect to the semantics can be proved easily by induction on the height of the computation.

THEOREM 7. If  $\Delta \vdash^? A$  succeeds then  $\Delta \models A$ .

The completeness can be proved in a number of ways. Here we give a semantic proof with respect to the Kripke semantics. The technique is standard: we define a canonical model and we show that provability by the proof procedure coincides with truth in the canonical model.

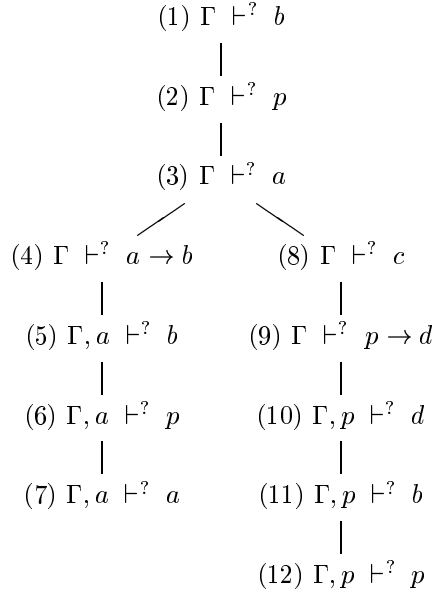


Figure 2. Derivation for Example 5.

The canonical model  $M$  is defined as follows:  $M = (W, \subseteq, \emptyset, V)$ , where  $W$  is the set of finite databases  $\Delta$  over  $\mathcal{L}$  and the evaluation function  $V$  is defined by stipulating

$$V(\Delta) = \{p \text{ atom} : \Delta \vdash^? p \text{ succeeds}\}.$$

By the (Monotony property) it is easy to see that  $M$  satisfies the increasingness condition (2) of Definition 2. The important property is expressed by the following proposition.

**PROPOSITION 8 (Canonical Model Property).** *For any  $\Delta \in W$  and formula  $A \in \mathcal{L}$ , we have:*

$$M, \Delta \models A \text{ iff } \Delta \vdash^? A \text{ succeeds}.$$

Let us attempt a proof of the above proposition, we prove the two directions by a simultaneous induction on the complexity of the formula  $A$ . If  $A$  is an atom then the claim holds by definition.

Let  $A \equiv B \rightarrow C$ . Consider first the direction ( $\Rightarrow$ ), suppose  $M, \Delta \models B \rightarrow C$ , consider  $\Delta' = \Delta, B$ . Then  $\Delta \subseteq \Delta'$ . By (Identity), we have  $\Delta' \vdash^? B$  succeeds, by induction hypothesis we get  $M, \Delta' \models B$ . Thus we get  $M, \Delta' \models C$

and by induction hypothesis  $\Delta' \vdash^? C$  succeeds. Thus  $\Delta, B \vdash^? C$  succeeds. By (Deduction Theorem), we finally get  $\Delta \vdash^? B \rightarrow C$  succeeds.

Conversely ( $\Leftarrow$ ), suppose that  $\Delta \vdash^? B \rightarrow C$  succeeds, by (Deduction theorem) we get that  $\Delta, B \vdash^? C$  succeeds. Let  $\Gamma$  be such that  $\Delta \subseteq \Gamma$  and  $M, \Gamma \models B$ , we have to show that  $M, \Gamma \models C$ . By induction hypothesis we have  $\Gamma \vdash^? B$  succeeds. We know that

- (1)  $\Gamma \vdash^? B$  and (2)  $\Delta, B \vdash^? C$  succeed.

We could easily conclude if from (1) and (2), we could infer that (3)  $\Delta, \Gamma \vdash^? C$  succeeds: namely, since  $\Delta \subseteq \Gamma$ , (3) is equivalent to  $\Gamma \vdash^? C$  succeeds. Thus by induction hypothesis we would get  $M, \Gamma \models C$ .

The question is: is it legitimate to conclude (3) from (1) and (2)? The answer is 'yes' and it will be shown hereafter. This property is well-known and is called *Cut*. Thus, given the properties of deduction theorem, identity, the canonical model property can be derived from cut. We may observe that the opposite also holds, i.e. that cut can be derived by the canonical model property. To see this suppose that  $\Gamma, B \vdash^? C$  and  $\Delta \vdash^? B$  succeeds. We get  $\Gamma \vdash^? B \rightarrow C$  succeeds. Thus by the canonical model property we get  $M, \Gamma \models B \rightarrow C$  and  $M, \Delta \models B$ . Since, trivially,  $\Delta \subseteq \Gamma \cup \Delta$ , by the condition (1) of Definition 2, we have  $M, \Gamma \cup \Delta \models B$ ; so that we obtain  $M, \Gamma \cup \Delta \models C$ , being  $\Gamma \subseteq \Gamma \cup \Delta$ . By the canonical model property we can conclude that  $\Gamma, \Delta \vdash^? C$  succeeds. The equivalence between the canonical model property and cut has been observed in [Miller, 1992].

The completeness is an immediate consequence of the canonical model property.

**THEOREM 9 (Completeness for I).** *If  $\Delta \models A$ , then  $\Delta \vdash^? A$  succeeds.*

**Proof.** If  $\Delta \models A$  holds, the entailment holds in particular in the canonical model  $M$ . thus for every  $\Gamma \in W$  if  $M, \Gamma \models B$  for every  $B \in \Delta$ , we have  $M, \Gamma \models A$ .

By (identity) we have  $\Delta \vdash^? B$  succeeds for every  $B \in \Delta$ . Thus by the canonical model property  $M, \Delta \models B$  for every  $B \in \Delta$ . We hence obtain  $M, \Delta \models A$  and by the canonical model property again  $\Delta \vdash^? A$  succeeds. ■

We have still to show that cut is admissible.

**THEOREM 10 (Admissibility of Cut).** *If  $\Delta, A \vdash^? B$  and  $\Gamma \vdash^? A$  succeed, then also  $\Delta, \Gamma \vdash^? B$  succeeds.*

**Proof.** Assume (1)  $\Delta, A \vdash^? B$  and (2)  $\Gamma \vdash^? A$  succeed.

The theorem is proved by induction on lexicographically-ordered pairs  $(c, h)$ , where  $c = cp(A)$ , and  $h$  is the height of a successful derivation of (1), that is of  $\Delta, A \vdash^? B$ . Suppose first  $c = 0$ , then  $A$  is an atom  $p$ , and we proceed by induction on  $h$ . If  $h = 0$ ,  $B$  is an atom  $q$  and either  $q \in \Delta$  or

$q = p = A$ . In the first case, the claim trivially follows by Proposition 6. In the second case it follows by hypothesis (2) and Proposition 6.

Let now  $h > 0$ , then (1) succeeds either by the implication rule or by the reduction rule. In the first case, we have that  $B = C \rightarrow D$  and from  $\Delta, A \vdash^? C \rightarrow D$  we step to  $\Delta, A, C \vdash^? D$ , which succeeds by a derivation  $h'$  shorter than  $h$ . Since  $(0, h') < (0, h)$ , by the induction hypothesis we get that  $\Delta, \Gamma, C \vdash^? D$ , succeeds, whence  $\Delta, \Gamma \vdash^? C \rightarrow D$  succeeds too. Let (1) succeed by reduction with respect to a formula  $C \in \Delta$ . Since  $A$  is an atom,  $C \neq A$ . Then  $B = q$  is an atom. We let  $C = D_1 \rightarrow \dots \rightarrow D_k \rightarrow q$ . We have for  $i = 1, \dots, k$

$$\Delta, A \vdash^? D_i \text{ succeeds by a derivation of height } h_i < h.$$

Since  $(0, h_i) < (0, h)$ , we may apply the induction hypothesis and obtain

$$(a_i) \Delta, \Gamma \vdash^? D_i \text{ succeeds, for } i = 1, \dots, k.$$

Since  $C \in \Delta \cup \Gamma$ , from  $\Delta, \Gamma \vdash^? q$  we can step to  $(a_i)$  and succeed. This concludes the case of  $(0, h)$ .

If  $c$  is arbitrary and  $h = 0$  the claim is trivial. Let  $c > 0$  and  $h > 0$ . The only difference with the previous cases is when (1) succeeds by reduction with respect to  $A$ . Let us see that case. Let

$$A = D_1 \rightarrow \dots \rightarrow D_k \rightarrow q \text{ and } B = q.$$

Then we have for  $i = 1, \dots, k$   $\Delta, A \vdash^? D_i$  succeeds by a derivation of height  $h_i < h$ . Since  $(c, h_i) < (c, h)$ , we may apply the induction hypothesis and obtain

$$(b_i) \Delta, \Gamma \vdash^? D_i \text{ succeeds for } i = 1, \dots, k.$$

By hypothesis (2) we can conclude that

$$(3) \Gamma, D_1, \dots, D_k \vdash^? q \text{ succeeds by a derivation of arbitrary height } h'.$$

Notice that each  $D_i$  has a smaller complexity than  $A$ , that is  $cp(D_i) = c_i < c$ . Thus  $(c_1, h') < (c, h)$ , and we can cut on (3) and  $(b_1)$ , so that we obtain

$$(4) \Delta, \Gamma, D_2, \dots, D_k \vdash^? q \text{ succeeds with some height } h''.$$

Again  $(c_2, h'') < (c, h)$ , so that we can cut  $(b_2)$  and (4). By repeating the same argument up to  $k$  we finally obtain

$$\Delta, \Gamma \vdash^? q \text{ succeeds.}$$

This concludes the proof. ■

We have given a semantic proof of the completeness of the proof procedure for the implicational fragment of intuitionistic logic. We have given all details (although they are rather easy) since this easy case works as a paradigm for other logics and richer languages. To summarize, the recipe for the semantic proof is the following: (1) give the right formulation of cut and show that it is admissible, (2) define a canonical model, (3) prove the canonical model property as in Proposition 8 and derive the completeness as in Theorem 9.

There are however other ways to prove the completeness depending on the chosen presentation of the logic. If we have a Hilbert-style axiomatization, we can show that every atomic instance of any axiom succeeds, and then show that the set of succeeding formulas is closed under formula substitution and modus ponens (supposing that these properties hold, which is the case for all the logics considered in this chapter). To prove the closure under MP we need again cut. Another possibility, if we have a presentation of the logic in terms of consequence relation rules, like a sequent calculus, we can show that every rule is admissible.

### 3.3 Loop-free and bounded resource deduction

In the previous section, we have introduced a goal-directed proof method for intuitionistic implicational logic. This method does not give a decision procedure, as it can easily loop: consider the trivial query

$$q \rightarrow q \vdash^? q$$

this query is reduced to itself by the reduction rule, so that the computation does not stop. There are two different strategies to deal with looping computation and termination.

One possibility is to detect the loop and stop the computation as soon as it is detected. The other possibility is to *prevent* any loop by constraining the use of the formulas. To perform loop-checking we need to consider the sequence of goals *and* the relative database from which they have been asked. A moment of reflection shows that it is enough to record only the atomic goals, since the non-atomic ones will always reduce to the same atomic goals by the implication rule. The other relevant information is the database from which they are asked. The same atomic goal may be asked from different databases and such a repetition does not mean that the computation necessarily loops:

EXAMPLE 11.

$$\begin{array}{rcl} & \vdash^? & ((c \rightarrow a) \rightarrow c) \rightarrow (c \rightarrow a) \rightarrow a \\ (c \rightarrow a) \rightarrow c & \vdash^? & (c \rightarrow a) \rightarrow a \\ (c \rightarrow a) \rightarrow c, c \rightarrow a & \vdash^? & a \end{array}$$

$$\begin{array}{rcl}
(c \rightarrow a) \rightarrow c, c \rightarrow a & \vdash^? & c \\
(c \rightarrow a) \rightarrow c, c \rightarrow a & \vdash^? & c \rightarrow a \\
(c \rightarrow a) \rightarrow c, c \rightarrow a, c & \vdash^? & a \\
(c \rightarrow a) \rightarrow c, c \rightarrow a, c & \vdash^? & c \\
& & \textit{success}
\end{array}$$

As you can see the goal **a** repeats twice along the same (and unique) branch, but the second time is asked from an increased database ( $c$  has been added). This does not represent a case of loop; we have a loop when we reask the same (atomic) goal from the same database. To detect the loop we need also to record the involved databases. For each computation branch we could record every pair (*atomic goal-database*) in a history list, so that, before making a reduction step, we check whether the same pair is already in the history list. This loop checking ensures termination. The reason is simple, but important. In the case of intuitionistic logic, the database can be regarded as a *set* of formulas (as we have done) so that adding one or more times the same formula does not matter, i.e. the database does not change. This gives decidability: there cannot be infinitely many different databases occurring in one computation branch (supposing that the initial one is finite) since at most a database can contain all subformulas of the initial query. Thus any loop will be detected.

However, the fact that the database is a *set* of formulas can be used to devise a more efficient loop checking mechanism in which we do not have to record the database itself. The idea is simple: we have a loop whenever we repeats the same goal from the same data. Thus we need to record only the atomic goals which are asked from the same database. Whenever we change the database we clear the history. The database changes (grows) when we add a formula not occurring in it by means of the  $\rightarrow$ -rule. This improved loop-checking procedure has been proposed in [Heudering *et al.*, 1996] to the purpose of obtaining a terminating sequent calculus for intuitionistic logic. Let  $H$  be the list of past atomic goals. The computation rules are modified as follows:

**Rule 1 for  $\rightarrow$**   
 $\Delta \vdash^? A \rightarrow B, H$  succeeds  
if  $A \notin \Delta$  and  $\Delta, A \vdash^? B, \emptyset$  succeeds.

**Rule 2 for  $\rightarrow$**   
 $\Delta \vdash^? A \rightarrow B, H$  succeeds  
if  $A \in \Delta$  and  $\Delta \vdash^? B, H$  succeeds.

**Reduction Rule**  
 $\Delta \vdash^? q, H$  succeeds  
if  $q \notin H$  and for some  $C_1 \rightarrow \dots \rightarrow C_n \rightarrow q$  in  $\Delta$  we have that for all  $i$   
 $\Delta \vdash^? C_i, H * (q)$  succeeds.

EXAMPLE 12.

$$\begin{array}{rcl}
& \vdash^? & ((p \rightarrow q) \rightarrow q) \rightarrow (q \rightarrow p) \rightarrow p, \emptyset \\
(p \rightarrow q) \rightarrow q & \vdash^? & (q \rightarrow p) \rightarrow p, \emptyset \\
(p \rightarrow q) \rightarrow q, q \rightarrow p & \vdash^? & p, \emptyset \\
(p \rightarrow q) \rightarrow q, q \rightarrow p & \vdash^? & q, (p) \\
(p \rightarrow q) \rightarrow q, q \rightarrow p & \vdash^? & p \rightarrow q, (p, q) \\
(p \rightarrow q) \rightarrow q, q \rightarrow p, p & \vdash^? & q, \emptyset \\
(p \rightarrow q) \rightarrow q, q \rightarrow p & \vdash^? & p \rightarrow q, (q) \\
(p \rightarrow q) \rightarrow q, q \rightarrow p & \vdash^? & q, (q) \\
& & \text{fail}
\end{array}$$

In this way one is able to detect a loop (the same atomic goal repeats from the same database), without having to record each pair (database goal).

As we have remarked at the beginning of this section loop-checking is not the only way to ensure termination. A loop is created because a formula used in one reduction step remains available for further reduction steps. It can be used as many times as we wish.

Let us adopt the point of view that each database formula can be used at most once. Thus our rule for reduction becomes

$$\begin{array}{l}
\text{from } \Delta \vdash^? q, \\
\text{if there is } B \in \Delta, \text{ with } B = C_1 \rightarrow \dots \rightarrow C_n \rightarrow q, \text{ step to} \\
\Delta - \{B\} \vdash^? C_i \text{ for } i = 1, \dots, n.
\end{array}$$

The item  $B$  is thus thrown out as soon as it is used.

Let us call such a computation *locally linear computation*, as each formula can be used at most once in each path of the computation. That is why we are using the word ‘locally’. One can also have the notion of (globally) linear computation, in which each formula can be used exactly once in the entire computation tree.

Since we take care of usage of formulas, it is natural to regard multiple copies of the same formula as *distinct*. This means that databases can now be considered as *multisets* of formulas. In order to keep the notation simple, we use the same notation as in the previous section. From now on,  $\Gamma, \Delta$ , etc. will range on multisets of formulas, and we will also write  $\Gamma, \Delta$  to denote the union multiset of  $\Gamma$  and  $\Delta$ , that is  $\Gamma \sqcup \Delta$ . To denote a multiset  $[A_1, \dots, A_n]$ , if there is no risk of confusion we will simply write  $A_1, \dots, A_n$ .

We present three notions of proof: (1) the goal-directed computation for intuitionistic logic, (2) the locally linear goal-directed computation (LL-computation), (3) linear goal-directed computation.

DEFINITION 13. We give the computation rules for a query:  $\Gamma \vdash^? G$ , where  $\Gamma$  is a multiset of formulas and  $G$  is a formula.

- (success)  $\Delta \vdash^? q$  immediately succeeds if the following holds:

1. for intuitionistic and LL- computation,  $q \in \Delta$ ,
2. for linear computation  $\Delta = q$ .

- (implication) From  $\Delta \vdash^? A \rightarrow B$ , we step to

$$\Delta, A \vdash^? B.$$

- (reduction) If there is a formula  $B \in \Delta$  with

$$B = C_1 \rightarrow \dots \rightarrow C_n \rightarrow q$$

then from  $\Delta \vdash^? q$ , we step, for  $i = 1, \dots, n$  to

$$\Delta_i \vdash^? C_i,$$

where the following holds

1. in the case of intuitionistic computation,  $\Delta_i = \Delta$ ;
2. in the case of locally linear computation,  $\Delta_i = \Delta - [B]$ ;
3. in the case of linear computation  $\sqcup_i \Delta_i = \Delta - [B]$ .

It can be shown that the monotony property holds for the LL-computation whereas it does not for the linear computation. Let  $\mathbf{Q}_L$ ,  $\mathbf{Q}_{LL}$  and  $\mathbf{Q}_I$  denote respectively the set of succeeding queries in the linear, in the locally linear, and in the intuitionistic computation, then we have

$$\mathbf{Q}_L \subseteq \mathbf{Q}_{LL} \subseteq \mathbf{Q}_I$$

The examples below shows that these inclusions are proper.

EXAMPLE 14.

1. We reconsider Example 11:  $c \rightarrow a, (c \rightarrow a) \rightarrow c \vdash^? a$   
The formula  $c \rightarrow a$  has to be used twice in order for  $a$  to succeed. Thus the query fails in the locally linear computation, but it succeeds in the intuitionistic one. This example can be generalized as follows, let:
2. Let  $A_0 = c$   
 $A_{n+1} = (A_n \rightarrow a) \rightarrow c$ .

Consider the following query:

$$A_n, c \rightarrow a \vdash^? a$$

The formula  $c \rightarrow a$  has to be used  $n + 1$  times locally.





We recall that a formula  $A$  is *Horn* if it is an atom, or has the form  $p_1 \rightarrow \dots \rightarrow p_n \rightarrow q$ , where all  $p_i$  and  $q$  are atoms.

PROPOSITION 16. *The locally-linear procedure is complete for Horn formulas.*

### 3.4 Bounded restart rule for intuitionistic logic

We have seen that the locally linear restriction does not retain completeness with respect to intuitionistic provability, as there are examples where formulas need to be used locally several times. We show that we can retain completeness for intuitionistic logic by adding another computation rule. The new rule is called the *bounded restart rule*.

Let us examine more closely why we needed in Example 11 the formula  $c \rightarrow a$  several times. The reason was that from other formulas, we got the goal  $\vdash^? a$  and we wanted to use  $c \rightarrow a$  to continue to the goal  $\vdash^? c$ . The formula  $c \rightarrow a$  was no longer available because it had already been used. In other words,  $\vdash^? a$  had already been asked and  $c \rightarrow a$  was used. This means that the next goal after  $\vdash^? a$  in the history was  $\vdash^? c$ .

If  $H$  is the history of the atomic goals asked, then somewhere in  $H$  there is  $\vdash^? a$  and immediately afterwards  $\vdash^? c$ .

We can therefore compensate for the reuse of  $c \rightarrow a$  by allowing ourselves to go back in the history to where  $\vdash^? a$  was, and allow ourselves to ask all atomic goals that come afterwards. We call this type of move *bounded restart*.

The previous example suggests the following new computation with bounded restart rule.

DEFINITION 17. [Locally linear computation with bounded restart] In the computation with bounded restart, the queries have the form  $\Delta \vdash^? G, H$ , where  $\Delta$  is a multiset of formulas and the history  $H$  is a sequence of atomic goals. The rules are as follows:

- (success)  $\Delta \vdash^? q, H$  succeeds if  $q \in \Delta$ ;
- (implication) from  $\Delta \vdash^? A \rightarrow B, H$  step to  $\Delta, A \vdash^? B, H$ ;
- (reduction) from  $\Delta \vdash^? q, H$  if  $C = D_1 \rightarrow D_2 \rightarrow \dots \rightarrow D_n \rightarrow q \in \Delta$ , then we step to

$$\Delta - [C] \vdash^? D_i, H * (q) \text{ for } i = 1, \dots, n;$$

- (bounded restart) from  $\Delta \vdash^? q, H$  step to

$$\Delta \vdash^? q_1, H * (q),$$

provided for some  $H_1, H_2, H_3$ , it holds  $H = H_1 * (q) * H_2 * (q_1) * H_3$ , where each  $H_i$  may be empty.

EXAMPLE 18.

$$\begin{array}{rcl}
& & \vdash^? \quad ((c \rightarrow a) \rightarrow c) \rightarrow (c \rightarrow a) \rightarrow a, \emptyset \\
(c \rightarrow a) \rightarrow c & \vdash^? & (c \rightarrow a) \rightarrow a, \emptyset \\
(c \rightarrow a) \rightarrow c, c \rightarrow a & \vdash^? & a, \emptyset \\
(c \rightarrow a) \rightarrow c & \vdash^? & c, (a) \\
& \vdash^? & c \rightarrow a, (a, c) \\
c & \vdash^? & a, (a, c) \\
c & \vdash^? & c, (a, c) \\
& & \text{success}
\end{array}$$

The last step is by bounded restart, it is legal since  $c$  follows  $a$  in the history.

The locally linear computation with bounded restart is sound and complete for intuitionistic logic. However, before stating the theorem, we want to remark about the meaning of the atoms in the history. To make the things easy, let the query be  $\Delta \vdash^? G, (p)$  and  $G$  be atomic. Suppose in addition that the query succeeds by a reduction step. Then  $G$  will be added to the history list just after  $p$ . Let the next goal be  $p$ , so that we can apply the bounded restart rule to the query  $\Delta \vdash^? p, (p, G)$ . The bounded restart step could be performed by reduction if we had the formula  $G \rightarrow p$  in the database. Thus the original query is equivalent to the query  $\Delta, G \rightarrow p \vdash^? G$ . The general correspondence is expressed in the next theorem.

**THEOREM 19** (Soundness and completeness of locally linear computation with bounded restart). *For the computation of Definition 17 we have:*  
 $\Delta \vdash^? G, (p_1, \dots, p_n)$  succeeds iff  $\Delta, G \rightarrow p_n, p_n \rightarrow p_{n-1}, \dots, p_2 \rightarrow p_1 \vdash G$  in intuitionistic logic.

### 3.5 Restart rule for classical logic

It is interesting to note that by adopting a variation of the bounded restart rule we can obtain a proof procedure for implicational classical logic. The variation is obtained by cancelling any restrictions and simply allowing us to ask any earlier atomic goal. We need not keep the history as a sequence, but only as a *set* of atomic goals. The rule becomes

**DEFINITION 20** (Restart rule in the LL-computation). If  $a \in H$ , from  $\Delta \vdash^? q, H$  step to  $\Delta \vdash^? a, H \cup \{q\}$ .

The formal definition of *locally linear computation with restart* is Definition 17 with the additional restart rule above in place of the bounded restart rule.

EXAMPLE 21.

$$\begin{array}{l}
\vdash^? \quad (a \rightarrow b) \rightarrow a \vdash^? a, \emptyset \\
\vdash^? \quad \vdash^? a \rightarrow b, \{a\} \\
a \vdash^? \quad a \vdash^? b, \{a\} \\
a \vdash^? \quad a, \{a, b\} \text{ by restart} \\
\text{success.}
\end{array}$$

The above query fails in the intuitionistic computation. Thus, this example shows that we are getting a logic which is stronger than intuitionistic logic. Namely, we are getting classical logic. This claim has to be properly proved, of course.

If we adopt the basic computation procedure for intuitionistic implication of Definition 4 rather than the LL-computation, we can restrict the restart rule to always choose the *initial goal* as the goal with which we restart. Thus, we do not need to keep the history, but only the initial goal and the rule becomes more deterministic. On the other hand, the price we pay is that we cannot throw out the formulas of database when they are used.

DEFINITION 22 (Simple computation with restart). The queries have the form

$$\Delta \vdash^? G, (G_0),$$

where  $G_0$  is a goal. The computation rules are the same as in the basic computation procedure for intuitionistic implication of Definition 4 plus the following rule

$$\text{(Restart) from } \Delta \vdash^? q, (G_0) \text{ step to } \Delta \vdash^? G_0, (G_0).$$

It is clear that the initial query of any derivation will have the form  $\Gamma \vdash^? A, (A)$ .

Given the underlying computation procedure of Definition 4, restarting from an arbitrary atomic goal in the history is equivalent to restart from the initial goal. This is expressed formally in the next proposition, where we let  $\vdash_{RI}^?$  and  $\vdash_{RA}^?$  be respectively the deduction procedure of Definition 22 and the deduction procedure of Definition 4 extended by the restart rule of Definition 20.

PROPOSITION 23. *For any database  $\Delta$  and formula  $G$ , we have*

$$(1) \Delta \vdash_{RA}^? G, \emptyset \text{ succeeds iff } (2) \Delta \vdash_{RI}^? G, (G) \text{ succeeds.}$$

We show that the proof-procedure obtained by adding the rule of restart from the initial goal to the basic procedure for intuitionistic logic defined in Definition 4 is sound and complete with respect to classical provability.

The idea is to replace the effect of restart from the initial goal by adding a suitable set of formulas which depends on the initial goal. This set of formulas can be seen as representing the negation (or complement) of the goal.

DEFINITION 24. Let  $A$  be any formula. The complement of  $A$ , denoted by  $Cop(A)$  is the following set of formulas:

$$Cop(A) = \{A \rightarrow p \mid p \text{ any atom of the language}\}$$

The set  $Cop(A)$  represents the negation of  $A$  in our implicational language which does not contain neither  $\neg$ , nor  $\perp$ .

The crucial, although easy, fact is that we can replace any application of the restart rule by a reduction step using a formula in  $Cop(A)$ .

LEMMA 25. (1)  $\Gamma \vdash^? A$  succeeds by using restart rule iff (2)  $\Gamma \cup Cop(A) \vdash^? A$  succeeds without using restart, that is by the intuitionistic procedure of Definition 4.

Now we only have to show that  $\Delta \vdash A$  in classical logic iff  $\Delta \cup Cop(A) \vdash^? A$  succeeds by the procedure defined in Definition 4. To this purpose we need the following lemma, which shows that  $Cop(A)$  works as the negation of  $A$ .

LEMMA 26. For any database  $\Delta$  and formulas  $G$  such that  $\Delta \supseteq Cop(G)$ , and for any goal  $A$ , we have if  $\Delta \cup \{A\} \vdash^? G$  and  $\Delta \cup Cop(A) \vdash^? G$  succeed then also  $\Delta \vdash^? G$  succeeds.

THEOREM 27. For any  $\Delta$  and  $A$ , (a) is equivalent to (b) below:

(a)  $\Delta \vdash A$  in classical logic,

(b)  $\Delta \cup Cop(A) \vdash^? A$  succeeds by the intuitionistic procedure defined in Definition 4.

**Proof.** ( $\Leftarrow$ ) Show (b) implies (a).

Assume  $\Delta \cup Cop(A) \vdash^? A$  succeeds. Then by the soundness of the computation procedure we get that  $\Delta \cup Cop(A) \vdash A$  in intuitionistic logic, and hence in classical logic. Since the *proof* is *finite* there is a finite set of the form  $\{A \rightarrow p_1, \dots, A \rightarrow p_n\}$  such that

(a1)  $\Delta, A \rightarrow p_1, \dots, A \rightarrow p_n \vdash A$  (in intuitionistic logic).

We must also have that  $\Delta \vdash A$ , in classical logic, because if there were an assignment  $h$  making  $\Delta$  true and  $A$  false, it would also make  $A \rightarrow p_i$  all true, contradicting (a1).

The above concludes the proof that (b) implies (a).

( $\Rightarrow$ ) Show that (a) implies (b).

We prove that if  $\Delta \cup Cop(A) \vdash^? A$  does not succeed then  $\Delta \not\vdash A$  in classical logic. Let  $\Delta_0 = \Delta \cup Cop(A)$ . We define a sequence of databases  $\Delta_n, n = 1, 2 \dots$  as follows:

Let  $B_1, B_2, B_3, \dots$  be an enumeration of all formulas of the language.

Assume  $\Delta_{n-1}$  has been defined and assume that  $\Delta_{n-1} \vdash^? A$  does not succeed. We define  $\Delta_n$ :

If  $\Delta_{n-1} \cup \{B_n\} \vdash^? A$  does not succeed, let  $\Delta_n = \Delta_{n-1} \cup \{B_n\}$ . Otherwise from Lemma 26 we must have:

$\Delta_{n-1} \cup Cop(B_n) \vdash^? A$  does not succeed.

and so let  $\Delta_n = \Delta_{n-1} \cup Cop(B_n)$ .

Finally, let  $\Delta' = \bigcup_n \Delta_n$

Clearly  $\Delta' \vdash^? A$  does not succeed.

Define an assignment of truth values  $h$  on the atoms of the language by  $h(p) = \mathbf{true}$  iff  $\Delta' \vdash^? p$  succeeds. We now prove that

for any  $B, h(B) = \mathbf{true}$  iff  $\Delta' \vdash^? B$  succeeds,

by induction on  $B$ . For atoms this is the definition.

Let  $B = C \rightarrow D$ . We prove the two directions by simultaneous induction. Suppose  $\Delta' \vdash^? C \rightarrow D$  succeeds. If  $h(C) = \mathbf{false}$ , then  $h(C \rightarrow D) = \mathbf{true}$  and we are done. Thus, assume  $h(C) = \mathbf{true}$ . By the induction hypothesis, it follows that  $\Delta' \vdash^? C$  succeeds. Since, by hypothesis we have that  $\Delta', C \vdash^? D$  succeeds, by cut we obtain that  $\Delta' \vdash^? D$  succeeds, and hence by the induction hypothesis  $h(D) = \mathbf{true}$ .

Conversely, if  $\Delta' \vdash^? C \rightarrow D$  does not succeed, we show that  $h(C \rightarrow D) = \mathbf{false}$ . Let  $Head(D) = q$ , we get

- (1)  $\Delta' \vdash^? D$  does not succeed
- (2)  $\Delta', C \vdash^? q$  does not succeed.

Hence by the induction hypothesis on (1) we have  $h(D) = \mathbf{false}$ . We show that  $\Delta' \vdash^? C$  must succeed. Suppose on the contrary that  $\Delta' \vdash^? C$  does not succeed. Hence  $C \notin \Delta'$ . Let  $B_n = C$  in the given enumeration. Since  $B_n \notin \Delta_n$ , by construction, it must be  $Cop(C) \subseteq \Delta'$ . In particular  $C \rightarrow q \in \Delta'$ , and hence  $\Delta', C \vdash^? q$  succeeds, against (2).

We have shown that  $\Delta' \vdash^? C$  succeeds, whence  $h(C) = \mathbf{true}$ , by the induction hypothesis. Since  $h(C) = \mathbf{false}$ , we obtain  $h(C \rightarrow D) = \mathbf{false}$ .

We can now complete the proof. Since  $\Delta' \vdash^? A$  does not succeed, we get  $h(A) = \mathbf{false}$ . On the other hand, for any  $B \in \Delta \cup Cop(A)$ ,  $h(B) = \mathbf{true}$  (since  $\Delta \cup Cop(A) \subseteq \Delta'$ ) and  $h(A) = \mathbf{false}$ . This means that  $\Delta \cup Cop(A) \not\vdash A$  in classical logic. This complete the proof.  $\blacksquare$

From the above theorem and Lemma 25 we immediately obtain the completeness of the proof procedure with restart from the initial goal.

**THEOREM 28.**  $\Delta \vdash A$  in classical logic iff  $\Delta \vdash^? A, (A)$  succeeds using the restart rule from the initial goal, added to the procedure of Definition 4.

Also the locally linear computation of Definition 17 with the restart rule from any previous goal is sound and complete. A proof is contained in [Gabbay and Olivetti, 2000].

**THEOREM 29.** [Soundness and completeness of locally linear computation with restart]

$\Delta \vdash^? G, H$  succeeds iff  $\Delta \vdash G \vee \bigvee H$  in classical logic.

Observe that we cannot restrict the application of restart to the first atomic goal occurring in the computation, consider:

$$(p \rightarrow q) \rightarrow p, p \rightarrow r \vdash^? r, \emptyset,$$

it succeeds by the following computation:

$$\begin{array}{l} (p \rightarrow q) \rightarrow p, p \rightarrow r \vdash^? r, \emptyset, \\ (p \rightarrow q) \rightarrow p \vdash^? p, \{r\}, \\ \vdash^? p \rightarrow q, \{r, p\}, \\ p \vdash^? q, \{r, p\}, \\ p \vdash^? p, \{r, p\}, \text{ restart from } p. \end{array}$$

However restarting from  $r$ , the first atomic goal would not help.

### 3.6 Termination and complexity

The proof systems based on locally linear computation are a good starting point for designing efficient automated-deduction procedures; on the one hand proof search is guided by the goal, on the other hand derivations have a smaller size since a formula that has to be reused does not create further branching. We now want to remark upon termination of the procedures.

The basic LL- procedure obviously terminates: since formulas are thrown out as soon as they are used in a reduction step, every branch of a given derivation eventually ends with a query which either immediately succeeds, or no further reduction step is possible from it. This was the motivation of the LL-procedure as an alternative to a loop-checking mechanism. Does the (bounded) restart rule preserve this property? As we have stated, it does not, in the sense that a silly kind of loop may be created by restart. Let us consider the following example, here we give the computation for intuitionistic logic, but the example works for the classical case as well:





$\Delta \vdash^? q', H * (q)$ , provided

1. there exists a formula  $C \in \Delta$ , with  $q' = \text{Head}(C)$ ,  $q' \neq q$  and
2.  $q \leq_H q'$  holds.

It is easy to see that the above rule ensures the termination of the procedure preserving its completeness.

Moreover we can reformulate the rules for success and reduction by building the bounded restart rule into them, to this purpose we simply restate:

1. (br-success)  $\Gamma \vdash^? q, H$  succeeds if there exists  $q'$  such that  $q \leq_H q'$  and  $q' \in \Gamma$ ;
2. (br-reduction) From  $\Gamma, C_1 \rightarrow \dots \rightarrow C_n \rightarrow q' \vdash^? q, H$  if  $q \leq_H q'$  steps for  $i = 1, \dots, n$  to  $\Gamma \vdash^? C_i, H * (q)$ .

The proof procedure can be further refined to match the known complexity bound for intuitionistic logic, namely  $O(n \log n)$  space. Observe that the history list may be kept linear in the length of the database+goal: only the leftmost and the rightmost occurrence of any atom in  $H$  are needed for determining  $\leq_H$ . Thus the history length is bounded by  $2 * k$  where  $k$  is the number of atoms occurring in the initial query. The length of each derivation branch is bounded by the length of the initial query and so is the length of each intermediate query. In searching a proof of a given query, we first apply the implication rule if the goal is an implication; if the goal is atomic we try first (br-success) and if it is not applicable we try (br-reduction).

The proof search space can be then described as a tree that contains AND branchings, corresponding to the (br-reduction steps) with multiple subgoals, and OR branchings corresponding to backtracking points, determined by alternative formulas which can be used in the (br-reduction) steps. The latter are branchings in the proof search space, not in the derivation tree. We assume that subgoals are examined and alternatives are scanned in a fixed manner (for instance from left to right).

To achieve a good space complexity bound, we do not store the whole derivation, we rather perform the proof search in a depth first manner expanding one query at a time. We only store one query at a time, the one which is going to be expanded by the rules. Moreover we keep a copy of the initial query. In addition we use a stack to keep track of the AND branchings and backtracking points, if any. We will not enter into the details of how to store the relevant information in the stack entries, it is described in [Gabbay *et al.*, 1999]. We only observe that: (1) since we can index formulas and subformulas of the initial query, each stack entry will not require more than  $O(\log n)$  bits, being  $n$  is the length of the initial query. (2) we have a stack entry for each query occurring along a derivation branch. Thus the depth of the stack is bounded by  $n$  the length of the initial query. In the whole, an

algorithm to search a derivation will not need more than  $O(n \log n)$  space, where  $n$  is the length of the initial query. This matches the known optimal upper-bound for intuitionistic logic.

**THEOREM 30.** *The procedure with bounded restart gives an  $O(n \log n)$ -space decision procedure for the implicational fragment of intuitionistic logic.*

The proof procedure based on bounded restart is almost deterministic except for one crucial point: the choice of the database formula to use in a reduction step. Here a sharp difference between classical and intuitionistic logic arises. In intuitionistic logic, the choice is critical: we could make the wrong choice and then have to backtrack to try an alternative formula. In the case of classical logic, backtracking is not necessary, that is, it does not matter which formula we choose to match an atomic goal in a reduction step.

**LEMMA 31.** *Let*

$$A = A_1 \rightarrow \dots \rightarrow A_n \rightarrow q \text{ and } B = B_1 \rightarrow \dots \rightarrow B_m \rightarrow q.$$

*Then (a) is equivalent to (b):*

(a)  $\Delta, A \vdash^? B_i, H \cup \{q\}$  succeeds for  $i = 1, \dots, m$ ;

(b)  $\Delta, B \vdash^? A_i, H \cup \{q\}$  succeeds for  $i = 1, \dots, n$ .

By the previous lemma we immediately have.

**PROPOSITION 32.** *In any computation of  $\Delta \vdash^? q, H$  with restart, no backtracking is necessary. The atom  $q$  can match with the head of any  $A_1 \rightarrow \dots \rightarrow A_n \rightarrow q \in \Delta$  and success or failure does not depend on the choice of such a formula.*

The parallel property to Lemma 31, Proposition 32 clearly does not hold for the intuitionistic case. This difference gives an intuitive account of the difference of complexity between the intuitionistic and the classical case.<sup>6</sup>

### 3.7 Extending the language

#### Conjunction and negation

In this and the next section we extend the language to the full propositional language. We start by considering conjunction. The addition of conjunction to the propositional language does not change the proof system much. Every formula  $A$  with conjunctions is equivalent in intuitionistic logic to a conjunction of formulas  $\bigwedge_i A_i$ , where  $A_i$  contain no conjunctions. If we

<sup>6</sup>We recall that intuitionistic provability is PSPACE-complete [Statman, 1979], whereas classical provability is CoNP-complete, although the space requirements are the same.

agree to represent a conjunction of formulas as a *set*, the handling of conjunction is straightforward, we can transform every database and goal into sets of formulas without conjunction. Then we extend the proof procedure to sets of goals  $S$ :

From  $\Delta \vdash^? S$  step to  $\Delta \vdash^? A$  for every  $A \in S$ .

The computation rule for conjunction can be stated directly:

$\Delta \vdash^? A \wedge B$  succeeds iff  $\Delta \vdash^? A$  succeeds and  $\Delta \vdash^? B$  succeeds.

We now turn to *negation*. As we have seen, negation can be introduced in classical and intuitionistic logic by adding a constant symbol  $\perp$  for falsity and defining the new connective  $\neg A$  for negation as  $A \rightarrow \perp$ . We will adopt this definition. However, we have to modify the computation rules, because we have to allow for the special nature of  $\perp$ , namely that  $\perp \vdash A$  holds for any  $A$ .

**DEFINITION 33.** [Computations for data and goal containing  $\perp$  for intuitionistic and classical logic] The basic procedure is that one defined in 4, (plus the restart rule for classical logic), with the following modifications:

1. Modify (success) rule to read:  $\Delta \vdash^? q$  immediately succeeds, if  $q \in \Delta$  or  $\perp \in \Delta$ .
2. Modify (reduction rule) to read: from  $\Delta \vdash^? q$  step, for  $i = 1, \dots, n$  to

$$\Delta \vdash^? B_i$$

if there is  $C \in \Delta$  such that  $Head(C) \in \{q, \perp\}$  and  $Body(C) = (B_1, \dots, B_n)$ .

In Definition 33 we have actually defined two procedures. One is the computation without the restart rule for intuitionistic logic with  $\perp$ , and the other is the computation with the restart rule for classical logic. We have to show that the two procedures indeed correctly capture the intended fragment of the respective systems. This is easy to see. The effect of the axiom  $\perp \vdash A$  is built into the computation via the modifications in 1. and 2. of Definition 33 and hence we know we are getting intuitionistic logic. To show that the restart rule yields classical logic, it is sufficient to show that the computation

$$(A \rightarrow \perp) \rightarrow \perp \vdash^? A$$

always succeeds with the restart rule. This can also be easily checked.

To complete the picture we show in the next proposition that the computation of  $\Delta \vdash^? A$  with restart is the same as the computation of  $\Delta \vdash^? (A \rightarrow \perp) \rightarrow A$  without restart. This means that the restart rule

(with original goal  $A$ ) can be effectively implemented by adding  $A \rightarrow \perp$  to the database and using the formula  $A \rightarrow \perp$  and the  $\perp$ -rules to replace uses of the restart rule. The above considerations correspond to the known translation from classical logic to intuitionistic logic, namely:

$\Delta \vdash A$  in classical logic iff  $\Delta \vdash \neg A \rightarrow A$  in intuitionistic logic.

The proof is similar to that one of Lemma 25, namely  $Cop(G)$  is a way of representing  $G \rightarrow \perp$  without using  $\perp$ .

PROPOSITION 34. *For any database  $\Delta$  and goal  $G$ :*

$\Delta \vdash^? G$  succeeds with restart iff  $\Delta \cup \{G \rightarrow \perp\} \vdash^? G$  succeeds without restart.

EXAMPLE 35. We check:

$$(q \rightarrow \perp) \rightarrow \perp \vdash^? q, (q)$$

$$(q \rightarrow \perp) \rightarrow \perp \vdash^? q \rightarrow \perp, (q)$$

$$(q \rightarrow \perp) \rightarrow \perp, q \vdash^? \perp, (q).$$

We cannot use the reduction rule here. So, we fail in intuitionistic logic. In classical logic we can use restart to obtain:

$$(q \rightarrow \perp) \rightarrow \perp, q \vdash^? q, (q)$$

and terminate successfully.

The locally linear computation with bounded restart (respectively restart) is complete for the  $(\rightarrow, \perp, \wedge)$ -fragment of intuitionistic (classical) logic. The termination and complexity analysis of the previous section applies also to this larger fragment.

### *Extension to the whole propositional intuitionistic logic*

To obtain a goal-directed proof method for full intuitionistic propositional logic we must find a way of handling disjunctive data. The handling of disjunction is more difficult than the handling of conjunction and negation. Consider the formula  $a \rightarrow (b \vee (c \rightarrow d))$ . We cannot rewrite this formula in intuitionistic logic to anything of the form  $B \rightarrow q$ , where  $q$  is atomic (or  $\perp$ ).

We therefore have to change our proof procedures to accommodate the general form of an intuitionistic formula with disjunction.

In classical logic disjunctions can be pulled to the outside of formulas using the following equivalences:

1.  $(A \vee B \rightarrow C) \equiv (A \rightarrow C) \wedge (B \rightarrow C)$

$$2. (C \rightarrow A \vee B) \equiv (C \rightarrow A) \vee (C \rightarrow B),$$

where  $\equiv$  denotes logical equivalence in classical logic. 1. is valid in intuitionistic logic but 2. is not valid. We have seen at the beginning of the section, the axioms governing disjunction. In view of those axioms, it is not difficult to devise rules to handle disjunction:

R1: from  $\Delta \vdash^? A \vee B$  step to  $\Delta \vdash^? A$  or to  $\Delta \vdash^? B$ .

R2: from  $\Delta, A \vee B \vdash^? C$  step to  $\Delta, A \vdash^? C$  and to  $\Delta, B \vdash^? C$ .

We can try to incorporate the two rules for disjunction within a goal-directed proof procedure for full intuitionistic logic.

DEFINITION 36. Computation rules for full intuitionistic logic with disjunction.

1. The propositional language contains the connectives  $\wedge, \vee, \rightarrow, \perp$ . Formulas are defined inductively as usual.
2. We define the operation  $\Delta + A$ , for any formula  $A = \bigwedge_i A_i$ , as follows:  $\Delta + A = \Delta \cup \{A_i\}$  provided  $A_i$  are not conjunctions.
3. The computation rules are as follows.

(suc)  $\Delta \vdash^? q$  succeeds if  $q \in \Delta$  or  $\perp \in \Delta$ ;

(conj) from  $\Delta \vdash^? A \wedge B$  step to  $\Delta \vdash^? A$  and to  $\Delta \vdash^? B$ ;

(g-dis) from  $\Delta \vdash^? A \vee B$  step to  $\Delta \vdash^? A$  or to  $\Delta \vdash^? B$ ;

(imp) from  $\Delta \vdash^? A \rightarrow B$  step to  $\Delta + A \vdash^? B$ ;

(red) from  $\Delta \vdash^? G$  if  $G$  is an atom  $q$  or  $G = A \vee B$ , if  $C \in \Delta$ , with  $C = A_1 \rightarrow \dots \rightarrow A_n \rightarrow D$  (where  $D$  is not an implication) step to

(a)  $\Delta \vdash^? A_i$ , for  $i = 1, \dots, n$ , and to

(b)  $\Delta + D \vdash^? G$ .

(c-dis) from  $\Delta, A \vee B \vdash^? C$  step to  $\Delta + A \vdash^? C$  and to  $\Delta + B \vdash^? C$ .

The above rules give a sound and complete system for full intuitionistic logic. However the rules are far from satisfactory, in the sense that the goal-directness is lost. For instance, we must be allowed to perform a reduction step not only when the goal is atomic, but also when it is a disjunction, as in the following case

$$A, A \rightarrow B \vee C \vdash^? B \vee C.$$

Similarly, even if the goal is an atom  $q$  in the reduction case (red) we cannot require that  $D$  in the formula  $A_1 \rightarrow \dots \rightarrow A_n \rightarrow D$  is atomic and  $D = q$ . If there are disjunctions in the database, at every step we can choose to work on the goal or to split the database by (c-dis) rule. Moreover, if there are  $n$  disjunctions a systematic application of (c-dis) rule yields  $2^n$  branches. All of this means that if we handle disjunction in the most obvious way we lose the goal-directedness of deduction and the computation becomes very inefficient.

The reason is that if *positive* disjunctions are allowed in a database  $\Gamma$  it is not true that

$$(dp) \Gamma \vdash A \vee B \text{ implies } \Gamma \vdash A \text{ or } \Gamma \vdash B.$$

This property, called *disjunction property*, holds when  $\Gamma$  does not contain positive disjunctions, but fails when it does contain them, as the example above shows. This means that the goal  $A \vee B$  cannot be decomposed/reduced to  $A$  and to  $B$ . In other words, we cannot proceed in a goal-directed fashion. There are three ways to overcome the problem of disjunction. The simplest solution is to kill the problem at the root: do not allow positive occurrences of disjunction in the database. To prevent positive disjunctions means to restrict our consideration to so-called *Harrop formulas*. To introduce them, let us define the two types of formulas D and G by mutual induction as follows:

$$\begin{aligned} D &:= q \mid \perp \mid G \rightarrow D \mid D \wedge D \\ G &:= q \mid \perp \mid G \wedge G \mid G \vee G \mid D \rightarrow G. \end{aligned}$$

A formula is *Harrop* if it is defined according to the D-clauses above. D-formulas are allowed as constituents of the database, whereas G formulas are allowed to be asked as goals. It is easy to extend the goal directed procedure to Harrop formulas. A database will be a set of D-formulas (which are not conjunctions themselves). We just add the rule R1 given above to handle disjunctive goals. This gives us a complete system, which can be optimized by adopting the diminishing resource approach and bounded restart.

Another solution, that we just mention, is to eliminate disjunction by adopting Statman's translation [Statman, 1979]: we can translate every pair database-goal  $(\Gamma, G)$  in a pair  $(\Gamma^*, G^*)$ , such that  $\Gamma^*, G^*$  do not contain disjunction, but contain additional atoms, and it holds (in intuitionistic logic)

$$\Gamma \vdash G \text{ iff } \Gamma^* \vdash G^*.$$

We can then use the proof procedure without disjunction.

However, one can try to cope with the whole propositional intuitionistic logic without limitations, by using additional machinery. There are two difficulties to define a goal-directed procedure. Consider the query  $\Gamma \vdash^?$

$A \vee B$ . We can adopt the rule R1 by continuing for instance with  $\Gamma \vdash^? A$ , and remember that at a previous step we could have chosen the disjunct  $B$ , thus we must be able to go back to  $\Gamma \vdash^? B$ . The use of history and restart can accomplish the necessary book-keeping mechanism. However, we must take into account that the database may have changed in the meantime, and we must go back to the right database; notice that in general

$$(A \rightarrow B) \vee C \not\equiv A \rightarrow B \vee C \not\equiv B \vee (A \rightarrow C)$$

(although these equivalences hold in classical logic). One way to keep track of the dependency between the goal and database from which it is asked is to use labels. This solution is developed in [Gabbay and Olivetti, 2000], where a labelled goal-directed proof procedure for full intuitionistic logic is given. The labels are partially ordered and can be interpreted as possible worlds.

The other technical trick is to extend suitably the notion of 'Head' of a formula to formulas with positive disjunctions; this is necessary to define the reduction step. For instance, ignoring the labelling and restart mechanism, the query  $\Gamma, K$ , where  $K = A \rightarrow (B \vee (C \rightarrow q)) \vdash^? q$ , and  $q$  is an atom, would be reduced to the queries:

$$\begin{aligned} \Gamma, K \vdash^? A, \\ \Gamma, K, B \vdash^? q, \\ \Gamma, K, C \rightarrow q \vdash^? C. \end{aligned}$$

and  $q$  would be recorded in the history.

We shall not give the details of the procedure which can be found in [Gabbay and Olivetti, 2000]. A similar, although much simpler, procedure can be given for classical logic. However, in classical logic the treatment of disjunctive data is not problematic, since on the one hand we can define disjunction using the other connectives (the  $\rightarrow$  connective alone suffices as  $A \vee B \equiv (A \rightarrow B) \rightarrow B$ ). On the other hand every formula can be rewritten as a set of clauses of the form:

$$p_1 \wedge \dots \wedge p_n \rightarrow q_1 \vee \dots \vee q_m$$

where  $n \geq 0$ , and  $m > 0$ , every  $p_i$  is an atom, and every  $q_j$  is an atom or is  $\perp$ . For data of this sort, a goal-directed procedure is easily designed, see [Gabbay and Olivetti, 2000; Nadathur, 1998; Loveland, 1991; Loveland, 1992].

### 3.8 Some history

A goal-directed proof system for a (first-order) fragment of intuitionistic and classical logic was first given by Gabbay [Gabbay and Reyle, 1984;

Gabbay, 1985] as a foundation of hypothetical logic programming. A similar system for intuitionistic logic was proposed in [McCarty, 1988a; McCarty, 1988b]. The restart rule was first proposed by Gabbay in his lecture notes in 1984 and the first theoretical results published in [Gabbay, 1985] and then 1985 in [Gabbay and Kriwaczek, 1991].<sup>7</sup> A similar idea to restart has been exploited by Loveland [1991; 1992], in order to extend conventional logic programming to non-Horn databases; in Loveland's proof procedure the restart rule is a way of implementing reasoning by case-analysis.

The concept of goal-directed computation can also be seen as a generalization of the notion of *uniform proof* as introduced in [Miller *et al.*, 1991]. A uniform proof system is called 'abstract logic programming' [Miller *et al.*, 1991]. The extension of the uniform proof paradigm to classical logic is recently discussed in [Harland, 1997] and [Nadathur, 1998]. The essence of a uniform-proof system is the same underlying the goal-directed paradigm: the proof-search is driven by the goal and the connectives can be interpreted directly as search instructions.

The locally linear computation with bounded restart was first presented by Gabbay [1991] and then further investigated in [Gabbay, 1992], where goal-directed procedures for classical and some intermediate logics are also presented. This refinement is strongly connected to the contraction-free sequent calculi for intuitionistic logic which have been proposed by many people: Dyckhoff [1992], Hudelmaier [1990], and Lincoln *et al.*, [1991]. To see the intuitive connection, let us consider the query:

$$(1) \Delta, (A \rightarrow p) \rightarrow q \vdash^? q, \emptyset$$

we can step by reduction to  $\Delta \vdash^? A \rightarrow p, (q)$  and then to  $\Delta, A \vdash^? p, (q)$ , which, by the soundness corresponds to

$$(2) \Delta, A, p \rightarrow q \vdash^? q.$$

In all the mentioned calculi (1) can be reduced to (2) by a sharpened left-implication rule (here used backwards). This modified rule is the essential ingredient to obtain a contraction-free sequent calculus for **I**, at least for its implicational fragment. A formal connection with these contraction-free calculi has not been studied yet. It might turn out that LL-computations correspond to *uniform proofs* (in the sense of [Miller *et al.*, 1991]) within these calculi.

In [Gabbay and Olivetti, 2000] the goal-directed methods are extended to some *intermediate logics*, i.e. logics which are between intuitionistic and classical logics. In particular, it is given a proof procedure for the family of intermediate logics of Kripke models with finite height, and for Dummett-Gödel logic LC. These proof systems are obtained by adding suitable restart rules to the intuitionistic system.

<sup>7</sup>The lecture notes have evolved also into the book [Gabbay, 1998].



## 4 MODAL LOGICS OF STRICT IMPLICATION

In this section we examine how to extend the goal-directed proof methods to modal logics. We begin considering a minimal language which contains only *strict implication*, we will then extend it to the full language of modal logics. Strict implication, denoted by  $A \Rightarrow B$  is read as ‘necessarily  $A$  implies  $B$ ’. The notion of necessity (and the dual notion of possibility) are the subject of modal logics. Strict implication can be regarded as a derived notion:  $A \Rightarrow B = \Box(A \rightarrow B)$ , where  $\rightarrow$  denotes material implication and  $\Box$  denotes modal necessity. However, strict implication can also be considered as a primitive notion, and has already been considered as such at the beginning of the century in many discussions about the paradoxes of material implication [Lewis, 1912; Lewis and Langford, 1932].

The extension of the goal-directed approach to strict implication and modal logics relies upon the *possible worlds* semantics of modal logics which is mainly due to Kripke.

The strict implication language  $\mathcal{L}(\Rightarrow)$  contains all formulas built out from a denumerable set  $Var$  of propositional variables by applying the strict implication connective, that is, if  $p \in Var$  then  $p$  is a formula of  $\mathcal{L}(\Rightarrow)$ , and if  $A$  and  $B$  are formulas of  $\mathcal{L}(\Rightarrow)$ , then so is  $A \Rightarrow B$ . Let us fix an atom  $p_0$ , we can define the constant  $\top \equiv p_0 \Rightarrow p_0$  and let  $\Box A \equiv \top \Rightarrow A$ .

**Semantics**

We review the standard Kripke semantics for  $\mathcal{L}(\Rightarrow)$ .

A Kripke structure  $M$  for  $\mathcal{L}(\Rightarrow)$  is a triple  $(W, R, V)$ , where  $W$  is a non-empty set (whose elements are called possible worlds),  $R$  is a binary relation on  $W$ , and  $V$  is a mapping from  $W$  to sets of propositional variables of  $\mathcal{L}$ . Truth conditions for formulas (of  $\mathcal{L}(\Rightarrow)$ ) are defined as follows:

- $M, w \models p$  iff  $p \in V(w)$ ;
- $M, w \models A \Rightarrow B$  iff for all  $w'$  such that  $wRw'$  and  $M, w' \models A$ , it holds  $M, w' \models B$ .

We say that a formula  $A$  is *valid* in a structure  $M$ , denoted by  $M \models A$ , if  $\forall w \in W, M, w \models A$ . We say that a formula  $A$  is *valid* with respect to a given class of structures  $\mathcal{K}$ , iff it is valid in every structure  $M \in \mathcal{K}$ . We sometimes use the notation  $\models_{\mathcal{K}} A$ . Let us fix a class of structures  $\mathcal{K}$ . Given two formulas  $A$  and  $B$ , we can define the *consequence relation*  $A \models_{\mathcal{K}} B$  as

$$\forall M = (W, R, V) \in \mathcal{K} \forall w \in W \text{ if } M, w \models_{\mathcal{K}} A \text{ then } M, w \models_{\mathcal{K}} B.$$

Different modal logics are obtained by considering classes of structures whose relation  $R$  satisfies some specific properties. The properties of the accessibility relations we consider are listed in Table 1.

Table 1. Standard properties of the accessibility relation.

Reflexivity	$\forall x xRx$
Transitivity	$\forall x\forall y\forall z xRy \wedge yRz \rightarrow xRz$
Symmetry	$\forall x\forall y xRy \rightarrow yRx$
Euclidean	$\forall x\forall y\forall z xRy \wedge xRz \rightarrow yRz$

Table 2. Some standard modal logics.

Name	Reflexivity	Transitivity	Symmetry	Euclidean
<b>K</b>				
<b>KT</b>	*			
<b>K4</b>		*		
<b>S4</b>	*	*		
<b>K5</b>				*
<b>K45</b>		*		*
<b>KB</b>			*	
<b>KTB</b>	*		*	
<b>S5</b>	*	*	*	*

We will take into consideration strict implication  $\Rightarrow$  as defined in systems **K**, **KT**,<sup>8</sup> **K4**, **S4**, **K5**, **K45**, **KB**, **KBT**, and **S5**.<sup>9</sup>

Properties of accessibility relation  $R$  in Kripke frames, corresponding to these systems are shown in Table 2.

Letting **S** be one of the modal systems above, we use the notation  $\models_{\mathbf{S}} A$  (and  $A \models_{\mathbf{S}} B$ ) to denote validity in (and the consequence relation determined by) the class of structures corresponding to **S**.

#### 4.1 Proof systems

In this section we present proof methods for all modal systems mentioned above. We regard a database as a set of labelled formulas  $x_i : A_i$  equipped by a relation  $\alpha$  giving connections between labels. The labels obviously represent worlds. Thus,  $x_i : A_i$  means that  $A_i$  holds at  $x_i$ . The goal of a query is asked with respect to a world. The form of databases and goals determine the notion of consequence relation

$$\frac{\{x_1 : A_1, \dots, x_n : A_n\}, \alpha \vdash x : A}{\text{consequence relation}}$$

<sup>8</sup>We use the acronym **KT** rather than the more common **T**, as the latter is also the name of a subrelevance logic we will meet in Section 5.

<sup>9</sup>We do not consider here systems containing  $D : \Box A \rightarrow \Diamond A$ , which correspond to the *seriality* of the accessibility relation, i.e.  $\forall x\exists y xRy$  in Kripke frames. The reason is that seriality cannot be expressed in the language of strict implication alone; moreover, it cannot be expressed in any modal language, unless  $\neg$  or  $\Diamond$  is allowed.

whose intended meaning is that if  $A_i$  holds at  $x_i$  (for  $i = 1, \dots, n$ ) and the  $x_i$  are connected as  $\alpha$  prescribes, then  $A$  must hold at  $x$ .

For different logics  $\alpha$  will be required to satisfy different properties such as reflexivity, transitivity, etc., depending on the properties of the accessibility relation of the system under consideration.

**DEFINITION 37.** Let us fix a denumerable alphabet  $\mathcal{A} = \{x_1, \dots, x_i, \dots\}$  of labels. A *database* is a finite graph of formulas labelled by  $\mathcal{A}$ . We denote a database as a pair  $(\Delta, \alpha)$ , where  $\Delta$  is a finite set of labelled formulas  $\Delta = \{x_1 : A_1, \dots, x_n : A_n\}$  and  $\alpha = \{(x_1, x'_1), \dots, (x_m, x'_m)\}$  is a set of links. Let  $Lab(E)$  denote the set of labels  $x \in \mathcal{A}$  occurring in  $E$ , and assume that (i)  $Lab(\Delta) = Lab(\alpha)$ , and (ii) if  $x : A \in \Delta, x : B \in \Delta$ , then  $A = B$ .<sup>10</sup>

A *trivial database* has the form  $(\{x_0 : A\}, \emptyset)$ .

The *expansion* of a database  $(\Gamma, \alpha)$  by  $y : C$  at  $x$ , with  $x \in Lab(\Gamma)$ ,  $y \notin Lab(\Gamma)$  is defined as follows:

$$(\Gamma, \alpha) \oplus_x (y : C) = (\Delta \cup \{y : C\}, \alpha \cup \{(x, y)\}).$$

**DEFINITION 38.** A query  $Q$  is an expression of the form:

$$Q = (\Delta, \alpha) \vdash^? x : G, H$$

where  $(\Delta, \alpha)$  is a database,  $x \in Lab(\Delta)$ ,  $G$  is a formula, and  $H$ , *the history*, is a set of pairs

$$H = \{(x_1, q_1), \dots, (x_m, q_m)\}$$

where  $x_i$  are labels and  $q_i$  are atoms. We will often omit the parentheses around the two components of a database and write  $Q = \Delta, \alpha \vdash^? x : G, H$ . A query from a trivial database  $\{x_0 : A\}$  will be written simply as:

$$x_0 : A \vdash^? x_0 : B, H,$$

and if  $A = \top$ , we sometimes just write  $\vdash^? x_0 : B, H$ .

**DEFINITION 39.** Let  $\alpha$  be a set of links, we introduce a family of relation symbols  $A_\alpha^S(x, y)$ , where  $x, y \in Lab(\alpha)$ . We consider the following conditions:

- (K)  $(x, y) \in \alpha \Rightarrow A_\alpha^S(x, y)$ ,
- (T)  $x = y \Rightarrow A_\alpha^S(x, y)$ ,
- (4)  $\exists z (A_\alpha^S(x, z) \wedge A_\alpha^S(z, y)) \Rightarrow A_\alpha^S(x, y)$ ,
- (5)  $\exists z (A_\alpha^S(z, x) \wedge A_\alpha^S(z, y)) \Rightarrow A_\alpha^S(x, y)$ ,
- (B)  $A_\alpha^S(x, y) \Rightarrow A_\alpha^S(y, x)$ .

<sup>10</sup>We will drop this condition in Section 4.3 when we extend the language by allowing conjunction.

For  $\mathbf{K} \in \mathbf{S} \subseteq \{\mathbf{K}, \mathbf{T}, \mathbf{4}, \mathbf{5}, \mathbf{B}\}$ , we let  $A^{\mathbf{S}}$  be the least relation satisfying all conditions in  $\mathbf{S}$ . Thus, for instance,  $A^{\mathbf{K45}}$  is the least relation such that:

$$\begin{aligned} A_{\alpha}^{\mathbf{K45}}(x, y) \quad \Leftrightarrow \quad & (x, y) \in \alpha \vee \\ & \vee \exists z (A_{\alpha}^{\mathbf{K45}}(x, z) \wedge A_{\alpha}^{\mathbf{K45}}(z, y)) \vee \\ & \vee \exists z (A_{\alpha}^{\mathbf{K45}}(z, x) \wedge A_{\alpha}^{\mathbf{K45}}(z, y)). \end{aligned}$$

We will use the standard abbreviations (i.e.  $A^{\mathbf{S5}} = A^{\mathbf{KT5}} = A^{\mathbf{KT45}}$ ).

**DEFINITION 40** (Modal Deduction Rules). For each modal system  $\mathbf{S}$ , the corresponding proof system, denoted by  $\mathbf{P}(\mathbf{S})$ , comprises the following rules parametrized to predicates  $A^{\mathbf{S}}$ :

- (success)  $\Delta, \alpha \vdash^? x : q, H$  immediately succeeds if  $q$  is an atom and  $x : q \in \Delta$ .
- (implication) From  $\Delta, \alpha \vdash^? x : A \Rightarrow B, H$ , step to

$$(\Delta, \alpha) \oplus_x (y : A) \vdash^? y : B, H,$$

where  $y \notin \text{Lab}(\Delta) \cup \text{Lab}(H)$ .

- (reduction) If  $y : C \in \Delta$ , with  $C = B_1 \Rightarrow B_2 \Rightarrow \dots \Rightarrow B_k \Rightarrow q$ , with  $q$  atomic, then from

$$\Delta, \alpha \vdash^? x : q, H$$

step to

$$\Delta, \alpha \vdash^? u_1 : B_1, H \cup \{(x, q)\},$$

$\vdots$ ,

$$\Delta, \alpha \vdash^? u_k : B_k, H \cup \{(x, q)\},$$

for some  $u_0, \dots, u_k \in \text{Lab}(\alpha)$ , with  $u_0 = y$ ,  $u_k = x$ , such that

for  $i = 0, \dots, k - 1$ ,  $A_{\alpha}^{\mathbf{S}}(u_i, u_{i+1})$  holds.

- (restart) If  $(y, r) \in H$ , then, from  $\Delta, \alpha \vdash^? x : q, H$ , with  $q$  atomic, step to

$$\Delta, \alpha \vdash^? y : r, H \cup \{(x, q)\}.$$

### Restricted restart

Similar to the case of classical logic, in any deduction of a query  $Q$  of the form  $\Delta, \alpha \vdash^? x : G, \emptyset$ , the *restart rule* can be restricted to the choice of the pair  $(y, r)$ , such that  $r$  is the uppermost atomic goal occurred in the deduction and  $y$  is the label associated to  $r$  (that is, the query in which  $r$  appears contains  $\dots \vdash^? y : r$ ). Hence, if the initial query is  $Q = \Delta, \alpha \vdash^?$

$x : G, \emptyset$  and  $G$  is an atom  $q$ , such a pair is  $(x, q)$ , if  $G$  has the form  $B_1 \Rightarrow \dots \Rightarrow B_k \Rightarrow r$ , then the first pair is obtained by repeatedly applying the implication rule until we reach the query  $\dots \vdash^? x_k : r$ , with  $x_k \notin Lab(\Delta)$ . With this restriction, we do not need to keep track of the history any more, but only of the first pair. An equivalent formulation is to allow restart from the initial goal (and its relative label) even if it is implicational, but the re-evaluation of an implication causes a redundant increase of the database, that is why we prefer the above formulation.

**PROPOSITION 41.** *If  $\Delta, \alpha \vdash^? x : G, \emptyset$  succeeds then it succeeds by a derivation in which every application of restart is restricted restart.*

We have omitted the reference to the specific proof system  $P(\mathbf{S})$ , since of the previous claim does not depend on the specific properties of the predicates  $A^{\mathbf{S}}$  involved in the definition of a proof system  $P(\mathbf{S})$ . We will omit the reference to  $P(\mathbf{S})$  whenever it is not necessary.

We show some examples of the proof procedure.

**EXAMPLE 42.** In Figure 4 we show a derivation of

$$((p \Rightarrow p) \Rightarrow a \Rightarrow b) \Rightarrow (b \Rightarrow c) \Rightarrow a \Rightarrow c.$$

in  $P(\mathbf{K})$ . By Proposition 41, we only record the first pair for restart, which, however, is not used in the derivation. As usual in each node we only show the additional data, if any. Thus the database in each node is given by the collection of the formulas from the root to that node. Here is an explanation of the steps: in step (2)  $\alpha = \{(x_0, x_1)\}$ ; in step (3)  $\alpha = \{(x_0, x_1), (x_1, x_2)\}$ ; in step (4)  $\alpha = \{(x_0, x_1), (x_1, x_2), (x_2, x_3)\}$ ; since  $A_{\alpha}^{\mathbf{K}}(x_2, x_3)$ , by reduction w.r.t.  $x_2 : b \Rightarrow c$  we get (5); since  $A_{\alpha}^{\mathbf{K}}(x_1, x_2)$  and  $A_{\alpha}^{\mathbf{K}}(x_2, x_3)$ , by reduction w.r.t.  $x_1 : (p \Rightarrow p) \Rightarrow a \Rightarrow b$  we get (6) and (8). the latter immediately succeeds as  $x_3 : a \in \Delta$ ; from (6) we step to (7) which immediately succeeds.

**EXAMPLE 43.** In Figure 5 we show a derivation of

$$(((a \Rightarrow a) \Rightarrow p) \Rightarrow q) \Rightarrow p) \Rightarrow p$$

in  $P(\mathbf{KBT})$ , we use restricted restart according to Proposition 41. In step (2),  $\alpha = \{(x_0, x_1)\}$ . Step (3) is obtained by reduction w.r.t.  $x_1 : (((a \Rightarrow a) \Rightarrow p) \Rightarrow q) \Rightarrow p$ , as  $A_{\alpha}^{\mathbf{KBT}}(x_1, x_1)$ . In step (4)  $\alpha = \{(x_0, x_1), (x_1, x_2)\}$ ; step (5) is obtained by restart; step (6) by reduction w.r.t.  $x_2 : (a \Rightarrow a) \Rightarrow p$ , as  $A_{\alpha}^{\mathbf{KBT}}(x_2, x_1)$ ; in step (7)  $\alpha = \{(x_0, x_1), (x_1, x_2), (x_1, x_3)\}$  and the query immediately succeeds.

In order to prove soundness and completeness, we need to give a formal meaning to queries, i.e. to define when a query is valid. We first introduce a notion of *realization* of a database in a model to give a semantic meaning to databases.

**DEFINITION 44 (Realization and validity).** Let  $A^{\mathbf{S}}$  be an accessibility predicate, given a database  $(\Gamma, \alpha)$  and a Kripke model  $M = (W, R, V)$ ,

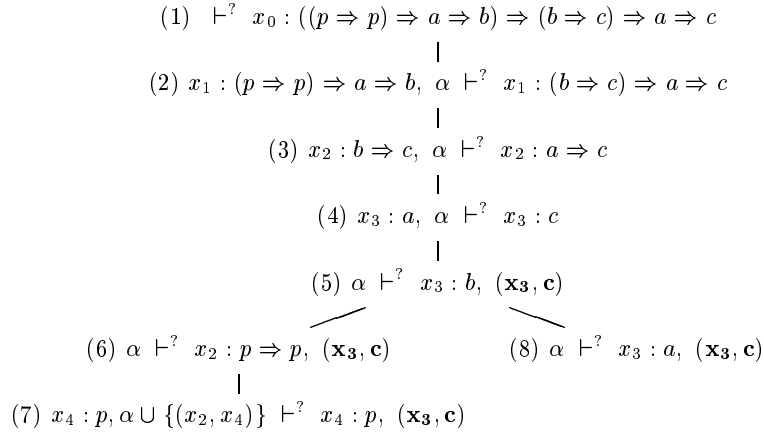


Figure 4. Derivation for Example 42.

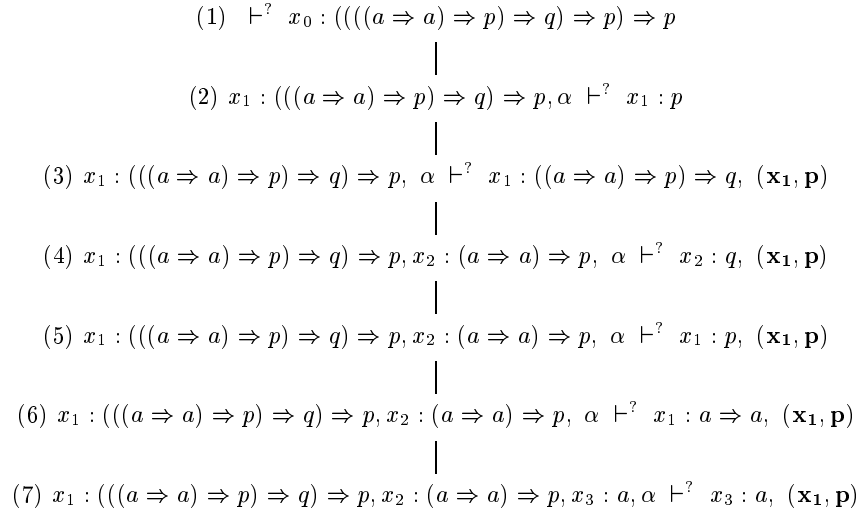


Figure 5. Derivation for Example 43.

a mapping  $f : Lab(\Gamma) \rightarrow W$  is called a *realization* of  $(\Gamma, \alpha)$  in  $M$  with respect to  $A^{\mathbf{S}}$ , if the following hold:

1.  $A_{\alpha}^{\mathbf{S}}(x, y)$  implies  $f(x)Rf(y)$ ;
2. if  $x : A \in \Gamma$ , then  $M, f(x) \models A$ .

We say that a query  $Q = \Gamma, \alpha \vdash^? x : G, H$  is *valid* in  $\mathbf{S}$  if for every  $\mathbf{S}$ -model  $M$  and every realization  $f$  of  $(\Gamma, \alpha)$ , we have either  $M, f(x) \models G$ , or for some  $(y, r) \in H$ ,  $M, f(y) \models r$ .

The soundness of the proof procedure can be proved easily by induction on the length of the computation.

**THEOREM 45 (Soundness).** *Let  $Q = \Gamma, \alpha \vdash^? x : G, H$  succeed in the proof system  $P(\mathbf{S})$ , then it is valid in  $\mathbf{S}$ .*

**COROLLARY 46.** *If  $x_0 : A \vdash^? x_0 : B, \emptyset$  succeeds in  $P(\mathbf{S})$ , then  $A \models_{\mathbf{S}} B$  holds. In particular, if  $\vdash^? x_0 : A, \emptyset$  succeeds in  $P(\mathbf{S})$ , then  $A$  is valid in  $\mathbf{S}$ .*

To prove completeness we proceed in a similar way to what we did for intuitionistic logic. First we show that cut is admissible. Then we prove completeness by a sort of canonical model construction. The cut rule states the following: let  $x : A \in \Gamma$ , then if (1)  $\Gamma \vdash y : B$  and (2)  $\Delta \vdash z : A$  succeed, we can ‘replace’  $x : A$  by  $\Delta$  in  $\Gamma$  and get a successful query from (1). Since there are labels and accessibility predicates, we must be careful. There are two points to clarify. First, we need to define the involved notion of substitution. Furthermore the proof systems  $P(\mathbf{S})$  depend uniformly on predicate  $A^{\mathbf{S}}$ , and we expect that the admissibility of cut depends on the properties of predicate  $A^{\mathbf{S}}$ . It turns out that the admissibility of cut (stated in Theorem 48) holds for every proof system  $P(\mathbf{S})$ , such that  $A^{\mathbf{S}}$  satisfies the following conditions:

- (i)  $A^{\mathbf{S}}$  is closed under substitution of labels;
- (ii)  $A_{\alpha}^{\mathbf{S}}(x, y)$  implies  $A_{\alpha \cup \beta}^{\mathbf{S}}(x, y)$ ;
- (iii)  $A_{\alpha}^{\mathbf{S}}(u, v)$  implies  $\forall x y (A_{\alpha \cup \{(u,v)\}}^{\mathbf{S}}(x, y) \leftrightarrow A_{\alpha}^{\mathbf{S}}(x, y))$ .

These conditions ensure the following properties.

**PROPOSITION 47.**

- (a) *If  $\Gamma, \alpha \vdash^? x : C, H$  succeeds then also  $\Gamma[u/v], \alpha[u/v] \vdash^? x[u/v] : C, H[u/v]$  succeeds.*
- (b) *If  $A_{\alpha}^{\mathbf{S}}(x, y)$  and  $\Gamma, \alpha \cup \{(x, y)\} \vdash^? u : G, H$  succeed, then also  $\Gamma, \alpha \vdash^? u : G, H$  succeeds.*

We say that two databases  $(\Gamma, \alpha)$ ,  $(\Delta, \beta)$  are *compatible for substitution*,<sup>11</sup> if

for every  $x \in \text{Lab}(\Gamma) \cap \text{Lab}(\Delta)$ , for all formulas  $C$ ,  $x : C \in \Gamma \Leftrightarrow x : C \in \Delta$ .

If  $(\Gamma, \alpha)$  and  $(\Delta, \beta)$  are compatible for substitution,  $x : A \in \Gamma$ , and  $y \in \text{Lab}(\Delta)$ , we denote by

$$(\Gamma, \alpha)[x : A/\Delta, \beta, y] = (\Gamma - \{x : A\} \cup \Delta, \alpha[x/y] \cup \beta).$$

the database which results by replacing  $x : A$  in  $(\Gamma, \alpha)$  by  $(\Delta, \beta)$  at point  $y$ .

At this point we can state precisely the result about cut.

**THEOREM 48** (Admissibility of cut). *Let predicate  $A^S$  satisfy the conditions (i), (ii), (iii) above. If the following queries succeed in the proof system  $P(\mathbf{S})$  :*

1.  $\Gamma[x : A] \vdash^? u : B, H_1$
2.  $\Delta, \beta \vdash^? y : A, H_2$ .

and  $(\Gamma, \alpha)$  and  $(\Delta, \beta)$  are compatible for substitution, then also

3.  $(\Gamma, \alpha)[x : A/\Delta, \beta, y] \vdash^? u[x/y] : B, H_1[x/y] \cup H_2$  succeeds in  $P(\mathbf{S})$  .

The proof proceeds similarly to the one of Theorem 10 and is given in [Gabbay and Olivetti, 2000]. From the theorem we immediately have the following two corollaries.

**COROLLARY 49.** *Under the same conditions as above, if  $x : A \vdash^? x : B$  succeeds and  $x : B \vdash^? x : C$  succeeds then also  $x : A \vdash^? x : C$  succeeds.*

**COROLLARY 50.** *If  $\mathbf{K} \in \mathbf{S} \subseteq \{\mathbf{K}, \mathbf{4}, \mathbf{5}, \mathbf{B}, \mathbf{T}\}$ , then in the proof system  $P(\mathbf{S})$  cut is admissible.*

As we have said, we can prove the completeness by a sort of canonical model construction, which is less constructive of the one of Theorem 9. The following properties will be used in the completeness proof.

**PROPOSITION 51.**

- (Identity) *If  $x : A \in \Gamma$ , then  $\Gamma, \alpha \vdash^? x : A, H$  succeeds.*
- (Monotony) *If  $Q = \Gamma, \alpha \vdash^? x : C, H$  succeeds and  $\Gamma \subseteq \Delta$ ,  $\alpha \subseteq \beta$ ,  $H \subseteq H'$ , then also  $\Delta, \beta \vdash^? x : C, H'$  succeeds.*

---

<sup>11</sup>This condition is not necessary if we allow the occurrence of several formulas with the same label in a database, as we will do in Section 4.3 when we add conjunction.



**THEOREM 52 (Completeness).** *Given a query  $Q = \Gamma, \alpha \vdash^? x : A, H$ , if  $Q$  is  $\mathbf{S}$ -valid then  $Q$  succeeds in the proof system  $P(\mathbf{S})$ .*

**Proof.** By contraposition, we prove that if  $Q = \Gamma, \alpha \vdash^? x : A, H$  does not succeed in one proof system  $P(\mathbf{S})$ , then there is an  $\mathbf{S}$ -model  $M$  and a realization  $f$  of  $(\Gamma, \alpha)$ , such that  $M, f(x) \not\models A$  and for any  $(y, r) \in H$ ,  $M, f(y) \not\models r$ .

We construct an  $\mathbf{S}$ -model by extending the database, through the evaluation of all possible formulas at every world (each represented by one label) of the database. Since such evaluation may lead, for implication formulas, to create new worlds, we must carry on the evaluation process on these new worlds. Therefore, in the construction we consider an enumeration of pairs  $(x_i, A_i)$ , where  $x_i$  is a label and  $A_i$  is a formula.

Assume  $\Gamma, \alpha \vdash^? x : A, H$  fails in  $P(\mathbf{S})$ . We let  $\mathcal{A}$  be a denumerable alphabet of labels and  $\mathcal{L}$  be the underlying propositional language. Let  $(x_i, A_i)$ , for  $i \in \omega$  be an enumeration of pairs of  $\mathcal{A} \times \mathcal{L}$ , starting with the pair  $(x, A)$  and containing infinitely many repetitions, that is

$$(x_0, A_0) = (x, A), \\ \forall y \in \mathcal{A}, \forall F \in \mathcal{L}, \forall n \exists m > n (y, F) = (x_m, A_m).$$

Given such enumeration we define (i) a sequence of databases  $(\Gamma_n, \alpha_n)$ , (ii) a sequence of histories  $H_n$ , (iii) a new enumeration of pairs  $(y_n, B_n)$ , as follows:

- (step 0) Let  $(\Gamma_0, \alpha_0) = (\Gamma, \alpha)$ ,  $H_0 = H$ ,  $(y_0, B_0) = (x, A)$ .
- (step n+1) Given  $(y_n, B_n)$ , if  $y_n \in \text{Lab}(\Gamma_n)$  and  $\Gamma_n, \alpha_n \vdash^? y_n : B_n, H_n$  fails then proceed according to (a) else to (b).

(a) if  $B_n$  is atomic, then we set

$$H_{n+1} = H_n \cup \{(y_n, B_n)\}, \\ (\Gamma_{n+1}, \alpha_{n+1}) = (\Gamma_n, \alpha_n), \\ (y_{n+1}, B_{n+1}) = (x_{k+1}, A_{k+1}), \\ \text{where } k = \max_{t \leq n} \exists s \leq n (y_s, B_s) = (x_t, A_t),$$

else let  $B_n = C \Rightarrow D$ , then we set

$$H_{n+1} = H_n, \\ (\Gamma_{n+1}, \alpha_{n+1}) = (\Gamma_n, \alpha_n) \oplus_{y_n} (x_m : C), \\ (y_{n+1}, B_{n+1}) = (x_m, D), \\ \text{where } x_m = \min\{x_t \in \mathcal{A} \mid x_t \notin \text{Lab}(\Gamma_n) \cup \text{Lab}(H_n)\}.$$

(b) We set

$$\begin{aligned} H_{n+1} &= H_n, \\ (\Gamma_{n+1}, \alpha_{n+1}) &= (\Gamma_n, \alpha_n), \\ (y_{n+1}, B_{n+1}) &= (x_{k+1}, A_{k+1}), \\ \text{where } k &= \max\{t \leq n \mid \exists s \leq n (y_s, B_s) = (x_t, A_t)\}, \quad \blacksquare \end{aligned}$$

The proof of completeness is made of several lemmas.

LEMMA 53.  $\forall k \exists n \geq k (x_k, A_k) = (y_n, B_n)$ .

**Proof.** By induction on  $k$ . If  $k = 0$ , the claim holds by definition. Let  $(x_k, A_k) = (y_n, B_n)$ .

- (i) if  $y_n \notin \text{Lab}(\Gamma_n)$ , or  $\Gamma_n, \alpha_n \vdash^? y_n : B_n, H_n$  succeeds, or  $B_n$  is atomic, then  $(x_{k+1}, A_{k+1}) = (y_{n+1}, B_{n+1})$ .
- (ii) Otherwise, let  $B_n = C_1 \Rightarrow \dots \Rightarrow C_t \Rightarrow r$ , ( $t > 0$ ), then  $(x_{k+1}, A_{k+1}) = (y_{n+t+1}, B_{n+t+1})$ .  $\blacksquare$

LEMMA 54. For all  $n \geq 0$ , if  $\Gamma_n, \alpha_n \vdash^? y_n : B_n, H_n$  fails, then:

$$\forall m \geq n \Gamma_m, \alpha_m \vdash^? y_n : B_n, H_m \text{ fails.}$$

**Proof.** By induction on  $cp(B_n) = c$ . if  $c = 0$ , that is  $B_n$  is an atom, say  $q$ , then we proceed by induction on  $m \geq n + 1$ .

- ( $m = n + 1$ ) we have  $\Gamma_n, \alpha_n \vdash^? y_n : q, H_n$  fails, then also  $\Gamma_n, \alpha_n \vdash^? y_n : q, H_n \cup \{(y_n, q)\}$  fails, whence, by construction,

$$\Gamma_{n+1}, \alpha_{n+1} \vdash^? y_n : q, H_{n+1} \text{ fails.}$$

- ( $m > n + 1$ ) Suppose we have proved the claim up to  $m \geq n + 1$ , and suppose by way of contradiction that  $\Gamma_m, \alpha_m \vdash^? y_n : q, H_m$  fails, but

$$(i) \Gamma_{m+1}, \alpha_{m+1} \vdash^? y_n : q, H_{m+1} \text{ succeeds.}$$

At step  $m$ ,  $(y_m, B_m)$  is considered; it must be  $y_m \in \text{Lab}(\Gamma_m)$  and

$$(ii) \Gamma_m, \alpha_m \vdash^? y_m : B_m, H_m \text{ fails.}$$

We have two cases, according to the form of  $B_m$ . If  $B_m$  is an atom  $r$ , as  $(y_n, q) \in H_m$ , from query (ii) by restart we can step to

$$\Gamma_m, \alpha_m \vdash^? y_n : q, H_m \cup \{(y_m, r)\},$$

that is the same as  $\Gamma_{m+1}, \alpha_{m+1} \vdash^? y_n : q, H_{m+1}$ , which succeeds and we get a contradiction. If  $B_m = C_1 \Rightarrow \dots \Rightarrow C_k \Rightarrow r$ , with  $k > 0$ , then from query (ii) we step in  $k$  steps to  $\Gamma_{m+k}, \alpha_{m+k} \vdash^? y_{m+k} : r, H_{m+k}$ , where  $(\Gamma_{m+k}, \alpha_{m+k}) = (\Gamma_m, \alpha_m) \oplus_{y_m} (y_{m+1} : C_1) \oplus_{y_{m+1}} \dots \oplus_{y_{m+k-1}} (y_{m+k} : C_k)$  and  $H_{m+k} = H_m$ ; then, by restart, since  $(y_n, q) \in H_{m+k}$ , we step to

$$(iii) \Gamma_{m+k}, \alpha_{m+k} \vdash^? y_n : q, H_{m+k} \cup \{(y_{m+k}, r)\}.$$

Since query (i) succeeds, by monotony we have that also query (iii) succeeds, whence query (ii) succeeds, contradicting the hypothesis.

Let  $cp(B_n) = c > 0$ , that is  $B_n = C \Rightarrow D$ . By hypothesis  $\Gamma_n, \alpha_n \vdash^? y_n : C \Rightarrow D, H_n$ , fails. Then by construction and by the computation rules  $\Gamma_{n+1}, \alpha_{n+1} \vdash^? y_{n+1} : D, H_{n+1}$ , fails, and hence, by the induction hypothesis,  $\forall m \geq n+1$ ,

$$\Gamma_m, \alpha_m \vdash^? y_{n+1} : D, H_m, \text{ fails.}$$

Suppose by way of contradiction that for some  $m \geq n+1$ ,  $\Gamma_m, \alpha_m \vdash^? y_n : C \Rightarrow D, H_m$ , succeeds. This implies that, for some  $z \notin \text{Lab}(\Gamma_m) \cup \text{Lab}(H_m)$ ,

$$(1) (\Gamma_m, \alpha_m) \oplus_{y_n} (z : C) \vdash^? z : D, H_m, \text{ succeeds.}$$

Since  $y_{n+1} : C \in \Gamma_{n+1} \subseteq \Gamma_m$ ,  $\alpha_{n+1} \subseteq \alpha_m$ ,  $H_{n+1} \subseteq H_m$ , by monotony, we get

$$(2) \Gamma_m, \alpha_m \vdash^? y_{n+1} : C, H_m, \text{ succeeds.}$$

The databases involved in queries (1) and (2) are clearly compatible for substitution, hence by cut we obtain that  $\Gamma_m, \alpha_m \vdash^? y_{n+1} : D, H_m$  succeeds, and we have a contradiction. ■

LEMMA 55.

- (i)  $\forall m, \Gamma_m, \alpha_m \vdash^? x : A, H_m$  fails, whence
- (ii)  $\forall m$ , if  $(y, r) \in H_m$ , then  $\Gamma_m, \alpha_m \vdash^? y : r, H_m$  fails.

**Proof.** Left to the reader. ■

LEMMA 56. If  $B_n = C \Rightarrow D$  and  $\Gamma_n, \alpha_n \vdash^? y_n : C \Rightarrow D, H_n$  fails, then there is a  $y \in \mathcal{A}$ , such that for  $k \leq n$ ,  $y \notin \text{Lab}(\Gamma_k)$  and  $\forall m > n$ : (i)  $(y_n, y) \in \alpha_m$ , (ii)  $\Gamma_m, \alpha_m \vdash^? y : C, H_m$  succeeds, (iii)  $\Gamma_m, \alpha_m \vdash^? y : D, H_m$  fails.

**Proof.** By construction, we can take  $y = y_{n+1}$ , the new point created at step  $n+1$ , so that (i), (ii), (iii) hold for  $m = n+1$ . In particular

(\*)  $\Gamma_{n+1}, \alpha_{n+1} \vdash^? y_{n+1} : D, H_{n+1}$  fails.

Since the  $(\Gamma, \alpha_m)$  are not decreasing (w.r.t. inclusion), we immediately have that (i) and (ii) also hold for every  $m > n + 1$ . By construction, we know that  $B_{n+1} = D$ , whence by (\*) and Lemma 54, (iii) also holds for every  $m > n + 1$ . ■

### Construction of the Canonical model

We define an **S**-model as follows  $M = (W, R, V)$ , such that

- $W = \bigcup_n \text{Lab}(\Gamma_n)$ ;
- $xRy \equiv \exists n A_{\alpha_n}^{\mathbf{S}}(x, y)$ ,
- $V(x) = \{q \mid \exists n x \in \text{Lab}(\Gamma_n) \wedge \Gamma_n, \alpha_n \vdash^? x : q, H_n \text{ succeeds}\}$ .

LEMMA 57. *The relation  $R$  as defined above has the same properties of  $A^{\mathbf{S}}$ , e.g. if  $\mathbf{S}=\mathbf{S4}$ , that is  $A^{\mathbf{S}}$  is transitive and reflexive, then so is  $R$  and the same happens in all other cases.*

**Proof.** Left to the reader. ■

LEMMA 58. *for all  $x \in W$  and formulas  $B$ ,*

$$M, x \models B \Leftrightarrow \exists n x \in \text{Lab}(\Gamma_n) \wedge \Gamma_n, \alpha_n \vdash^? x : B, H_n \text{ succeeds}.$$

**Proof.** We prove both directions by mutual induction on  $cp(B)$ . If  $B$  is an atom then the claim holds by definition. Thus, assume  $B = C \Rightarrow D$ .

( $\Leftarrow$ ) Suppose for some  $m$   $\Gamma_m, \alpha_m \vdash^? x : C \Rightarrow D, H_m$  succeeds. Let  $xRy$  and  $M, y \models C$ , for some  $y$ . By definition of  $R$ , we have that for some  $n_1$ ,  $A_{\alpha_{n_1}}^{\mathbf{S}}(x, y)$  holds. Moreover, by the induction hypothesis, for some  $n_2$ ,  $\Gamma_{n_2}, \alpha_{n_2} \vdash^? y : C, H_{n_2}$  succeeds. Let  $k = \max\{n_1, n_2, m\}$ , then we have

1.  $\Gamma_k, \alpha_k \vdash^? x : C \Rightarrow D, H_k$  succeeds,
2.  $\Gamma_k, \alpha_k \vdash^? y : C, H_k$  succeeds,
3.  $A_{\alpha_k}^{\mathbf{S}}(x, y)$ .

So that from 1. we also have:

- 1'.  $(\Gamma_k, \alpha_k) \oplus_x (z : C) \vdash^? z : D, H_k$  succeeds, (with  $z \notin \text{Lab}(\Gamma_k) \cup \text{Lab}(H_k)$ ).

We can cut 1'. and 2., and obtain:

$$\Gamma_k, \alpha_k \cup \{(x, y)\} \vdash^? y : D, H_k \text{ succeeds}.$$

Hence, by 3. and Proposition 47(b) we get  $\Gamma_k, \alpha_k \vdash^? y : D, H_k$  succeeds, and by the induction hypothesis,  $M, y \models D$ ,

( $\Rightarrow$ ) Suppose by way of contradiction that  $M, x \models C \Rightarrow D$ , but for all  $n$  if  $x \in \text{Lab}(\Gamma_n)$ , then  $\Gamma_n, \alpha_n \vdash^? x : C \Rightarrow D, H_n$  fails. Let  $x \in \text{Lab}(\Gamma_n)$ , then there are  $m \geq k > n$ , such that  $(x, C \Rightarrow D) = (x_k, A_k) = (y_m, B_m)$  is considered at step  $m + 1$ , so that we have:

$$\Gamma_m, \alpha_m \vdash^? y_m : C \Rightarrow D, H_m \text{ fails.}$$

By Lemma 56, there is a  $y \in \mathcal{A}$ , such that (a) for  $t \leq m$ ,  $y \notin \text{Lab}(\Gamma_t)$  and (b):  $\forall m' > m$  (i)  $(y_n, y) \in \alpha_{m'}$ , (ii)  $\Gamma_{m'}, \alpha_{m'} \vdash^? y : C, H_{m'}$  succeeds, (iii)  $\Gamma_{m'}, \alpha_{m'} \vdash^? y : D, H_{m'}$  fails.

By (i) we have  $xRy$  holds, by (ii) and the induction hypothesis, we have  $M, y \models C$ . By (a) and (iii), we get:  $\forall n$  if  $y \in \text{Lab}(\Gamma_n)$ , then  $\Gamma_n, \alpha_n \vdash^? y : D, H_n$  fails. Hence, by the induction hypothesis, we have  $M, y \not\models D$ , and we get a contradiction. ■

**Proof** of The Completeness Theorem, 52. We are now able to conclude the proof of the completeness theorem. Let  $f(z) = z$ , for every  $z \in \text{Lab}(\Gamma_0)$ , where  $(\Gamma_0, \alpha_0) = (\Gamma, \alpha)$  is the original database. It is easy to see that  $f$  is a realization of  $(\Gamma, \alpha)$  in  $M$ : if  $A_\alpha^S(u, v)$  then  $A_{\alpha_0}^S(u, v)$ , hence  $f(u)Rf(v)$ . If  $u : C \in \Gamma = \Gamma_0$ , then by identity and the previous lemma we have  $M, f(u) \models C$ . On the other hand, by Lemma 55, and the previous lemma we have  $M, f(x) \not\models A$  and  $M, f(y) \not\models r$  for every  $(y, r) \in H$ . This concludes the proof. ■

By the previous theorem we immediately have the corollary.

**COROLLARY 59.** *If  $A \models_S B$  holds, then  $A \vdash^? x_0 : B, \emptyset$ , succeeds in  $P(\mathbf{S})$ . In particular, if  $A$  is valid in the modal system  $\mathbf{S}$ , then  $\vdash^? x_0 : A, \emptyset$ , succeeds in  $P(\mathbf{S})$ .*

## 4.2 Simplification for specific systems

In this section we show that for most of the modal logics we have considered, the use of labelled databases is not necessary and we can simplify either the structure of databases, or the deduction rules.

If we want to check the validity of a formula  $A$ , we evaluate  $A$  from a trivial database  $\vdash^? x_0 : A, \emptyset$ . Restricting our attention to computations from trivial databases, we observe that we can only generate databases which have the form of trees.

**DEFINITION 60.** A database  $(\Delta, \alpha)$  is called a *tree-database* if the set of links  $\alpha$  forms a tree.

Let  $(\Delta, \alpha)$  be a tree database and  $x \in \text{Lab}(\Delta)$ , we define the subdatabase  $\text{Path}(\Delta, \alpha, x)$  as the *list* of labelled formulas lying on the path from the root

of  $\alpha$ , say  $x_0$ , up to  $x$ , that is:  $Path(\Delta, \alpha, x) = (\Delta', \alpha')$ , where:

$$\begin{aligned} \alpha' &= \{(x_0, x_1), (x_1, x_2), \dots, (x_{n-1}, x_n) \mid x_n = x \\ &\quad \text{and for } i = 1, \dots, n, (x_{i-1}, x_i) \in \alpha\} \\ \Delta' &= \{y : A \in \Delta \mid y \in Lab(\alpha')\}. \end{aligned}$$

**PROPOSITION 61.** *If a query  $Q$  occurs in any derivation from a trivial database, then  $Q = \Delta, \alpha \vdash^? z : B, H$ , where  $(\Delta, \alpha)$  is a tree-database.*

From now on we restrict our consideration to tree-databases.

*Simplification for **K**, **K4**, **S4**, **KT**: Databases as Lists*

For systems **K**, **K4**, **S4**, **KT** the proof procedure can be simplified in the sense that: (i) the databases are lists of formulas, (ii) the restart rule is not needed. The key fact is expressed by the following theorem.

**THEOREM 62.** *If  $\Delta, \alpha \vdash^? x : A, \emptyset$  succeeds, then  $Path(\Delta, \alpha, x) \vdash^? x : A, \emptyset$  succeeds without using restart.*

Intuitively, only the formulas laying on the path from the root to  $x : A$  can be used in a proof of  $\Delta, \alpha \vdash^? x : A, \emptyset$ . The reason why restart is not needed is related: a restart step, say a restart from  $x : q$ , is useful only if we can take advantage of formulas at worlds created after the first call of  $x : q$  by means of the evaluation of an implicational goal. But these new worlds (being new) do not lay on the path from the root to  $x : q$ , thus they can be ignored and so can the restart step.

By virtue of this theorem we can reformulate the proof system for logics from **K** to **S4** as follows. A database is simply a list of formulas  $A_1, \dots, A_n$ , which stands for the labelled database  $(\{x_1 : A_1, \dots, x_n : A_n\}, \alpha)$ , where  $\alpha = \{(x_1, x_2), \dots, (x_{n-1}, x_n)\}$ . A query has the form:

$$A_1, \dots, A_n \vdash^? B$$

which represents  $\{x_1 : A_1, \dots, x_n : A_n\}, \alpha \vdash^? x_n : B$ . The history has been omitted since restart is not needed. Letting  $\Delta = A_1, \dots, A_n$ , we reformulate the predicates  $A^S$  as relations between formulas within a database  $A^S(\Delta, A_i, A_j)$ , in particular we can define:

$$\begin{aligned} A^K(\Delta, A_i, A_j) &\equiv i + 1 = j \\ A^{KT}(\Delta, A_i, A_j) &\equiv i = j \vee i + 1 = j \\ A^{K4}(\Delta, A_i, A_j) &\equiv i < j \\ A^{S4}(\Delta, A_i, A_j) &\equiv i \leq j \end{aligned}$$

The rules become:

- (success)  $\Delta \vdash^? q$  succeeds if  $\Delta = A_1, \dots, A_n$ , and  $A_n = q$ ;

- (implication) from  $\Delta \vdash^? A \Rightarrow B$  step to  $\Delta, A \vdash^? B$ ;
- (reduction) from  $\Delta \vdash^? q$  step to  $\Delta_i \vdash^? D_i$ , for  $i = 1, \dots, k$ ,  
if there is a formula  $A_j = D_1 \Rightarrow \dots \Rightarrow D_k \Rightarrow q \in \Delta$ , and there are integers  $j = j_0 \leq j_1 \leq \dots \leq j_k = n$ , such that

$$i = 1, \dots, k, A^{\mathbf{S}}(\Delta, A_{j_{i-1}}, A_{j_i}) \text{ holds and } \Delta_i = A_1, \dots, A_{j_i}.$$

EXAMPLE 63. We show that  $((b \Rightarrow a) \Rightarrow b) \Rightarrow c \Rightarrow (b \Rightarrow a) \Rightarrow a$  is a theorem of **S4**.

$$\begin{array}{rcl} & \vdash^? & ((b \Rightarrow a) \Rightarrow b) \Rightarrow c \Rightarrow (b \Rightarrow a) \Rightarrow a \\ (b \Rightarrow a) \Rightarrow b & \vdash^? & c \Rightarrow (b \Rightarrow a) \Rightarrow a \\ (b \Rightarrow a) \Rightarrow b, c & \vdash^? & (b \Rightarrow a) \Rightarrow a \\ (b \Rightarrow a) \Rightarrow b, c, b \Rightarrow a & \vdash^? & a \text{ reduction w.r.t. } b \Rightarrow a \text{ (1)} \\ (b \Rightarrow a) \Rightarrow b, c, b \Rightarrow a & \vdash^? & b \text{ reduction w.r.t. } (b \Rightarrow a) \Rightarrow b \text{ (2)} \\ (b \Rightarrow a) \Rightarrow b, c, b \Rightarrow a & \vdash^? & b \Rightarrow a \\ (b \Rightarrow a) \Rightarrow b, c, b \Rightarrow a, b & \vdash^? & a \text{ reduction w.r.t. } b \Rightarrow a \\ (b \Rightarrow a) \Rightarrow b, c, b \Rightarrow a, b & \vdash^? & b. \end{array}$$

This formula fails in both **KT** and **K4**, and therefore also fails in **K**: reduction at step (1) is allowed in **KT** but not in **K4**; on the contrary, reduction at step (2) is allowed in **K4** but not in **KT**.

#### *Simplification for K5, K45, S5: Databases as Clusters*

We can also give an unlabelled formulation of logics **K5**, **K45**, **S5**. The simplification is allowed by the fact that we can define explicitly the accessibility relation.

PROPOSITION 64. Let  $Q = \Delta, \alpha \vdash^? x : G, H$  be any query which occurs in a  $P(\mathbf{K5})$  deduction from a trivial database  $x_0 : A \vdash^? x_0 : B, H_0$ . Let  $R_\alpha^{\mathbf{K5}}(x, y)$  be defined as follows:

$$\begin{aligned} R_\alpha^{\mathbf{K5}}(x, y) \equiv & (x = x_0 \wedge (x_0, y) \in \alpha) \\ & \vee (\{x, y\} \subseteq \text{Lab}(\alpha) \wedge x \neq x_0 \wedge y \neq x_0). \end{aligned}$$

Then we have  $R_\alpha^{\mathbf{K5}}(x, y) \equiv A_\alpha^{\mathbf{K5}}(x, y)$ .

COROLLARY 65. Under the same conditions as the last proposition, we have:

$$R_\alpha^{\mathbf{K5}}(x_0, x) \text{ and } R_\alpha^{\mathbf{K5}}(x_0, y) \text{ implies } x = y.$$

PROPOSITION 66. Let  $Q = \Delta, \alpha \vdash^? x : G, H$  be any query which occurs in a  $P(\mathbf{K45})$  deduction from a trivial database  $x_0 : A \vdash^? x_0 : B, H_0$ . Let  $R_\alpha^{\mathbf{K45}}(x, y)$  be defined as follows:

$$R_{\alpha}^{\mathbf{K45}}(x, y) \equiv \{x, y\} \subseteq \text{Lab}(\alpha) \wedge y \neq x_0.$$

Then we have  $R_{\alpha}^{\mathbf{K45}}(x, y) \equiv A_{\alpha}^{\mathbf{K45}}(x, y)$ .

**PROPOSITION 67.** *Let  $Q = \Delta, \alpha \vdash^? x : G, H$  be any query which occurs in a  $P(\mathbf{S5})$  deduction from a trivial database  $x_0 : A \vdash^? x_0 : B, H_0$ . Let  $R_{\alpha}^{\mathbf{S5}}(x, y)$  be defined as follows:*

$$R_{\alpha}^{\mathbf{S5}}(x, y) \equiv \{x, y\} \subseteq \text{Lab}(\alpha).$$

Then we have  $R_{\alpha}^{\mathbf{S5}}(x, y) \equiv A_{\alpha}^{\mathbf{S5}}(x, y)$ .

From the previous propositions we can reformulate the proof systems for **K5**, **K45** and **S5** without making use of labels. For **K5** the picture is as follows: either a database contains just one point  $x_0$ , or there is an initial point  $x_0$  which is connected to another point  $x_1$ , and any point excluding  $x_0$  is connected to any other. In the case of **K45**,  $x_0$  is connected also to any point other than itself. Thus, in order to get a concrete structure without labels we must keep distinct the initial world from all the others, and we must indicate what is the current world, that is the world in which the goal formula is evaluated. In case of **K5** we must also identify the (only) world to which the initial world is connected. We are thus led to consider the following structure.

A non-empty database has the form:

$$\Delta = B_0 \mid \mid \quad \text{or} \quad \Delta = B_0 \mid B_1, \dots, B_n \mid B_i, \text{ where } 1 \leq i \leq n,$$

and  $B_0, B_1, \dots, B_n$  are formulas. We also define

$$\text{Actual}(\Delta) = \begin{cases} B_0 & \text{if } \Delta = B_0 \mid \mid, \\ B_i & \text{if } \Delta = B_0 \mid B_1, \dots, B_n \mid B_i. \end{cases}$$

This rather odd structure is forced by the fact that in **K5** and **K45** we have reflexivity in all worlds, except in the initial one and therefore, in contrast to all other systems, we have considered so far, the success of

$$\vdash^? x_0 : A \Rightarrow B, \text{ which means that } A \Rightarrow B \text{ is valid,}$$

does not imply the success of

$$x_0 : A \vdash^? x_0 : B, \text{ which means that } A \rightarrow B \text{ is valid (material implication).}^{12}$$

The addition operation is defined as follows:

$$\Delta \oplus A = \begin{cases} B_0 \mid B_1, \dots, B_n, A \mid A & \text{if } \Delta = B_0 \mid B_1, \dots, B_n \mid B_i \\ B_0 \mid A \mid A & \text{if } \Delta = B_0 \mid \mid \\ \top \mid A \mid A & \text{if } \Delta = \emptyset \end{cases}$$

<sup>12</sup>In these two systems the validity of  $\Box C$  does not imply the validity of  $C$ , as it holds for all the other systems considered in this section.



A query has the form

$$\Delta \vdash^? G, H, \text{ where } H = \{(A_1, q_1), \dots, (A_k, q_k)\}, \text{ with } A_j \in \Delta.$$

DEFINITION 68 (Deduction Rules for **K5** and **K45**).

Given  $\Delta = B_0 \mid B_1, \dots, B_n \mid B$ , let

$$\begin{aligned} A^{\mathbf{K5}}(\Delta, X, Y) &\equiv (X = B_0 \wedge Y = B_1) \\ &\quad \vee (X = B_i \wedge Y = B_j \wedge i, j > 0) \text{ and} \\ A^{\mathbf{K45}}(\Delta, X, Y) &\equiv (X = B_i \wedge Y = B_j \text{ with } j > 0) \end{aligned}$$

- (success)  $\Delta \vdash^? q, H$  succeeds if  $Actual(\Delta) = q$ .
- (implication) From  $\Delta \vdash^? A \Rightarrow B, H$  step to  $\Delta \oplus A \vdash^? B, H$ .
- (reduction) if  $\Delta = B_0 \mid B_1, \dots, B_n \mid B$  and  $C = D_1 \Rightarrow \dots \Rightarrow D_k \Rightarrow q \in \Delta$ , from  $\Delta \vdash^? G, H$  step to

$$B_0 \mid B_1, \dots, B_n \mid C_i \vdash^? D_i, H \cup \{(B, q)\} \text{ for } i = 1, \dots, k,$$

for some  $C_0, \dots, C_k \in \Delta$ , such that  $C_0 = C$ ,  $C_k = B_n$ , and  $A^{\mathbf{K5}}(\Delta, C_{i-1}, C_i)$  (respectively  $A^{\mathbf{K45}}(\Delta, C_{i-1}, C_i)$ ) holds.

- (restart) If  $\Delta = B_0 \mid B_1, \dots, B_n \mid B_i$  and  $(B_j, r) \in H$ , with  $j > 0$ , then from  $\Delta \vdash^? q, H$ , step to

$$B_0 \mid B_1, \dots, B_n \mid B_j \vdash^? r, H \cup \{(B_i, q)\},$$

According to the above discussion, we observe that the check of the validity of  $\models A \Rightarrow B$ , corresponds to the query

$$\emptyset \vdash^? A \Rightarrow B, \emptyset,$$

which (by the implication rule) is reduced to the query

$$\top \mid A \mid A \vdash^? B, \emptyset.$$

This is different from checking the validity of  $A \rightarrow B$  ( $\rightarrow$  is the material implication), which corresponds to the query

$$A \mid \mid \vdash^? B, \emptyset.$$

The success of the former query does not imply the success of the latter. For instance in **K5**,

$$\not\models (\top \Rightarrow p) \rightarrow p \text{ and indeed } \top \Rightarrow p \mid \mid \vdash^? p, \emptyset \text{ fails.}$$

On the other hand we have

$\models (\top \Rightarrow p) \Rightarrow p$  and indeed  $\top \mid \top \Rightarrow p \mid \top \Rightarrow p \vdash^? p, \emptyset$  succeeds.

The reformulation of the proof system for **S5** is similar, but simpler. In the case of **S5**, there is no need to keep the first formula/world apart from the others. Thus, we may simply define a non-empty database as a pair  $\Delta = (S, A)$ , where  $S$  is a set of formulas and  $A \in S$ . If  $\Delta = (S, A)$ , we let

$$Actual(\Delta) = A \text{ and } \Delta \oplus B = (S \cup \{B\}, B).$$

For  $\Delta = \emptyset$ , we define  $\emptyset \oplus A = (\{A\}, A)$ . With these definitions the rules are similar to those of **K5** and **K45**, with the following simplifications:

- (reduction) if  $\Delta = (S, B)$  and  $C = D_1 \Rightarrow \dots \Rightarrow D_k \Rightarrow q \in \Delta$ , then from  $\Delta \vdash^? G, H$  step to

$$(S, C_i) \vdash^? D_i, H \cup \{(B, q)\}, \text{ where for } i = 1, \dots, k \ C_i \in \Delta \text{ and } C_k = B.$$

- (restart) If  $\Delta = (S, B)$  and  $(C, r) \in H$ , then from  $(S, B) \vdash^? q, H$ , step to

$$(S, C) \vdash^? r, H \cup \{(B, q)\},$$

**EXAMPLE 69.** In Figure 6 we show a derivation of the following formula in **S5**

$$((a \Rightarrow b) \Rightarrow c) \Rightarrow (a \Rightarrow d \Rightarrow c) \Rightarrow (d \Rightarrow c).$$

In the derivation we make use of restricted restart, according to Proposition 41. A brief explanation of the derivation: step (5) is obtained by reduction w.r.t.  $(a \Rightarrow b) \Rightarrow c$ , step (7) by restart, steps (8) and (9) by reduction w.r.t.  $a \Rightarrow d \Rightarrow c$ , and they both succeed immediately.

### 4.3 Extending the language

In this section we extend the proof procedures to broader fragments. We first consider a simple extension allowing conjunction. To handle conjunction in the labelled formulation, we simply drop the condition that a label  $x$  may be attached to only one formula, thus formulas with the same label can be thought as logically conjuncted. In the unlabelled formulation, for those systems enjoying such a formulation, the general principle is to deal with *sets* of formulas, instead of *single formulas*. A database will be a structured collection of *sets* of formulas, rather than a collection of formulas. The structure is always the same, but the constituents are now sets. Thus, in the cases of **K**, **KT**, **K4** and **S4**, databases will be lists of *sets* of formulas,



For the deduction procedures all we have to do is to handle sets of formulas. We define, for  $x \in Lab(\Gamma) \cup Lab(H)$ ,  $y \notin Lab(\Gamma) \cup Lab(H)$  and finite set of formulas  $S = \{D_1, \dots, D_t\}$ ,

$$(\Gamma, \alpha) \oplus_x y : S = (\Delta \cup \{y : D_1, \dots, y : D_t\}, \alpha \cup \{(x, y)\}),$$

then we change the (*implication*) rule in the obvious way:

$$\text{from } \Delta, \alpha \vdash^? x : S \Rightarrow B, H,$$

step to

$$(\Delta, \alpha) \oplus_x y : S \vdash^? y : B, H,$$

where  $S$  is a set of formulas in NF and  $y \notin Lab(\Gamma)$ , and we add a rule for proving sets of formulas:

$$\text{from } (\Delta, \alpha) \vdash^? x : \{B_1, \dots, B_k\}, H$$

step to

$$(\Delta, \alpha) \vdash^? x : B_i, H \text{ for } i = 1, \dots, k.$$

Regarding the simplified formulations without labels, the structural restrictions in the rules (reduction and success) are applied to the sets of formulas, which are now the constituents of databases, considered as units; the history  $H$ , when needed, becomes a set of pairs  $(S_i, A_i)$ , where  $S_i$  is a set and  $A_i$  is a formula. The property of restricted restart still holds for this formulation.

We can extend further extend the  $\mathcal{L}(\Rightarrow, \wedge)$ -fragment in two directions. In one direction, we can define a modal analogue of Harrop formulas for intuitionistic logic that we have introduced in Section 3.7. This extension is relevant for logic programming applications [Giordano *et al.*, 1992; Giordano and Martelli, 1994]. In the other direction, we can define a proof system for the whole propositional modal language via a translation into an implicative normal form.

### *Modal Harrop formulas*

We can define a modal analogue of Harrop formulas by allowing disjunction of goals and *local clauses* of the form

$$G \rightarrow q,$$

where  $\rightarrow$  denotes ordinary (material) implication. We call them ‘local’, since  $x : G \rightarrow q$  can be used only in world  $x$  to reduce the goal  $x : q$  and it is not usable/visible in any other world. It is ‘private’ to  $x$ , whereas the ‘global’ clause  $G \Rightarrow q$  can be used to reduce  $q$  in any world  $y$  accessible from  $x$ . It is not a case that modalities have been used to implement visibility rules and structuring mechanisms in logic programming [Giordano *et al.*, 1992;

Giordano and Martelli, 1994]. In order to define a modal Harrop fragment we distinguish D-formulas, which are the constituents of databases, and G-formulas which can occur as goals. The former are further distinct in *modal* D-formulas (MD) and *local* D-formulas (LD).

$$\begin{aligned}
LD &:= G \rightarrow q, \\
MD &:= \top \mid q \mid G \Rightarrow MD, \\
D &:= LD \mid MD, \\
CD &:= D \mid CD \wedge CD; \\
G &:= \top \mid q \mid G \wedge G \mid G \vee G \mid CD \Rightarrow G.
\end{aligned}$$

We also use  $\Box G$  and  $\Box D$  as syntactic sugar for  $\top \Rightarrow G$  and  $\top \Rightarrow D$ . Notice that atoms are both LD- and MD-formulas (as  $\top \rightarrow q \equiv q$ ); moreover, any non-atomic MD-formula can be written as  $G_1 \Rightarrow \dots \Rightarrow G_k \Rightarrow q$ . Finally, CD formulas are just conjunction of D-formulas.

For D- and G-formulas as defined above we can easily extend the proof procedure. We give it in the most general formulation for labelled databases. It is clear that one can derive an unlabelled formulation for systems which allow it, as explained in the previous section. In the labelled formulation, queries have the form

$$\Delta, \alpha \vdash^? x : G, H$$

where  $\Delta$  is a set of D-formulas,  $G$  is a G-formula, and  $H = \{(x_1, G_1), \dots, (x_k, G_k)\}$ , where  $G_i$  are G-formulas. The **additional** rules are:

- (true)  $\Delta, \alpha \vdash^? x : \top, H$  immediately succeeds.
- (local-reduction) From  $\Delta, \alpha \vdash^? x : q, H$  step to
$$\Delta, \alpha \vdash^? x : G, H \cup \{(x : q)\}$$
if  $x : G \rightarrow q \in \Delta$ .
- (and) From  $\Delta, \alpha \vdash^? x : G_1 \wedge G_2, H$  step to
$$\Delta, \alpha \vdash^? x : G_1, H \text{ and } \Delta, \alpha \vdash^? x : G_2, H.$$
- (or) From  $\Delta, \alpha \vdash^? x : G_1 \vee G_2, H$  step to
$$\Delta, \alpha \vdash^? x : G_1, H \cup \{(x, G_2)\} \text{ or to } \Delta, \alpha \vdash^? x : G_2, H \cup \{(x, G_1)\}.$$

EXAMPLE 70. Let  $\Delta$  be the following database

$$\begin{aligned}
x_0 &: [(\Box p \Rightarrow s) \wedge b] \rightarrow q, \\
x_0 &: ((p \Rightarrow q) \wedge \Box a] \Rightarrow r) \rightarrow q, \\
x_0 &: a \rightarrow b.
\end{aligned}$$



*Extension to the whole propositional language*

We can easily extend the procedure to the whole propositional modal language: we just consider the computation procedure for classical logic and we combine it with the modal procedure for  $\mathcal{L}(\Rightarrow, \wedge)$ . To minimize the work, we can introduce a simple normal form on the set of connectives  $(\Rightarrow, \rightarrow, \wedge, \top, \perp)$ . It is obvious that this set forms a complete base for modal logic. The normal form is an immediate extension of the normal form for  $\Rightarrow, \wedge$ .

**PROPOSITION 72.** *Every modal formula over the language  $(\rightarrow, \neg, \diamond, \square)$  is equivalent to a set (conjunction) of NF-formulas of the form*

$$S_0 \rightarrow (S_1 \Rightarrow (S_2 \Rightarrow \dots \Rightarrow (S_n \Rightarrow q) \dots))$$

where  $q$  is an atom,  $\top$ , or  $\perp$ ,  $n \geq 0$ , and each  $S_i$  is a conjunction (set) of NF-formulas.

As usual we omit parentheses, so that the above will be written as

$$S_0 \rightarrow S_1 \Rightarrow S_2 \Rightarrow \dots \Rightarrow S_n \Rightarrow q.$$

In practice we will replace the conjunction by the set notation as we wish. When we need it, we distinguish two types of NF-formulas, (i) those with non-empty  $S_0$ , which are written as above, and (ii) those with empty  $S_0$ , which are simplified to  $S_1 \Rightarrow S_2 \Rightarrow \dots \Rightarrow S_n \Rightarrow q$ . For a quick case analysis, we can also say that type (i) formulas have the form  $S \rightarrow D$ , and type (ii) have the form  $S \Rightarrow D$ , where  $D$  is always of type (ii).

For instance, the NF-form of  $p \rightarrow \diamond r$  is

$$(p \wedge (r \Rightarrow \perp)) \rightarrow \perp, \text{ or equivalently } \{p, r \Rightarrow \perp\} \rightarrow \perp.$$

This formula has the structure  $S_0 \rightarrow q$ , where  $S_0 = \{p, r \Rightarrow \perp\}$  and  $q = \perp$ . The NF-form of  $\square(\diamond a \rightarrow \diamond(b \wedge c))$  is given by  $((a \Rightarrow \perp) \rightarrow \perp) \Rightarrow ((b \wedge c) \Rightarrow \perp) \Rightarrow \perp$ . We give below the rules for queries of the form

$$\Gamma, \alpha \vdash^? x : G, H,$$

where  $\Gamma$  is a labelled set of NF-formulas,  $G$  is a NF-formula,  $\alpha$  is a set of links (as usual),  $H$  is a set of pairs  $\{(x_1, q_1), \dots, (x_k, q_k)\}$ , where  $q_i$  is an atom.

**DEFINITION 73** (Deduction rules for whole modal logics). For each modal system  $\mathbf{S}$ , the corresponding proof system, denoted by  $\mathbf{P}(\mathbf{S})$ , comprises the following rules, parametrized to predicates  $A^{\mathbf{S}}$ .

- (success)  $\Delta, \alpha \vdash^? x : q, H$  immediately succeeds if  $q$  is an atom and  $x : q \in \Delta$ .
- (strict implication) From  $\Delta, \alpha \vdash^? x : S \Rightarrow D, H$  step to

$$(\Delta, \alpha) \oplus_x (y : S) \vdash^? y : D, H,$$

where  $y \notin \text{Lab}(\Delta) \cup \text{Lab}(H)$ .

- (implication) From  $\Delta, \alpha \vdash^? x : S \rightarrow D, H$  step to

$$\Delta \cup \{x : A \mid A \in S\}, \alpha \vdash^? x : D, H.$$

- (reduction) If  $y : C \in \Delta$ , with  $C = S_0 \rightarrow S_1 \Rightarrow S_2 \Rightarrow \dots \Rightarrow S_k \Rightarrow q$ , with  $q$  atomic, then from

$$\Delta, \alpha \vdash^? x : q, H$$

step to

$$\begin{aligned} \Delta, \alpha \vdash^? u_0 : S_0, H' \\ \Delta, \alpha \vdash^? u_1 : S_1, H' \\ \vdots \\ \Delta, \alpha \vdash^? u_k : S_k, H' \end{aligned}$$

where  $H' = H$  if  $q = \perp$ , and  $H' = H \cup \{(x, q)\}$  otherwise, for some  $u_0, \dots, u_k \in \text{Lab}(\alpha)$ , such that  $u_0 = y$ ,  $u_k = x$ , and

$$A_\alpha^S(u_i, u_{i+1}) \text{ holds, for } i = 0, \dots, k-1.$$

- (restart) If  $(y, r) \in H$ , then, from  $\Delta, \alpha \vdash^? x : q, H$ , with  $q$  atomic, step to

$$\Delta, \alpha \vdash^? y : r, H \cup \{(x, q)\}.$$

- (falsity) From  $\Delta, \alpha \vdash^? x : q, H$ , if  $y \in \text{Lab}(\Gamma)$  step to

$$\Delta, \alpha \vdash^? y : \perp, H \cup \{(x : q)\}.$$

- (conjunction) From  $(\Delta, \alpha) \vdash^? x : \{B_1, \dots, B_k\}, H$  step to

$$(\Delta, \alpha) \vdash^? x : B_i, H \text{ for } i = 1, \dots, k.$$

If the number of subgoals is 0, i.e.  $k = 0$ , the above reduction rule becomes the rule for *local clauses* of the previous section. On the other hand if  $S_0 = \emptyset$ , then the query with goal  $S_0$  is omitted and we have the rule of Section 4.1.

The proof procedure is sound and complete, as asserted in the next theorem, and the completeness proof is just a minor extension of the one of Theorem 52.

**THEOREM 74.**  $\Gamma, \gamma \vdash^? x : G, H$  succeeds if and only if it is valid.

**EXAMPLE 75.** In **K5** we have  $\diamond p \rightarrow \Box \diamond p$ . This is translated as  $((p \Rightarrow \perp) \rightarrow \perp) \rightarrow ((p \Rightarrow \perp) \Rightarrow \perp)$ . Below we show a derivation. Some



explanation and remarks: at step (1) we can only apply the rule for falsity, or reduction w.r.t.  $y : p \Rightarrow \perp$ , since  $A_\alpha^{\mathbf{K5}}(x_0, y)$  implies  $A_\alpha^{\mathbf{K5}}(y, y)$ . We apply the rule for falsity. The reduction at step (2) is legitimate as  $A_\alpha^{\mathbf{K5}}(x_0, y)$  and  $A_\alpha^{\mathbf{K5}}(x_0, z)$  implies  $A_\alpha^{\mathbf{K5}}(y, z)$ .

$$\begin{array}{lcl}
& \vdash^? x_0 : ((p \Rightarrow \perp) \rightarrow \perp) \rightarrow ((p \Rightarrow \perp) \Rightarrow \perp) & \\
x_0 : (p \Rightarrow \perp) \rightarrow \perp & \vdash^? x_0 : (p \Rightarrow \perp) \Rightarrow \perp & \\
y : p \Rightarrow \perp, \alpha = \{(x_0, y)\} & \vdash^? y : \perp & (1) \\
& \vdash^? x_0 : \perp \quad \text{rule for } \perp & \\
& \vdash^? x_0 : p \Rightarrow \perp & \\
z : p, \alpha = \{(x_0, y), (x_0, z)\} & \vdash^? z : \perp & (2) \\
& \vdash^? z : p & \\
& \text{success} & 
\end{array}$$

This proof procedure is actually a minor extension of the one based on strict-implication/conjunction. The rule for falsity may be source of non-determinism, as it can be applied to any label  $y$ . Further investigation should clarify to what extent this rule is needed and if it is possible to restrict its applications to special cases. Another point which deserve investigation is *termination*. The proof procedure we have described may not terminate. The two standard techniques to ensure termination, loop-checking and diminishing-resources, could be possibly applied in this context. Again further investigation is needed to clarify this point, taking into account the known results (see [Viganò, 1999; Heudering *et al.*, 1996]).

#### 4.4 Some history

Many authors have developed analytic proof methods for modal logics, (see the fundamental book by Fitting [1983], and Goré [1999] for a recent and comprehensive survey).

The use of goal-directed methods in modal logic has not been fully explored. The most relevant work in this area is the one by Giordano, Martelli and colleagues [Giordano *et al.*, 1992; Giordano and Martelli, 1994; Baldoni *et al.*, 1998] who have developed goal-directed methods for fragments of first-order (multi-)modal logics. Their work is motivated by several purposes: introducing scoping constructs (such as blocks and modules) in logic programming, representing epistemic and inheritance reasoning. In particular in [Giordano and Martelli, 1994] a family of first-order logic programming languages is defined, based on the modal logic **S4** with the aim of representing a variety of scoping mechanisms. If we restrict our consideration to the propositional level, their languages are strongly related to the one defined in Section 4.3 in the case the underlying logic is **S4**. The largest (propositional) fragment of **S4** they consider, called  $L_4$ , is very close

to the one defined in the previous Section for modal-Harrop formulas, although neither one of the two is contained in the other. The proof procedure they give for  $L_4$  (at the propositional level) is essentially the same as the unlabelled version of P(**S4**) for modal Harrop formulas.

Abadi and Manna [1989] have defined an extension of PROLOG, called *TEMPLOG* based on a fragment of first-order temporal logic. Their language contains the modalities  $\diamond$ ,  $\square$ , and the temporal operator  $\bigcirc$  (next). They introduce a notion of temporal Horn clause whose constituents are atoms  $B$  possibly prefixed by an arbitrary sequence of next, i.e.  $B \equiv \bigcirc^k A$  (with  $k \geq 0$ ). The modality  $\square$  is allowed in front of clauses (permanent clauses) and clause-heads, whereas the modality  $\diamond$  is allowed in front of goals. The restricted format of the rules allows one to define an efficient and simple goal-directed procedure without the need of any syntactic structuring or labelling. An alternative, although related, extension based on temporal logic has been studied in [Gabbay, 1987].

Farinàs in [1986] describes MOLOG a (multi)-modal extension of PROLOG. His proposal is more a general framework than a specific language, in the sense that the language can support different modalities governed by different logics. The underlying idea is to extend classical resolution by special rules of the following pattern: let  $B, B'$  be modal atoms (i.e. atomic formulas possibly prefixed by modal operators), then if  $G \rightarrow B$  is a clause and  $B' \wedge C_1 \wedge \dots \wedge C_k$  is the current goal, and

$$(*) \models_{\mathbf{S}} B \equiv B' \text{ holds}$$

then the goal can be reduced to  $G \wedge C_1 \wedge \dots \wedge C_k$ . It is clear that the effectiveness of the method depends on how difficult it is to check (\*); in case of conventional logic programming the (\*) test is reduced to unification. The proposed framework is exemplified in [Farinàs, 1986] by defining a multi-modal language based on **S5** with necessity operators such as *Knows*( $a$ ). In this case one can define a simple matching predicate for the test in (\*), and hence an effective resolution rule.

In general, we can distinguish two paradigms in proof systems for modal logics: on the one hand we have *implicit* calculi in which each proof configuration contains a set of formulas implicitly representing a single possible world; the modal rules encodes the shifts of world by manipulating sets of formulas and formulas therein. On the other hand we have explicit methods in which the possible world structure is explicitly represented using labels and relations among them; the rules can create new worlds, or move formulas around them. In between there are ‘intermediate’ proof methods which add some semantic structure to flat sequents, but they do not explicitly represent a Kripke model [Masini, 1992; Wansing, 1994; Goré, 1999].

The use of labels to represent worlds for modal logics is rather old and goes back to Kripke himself. In the seminal work [Fitting, 1983] formulas

are labelled by strings of atomic labels (world prefixes), which represent paths of accessible worlds. The rules for modalities are the same for every system: for instance if a branch contains  $\sigma : \Box A$ , and  $\sigma'$  is accessible from  $\sigma$ , then one can add  $\sigma' : A$  to the same branch. For each system, there are some specific accessibility conditions on prefixes which constraint the propagation of modal formulas. This approach has been recently improved by Massacci [1994]. Basin, Matthews and Viganò have developed a proof-theory for modal logics making use of labels and an explicit accessibility relation [Basin *et al.*, 1997a; Basin *et al.*, 1999; Viganò, 1999]. A related approach was presented in [Gabbay, 1996] and [Russo, 1996]. These authors have developed both sequent and natural deduction systems for several modal logics which are completely uniform.

If we forget the goal-directed feature the proof methods presented in this section clearly belongs to the ‘explicit’-calculi tradition in their labelled version, and to the ‘intermediate’-calculi tradition calculi in their unlabelled version. The sequence of (sets of) formulas represent a sequence of possible worlds. It is not a case that the unlabelled version of  $\mathbf{K}$  is strongly related to the two-dimensional sequent calculus by Masini [1992].

## 5 SUBSTRUCTURAL LOGICS

### 5.1 Introduction

In this section we consider substructural logics. The denomination *substructural logics* comes from sequent calculi terminology. In sequent calculi, there are rules which introduce logical operators and rules which modify the structure of sequents. The latter are called the *structural rules*. In case of classical and intuitionistic logic these rules are *contraction*, *weakening* and *exchange*. Substructural logics restrict or allow a finer control on structural rules. More generally, substructural logics restricts *the use* of formulas in a deduction. The restrictions may require either that every formula of the database must be used, or that it cannot be used more than once, or that it must be used according to a given ordering of database formulas.

We present the systems of substructural logics, restricted to their implicational fragment, by means of a Hilbert axiomatization and a possible-world semantics.

**DEFINITION 76** (Axiomatization of implication). We consider the following list of axioms:

- (id)  $A \rightarrow A$ ;
- (h1)  $(B \rightarrow C) \rightarrow (A \rightarrow B) \rightarrow A \rightarrow C$ ;
- (h2)  $(A \rightarrow B) \rightarrow (B \rightarrow C) \rightarrow A \rightarrow C$ ;

- (h3)  $(A \rightarrow A \rightarrow B) \rightarrow A \rightarrow B$ ;  
 (h4)  $(A \rightarrow B) \rightarrow ((A \rightarrow B) \rightarrow C) \rightarrow C$ ;  
 (h5)  $A \rightarrow (A \rightarrow B) \rightarrow B$ ;  
 (h6)  $A \rightarrow B \rightarrow B$ .

Together with the following rules:

$$\frac{A \rightarrow B \quad A}{B} (MP)$$

$$\frac{A \rightarrow B}{(B \rightarrow C) \rightarrow A \rightarrow C} (Suff).$$

Each system is axiomatized by taking the closure under modus ponens (MP) and under substitution of the combinations of axioms/rules of Table 3.

Table 3. Axioms for substructural implication.

Logic	Axioms
<b>FL</b>	(id), (h1), ( <i>Suff</i> )
<b>T-W</b>	(id), (h1), (h2)
<b>T</b>	(id), (h1), (h2), (h3)
<b>E-W</b>	(id), (h1), (h2), (h4)
<b>E</b>	(id), (h1), (h2), (h3), (h4)
<b>L</b>	(id), (h1), (h2), (h5)
<b>R</b>	(id), (h1), (h2), (h3), (h5)
<b>BCK</b>	(id), (h1), (h2), (h5), (h6)
<b>I</b>	(id), (h1), (h2), (h3), (h5), (h6)

In the above axiomatization, we have not worried about the minimality and independence of the group of axioms for each system. For some systems the corresponding list of axioms given above is redundant, but it quickly shows some inclusion relations among the systems. We just remark that in presence of (h4), (h2) can be obtained by (h1). Moreover, (h4) is a weakening of (h5). The rule of (*Suff*) is clearly obtainable from (h2) and (MP). To have a complete picture we have included also intuitionistic logic **I**, although the axiomatization above is highly redundant (see Section 3.1).

We give a brief explanation of the names of the systems and how they are known in the literature.<sup>13</sup> **R** is the most important system of *relevant*

<sup>13</sup>The names **R**, **E**, **T**, **BCK**, etc. in this section refer mainly to *the implicative fragment* of the logical systems known in the literature [Anderson and Belnap, 1975]

logic and it is axiomatized by dropping the irrelevant axiom (h6) from the axiomatization of intuitionistic implication.

The system **E** combines relevance and necessity. The implication of **E** can be read at the same time as relevant implication and strict implication. Moreover, we can define

$$\Box A =_{def} (A \rightarrow A) \rightarrow A,$$

and **E** interprets  $\Box$  the same as **S4**. **E** is axiomatized by restricting the exchange axiom (h5) to implicational formulas (the axiom (h4)).

The weaker **T** stems from a concern about the use of the two hypotheses in an inference by Modus Ponens: the restriction is that the minor  $A$  must not be derived ‘before’ the ticket  $A \rightarrow B$ . This is clarified by the Fitch-style natural deduction of **T**, for which we refer to [Anderson and Belnap, 1975].

**BCK** is the system which result from intuitionistic implicational logic by dropping contraction. **L** rejects both weakening and contraction and it is the implicational fragment of linear logic [Girard, 1987] (also commonly known as **BCI** logic).

We will also consider contractionless versions of **E** and **T**, , namely **E-W** and **T-W** respectively.

The weakest system we consider is **FL**,<sup>14</sup> which is related to the *right implicational* fragment of Lambek calculus. This system rejects all sub-structural rules.

We will mainly concentrate on the implicational fragment of the systems mentioned. In Section 5.3, we will extend the proof systems to a fragment similar to Harrop-formulas. For the fragment considered, all the logics studied in this section are subsystems of intuitionistic logic. However, this is no longer true for the fragment comprising an involutive negation, which can be added (and has been added) to each system. In this section we do not consider the treatment of negation. We refer the reader to [Anderson and Belnap, 1975; Anderson *et al.*, 1992] for an extensive discussion.

In Figure 5.1 we show the inclusion relation of the systems we consider in this section.

We give a corresponding semantics for this set of systems. The semantics we refer is a simplification of the one proposed in [Fine, 1974], [Anderson *et al.*, 1992] and elaborated more recently by Došen [1988; 1989].<sup>15</sup>

---

with the corresponding names. The implicational fragments are usually denoted with the subscript  $\rightarrow$ . Thus, what we call **R** is denoted in the literature by **R** $\rightarrow$  and so forth; since we are mainly concerned with the implicational systems we have preferred to minimize the notation, stating explicitly when we make exception to this convention.

<sup>14</sup>The denomination of the system is taken from [Ono, 1998; Ono, 1993].

<sup>15</sup>Dealing only with the implicational fragment, we have simplified Fine semantics: we do not have *prime* or maximal elements.

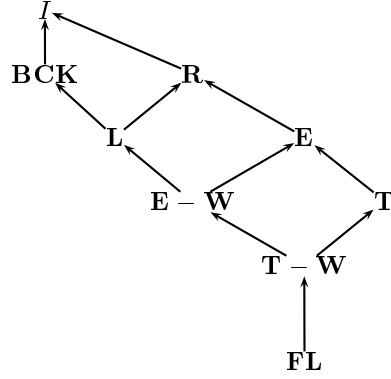


Figure 8. Lattice of Substructural Logics.

DEFINITION 77. Let us fix a language  $\mathcal{L}$ , a Fine  $\mathbf{S}$ -structure<sup>16</sup>  $M$  is a tuple of the form:

$$M = (W, \leq, \circ, 0, V),$$

where  $W$  is a non empty set,  $\circ$  is a binary operation on  $W$ ,  $0 \in W$ ,  $\leq$  is a partial order relation on  $W$ ,  $V$  is a function of type  $W \rightarrow Pow(Var)$ . In all structures the following properties are assumed to hold:

$$\begin{aligned} 0 \circ a &= a, \\ a \leq b &\text{ implies } a \circ c \leq b \circ c, \\ a \leq b &\text{ implies } V(a) \subseteq V(b). \end{aligned}$$

For each system  $\mathbf{S}$ , a  $\mathbf{S}$ -structure satisfies a subset of the following conditions, as specified in Table 4

- (a1)  $a \circ (b \circ c) \leq (a \circ b) \circ c$ ;
- (a2)  $a \circ (b \circ c) \leq (b \circ a) \circ c$ ;
- (a3)  $(a \circ b) \circ b \leq a \circ b$ ;
- (a4)  $a \circ 0 \leq a$ ;
- (a5)  $a \circ b \leq b \circ a$ ;
- (a6)  $0 \leq a$ .

**Truth conditions** for  $a \in W$ , we define

- $M, a \models p$  if  $p \in V(a)$ ;

<sup>16</sup>We just write  $\mathbf{S}$ -structure if there is no risk of confusion.

- $M, a \models A \rightarrow B$  if

$$\forall b \in W (M, b \models A \Rightarrow M, a \circ b \models B).$$

We say that  $A$  is *valid* in  $M$  (denoted by  $M \models A$ ) if  $M, 0 \models A$ . We say that  $A$  is **S**-valid, denoted by  $\models_{\mathbf{S}}^{Fine} A$  if  $A$  is valid in every **S**-structure.

Table 4. Algebraic conditions of Fine Semantics.

Logic	(a1)	(a2)	(a3)	(a4)	(a5)	(a6)
<b>FL</b>	*					
<b>T-W</b>	*	*				
<b>T</b>	*	*	*			
<b>E-W</b>	*	*		*		
<b>E</b>	*	*	*	*		
<b>L</b>	*	*		*	*	
<b>R</b>	*	*	*	*	*	
<b>BCK</b>	*	*		*	*	*
<b>I</b>	*	*	*	*	*	*

We have included again intuitionistic logic **I** in Table 4 to show its proper place within this framework. Again this list of semantical conditions is deliberately redundant in order to show quickly the inclusion relation among the systems. The axiomatization given above is sound and complete with respect to this semantics. In particular each axiom (hi) corresponds to the semantical condition (ai).

**THEOREM 78** (Anderson *et al.*, 1992, Fine, 1974, Došen, 1989).  $\models_{\mathbf{S}} A$  if and only if  $A$  is derivable in the corresponding axiom system of Definition 76.

We assume that  $\circ$  associates to the left, so we write

$$a \circ b \circ c = (a \circ b) \circ c.$$

## 5.2 Proof systems

We develop proof methods for the implicative logics: **R**, **BCK**, **E**, **T**, **E-W**, **T-W**, **FL**. As we have seen in the section about modal logics, we can control the use of formulas by labelling data and putting constraints on the labels. In this specific context by labelling data, we are able to record whether they have been used or not and to express the additional conditions needed for each specific system. Formulas are labelled with atomic labels  $x, y, z$ . Intuitively these labels can be read as representing at the same time *resources* and *positions* within a database.

DEFINITION 79. Let us fix a denumerable alphabet  $\mathcal{A} = \{x_1, \dots, x_i, \dots\}$  of labels. We assume that labels are totally ordered as shown in the enumeration,  $\mathbf{v}_0$  is the first label. A *database* is a finite set of labelled formulas  $\Delta = \{x_1 : A_1, \dots, x_n : A_n\}$ . We assume that

$$\text{if } x : A \in \Delta \text{ and } x : B \in \Delta, \text{ then } A = B.^{17}$$

We use the notation  $Lab(E)$  for the set of labels occurring in an expression  $E$ , and we finally assume that  $\mathbf{v}_0 \notin Lab(\Delta)$ . Label  $\mathbf{v}_0$  will be used for queries from the empty database.

DEFINITION 80. A query  $Q$  is an expression of the form:

$$\Delta, \delta \vdash^? x : G$$

where  $\Delta$  is a database,  $\delta$  is a finite set of labels not containing  $\mathbf{v}_0$ ; moreover if  $x \neq \mathbf{v}_0$  then  $x \in Lab(\Delta)$ , and  $G$  is a formula.

A query from the empty database has the form:

$$\vdash^? \mathbf{v}_0 : G.$$

Let  $\max(\delta)$  denote the maximum label in  $\delta$  according to the enumeration of the labels. By convention, we stipulate that if  $\delta = \emptyset$ , then  $\max(\delta) = \mathbf{v}_0$ . The set of labels  $\delta$  may be thought as denoting the set of resources that are available to prove the goal. Label  $x$  in front of the goal has a double role as a ‘position’ in the database from which the goal is asked, and as available resource.

The rules for success and reduction are parametrized to some conditions  $Succ^S$  and  $Red^S$  that will be defined below.

- (success)  $\Delta, \delta \vdash^? x : q$ ; succeeds if  $x : q \in \Delta$  and  $Succ^S(\delta, x)$ .
- (implication) from  $\Delta, \delta \vdash^? x : C \rightarrow G$  step to

$$\Delta \cup \{y : C\}, \delta \cup \{y\} \vdash^? y : G,$$

where  $y > \max(Lab(\Delta))$ , (whence  $y \notin Lab(\Delta)$ );

- (reduction) from

$$\Delta, \delta \vdash^? x : q,$$

if there is some  $z : C \in \Delta$ , with  $C = A_1 \rightarrow \dots \rightarrow A_k \rightarrow q$ , and there are  $\delta_i$ , and  $x_i$  for  $i = 0, \dots, k$  such that:

1.  $\delta_0 = \{z\}$ ,  $x_0 = z$ ,

---

<sup>17</sup>This restriction will be lifted in Section 5.3 where conjunction is introduced in the language.



2.  $\bigcup_{i=0}^k \delta_i = \delta$ ,
3.  $Red^S(\delta_0, \dots, \delta_k, x_0, \dots, x_k; x)$

then for  $i = 1, \dots, k$ , we step to

$$\Delta, \delta_i, \vdash^? x_i : A_i.$$

The conditions for success are either (s1) or (s2) according to each system:

$$(s1) \text{ Succ}^S(\delta, x) \equiv x \in \delta,$$

$$(s2) \text{ Succ}^S(\delta, x) \equiv \delta = \{x\}.$$

The conditions  $Red^S$  are obtained as combination of the following clauses:

$$(r0) \ x_k = x;$$

$$(r1) \ \text{for } i, j = 0, \dots, k, \delta_i \cap \delta_j = \emptyset;$$

$$(r2) \ \text{for } i = 1, \dots, k, x_{i-1} \leq x_i \text{ and } \max(\delta_i) \leq x_i;$$

$$(r3) \ \text{for } i = 1, \dots, k, x_{i-1} \leq x_i \text{ and } \max(\delta_i) = x_i;$$

$$(r4) \ \text{for } i = 1, \dots, k, x_{i-1} < x_i, \max(\delta_{i-1}) = x_{i-1} < \min(\delta_i) \text{ and } \max(\delta_k) = x_k.$$

The conditions  $Red^S$  are then defined according to Table 5.

Table 5. Restrictions on reduction and success.

Condition	(r0)	(r1)	(r2)	(r3)	(r4)	(Success)
<b>FL</b>	*				*	(s2)
<b>T-W</b>	*	*		*		(s2)
<b>T</b>	*			*		(s2)
<b>E-W</b>	*	*	*			(s2)
<b>E</b>	*		*			(s2)
<b>L</b>		*				(s2)
<b>R</b>						(s2)
<b>BCK</b>		*				(s1)

Notice that

$$(r4) \Rightarrow (r3) \Rightarrow (r2), \text{ and}$$

$$(r4) \Rightarrow (r1).$$

We give a quick explanation of the conditions  $Succ^S$  and  $Red^S$ . We recall that the component  $\delta$  represents the set of available resources which must/can be used in a derivation.

For the success rule, in all cases but **BCK**, we have that we can succeed if  $x : q$  is in the database,  $x$  is the only resource left, and  $q$  is asked from position  $x$ ; in the case of **BCK**  $x$  must be among the available resources, but we do not require that  $x$  is the only one left.

The conditions for the reduction rule can be explained intuitively as follows: resources  $\delta$  are split in several  $\delta_i$ , for  $i = 1, \dots, k$  and each part  $\delta_i$  must be used in a derivation of a subgoal  $A_i$ .

In the case of logics *without contraction* we cannot use a resource twice, therefore by restriction (r1), the  $\delta_i$ s must be disjointed and  $z$ , the label of the formula we are using in the reduction step, is no longer available.

Restriction (r2) imposes that successive subgoals are to be proved from successive positions in the database: only positions  $y \geq x$  are ‘accessible’ from  $x$ ; moreover each  $x_i$  must be accessible from resources in  $\delta_i$ . Notice that the last subgoal  $A_k$  must be proved from  $x$ , the position from which the atomic goal  $q$  is asked.

Restriction (r3) is similar to (r4), but it further requires that the position  $x_i$  is among the available resources  $\delta_i$ .

Restriction (r4) forces the goal  $A_i$  to be proved by using successive disjointed *segments*  $\delta_i$  of  $\delta$ . Moreover,  $z$  which labels the formula used in the reduction step must be the first (or least) resource among the available ones.

It is not difficult to see that intuitionistic (implicational) logic is obtained by considering success condition (s1) and no other constraint. More interestingly, we can see that **S4**-strict implication is given by considering success condition (s1) and restrictions (r0) and (r2) on reduction. We leave the reader to check that the above formulation coincides with the database-as-list formulation of **S4** we have seen in the previous section. We can therefore consider **S4** as a substructural logic obtained by imposing a restriction on the weakening and the exchange rules. On the other hand, the relation between **S4** and **E** should be apparent: the only difference is the condition on the success rule which controls the weakening restriction.

We can prove that each system is complete with respect to the its axiomatization by a syntactic proof. To this aim, we need to show that every axiom/rule is derivable, and the sets of derivable formulas is closed under substitution and Modus Ponens. The former property is proved by induction on the length of a derivation. The latter property is as usual a straightforward consequence of cut admissibility. This property is proved similarly to Theorem 10, although the details of the proof are more complex, because of the various restrictions on the reduction rule (see [Gabbay and Olivetti, 2000] pages 181–191, Theorem 5.19).

PROPOSITION 81 (Substitution). *If  $Q = \Gamma, \gamma \vdash^? x : A$  succeeds, then also  $Q' = \Gamma[q/B], \gamma \vdash^? x : A[q/B]$  succeeds.*

PROPOSITION 82 (Modus Ponens). *If  $\vdash^? \mathbf{v}_0 : A \rightarrow B$  and  $\vdash^? \mathbf{v}_0 : A$  succeed then also  $\vdash^? \mathbf{v}_0 : B$  succeeds.*

PROPOSITION 83 (Identity). *If  $x : A \in \Gamma$  and  $\text{Succ}^{\mathbf{S}}(\gamma, x)$  then  $\Gamma, \gamma \vdash^? x : A$  succeeds.*

THEOREM 84 (Completeness). *For every system  $\mathbf{S}$ , if  $A$  is a theorem of  $\mathbf{S}$ , then  $\vdash \mathbf{v}_0 : A$  succeeds in the corresponding proof system for  $\mathbf{S}$ .*

**Proof.** By Propositions 81, 82, we only need to show a derivation of an arbitrary atomic instance of each axiom in the relative proof system. In the case of reduction, the condition  $\gamma = \bigcup \gamma_i$ , will not be explicitly shown, as its truth will be apparent by the choice of  $\gamma_i$ . We assume that the truth of the condition for the success rule is evident and we do not mention it. At each step we only show the current goal, the available resources and the *new* data introduced in the database, if any. Moreover, we justify the queries obtained by a reduction step by writing the relation  $\text{Red}^{\mathbf{S}}(\gamma_0, \dots, \gamma_n, x_0, \dots, x_n; x)$  (for suitable  $\gamma_i, x_i$ ) under them; the database formula used in the reduction step is identified by  $\gamma_0$ .

(id) In all systems:

$$\vdash^? \mathbf{v}_0 : a \rightarrow a$$

we step to

$$u : a, \{u\} \vdash^? u : a,$$

which immediately succeeds in all systems.

(h1) In all systems:

$$\vdash^? \mathbf{v}_0 : (b \rightarrow c) \rightarrow (a \rightarrow b) \rightarrow a \rightarrow c.$$

three steps of the implication rule leads to:

$$\begin{array}{r} x_1 : b \rightarrow c, x_2 : a \rightarrow b, x_3 : a, \{x_1, x_2, x_3\} \vdash^? x_3 : c \\ \{x_2, x_3\} \vdash^? x_3 : b \\ \text{Red}^{\mathbf{S}}(\{x_1\}, \{x_2, x_3\}, x_1, x_3; x_3) \\ \{x_3\} \vdash^? x_3 : a \\ \text{Red}^{\mathbf{S}}(\{x_2\}, \{x_3\}, x_2, x_3; x_3). \end{array}$$

(h2) In all systems, but **FL**:

$$\vdash^? \mathbf{v}_0 : (a \rightarrow b) \rightarrow (b \rightarrow c) \rightarrow a \rightarrow c.$$

three steps of the implication rule leads to:

$$\begin{array}{l} x_1 : a \rightarrow b, x_2 : b \rightarrow c, x_3 : a, \{x_1, x_2, x_3\} \vdash^? x_3 : c \\ \quad (*) \{x_1, x_3\} \vdash^? x_3 : b, \\ \quad \text{Red}^{\mathbf{S}}(\{x_2\}, \{x_1, x_3\}, x_2, x_3; x_3) \\ \quad \quad \{x_3\} \vdash^? x_3 : a, \\ \quad \text{Red}^{\mathbf{S}}(\{x_1\}, \{x_3\}, x_1, x_3; x_3). \end{array}$$

the step (\*) is allowed in all systems, but those with (r4), namely **FL**.

(h3) In all systems, but those with (r1) or (r4):

$$\vdash^? \mathbf{v}_0 : (a \rightarrow a \rightarrow b) \rightarrow a \rightarrow b.$$

Two steps of the implication rule leads to:

$$x_1 : a \rightarrow a \rightarrow b, x_2 : a, \{x_1, x_2\} \vdash^? x_2 : b.$$

By reduction we step to:

$$\{x_2\} \vdash^? x_2 : a \quad \text{and} \quad \{x_2\} \vdash^? x_2 : a$$

since  $\text{Red}^{\mathbf{S}}(\{x_1\}, \{x_2\}, \{x_2\}, x_1, x_2, x_2; x_2)$  holds in all systems without (r1) and (r4).

(h4) In all systems, but those with (r3) or (r4):

$$\vdash^? \mathbf{v}_0 : (a \rightarrow b) \rightarrow ((a \rightarrow b) \rightarrow c) \rightarrow c.$$

two steps of the implication rule leads to:

$$\begin{array}{l} x_1 : a \rightarrow b, x_2 : (a \rightarrow b) \rightarrow c, \{x_1, x_2\} \vdash^? x_2 : c \\ \quad (*) \{x_1\} \vdash^? x_2 : a \rightarrow b \\ \quad \text{Red}^{\mathbf{S}}(\{x_2\}, \{x_1\}, x_2, x_2; x_2) \\ \quad \quad x_3 : a, \{x_1, x_3\} \vdash^? x_3 : b \\ \quad \quad \quad \{x_3\} \vdash^? x_3 : a \\ \quad \text{Red}^{\mathbf{S}}(\{x_1\}, \{x_3\}, x_1, x_3; x_3) \end{array}$$

The step (\*) is allowed by (r2), but not by (r3) or (r4) since  $\max(\{x_1\}) = x_1 < x_2$ .

(h5) In **L,R,BCK**:

$$\vdash^? \mathbf{v}_0 : a \rightarrow (a \rightarrow b) \rightarrow b.$$

two steps of implication rule leads to:

$$\begin{array}{l} x_1 : a, x_2 : a \rightarrow b, \{x_1, x_2\} \vdash^? x_2 : b \\ \{x_1\} \vdash^? x_1 : a \\ Red^S(\{x_2\}, \{x_1\}, x_2, x_1; x_2). \end{array}$$

(h6) In **BCK** we have:

$$\vdash^? \mathbf{v}_0 : a \rightarrow b \rightarrow b.$$

two steps by implication rule leads to:

$$x_1 : a, x_2 : b, \{x_1, x_2\} \vdash^? x_2 : b$$

which succeeds by the success condition of **BCK**. This formula does not succeed in any other system.

(*Suff*) We prove the admissibility of (*Suff*) rule in **FL**. Let  $\vdash^? \mathbf{v}_0 : A \rightarrow B$  succeed. Then for any formula  $C$ , we have to show that

$$\vdash^? \mathbf{v}_0 : (B \rightarrow C) \rightarrow A \rightarrow C \text{ succeeds.}$$

Let  $C = C_1 \rightarrow \dots \rightarrow C_n \rightarrow q$ . Starting from

$$\vdash^? \mathbf{v}_0 : (B \rightarrow C_1 \rightarrow \dots \rightarrow C_n \rightarrow q) \rightarrow A \rightarrow C_1 \rightarrow \dots \rightarrow C_n \rightarrow q$$

by the implication rule, we step to  $\Delta, \{x_1, \dots, x_{n+2}\} \vdash x_{n+2} : q$ , where

$$\begin{array}{l} \Delta = \{x_1 : B \rightarrow C_1 \rightarrow \dots \rightarrow C_n \rightarrow q, x_2 : A, \\ x_3 : C_1, \dots, x_{n+2} : C_n\}. \end{array}$$

From the above query we step by reduction to:

$$\begin{array}{l} Q' = \Delta, \{x_2\} \vdash^? x_2 : B \text{ and} \\ Q_i = \Delta, \{x_{i+2}\} \vdash^? x_{i+2} : C_i \text{ for } i = 1, \dots, n. \end{array}$$

since the conditions for reduction are satisfied. By hypothesis,  $\vdash^? \mathbf{v}_0 : A \rightarrow B$  succeeds, which implies, by the implicational rule, that  $x_2 : A, \{x_2\} \vdash^? x_2 : B$  succeeds, but then  $Q'$  succeeds by monotony. Queries  $Q_i$  succeed by Proposition 83.  $\blacksquare$

We can prove the soundness semantically. To this purpose we need to interpret databases and queries in the semantics. As usual, we introduce the notion of realization of a database and then of validity of a query.

**DEFINITION 85 (Realization).** Given a database  $\Gamma$ , and a set of labels  $\gamma$ , an **S**-realization of  $(\Gamma, \gamma)$  in an **S**-structure  $M = (W, \circ, \leq, 0, V)$ , is a mapping  $\rho : \mathcal{A} \rightarrow W$  such that:

1.  $\rho(\mathbf{v}_0) = 0$ ;
2. if  $y : B \in \Gamma$  then  $M, \rho(y) \models B$ .

In order to define the notion of validity of a query, we need to introduce some further notation. Given an **S**-realization  $\rho$ ,  $\gamma$  and  $x$ , we define

$$\begin{aligned} \rho(\gamma) &= 0 \text{ if } \gamma = \emptyset, \\ \rho(\gamma) &= \rho(x_1) \circ \dots \circ \rho(x_n) \text{ if } \gamma = \{x_1, \dots, x_n\}, \text{ where } x_1 < \dots < x_n \\ \rho(\langle \gamma, x \rangle) &= \rho(\gamma) \text{ if } x \in \gamma, \\ \rho(\langle \gamma, x \rangle) &= \rho(\gamma) \circ 0 \text{ if } x \notin \gamma. \end{aligned}$$

**DEFINITION 86 (Valid query).** Let  $Q = \Gamma, \gamma \vdash^? x : A$ , we say that  $Q$  is **S**-valid if for every **S**-structure  $M$ , for every realization  $\rho$  of  $\Gamma$  in  $M$ , we have

$$M, \rho(\langle \gamma, x \rangle) \models A.$$

According to the definition above, the **S**-validity of the query  $\vdash^? \mathbf{v}_0 : A$  means that the formula  $A$  is **S**-valid (i.e.  $\models_{\mathbf{S}}^{Fine} A$ ).

**THEOREM 87.** *If  $Q = \Gamma, \gamma \vdash^? x : A$  succeeds in the proof system for **S** then  $Q$  is **S**-valid. In particular, if  $\mathbf{v}_0 : A$  succeeds in the proof system for **S**, then  $\models_{\mathbf{S}}^{Fine} A$ .*

The proof can be done by induction on the length of derivations, by suitably relating the constraints of the reduction rule to the algebraic semantic conditions.

### 5.3 Extending the language

In this section we show how we can extend the language by some other connectives. We allow extensional conjunction ( $\wedge$ ), disjunction ( $\vee$ ), and intensional conjunction or *tensor* ( $\otimes$ ). The distinction between  $\wedge$  and  $\otimes$  is typical of substructural logics and it comes from the rejection of some structural rule:  $\wedge$  is the usual lattice-inf connective,  $\otimes$  is close to a residual operator with respect to  $\rightarrow$ . In relevant logic literature  $\otimes$  is often called

*fusion* or *cotenableity* and denoted by  $\circ$ .<sup>18</sup> The addition of the above connectives presents some semantic options. The most important one is whether *distribution* (dist) of  $\wedge$  and  $\vee$  is assumed or not. The list of axioms/rules below characterizes distributive substructural logics.

DEFINITION 88 (Axioms for  $\wedge, \otimes, \vee$ ).

1.  $A \wedge B \rightarrow A$ ,
2.  $A \wedge B \rightarrow B$ ,
3.  $(C \rightarrow A) \wedge (C \rightarrow B) \rightarrow (C \rightarrow A \wedge B)$ ,
4.  $A \rightarrow A \vee B$ ,
5.  $B \rightarrow A \vee B$ ,
6.  $(A \rightarrow C) \wedge (B \rightarrow C) \rightarrow (A \vee B \rightarrow C)$
7.  $\frac{A \quad B}{A \wedge B}$
8.  $\frac{A \rightarrow B \rightarrow C}{A \otimes B \rightarrow C}$
9.  $\frac{A \otimes B \rightarrow C}{A \rightarrow B \rightarrow C}$

(e- $\wedge$ ) For **E** and **E-W** only

$$\Box A \wedge \Box B \rightarrow \Box(A \wedge B)$$

where  $\Box C =_{def} (C \rightarrow C) \rightarrow C$ .

(dist)  $A \wedge (B \vee C) \rightarrow (A \wedge B) \vee C$ .

As we have said, the addition of distribution (dist) is a semantic choice, which may be argued. However, for the fragment of the language we consider in this section it does not really matter whether distribution is assumed or not. This fragment roughly corresponds to the Harrop fragment of the section of modal logics (see Section 4.3); since we do not allow positive occurrences of disjunction, the presence of distribution is immaterial.<sup>19</sup> We have included distribution to have a complete axiomatization with respect to the semantics we adopt.

<sup>18</sup>We follow here the terminology and notation of linear logic [Girard, 1987].

<sup>19</sup>In the fragment we consider we trivially have (for any **S**)  $\Gamma, \alpha \vdash^? x : A \wedge (B \vee C)$  implies  $\Gamma, \alpha \vdash^? x : A \vee (B \wedge C)$ , where  $\Gamma$  is a set of D-formulas and  $A, B, C$  are G-formulas (see below).

As in the case of modal logics (see Section 4.3), we distinguish D-formulas, which are the constituents of databases, and G-formulas which can be asked as goals.

DEFINITION 89. Let D-formulas and G-formulas be defined as follows:

$$\begin{aligned} D &:= q \mid G \rightarrow D, \\ CD &:= D \mid CD \wedge CD, \\ G &:= q \mid G \wedge G \mid G \vee G \mid G \otimes G \mid CD \rightarrow G. \end{aligned}$$

A database  $\Delta$  is a finite set of *labelled* D-formulas.

A database corresponds to a  $\otimes$ -composition of conjunctions of D-formulas. Formulas with the same label are thought as  $\wedge$ -conjoined. Every D-formula has the form

$$G_1 \rightarrow \dots \rightarrow G_k \rightarrow q,$$

In the systems **R**, **L**, **BCK**, we have the theorems

$$(A \rightarrow B \rightarrow C) \rightarrow (A \otimes B \rightarrow C) \quad \text{and} \quad (A \otimes B \rightarrow C) \rightarrow (A \rightarrow B \rightarrow C)$$

Thus, in these systems we can simplify the syntax of (non atomic) D-formulas to  $G \rightarrow q$  rather than  $G \rightarrow D$ . This simplification is not allowed in the other systems where we only have the weaker relation

$$\vdash (A \rightarrow B \rightarrow C) \Leftrightarrow \vdash A \otimes B \rightarrow C.$$

The extent of Definition 89 is shown by the following proposition.

PROPOSITION 90. *Every formula on  $(\wedge, \vee, \rightarrow, \otimes)$  without*

- *positive<sup>20</sup> (negative) occurrences of  $\otimes$  and  $\vee$  and*
- *occurrences of  $\otimes$  within a negative (positive) occurrence of  $\wedge$*

*is equivalent to a  $\wedge$ -conjunction of D-formulas (G-formulas).*

The reason we have put the restriction on nested occurrences of  $\otimes$  within  $\wedge$  is that, on the one hand, we want to keep the simple labelling mechanism we have used for the implicational fragment, and on the other we want to identify a common fragment for all systems to which the computation rules are easily extended. The labelling mechanism no longer works if we relax

<sup>20</sup>Positive and negative occurrences are defined as follows:  $A$  occurs positively in  $A$ ; if  $B\#C$  occurs positively (negatively) in  $A$  (where  $\# \in \{\wedge, \vee, \otimes\}$ ), then  $B$  and  $C$  occur positively (negatively) in  $A$ ; if  $B \rightarrow C$  occurs positively (negatively) in  $A$ , then  $B$  occurs negatively (positively) in  $A$  and  $C$  occurs positively (negatively) in  $A$ . We say that a connective  $\#$  has a positive (negative) occurrence in a formula  $A$  if there is a formula  $B\#C$  which occurs positively (negatively) in  $A$ .



this restriction. For instance, how could we handle  $A \wedge (B \otimes C)$  as a D-formula? We should add  $x : A$  and  $x : B \otimes C$  in the database. The formula  $x : B \otimes C$  cannot be decomposed, unless we use complex labels: intuitively we should split  $x$  into some  $y$  and  $z$ , add  $y : B$  and  $z : C$ , and remember that  $x, y, z$  are connected (in terms of Fine semantics the connection would be expressed as  $x = y \circ z$ ).<sup>21</sup>

The computation rules can be extended to this fragment without great effort.

**DEFINITION 91** (Proof system for the extended language). We give the rules for queries of the form

$$\Delta, \delta \vdash^? x : G,$$

where  $\Delta$  is a set of D-formulas and  $G$  is a G-formula.

- (success)  $\Delta, \delta \vdash^? x : q$  succeeds if  $x; q \in \Delta$  and  $Succ^S(\delta, x)$ .
- (implication) from  $\Delta, \delta \vdash^? x : CD \rightarrow G$   
if  $CD = D_1 \wedge \dots \wedge D_n$ , we step to

$$\Delta \cup \{y : D_1, \dots, y : D_n\}, \delta \cup \{y\} \vdash^? y : G$$

where  $y > \max(Lab(\Delta))$ , (hence  $y \notin Lab(\Delta)$ ).

- (reduction) from  $\Delta, \delta \vdash^? x : q$   
if there is  $z : G_1 \rightarrow \dots \rightarrow G_k \rightarrow q \in \Delta$  and there are  $\delta_i$ , and  $x_i$  for  $i = 0, \dots, k$  such that:

1.  $\delta_0 = \{z\}$ ,  $x_0 = z$ ,
2.  $\bigcup_{i=0}^k \delta_i = \delta$ ,
3.  $Red^S(\delta_0, \dots, \delta_k, x_0, \dots, x_k; x)$ ,

then for  $i = 1, \dots, k$ , we step to

$$\Delta, \delta_i, \vdash^? x_i : G_i.$$

- (conjunction) from  $\Delta, \delta \vdash^? x : G_1 \wedge G_2$  step to

$$\Delta, \delta \vdash^? x : G_1 \text{ and } \Delta, \delta \vdash^? x : G_2.$$

- (disjunction) from  $\Delta, \delta \vdash^? x : G_1 \vee G_2$  step to

$$\Delta, \delta \vdash^? x : G_i \text{ for } i = 1 \text{ or } i = 2.$$

---

<sup>21</sup>In some logics, such as **L**, we do not need this restriction since we have the following property:  $\Gamma, A \wedge B \vdash C$  implies  $\Gamma, A \vdash C$  or  $\Gamma, B \vdash C$ . Thus, we can avoid introducing extensional conjunctions into the database, and instead introduce only one of the two conjuncts (at choice). This approach is followed by Harland and Pym [1991]. However the above property does not hold for **R** and other logics.

- (tensor) from  $\Delta, \delta \vdash^? x : G_1 \otimes G_2$   
if there are  $\delta_1, \delta_2, x_1$  and  $x_2$  such that
  1.  $\delta = \delta_1 \cup \delta_2$ ,
  2.  $Red^S(\delta_1, \delta_2, x_1, x_2; x)$ ,

step to

$$\Delta, \delta_1 \vdash^? x_1 : G_1 \text{ and } \Delta, \delta_2 \vdash^? x_2 : G_2.$$

An easy extension of the method is the addition of the truth constants  $\mathbf{t}$ , and  $\top$  which are governed by the following axioms/rules

$$\begin{aligned} A &\rightarrow \top, \\ \vdash \mathbf{t} &\rightarrow A \text{ iff } \vdash A. \end{aligned}$$

Plus the axiom of  $(\mathbf{t} \rightarrow A) \rightarrow A$  for **E** and **E-W**. We can think of  $\mathbf{t}$  as defined by propositional quantification

$$\mathbf{t} =_{def} \forall p(p \rightarrow p).$$

Equivalently, given any formula  $A$ , we can assume that  $\mathbf{t}$  is the conjunction of all  $p \rightarrow p$  such that the atom  $p$  occurs in  $A$ . Basing on this definition, it is not difficult to handle  $\mathbf{t}$  in the goal-directed way and we leave to the reader to work out the rules. The treatment of  $\top$  is straightforward.

EXAMPLE 92. Let  $\Delta$  be the following database:

$$\begin{aligned} x_1 &: e \wedge g \rightarrow d, \\ x_2 &: (c \rightarrow d) \otimes (a \vee b) \rightarrow p, \\ x_3 &: c \rightarrow e, \\ x_3 &: c \rightarrow g, \\ x_4 &: (c \rightarrow g) \rightarrow b. \end{aligned}$$

In Figure 9, we show a successful derivation of

$$\Delta, \{x_1, x_2, x_3, x_4\} \vdash^? x_4 : p$$

in relevant logic **E** (and stronger systems). We leave to the reader to justify the steps according to the rules. The success of this query corresponds to the validity of the following formula in **E**:

$$\begin{aligned} &[(e \wedge g \rightarrow d) \otimes ((c \rightarrow d) \otimes (a \vee b) \rightarrow p) \otimes ((c \rightarrow e) \wedge \\ &\quad \wedge (c \rightarrow g)) \otimes ((c \rightarrow g) \rightarrow b)] \rightarrow p. \end{aligned}$$

We can extend the soundness and completeness result to this larger fragment. We first extend Definition 77 by giving the truth conditions for the additional connectives.

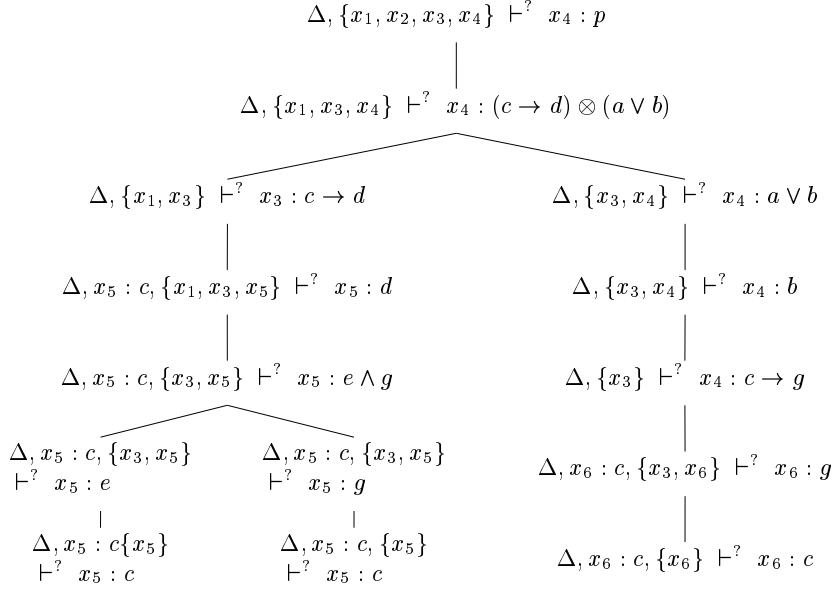


Figure 9. Derivation for Example 92.

**DEFINITION 93.** Let  $M = (W, \circ, 0, V)$  be a **S**-structure, let  $a \in W$  we stipulate:

$$M, a \models A \wedge B \text{ iff } M, a \models A \text{ and } M, a \models B,$$

$$M, a \models A \vee B \text{ iff } M, a \models A \text{ or } M, a \models B,$$

$$M, a \models A \otimes B \text{ iff there are } b, c \in W, \text{ s.t. } b \circ c \leq a \text{ and } M, a \models A \text{ and } M, a \models B.$$

It is straightforward to extend Theorem 87 obtaining the soundness of the proof procedure with respect to this semantics. The completeness can be proved by a canonical model construction. The details are given in [Gabbay and Olivetti, 2000], (see Section 6.1). Putting the two results together we obtain:

**THEOREM 94.** Let  $\Gamma, \gamma \vdash^? x : G$  be a query with  $\gamma = \{x_1, \dots, x_k\}$  (ordered as shown), and let  $S_i = \{A \mid x_i : A \in \Gamma\}$ ,  $i = 1, \dots, k$ . The following are equivalent:

1.  $\models_{\mathbf{S}}^{Fine} (\bigwedge S_1 \otimes \dots \otimes \bigwedge S_k) \rightarrow G$
2.  $\Gamma, \gamma \vdash^? x : G$  succeeds in the system **S**.

#### 5.4 Eliminating the labels

As in the case of modal logics, it turns out that for most of the systems (at least if we consider the implicational fragment), labels are not needed.

PROPOSITION 95.

- If  $\Gamma, \gamma \vdash^? x : B$  succeeds,  $u : A \in \Gamma$  but  $u \notin \gamma$ , then  $\Gamma - \{u : A\}, \gamma \vdash^? x : B$  succeeds.
- For **R**, **L**, **BCK**, if  $\Gamma, \gamma \vdash^? x : B$  succeeds, then for every  $y \in \gamma$ ,  $\Gamma, \gamma \vdash^? y : B$  succeeds.
- For **T**, **T-W**, **FL**, if  $\Gamma, \gamma \vdash^? x : B$  succeeds then it must be  $x = \max(\gamma)$ .

Because of the previous proposition, the formulas which can contribute to the proof are those that are listed in  $\gamma$ . Since every copy of a formula gets a different label, labels and ‘usable’ formulas are into a 1-1 correspondence. Moreover in all cases, but **E** and **E-W**, the label  $x$  in front of the goal is either irrelevant or it is determined by  $\gamma$ . Putting these facts together, we can reformulate the proof systems for all logics, but **E** and **E-W** without using labels. The restrictions on reduction can be expressed directly in terms of the sequence of database formulas. In other words, formulas and labels become the same things. As an example we give the reformulation of **L** and **FL**. In case of **L**, it is easily seen that the order of formulas does not matter (it can be proved that permuting the database does not affect the success of a query), thus the database can be thought as a multiset of formulas, and the rules become as follows:

1. (success)  $\Delta \vdash^? q$  succeeds if  $\Delta = q$ ,
2. (implication) from  $\Delta \vdash^? A \rightarrow B$  step to  $\Delta, A \vdash^? B$ ,
3. (reduction) from  $\Delta, C \rightarrow \dots \rightarrow C_n \rightarrow q \vdash^? q$  step to  $\Delta_i \vdash^? C_i$ , where  $\Delta = \sqcup_i \Delta_i$ , ( $\sqcup$  denotes multiset union).

In other words we obtain what we have called the *linear computation* in Section 3.3, Definition 13.

In case of **FL**, the order of the formulas is significant. Thus the rules for success and implication are the same, but in case of implication the formula  $A$  is added to the *end* of  $\Delta$ . The rule of reduction becomes:

(**FL**-reduction) from  $C \rightarrow \dots \rightarrow C_n \rightarrow q, \Delta \vdash^? q$  step to  $\Delta_i \vdash^? C_i$ , where  $\Delta = \Delta_1, \dots, \Delta_n$ ,

in this case “,” denotes concatenation. The reformulation without labels for **R**, **T**, and **T-W**, **BCK** is similar and left to the reader.

### 5.5 Some history

The use of labels to deal with substructural logics is not a novelty: it has been used for instance by Prawitz [1965], by Anderson and Belnap [1975], to develop a natural-deduction formulation of most relevant logics. The goal-directed proof systems we have presented are similar to their natural deduction systems in this respect: there are not explicit structural rules. The structural rules are *internalized* as restrictions in the logical rules.

Bollen has developed a goal-directed procedure for a fragment of first-order  $\mathbf{R}$  [Bollen, 1991]). His method avoids splitting derivations in several branches by maintaining a global proof-state  $\Delta \vdash^? [A_1, \dots, A_n]$ , where all  $A_1, \dots, A_n$  have to be proved (they can be thought as linked by  $\otimes$ ). If we want to keep all subgoals together, we must take care that different subgoals  $A_i$  may happen to be evaluated in different contexts. For instance in

$$C \vdash^? D \rightarrow E, F \rightarrow G$$

according to the implication rule, we must evaluate  $E$  from  $\{C, D\}$ , and  $G$  from  $\{C, F\}$ . Bollen accommodates this kind of context-dependency by indexing subgoals with a number which refers to the part of the database that can be used to prove it. Furthermore, a list of numbers is maintained to remember the usage; in a successful derivation the usage list must contain the numbers of all formulas of the database.

Many people have developed goal-directed procedures for fragments of linear logic leading to the definition of logic programming languages based on linear logic. This follows the tradition of the *uniform proof paradigm* proposed by Miller [Miller *et al.*, 1991]. We notice, in passing, that implicational  $\mathbf{R}$  can be encoded in linear logic by defining  $A \rightarrow B$  by  $A \multimap !A \multimap B$ , where  $!$  is the exponential operator which enables the contraction of the formula to which is applied. The various proposals differ in the choice of the language fragment. Much emphasis is given to the treatment of the exponential  $!$ , as it is needed for defining logic programming languages of some utility: in most applications, we need permanent resources or data, (i.e. data that are not ‘consumed’ along a deduction); permanent data can be represented by using the  $!$  operator.

Some proposals, such as [Harland and Pym, 1991] and [Andreoli and Pareschi, 1991; Andreoli, 1992] take as basis the multi-consequent (or classical) version of linear logic. Moreover, mixed systems have been studied [Hodas and Miller, 1994] and [Hodas, 1993] which combine different logics into a single language: linear implication, intuitionistic implication and more.<sup>22</sup>

---

<sup>22</sup>In [Hodas, 1993], Hodas has proposed a language called  $\mathbf{O}$  which combines intuitionistic, linear, affine and relevant implication. The idea is to partition the ‘context’, i.e. the database in several (multi) sets of data corresponding to the different handling of the data according to each implicational logic.

O’Hearn and Pym [1999] have recently proposed an interesting development of linear logic called the *logic of bunched implications*. Their system has strong semantical and categorical motivations. The logic of bunched implications combines a multiplicative implication, namely the one of linear logic and an additive (or intuitionistic) one. The proof contexts, that is the antecedents of a sequent, are structures built by two operators: ‘;’ corresponds to the multiplicative conjunction  $\otimes$  and ‘,’ corresponds to the additive conjunction  $\wedge$ . These structures are called *bunches*. Similar structures have been used to obtain calculi for distributive relevant logics [Dunn, 1986]. The authors define also an interesting extension to the first-order case by introducing intensional quantifiers. Moreover they develop a goal-directed proof procedure for a Harrop-fragment of this logic and show significant applications to logic programming.

In [Gabbay and Olivetti, 2000], the goal-directed systems are extended to **RM0** where a suitable restart rule takes care of the mingle rule. Moreover, it is shown how the goal-directed method for **R** can be turned into a decision procedure (for the pure implicational part) by adding a suitable loop-checking mechanism. We notice that for contractionless logics, the goal-directed proof-methods of this section can be the base of decision procedures.

To implement goal-directed proof systems for substructural logics, one can import solutions and techniques developed in the context of linear logic programming. For instance, in the reduction and in the  $\otimes$ -rule we have to guess a portion of the goal label. A similar problem has been discussed in the linear logic programming community [Hodas and Miller, 1994; Harland and Pym, 1997] where one has to guess a split of the sequent (only the antecedent in the intuitionistic version, both the antecedent and consequent in the ‘classical’ version). A number of solutions have been provided, most of them based on a *lazy* computation of the split parts. Perhaps the most general way to handle this problem has been addressed in [Harland and Pym, 1997] where it is proposed to represent the sequent split by means of Boolean constraints expressed by labels attached to the formulas. Different strategies of searching the partitions correspond to different strategies of solving the constraints (lazy, eager and mixed).

## 6 DEVELOPMENTS, APPLICATIONS AND MECHANISMS

The goal-directed proof methods presented in this chapter may be useful for several purposes, beyond the mere deductive task. The deductive procedures can be easily extended in order to compute further information, or to support other logical tasks, such as abduction. We show two examples: the computation of interpolants in implicational linear logic, and the computation of abductive explanations in intuitionistic logic. Both of them are

simple extensions of the goal-directed procedures that we have seen in the previous sections. We conclude by a discussion about the extension of the goal-directed methods to the first order case.

### 6.1 Interpolation for linear implication

By interpolation we mean the following property. Let  $Q_1$  and  $Q_2$  be two sets of atoms; let  $\mathcal{L}_1$  be the language generated by  $Q_1$  and  $\mathcal{L}_2$  by  $Q_2$ ; finally let  $A, B$  be formulas with  $A \in \mathcal{L}_1$  and  $B \in \mathcal{L}_2$ . If  $A \vdash B$ , then there is a formula  $C \in \mathcal{L}_1 \cap \mathcal{L}_2$ , such that  $A \vdash C$  and  $C \vdash B$ . The formula  $C$  is called an *interpolant* of  $A$  and  $B$ . Here, we consider the interpolation property for implicational linear logic, i.e.  $\vdash$  denotes provability in (implicational) linear logic, and  $A, B$  contain only implication. We show how the procedure of the previous section can be modified to compute an interpolant. Since the computation is defined for sequents of the form  $\Gamma \vdash A$ , where  $\Gamma$  is a database (a multiset of formulas in this case), we need to generalize the notion of interpolation to provability statements of this form. As a particular case we will have the usual interpolation property for pair of formulas. To generalize interpolation to database we need first to generalize the provability relation to databases. Let  $\Delta = A_1, \dots, A_n$ , we define:

$$\begin{aligned} \Gamma \vdash \Delta \text{ iff there are } \Delta_1, \dots, \Delta_n, \text{ such that } \Delta &= \Delta_1 \sqcup \dots \sqcup \Delta_n \text{ and} \\ \Delta_i \vdash A_i \text{ for } i &= 1, \dots, n. \end{aligned}$$

The formulas  $A_i$  can be thought as conjuncted by  $\otimes$ . Moreover, since  $A_i \rightarrow \dots \rightarrow A_n \rightarrow B$ , is equivalent to  $A_1 \otimes \dots \otimes A_n \rightarrow B$ , and  $\otimes$  is associative and commutative, we abbreviate the above formula with  $\Delta \rightarrow B$ , where  $\Delta$  is the multiset  $A_1, \dots, A_n$ . Let  $\Gamma \in \mathcal{L}_1$  and  $A \in \mathcal{L}_2$ , suppose that  $\Gamma \vdash A$  holds; an *interpolant* for  $\Gamma \vdash A$  is a database  $\Pi \in \mathcal{L}_1 \cap \mathcal{L}_2$  such that  $\Gamma \vdash \Pi$  and  $\Pi \vdash A$ . Observe that if  $|\Gamma| = 1$ , it must be also  $|\Pi| = 1$ , and we have the ordinary notion of interpolant.

We give a recursive procedure to calculate an interpolant for  $\Gamma \vdash A$ . We do this by defining a recursive predicate  $Inter(\Gamma; \Delta; A; \Pi)$  which is true whenever:  $\Gamma \in \mathcal{L}_1$ ,  $\Delta, A \in \mathcal{L}_2$ ,  $\Gamma, \Delta \vdash A$  holds, and  $\Pi$  is an interpolant for  $\Gamma \vdash \Delta \rightarrow A$ .

- (Succ 1)  $Inter(q; \emptyset; q; q)$ .
- (Succ 2)  $Inter(\emptyset; q; q; \emptyset)$ .
- (Imp)  $Inter(\Gamma; \Delta; A \rightarrow B, \Pi)$  if  $Inter(\Gamma; \Delta, A; B, \Pi)$ .
- (Red 1)  $Inter(\Gamma, C_1 \rightarrow \dots C_n \rightarrow q; \Delta; q; \Pi)$ , if there are  $\Gamma_i, \Delta_i, \Pi_i$  for  $i = 1, \dots, n$  such that
  1.  $\Gamma = \sqcup_i \Gamma_i, \Delta = \sqcup_i \Delta_i,$

2.  $Inter(\Delta_i; \Gamma_i; C_i; \Pi_i)$  for  $i = 1, \dots, n$ ,
  3.  $\Pi = \Pi_1 \sqcup \dots \sqcup \Pi_n \rightarrow q$ .
- (Red 2)  $Inter(\Gamma; \Delta, C_1 \rightarrow \dots \rightarrow C_n \rightarrow q; q; \Pi)$ , if there are  $\Gamma_i, \Delta_i, \Pi_i$  for  $i = 1, \dots, n$  such that
    1.  $\Gamma = \sqcup_i \Gamma_i, \Delta = \sqcup_i \Delta_i, \Pi = \sqcup_i \Pi_i$ ,
    2.  $Inter(\Gamma_i; \Delta_i; C_i; \Pi_i)$  for  $i = 1, \dots, n$ .

If we want to find an interpolant for  $\Gamma \vdash A$ , the initial call is  $Inter(\Gamma; \emptyset; A; \Pi)$ , where  $\Pi$  is computed along the derivation. One can observe that the predicate  $Inter$  is defined by following the definition of the goal-directed proof procedure. (Succ 1) and (Succ 2) apply to the case of immediate success, respectively when the atom is in  $\mathcal{L}_1$ , and when it is in  $\mathcal{L}_2$ . Analogously, (Red 1) deals with the case the atomic goal unifies with the head of a formula in  $\mathcal{L}_1$  and (Red 2) when it unifies with the head of a formula in  $\mathcal{L}_2$ . It can be proved that if  $\Gamma \vdash A$  the algorithm computes an interpolant. Moreover all interpolants can be obtained by selecting different formulas in the reduction steps. Here is a non trivial example.

EXAMPLE 96. Let

$$A = ((f \rightarrow e) \rightarrow ((a \rightarrow b) \rightarrow c) \rightarrow (a \rightarrow q) \rightarrow (q \rightarrow b) \rightarrow f \rightarrow b) \rightarrow g,$$

and

$$B = (e \rightarrow c \rightarrow d) \rightarrow (d \rightarrow a) \rightarrow (a \rightarrow b) \rightarrow g.$$

One can check that  $A \vdash B$  holds; we compute an interpolant, by calling  $Inter(A; \emptyset; B; \Pi)$ . Let us abbreviate the antecedent of  $A$  by  $A'$ . Here are the steps, we underline the formula used in a reduction step:

- (1)  $Inter(\underline{A}; \emptyset; B; \Pi)$
- (2)  $Inter(A; e \rightarrow c \rightarrow d, d \rightarrow a, a \rightarrow b; g; \Pi)$
- (3)  $Inter(e \rightarrow c \rightarrow d, d \rightarrow a, a \rightarrow b; \emptyset; A', \Pi_1) \quad \Pi = \Pi_1 \rightarrow g$
- (4)  $Inter(e \rightarrow c \rightarrow d, d \rightarrow a, a \rightarrow b; f \rightarrow e, (a \rightarrow b) \rightarrow c, a \rightarrow q, \underline{q \rightarrow b}, f; b, \Pi_1)$
- (5)  $Inter(e \rightarrow c \rightarrow d, d \rightarrow a, a \rightarrow b; f \rightarrow e, (a \rightarrow b) \rightarrow c, \underline{a \rightarrow q}, f; q, \Pi_1)$
- (6)  $Inter(e \rightarrow c \rightarrow d, \underline{d \rightarrow a}, a \rightarrow b; f \rightarrow e, (a \rightarrow b) \rightarrow c, f; a, \Pi_1)$
- (7)  $Inter(f \rightarrow e, (a \rightarrow b) \rightarrow c, f; \underline{e \rightarrow c \rightarrow d}, a \rightarrow b; ; d, \Pi_2)$   
 $\Pi_1 = \Pi_2 \rightarrow a$



Now we have a split with  $\Pi_2 = \Pi_3, \Pi_4$  and

$$(8) \text{ Inter}(\underline{f \rightarrow e}, f; \emptyset; e; \Pi_3) \quad \text{and} \quad (9) \text{ Inter}(\underline{(a \rightarrow b) \rightarrow c}; a \rightarrow b; ; c, \Pi_4).$$

(8) gives

$$(10) \text{ Inter}(\emptyset; f; f; \Pi_5) \quad \Pi_3 = \Pi_5 \rightarrow e$$

whence  $\Pi_5 = \emptyset$ . (9) gives

$$(11) \text{ Inter}(a \rightarrow b; \emptyset; a \rightarrow b; \Pi_6) \quad \Pi_4 = \Pi_6 \rightarrow c$$

$$(12) \text{ Inter}(\underline{a \rightarrow b}; a; b; \Pi_6)$$

$$(13) \text{ Inter}(a; \emptyset; a; \Pi_7) \quad \Pi_6 = \Pi_7 \rightarrow b$$

Thus  $\Pi_7 = a$  and we have:

$$\Pi_6 = a \rightarrow b,$$

$$\Pi_3 = e,$$

$$\Pi_4 = (a \rightarrow b) \rightarrow c,$$

$$\Pi_2 = e, (a \rightarrow b) \rightarrow c,$$

$$\Pi_1 = e \rightarrow ((a \rightarrow b) \rightarrow c) \rightarrow a,$$

$$\Pi = (e \rightarrow ((a \rightarrow b) \rightarrow c) \rightarrow a) \rightarrow g.$$

$\Pi$  is evidently in the common language of  $A$  and  $B$ ; we leave to the reader to check that  $A \vdash \Pi$  and  $\Pi \vdash B$ .

It is likely that similar procedures based on the goal-directed computation can be devised for other logics admitting a goal directed presentation. However, further investigation is needed to see to what extent this approach works for other cases. For instance, the step in (Red 1) is justified by the structural *exchange* law which holds for linear logic. For non-commutative logics the suitable generalization of the interpolation property, which is the base of the inductive procedure, might be more difficult to find.

## 6.2 Abduction for intuitionistic implication

Abduction is an important kind of inference for many applications. It has been widely and deeply studied in logic programming context (see [Eshghi, 1989] for a seminal work). Abductive inference can be described as follows: given a set of data  $\Gamma$  and a formula  $A$  such that  $\Gamma \not\vdash A$ , find a set of formulas  $\Pi$  such that  $\Gamma, \Pi \vdash A$ . Usually, there are some further requirements on the possible  $\Pi$ 's, such as *minimality*. The set  $\Pi$  is called an abductive solution for  $\Gamma \vdash A$ . We define below a metapredicate  $Abduce(\Gamma; A; \Pi)$  whose meaning, is that  $\Pi$  is an abductive solution for  $\Gamma \vdash A$ . The predicate is defined by induction on the goal-directed derivation. Given a formula  $C$  and a database  $\Gamma$ , we write  $C \rightarrow \Gamma$  to denote the set  $\{C \rightarrow D \mid D \in \Gamma\}$ .

1.  $Abduce(\Gamma; q; \Pi)$  if  $q \in \Gamma$  and  $\Pi = \emptyset$ .
2.  $Abduce(\Gamma; q; \Pi)$  if  $\forall C \in \Gamma \ q \neq Head(C)$  and  $\Pi = \{q\}$ .
3.  $Abduce(\Gamma; A \rightarrow B; \Pi)$  if  $Abduce(\Gamma, A; B; \Pi')$  and  $\Pi = A \rightarrow \Pi'$ .
4.  $Abduce(\Gamma; q; \Pi)$  if there is  $C_1 \rightarrow \dots \rightarrow C_n \rightarrow q \in \Gamma$  and  $\Pi_1, \dots, \Pi_n$  such that
  - (a)  $Abduce(\Gamma; C_i, \Pi_i)$  for  $i = 1, \dots, n$ .
  - (b)  $\Pi = \bigcup_i \Pi_i$ .

It can be shown that if  $Abduce(\Gamma; q; \Pi)$  holds then  $\Gamma, \Pi \vdash A$ .

EXAMPLE 97. Let  $\Gamma = (a \rightarrow c) \rightarrow b, (d \rightarrow b) \rightarrow s \rightarrow q, (p \rightarrow q) \rightarrow t \rightarrow r$ . We compute  $Abduce(\Gamma; r; \Pi)$ . We have

- (1)  $Abduce(\Gamma; r; \Pi)$  reduces to
- (2)  $Abduce(\Gamma; p \rightarrow q; \Pi_1)$  and (3)  $Abduce(\Gamma; t; \Pi_2)$   
with  $\Pi = \Pi_1 \cup \Pi_2$ ; (3) gives  $\Pi_2 = \{t\}$ . (2) is reduced as follows
- (4)  $Abduce(\Gamma, p; q; \Pi_3)$  and  $\Pi_1 = p \rightarrow \Pi_3$
- (5)  $Abduce(\Gamma, p; d \rightarrow b; \Pi_4)$  and (6)  $Abduce(\Gamma, p; s; \Pi_5)$   
with  $\Pi_3 = \Pi_4 \cup \Pi_5$ ; (6) gives  $\Pi_5 = \{s\}$ . (5) is reduced as follows
- (7)  $Abduce(\Gamma, p, d; b; \Pi_6)$  and  $\Pi_4 = b \rightarrow \Pi_6$
- (8)  $Abduce(\Gamma, p, d; a \rightarrow c; \Pi_6)$
- (9)  $Abduce(\Gamma, p, d, a; c; \Pi_7)$  and  $\Pi_6 = a \rightarrow \Pi_7$   
(9) gives  $\Pi_7 = \{c\}$ .

Thus

$$\Pi_6 = \{a \rightarrow c\}$$

$$\Pi_4 = \{b \rightarrow a \rightarrow c\}$$

$$\Pi_3 = \{b \rightarrow a \rightarrow c, s\}$$

$$\Pi_1 = \{p \rightarrow b \rightarrow a \rightarrow c, p \rightarrow s\}$$

$$\Pi = \{p \rightarrow b \rightarrow a \rightarrow c, p \rightarrow s, t\}.$$

One can easily check that  $\Gamma, \Pi \vdash r$ .

The abductive proof procedure we have exemplified is a sort of metapredicate defined from the goal-directed computation. It can be used to generate possible abductive solutions, that might be then compared and processed further. Of course variants and refinements of the abductive procedures are possible for specific purposes.

### 6.3 First-order extension

We conclude this chapter by some remarks on the major extension of these methods: the one to the first-order level. The extension is not straightforward. In general, in most non-classical logics there are several options in the interpretation of quantifiers and terms according to the intended semantics (typically, constant domain, increasing domains etc.); moreover, one may adopt either rigid, or non-rigid interpretation of terms. Sometimes the interpretation of quantifiers is not entirely clear, as it happens in substructural logics. However, even when quantifiers are well understood as in intuitionistic and modal logics, a goal-directed treatment of the full language creates troubles analogously to the treatment of disjunction. In particular the treatment of positive occurrences of the existential quantifier is problematic. In classical logic such a problem does not arise as one can always transform the pair (Database, Goal) into a set of universal sentences using Skolem transformation. A similar method cannot be applied to most non-classical logics, where one cannot skolemize the data before the computation starts.

As a difference with the theorem proving view, in the goal-directed approach (following the line of logic programming) one would like to define proof procedures which compute answer-substitutions. This means that the outcome of a successful computation of

$$\Delta \vdash^? G[X],$$

where  $G[X]$  stands for  $\exists X G[X]$  is not only 'yes', but it is (a most general) substitution  $X/t$ , such that  $\Delta \vdash G[X/t]$  holds. In [Gabbay and Olivetti, 2000] it is presented a proof procedure for  $(\rightarrow, \forall)$  fragment of intuitionistic logic. The procedure is in the style of a logic programming and uses unification to compute answer-substitutions. The proof-procedure checks the simultaneous success of a set of pairs (Database, Goals); this structure is enforced by the presence of shared free variables occurring in different databases arising during the computation. The approach of 'Run-time skolemization' is used to handle universally quantified goals.<sup>23</sup> That is to say the process of eliminating universal quantifiers on the goal by Skolem functions is carried on in parallel with goal reduction. The 'Run-time Skolemisation' approach has been presented in [Gabbay, 1992; Gabbay and Reyle, 1993] and adopted in N-Prolog [Gabbay and Reyle, 1993], a hypothetical extension of Prolog based on intuitionistic logic. A similar idea for classical logic is embodied in the free-variable tableaux, see [Fitting, 1990] and [Hähnle and Schmitt, 1994] for an improved rule. The use of Skolem functions and normal form for intuitionistic proof-search has been studied by many authors, we just mention: [Shankar, 1992], [Pym and Wallen, 1990], and [Sahlin *et al.*, 1992].

<sup>23</sup>These would be existentially quantified formulas in classical logic as checking  $\Delta \vdash \forall x A$  is equivalent to check  $\Delta \cup \{\exists x \neg A\}$  for inconsistency.

It is likely that in order to develop the first-order extensions for others non-classical logics one could take advantage of the labelled proof system. One would like to represent the semantic options on the interpretation of quantifiers by tinkering with the unification and the Skolemisation mechanism. The constraints on the unification and Skolemisation might perhaps be expressed in terms of the labels associated with the formulas and their dependencies. At present the extension of the goal-directed methods to the main families of non-classical logics along these lines is not at hand and it is a major topic of future investigation.

Dov M. Gabbay  
*King's College London, UK.*

Nicola Olivetti  
*Università di Torino, Italy.*

### BIBLIOGRAPHY

- [Abadi and Manna, 1989] M. Abadi and Z. Manna. Temporal logic programming. *Journal of Symbolic Computation*, **8**, 277–295, 1989.
- [Andreoli, 1992] J. M. Andreoli. Logic programming with focusing proofs in linear logic. *Journal Logic and Computation*, **2**, 297–347, 1992.
- [Andreoli and Pareschi, 1991] J. M. Andreoli and R. Pareschi. Linear objects: logical processes with built in inheritance. *New Generation Computing*, **9**, 445–474, 1991.
- [Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap. *Entailment, The Logic of Relevance and Necessity*, Vol. 1. Princeton University Press, New Jersey, 1975.
- [Anderson *et al.*, 1992] A. R. Anderson, N. D. Belnap and J. M. Dunn. *Entailment, The Logic of Relevance and Necessity*, Vol. 2. Princeton University Press, New Jersey, 1992.
- [Baldoni *et al.*, 1998] M. Baldoni, L. Giordano and A. Martelli. A modal extension of logic programming, modularity, beliefs and hypothetical reasoning. *Journal of Logic and Computation*, **8**, 597–635, 1998.
- [Basin *et al.*, 1997a] D. Basin, S. Matthews and L. Viganò. Labelled propositional modal logics: theory and practice. *Journal of Logic and Computation*, **7**, 685–717, 1997.
- [Basin *et al.*, 1997b] D. Basin, S. Matthews and L. Viganò. A new method for bounding the complexity of modal logics. In *Proceedings of the Fifth Gödel Colloquium*, pp. 89–102, LNCS 1289, Springer-Verlag, 1997.
- [Basin *et al.*, 1999] D. Basin, S. Matthews and L. Viganò. Natural deduction for non-classical logics. *Studia Logica*, **60**, 119–160, 1998.
- [Bollen, 1991] A. W. Bollen. Relevant logic programming, *Journal of Automated Reasoning*, **7**, 563–585, 1991.
- [Došen, 1988] K. Došen. Sequent systems and grupoid models I. *Studia Logica*, **47**, 353–385, 1988.
- [Došen, 1989] K. Došen. Sequent systems and grupoid models II. *Studia Logica*, **48**, 41–65, 1989.
- [Dummett, 2001] M. Dummett. *Elements of Intuitionism*, Oxford University Press, (First edn. 1977), second edn., 2001.
- [Dunn, 1986] J. M. Dunn. Relevance logic and entailment. In *Handbook of Philosophical Logic*, vol III, D. Gabbay and F. Guenther, eds. pp. 117–224. D. Reidel, Dordrecht, 1986.
- [Dyckhoff, 1992] R. Dyckhoff. Contraction-free sequent calculi for intuitionistic logic, *Journal of Symbolic Logic*, **57**, 795–807, 1992.

- [Eshghi, 1989] K. Eshghi and R. Kowalski. Abduction compared with negation by failure. In *Proceedings of the 6th ICLP*, pp. 234–254, Lisbon, 1989.
- [Farinãs, 1986] L. Farinãs del Cerro. MOLOG: a system that extends Prolog with modal logic. *New Generation Computing*, **4**, 35–50, 1986.
- [Fine, 1974] K. Fine. Models for entailment, *Journal of Philosophical Logic*, **3**, 347–372, 1974.
- [Fitting, 1983] M. Fitting. *Proof methods for Modal and Intuitionistic Logic*, vol 169 of Synthese library, D. Reidel, Dordrecht, 1983.
- [Fitting, 1990] M. Fitting. *First-order Logic and Automated Theorem Proving*. Springer-Verlag, 1990.
- [Gabbay, 1981] D. M. Gabbay. *Semantical Investigations in Heyting's Intuitionistic Logic*. D. Reidel, Dordrecht, 1981.
- [Gabbay, 1985] D. M. Gabbay. *N-Prolog Part 2*. *Journal of Logic Programming*, 251–283, 1985.
- [Gabbay, 1987] D. M. Gabbay. Modal and temporal logic programming. In *Temporal Logics and their Applications*, A. Galton, ed. pp/ 197–237. Academic Press, 1987.
- [Gabbay, 1991] D. M. Gabbay. Algorithmic proof with diminishing resources. In *Proceedings of CSL'90*, pp. 156–173. LNCS vol 533, Springer-Verlag, 1991.
- [Gabbay, 1992] D. M. Gabbay. Elements of algorithmic proof. In *Handbook of Logic in Theoretical Computer Science*, vol 2. S. Abramsky, D. M. Gabbay and T. S. E. Maibaum, eds., pp. 307–408, Oxford University Press, 1992.
- [Gabbay, 1996] D. M. Gabbay, *Labelled Deductive Systems* (vol I), Oxford Logic Guides, Oxford University Press, 1996.
- [Gabbay, 1998] D. M. Gabbay. *Elementary Logic*, Prentice Hall, 1998.
- [Gabbay and Kriwaczek, 1991] D. M. Gabbay and F. Kriwaczek. A family of goal-directed theorem-provers based on conjunction and implication. *Journal of Automated Reasoning*, **7**, 511–536, 1991.
- [Gabbay and Olivetti, 1997] D. M. Gabbay and N. Olivetti. Algorithmic proof methods and cut elimination for implicational logics - part I, modal implication. *Studia Logica*, **61**, 237–280, 1998.
- [Gabbay and Olivetti, 1998] D. Gabbay and N. Olivetti. Goal-directed proof-procedures for intermediate logics. In *Proc. of LD'98 First International Workshop on Labelled Deduction*, Freiburg, 1998.
- [Gabbay et al., 1999] D. M. Gabbay, N. Olivetti and S. Vorobyov. Goal-directed proof method is optimal for intuitionistic propositional logic. Manuscript, 1999.
- [Gabbay and Olivetti, 2000] D.M. Gabbay and N. Olivetti. *Goal-directed proof theory*, Kluwer Academic Publishers, Applied Logic Series, Dordrecht, 2000.
- [Gabbay and Reyle, 1984] D. M. Gabbay and U. Reyle. *N-Prolog: an Extension of Prolog with hypothetical implications*, I. *Journal of Logic Programming*, **4**, 319–355, 1984.
- [Gabbay and Reyle, 1993] D. M. Gabbay and U. Reyle. Computation with run time Skolemisation (*N-Prolog*, Part 3). *Journal of Applied Non Classical Logics*, **3**, 93–128, 1993.
- [Gallier, 1987] J. H. Gallier. *Logic for Computer Science*, John Wiley, New York, 1987.
- [Galmiche and Pym, 1999] D. Galmiche and D. Pym. Proof-Search in type-theoretic languages. To appear in *Theoretical Computer Science*, 1999.
- [Giordano et al., 1992] L. Giordano, A. Martelli and G. F. Rossi. Extending Horn clause logic with implication goals. *Theoretical Computer Science*, **95**, 43–74, 1992.
- [Giordano and Martelli, 1994] L. Giordano and A. Martelli. Structuring logic programs: a modal approach. *Journal of Logic Programming*, **21**, 59–94, 1994.
- [Girard, 1987] J. Y. Girard. Linear logic. *Theoretical Computer Science* **50**, 1–101, 1987.
- [Goré, 1999] R. Goré. Tableaux methods for modal and temporal logics. In *Handbook of Tableau Methods*, M. D'Agostino et al., eds. Kluwer Academic Publishers, 1999.
- [Hähnle and Schmitt, 1994] R. Hähnle and P. H. Schmitt. The liberalized  $\delta$ -rule in free-variables semantic tableaux. *Journal Automated Reasoning*, **13**, 211–221, 1994.
- [Harland and Pym, 1991] J. Harland and D. Pym. The uniform proof-theoretic foundation of linear logic programming. In *Proceedings of the 1991 International Logic Programming Symposium*, San Diego, pp. 304–318, 1991.

- [Harland, 1997] J. Harland. On Goal-Directed Provability in Classical Logic, *Computer Languages*, pp. 23:2-4:161-178, 1997.
- [Harland and Pym, 1997] J. Harland and D. Pym. Resource-distribution by Boolean constraints. In *Proceedings of CADE 1997*, pp. 222–236, Springer Verlag, 1997.
- [Heudering *et al.*, 1996] A. Heudering, M. Seyfried and H. Zimmerman. Efficient loop-check for backward proof search in some non-classical propositional logics. In (P. Miglioli *et al.* eds.) *Tableaux 96*, pp. 210–225. LNCS 1071, Springer Verlag, 1996.
- [Hodas, 1993] J. Hodas. *Logic programming in intuitionistic linear logic: theory, design, and implementation*. PhD Thesis, University of Pennsylvania, Department of Computer and Information Sciences, 1993.
- [Hodas and Miller, 1994] J. Hodas and D. Miller. Logic programming in a fragment of intuitionistic linear logic. *Information and Computation*, **110**, 327–365, 1994.
- [Hudelmaier, 1990] J. Hudelmaier. Decision procedure for propositional  $n$ -prolog. In (P. Schroeder-Heister ed.) *Extensions of Logic Programming*, pp. 245–251, Springer Verlag, 1990.
- [Hudelmaier, 1990] J. Hudelmaier. An  $O(n \log n)$ -space decision procedure for intuitionistic propositional logic. *Journal of Logic and Computation*, **3**, 63–75, 1993.
- [Lambek, 1958] J. Lambek. The mathematics of sentence structure, *American Mathematics Monthly*, **65**, 154–169, 1958.
- [Lewis, 1912] C. I. Lewis. Implication and the algebra of logic. *Mind*, **21**, 522–531, 1912.
- [Lewis and Langford, 1932] C. I. Lewis and C. H. Langford. *Symbolic Logic*, The Century Co., New York, London, 1932. Second edition, Dover, New York, 1959.
- [Lincoln *et al.*, 1991] P. Lincoln, P. Scedrov and N. Shankar. Linearizing intuitionistic implication. In *Proceedings of LICS'91*, G. Kahn ed., pp. 51–62, IEEE, 1991.
- [Lloyd, 1984] J. W. Lloyd, *Foundations of Logic Programming*, Springer, Berlin, 1984.
- [Loveland, 1991] D. W. Loveland. Near-Horn prolog and beyond. *Journal of Automated Reasoning*, **7**, 1–26, 1991.
- [Loveland, 1992] D. W. Loveland. A comparison of three Prolog extensions. *Journal of Logic Programming*, **12**, 25–50, 1992.
- [Masini, 1992] A. Masini. 2-Sequent calculus: a proof theory of modality. *Annals of Pure and Applied Logics*, **59**, 115–149, 1992.
- [Massacci, 1994] F. Massacci. Strongly analytic tableau for normal modal logics. In *Proceedings of CADE '94*, LNAI 814, pp. 723–737, Springer-Verlag, 1994.
- [McCarty, 1988a] L. T. McCarty. Clausal intuitionistic logic. I. Fixed-point semantics. *Journal of Logic Programming*, **5**, 1–31, 1988.
- [McCarty, 1988b] L. T. McCarty. Clausal intuitionistic logic. II. Tableau proof procedures. *Journal of Logic Programming*, **5**, 93–132, 1988.
- [McRobbie and Belnap, 1979] M. A. McRobbie and N. D. Belnap. Relevant analytic tableaux. *Studia Logica*, **38**, 187–200, 1979.
- [Miller *et al.*, 1991] D. Miller, G. Nadathur, F. Pfenning and A. Scedrov. Uniform proofs as a foundation for logic programming. *Annals of Pure and Applied Logic*, **51**, 125–157, 1991.
- [Miller, 1992] D. A. Miller. Abstract syntax and logic programming. In *Logic Programming: Proceedings of the 2nd Russian Conference*, pp. 322–337, LNAI 592, Springer Verlag, 1992.
- [Nadathur, 1998] G. Nadathur. Uniform provability in classical logic. *Journal of Logic and Computation*, **8**, 209–229, 1998.
- [O'Hearn and Pym, 1999] P. W. O'Hearn and D. Pym. The logic of bunched implications. *Bulletin of Symbolic Logic*, **5**, 215–244, 1999.
- [Olivetti, 1995] N. Olivetti. *Algorithmic Proof-theory for Non-classical and Modal Logics*. PhD Thesis, Dipartimento di Informatica, Università di Torino, 1995.
- [Ono, 1993] H. Ono. Semantics for substructural logics. In P. Schroeder-Heister and K. Došen (eds.) *Substructural Logics*, pp. 259–291, Oxford University Press, 1993.
- [Ono, 1998] H. Ono. Proof-theoretic methods in non-classical logics—an introduction. In (M. Takahashi *et al.* eds.) *Theories of Types and Proofs*, pp. 267–254, Mathematical Society of Japan, 1998.
- [Prawitz, 1965] D. Prawitz. *Natural Deduction*. Almqvist & Wiksell, 1965.

- [Pym and Wallen, 1990] D. J. Pym and L. A. Wallen. Investigation into proof-search in a system of first-order dependent function types. In Proc. of CADE'90, pp. 236–250. LNCS vol 449, Springer-Verlag, 1990.
- [Restall, 1999] G. Restall. *An introduction to substructural logics*. Routledge, to appear, 1999.
- [Russo, 1996] A. Russo. Generalizing propositional modal logics using labelled deductive systems. In (F. Baader and K. Schulz, eds.) *Frontiers of Combining Systems*, pp. 57–73. Vol. 3 of Applied Logic Series, Kluwer Academic Publishers, 1996.
- [Sahlin *et al.*, 1992] D. Sahlin, T. Franzén, and S. Haridi. An intuitionistic predicate logic theorem prover. *Journal of Logic and Computation*, **2**, 619–656, 1992.
- [Schroeder-Heister and Došen, 1993] P. Schroeder-Heister and K. Došen(eds.) *Substructural Logics*. Oxford University Press, 1993.
- [Shankar, 1992] N. Shankar. Proof search in the intuitionistic sequent calculus. In Proc. of CADE'92, pp. 522–536. LNAI 607, Springer Verlag, 1992.
- [Statman, 1979] R. Statman. Intuitionistic propositional logic is polynomial-space complete. *Theoretical Computer Science*, **9**, 67–72, 1979.
- [Thistlewaite *et al.*, 1988] P. B. Thistlewaite, M. A. McRobbie and B. K. Meyer. *Automated Theorem Proving in Non Classical Logics*, Pitman, 1988.
- [Turner, 1985] R. Turner. *Logics for Artificial Intelligence*, Ellis Horwood Ltd., 1985.
- [Troelstra, 1969] A. S. Troelstra. *Principles of Intuitionism*. Springer-Verlag, Berlin, 1969.
- [Urquhart, 1972] A. Urquhart. The semantics of relevance logics. *The Journal of Symbolic Logic*, **37**, 159–170, 1972.
- [Urquhart, 1984] A. Urquhart. The undecidability of entailment and relevant implication. *The Journal of Symbolic Logic*, **49**, 1059–1073, 1984.
- [van Dalen, 1986] D. Van Dalen. Intuitionistic logic. In *Handbook of Philosophical logic, Vol 3*, D. Gabbay and F. Guenther, eds. pp. 225–239. D. Reidel, Dordrecht, 1986.
- [Viganò, 1999] L. Viganò. *Labelled Non-classical Logics*. Kluwer Academic Publishers, 2000. to appear.
- [Wallen, 1990] L. A. Wallen, *Automated Deduction in Nonclassical Logics*, MIT Press, 1990.
- [Wansing, 1994] H. Wansing. Sequent Calculi for normal propositional modal logics. *Journal of Logic and Computation*, **4**, 125–142, 1994.





## ON NEGATION, COMPLETENESS AND CONSISTENCY

### 1 INTRODUCTION

In this Chapter we try to understand negation from two different points of view: a syntactical one and a semantic one. Accordingly, we identify two different types of negation. The same connective of a given logic might be of both types, but this might not always be the case.

The syntactical point of view is an abstract one. It characterizes connectives according to the internal *role* they have inside a logic, regardless of any meaning they are intended to have (if any). With regard to negation our main thesis is that the availability of what we call below an internal negation is what makes a logic essentially *multiple-conclusion*.

The semantic point of view, in contrast, is based on the intuitive meaning of a given connective. In the case of negation this is simply the intuition that the negation of a proposition  $A$  is true if  $A$  is not, and not true if  $A$  is true.<sup>1</sup>

Like in most modern treatments of logics (see, e.g., [Scott, 1974; Scott, 1974b; Hacking, 1979; Gabbay, 1981; Urquhart, 1984; Wojcicki, 1988; Epstein, 1995; Avron, 1991a; Cleave, 1991; Fagin *et al.*, 1992]), our study of negation will be in the framework of Consequence Relations (CRs). Following [Avron, 1991a], we use the following rather general meaning of this term:

DEFINITION.

(1) A *Consequence Relation* (CR) on a set of formulas is a binary relation  $\vdash$  between (finite) multisets of formulas s.t.:

(I) Reflexivity:  $A \vdash A$  for every formula  $A$ .

(II) Transitivity, or “Cut”: if  $\Gamma_1 \vdash \Delta_1, A$  and  $A, \Gamma_2 \vdash \Delta_2$ , then  $\Gamma_1, \Gamma_2 \vdash \Delta_1, \Delta_2$ .

(III) Consistency:  $\emptyset \not\vdash \emptyset$  (where  $\emptyset$  is the empty multiset).

---

<sup>1</sup>We have avoided here the term “false”, since we do not want to commit ourselves to the view that  $A$  is false precisely when it is not true. Our formulation of the intuition is therefore obviously circular, but this is unavoidable in intuitive informal characterizations of basic connectives and quantifiers.

(2) A single-conclusion CR is a CR  $\vdash$  such that  $\Gamma \vdash \Delta$  only if  $\Delta$  consists of a single formula.

The notion of a (multiple-conclusion) CR was introduced in [Scott, 1974] and [Scott, 1974b]. It was a generalization of Tarski's notion of a consequence relation, which was single-conclusion. Our notions are, however, not identical to the original ones of Tarski and Scott. First, they both considered *sets* (rather than multisets) of formulas. Second, they impose a third demand on CRs: monotonicity. We shall call a (single-conclusion or multiple-conclusion) CR which satisfies these two extra conditions *ordinary*. A single-conclusion, ordinary CR will be called *Tarskian*.<sup>2</sup>

The notion of a “logic” is in practice broader than that of a CR, since usually several CRs are associated with a given logic. Given a logic  $\mathcal{L}$  there are in most cases two major single-conclusion CRs which are naturally associated with it: the external CR  $\vdash_{\mathcal{L}}^e$  and the internal CR  $\vdash_{\mathcal{L}}^i$ . For example, if  $\mathcal{L}$  is defined by some axiomatic system  $AS$  then  $A_1, \dots, A_n \vdash_{\mathcal{L}}^e B$  iff there exists a proof in  $AS$  of  $B$  from  $A_1, \dots, A_n$  (according to the most standard meaning of this notion as defined in undergraduate textbooks on mathematical logic), while  $A_1, \dots, A_n \vdash_{\mathcal{L}}^i B$  iff  $A_1 \rightarrow (A_2 \rightarrow \dots \rightarrow (A_n \rightarrow B) \dots)$  is a theorem of  $AS$  (where  $\rightarrow$  is an appropriate “implication” connective of the logic). Similarly if  $\mathcal{L}$  is defined using a Gentzen-type system  $G$  then  $A_1, \dots, A_n \vdash_{\mathcal{L}}^i B$  if the sequent  $A_1, \dots, A_n \Rightarrow B$  is provable in  $G$ , while  $A_1, \dots, A_n \vdash_{\mathcal{L}}^e B$  iff there exists a proof in  $G$  of  $\Rightarrow B$  from the assumptions  $\Rightarrow A_1, \dots, \Rightarrow A_n$  (perhaps with cuts).  $\vdash_{\mathcal{L}}^e$  is always a Tarskian relation,  $\vdash_{\mathcal{L}}^i$  frequently is not. The existence (again, in most cases) of these two CRs should be kept in mind in what follows. The reason is that semantic characterizations of connectives are almost always done w.r.t. Tarskian CRs (and so here  $\vdash_{\mathcal{L}}^e$  is usually relevant). This is not the case with syntactical characterizations, and here frequently  $\vdash_{\mathcal{L}}^i$  is more suitable.

## 2 THE SYNTACTICAL POINT OF VIEW

### 2.1 Classification of basic connectives

Our general framework allows us to give a completely abstract definition, *independent of any semantic interpretation*, of standard connectives. These characterizations explain why these connectives are so important in almost every logical system.

In what follows  $\vdash$  is a fixed CR. All definitions are taken to be relative to  $\vdash$  (the definitions are taken from [Avron, 1991a]).

---

<sup>2</sup>What we call a Tarskian CR is exactly Tarski's original notion. In [Avron, 1994] we argue at length why the notion of a proof in an axiomatic system naturally leads to *our* notion of single-conclusion CR, and why the further generalization to multiple-conclusion CR is also very reasonable.

We consider two types of connectives. The *internal* connectives, which make it possible to transform a given sequent into an equivalent one that has a special required form, and the *combining* connectives, which allow us to combine (under certain circumstances) two sequents into one which contains exactly the same information. The most common (and useful) among these are the following connectives:

**Internal Disjunction:**  $+$  is an internal disjunction if for all  $\Gamma, \Delta, A, B$ :

$$\Gamma \vdash \Delta, A, B \quad \text{iff} \quad \Gamma \vdash \Delta, A + B .$$

**Internal Conjunction:**  $\otimes$  is an internal conjunction if for all  $\Gamma, \Delta, A, B$ :

$$\Gamma, A, B \vdash \Delta \quad \text{iff} \quad \Gamma, A \otimes B \vdash \Delta .$$

**Internal Implication:**  $\rightarrow$  is an internal implication if for all  $\Gamma, \Delta, A, B$ :

$$\Gamma, A \vdash B, \Delta \quad \text{iff} \quad \Gamma \vdash A \rightarrow B, \Delta .$$

**Internal Negation:**  $\neg$  is an internal negation if the following two conditions are satisfied by all  $\Gamma, \Delta$  and  $A$ :

- (1)  $A, \Gamma \vdash \Delta \quad \text{iff} \quad \Gamma \vdash \Delta, \neg A$
- (2)  $\Gamma \vdash \Delta, A \quad \text{iff} \quad \neg A, \Gamma \vdash \Delta .$

**Combining Conjunction:**  $\wedge$  is a combining conjunction iff for all  $\Gamma, \Delta, A, B$ :

$$\Gamma \vdash \Delta, A \wedge B \quad \text{iff} \quad \Gamma \vdash \Delta, A \quad \text{and} \quad \Gamma \vdash \Delta, B .$$

**Combining Disjunction:**  $\vee$  is a combining disjunction iff for all  $\Gamma, \Delta, A, B$

$$A \vee B, \Gamma \vdash \Delta \quad \text{iff} \quad A, \Gamma \vdash \Delta \quad \text{and} \quad B, \Gamma \vdash \Delta .$$

**Note:** The combining connectives are called “additives” in Linear logic (see [Girard, 1987]) and “extensional” in Relevance logic. The internal ones correspond, respectively, to the “multiplicative” and the “intensional” connectives.

Several well-known logics can be defined using the above connectives:

$LL_m$  — **Multiplicative Linear Logic** (without the propositional constants): This is the logic which corresponds to the *minimal* (multiset) CR which includes all the internal connectives.

$LL_{ma}$  — **Propositional Linear Logic** (without the “exponentials” and the propositional constants): This corresponds to the minimal consequence relation which contains all the connectives introduced above.

$R_m$  — **the Intensional Fragment of the Relevance Logic  $R$** :<sup>3</sup> This corresponds to the minimal CR which contains all the internal connectives and is *closed under contraction*.

<sup>3</sup>see [Anderson and Belnap, 1975] or [Dunn, 1986].

***R* without Distribution:** This corresponds to the minimal CR which contains all the connectives which were described above and is closed under contraction.

***RM<sub>I<sub>m</sub></sub>*** — **the Intensional Fragment of the Relevance Logic *RM<sub>I</sub>*:**<sup>4</sup> This corresponds to the minimal sets-CR which contains all the internal connectives.

**Classical Proposition Logic:** This of course corresponds to the minimal ordinary CR which has all the above connectives. Unlike the previous logics there is no difference in it between the combining connectives and the corresponding internal ones.

In all these examples we refer, of course, to the *internal* consequence relations which naturally correspond to these logics (In all of them it can be defined by either of the methods described above).

## 2.2 Internal Negation and Strong Symmetry

Among the various connectives defined above only negation essentially demands the use of multiple-conclusion CRs (even the existence of an internal disjunction does not *force* multiple-conclusions, although its existence is trivial otherwise.). Moreover, its existence creates full symmetry between the two sides of the turnstyle. Thus in its presence, closure under any of the structural rules on one side entails closure under the same rule on the other, the existence of any of the binary internal connectives defined above implies the existence of the rest, and the same is true for the combining connectives.

To sum up: internal negation is the connective with which “the hidden symmetries of logic” [Girard, 1987] are explicitly represented. We shall call, therefore, any multiple-conclusion CR which possesses it *strongly symmetric*.

Some alternative characterizations of an internal negation are given in the following easy proposition.

PROPOSITION 1. *The following conditions on  $\vdash$  are all equivalent:*

- (1)  $\neg$  is an internal negation for  $\vdash$ .
- (2)  $\Gamma \vdash \Delta, A$  iff  $\Gamma, \neg A \vdash \Delta$
- (3)  $A, \Gamma \vdash \Delta$  iff  $\Gamma \vdash \Delta, \neg A$
- (4)  $A, \neg A \vdash$  and  $\vdash \neg A, A$
- (5)  $\vdash$  is closed under the rules:

$$\frac{A, \Gamma \vdash \Delta}{\Gamma \vdash \Delta, \neg A} \qquad \frac{\Gamma \vdash \Delta, A}{\neg A, \Gamma \vdash \Delta} .$$

Our characterization of internal negation and of symmetry has been done within the framework of multiple-conclusion relations. Single-conclusion

<sup>4</sup>see [Avron, 1990a; Avron, 1990b].

CRs are, however, more natural. We proceed next to introduce corresponding notions for them.

DEFINITION.

(1) Let  $\vdash_{\mathcal{L}}$  be a single-conclusion CR (in a language  $\mathcal{L}$ ), and let  $\neg$  be a unary connective of  $\mathcal{L}$ .  $\vdash_{\mathcal{L}}$  is called *strongly symmetric* w.r.t. to  $\neg$ , and  $\neg$  is called an *internal negation* for  $\vdash_{\mathcal{L}}$ , if there exists a multiple-conclusion CR  $\vdash_{\mathcal{L}}^*$  with the following properties:

- (i)  $\Gamma \vdash_{\mathcal{L}}^* A$  iff  $\Gamma \vdash_{\mathcal{L}} A$
- (ii)  $\neg$  is an internal negation for  $\vdash_{\mathcal{L}}^*$

(2) A single-conclusion CR  $\vdash_{\mathcal{L}}$  is called *essentially multiple-conclusion* iff it has an internal negation.

Obviously, if a CR  $\vdash_{\mathcal{L}}^*$  like in the last definition exists then it is unique. We now formulate sufficient and necessary conditions for its existence.

THEOREM 2.  $\vdash_{\mathcal{L}}$  is strongly symmetric w.r.t.  $\neg$  iff the following conditions are satisfied:

- (i)  $A \vdash_{\mathcal{L}} \neg\neg A$
- (ii)  $\neg\neg A \vdash_{\mathcal{L}} A$
- (iii) If  $\Gamma, A \vdash_{\mathcal{L}} B$  then  $\Gamma, \neg B \vdash_{\mathcal{L}} \neg A$ .

**Proof.** The conditions are obviously necessary. Assume, for the converse, that  $\vdash_{\mathcal{L}}$  satisfies the conditions. Define:  $A_1, \dots, A_n \vdash_{\mathcal{L}}^s B_1, \dots, B_k$  iff for every  $1 \leq i \leq n$  and  $1 \leq j \leq k$ :

$$A_1, \dots, A_{i-1}, \neg B_1, \dots, \neg B_k, A_{i+1}, \dots, A_n \vdash \neg A_i$$

$$A_1, \dots, A_n, \neg B_1, \dots, \neg B_{j-1}, \neg B_{j+1}, \dots, \neg B_k \vdash B_j.$$

It is easy to check that  $\vdash_{\mathcal{L}}^s$  is a CR whenever  $\vdash_{\mathcal{L}}$  is a CR (whether single-conclusion or multiple-conclusion), and that if  $\Gamma \vdash_{\mathcal{L}}^s A$  then  $\Gamma \vdash_{\mathcal{L}} A$ . The first two conditions imply (together) that  $\neg$  is an internal negation for  $\vdash_{\mathcal{L}}^s$  (in particular: the second entails that if  $A, \Gamma \vdash_{\mathcal{L}}^s \Delta$  then  $\Gamma \vdash_{\mathcal{L}}^s \Delta, \neg A$  and the first that if  $\Gamma \vdash_{\mathcal{L}}^s \Delta, A$  then  $\neg A, \Gamma \vdash_{\mathcal{L}}^s \Delta$ ). Finally, the third condition entails that  $\vdash_{\mathcal{L}}^s$  is conservative over  $\vdash_{\mathcal{L}}$ . ■

**Examples of logics with an internal negation.**

1. Classical logic.
2. Extensions of classical logic, like the various modal logics.
3. Linear logic and its various fragments.

4. The various Relevance logics (like  $R$  and  $RM$  (see [Anderson and Belnap, 1975; Dunn, 1986; Anderson and Belnap, 1992] or  $RMI$  [Avron, 1990a; Avron, 1990b]) and their fragments.
5. The various many-valued logics of Łukasiewicz, as well as Sobociński 3-valued logic [Sobociński, 1952].

**Examples of logics without an internal negation.**

1. Intuitionistic logic.
2. Kleene's 3-valued logic and its extension  $LPF$  [Jones, 1986].

**Note:** Again, in all these examples above it is the *internal* CR which is essentially multiple-conclusion (or not) and has an internal negation. This is true even for classical predicate calculus: There, e.g.,  $\forall xA(x)$  follows from  $A(x)$  according to the *external* CR, but  $\neg A(x)$  does not follow from  $\neg\forall xA(x)$ .<sup>5</sup>

All the positive examples above are instances of the following proposition, the easy proof of which we leave to the reader:

**PROPOSITION 3.** *Let  $\mathcal{L}$  be any logic in a language containing  $\neg$  and  $\rightarrow$ . Suppose that the set of valid formulae of  $\mathcal{L}$  includes the set of formulae in the language of  $\{\neg, \rightarrow\}$  which are theorems of Linear Logic,<sup>6</sup> and that it is closed under MP for  $\rightarrow$ . Then the internal consequence relation of  $\mathcal{L}$  (defined using  $\rightarrow$  as in the introduction) is strongly symmetric (with respect to  $\neg$ ).*

The next two theorems discuss what properties of  $\vdash_{\mathcal{L}}$  are preserved by  $\vdash_{\mathcal{L}}^s$ . The proofs are straightforward.

**THEOREM 4.** *Assume  $\vdash_{\mathcal{L}}$  is essentially multiple-conclusion.*

1.  $\vdash_{\mathcal{L}}^s$  is monotonic iff so is  $\vdash_{\mathcal{L}}$ .
2.  $\vdash_{\mathcal{L}}^s$  is closed under expansion (the converse of contraction) iff so is  $\vdash_{\mathcal{L}}$ .
3.  $\wedge$  is a combining conjunction for  $\vdash_{\mathcal{L}}^s$  iff it is a combining conjunction for  $\vdash_{\mathcal{L}}$ .
4.  $\rightarrow$  is an internal implication for  $\vdash_{\mathcal{L}}^s$  iff it is an internal implication for  $\vdash_{\mathcal{L}}$ .

<sup>5</sup>The internal CR of classical logic has been called the “truth” CR in [Avron, 1991a] and was denoted there by  $\vdash^t$ , while the external one was called the “validity” CR and was denoted by  $\vdash^v$ . On the propositional level there is no difference between the two.

<sup>6</sup>Here  $\neg$  should be translated into linear negation,  $\rightarrow$  – into linear implication.

**Notes:**

- 1) Because  $\vdash_{\mathcal{L}}^s$  is strongly symmetric, Parts (3) and (4) can be formulated as follows:  $\vdash_{\mathcal{L}}^s$  has the internal connectives iff  $\vdash_{\mathcal{L}}$  has an internal implication and it has the combining connectives iff  $\vdash_{\mathcal{L}}$  has a combining conjunction.
- 2) In contrast, a combining disjunction for  $\vdash_{\mathcal{L}}$  is not necessarily a combining disjunction for  $\vdash_{\mathcal{L}}^s$ . It is easy to see that a necessary and sufficient condition for this to happen is that  $\vdash_{\mathcal{L}} \neg(A \vee B)$  whenever  $\vdash_{\mathcal{L}} \neg A$  and  $\vdash_{\mathcal{L}} \neg B$ . An example of an essentially multiple-conclusion system with a combining disjunction which does not satisfy the above condition is *RMI* of [Avron, 1990a; Avron, 1990b]. That system indeed does not have a combining conjunction. This shows that a *single-conclusion* logic  $\mathcal{L}$  with an internal negation and a combining disjunction does not necessarily have a combining conjunction (unless  $\mathcal{L}$  is monotonic). The converse situation is not possible, though: If  $\neg$  is an internal negation and  $\wedge$  is a combining conjunction then  $\neg(\neg A \wedge \neg B)$  defines a combining disjunction even in the single-conclusion case.
- 3) An internal conjunction  $\otimes$  for  $\vdash_{\mathcal{L}}$  is also not necessarily an internal conjunction for  $\vdash_{\mathcal{L}}^s$ . We need here the extra condition that if  $A \vdash_{\mathcal{L}} \neg B$  then  $\vdash_{\mathcal{L}} \neg(A \otimes B)$ . An example which shows that this condition does not necessarily obtain even if  $\vdash_{\mathcal{L}}$  is an ordinary CR, is given by the following CR  $\vdash_{triv}$ :

$$A_1, \dots, A_n \vdash_{triv} B \quad \text{iff} \quad n \geq 1 .$$

It is obvious that  $\vdash_{triv}$  is a Tarskian CR and that every unary connective of its language is an internal negation for it, while every binary connective is an internal conjunction. The condition above fails, however, for  $\vdash_{triv}$ .

- 4) The last example shows also that  $\vdash_{\mathcal{L}}^s$  may not be closed under contraction when  $\vdash_{\mathcal{L}}$  does, even if  $\vdash_{\mathcal{L}}$  is Tarskian. Obviously,  $\Gamma \vdash_{triv}^s \Delta$  iff  $|\Gamma \cup \Delta| \geq 2$ . Hence  $\vdash_{triv}^s A, A$  but  $\not\vdash_{triv}^s A$ . The exact situation about contraction is given in the next proposition.

**PROPOSITION 5.** *If  $\vdash_{\mathcal{L}}$  is essentially multiple-conclusion then  $\vdash_{\mathcal{L}}^s$  is closed under contraction iff  $\vdash_{\mathcal{L}}$  is closed under contraction and satisfies the following condition:*

$$\text{If } A \vdash_{\mathcal{L}} B \text{ and } \neg A \vdash_{\mathcal{L}} B \text{ then } \vdash_{\mathcal{L}} B .$$

*In case  $\vdash_{\mathcal{L}}$  has a combining disjunction this is equivalent to:*

$$\vdash_{\mathcal{L}} \neg A \vee A .$$

**Proof.** Suppose first that  $\vdash_{\mathcal{L}}$  is closed under contraction and satisfies the condition. Assume that  $\Gamma \vdash_{\mathcal{L}}^s \Delta, A, A$ . If either  $\Gamma$  or  $\Delta$  is not empty then this is equivalent to  $\neg A, \neg A, \Gamma^* \vdash_{\mathcal{L}} B$  for some  $\Gamma^*$  and  $B$ . Since  $\vdash_{\mathcal{L}}$  is closed under contraction, this implies that  $\neg A, \Gamma^* \vdash_{\mathcal{L}} B$ , and so  $\Gamma \vdash_{\mathcal{L}}^s \Delta, A$ . If both  $\Gamma$  and  $\Delta$  are empty then we have  $\neg A \vdash_{\mathcal{L}} A$ . Since also  $A \vdash_{\mathcal{L}} A$ , the condition implies that  $\vdash_{\mathcal{L}} A$ , and so  $\vdash_{\mathcal{L}}^s A$ .

For the converse, suppose  $\vdash_{\mathcal{L}}^s$  is closed under contraction. This obviously entails that so is also  $\vdash_{\mathcal{L}}$ . Assume now that  $A \vdash_{\mathcal{L}} B$  and  $\neg A \vdash_{\mathcal{L}} B$ . Then  $A \vdash_{\mathcal{L}}^s B$  and  $\vdash_{\mathcal{L}}^s B, A$ . Applying cut we get that  $\vdash_{\mathcal{L}}^s B, B$ , and so  $\vdash_{\mathcal{L}}^s B$ . It follows that  $\vdash_{\mathcal{L}} B$ . ■

### 3 THE SEMANTIC POINT OF VIEW

We turn in this section to the semantic aspect of negation.

#### 3.1 *The General Framework*

A “semantics” for a logic consists of a set of “models”. The main property of a model is that every sentence of a logic is either true in it or not (and not both). The logic is sound with respect to the semantics if the set of sentences which are true in each model is closed under the CR of the logic, and complete if a sentence  $\varphi$  follows (according to the logic) from a set  $T$  of assumptions iff every model of  $T$  is a model of  $\varphi$ . Such a characterization is, of course, possible only if the CR we consider is Tarskian. *In this section we assume, therefore, that we deal only with Tarskian CRs.* For logics like Linear Logic and Relevance logics this means that we consider only the *external* CRs which are associated with them (see the Introduction).

Obviously, the essence of a “model” is given by the set of sentences which are true in it. Hence a semantics is, essentially, just a set  $S$  of theories. Intuitively, these are the theories which (according to the semantics) provide a full description of a possible state of affairs. Every other theory can be understood as a partial description of such a state, or as an approximation of a full description. Completeness means, then, that a sentence  $\varphi$  follows from a theory  $T$  iff  $\varphi$  belongs to every superset of  $T$  which is in  $S$  (in other words: iff  $\varphi$  is true in any possible state of affairs of which  $T$  is an approximation).

Now what constitutes a “model” is frequently defined using some kind of algebraic structures. Which kind (matrices with designated values, possible worlds semantics and so on) varies from one logic to another. It is difficult, therefore, to base a general, uniform theory on the use of such structures. Semantics (= a set of theories!) can also be defined, however, purely syntactically. Indeed, below we introduce several types of syntactically defined semantics which are very natural for *every* logic with “negation”. Our investigations will be based on these types.

Our description of the notion of a model reveals that *externally* it is based on two classical “laws of thought”: the law of contradiction and the law of excluded middle. When this external point of view is reflected inside the logic with the help of a unary connective  $\neg$  we call this connective a (strong) *semantic negation*. Its intended meaning is that  $\neg A$  should be true precisely



when  $A$  is not. The law of contradiction means then that only consistent theories may have a model, while the law of excluded middle means that the set of sentences which are true in some given model should be negation-complete. The sets of consistent theories, of complete theories and of normal theories (theories that are both) have, therefore a crucial importance when we want to find out to what degree a given unary connective of a logic can be taken as a semantic negation. Thus complete theories reflect a state of affairs in which the law of excluded middle holds. It is reasonable, therefore, to say that this law semantically obtains for a logic  $\mathcal{L}$  if its consequence relation  $\vdash_{\mathcal{L}}$  is *determined* by its set of complete theories. Similarly,  $\mathcal{L}$  (strongly) satisfies the law of contradiction iff  $\vdash_{\mathcal{L}}$  is determined by its set of consistent theories, and it semantically satisfies both laws iff  $\vdash_{\mathcal{L}}$  is determined by its set of normal theories.

The above characterizations might seem unjustifiably strong for logics which are designed to allow non-trivial inconsistent theories. For such logics the demand that  $\vdash_{\mathcal{L}}$  should be determined by its set of normal theories is reasonable only if we start with a consistent set of assumptions (this is called strong  $\mathcal{C}$ -normality below). A still weaker demand ( $\mathcal{C}$ -normality) is that any consistent set of assumptions should be an approximation of at least one normal state of affairs (in other words: it should have at least one normal extension).

It is important to note that the above characterizations are independent of the existence of any internal reflection of the laws (for example: in the forms  $\neg(\neg A \wedge A)$  and  $\neg A \vee A$ , for suitable  $\wedge$  and  $\vee$ ). There might be strong connections, of course, in many important cases, but they are neither necessary nor always simple.

We next define our general notion of semantics in precise terms.

**DEFINITION.** Let  $\mathcal{L}$  be a logic in  $L$  and let  $\vdash_{\mathcal{L}}$  be its associated CR.

1. A *setup* for  $\vdash_{\mathcal{L}}$  is a set of formulae in  $L$  which is closed under  $\vdash_{\mathcal{L}}$ . A *semantics* for  $\vdash_{\mathcal{L}}$  is a nonempty set of setups which does not include the trivial setup (i.e., the set of all formulae).
2. Let  $S$  be a semantics for  $\vdash_{\mathcal{L}}$ . An *S-model* for a formula  $A$  is any setup in  $S$  to which  $A$  belongs. An *S-model* of a theory  $T$  is any setup in  $S$  which is a superset of  $T$ . A formula is called *S-valid* iff every setup in  $S$  is a model of it. A formula  $A$  *S-follows* from a theory  $T$  ( $T \vdash_{\mathcal{L}}^S A$ ) iff every *S-model* of  $T$  is an *S-model* of  $A$ .

**PROPOSITION 6.**  $\vdash_{\mathcal{L}}^S$  is a (Tarskian) consequence relation and  $\vdash_{\mathcal{L}} \subseteq \vdash_{\mathcal{L}}^S$ .

**Notes:**

1.  $\vdash_{\mathcal{L}}^S$  is not necessarily finitary even if  $\vdash$  is.
2.  $\vdash_{\mathcal{L}}$  is just  $\vdash_{\mathcal{L}}^{S(\mathcal{L})}$  where  $S(\mathcal{L})$  is the set of all setups for  $\vdash_{\mathcal{L}}$ .

3. If  $S_1 \subseteq S_2$  then  $\vdash_{\mathcal{L}}^{S_2} \subseteq \vdash_{\mathcal{L}}^{S_1}$ .

#### EXAMPLES.

1. For classical propositional logic the standard semantics consists of the setups which are induced by some valuation in  $\{t, f\}$ . These setups can be characterized as theories  $T$  such that

$$(i) \quad \neg A \in T \text{ iff } A \notin T \quad (ii) \quad A \wedge B \in T \text{ iff both } A \in T \text{ and } B \in T$$

(and similar conditions for the other connectives).

2. In classical predicate logic we can define a setup in  $S$  to be any set of formulae which consists of the formulae which are true in some given first-order structure relative to some given assignment. Alternatively we can take a setup to consist of the formulae which are *valid* in some given first-order structure. In the first case  $\vdash^S = \vdash^t$ , in the second  $\vdash^S = \vdash^v$ , where  $\vdash^t$  and  $\vdash^v$  are the “truth” and “validity” consequence relations of classical logic (see [Avron, 1991a] for more details).

3. In modal logics we can define a “model” as the set of all the formulae which are true in some world in some Kripke frame according to some valuation. Alternatively, we can take a model as the set of all formulae which are valid in some Kripke frame, relative to some valuation. Again we get the two most usual consequence relations which are used in modal logics (see [Avron, 1991a] or [Fagin *et al.*, 1992]).

From now on the *following two conditions will be assumed in all our general definitions and propositions*:

1. The language contains a negation connective  $\neg$ .
2. For no  $A$  are both  $A$  and  $\neg A$  theorems of the logic.

DEFINITION. Let  $S$  be a semantics for a CR  $\vdash_{\mathcal{L}}$

1.  $\vdash_{\mathcal{L}}$  is strongly complete relative to  $S$  if  $\vdash_{\mathcal{L}}^S = \vdash_{\mathcal{L}}$ .
2.  $\vdash_{\mathcal{L}}$  is weakly complete relative to  $S$  if for all  $A$ ,  $\vdash_{\mathcal{L}} A$  iff  $\vdash_{\mathcal{L}}^S A$ .
3.  $\vdash_{\mathcal{L}}$  is  $c$ -complete relative to  $S$  if every consistent theory of  $\vdash_{\mathcal{L}}$  has a model in  $S$ .
4.  $\vdash_{\mathcal{L}}$  is strongly  $c$ -complete relative to  $S$  if for every  $A$  and every *consistent*  $T$ ,  $T \vdash_{\mathcal{L}}^S A$  iff  $T \vdash_{\mathcal{L}} A$ .

**Notes:**

1. Obviously, strong completeness implies strong  $c$ -completeness, while strong  $c$ -completeness implies both  $c$ -completeness and weak completeness.
2. Strong completeness means that deducibility in  $\vdash_{\mathcal{L}}$  is equivalent to semantic consequence in  $S$ . Weak completeness means that theoremhood in  $\vdash_{\mathcal{L}}$  (i.e., derivability from the empty set of assumptions) is equivalent to semantic validity (= truth in all models).  $c$ -completeness means that consistency implies satisfiability. It becomes identity if only consistent sets can be satisfiable, i.e., if  $\{\neg A, A\}$  has a model for no  $A$ . This is obviously too strong a demand for paraconsistent logics. Finally, strong  $c$ -completeness means that if we restrict ourselves to *normal* situations (i.e., consistent theories) then  $\vdash_{\mathcal{L}}$  and  $\vdash_{\mathcal{L}}^S$  are the same. This might sometimes be weaker than full strong completeness.

The last definition uses the concepts of “consistent” theory. The next definition clarifies (among other things) the meaning of this notion as we are going to use in it this paper.

DEFINITION. Let  $\mathcal{L}$  and  $\vdash_{\mathcal{L}}$  be as above. A theory in  $L$  *consistent* if for no  $A$  it is the case that  $T \vdash_{\mathcal{L}} A$  and  $T \vdash_{\mathcal{L}} \neg A$ , *complete* if for all  $A$ , either  $T \vdash_{\mathcal{L}} A$  or  $T \vdash_{\mathcal{L}} \neg A$ , *normal* if it is both consistent and complete.  $CS_{\mathcal{L}}$ ,  $CP_{\mathcal{L}}$  and  $N_{\mathcal{L}}$  will denote, respectively, the sets of its consistent, complete and normal theories.

Given  $\vdash_{\mathcal{L}}$ , the three classes,  $CS_{\mathcal{L}}$ ,  $CP_{\mathcal{L}}$  and  $N_{\mathcal{L}}$ , provide 3 different syntactically defined semantics for  $\vdash_{\mathcal{L}}$ , and 3 corresponding consequence relations  $\vdash_{\mathcal{L}}^{CS_{\mathcal{L}}}$ ,  $\vdash_{\mathcal{L}}^{CP_{\mathcal{L}}}$  and  $\vdash_{\mathcal{L}}^{N_{\mathcal{L}}}$ . We shall henceforth denote these CRs by  $\vdash_{\mathcal{L}}^{CS}$ ,  $\vdash_{\mathcal{L}}^{CP}$  and  $\vdash_{\mathcal{L}}^N$ , respectively. Obviously,  $\vdash_{\mathcal{L}}^{CS} \subseteq \vdash_{\mathcal{L}}^N$  and  $\vdash_{\mathcal{L}}^{CP} \subseteq \vdash_{\mathcal{L}}^N$ . In the rest of this section we investigate these relations and the completeness properties they induce.

Let us start with the easier case: that of  $\vdash_{\mathcal{L}}^{CS}$ . It immediately follows from the definitions (and our assumptions) that relative to it every logic is strongly  $c$ -complete (and so also  $c$ -complete and weakly complete). Hence the only completeness notion it induces is the following:

DEFINITION. A logic  $\mathcal{L}$  with a consequence relation  $\vdash_{\mathcal{L}}$  is strongly consistent if  $\vdash_{\mathcal{L}}^{CS} = \vdash_{\mathcal{L}}$ .

$\vdash_{\mathcal{L}}^{CS}$  is not a really interesting CR. As the next theorem shows, what it does is just to trivialize inconsistent  $\vdash_{\mathcal{L}}$ -theories. Strong consistency, accordingly, might not be a desirable property, certainly not a property that any logic with negation should have.

## PROPOSITION 7.

1.  $T \vdash_{\mathcal{L}}^{CS} A$  iff either  $T$  is inconsistent in  $\mathcal{L}$  or  $T \vdash_{\mathcal{L}} A$ . In particular,  $T$  is  $\vdash_{\mathcal{L}}^{CS}$ -consistent iff it is  $\vdash_{\mathcal{L}}$ -consistent.
2.  $\mathcal{L}$  is strongly consistent iff  $\neg A, A \vdash_{\mathcal{L}} B$  for all  $A, B$  (iff  $T$  is consistent whenever  $T \not\vdash_{\mathcal{L}} A$ ).
3. Let  $\mathcal{L}^{CS}$  be obtained from  $\mathcal{L}$  by adding the rule: from  $\neg A$  and  $A$  infer  $B$ . Then  $\vdash_{\mathcal{L}}^{CS} = \vdash_{\mathcal{L}^{CS}}$ . In particular: if  $\vdash_{\mathcal{L}}$  is finitary then so is  $\vdash_{\mathcal{L}}^{CS}$ .
4.  $\vdash_{\mathcal{L}}^{CS}$  is strongly consistent.

We turn now to  $\vdash^{CP}$  and  $\vdash^N$ . In principle, each provides 4 notions of completeness. We don't believe, however, that considering the two notions of  $c$ -consistency is natural or interesting in the framework of  $\vdash^{CP}$  ( $c$ -completeness, e.g., means there that every consistent theory has a complete extension, but that extension might not be consistent itself). Accordingly we shall deal with the following 6 notions of syntactical completeness.<sup>7</sup>

DEFINITION. Let  $\mathcal{L}$  be a logic and let  $\vdash_{\mathcal{L}}$  be its consequence relation.

1.  $\mathcal{L}$  is *strongly complete* if it is strongly complete relative to  $CP$ .
2.  $\mathcal{L}$  is *weakly complete* if it is weakly complete relative to  $CP$ .
3.  $\mathcal{L}$  is *strongly normal* if it is strongly complete relative to  $N$ .
4.  $\mathcal{L}$  is *weakly normal* if it is weakly complete relative to  $N$ .
5.  $\mathcal{L}$  is  *$c$ -normal* if it is  $c$ -complete relative to  $N$ .
6.  $\mathcal{L}$  is *strongly  $c$ -normal* if it is strongly  $c$ -complete relative to  $N$  (this is easily seen to be equivalent to  $\vdash_{\mathcal{L}}^N = \vdash_{\mathcal{L}}^{CS}$ ).

For the reader's convenience we repeat what these definitions actually mean:

1.  $\mathcal{L}$  is strongly complete iff whenever  $T \not\vdash_{\mathcal{L}} A$  there exists a complete extension  $T^*$  of  $T$  such that  $T^* \not\vdash_{\mathcal{L}} A$ .
2.  $\mathcal{L}$  is weakly complete iff whenever  $A$  is not a theorem of  $\mathcal{L}$  there exists a complete  $T^*$  such that  $T^* \not\vdash_{\mathcal{L}} A$ .
3.  $\mathcal{L}$  is strongly normal iff whenever  $T \not\vdash_{\mathcal{L}} A$  there exists a complete and consistent extension  $T^*$  of  $T$  such that  $T^* \not\vdash_{\mathcal{L}} A$ .
4.  $\mathcal{L}$  is weakly normal iff whenever  $A$  is not a theorem of  $\mathcal{L}$  there exists a complete and consistent theory  $T^*$  such that  $T^* \not\vdash_{\mathcal{L}} A$ .

<sup>7</sup>In [Anderson and Belnap, 1975] the term "syntactically complete" was used for what we call below "strongly  $c$ -normal".

5.  $\mathcal{L}$  is  $c$ -normal if every consistent theory of  $\mathcal{L}$  has a complete and consistent extension.
6.  $\mathcal{L}$  is strongly  $c$ -normal iff whenever  $T$  is consistent and  $T \not\vdash_{\mathcal{L}} A$  there exists a complete and consistent extension  $T^*$  of  $T$  such that  $T^* \not\vdash_{\mathcal{L}} A$ .

Our next proposition provides simpler syntactical characterizations of some of these notions in case  $\vdash_{\mathcal{L}}$  is finitary.

**PROPOSITION 8.** *Assume that  $\vdash_{\mathcal{L}}$  is finitary.*

1.  $\mathcal{L}$  is strongly complete iff for all  $T, A$  and  $B$ :

$$(*) \quad T, A \vdash_{\mathcal{L}} B \quad \text{and} \quad T, \neg A \vdash_{\mathcal{L}} B \quad \text{imply} \quad T \vdash_{\mathcal{L}} B$$

*In case  $\mathcal{L}$  has a combining disjunction  $\vee$  then  $(*)$  is equivalent to the theoremhood of  $\neg A \vee A$  (excluded middle).*

2.  $\mathcal{L}$  is strongly normal if for all  $T$  and  $A$ :

$$(**) \quad T \vdash_{\mathcal{L}} A \quad \text{iff} \quad T \cup \{\neg A\} \quad \text{is inconsistent.}$$

3.  $\mathcal{L}$  is strongly  $c$ -normal iff  $(**)$  obtains for every consistent  $T$ .

4.  $\mathcal{L}$  is  $c$ -normal iff for every consistent  $T$  and every  $A$  either  $T \cup \{A\}$  or  $T \cup \{\neg A\}$  is consistent.

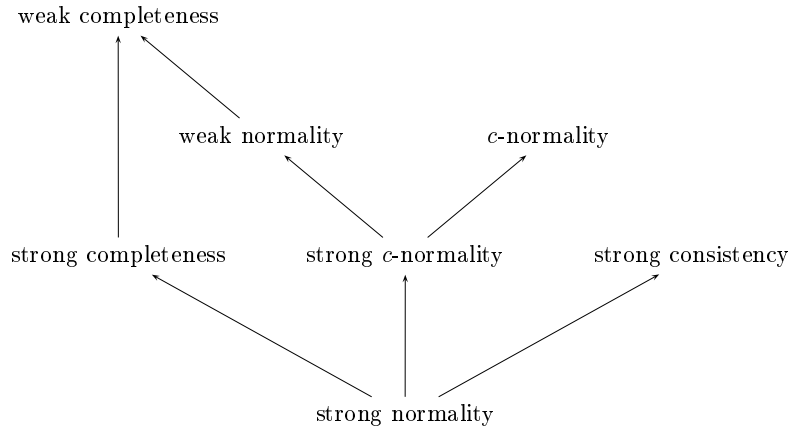
**Proof.** Obviously, strong completeness implies  $(*)$ . For the converse, assume that  $T \not\vdash B$ . Using  $(*)$ , we extend  $T$  in stages to a complete theory such that  $T^* \not\vdash B$ . This proves part 1. The other parts are straightforward. ■

#### COROLLARIES.

1. If  $\mathcal{L}$  is strongly normal then it is strongly symmetric w.r.t.  $\neg$ . Moreover:  $\vdash_{\mathcal{L}}^s$  is an ordinary multiple-conclusion CR.
2. If  $\mathcal{L}$  is strongly symmetric w.r.t.  $\neg$  then it is strongly complete iff  $\vdash_{\mathcal{L}}^s$  is closed under contraction.

**Proof.** These results easily follows from the last proposition and Theorems 2, 4 and 5 above. ■

In the figure below we display the obvious relations between the seven properties of logics which were introduced here (where an arrow means “contained in”). The next theorem shows that no arrow can be added to it:



**THEOREM 9.** *A logic can be:*

1. *strongly consistent and c-normal without even being weakly complete*
2. *strongly complete and strongly c-normal without being strongly consistent (and so without being strongly normal)*
3. *strongly consistent without being c-normal*
4. *strongly complete, weakly normal and c-normal without being strongly c-normal*
5. *strongly complete and c-normal without being weakly normal*
6. *strongly consistent, c-normal and weakly normal without being strongly c-normal (=strongly normal in this case, because of strong consistency)*
7. *strongly complete without being c-normal.*<sup>8</sup>

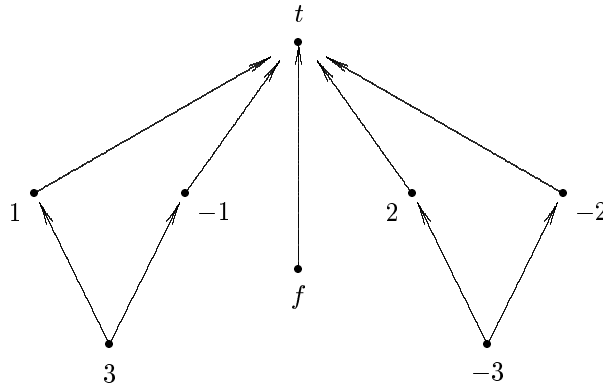
**Proof.** Appropriate examples for 1-6 are given below, respectively, in theorems 12, 18, 33, 19, 35 and the corollary to theorem 19. As for the last part, let  $\mathcal{L}$  be the following system in the language of  $\{\neg, \rightarrow\}$ :<sup>9</sup>

<sup>8</sup>Hence the two standard formulations of the “strong consistency” of classical logic are *not* equivalent in general.

<sup>9</sup>Classical logic is obtained from it by adding  $\neg A \rightarrow (A \rightarrow B)$  as axiom (see [Epstein, 1995, Ch. 2L]).

- $Ax1:$      $A \rightarrow (B \rightarrow A)$
- $Ax2:$      $A \rightarrow (B \rightarrow C) \rightarrow (A \rightarrow B) \rightarrow (A \rightarrow C)$
- $Ax3:$      $(\neg A \rightarrow B) \rightarrow ((A \rightarrow B) \rightarrow B)$
- (MP)      $\frac{A \quad A \rightarrow B}{B}$  .

Obviously, the deduction theorem for  $\rightarrow$  holds for this system, since  $MP$  is the only rule of inference, and we have  $Ax1$  and  $Ax2$ . This fact,  $Ax3$  and proposition 8 guarantee that it is strongly complete. To show that it is not  $c$ -normal, we consider the theory  $T_0 = \{p \rightarrow q, p \rightarrow \neg q, \neg p \rightarrow r, \neg p \rightarrow \neg r\}$ . Obviously,  $T_0$  has no complete and consistent extension. We show that it is consistent nevertheless. For this we use the following structure:



Define in this structure  $a \rightarrow b$  as  $t$  if  $a \leq b$ ,  $b$  otherwise,  $\neg x$  as  $f$  if  $x = t$ ,  $t$  if  $x = f$  and  $\neg x$  otherwise. It is not difficult now to show that if  $T \vdash A$  in the present logic for some  $T$  and  $A$ , and  $v$  is a valuation in this structure such that  $v(B) = t$  for all  $B \in T$ , then  $v(A) = t$ . Take now  $v(p) = 3$ ,  $v(q) = 1$ ,  $v(r) = 2$ . Then  $v(B) = t$  for all  $B \in T_0$ , but obviously there is no  $A$  such that  $v(A) = v(\neg A) = t$ . Hence  $T_0$  is consistent. ■

We end this introductory subsection with a characterization of  $\vdash_{\mathcal{L}}^{CP}$  and  $\vdash_{\mathcal{L}}^N$ . The proofs are left to the reader.

**PROPOSITION 10.**

1.  $\vdash_{\mathcal{L}}^{CP}$  is strongly complete, and is contained in any strongly complete extension of  $\vdash_{\mathcal{L}}$ .
2. Suppose  $\vdash_{\mathcal{L}}$  is finitary.  $T \vdash_{\mathcal{L}}^{CP} A$  iff for some  $B_1, \dots, B_n$  ( $n \geq 0$ ) we have that  $T \cup \{B_1^*, \dots, B_n^*\} \vdash_{\mathcal{L}} A$  for every set  $\{B_1^*, \dots, B_n^*\}$  such that for all  $i$ ,  $B_i^* = B_i$  or  $B_i^* = \neg B_i$ .

3. If  $\vdash_{\mathcal{L}}$  is finitary, then so is  $\vdash_{\mathcal{L}}^{CP}$ .

PROPOSITION 11.

1.  $\vdash_{\mathcal{L}}^N$  is strongly normal, and is contained in every strongly normal extension of  $\vdash_{\mathcal{L}}$ .
2. If  $\vdash_{\mathcal{L}}$  is finitary then  $T \vdash_{\mathcal{L}}^N A$  iff for some  $B_1, \dots, B_n$  we have that for all  $\{B_1^*, \dots, B_n^*\}$  where  $B_i^* \in \{B_i, \neg B_i\}$  ( $i = 1, \dots, n$ ), either  $T \cup \{B_1^*, \dots, B_n^*\}$  is inconsistent or  $T \cup \{B_1^*, \dots, B_n^*\} \vdash_{\mathcal{L}} A$
3.  $\vdash_{\mathcal{L}}^N$  is finitary if  $\vdash_{\mathcal{L}}$  is.

### 3.2 Classical and Intuitionistic Logics

Obviously, classical propositional logic is strongly normal. In fact, most of the proofs of the completeness of classical logic relative to its standard two-valued semantics begin with demonstrating the condition (\*\*) in Proposition 8, and are based on the fact that every complete and consistent theory determines a unique valuation in  $\{t, f\}$  - and vice versa. In other words:  $N$  here is exactly the usual semantics of classical logic, only it can be characterized also using an especially simple algebraic structure (and valuations in it). One can argue that this strong normality *characterizes* classical logic. To be specific, it is not difficult to show the following claims:

1. classical logic is the only logic in the language of  $\{\neg, \wedge\}$  which is strongly normal w.r.t.  $\neg$  and for which  $\wedge$  is an internal conjunction. Similar claims hold for the  $\{\neg, \rightarrow\}$  language, if we demand  $\rightarrow$  to be an internal implication and for the  $\{\neg, \vee\}$  language, if we demand  $\vee$  to be a combining disjunction.
2. Any logic which is strongly normal and has either an internal implication, or an internal conjunction or a combining disjunction contains classical propositional logic.

The next proposition summarizes the relevant facts concerning intuitionistic logic. The obvious conclusion is that although the official intuitionistic negation has some features of negation, it still lacks most. Hence, it cannot be taken as a real negation from our semantic point of view.

PROPOSITION 12. *Intuitionistic logic is strongly consistent and c-normal, but it is not even weakly complete.*

**Proof.** Strong consistency follows from part 3 of Proposition 7. *c*-normality follows from part 4 of Proposition 8, since in intuitionistic logic if both  $T \cup \{A\}$  and  $T \cup \{\neg A\}$  are inconsistent then  $T \vdash_H \neg A$  and  $T \vdash_H \neg \neg A$ , and so  $T$  is inconsistent. Finally,  $\neg A \vee A$  belongs to every complete setup, but is not intuitionistically valid. ■



**Note:** Intuitionistic logic and classical logic have exactly the same consistent and complete setups, since any complete intuitionistic theory is closed under the elimination rule of double negation. Hence any consistent intuitionistic theory has a classical two-valued model.

What about fragments (with negation) of Intuitionistic Logic? Well, they are also strongly consistent and  $c$ -normal, by the same proof. Moreover,  $((A \rightarrow B) \rightarrow A) \rightarrow A$  is another example of a sentence which belongs to every complete setup (since  $A \vdash_H ((A \rightarrow B) \rightarrow A) \rightarrow A$  and  $\neg A \vdash_H ((A \rightarrow B) \rightarrow A) \rightarrow A$ ), but is not provable. The set of theorems of the pure  $\{\neg, \wedge\}$  fragment, on the other hand, is identical to that of classical logic, as is well known. This fragment is, therefore, easily seen to be weakly normal. It is still neither strongly complete nor strongly  $c$ -normal, since  $\neg\neg A \vdash_H^{CP} A$ . ■

Finally, we note the important fact that classical logic can be viewed as the completion of intuitionistic logic. More precisely:

PROPOSITION 13.

1.  $\vdash_H^{CS} = \vdash_H$
2.  $\vdash_H^{CP} = \vdash_H^N = \text{classical logic}$ .

**Proof.**

2.  $\vdash_L^{CP} = \vdash_L^N$  whenever  $L$  is strongly consistent (i.e., all nontrivial theories are consistent). In the proof of the previous proposition we have seen also that  $\vdash_H^{CP} \neg A \vee A$  and  $\vdash_H^{CP} ((A \rightarrow B) \rightarrow A) \rightarrow A$ . It is well known, however, that by adding either of these schemes to intuitionistic logic we get classical logic. Hence classical logic is contained in  $\vdash_H^{CP}$ . Since classical logic is already strongly complete,  $\vdash_H^{CP}$  is exactly classical logic. (Note that this is true for any fragment of the language which includes negation.) ■

### 3.3 Linear Logic (LL)

In the next 3 subsections we are going to investigate some known substructural logics [Schroeder-Heister and Došen, 1993]. Before doing it we must emphasize again that in this section it is only the external, Tarskian consequence relation of these logics which can be relevant. This consequence relation can very naturally be defined by using the standard Hilbert-type formulations of these logics:  $A_1, \dots, A_n \vdash_{\mathcal{L}}^e B$  ( $\mathcal{L} = LL, R, RM, RMI$ , etc.) iff there exists an ordinary deduction of  $B$  from  $A_1, \dots, A_n$  in the corresponding Hilbert-type system. This definition is insensitive to the exact choice of axioms (or even rules), provided we take all the rules as rules of derivation and not just as rules of proof. In the case of Linear Logic one can use for this the systems given in [Avron, 1988] or in [Troelstra, 1992]. An

alternative equivalent definition of the various external CRs can be given using the standard Gentzen-types systems for these logics (in case such exist), as explained in the introduction. Still another characterization in the case of Linear Logic can be given using the phase semantics of [Girard, 1987]:  $A_1, \dots, A_n \vdash_{LL}^e B$  iff  $B$  is true in every phase model of  $A_1, \dots, A_n$ . In what follows we shall omit the superscript “ $e$ ” and write just  $\vdash_{LL}, \vdash_{LL_m}$ , etc.

Unlike in [Girard, 1987] we shall take below negation as one of the connectives of the language of linear logic and write  $\neg A$  for the negation of  $A$  (this corresponds to Girard’s  $A^\perp$ ). As in [Avron, 1988] and in the relevance logic literature, we use arrow ( $\rightarrow$ ) for linear implication.

We show now that linear logic is incomplete with respect to our various notions.

**PROPOSITION 14.**  *$LL_m (LL_{ma}, LL)$  is not strongly consistent.*

**PROPOSITION 15.**  *$LL_m (LL_{ma}, LL)$  is neither strongly complete nor  $c$ -normal.*

**Proof.** Consider the following theory:

$$T = \{p \rightarrow \neg p, \neg p \rightarrow p\} .$$

From the characterization of  $\vdash_{LL_m}$  given in [Avron, 1992] it easily follows that has  $T$  been inconsistent then there would be a provable sequent of the form:  $\neg p \rightarrow p, \neg p \rightarrow p, \dots, \neg p \rightarrow p, p \rightarrow \neg p, \dots, p \rightarrow \neg p \Rightarrow$ . But in any cut-free proof of such a sequent the premises of the last applied rule should have an odd number of occurrences of  $p$ , which is impossible in a provable sequent of the purely multiplicative linear logic. Hence  $T$  is consistent. Obviously, every complete extension of  $T$  proves  $p$  and  $\neg p$  and so is inconsistent. This shows that  $LL_m$  is not  $c$ -normal. It also shows that  $p$  is not provable from  $T$ , although it is provable from any complete extension of it, and so  $LL_m$  is not strongly complete. ■

**PROPOSITION 16.**  *$LL_{ma}$  (and so also  $LL$ ) is not weakly complete.*

**Proof.**  $\sim A \oplus A$  is not a theorem of linear logic, but it belongs to any complete theory. ■

It follows that Linear logic (and its multiplicative-additive fragment) has none of the properties we have defined in this section. Its negation is therefore not really a negation from our present *semantic* point of view.

Our results still leave the possibility that  $LL_m$  might be weakly complete or even weakly normal. We conjecture that it is not, but we have no counterexample.

We end this section by giving axiomatizations of  $\vdash_{LL}^{CP}$  and  $\vdash_{LL}^N$ .

## PROPOSITION 17.

1. Let  $LL^{CP}$  be the full Hilbert-type system for linear logic (as given in [Avron, 1988]) together with the rule: from  $!A \rightarrow B$  and  $!\neg A \rightarrow B$  infer  $B$ . Then  $\vdash_{LL}^{CP} = \vdash_{LL^{CP}}$ .
2. Let  $LL^N$  be  $LL^{CP}$  together with the disjunctive syllogism for  $\oplus$  (from  $\neg A$  and  $A \oplus B$  infer  $B$ ). Then  $\vdash_{LL}^N = \vdash_{LL^N}$ .

**Proof.**

1. The necessitation rule (from  $A$  infer  $!A$ ) is one of the rules of  $LL$ .<sup>10</sup> It follows therefore that  $B$  should belong to any complete setup which contains both  $!A \rightarrow B$  and  $!\neg A \rightarrow B$ . Hence the new rule is valid for  $\vdash_{LL}^{CP}$  and  $\vdash_{LL^{CP}} \subseteq \vdash_{LL}^{CP}$ .

For the converse, assume  $T \vdash_{LL}^{CP} A$ . Then there exist  $B_1, \dots, B_n$  like in proposition 10(2). We prove by induction on  $n$  that  $T \vdash_{LL^{CP}} A$ . The case  $n = 0$  is obvious. Suppose the claim is true for  $n - 1$ . We show it for  $n$ . By the deduction theorem for  $LL$ ,  $!B_1^*, \dots, !B_n^* \Rightarrow A$  is derivable from  $T$  in  $LL^{CP}$ .<sup>11</sup> More precisely:  $!B_1^* \otimes !B_2^* \dots \otimes !B_n^* \rightarrow A$  is derivable from  $T$  for any choice of  $B_1^*, \dots, B_n^*$ . Since  $!C \otimes !D \leftrightarrow !(C \& D)$  is a theorem of  $LL$ , this means that both  $!B_n \rightarrow (!(B_1^* \& \dots \& B_{n-1}^*) \rightarrow A)$  and  $!\neg B_n \rightarrow (!(B_1^* \& \dots \& B_{n-1}^*) \rightarrow A)$ . By the new rule of  $LL^{CP}$  we get therefore that  $T \vdash_{LL^{CP}} !(B_1^* \& \dots \& B_{n-1}^*) \rightarrow A$ , and so  $T \vdash_{LL^{CP}} !B_1^* \otimes !B_2^* \otimes \dots \otimes !B_{n-1}^* \rightarrow A$  for all choices of  $B_1^*, \dots, B_{n-1}^*$ . An application of the induction hypothesis gives  $T \vdash_{LL^{CP}} A$ .

2. The proof is similar, only this time we should have (by proposition 11) that  $T \cup \{B_1^*, \dots, B_n^*\}$  is either inconsistent in  $L^N$  or proves  $A$  there. In both cases it proves  $A \oplus \perp$  in  $LL^{CP}$ . The same argument as before will show that  $T \vdash_{LL^{CP}} A \oplus \perp$ . Since  $\vdash_{LL} \neg \perp$ , one application of the disjunctive syllogism will give  $T \vdash_{LL^{CP}} A$ . It remains to show that the disjunctive syllogism is valid for  $\vdash_{LL}^N$ . This is easy, since  $\{\neg A, A \oplus B, \neg B\}$  is inconsistent in  $LL$ , and so any complete and consistent extension of  $\{\neg A, A \oplus B\}$  necessarily contains  $B$ . ■

### 3.4 The Standard Relevance Logic $R$ and its Relatives

In this section we investigate the standard relevance logic  $R$  of Anderson and Belnap [Anderson and Belnap, 1975; Dunn, 1986] and its various extensions and fragments. Before doing this we should again remind the reader what consequence relation we have in mind: the ordinary one which is associated

<sup>10</sup>Note again that we are talking here about  $\vdash_{LL}^c$ !

<sup>11</sup>In fact, at the beginning it is derivable from  $T$  in  $LL$ , but for the induction to go through we need to assume derivability in  $LL^{CP}$  at each step.

with the standard Hilbert-type formulations of these logics. As in the case of linear logic, this means that we take both rules of  $R$  ( $MP$  and adjunction) as rules of derivation and define  $T \vdash_R A$  in the most straightforward way.

Let us begin with the purely intensional (=multiplicative) fragment of  $R$ :  $R_m$ . We state the results for this system, but they hold for all its nonclassical various extensions (by axioms) which are discussed in the literature.

**THEOREM 18.**  *$R_m$  is not strongly consistent, but it is strongly complete and strongly  $c$ -normal.*

**Proof.** It is well-known that  $R_m$  is not strongly consistent in our sense. Its main property that we need for the other claims is that  $T, A \vdash_{R_m} B$  iff either  $T \vdash_{R_m} B$  or  $T \vdash_{R_m} A \rightarrow B$ . The strong completeness of  $R_m$  follows from this property by the provability of  $(\neg A \rightarrow B) \rightarrow ((A \rightarrow B) \rightarrow B)$  and proposition 8(1).

To show strong  $c$ -normality, we note first that a theory  $T$  is inconsistent in  $R_m$  iff  $T \vdash_{R_m} \neg(B \rightarrow B)$  for some  $B$  (because  $\vdash_{R_m} \neg B \rightarrow (B \rightarrow \neg(B \rightarrow B))$ ). Suppose now that  $T$  is consistent and  $T \not\vdash_{R_m} A$ . Were  $T \cup \{\neg A\}$  inconsistent then by the same main property and the consistency of  $T$  we would have that  $T \vdash_{R_m} \neg A \rightarrow \neg(B \rightarrow B)$  for some  $B$ , and so that  $T \vdash_{R_m} (B \rightarrow B) \rightarrow A$  and  $T \vdash_{R_m} A$ . A contradiction. Hence  $T \cup \{\neg A\}$  is consistent and we are done by proposition 8(3). ■

The last theorem is the optimal theorem concerning negation that one can expect from a logic which was designed to be paraconsistent. It shows that with respect to normal “situations” (i.e., consistent theories) the negation connective of  $R_m$  behaves exactly as in classical logic. The difference, therefore, is mainly w.r.t. inconsistent theories. Unlike classical logic they are not necessarily trivial in  $R_m$ . Strong completeness means, though, that excluded middle, at least, can be assumed even in the abnormal situations.

When we come to  $R$  as a whole the situation is not as good as for the purely intensional fragments. Strong  $c$ -normality is lost. What we do have is the following:

**THEOREM 19.**  *$R$  is strongly complete,  $c$ -normal and weakly normal,<sup>12</sup> but it is neither strongly consistent nor strongly  $c$ -normal.*

**Proof.** Obviously,  $R$  is not strongly consistent. It is also well known that  $\neg p, p \vee q \not\vdash_R q$ . Still  $q$  belongs to any complete and consistent extension of the (even classically!) consistent theory  $\{\neg p, p \vee q\}$ , since  $\{\neg p, p \vee q, \neg q\}$  is not consistent in  $R$ . It follows that  $R$  is not strongly  $c$ -normal. On the other hand, to any extension  $\mathcal{L}$  of  $R$  by axiom schemes it is true that if  $T, A \vdash_{\mathcal{L}} C$  and  $T, B \vdash_{\mathcal{L}} C$ , then  $T, A \vee B \vdash_{\mathcal{L}} C$  [Anderson and Belnap, 1975]. Since  $\vdash_R A \vee \neg A$ , this and proposition 8(1) entail that any such extension is strongly complete. Suppose, next, that  $T$  is theory and  $A$  a

<sup>12</sup>Weak normality is proved in [Anderson and Belnap, 1975] under the name “syntactical completeness”.

formula such that  $T \cup \{A\}$  and  $T \cup \{\neg A\}$  are inconsistent ( $\mathcal{L}$  as above). Then for some  $B$  and  $C$  it is the case that  $T, A \vdash_{\mathcal{L}} \neg B \wedge B$  and  $T, \neg A \vdash_{\mathcal{L}} \neg C \wedge C$ . It follows that  $T, A \vee \neg A \vdash_{\mathcal{L}} (\neg B \wedge B) \vee (\neg C \wedge C)$ . Since  $A \vee \neg A$  and  $\neg[(\neg B \wedge B) \vee (\neg C \wedge C)]$  are both theorems of  $R$ ,  $T$  is inconsistent in  $\mathcal{L}$ . By proposition 8(4) this shows that any such logic is  $c$ -normal. Suppose, finally, that  $\not\vdash_R A$ . Had  $\{\neg A\}$  been inconsistent, we would have that for some  $B$ ,  $\neg A \vdash_R \neg B \wedge B$ . This, in turn, entails that  $A \vee \neg A \vdash_R A \vee (\neg B \wedge B)$ , and so that  $\vdash_R A \vee (\neg B \wedge B)$ . On the other hand,  $\vdash_R \neg(\neg B \wedge B)$ . By the famous theorem of Meyer and Dunn concerning the admissibility of the disjunctive syllogism in  $R$  [Anderson and Belnap, 1975; Dunn, 1986] it would follow, therefore, that  $\vdash_R A$ , contradicting our assumption. Hence  $\{\neg A\}$  is consistent, and so, by the  $c$ -normality of  $R$  which we have just proved, it has a consistent and complete extension which obviously does not prove  $A$ . This shows that  $R$  is weakly normal (the proof for  $RM$  is identical). ■

**COROLLARY.**  $\vdash_R^{CS}$  is strongly consistent,  $c$ -normal and weakly normal, but it is not strongly  $c$ -normal.

**Note:** A close examination of the proof of the last theorem shows that the properties of  $R$  which are described there are shared by many of its relatives (like  $RM$ , for example). We have, in fact, the following generalizations:

1. Every extension of  $R$  which is not strongly consistent is also not strongly  $c$ -normal.
2. Every extension of  $R$  by axiom-schemes is both strongly complete and  $c$ -normal.
3. Every extension of  $R$  by axiom schemes for which the disjunctive syllogism is an admissible rule<sup>13</sup> is weakly normal.

In fact, (1)–(3) are true (with similar proofs) also for many systems weaker than  $R$  in the relevance family, like  $E$ .

Our results show that  $\vdash_R^{CP} = \vdash_R$ , but  $\vdash_R^N \neq \vdash_R^{CS}$  (since  $R$  is not strongly  $c$ -normal). Hence  $\vdash_R^N$  is a new consequence relation, and we turn next to axiomatize it.

**DEFINITION.** Let  $\mathcal{L}$  be an extension of  $R$  by axiom schemes and let  $\mathcal{L}^N$  be the system which is obtained from  $\mathcal{L}$  by adding to it the disjunctive syllogism ( $\gamma$ ) as an extra rule: from  $\neg A$  and  $A \vee B$  infer  $B$ .

**THEOREM 20.**  $\vdash_{\mathcal{L}}^N = \vdash_{\mathcal{L}^N}$ .

**Proof.** To show that  $\vdash_{\mathcal{L}^N} \subseteq \vdash_{\mathcal{L}}^N$  it is enough to show that  $\neg A, A \vee B \vdash_{\mathcal{L}}^N B$ . This was already done, in fact, in the proof of the last theorem. For the converse, assume  $T \vdash_{\mathcal{L}}^N A$ . Since  $\mathcal{L}$  is  $c$ -normal (see last note),  $T \cup \{\neg A\}$

<sup>13</sup>See [Anderson and Belnap, 1975] and [Dunn, 1986] for examples and criteria when this is the case.

cannot be  $\mathcal{L}$ -consistent. Hence  $T \cup \{\neg A\} \vdash_{\mathcal{L}} \neg B \wedge B$  for some  $B$ . This entails that  $T \vdash_{\mathcal{L}} A \vee (\neg B \wedge B)$  and that  $T \vdash_{\mathcal{L}^N} A$  exactly as in the proof of the weak normality of  $R$ . ■

### 3.5 The Purely Relevant Logic $RMI$

The purely relevant logic  $RMI$  was introduced in citeAv90a,Av90b. Proof-theoretically it differs from  $R$  in that:

- (i) The converse of contraction (or, equivalently, the mingle axiom of  $RM$ ) is valid in it. This is equivalent to the idempotency of the intensional disjunction  $+$  (=“par” of Girard). In the purely multiplicative fragment  $RMI_m$  it means also that assumptions with respect to  $\rightarrow$  can be taken as coming in *sets* (rather than multisets, as in  $LL_m$  or  $R_m$ ).
- (ii) The adjunction rule  $(B, C \vdash B \wedge C)$  as well as the distribution axiom  $(A \wedge (B \vee C) \rightarrow (A \wedge B) \vee (A \wedge C))$  are accepted only if  $B$  and  $C$  are “relevant”. This relevance relation can be expressed in the logic by the sentence  $R^+(A, B) = (A \rightarrow A) + (B \rightarrow B)$ , which should be added as an extra premise to adjunction and distribution (this sentence is the counterpart of the “mix” rule of [Girard, 1987]).

We start our investigation with the easier case of  $RMI_m$ .

**THEOREM 21.** *Exactly like  $R_m$ ,  $RMI_m$  is not strongly consistent, but it is both strongly complete and strongly c-normal.*

**Proof.** Exactly like in the case of  $R_m$ . ■

Like in classical logic, and unlike the case of  $R_m$ , these two main properties of  $RMI_m$  are strongly related to simple, intuitive, algebraic semantics. Originally, in fact,  $RMI_m$  was designed to correspond to a class of structures which are called in [Avron, 1990a] “full relevant disjunctive lattices” (full r.d.l.). A full r.d.l. is a structure which results if we take a tree and attach to each node  $b$  its own two basic truth-values  $\{t_b, f_b\}$ . To a leaf  $b$  of the tree we *can* attach instead a single truth-value  $I_b$  which is the negation of itself (its meaning is “both true and false” or “degenerate”).  $b$  is called abnormal in this case. Intuitively, the nodes of the tree represent “domains of discourse”. Two domains are relevant to each other if they have a common branch, while  $b$  being nearer than  $a$  to the root on a branch intuitively means that  $b$  has a higher “degree of reality” (or higher “degree of significance”) than  $a$  (we write  $a < b$  in this case). The operation of  $\neg$  (negation) is defined on a full r.d.l.  $M$  in the obvious way, while  $+$  (relevant disjunction) is defined as follows: Let  $|t_a| = |f_a| = |I_a| = a$ , and let  $\text{val}(t_b) = t$ ,  $\text{val}(f_b) = f$  and  $\text{val}(I_b) = I$ . Define  $x_{\leq+} y$  if either  $x = y$  or  $|x| < |y|$  or  $|x| = |y|$  and  $\text{val}(y) = t$ .  $(M, \leq_+)$  is an upper semilattice. Let

$x + y = \sup_{\leq_+} (x, y)$ . An  $RM I_m$ -model is a pair  $(M, v)$  where  $M$  is a full r.d.l. and  $v$  a valuation in it (which respects the operations). A sentence  $A$  is true in a model  $(M, v)$  if  $\text{val}(v(A)) \neq f$ . Obviously, every model  $(M, v)$  determines an  $RM I_m$ -setup of all the formulae which are true in it. Denote the collection of all these setups by  $RDL_m$ .

**PROPOSITION 22.**  $CP_{RM I_m} = RDL_m$

**Proof.** It is shown in [Avron, 1990b] that the Lindenbaum algebra of any complete  $RM I_m$ -theory determines a model in which exactly its sentences are true. This implies that  $CP_{RM I_m} \subseteq RDL_m$ . The converse is obvious from the definitions. ■

**CROLLARY.** [Avron, 1990b]:  $RM I_m$  is sound and complete for the semantics of full r.d.l.s. In other words:  $T \vdash_{RM I_m} A$  iff  $A$  is true in every model of  $T$ .

**Proof.** Checking soundness is straightforward, while completeness follows from the syntactic strong completeness of  $RM I_m$  (theorem 21) and the last theorem. ■

The strong  $c$ -normality of  $RM I_m$  also has an interpretation in terms of the semantics of full r.d.l.s. In order to describe it we need first some definitions:

**DEFINITION.**

1. A full r.d.l is consistent iff for every  $x$  in it  $\text{val}(x) \in \{t, f\}$  (i.e., the intermediate truth-value  $I$  is not used in its construction). This is equivalent to:  $x \neq \neg x$  for all  $x$ .
2. A model  $(M, v)$  is consistent iff  $M$  is consistent.
3.  $CRDL_m$  is the collection of the  $RM I_m$ -setups which are determined by some consistent model.

**Note:** On every tree one can base exactly one consistent full r.d.l. (but in general many inconsistent ones).

**PROPOSITION 23.**  $N_{RM I_m} = CRDL_m$ .

**Proof.** In the construction from [Avron, 1990b] which is mentioned in the proof of proposition 22, a complete and consistent theory is easily seen to determine a consistent model. The converse is obvious. ■

In view of the last proposition, the strong  $c$ -normality of  $RM I_m$  and its two obvious corollaries (weak normality and  $c$ -normality) can be reformulated in terms of the algebraic models as follows:

**PROPOSITION 24.**

1. If  $T$  is consistent then  $T \vdash_{RM I_m} A$  iff  $A$  is true in any consistent model of  $T$ .

2.  $\vdash_{RMI_m} A$  iff  $A$  is true in any consistent model.
3. Every consistent  $RMI_m$ -theory has a consistent model.

It follows that if we restrict our attention to consistent  $RMI_m$ -theories, we can also restrict our semantics to consistent full r.d.l.s, needing, therefore, only the classical two truth-values  $t$  and  $f$ , but not  $I$ .

Exactly as in the case of  $R$ , when we pass to  $RMI$  things become more complicated. Moreover, although we are going to show that  $RMI$  has *exactly* the same properties as  $R$ , the proofs are harder.

**THEOREM 25.**  *$RMI$  is strongly complete.*

**Proof.** The proof is like the one for  $R$  given above, since  $RMI$  has the relevant properties of  $R$  which were used there (see [Avron, 1990b]). ■

Like in the case of  $RMI_m$ , the strong completeness of  $RMI$  is directly connected to the semantics of full r.d.l.s. This semantics is extended in [Avron, 1990a; Avron, 1990b] to the full language by defining the operator  $\wedge$  on a full r.d.l. as follows: define  $\leq$  on  $M$  by:  $x \leq y$  iff  $\text{val}(\neg x + y) \neq f$ .  $(M, \leq)$  is a lattice. Let  $x \wedge y = \inf_{\leq}(x, y)$ . The notions of an  $RMI$ -model, consistent  $RMI$ -model and the truth of a formula  $A$  (of the language of  $RMI$ ) in such models are defined as in the case of  $RMI_m$ . The classes of setups  $RDL$  and  $CRDL$  are also defined like their counterparts in the case of  $RMI_m$ . Again we have:

**PROPOSITION 26.**

1.  $CP_{RMI} = RDL$ .
2.  $N_{RMI} = CRDL$ .

**Proof.** Similar to the proofs of propositions 22 and 23. ■

Again, theorem 25 and 26(1) entail the following result of [Avron, 1990b]:

**COROLLARY.**  *$RMI$  is sound and complete for the semantics of full r.d.l.s.*

**THEOREM 27.**

1.  $\vdash_{RMI} A$  iff  $A$  is valid in all the consistent models.
2.  $RMI$  is weakly normal.

**Proof.**

1. Suppose that  $\not\vdash_{RMI} A$ . Then there is a model  $(M, v)$  in which  $A$  is not true. Let  $M'$  be the consistent full r.d.l based on  $T_M$  (the tree on which  $M$  is based). Let  $v'$  be any valuation in  $M'$  which satisfies the following conditions: (i)  $|v'(P)| = |v(P)|$  for every atomic  $P$ ,



(ii)  $v'(P) = v(P)$  whenever  $|v(P)|$  is normal in  $M$ . It is easy to see that conditions (i) and (ii) are preserved if we replace  $P$  by any sentence. In particular  $v'(A) = v(A)$  and so  $A$  is not valid in the consistent model  $M'$ .

2. Immediate from part (1) and proposition 26(2) ■

THEOREM 28.

1. *RMI is c-normal.*

2. *Every consistent RMI-Theory has consistent model.*

**Proof.** (1) By proposition 8(4) it suffices to prove that if  $T$  is consistent and  $A$  a sentence then either  $T \cup \{A\}$  or  $T \cup \{\neg A\}$  is consistent. This is not so easy, however, since like in  $R$ ,  $T \cup \{A\}$  might be inconsistent even if  $T \not\vdash \neg A$ , while unlike in  $R$ ,  $(\gamma)$  for  $\vee$  is *not* sound for  $\vdash_{RMI}^N$ .

Suppose then that  $T \cup \{A\}$  and  $T \cup \{\neg A\}$  are both inconsistent. Since  $\neg B, B \vdash_{RMI} \neg(B \rightarrow B)$ , this means, by *RMI* deduction theorem for  $\supset$ <sup>14</sup> that there exist sentences  $B$  and  $C$  such that  $T \vdash_{RMI} A \supset \neg(B \rightarrow B)$ ,  $T \vdash_{RMI} \neg A \supset \neg(C \rightarrow C)$ . In order to prove that  $T$  is inconsistent it is enough therefore to show that the following theory  $F_0$  is inconsistent:

$$F_0 = \{A \supset \neg(B \rightarrow B) \ , \ \neg A \supset \neg(C \rightarrow C)\} .$$

For this we show that the following sentence  $\varphi$  and its negation are theorems of  $F_0$  (where  $a \circ b = \neg(\neg a + \neg b)$ ):

$$\varphi = (B \vee [\neg A \circ R^+(A + C, B)]) \wedge (C \vee [(A + C) \circ R^+(A + B, C)]) .$$

By the completeness theorem it suffices to show that  $\varphi$  gets a neutral value ( $I$ ) in every model of  $F_0$ . Let  $(M, v)$  be such a model, and denote by  $R$  the relevance relation between the nodes of the tree on which  $M$  is based. It is easy to see that:

- a)  $|v(A)| \not\prec |v(B)|$       $|v(A)| \not\prec |v(C)|$
- b) If  $|v(A)| \not\# |v(B)|$  or if  $v(A)$  is designated then  $v(B)$  is neutral.
- c) If  $|v(A)| \not\# |v(C)|$  or if  $v(\neg A)$  is designated then  $v(C)$  is neutral.

Denote, for convenience,  $v(A)$  by  $a$ ,  $v(B)$  by  $b$ ,  $v(C)$  by  $c$ , and the two conjuncts of  $\varphi$  by  $\varphi_1$  and  $\varphi_2$  respectively. Then:

- (i) If  $|b| \not\# (|a| \vee |c|)$  then  $v(\varphi_1) = b$ . Also we have then that  $|c| \leq |a| \vee |c| < |a| \vee |b| = |a + b|$  (since always  $(|a| \vee |b|) R (|a| \vee |c|)$ ). Hence  $|c| R |a + b|$  and so  $v(\varphi_2) = t_{|a| \vee |b| \vee |c|}$ . It follows that  $v(\varphi) = b$  and so  $v(\varphi)$  is neutral by b) above.

---

<sup>14</sup>See [Avron, 1990b]. The connective  $\supset$  is defined there by  $a \supset b = b \vee (a \rightarrow b)$ .

- (ii) If  $|b| R(|a| \vee |c|)$  and either  $|a| < |a| \vee |c|$  or  $\text{val}(a) = f$  then, by a),  $|b| \leq |a| \vee |c|$  and either  $|a| \not R |c|$  or  $v(\neg A)$  is designated. Hence  $c$  is neutral by c). It follows (since either  $|a| \not R |c|$  or  $\text{val}(a) = f$ ), that either  $|a| \not R |c|$  or  $|c| < |a|$ . In both cases  $v(A + C) = f_{|a| \vee |b| \vee |c|}$ ,  $v(\varphi_2) = c$ , and  $v(\varphi_1) = t_{|a| \vee |b| \vee |c|}$ . Hence  $v(\varphi) = c$ , which is neutral.
- (iii) If  $|b| R(|a| \vee |c|)$ ,  $|a| = |a| \vee |c|$  and  $a$  is designated then, by a),  $|a| = |a| \vee |b| \vee |c|$ . If  $\text{val}(a) = I$  then also  $\text{val}(b) = I$  and  $\text{val}(c) = I$ , and so  $\text{val}(v(\varphi)) = I$ . If  $\text{val}(a) = t$  then by b)  $b$  is neutral and so  $|b| < |a|$  ( $|a|$  is normal!). Obviously  $|c| < |a|$  in this case, and so  $v(\varphi_1) = b$ ,  $v(\varphi_2) = t_{|a|} = a$  and  $v(\varphi) = b$ , which is neutral.

(2) Immediate from (1) and proposition 26(2). ■

**PROPOSITION 29.** *RMI is not strongly c-normal.*

**Proof.** Let  $\psi_1$  and  $\psi_2$  be the two elements of the theory  $F_0$  from the last proof. Let  $T = \{\psi_1\}$ ,  $A = \neg\psi_2$ . Then  $T$  is consistent (even classically!) and  $A$  is provable in every consistent and complete extension of  $T$  (since  $F_0$  is inconsistent). Hence  $T \vdash_{RMI}^N A$ . However,  $T \not\vdash_{RMI} A$  since it is easy to construct a full model of  $\psi_1$  in which  $\neg\psi_2$  is not true. ( $\psi_1$  is neutral in this model.) ■

Like in the case of  $R$ , our results show that  $\vdash_{RMI}^N$  is stronger than  $\vdash_{RMI}$  and  $\vdash_{RMI}^{CS}$ . We now construct a formal system for this consequence relation.

**DEFINITION.** The system *RMIC* is *RMI* strengthened by *M.T.* for  $\supset$ :

$$A \supset B, \quad \neg B \vdash \neg A.$$

**THEOREM 30.**

1.  $T \vdash_{RMIC} A$  iff  $T \vdash_{RMI}^N A$
2.  $\vdash_{RMIC} A$  iff  $\vdash_{RMI} A$ .

**Proof.**

1. Obviously, if both  $A \supset B$  and  $\neg B$  are true in a consistent model  $(M, v)$  then so is  $\neg A$ . Hence if  $T \vdash_{RMIC} A$  then  $T \vdash_{RMI}^N A$ . For the converse, suppose  $T \vdash_{RMI}^N A$ . Then by Theorem 26  $T \cup \{\neg A\}$  has no consistent model. This means, by Theorem 28, that  $T \cup \{\neg A\}$  is inconsistent. Hence  $T \vdash_{RMI} \neg A \supset \neg(B \rightarrow B)$  for some  $B$ . Since also  $\vdash_{RMI} \neg\neg(B \rightarrow B)$ , we have that  $T \vdash_{RMIC} \neg\neg A$ , by applying M.T. Hence  $T \vdash_{RMIC} A$ .

2. Immediate from 1) and theorem 27(2). ■

**Notes:**

1. From 30(2) it is clear that the system  $RMI$  is closed under M.T. for  $\supset$ . By applying this rule to theories we can make, however, any inconsistent theory trivial. This resembles the status of  $(\gamma)$  in  $R$  and  $E$ . Indeed  $(\gamma)$  may be viewed as M.T. for the usual implication as defined in classical logic. A comparison of theorems 30 and 20 deepens the analogy (note that  $RMI$  is *not* an extension of  $R$  and 20 fails for it!).
2. Despite 30(2)  $RMI$  and  $RMIC$  are totally different even for consistent theories, as we have seen in prop. 29. *It is important, however, to note that theory  $T$  is consistent in  $RMI$  iff it is consistent in  $RMIC$ .* This follows easily from theorem 28.

*3.6 Three Valued Logics*

Like in section 2, we consider here only the 3-valued logic which we call in [Avron, 1991b] “natural” (in fact, only those with Tarskian CR). All these logics have the connectives  $\{\neg, \wedge, \vee\}$  as defined by Kleene. The weaker ones have only these connectives as primitive. The stronger ones have also an implication connective which reflect their consequence relation.

Suppose the truth-values are  $\{t, f, I\}$ .  $t$  and  $f$  correspond to the classical truth values. Hence  $t$  is designated,  $f$  is not. The 3-valued logics are therefore naturally divided into two main classes: those in which  $I$  is not designated, and those in which it is. The first type of logics can be understood as those in which the law of contradiction is valid, but excluded middle is not. The second type – the other way around.

*Kleene’s basic 3-valued logic*

This logic, which we denote by  $K\ell$ , has only  $t$  as designated and  $\{\neg, \vee, \wedge\}$  as primitives. It has no valid formula, but it does have a non-trivial consequence relation, defined by the 3-valued semantics. A setup in this semantics is any set of the form  $\{A \mid v(A) = t\}$  where  $v$  is a 3-valued valuation, and the consequence relation  $\vdash_{K\ell}$  is defined by this semantics. A sound and strongly complete Gentzen-type or natural deduction formulations have been given in several places (see, e.g., [Barringer *et al.*, 1984] or [Avron, 1991b]).

The properties of  $\vdash_{K\ell}$  which are relevant to the present paper are summarized in the following theorem:

**THEOREM 31.**

1. *Like intuitionistic logic,  $\vdash_{K\ell}$  is strongly consistent, c-normal but not even weakly complete.*
2.  *$\vdash_{K\ell}^{CP}$  is classical logic.*

**Proof.**

1. Since  $\neg A, A \vdash_{K\ell} B$ ,  $\vdash_{K\ell}$  is strongly consistent. Since  $\vdash_{K\ell}^{CP} A \vee \neg A$  but  $\not\vdash_{K\ell} A \vee \neg A$ ,  $\vdash_{K\ell}$  is not weakly complete.

We turn now to  $c$ -normality. First we need a lemma

LEMMA 32. *If  $T$  has a 3-valued model then it has also a classical, two valued model.*

**Proof of the lemma:** It is enough to show that every finite subset of  $T$  has a two-valued model (by compactness of classical logic). So let  $\Gamma$  be a finite set which has a 3-valued model. Since De-Morgan laws and the double-negation laws are valid for the three-valued truth tables, we may assume that all the formulas in  $\Gamma$  are in negation normal form. We prove now the claim by induction on the number of  $\wedge$  and  $\vee$  in  $\Gamma$ . If all the formulas in  $\Gamma$  are either atomic or negations of atomic formula, then the claim is obvious. If  $\Gamma = \Gamma_1 \cup \{A \wedge B\}$  then  $\Gamma$  has a model iff  $\Gamma_1 \cup \{A, B\}$  has a model, and so we can apply the induction hypothesis to  $\Gamma_1 \cup \{A, B\}$ . If  $\Gamma = \Gamma_1 \cup \{A \vee B\}$  then  $\Gamma$  has a model iff either  $\Gamma_1 \cup \{A\}$  or  $\Gamma_1 \cup \{B\}$  has, and we can apply the induction hypothesis to the one which does, getting by this a two-valued model for  $\Gamma$ . ■

To complete the proof of the theorem, let  $T$  be a consistent  $\vdash_{K\ell}$ -theory. The definitions of consistency and of  $\vdash_{K\ell}$  imply in this case that it has some 3-valued model. By the lemma it has also a two-valued model. Let  $T^*$  be the set of all the formulae that are true in that two-valued model. Then  $T^*$  is a  $\vdash_{K\ell}$ -setup which is consistent (even classically), complete, and an extension of  $T$ .

2. Since  $\vdash_{K\ell}^{CP} \neg A \vee A$  and  $\neg A \vee C, A \vee B \vdash_{K\ell} C \vee B$ , it is easy to show, using (for example) Shoenfield's axiomatization of classical logic in [Shoenfield, 1967] that  $\vdash_{C\ell} \subseteq \vdash_{K\ell}^{CP}$ . The converse is obvious, since  $\vdash_{K\ell} \subseteq \vdash_{C\ell}$  and  $\vdash_{C\ell}$  is strongly complete (by  $\vdash_{C\ell}$  we mean here classical logic). ■

**LPF/L<sub>3</sub>**

*LPF* was developed in [Barringer *et al.*, 1984] for the VDM Project. As explained in [Avron, 1991b], it can be obtained from  $\vdash_{K\ell}$  by adding an internal implication  $\supset$  so that  $T, A \vdash_{LPF} B$  iff  $T \vdash_{LPF} A \supset B$ . The definition of  $\supset$  is:  $a \supset b = t$  if  $a \neq t, b$  if  $a = t$ . Alternatively one can add to the language Łukasiewicz's implication, or the operator  $\Delta$  used in [Barringer *et al.*, 1984]. All these connectives are definable from one another with the help of  $\neg, \wedge$  and  $\vee$ .

## THEOREM 33.

1.  $\vdash_{LPF}$  is strongly consistent but neither weakly complete nor *c-normal*.
2.  $\vdash_{LPF}^{CP}$  is classical logic.

**Proof.**

1. That  $\vdash_{LPF}$  is strongly consistent but not weakly normal follows from the corresponding fact for  $\vdash_{K\ell}$ , since  $\vdash_{LPF}$  is a conservative extension of  $\vdash_{K\ell}$ . As for *c-normality*, it is enough to note that  $\{(A \vee \neg A) \supset B, \neg B\}$  is consistent in *LPF* (take  $v(A) = I, v(B) = f$ ) but obviously has no consistent and complete extension.
2. Again, take any axiomatization of classical logic in the *LPF*-language and check that all the axioms and rules are valid in  $\vdash_{LPF}^{CP}$ . ■

*The Basic Paraconsistent 3-valued logic PAC*

This logic, which we call *PAC* in [Avron, 1991b]<sup>15</sup>, has the same language (with the same definitions of the connectives) as  $\vdash_{K\ell}$ . The difference is that here both *t* and *I* are designated. A setup in the intended semantics is, therefore, this time a set of the form  $\{A \mid v(A) = t \text{ or } v(A) = I\}$ , where *v* is a three-valued valuation. A sound and strongly complete (relative to the 3-valued semantics) Gentzen-type axiomatization is given in [Avron, 1991b].<sup>16</sup>

## THEOREM 34.

1.  $\vdash_{PAC}$  is strongly complete, weakly normal and *c-normal*. It is neither strongly consistent nor strongly *c-normal*.
2.  $\vdash_{PAC}^N$  is identical to classical logic.

---

<sup>15</sup>It is a fragment of several logics which got several names in the literature – see next subsection.

<sup>16</sup>Giving a faithful Hilbert-type system is somewhat a problem here, since the set of valid formulas is identical to that of classical logic, but the consequence relation is not.

**Proof.**

1. The strong completeness theorem for the Gentzen-type system entails that  $\vdash_{PAC}$  is finitary. Hence to show strong syntactical completeness it is enough to show that the condition in 8(1) obtains. This is easy. Weak normality is immediate from the fact that  $\vdash_{PAC} A$  iff  $A$  is a classical tautology (see [Avron, 1991b]) and that  $\vdash_{PAC} \subseteq \vdash_{Cl}$ .  $c$ -normality is proved exactly as for  $R$  (it is easy to check that  $\vdash_{PAC}$  has all the properties which are used in that proof). It is also easy to check that  $\neg p, p \not\vdash_{PAC} q$  and that  $\{\neg p, p \vee q\}$  is consistent, that  $\neg p, p \vee q \vdash_{PAC}^N q$  but  $\neg p, p \vee q \not\vdash_{PAC} q$  (take  $v(p) = I, v(q) = f$ ). Hence  $\vdash_{PAC}$  is not strongly  $c$ -normal and not strongly consistent.
2. Since all classical tautologies are valid in  $\vdash_{PAC}$  and  $MP$  for classical implication is valid for  $\vdash_{PAC}^N, \vdash_{Cl} \subseteq \vdash_{PAC}^N$ . The converse is obvious, since  $\vdash_{Cl}$  is strongly  $c$ -normal and  $\vdash_{PAC} \subseteq \vdash_{Cl}$ . ■

 $RM_3/J_3$ 

This logic is obtained from  $PAC$  by the addition of certain connectives while keeping the same CR. There are two essential ways that this has been done (independently) in the literature (they were shown equivalent in [Avron, 1991b]):

- (i) Adding an implication  $\rightarrow$ , defined as in [Sobociński, 1952]. In this way we get the strongest logic in the relevance family: the three-valued extension of  $RM$ . It is in this way that this logic arose in the relevance literature. The corresponding matrix is called there  $M_3$  and the logic  $RM_3$ . It can be axiomatized by adding to  $R$  the axioms  $A \rightarrow (A \rightarrow A)$  and  $A \vee (A \rightarrow B)$ .
- (ii) Adding an implication  $\supset$ , defined by (see [da Costa, 1974])  $a \supset b = t$  if  $a = f, a \supset b = b$  otherwise. For this connective the deduction theorem holds. In this form the logic was called  $J_3$  in [D'Ottaviano, 1985] (see also [Epstein, 1995])<sup>17</sup>. It was independently investigated also in [Avron, 1986] and in [Rozonoer, 1989]. Strongly complete Hilbert-type formulations with M.P. for  $\supset$  as the only rule of inference were given in those papers, and a cut-free Gentzen-type formulation can be found in [Avron, 1991b].

In what follows we shall use the neutral name  $Pac^*$  for the CR of  $PAC$  in the extended language. The next theorem shows that the main difference between  $Pac^*$  and  $PAC$  is that  $Pac^*$  is *not* weakly normal.

<sup>17</sup>[D'Ottaviano, 1985] and [Epstein, 1995] consider a language with more connectives, but we shall not treat them here.

THEOREM 35.

1.  $Pac^*$  is strongly complete and  $c$ -normal. It is neither strongly consistent nor weakly normal.
2.  $\vdash_{Pac^*}^N$  is identical to classical logic.

**Proof.**

1. Strong completeness and  $c$ -normality can easily be proved. Since  $\vdash_{Pac^*}$  is a conservative extension of  $\vdash_{Pac}$ , it is not strongly consistent. Finally  $\vdash_{Pac^*}^N A \wedge \neg A \supset B$ , since  $\neg(A \wedge \neg A \supset B) \vdash_{Pac^*} A \wedge \neg A$ , but  $\not\vdash_{Pac^*} A \wedge \neg A \supset B$  (the same argument applies to  $(A \wedge \neg A \rightarrow B)$ ).
2. It is provable in [Dunn, 1970] that classical logic is the only proper extension of  $RM_3$  in the language of  $\{\neg, \vee, \wedge, \rightarrow\}$  (from the point of view of theoremhood). Since we have just seen that the set of valid sentences in  $\vdash_{Pac^*}^N$  is such a proper extension, and since  $MP$  for  $\rightarrow$  is valid for it,  $\vdash_{Pac^*}^N$  should be identical to  $\vdash_{Cl}$  (in this language). The same argument works for the  $\{\neg, \vee, \wedge, \supset\}$  language using the results of [Avron, 1986]. Alternatively, it is not difficult to show that by adding  $\neg A \wedge A \rightarrow B$  to the Hilbert-type formulation of  $RM_3$  or  $\neg A \wedge A \supset B$  to that of  $J_3$  we get classical logic in the corresponding languages. ■

#### 4 CONCLUSION

We have seen two different aspects of negation. From our two points of view the major conclusions are:

- The negation of classical logic is a perfect negation from both syntactical and semantic points of view.
- Next come the intensional fragments of the standard relevance logics ( $R_m, RMI_m, RM_m$ ). Their negation is an internal negation for their associated internal CR. Relative to the external one, on the other hand, it has the optimal properties one may expect a semantic negation to have in a paraconsistent logic. In the full systems ( $R, RMI, RM$ ) the situation is similar, though less perfect (from the semantic point of view). It is even less perfect for the 3-valued paraconsistent logic.
- The negation of Linear Logic is a perfect internal negation w.r.t. its associated internal CR. It is not, however, a negation from the semantic point of view. The same applies to Łukasiewicz 3-valued logic.
- The negations of intuitionistic logic and of Kleen's 3-valued logic are not really negations from the two points of view presented here.

In addition we have seen that within our general semantic framework, any consequence relation which is not strongly normal naturally induces one or more derived consequence relations in which its negation better deserves this name. We gave sound and complete axiomatic systems for these derived relations for all the substructural logics we have investigated.

*Department of Computer Science, Tel Aviv University, Israel.*

## BIBLIOGRAPHY

- [Anderson and Belnap, 1975] A. R. Anderson and N. D. Belnap. *Entailment* vol. 1, Princeton University Press, Princeton, NJ, 1975.
- [Anderson and Belnap, 1992] A. R. Anderson and N. D. Belnap. *Entailment* vol. 2, Princeton University Press, Princeton, NJ, 1992.
- [Avron, 1986] A. Avron. On an Implication Connective of RM, *Notre Dame Journal of Formal Logic*, **27**, 201–209, 1986.
- [Avron, 1988] A. Avron. The Semantics and Proof Theory of Linear Logic, *Journal of Theoretical Computer Science*, **57**, 161–184, 1988.
- [Avron, 1990a] A. Avron. Relevance and Paraconsistency - A New Approach., *Journal of Symbolic Logic*, **55**, 707–773, 1990.
- [Avron, 1990b] A. Avron. Relevance and Paraconsistency - A New Approach. Part II: the Formal systems, *Notre Dame Journal of Formal Logic*, **31**, 169–202, 1990.
- [Avron, 1991a] A. Avron. Simple Consequence relations, *Information and Computation*, **92**, 105–139, 1991.
- [Avron, 1991b] A. Avron. Natural 3-valued Logics— Characterization and Proof Theory, *Journal of Symbolic Logic*, **56**, 276–294, 1991.
- [Avron, 1992] A. Avron. Axiomatic Systems, Deduction and Implication *Journal of Logic and Computation*, **2**, 51–98, 1992.
- [Avron, 1994] A. Avron. What is a Logical System?, in [Gabbay, 1994; pp. 217–238].
- [Barringer *et al.*, 1984] H. Barringer, J. H. Cheng and C. B. Jones. A Logic Covering Undefinability in Program Proofs, *Acta Informatica*, **21**, 251–269, 1984.
- [Cleave, 1991] J. P. Cleave. *A Study of Logics*, Oxford Logic Guides, Clarendon Press, Oxford, 1991.
- [da Costa, 1974] N. C. A. da Costa. Theory of Inconsistent Formal Systems, *Notre Dame Journal of Formal Logic*, **15**, 497–510, 1974.
- [D’Ottaviano, 1985] I. M. L. D’Ottaviano. The completeness and compactness of a three-valued first-order logic, *Revista Colombiana de Matematicas*, **XIX**, 31–42, 1985.
- [Dunn, 1970] J. M. Dunn. Algebraic completeness results for R-mingle and its extensions, *The Journal of Symbolic Logic*, **24**, 1–13, 1970.
- [Dunn, 1986] J. M. Dunn. Relevant logic and entailment. In *Handbook of Philosophical Logic*, 1st edition, Vol III, D. M. Gabbay and F. Guenther, eds. pp. 117–224. Reidel: Dordrecht, 1986.
- [Epstein, 1995] R. L. Epstein. *The Semantic Foundations of Logic, vol. 1: Propositional Logics*, 2nd edition, Oxford University Press, 1995.
- [Fagin *et al.*, 1992] R. Fagin, J. Y. Halpern and Y. Vardi. What is an Inference Rule? *Journal of Symbolic Logic*, **57**, 1017–1045, 1992.
- [Gabbay, 1981] D. M. Gabbay. *Semantical investigations in Heyting’s intuitionistic logic*, Reidel: Dordrecht, 1981.
- [Gabbay, 1994] D. M. Gabbay, ed. *What is a Logical System?* Oxford Science Publications, Clarendon Press, Oxford, 1994.
- [Girard, 1987] J.-Y. Girard. Linear Logic, *Theoretical Computer Science*, **50**, 1–101, 1987.
- [Hacking, 1979] I. Hacking. What is logic? *The Journal of Philosophy*, **76**, 185–318, 1979. Reprinted in [Gabbay, 1994].



- [Jones, 1986] C. B. Jones. *Systematic Software Development Using VDM*, Prentice-Hall International, UK, 1986.
- [Rozonoer, 1989] L. I. Rozonoer. On Interpretation of Inconsistent Theories, *Information Sciences*, **47**, 243–266, 1989.
- [Scott, 1974] D. Scott. Rules and derived rules. In *Logical Theory and Semantical Analysis*, S. Stenlund, ed. pp. 147–161, Reidel: Dordrecht, 1974.
- [Scott, 1974b] D. Scott. Completeness and axiomatizability in many-valued logic. In *Proceeding of the Tarski Symposium*, Proceeding of Symposia in Pure Mathematics, vol. XXV, American Mathematical Society, Rhode Island, pp. 411–435, 1974.
- [Schroeder-Heister and Došen, 1993] P. Schroeder-Heister and K. Došen, eds. *Substructural Logics*, Oxford Science Publications, Clarendon Press, Oxford, 1993.
- [Shoenfield, 1967] J. R. Shoenfield. *Mathematical Logic*, Addison-Wesley, Reading, Mass., 1967.
- [Sobociński, 1952] B. Sobociński. Axiomatization of partial system of three-valued calculus of propositions, *The Journal of Computing Systems*, **11**, 23–55, 1952.
- [Troelstra, 1992] A. S. Troelstra. *Lectures on Linear Logic*, CSLI Lecture Notes No. 29, Center for the Study of Language and Information, Stanford University, 1992.
- [Urquhart, 1984] A. Urquhart. Many-valued Logic. In *Handbook of Philosophical Logic*, Vol III, first edition. D. Gabbay and F. Guentner, eds. pp. 71–116. Reidel: Dordrecht, 1984.
- [Wojcicki, 1988] R. Wojcicki. *Theory of Logical Calculi*, Synthese Library, vol. 199, Kluwer Academic Publishers, 1988.



TON SALES

## LOGIC AS GENERAL RATIONALITY: A SURVEY

Logic today is urged to confront and solve the problem of reasoning under non-ideal conditions, such as *incomplete information* or *imprecisely formulated statements*, as is the case with *uncertainty*, *approximate* descriptions or linguistic *vagueness*. At the same time, Probability theory has widened its traditional field of analysis (the *expected frequency* of physical phenomena) so as to encompass and analyze general *rational expectations*. Thus, Probability has placed itself in the position of offering Logic a solution for its own long-awaited generalization. The basis for that turns out to be precisely the shared base underlying the two disciplines. This theoretical base predates their common birth, as seen in the early efforts of Bernoulli and Laplace, as well as in Boole's 1854 attempt to formalize the "laws of thought" and then, as he claimed, to "derive Logic and Probability" from them. Once we recover (following Popper's 1938 advice) the underlying *formalism*, we come, by interpreting it in two different directions, back into either *Logic* or *Probability*. The present survey explains the story so far and does the reconstruction work from the logical point of view. The stated aim is to generalize *Logic* so as to cover, as Boole intended, the whole of *rationality*.

### INTRODUCTION

This survey could as well be entitled: "*How Logic was once the same as Probability, and then they diverged —and how they may again be formally the same*", or "*Logic and classical Probability: recovering the lost common ground*". Before we begin, let us say that the implied desideratum of the title(s) is long overdue. Indeed, that (a) standard Logic *can* be generalized, and that (b) the natural generalization of Logic *is* —or derives from, or is suggested by— Probability theory seems at present the shared conviction of a number of logicians and probabilists. Thus, to cite a few of the latter, Ramsey wrote (in 1926) that the laws of probability are actually laws of *consistency* (or *rational* behavior), an extension of Formal Logic to cover partial information, and that Probability theory could become the "logic of consistency" which would control and guarantee, as Mathematics does, that our beliefs are not self-contradictory. At about the same time de Finetti concluded that Probability theory is the only possible "logic" to generalize standard Logic. All the same, Patrick Suppes was considering in 1979 that Probability theory is the natural extension of classical deductive inference

*D. Gabbay and F. Guentner (eds.),  
Handbook of Philosophical Logic, Volume 9, 321–366.  
© 2002, Kluwer Academic Publishers. Printed in the Netherlands.*

rules, while, more recently, Glenn Shafer declared that “probability is not really about numbers; it is about the structure of reasoning”.

Why the technicalities of Probability theory should be viewed today as a powerful *generalizer* of standard Logic and a suitable formal *unifier* of the two may come as a surprise both to probabilists and logicians. Actually, the idea—that we pursue in our generalization below—is very simple: Probability and Logic are but *two interpretations of a same underlying concept*. This is how the founders of both theories saw it and what Popper later explicitly said (and Kolmogorov claimed he had done). But this old notion, over which notable thinkers like Reichenbach or Carnap agonized after the 1930s, is shared nowadays by a surprisingly exiguous minority of specialists (in both disciplines).

Logic and Probability are overwhelmingly seen today as two completely disparate fields, with a very few, if any, points of contact. Logic deals with *reasoning* and *truth*, Probability with inference on poor data. They seem to have nothing in common. Though they both start with a set  $\mathcal{B}$  of Boolean-structured objects (respectively *sentences* and *events*—or, confusingly, “propositions”) and though they assign them values in a simple number system containing the one and the zero (here logicians prefer ‘truth’ and ‘falsity’, though), at this point the similarity apparently ends, for the two valuations are perceived to be very different: the probabilistic  $P : \mathcal{B} \rightarrow [0, 1]$  obeys a set of axioms set forth by Kolmogorov in the 1930s (that do not actually follow from any particularly “probabilistic” rules but rather reduce it to a simple ‘measure’—in the technical sense—of the “event” objects), while the *logical* valuation is felt to be of a quite different nature and regulated by a semantics set forth by Tarski, also in the thirties, and buttressed by elaborate, specialized logical considerations. Moreover, either field not only has a different type of problem to solve, it also has a different, incompatible set of base concepts and interpretations to work on. And the diverging traditions have bred different strokes of unrelated practitioners and two methodologies that are seen by the mainstream mathematician as far distant (even lying at opposite fields of Mathematics, i.e. *real* vs. *discrete*).

However, this is not how things were seen in the first stages of the modern theories of Probability and Logic. Their founders, notably Bernoulli and Laplace, or Boole and Peirce, dithered a lot on what might “probability” or “truth” mean, and often tended to explain one through the other in incipient, half-baked intersecting intuitions, as can be readily seen by browsing into the original literature. Later developments, as well as the progressively firmer foundations and the more specialized and mutually deviating interpretations that either field painstakingly acquired, created an increasing gap between Probability and Logic, in which both contenders apparently never found a ground or occasion to reconcile into one unified approach (which, as we suggest below, is not only desirable but feasible and even natural).

Why striving to recover an encompassing view should be interesting at all may be not obvious at first, but there are strong reasons for it. First, because, as we said, both theories are two differing interpretations of the same idea. Second, because both probabilists and logicians have recently been hard pressed against the limits of their own disciplines when confronting new challenges with consecrated methods. Prominent challenges include: (1) for probabilists, how to clarify the ultimate meaning, and the practical import, of the apparently obvious idea of “probability” (doubters here include names as Keynes [1921], Ramsey [1926] or de Finetti [1931], and the questions raised prolong well to this day into widely-discussed conundrums as the status of *subjective probability*, *rational belief* or *bayesianism*); or (2) for logicians, how to validate reasoning under *uncertainty* or with *incomplete* or *approximate* information (a problem that eventually gave rise, also around the 1930s, to non-standard formalisms such as the *many-valued logics* of Łukasiewicz [1920] (and [1930], with Tarski) or Kleene [1938], or the attempts at defining a *probability logic* by Reichenbach [1935a,b], discussed by Carnap [1950].

The material below is structured in two parts: the first is a short survey explaining why Logic and classical Probability were once the same thing—and gave the (common) pioneers (Bernoulli, Hume, Laplace, Boole) lots of cross-supporting arguments—and why they soon diverged to the point of being considered unrelated. The second part—considerably longer—is a summary of how we locate Logic firmly in the Logic/Probability common heritage; it is based on former work by the author (Sales [1982a,b, 92,94,96]) and the starting point is Popper’s 1938 suggestion (see Popper [1959]) to set forth a *unique algebraic uninterpreted formalism* as the common source from which, through distinct interpretations, both Logic *and* Probability can be formally derived as particular instances. The common formal idea we advance is, as will be later explained, that we can postulate an *additive valuation* (in e.g.  $[0,1]$ ) of the elements of a given abstract Boolean structure  $\mathcal{B}$ —that is later interpreted by Probability as a (set-theoretic) *event*, and by Logic as a (non-set) *sentence*. Our generalization proceeds from this point on as an exclusively *logical* reading of the common uninterpreted formalism. (The development is satisfactory also in a second, non-formal sense, since it can be seen as a vindication and reconstruction of the pioneers’ historical common source of insights.)

## A. The Probability/Logic Interface

### 1 THE VIEW FROM PROBABILITY

#### 1.1 *Classical Probability*

Around the year 1700, Jakob Bernoulli tried to define the *probability* of a phenomenon as a non-evident and non-subjective something that fortunately had effects: the observable frequency, or relative number of cases in which the phenomenon manifested itself. This number was supposed fixed and objective. So, measuring frequencies was the way to estimate “probability”. Conversely, knowing the probability of a phenomenon allowed to predict its expected frequency. This is, in essence, Bernoulli’s theorem. It is the first clear, albeit implicit, definition of probability. It is also the first instance of a duality that is present since in Probability theory: probability  $P$  —a supposedly *objective* property of phenomena— is conceived simultaneously as (1) the ratio of positive cases (call it  $P_c$ ) and (2) the number we have (call it  $P_b$ ,  $b$  for ‘belief’) to estimate  $P_c$ . The first is assumedly an objective reality, the second an inevitably subjective entity that depends on our past history of observations (a paradox that is the common theme of many reflections, like those of e.g. de Finetti). Obviously, the fewer our interactions, the more subjective our  $P_b$  estimate is. The idea is that  $P_b$  “approximates”  $P_c$ , and the aim is getting  $P_b = P_c$  (in some limit situation).

The Rev. Bayes developed Bernoulli’s idea of  $P_b$  converging to  $P_c$  through observational updates and came up with his celebrated formula (posthumously revealed in 1763) to compute  $P$ . Hailed by observational scientists for more than a century, it is now the heart of a debate about what is this Bayes-computed probability. Called “a priori” probability, anti-Bayesians contend it is nothing more than simple, non-objective belief based on a hypothetical view of our ignorance.

Laplace, in 1774 and later, defined probability as  $P_c$ , the ratio of favorable *cases*, all assumed having “the same probability”. The obvious circularity raised some eyebrows in the 1920s, but Laplace’s has been the standard and successful definition since, at least for non-sophisticated applications. Note that it places probability clearly on the *frequency* side, and cavalierly dismisses any subjective-sounding *belief* content, perhaps the reason for its long-standing success. Note, too, that *cases* are a logical notion, since they can be defined —and were by Laplace himself— as the *true* instances of a proposition. Thus, Laplace’s [1774,1820] probability can be seen as an early generalization of Logic, particularly of the concept of validity (resp. consistency), now interpretable as “true in all (resp. some) cases”. Some years before this, Hume had already implied too that probability was a generalization of logical inference by considering that, given a proposition

obtained from some conditions or premises, its probability was the proportion of premises (or premise extensions) in which the proposition was satisfied.

The classical probabilists thus conceived probability, more than as something associated with indeterminism or uncertainty, as a measure of our knowledge of phenomena in the presence of incomplete information (or, dually, as a measure of our partial ignorance). This concept was considered objective because, though not directly measurable, (1) it referred to a supposedly objective situation indirectly parameterizable through its observable effects, and (2) it was manipulable through rules of an objective calculus. So, after Laplace, Probability theory came to be dominated by the probability-as-frequency view. This was convenient as it was objective and “scientific” and adequately eschewed the estimation or “belief” problem.

It lasted until the 1920s, when von Mises, dissatisfied with the classical solution, formalized (in 1928) the estimation or approximation problem by postulating a “sample space”  $\Omega$ , defining frequency in it and computing probability as the result of some limit process, in which the number of observations tended to infinity. Since this was no ordinary limit and the process not quite satisfactory, Kolmogorov [1933] came up with the now universally accepted solution: probability is just the *measure* of an *event* (an event being a set of outcomes); this measure is taken in the mathematical sense, i.e. as a countably additive valuation (though the need for countable additivity has been challenged by many, notably de Finetti [1970] or Popper [1959]). An interesting thing to note: Kolmogorov declared that his formalization was “neutral” in the sense that it was abstract (and thus previous to any interpretation); in his words, probability had to be formal, pure mathematics, merely ruled by axioms. Nevertheless he also declared that his measured entities (in theory merely the members of a  $\sigma$ -algebra over  $\Omega$ ) *are* actually sets. And though he added that this was irrelevant, Popper protested (in 1955) that it *is* relevant in some important cases, and noted that a truly abstract formalization must admit *any* interpretations, including those in which the measured entities are not sets. (But, we add, this is precisely the case of Logic, where we have non-set entities, namely *sentences* from a *language*, not *events* from a *sample space*.) Popper [1959] offered an axiomatic alternative first suggested in 1938, fully developed in 1955, and now fashionable (under the guise of “probabilistic semantics” or “Popper functions”).

## 1.2 “Subjective probability”, “Probability logic” and “Logical probability”

Following Keynes’s [1921] lead, Ramsey [1926] was the first to consider that the belief side of probability, already present in Bernoulli or the Bayes’ formula, was the core of the concept, since the “true” value of probability was

beyond our reach and the best we could do is approximate it through a careful, consistent, rational procedure; so he defined probability as *belief* and defined this as a number obtained on the basis of *consistency* considerations about the belief-holder's *rational* behavior (as deducible from betting protocols); the rationality-induced consistency insured that the number, though inevitably "subjective", was nevertheless the most objective measure one could obtain. In a similar spirit, and in same years, Bruno de Finetti [1931,37] approached probability as an inherently non-objective concept. He summarized it in his well-known slogan "probability does not exist" (i.e. objectively, at least "not more than the cosmic ether"), and nothing beyond consistency assures its imagined objectivity. According to him, probability is merely what we expect on the basis of past experience and the assumed consistency of what we do. As a number, probability (which de Finetti [1937,70] constructed formally on the basis of a vector space of rational expectations), is "objective" as far as the procedure to obtain it obeys coherent assumptions.

The "subjective probability" thesis of Ramsey/de Finetti has found continuation till now in the work of Jeffreys [1939], Koopman [1940], Savage [1954] or Jeffrey [1965], to name a few, all of which reject the epithet "subjective"; they prefer to be called simply probabilists and at most admit that the probability they deal with is a (non-subjective) partial or rational belief, i.e. the value we assign propositions in absence of complete information.

On the other hand, in a series of studies beginning in 1932 Hans Reichenbach [1935a,b], a physicist with an interest in foundations, interpreted probability as a logic. The logic (*probability logic* he called it) was not truth-functional, but he could subsume all classical tautologies as particular cases of propositions  $p$  that had unit probability (i.e.  $|p| = 1$ ). He obtained formulas for the value (probability) of the connectives which are basically like the ones we obtain below; he says that e.g.  $|p \vee q|$  is a function of  $|p|$  and  $|q|$  plus a third parameter  $k$  he calls *Kopplungsgrad* or 'degree of coupling' (defined roughly as the relative size of the intersection of overlapping areas or classes to which a measure is applied that coincides with the conditional probability). This is equivalent to what we obtain in our generalization below, but note that Reichenbach never moves out of probability and events: he always speaks of probability in its standard meaning and only in a translation of senses he says he can interpret the probability of an event sequence — a sequence of binary truth values — as its (non-binary) "truth". His world is clearly that of Probability, and what he obtains is a Logic only in the sense that he speaks of truth, albeit probability by another name. Moreover, though his contemporary critics (including Tarski [1935a]) argued against the construction, they did not because of the subsidiary role of truth in it (as a surrogate for a probability of one) but on the arguable ground that a proper logic ought to be truth-functional (see Urquhart's [1986] comment below as to the contrary).



(Work on ‘probability logic’ by Kemeny *et al.* (see Leblanc [1960]), Bacchus [1990], Halpern [1990] or Fagin *et al.* [1990] is not mentioned here because what these authors deal with is not what is understood by that name in the the classical tradition. Instead, they apply the standard treatment of any first-order theory *inside* Logic, i.e. augmenting ordinary first-order logic with a number of specific axioms that syntactically describe the real numbers (or an ordered field) and the probability operations on them, so that the Probability laws can be derived formally as theorems.)

Another indirect view of Logic-as-valuation was the one adopted by Rudolf Carnap [1950,62], starting in the 1940s. The Logic/Probability link began in his case by trying to justify probability logically. In his view (that he called *logical probability* and surmised as theoretical ground on which to base a “logic of induction” to which Popper came to be fiercely opposed), the probability of an event was the proportion or, more generally, the measure (“ratio of ranges of propositions”) of the intervening circumstances (described as logical sentences) concurring in the event. For this measure he said he was inspired by a definition of Wajsberg—which was inspired in turn by proposition \*5.15 of Wittgenstein’s [1922] *Tractatus* (ultimately Bolzano-inspired, see below). Carnap hesitated and changed his approach often along the 1950s; for instance, notably, he came to value sentences instead of events, but came back to events later, shortly before giving up the whole scheme. Wittgenstein and Wajsberg’s extensional rendition and Carnap’s use of them is, like Reichenbach’s implicit grounding of probability on rather obscure “overlapping classes”, strongly reminiscent of the Stone representation we obtain below out of our general, non-probabilistic truth valuation of logical sentences. It is interesting to note that Carnap’s logically-described components of events correspond rather precisely to what Laplace had called the (positive i.e. true) “cases” concurring in an event. They are also almost interchangeable with Boole’s *cases* (his “conceivable sets of circumstances”) underlying a logical proposition (or with equivalent descriptions by McColl and Peirce and Wittgenstein, see below).

## 2 THE VIEW FROM LOGIC

### 2.1 *The pioneer logicians*

The laplacian idea of having “cases” (a logical concept, we noted) and then measuring the proportion of the true ones seem to have been floating all over. Laplace’s uninfluential contemporary, the Austrian philosopher Bolzano, had this to say about first-order propositions and truth: propositions, he says, have an associated “degree of validity”, a number in  $[0,1]$  which equals “the proportion of true ‘variants’” (Bolzano’s “variants” are our term substitutions).

And then Boole [1854], when just after studying classes he sets out to analyze *propositions* (in 1847), conceives them by means of an alternative interpretation of his elective symbol  $x$  (already introduced for classes) and says it now stands for the *cases* (defined informally as “conceivable sets of circumstances”) —out of a given hypothetical “universe” (a De Morgan’s idea)— in which the proposition is *true*. In the last chapters of his 1854 book (significantly entitled “An investigation of the laws of thought, on which are founded the mathematical theories of logic and probabilities”) Boole even likens the product  $x \cdot y$  of two propositions (i.e. the conjunction value, actually) to the *probability* of simultaneously having them both and the (value of) the sum  $x + y$  to the *probability* of having either (provided they are mutually exclusive).

An equivalent idea is present in MacColl’s [1906] partially published reflections (started before 1897), where he says that propositions are generally “variable”, meaning they are sometimes the case, depending on their (basically probabilistic) modality. Peirce [1902], in an unpublished work that deliberately follows MacColl’s steps, sets out to distinguish “necessary” from “contingent” propositions, most being the latter sort, characterized by their (probabilistic) occurrence.

In a similar vein Wittgenstein [1922] considers a little later that propositions reflect —and are basically decomposable into— “states of affairs” (an idea borrowed from Leibniz). That those states of affairs (reminiscent of Laplace’s or Boole’s cases) are in some way a measurable universe whose proportions gave information on the truth of the composite propositions is obvious from the *Tractatus* (and is the inspiration of the extensional view of Wajsberg and Carnap mentioned above).

## 2.2 “Multi-valued logic”

While some probabilists (from Ramsey to Reichenbach) agonized in the 1930s over their base concepts, there was intense soul searching also in the logicians’ camp. The main new idea came from Łukasiewicz in 1930 (down from antecedents since 1918) when he postulated a logic in which “truth” values could take any value from the infinite real  $[0,1]$  interval (see Łukasiewicz & Tarski [1930]). To compute the value of composite propositions in his “many-valued logic” he obtained (truth-functional) formulas which are exactly the ones we obtain below except for the fact that they presuppose full *compatibility* (a concept we explain below) among *all* propositions, no matter some are the negation of others. This was not perceived as a problem at the time but *it was*, as some fellow logicians pointed out to him at the 1938 Zurich workshop (see Łukasiewicz [1938]). They considered that either we assign  $p \wedge \neg p$  the value  $\min(|p|, 1 - |p|)$  given by the formulas (thus blatantly contradicting the basic logical *law of non-contradiction*) or else we assign it the —correct— zero value (meaning *falsity*, so in accor-

dance with ordinary Logic) but then we arbitrarily disobey the postulated truth-functionality. As this predicament found no decisive solution, many-valued logic has continued to this day —along with unexpected offshoots like “fuzzy logic”— consecrating truth-functionality as a rigid principle and thus putting itself *out* of mainstream Logic, of which is a weak incompatible variety (unless we add, as suggested by van Fraassen [1968], some super-valuation mechanism to it). At the end of a detailed survey of multi-valued logic, Alasdair Urquhart [1986] comments that it is hardly surprising that those systems have remained logical toys or curiosities since “there seems to be a fundamental error [truth functionality] at the root”. Some modern cultivators wonder whether is it possible to combine the two best-known formalisms, Probability and Logic, in any way (but Lee [1972] admits that “we do not seem how to do this”); others, like Hamacher [1976], Zimmerman [1977] or Trillas *et al.* [1982], in asking what are the correct truth-value formulas for the fuzzy calculi, hesitate among a variety of candidates, while still another, Minker (with Aronson *et al.* [1980]), would like to know the “truth bounds” of many-valued conclusions obtained from premises.

### 3 THE VIEW FROM THE OUTSIDE

#### 3.1 Mathematics

Alfred Tarski straightforwardly supposed (in Tarski & Horn [1948]) he had simply a *Boolean algebra* and then set out to analyze thoroughly all possible *measures* in it. So did Gaifman [1962] —and Scott (with Krauss, [1965])— who extended this analysis to first-order logical formulas; these were assigned (additive) values, that were called ‘measures’ by Gaifman (and ‘probabilities’ by Scott). True to mathematicians’ fashion (i.e. approaching topics in uninterpreted, “abstract” formalisms), they did not understand ‘probability’ as other people do; they just used the word as a synonym for *normalized measure* (a *measure* being a  $\sigma$ -additive valuation on the positive reals). In this sense, their “probability” is a blanket term for any common generalization —such as the one we attempt here— for the two (heavily interpreted) fields of Logic and Probability.

Also in this line, J. Łoś [1962] explored general “probability” valuations of logical *sentences* and came up with a (reasonably unsurprising) *representation theorem* of probabilities on a (set-theoretical) space of *models* (or *interpretations*) in the logical sense. Łoś’s line has been consistently followed by Fenstad since 1965. (Fenstad’s papers [1967,68,80,81] have been a source of inspiration for our generalization below.)

### 3.2 *Philosophy*

Philosophers have also been exploring the common material. A few representative examples are Hintikka and Suppes (both presenting first results in 1965), Stalnaker (in 1968–70), Lewis (in 1972) or Popper (e.g. in 1987, with Miller). The first (Jaakko Hintikka [1968]), inspired by Bar-Hillel’s (and Carnap’s, [1952]) information-measure ideas, was suggesting in 1965 (with Pietarinen, [1968]) various formulas to parameterize the information contained in sentences. Also in 1965, the second (Patrick Suppes [1968]) did a circumscribed analysis of the Modus Ponens logical inference rule from a generalized perspective (what he called ‘probabilistic inference’) in which he got formulas fully consistent with the ones we obtain below. Another philosopher traditionally preoccupied with logic/probability differences (especially those centered on the conditional/conditioning operation), Robert Stalnaker [1970], revealed some fine points (among which our “ $A \rightarrow B$ ”  $\neq$  “ $B|A$ ” conceptual and practical distinction). His work and David Lewis’s [1976] have done much to clarify and distinguish concepts shared by logicians and probabilists.

But, prior to these 1960s efforts, the single philosopher to do this most explicitly is surely Popper [1959], in lucid but little known pioneering work. He did not only see (in 1938) that the two concepts were different interpretations of a (yet to be written) formalism —Kolmogorov [1933] also saw this— but he designed one in a very simple and intuitive way by defining a valuation in  $[0,1]$  on pairs  $(a, b)$  of sentences (of a very elementary language) that was directly constructible by users (i.e. reasoners and probability-estimators alike) and that gave way naturally to a Boolean structure with the usual properties (including measurability). Whatever sense the user gave the valuation (“probability”, “truth likelihood”, “truth content” or simply “truth”) it was the user’s concern. Popper later used his own formalism (and the derived Booleanity assumption) to deduce properties of his ‘truth content’ measure and so emit (with David Miller, [1987]) a post-mortem indictment against Carnap’s [1950,62] “inductive logic”.

### 3.3 *“Fuzzy Logic”*

From a logical point of view, Fuzzy Logic (under development since 1965) can be considered as an “interpreted” variety of Łukasiewicz & Tarski’s [1930] infinite-valued logic. (“Interpreted” because it adds to many-valued logic an extensional interpretation of predicates in terms of non-standard sets.) Thus, it was already mentioned in a former section, where we considered it as an (unexpected) offshoot of the many-valued logic family, and we dedicated it some short comments. Nevertheless, the overgrown “fuzzy” tradition, now largely applications-oriented, has its own self-contained rules and momentum and is *not* exactly logic nor probability. Nor, it claims, has

barely anything to do with them, with which it is pretendedly “orthogonal”, only devoted to linguistically-motivated imprecision (i.e. *vagueness*).

One may doubt the claim by fuzzy theorists that the issues they currently discuss have no *logical* bearing. On the contrary, they seem fully relevant for logical discussion, so we have dedicated an appendix to comment rather expansively on fuzzy ‘logic’, as well as to mention an early generalization of Logic that arose inside the fuzzy tradition (by B.R. Gaines [1978]).

### 3.4 *Artificial Intelligence*

Since the moment the first “expert systems” dealt with uncertain information (the obvious cases are Mycin and Prospector), AI as a discipline got involved too in the Probability/Logic dilemma about what is the ultimate nature of “truth” measures of sentences presented to the expert system user (see Shortliffe [1976]). Leaving aside Mycin’s “uncertainty factors” (later revealed to be actually measuring belief *change*, see Heckermann [1986]), the typical measures are Prospector’s “*probability assignments*” (see Duda *et al.* [1976]), that are considered unproblematic and intuitive (to the user, who can easily estimate them), and are combined according to Bayes’ formula as though they were really what their chosen name implied. Whatever the true status (probabilistic, or logical) of the calculus, the formulas on offer happen to become corollaries of our generalized calculus below (where, unlike in AI’s rather *ad hoc* formalisms, nothing is assumed about whether the measures are actually “probabilities” or “truths” or something else).

Nils Nilsson’s [1986] stated goal in his ‘probabilistic logic’ paper (orally anticipated in 1983) was to rationalize past work in the Prospector expert system project (1976-80) and give it a formal background by propounding a ‘probabilistic entailment’ that would do for this formalism what the Modus Ponens rule ( $A, A \rightarrow B \vdash B$ ) does for ordinary Logic. He obtained the well-known bounds for the probability of  $B$  (see e.g. our formula (8) below) by Venn diagram techniques, that he extended to the study of convex hulls in a “probability space” of “possible worlds” (the latter terms are both familiar terminology to de Finetti and Łoś readers). As Nilsson [1993] acknowledged later, his method is similar to work by Good [1950] and Smith [1961] (not to mention de Finetti [1937,70]), authors of whom he was unaware at the time. His goal is, in fact, shared by many since the first 1980s—including the present author and others mentioned in previous and later paragraphs. (The Nilsson effort is briefly discussed in the next section.)

The Artificial Intelligence context has continued to breed practical motivation for the Logic/Probability demarcation. A series of special conferences (*‘Uncertainty in A.I.’*) has been called (beginning in 1986, see Kanal & Lemmer [1986]) and given useful insights into the differences and similarity of the once-separate fields, including expert system coefficient analysis by Heckermann [1986], Grosz [1986] and others or the theoretical framework

called *belief networks*, developed for the efficient computation of “probabilities” (or whatever they are) by Judea Pearl [1988]. These contributions to general and practical Logic, worthwhile as they are, have the implicit bias that what is actually manipulated is the *probability* of distribution-driven events (rather than the *belief*, *commitment* or *assertiveness* of linguistic sentences), and thus the interpretations are always loaded with unnecessary concessions to probabilistic terminology and methods. So, for instance, Pearl, whose formalism is nominally about “beliefs”, is nevertheless overwhelmed with computing probability distributions of *facts* and with assuming simplifying conditions (such as independence or conditioning) to obtain the final value; if this assignment is to be a real “belief”, as stated, then presumably the “facts” and their distributional assumptions should be less real and objective than supposed: probably a consistent calculus (consistent in the Ramsey/de Finetti sense) based on possible or estimated (rather than actual) “facts” would suffice—and for this the possible-worlds or the rational-expectations analyses are already at hand (and ready to be usefully supplemented by a practical procedure such as Pearl’s).

An unsuspected benefit of Logic-oriented analysis by Artificial Intelligence practitioners has been their growing awareness that a system of premises (what they tend to call “knowledge base”) from which predictions are made (or actions are taken) is essentially a set of *beliefs* to which the agent is committed. This is now already clear in classic AI textbooks as Genesereth & Nilsson’s [1987], where the distinction is made between inference procedures where the user’s *full* commitment must be kept *throughout* the inference process and those where the belief premises are “qualified” (e.g. modally, with a belief operator) or “quantified” (with a “probability” assignment); in the latter cases, it is assumed, the commitment is less than absolute and the conclusion strength, therefore, less than guaranteed—however formally valid the reasoning may be. This approach is welcome, since it implies that however we treat premises in a logical argument—either as commitment-inducing beliefs or as admittedly weak probes—they all take part in the inference process and share with it a common goal: knowing to what extent can we *rely* on conclusions.

### 3.5 Probabilistic logic

As mentioned in the previous section, Nilsson [1986] sets out to investigate how Logic would generalize if one were to “assign probabilities instead of truth values to sentences”. Though he calls his probabilities *probabilistic truth values* he treats them as real *probabilities* (at least to the extent that Prospector’s numerical assignments are). This shows clearly in his subsequent treatment of (sentence) conditioning, that he considers plainly a Bayes process and relates to considerations by authors as Pearl, Heckermann, Grosz or Cheeseman (who explicitly deal with *probability* distribu-

tions). Based on entropy considerations by the latter author, Nilsson refines his bounding formulas so as to spot an exact value for the  $B$  in his ‘probabilistic entailment’ (= generalized Modus Ponens) rule—that happens to be the midpoint of the bounding interval (compare that with formulas (9-11) below).

Nilsson’s grounding for his ‘probabilistic logic’ is basically semantic: he exhaustively generates and examines the “possible worlds” inherent in a formula; but the way he then discards some of the worlds—before assigning values to them—amounts to introduce consistency considerations. Akin to this method is what Paass [1988] proposes in a survey: assign basic “subjective probabilities”, construct a universe of “relevant propositions”—which turns out to be isomorphic to Shafer’s “frame of discernment” (or to our  $\Theta$  below)—and then evaluate the resulting probability distributions on it. The computation may be done in Dempster-Shafer’s terms (see Shafer [1976]) or by other methods: linear programming, stepwise simplification, Pearl’s “belief networks” (with interactions) or statistical simulation (see Paass [1988]).

Though they may not use the name (invented by Nilsson), many so-called “probability logics” do not descend from Reichenbach but are really *probabilistic logics* and share Nilsson’s conception and aim: finding a logical foundation for the use of  $[0,1]$  *probability* assignments to sentences taking part in an inference. Most of them formally derive their technical motivation and analysis from Gaifman [1962] and Scott & Krauss [1968]. The author of one of the first such ‘probability logics’, Theodore Hailperin [1984]—who also motivates his analysis historically (and also mentions the classics, from Bernoulli to Keynes and beyond)—sets out to generalize “truth” values and Logic in model-theoretical style, through the use of a modified version of the *model* and *consequence* concepts. This is done too by Bolc & Borowik [1992], who base their analysis on Scott & Krauss [1968] and Adams [1966], and by Gerla [1994] in an interesting attempt parallel to ours below.

## 4 BRIDGING THE GAP

### 4.1 *Attempts at a synthesis*

That the need for a generalization of Logic is widely felt, and that the time is now come to try it, is attested by the many surveys—and attempts at Logic/Probability synthesis (like the present one)—that are appearing of late (see for instance Gärdenfors [1988], Paass [1988], Garbolino *et al.* [1991], Bolc & Borowik [1992] or Gerla [1994]). But other such efforts deserve mention: Kyburg’s [1993] is an exhaustive survey on logics of “uncertainty” where the author probably respects too much the usual division that insulates the surveyed authors’ self-assigned topics, as he divides his

survey, too prolixly, in “objectivism”, “subjectivism”, Nilsson’s “probabilistic logic”, “belief functions”, “measures”, “probabilities”, “statistical facts”, “updating” and “inference”, where the very exhaustivity gets in the way of a comprehensive attempt at synthesis. Similarly division-respecting are several 1995 drafts by Friedman & Halpern on plausibility measures, where the probabilistic bias dominates—in the terminology, in the chosen operations, etc.—though apparently the original intention was to widen “the measure” to a general “plausibility” concept. In the case of the swift and consistent work done by Dubois & Prade [1987,93], now a respected tradition, here the drawback to attain wide method-independent generality is the self-imposed limitation to (fuzzy) possibility measures—though some convergence with non-fuzzy approaches may occur in the future (see below on subadditivity). For the sake of completeness, we must add here work in progress by two researchers with a long tradition in trying to bridge the Probability/Logic gap: Richard Jeffrey [1995] and Glenn Shafer [1996].

#### 4.2 Attempts at finding a meaning for the value

Confronted with the meaning that a “truth value” may have—or may be given—when extended to points in the  $[0,1]$  interval, different people have reacted in a number of ways. Here we mention only those who did not surrender to the temptation of subsuming truth value into *probability* (carrying with it a heavily loaded interpretation of theory). Popper [1972] thought that, in terms of theories rather than sentences, truth value could be made to mean *truth content* (of the theory), degree of *approximation* to truth or, interchangeably, its (appropriately defined) *distance to falsehood*. Haack [1974] saw it could also be interpreted as *partial truth*, roughly defined as the proportion of true components of a sentence or theory, or—equivalently—the “truth” of their conjunction. (An unwilling distant relative of Haack’s is the quantum physicists’ interpretation of the “truth” of a probabilistic quantum event sequence, which is similarly defined by reduction—conjunction, actually—to its elementary event components; see e.g. Reichenbach [1935a,b] or Watanabe [1969]). Dana Scott [1973] tried to answer by defining *truth value* as one minus the *error* we commit when ascertaining or deciding it, clearly in analogy with what we do in the observational sciences. Based on ideas advanced in the 1950s by Bar Hillel (with Carnap, [1952]), Johnson-Laird [1975,83] gave too a definition of truth value, albeit indirectly, by positing as a new concept the *informativeness* or “degree of information” of a sentence (a quantity negatively correlated with the “truth-table probability”) to see how it evolves through the reasoning process, with an eye more on guiding the process than on controlling the *degree of truth* (or *assertiveness*, or whatever), which is what Logic puts the proper emphasis on.



### 4.3 Popper, and Probabilistic Semantics

As repeatedly mentioned above, Popper [1959] (in 1938 and 1955) had decided that Probability and Logic had to be given at last their long-overdue common formalism. Disagreeing with Kolmogorov's [1933] solution (a  $[0,1]$ -valuation on sets) because it was already (semi)interpreted—the valued objects were *sets*—and had a bias toward Probability, he proposed instead a totally abstract, uninterpreted system consisting of (1) an elementary algebra of “sentences” (not necessarily Boolean, merely closed by “conjunction” and “negation”), and (2) a  $[0,1]$ -valuation  $v(a, b)$  on *pairs* of such sentences satisfying very basic and reasonable conditions (see the appendices in Popper [1959]). Such valuations, called “Popper functions”, have become a vogue now (under the name of “probabilistic semantics”), following efforts by Harper [1975], Field [1976] or Leblanc [1979,83] to base ordinary (i.e. “unary”) probability on it. Great advantages of the Popper formalism are:

- the formulas may be interpreted at will either “logically” or “probabilistically”: when in the latter mode, the “sentences” are elementary *events*, the basic operations are *intersection* and *complementation*, and the valuation is just plain *conditional probability* (but with an unsuspected plus: there is no need that the “conditioning” event should be assigned a non-zero (unary) probability)
- there is no need for  $\sigma$ -additivity (as Popper [1959] himself bothers to show), nor is  $\sigma$ -additivity abstract enough for Kolmogorov's [1933] pretendedly neutral formalism to qualify as really neutral (since it is satisfied in an interpretation but discards certain others)
- the resulting quotient algebra *modulo* equi-valuation *is* automatically a *Boolean algebra*, which is not only simple and extraordinarily convenient but, because obtained from very simple assumptions, disarmingly *natural*
- each quotient algebra *class* is interpretable, at will, as an ordinary logical *sentence* or an ordinary probabilistic *event*, and its value turns out to be automatically its *truth value* or, respectively, its (so-called “unary”, i.e. ordinary) *probability*
- the formalism being completely abstract (i.e. uninterpreted) and the interpretation totally free, the “probability” may be—with equal legitimacy—subjective, objective or whatever; in particular it may be Popper's [1962] *truth content*, or its *probability* (the latter is, according to him, the value we give a theory when nothing is known about its *content*, which correlates *negatively* with it)

- likewise, the “truth value” may be *truth* or simple *belief*; in the present author’s view, many other interpretations are equally legitimate, such as “degree of commitment”, “assertive value” or any other that gives us information on the *reliability* —subsuming *truth*— of sentences (including premises and conclusion), and that allows us to have *control* over their behavior along a chain of reasoning
- if one does not accept Popper’s simple conditions, the same result can provably be obtained by accepting alternative and equally simple conditions set forth —independently— by Cox [1961] (and, lately, also the ones by Woodruff [1995]).

As stated, “truth values” may be interpreted in various legitimate ways. Furthermore, any truly abstract formalism requires that they must. In the following sections —to the end of the article— we consider, in genuine Popperian fashion, several fully *logical* interpretations of the “truth value” concept (*belief*, *assertiveness*, etc.), all motivated by what should be included in any study of Logic: *invariance* (of *truth* —or of what the truth value stands for, be it *approximation* to truth, *reliability* or whatever else) along the whole *reasoning* process (assumed formally *valid*). But, compared to the standard Popper schema, our method proceeds just in the reverse direction: where Popper first defines the two-place probability function and then the unary probability is obtained by taking the quotient, we begin instead by a unary valuation and then we subsidiarily define conditioning (that we prefer to call “truth relativity” or “relative truth”) to obtain Popper’s basic two-placed function. The process inversion is unimportant, as we could as well have begun by a two-place valuation (of the *assertive value* —or whatever else— of a sentence *relatively* to the others) and then obtain its *absolute* assertive value by taking the quotient. And note that when we proceed in our direction rather than Popper’s, evaluating the mutually *relative* position of sentences through the  $\alpha$  and  $\sigma$  parameters (see below) amounts to just computing the basic two-place Popper function.

## B. Steps toward a General Logic of Rationality

### 5 MOTIVATION

Let us advance what is our aim here by starting with an obvious remark. When we argue, we do not always fully assert what we say. We often make half-hearted assertions of sentences we are not sure about, or we even use as assertions sentences we hardly believe to be the case. And yet we proceed by reasoning from such weak premises. If we admit we do, and want to treat this inside Logic, we need to qualify assertions, or, if possible, to quantify

their strength, and try to follow and control what effects weak assertions may have in the reasoning process, whether and how they affect its logical validity and how we can tell the strength of the conclusion. All this is indeed a proper *logical* subject (that, however, classical logic never set out to confront).

By tradition, Logic is about *truth*; or, more precisely, it is about truth-preserving manipulations that allow us to validate arguments. *Arguments* are lists of sentences that we note by “ $\Gamma \vdash C$ ”, where  $C$  is called the *conclusion* and  $\Gamma$  is a –possibly empty, or infinite– list of sentences called *premises*. It is no obligation for sentences to be true (or even to have meaning). We merely use them to see whether certain formal manipulations—the *inference rules*—assure us that the prediction embodied by the conclusion  $C$  is *true* whenever the premises are. The whole process is dependent on the truth of the premises: if we cannot assure the truth of *all* of them, the whole procedure becomes redundant. This is how ordinary Logic approaches reasoning.

Now, at least two questions arise. First, suppose we are not sure whether some premise applies, yet we want to know the “truth” of  $C$  (or what is left of it, or, in other words, the *reliability* we can still attach to  $C$  as a prediction). Second, suppose that we deliberately want to *weaken* some true premise to see whether –or up to what point– the conclusion still *holds* (this is : probe the conclusion’s dependence on its premises, or, in other words, the argument’s “robustness”). By tradition, none of this is approached by Logic (so far). To see how we can generalize Logic to cover weak premises we must get a closer look on how we assign truth to sentences. In Logic the

base material is the *sentence*, say  $A$ . (Note  $A$  is *not* a set, but merely a member of a given language  $\mathcal{L}$ .) Once we interpret it we get what we can call a *proposition*. Then, by looking at what is the case, we get a value (a “truth value”). Following Tarski’s [1935b] well-known schema

(T) ‘ $A$ ’ is true if and only if  $A$  is true

the reasoner can verify the sentence (i.e examine the proposition  $A$ , obtained by “unquoting” the sentence  $A$ ) and declare it true whenever the translation  $A$  of the object-language sentence  $A$  is found true.  $A$  is thus assigned the one or *true* value, and we say that  $A$  has full credibility and eventually we assert it with the full confidence that truth warrants. If we find  $A$  false, we assign it the zero or *false* value and give it null credibility. Note it is the reasoner who is full command of the sentences and the translation process, and thus the only one who can validate their truth. As sentences are used to *assert*, and assertions are defined as “true, believed or merely hypothesized” sentences, it is left to the reasoner who uses them to count actively on the quality of assertions as part of the reasoning process itself (so, for instance,

the reasoner usually qualifies the conclusions, conditioned on the strength the original assertions carried).

Now suppose we are reasoning in Physics and  $A$  (a premise) is the positive result of an experiment. If we are sure that  $A$  is true, then we are done: we can use it in a reasoning as a true premise and proceed with the assumedly valid argument to obtain the conclusion. But suppose that, as is usually the case in Physics, we have some qualms about the truth of  $A$ , so we quantify the error  $\varepsilon$  of the experiment. Now the “truth” of the assertion ‘result is positive’ is no longer 1 as before but, say, “ $1 - \varepsilon$ ”. We then perform the formal—valid—reasoning. The question now is : what *confidence*—as a function  $\varphi(\varepsilon)$  of  $\varepsilon$ — may we have in the conclusion?

Most reasonings are like this. We perform as though premises were really true, often unconvincedly. They are provably true sometimes, but most of the time they are ‘assumed true’ for the sake of the argument. Now, because the (T) validation process to declare a sentence true is under control of the user (who performs the translation and decides whether the unquoted statement is observed to be the case), so the reasoner is the only one who can qualify the truth with the appropriate provisos, according to the difficulties met in the validation (unquoting) process. It seems only natural to ask this user not only to qualify but to *quantify* (with a number in e.g. [0,1]) what is the degree of credibility (or belief) (s)he assigns it. The user *can* usually do it consistently (this is the “rational” behavior studied by Ramsey [1926]), thereby defining (by de Finetti’s [1937] theorem) an *additive* valuation. (S)he can always assign the sentence  $A$  the *value*  $v(A)$ —or, as we will write hereafter,  $[[A]]$ —; this value (“truth value” we will call it) may be computed in an unspecified way (by betting preferences, belief networks, simulation, statistical survey, or whatever) and based on *any* preferred interpretation of  $A$ , be it standard *probability* of  $A$  as an event, or Popper’s *truth likelihood* or *truth content* of  $A$  as a proposition or theory (that, as Popper found in the 1930s, is negatively correlated with its probability), or its *partial truth* in Haack’s [1974] sense, or Shafer’s [1976] *belief* (and its dual, *plausibility*), or its *reliability*, or the *credibility* of—or the (user’s) *belief* in— $A$ , or whatever (provided the assignment is done consistently).  $[[A]]$  represents a rough index of the *confidence* we have in  $A$  being the case and—consequently—the force with which we feel we can *assert*  $A$  in a particular argument (or the assertiveness we can commit into it). This measure is always possible, provided the user is “rational” in Ramsey’s sense (but “non-rational” measures are also possible: they merely give *non-additive* values; see the final section, on subadditivity). What is measured is the user’s *belief* in and *commitment* to  $A$ : a zero value means that the premise is to be taken as false, 1 means that it is a true (and therefore fully assertable) premise or—more often—that it is to be *assumed true* (and fully endorsed), and  $v(A) = 1 - \varepsilon$  ( $\varepsilon > 0$ ) means that we can assert  $A$  but with some apprehension or risk (that we assume)  $\varepsilon$ .

Now, approached in a most general way, the problem to solve is as follows. Suppose we have the reasoning  $A_1, A_2, \dots \vdash B$ . Suppose we assign degrees of confidence or assertiveness to the premises. The question is: what will be the effect of those degrees in the confidence or reliability of  $B$ ? (We would thus probe the argument's robustness.) And what if we vary our confidence levels in some premises? Can it happen that, though the reasoning may be formally *valid*, the reliability of  $B$  turns out to be zero (thus making the argument *unsound*)? Or does  $B$  maintain its "truth" (or reliability = 1) though all the premises get null confidence themselves (thus making  $B$ 's truth independent from the premises)? This analysis, like the physicist's  $\varphi(\varepsilon)$  estimation problem above, *is* a legitimate logician's concern. (It is what we proceed to develop in our *proof theory* below.)

We illustrate this with an example. This is the well-known sorites about bald men: "If a man with  $i$  hairs is not bald then a man with  $i - 1$  hairs is still not bald. Suppose a man has  $n$  hairs. Therefore, a man with 0 hairs is still not bald". Formally:

$$\frac{A_i \rightarrow A_{i-1} \quad (i : 1, \dots, n)}{A_n} \quad \frac{}{A_0}$$

This is a paradox because the reasoning is formally correct (it consists of merely  $n$  applications of the Modus Ponens rule), the  $n + 1$  premises are deemed flawless, but the conclusion is outright false (or, more precisely, a *contradictio in terminis*). Usually, it is the length of the argument that is put to blame. There is, however, a more concrete and satisfactory answer we can offer. The  $n$  premises  $A_i \rightarrow A_{i-1}$  cannot obviously be asserted with the same assurance whatever the index value. That's why the argument fails: for low values of  $i$  the premises simply cannot be asserted, even if the rest can, so we can *never* have all premises asserted, and the reasoning is formally valid but vacuously so.

Formally, what happens is that the value  $\llbracket A_i \rightarrow A_{i-1} \rrbracket$  decreases with  $i$ , so that when  $i$  is  $n$  (or even, say, around  $n/2$  or  $n/3$ ) it is 1 or very near 1, but when  $i$  approaches, say,  $n/10$  —and surely when it becomes zero—the value of  $A_i \rightarrow A_{i-1}$  (= the predisposition we have to assert it —or the willingness to assume the risk) comes down to an exceedingly low number. According to a simple proof theory (that we describe below), the conclusion  $A_0$  has the same truth value, at best, as that lowest of numbers (and, thus, the reasoner would be willing to assert the conclusion just no more than he or she would willing to assert  $A_1 \rightarrow A_0$ ).

Note that, though we use words such as 'belief' or 'commitment' as surrogates for truth, *truth* itself is not merely reducible to belief (or probability): compare de Finetti's well-known position ("probability tells us only what to expect, not what will actually be the case") with the more recent comments of Cohen [1990] (a probabilist), who admits that when we say that

something is true with probability, say, .26, “this result tells us nothing about the truth” of the predicted fact or the postulated hypothesis, which will be –is, actually— true or not regardless of our (expectation-inducing) probability computations. This is not to say that Logic should continue to treat truth exclusively, as it now does. On the contrary, we contend that Logic, as it becomes a *general theory of rationality*, should center on a new object of study: the *assertive value* of sentences (that subsumes *truth* and which we will hereafter call –somewhat misleadingly— *truth value*), because this is what is really manipulated in arguments, and because this concept may let us analyze them in full generality, be it through weak premises, strong conclusions or argument robustness.

## 6 “TRUTH” VALUATIONS OVER SENTENCES

We assume we have a set  $\mathcal{L}$  of sentences that form a Boolean algebra (with respect to the three connectives and two special sentences  $\perp$  and  $\top$ ). Now we have the whole Proof Theory of Sentential Logic by identifying the “ $\vdash$ ” order defined by the Boolean algebra with the deductive consequence relation. Thus the algebra of sentences we started with automatically becomes the Lindenbaum-Tarski algebra of all sentences modulo the interderivability relation “ $\dashv\vdash$ ” given by the  $\vdash$  order (i.e.  $A \dashv\vdash B$  iff  $A = B$ ). We then assume that all sentences are valued in  $[0, 1]$ , which we do in the standard way of a *normalized measure*  $v : \mathcal{L} \rightarrow [0, 1] : A \mapsto \llbracket A \rrbracket$ , by just requiring that  $\top$  gets a value of 1 (1 is the only ‘designated value’ we consider) and that the valuation  $v$  is *additive* (i.e.  $\llbracket A \vee B \rrbracket = \llbracket A \rrbracket + \llbracket B \rrbracket - \llbracket A \wedge B \rrbracket$ ). So we now have also the whole Model Theory of Sentential Logic.

This “truth” valuation is merely a (finitely additive) *probability* in all technical senses, but here  $A$  is a *sentence* (in a language  $\mathcal{L}$ ), not an *event* (in a sample space  $\Omega$ ).  $\llbracket A \rrbracket = 1$ ,  $\llbracket A \rrbracket = 0$  and  $\llbracket A \rrbracket = 1/2$  here just mean – respectively— *truth* (or, more precisely, that “ $A$  is taken as truth”), *falsity* and *undecided belief* (when expressly *asserting* the  $A$  sentence); this is to be contrasted with (respectively) probabilistic “certainty”, zero-probability or balanced odds (when evaluating the *uncertain* outcome of  $A$  as an event). We do not require that the valuations—even when interpreted fully as “truth” valuations—to be “extensional” or “truth-functional” as done in many-valued logics. As for the Booleanity of the sentences, either this is assumed (which is undemanding) or it derives from the “minimal algebra” of sentences suggested by Popper [1959]—or from provably equivalent simple assumptions (e.g. by Cox [1961] or Woodruff [1995]). Also, the additive character of the valuation amounts to having a ‘rational’ (Ramsey [1926]) or ‘coherent’ (De Finetti [1937,70]) *belief*, a concept so akin to ‘strength of assertion’ in Logic as to be all but exchangeable.

From the Booleanity of  $\mathcal{L}$  and the above properties of the  $v$  valuation we

immediately obtain:

- (1)  $\llbracket \neg A \rrbracket = 1 - \llbracket A \rrbracket$
- (2)  $\llbracket A \wedge B \rrbracket \leq \min(\llbracket A \rrbracket, \llbracket B \rrbracket)$   
 $\llbracket A \vee B \rrbracket \geq \max(\llbracket A \rrbracket, \llbracket B \rrbracket)$   
 $\llbracket A \rightarrow B \rrbracket = 1 - \llbracket A \rrbracket + \llbracket A \wedge B \rrbracket$   
 $\llbracket A \leftrightarrow B \rrbracket = 1 - \llbracket A \vee B \rrbracket + \llbracket A \wedge B \rrbracket$

## 7 SENTENCES AS SET EXTENSIONS, AND TRUTH AS MEASURE

Any Boolean Algebra has a *representation* on a set structure (a field of sets) as Stone proved long ago in a famous theorem (see, for example, Koppelberg *et al.* [1989]). Thus, given the Boolean sentence algebra  $\mathcal{L}$ , there exist both a set  $\Theta$  (whatever the meaning we give to its elements  $\theta$ ) and a ‘representation’ function that can be characterized as an isomorphism of  $\mathcal{L}$  into the Boolean subalgebra  $\mathcal{B}$  of clopens in  $\mathcal{P}(\Theta)$ , i.e.

$$\rho : \mathcal{L} \longleftrightarrow \mathcal{B} : A \mapsto \mathbf{A} \quad (\mathcal{B} \subset \mathcal{P}(\Theta), \mathbf{A} \subset \Theta).$$

We call the members of  $\Theta$  *possible worlds*, or *cases* (as Laplace [1774] or Boole [1854]) or *possibilities* (Shafer [1976]) or even *observers*, *states*, etc.  $\Theta$  is the *universe of discourse* or *reference frame* (the set of *possible worlds*). It coincides with Fenstad’s [1968] *model space* (where the  $\theta$ s are *interpretations* in the standard logic sense).

We can establish a general, one-to-one correspondence between the two worlds (the language world  $\mathcal{L}$  and the referential universe  $\Theta$ , both made up of “propositions”) and their constituent parts, thus:

$$\begin{array}{lll} \mathcal{L} & \iff & \mathcal{B} \\ A(A \in \mathcal{L}) & \iff & \mathbf{A}(\mathbf{A} \subset \Theta) \\ A \wedge B & \iff & \mathbf{A} \cap \mathbf{B} \\ A \vee B & \iff & \mathbf{A} \cup \mathbf{B} \\ \neg A & \iff & \mathbf{A}^c \\ \top & \iff & \Theta \\ \perp & \iff & \emptyset \\ A \vdash B & \iff & \mathbf{A} \subset \mathbf{B} \end{array}$$

If  $\mathcal{L}$  has a finite number of generators, then it has  $2^n$  atoms  $a$  and the two bijective correspondences  $\mathcal{L} \iff \mathcal{P}(\Theta)$  and  $a \iff \{\theta\}$  also hold.

The valuation  $v$  and the representation isomorphism  $\rho$  induce a  $[0,1]$ -valued measure  $\mu$  in  $\mathcal{B} \subset \mathcal{P}(\Theta)$ , in such a way that  $\mu = v \circ \rho^{-1}$ , i.e.  $\mu(\mathbf{A}) = \llbracket A \rrbracket$ .

Intuitively, the measure  $\mu(\{\theta\})$  of each individual  $\theta$  in a *finite*  $\Theta$  universe is the relevance or the degree of realizability of the given possible world. The  $\mu$  measure corresponds to the weighing function  $\lambda$  in Fenstad's [1968] *model space*. As it is known,  $\mu$  (or  $\lambda$ ) is not only additive but –by the compactness property– countably so; thus  $\mu$  is eligible as a standard “probability” measure (in the technical sense).

## 8 CONNECTIVES AND SENTENTIAL STRUCTURE

If we want to compute the truth value of composite sentences, the task is easy. For the negation connective, the formula is given by (1) above. For the binary sentences composed of  $A$  and  $B$  we have the formulas below, where we observe that, besides  $\llbracket A \rrbracket$  and  $\llbracket B \rrbracket$ , we now need a third parameter that we note “ $\alpha_{AB}$ ” and that we call “*compatibility* between  $A$  and  $B$ ”; its value is defined by  $\alpha_{AB} =_{df} 1 - \beta_{AB}$ , where:

$$\beta_{AB} = \frac{\min(\llbracket A \rrbracket, \llbracket B \rrbracket) - \llbracket A \wedge B \rrbracket}{\min(\llbracket A \rrbracket, \llbracket B \rrbracket, 1 - \llbracket A \rrbracket, 1 - \llbracket B \rrbracket)}.$$

We call  $\beta_{AB}$  the “*degree of incompatibility* of sentences  $A$  and  $B$ ” and we rename the denominator by calling it “ $\Delta_{AB}$ ”. Then, the formulas for the connectives are:

$\llbracket A \wedge B \rrbracket$	$=$	$\min(\llbracket A \rrbracket, \llbracket B \rrbracket) - \beta_{AB} \cdot \Delta_{AB}$
$\llbracket A \vee B \rrbracket$	$=$	$\max(\llbracket A \rrbracket, \llbracket B \rrbracket) + \beta_{AB} \cdot \Delta_{AB}$
$\llbracket A \rightarrow B \rrbracket$	$=$	$\min(1, 1 - \llbracket A \rrbracket + \llbracket B \rrbracket) - \beta_{AB} \cdot \Delta_{AB}$
$\llbracket A \leftrightarrow B \rrbracket$	$=$	$1 -  \llbracket A \rrbracket - \llbracket B \rrbracket  - 2 \cdot \beta_{AB} \cdot \Delta_{AB}$

So, by knowing a single value (either of  $\llbracket A \wedge B \rrbracket$ ,  $\llbracket A \vee B \rrbracket$ ,  $\llbracket A \rightarrow B \rrbracket$ ,  $\llbracket A \leftrightarrow B \rrbracket$ ,  $\alpha_{AB}$  or  $\beta_{AB}$  —or  $\llbracket A|B \rrbracket$  or  $\llbracket B|A \rrbracket$ , see below) we can compute, via  $\alpha_{AB}$  (or  $\beta_{AB}$ ), the other seven. The parameter  $\alpha_{AB}$  (which is a modern version of Reichenbach's [1935b] “Kopplungsgrad”) acts as an indicator or measure of the “relative position” of  $\mathbf{A}$  and  $\mathbf{B}$  inside  $\Theta$ , while “ $\Delta_{AB}$ ” is a quadruple minimum that only depends on the values of  $\llbracket A \rrbracket$  and  $\llbracket B \rrbracket$ . Note that if we suppose that  $\beta_{AB} = 0$  for *all*  $A$  and  $B$  then the above formulas are

$$\begin{aligned} \llbracket A \wedge B \rrbracket &= \min(\llbracket A \rrbracket, \llbracket B \rrbracket) \\ \llbracket A \vee B \rrbracket &= \max(\llbracket A \rrbracket, \llbracket B \rrbracket) \\ \llbracket A \rightarrow B \rrbracket &= \min(1, 1 - \llbracket A \rrbracket + \llbracket B \rrbracket) \\ \llbracket A \leftrightarrow B \rrbracket &= 1 - |\llbracket A \rrbracket - \llbracket B \rrbracket| \end{aligned}$$



and coincide with those given in ordinary *many-valued logics*. Instead, if  $\beta_{AB} = 1$ , the connectives are

$$\begin{aligned} \llbracket A \wedge B \rrbracket &= \max(0, \llbracket A \rrbracket + \llbracket B \rrbracket - 1) \\ \llbracket A \vee B \rrbracket &= \min(1, \llbracket A \rrbracket + \llbracket B \rrbracket) \\ \llbracket A \rightarrow B \rrbracket &= \max(1 - \llbracket A \rrbracket, \llbracket B \rrbracket) \\ \llbracket A \leftrightarrow B \rrbracket &= |\llbracket A \rrbracket + \llbracket B \rrbracket - 1| \end{aligned}$$

and coincide with those given in *threshold logics*.

## 9 RELATIVE TRUTH

Now suppose we want to express the conjunction value as a product:

$$\llbracket A \wedge B \rrbracket = \llbracket A \rrbracket \cdot \tau.$$

With the current  $\mathcal{L}/\mathcal{P}(\Theta)$  representation in mind, we obtain:

$$(3) \quad \tau = \frac{\mu(\mathbf{A} \cap \mathbf{B})}{\mu(\mathbf{A})}$$

We define  $\tau$  as the *relative truth* “ $\llbracket B|A \rrbracket$ ” (i.e. the “truth of  $B$  relative to  $A$ ”). Though this definition exactly parallels that of conditional probability, the account we give leaves out any probabilistic interpretation of the concept and retrieves it for exclusively logical contexts.

In particular, if  $\llbracket B|A \rrbracket = \llbracket B \rrbracket$  then we say that  $A$  and  $B$  are *independent*. In that case, the conjunction can be expressed as the product:

$$\llbracket A \wedge B \rrbracket = \llbracket A \rrbracket \cdot \llbracket B \rrbracket.$$

In any other case we say that  $A$  and  $B$  are mutually *dependent* and speak of the *relative truth* of one with respect to the other. Note the dependence goes both ways and the two situations are symmetric. We have (assuming  $\llbracket A \rrbracket \neq 0$ ):

$$\begin{aligned} \llbracket A \rrbracket \cdot \llbracket B|A \rrbracket &= \llbracket B \rrbracket \cdot \llbracket A|B \rrbracket = \llbracket A \wedge B \rrbracket \quad (\text{a logical “Bayes formula”}) \\ \llbracket B|A \rrbracket &= 1 - \frac{1 - \llbracket A \rightarrow B \rrbracket}{\llbracket A \rrbracket}. \end{aligned}$$

From the latter, note that, in general,

$$\llbracket B|A \rrbracket \neq \llbracket A \rightarrow B \rrbracket.$$

Particularly, we have *always*

$$\llbracket B|A \rrbracket < \llbracket A \rightarrow B \rrbracket$$

*except* when either  $\llbracket A \rrbracket = 1$  or  $\llbracket A \rightarrow B \rrbracket = 1$ , in which cases (and they are the *only* ones)  $\llbracket B|A \rrbracket = \llbracket A \rightarrow B \rrbracket$ . (These facts have been repeatedly noticed by many people, notably by Reichenbach [1935b], Popper [1959], Stalnaker [1970] or Lewis [1976].)

When two sentences  $A$  and  $B$  are *independent* then (and this is a necessary and sufficient condition for that to happen):  $\alpha_{AB} = \max(\llbracket A \rrbracket, \llbracket B \rrbracket)$  if  $\llbracket A \rrbracket + \llbracket B \rrbracket \leq 1$   $= \max(\llbracket \neg A \rrbracket, \llbracket \neg B \rrbracket)$  if  $\llbracket A \rrbracket + \llbracket B \rrbracket \geq 1$  and then the connectives obey the formulas

$$\begin{aligned}\llbracket A \wedge B \rrbracket &= \llbracket A \rrbracket \cdot \llbracket B \rrbracket \\ \llbracket A \vee B \rrbracket &= \llbracket A \rrbracket + \llbracket B \rrbracket - \llbracket A \rrbracket \cdot \llbracket B \rrbracket \\ \llbracket A \rightarrow B \rrbracket &= 1 - \llbracket A \rrbracket + \llbracket A \rrbracket \cdot \llbracket B \rrbracket.\end{aligned}$$

The statement ‘ $A \rightarrow B$ ’ can have, among other readings, one logical (“ $A$  is sufficient for  $B$ ” or “ $B$  is necessary for  $A$ ”), another (loosely) “causal” (“ $A$  occurs and  $B$  follows”). But because  $A \rightarrow B$  is valued in  $[0,1]$ , its value  $\llbracket A \rightarrow B \rrbracket$  (and the values  $\llbracket B|A \rrbracket$  and  $\llbracket A|B \rrbracket$ ) now mean only *degrees*, and so  $B \rightarrow A$  may be—and usually is—read “evidentially” (“ $B$  is evidence for  $A$ ”). Within such a frame of mind,

- $\llbracket B|A \rrbracket$  (or “ $\sigma_{A(B)}$ ”) could be termed “degree of *sufficiency* or *causality*” of  $A$  (or “*causal support* for  $B$ ”), to be read as “degree in which  $A$  is sufficient for  $B$ ” or “degree in which  $A$  is a cause of  $B$ ”. In view of (3), it is roughly a measure of how much of  $\mathbf{A}$  is contained in  $\mathbf{B}$ .
- $\llbracket A|B \rrbracket$  (or “ $\nu_{A(B)}$ ”) could be termed “degree of *necessity*” or “*evidence*” of  $A$  (or “*evidential support* for  $A$ ”), to be read as “degree in which  $A$  is necessary for  $B$ ” or “degree in which  $B$  is evidence (= support of hypothesis) for  $A$  (=the hypothesis)”. With (3) in mind, it can be seen as how much of  $\mathbf{B}$  overlaps with  $\mathbf{A}$ .

Such measures may be directly estimated by experts, normally by interpreting the  $\theta$ s frequently, in terms of *cases*, like Boole [1854], *possibilities* (Shafer [1976]), *elementary events* in  $\Theta$ , or possible *interpretations*; at any rate, they may be statistically-based or simply imagined, presumably on the basis of past experience or sheer plausibility. Thus,  $\sigma_{A(B)}$  in a causal reading of “ $A \rightarrow B$ ” would be determined by answering the question: “How many times (proportionally) —experience shows—  $A$  occurs and  $B$  follows?” For  $\nu_{A(B)}$ , the question would be: “How many times effect  $B$  occurs and  $A$  has occurred previously?” (Similarly for the evidential reading of “ $A \rightarrow B$ ”). Once  $\sigma$  and  $\nu$  have been guessed, they may be adjusted (via the

$$\frac{\sigma_{A(B)}}{\nu_{A(B)}} = \frac{\llbracket B \rrbracket}{\llbracket A \rrbracket}$$

relation) and then lead —by straightforward computation— to  $\llbracket A \rightarrow B \rrbracket$ ,  $\llbracket B \rightarrow A \rrbracket$  and  $\alpha_{AB}$ , which allows one to compute all other values for connectives and also to get a picture of the structural relations linking  $A$  and  $B$ . This process of eliciting the  $\sigma$  value for every  $\langle A, B \rangle$  sentence pair is closely equivalent to computing Popper’s *binary probability* function.

Note first that a similar computing process takes place in “Bayesian reasoning”, and in the “approximate reasoning” methods as implemented in marketable expert systems, though none of them satisfactorily explains on what *logical* grounds are the procedures justified, nor can they avoid supplying a (nominally) *probabilistic* account of them. Such an account would be here clearly misplaced, since there is usually neither (a) a sample space of *events*, but a language of *sentences* (i.e. linguistic descriptions —not necessarily of any “event”), nor (b) a measure based on *uncertainty* and *outcomes* (the topics Probability is supposed to deal with) but rather simple *beliefs* or, at most, mere *a priori* estimates, nor (c) an adequate updatable statistical or probabilistic basis to compute the values of the “events” (and their ongoing, dynamical change).

Note also that any reasoning can proceed here in both directions (from  $A$  to  $B$  and from  $B$  to  $A$ ), because both conditionals  $A \rightarrow B$  and  $B \rightarrow A$  claim a non-zero value and thus a “causal” top-down reasoning can be complemented by an “evidential” bottom-up reasoning on the same set of given sentences (as also happens in Pearl’s [1988] belief networks).

## 10 THE GEOMETRY OF LOGIC: DISTANCE, TRUTH LIKELIHOOD, INFORMATION CONTENT AND ENTROPY IN $\mathcal{L}$

The fact that we have:

$$\llbracket A \leftrightarrow B \rrbracket = 1 - (\llbracket A \vee B \rrbracket - \llbracket A \wedge B \rrbracket)$$

strongly suggests using  $1 - \llbracket A \leftrightarrow B \rrbracket = \llbracket A \vee B \rrbracket - \llbracket A \wedge B \rrbracket$  as a measure of the *distance*  $\overline{AB}$  (under a given valuation  $v$ ). So we do. (We remark that all definitions we give here of distance and related concepts are not only applicable to sentences but to *theories* as well, because for a general lattice  $\mathcal{L}$  the lattice  $\hat{\mathcal{L}}$  of theories derived from each sentence in  $\mathcal{L}$  is isomorphic to  $\mathcal{L}$ .)

DEFINITION 1. *Distance* (or *Boolean distance*) between two sentences or theories  $A$  and  $B$  is:

$$\begin{aligned} d(A, B) &=_{df} 1 - \llbracket A \leftrightarrow B \rrbracket \\ &= \llbracket A \vee B \rrbracket - \llbracket A \wedge B \rrbracket \\ &= |\llbracket A \rrbracket - \llbracket B \rrbracket| + 2 \cdot \beta_{AB} \cdot \Delta_{AB} \end{aligned}$$

DEFINITION 2. *Compatible distance* between two sentences or theories  $A$  and  $B$  is:

$$d^+(A, B) =_{df} |\llbracket A \rrbracket - \llbracket B \rrbracket| = \llbracket A \rrbracket + \llbracket B \rrbracket - 2 \min(\llbracket A \rrbracket, \llbracket B \rrbracket)$$

We now define a *truth likelihood* value for  $A$ —approximating Popper’s [1972] (and Miller’s [1978]) *truth likelihood* or *verisimilitude* measure—by making it to equal the distance from  $A$  to falsehood, i.e.  $d(A, \perp)$ . We obtain, immediately:

$$\begin{aligned} d(A, \perp) &= d(\top, \perp) - d(\top, A) = 1 - d(A, \top) \\ &= 1 - d(A \Delta \top, \perp) \\ &= 1 - d(\neg A, \perp) = 1 - \llbracket \neg A \rrbracket = \llbracket A \rrbracket \end{aligned}$$

So here we have a further interpretation of our “truth values”  $\llbracket A \rrbracket$  in terms of Popper’s [1972] *truth likelihood* or *verisimilitude*. We might as well consider  $\llbracket A \rrbracket$  as a rough measure of *partial truth* or *truth content* of  $A$ . In a similar vein, we may recall that Scott [1973] suggested the “truth value”  $\llbracket A \rrbracket$  of many-valued logics could be interpreted as *one* (meaning *truth*) less the *error* of  $A$  (or rather of a measure settling the truth of  $A$ ) or the *inexactness* of  $A$  (as a theory); in this framework, it comes out that, in our terms,  $\llbracket A \rrbracket = 1 - \varepsilon_A$  and  $\varepsilon_A = 1 - \llbracket A \rrbracket = d(A, \top)$ .

We further observe that, for any sentential letters  $P$  and  $Q$ , any uniform truth valuation yields  $\llbracket P \rrbracket = \llbracket \neg P \rrbracket = .50$ ,  $\llbracket P \wedge Q \rrbracket = .25$  and  $\llbracket P \vee Q \rrbracket = .75$ , which is like saying that, if all letters are equiprobable, the given values are the probability of the given sentence being true (a number that Johnson-Laird [1975,83] calls, appropriately, “truth-table probability”). This value’s complement to one is reasonably made to correspond to the amount of information—in a loose sense—we have when the sentence is true. This is precisely what Johnson-Laird [1975,83] defines as “degree of information”, “semantical information” or *informativeness*  $I(A)$  of a sentence  $A$ . (Viewed in our terms, this *information content*  $I(A)$  equals  $1 - \llbracket A \rrbracket$ , or  $I(A) = \llbracket \neg A \rrbracket = d(A, \top) = \varepsilon_A$ .) The concept, based on Bar-Hillel’s (and Carnap’s, [1952]) ideas, was originally designed to model the reasoning process, assumed to be driven by an increase both of informativeness and parsimony. What is interesting is that the informativeness of composite sentences is computed by combining them according to non-truth-functional rules which yield values that coincide with those predicted by our formulas.

A new measure we can define, which can also be used as an *entropy* (in the sense of De Luca & Termini [1972]), is this one:

DEFINITION 3. *Imprecision* (or, perhaps, “fuzziness”) of a sentence (or theory)  $A$  is the value for  $A$  of the function

$$f : \mathcal{L} \longrightarrow [0, 1] \text{ such that } f(A) = 1 - d^+(A, \neg A)$$

It is immediate that:

$$f(A) = 2 \min (\llbracket A \rrbracket, 1 - \llbracket A \rrbracket),$$

which is equivalent to saying that the imprecision of a sentence  $A$  equals twice the error we make when we evaluate on  $A$  the truth of the law of non-contradiction or of the excluded middle by considering there really *is* maximum compatibility between  $A$  and  $\neg A$ . (Actually, there is *null* compatibility, as  $\alpha_{AB}$  is provably *zero* when  $B$  is  $\neg A$ .) In connection with this measure, we note in passing that:

- *classical* (two-valued) *logic* is the special case of ours in which *all* sentences in  $\mathcal{L}$  have *zero* imprecision.
- ordinary *multi-valued logics* —like Łukasiewicz-Tarski’s  $L_\infty$ — are the ones where at least *one* sentence in  $\mathcal{L}$  has non-zero imprecision. This measure being —as it is— an error function, imprecision is here just the degree in which these logics fail to distinguish contradictions (their lack of “resolving power”).

## 11 ELEMENTARY PROOF THEORY FOR GENERAL ASSERTIONS

Once we have valued sentences in  $[0, 1]$  we now need a Proof Theory that in the most natural way extends standard logic so as to treat imprecise statements or weak assertions, and measure and control whatever effect they may have on reasoning (as well as to explain some results in approximate reasoning methods from Artificial Intelligence).

To begin with, suppose a valid argument, noted  $\Gamma \vdash B$  (where  $\Gamma$  are the premises, or a finite subset of them). Classical logic declares it *valid* if  $B$  is derivable from  $\Gamma$  in an appropriate deduction calculus. By the completeness property, this amounts to assert the truth of  $B$  whenever the premises in  $\Gamma$  are true. Now, as we said, the ultimate judge of the truth of the premises is the reasoner. It is the reasoner who decides that each premise used is true (or to be considered true). To justify such a decision, the reasoner applies a truth criterion such as Tarski’s [1935b] (T) schema. Thus the reasoner declares  $A$  true when assured that what  $A$  describes is precisely the case. If the reasoner is not sure of the result of his/her validation or does not want to commit him/herself to it, then the reasoner may choose not to make a full assertion by claiming that  $A$ ’s verification does not yield an obvious result. In that case, the reasoner may rather easily “qualify” the assertion by assigning numbers in  $[0, 1]$  such as  $v(A)$  —that we noted “ $\llbracket A \rrbracket$ ”— or  $\varepsilon(A)$  [ =  $1 - v(A)$  ] meaning that the reasoner believes or is willing to assert  $A$  to the degree  $v(A)$  or assume it with a risk or estimated error of  $\varepsilon(A)$ .

The proof theory we now sketch is a slightly extended version of the standard one. Here we understand by *proof theory* the usual syntactical deduction procedures *plus* the computation of numerical coefficients that we must perform alongside the standard deductive process. We do that because a final value of zero for the conclusion would invalidate the whole argument as thoroughly as though the reasoning were formally —syntactically— invalid. As always, any formally valid *argument* will have, by definition, the following *sequent* form:

$$\Gamma \vdash B$$

where  $B$  is the conclusion and  $\Gamma$  stands for the list —or, rather, the conjunction— of the premises (or, by the compactness property, of a finite number of them). We have, elementarily:

$$(4) \quad \Gamma \vdash B \Rightarrow \llbracket \Gamma \rrbracket \leq \llbracket B \rrbracket.$$

We henceforth assume that we have a *valid argument* (so  $\Gamma \vdash B$  will always hold), and that all premises are non-zero (i.e.  $\forall i \llbracket A_i \rrbracket > 0$ ). We distinguish four possible cases:

1.  $\llbracket \Gamma \rrbracket = 0$  (i.e. the premises are —materially— *inconsistent*). Here by (4)  $\llbracket B \rrbracket$  can be anywhere between 0 and 1; this value is in principle undetermined, and uncontrollably so.
2.  $\llbracket B \rrbracket = 0$ . This entails, by (4),  $\llbracket \Gamma \rrbracket = 0$  and we are in a special instance of the previous case. The reasoning is formally valid, the premises are not asserted, and the conclusion is false.
3.  $\llbracket \Gamma \rrbracket \in (0, 1)$  (i.e. the premises are consistent). Then, by (4),  $\llbracket B \rrbracket > 0$ . We have a formally valid argument, we risk assessing the premises (though with some apprehension) and get a conclusion which can be effectively asserted though by assuming a —bounded— risk. This will be the case we will set to explore below.
4.  $\llbracket \Gamma \rrbracket = 1$ . This condition means that  $\llbracket A_1 \rrbracket = \dots = \llbracket A_n \rrbracket = 1$  and, by (4),  $\llbracket B \rrbracket = 1$ . So the premises are *all* asserted —with no risk incurred— and the conclusion *holds* unconditionally (remember  $\Gamma \vdash B$  is formally valid). This is the classical case studied by ordinary two-valued Logic.

We are interested in examining *case 3* above, i.e. formally valid reasoning *plus* assertable premises (though not risk-free assertions) *plus* assertable conclusion (but at some measurable cost). Cases 3 and 4 characterize in a most general way all *sound* reasoning. Case 2 characterizes *unsound* arguments (since in this case having a formally valid argument  $\Gamma \vdash B$  does

not preclude getting an irrelevant conclusion ( $\llbracket B \rrbracket = 0$ ). As this case is the one to avoid, we have:

DEFINITION 4. *Unsoundness* of a valid argument  $\Gamma \vdash B$  is having  $\llbracket B \rrbracket = 0$  though the premises are themselves non-zero. So:

DEFINITION 5. *Soundness* of a valid argument  $\Gamma \vdash B$  is having  $\llbracket B \rrbracket > 0$  whenever the premises non-zero.

With that in mind, we can now turn to the basic inference rule, the Modus Ponens (MP). From a strictly logic point of view, this rule is

$$(5) \quad \begin{array}{r} A \quad m \\ A \rightarrow B \quad n \\ \hline B \quad p \end{array}$$

where  $m$ ,  $n$  and  $p$  stand for the strength or force (or “truth value”) we are willing to assign each assertion; so, in our terms,  $m$ ,  $n$  and  $p$  are just our  $\llbracket A \rrbracket$ ,  $\llbracket A \rightarrow B \rrbracket$  and  $\llbracket B \rrbracket$ . They are numbers in  $[0,1]$  that take part in a (numerical) computation which parallels and runs along the logical, purely syntactical deduction process. This is well understood and currently exploited by reasoning systems in Artificial Intelligence that must rely on numerical evaluations —given by users— that amount to *credibility* assignments (or “certainty factors”), *belief* coefficients, or even —rather confusingly— *probabilities* (often just *a priori* probability estimates); this is the case of successful *expert systems* such as Prospector or Mycin. The trouble with such systems is that they tend to view Modus Ponens as a probability rule (this is made explicit in systems of the Prospector type, see Duda *et al.* [1976]). They use it to present the MP rule in this way:

$$(6) \quad \begin{array}{r} A(m) \\ A \rightarrow B(\sigma) \\ \hline B(p) \end{array}$$

where  $m$  and  $p$  are the ‘probability’ (a rather loose term here) of  $A$  and  $B$ , and “ $A \rightarrow B(\sigma)$ ” means that “whenever  $A$  happens,  $B$  happens with probability  $\sigma$ ”. Here  $\sigma$  turns out to be just  $v(B|A)$  or “ $\llbracket B|A \rrbracket$ ”, the “relative truth” of  $B$  given  $A$ ; this concept, modelled on a close analogy —*ceteris paribus*— with that of (probabilistic) conditioning, is what we have defined above (in (3)) and called “degree of sufficiency”  $\sigma$  of  $A$  —or of necessity of  $B$ —, assumed easily elicitable by experts. So it is just natural, and immediate, to compute the  $p$  value thus:

$$p \geq \sigma \cdot m$$

or, in our notation,

$$\llbracket B \rrbracket \geq \llbracket B|A \rrbracket \cdot \llbracket A \rrbracket$$

which is just another version of formula (2).

The problem is that what we have, from our purely logical, probability-rid standpoint, is (5), not (6), and in (5)  $n$  is not  $\llbracket B|A \rrbracket$  but  $\llbracket A \rightarrow B \rrbracket$ . Recall that  $\llbracket B|A \rrbracket$  and  $\llbracket A \rightarrow B \rrbracket$  not only do *not* coincide but mean different things (as repeatedly noticed by logicians, and as explained above). Indeed,  $\llbracket A \rightarrow B \rrbracket$  is the value (“truth” we may call it, or “truth minus risk”) we assign to the (logical) assertion  $A \rightarrow B$ . Instead,  $\llbracket B|A \rrbracket$  is a relative measure linking materially, *factually*,  $A$  and  $B$  (or, better still, the  $\mathbf{A}$  and  $\mathbf{B}$  sets), with no concern whether a true *logical* relation between them exists; we might even have  $\llbracket B|A \rrbracket < \llbracket B \rrbracket$ , thereby indicating there exists an *anticorrelation* (thus rather contradicting any —logical or other— reasonable kind of relationship between  $A$  and  $B$ ). So we turn back to our (5) rule; note that  $m + n \geq 1$  (this *always* holds) and that  $\llbracket B|A \rrbracket$  can be obtained from  $\llbracket A \rightarrow B \rrbracket$ , or vice versa:  $\llbracket A \rightarrow B \rrbracket$  from  $\llbracket B|A \rrbracket$  through

$$(7) \quad \llbracket A \rightarrow B \rrbracket = 1 - \llbracket A \rrbracket \cdot (1 - \llbracket B|A \rrbracket)$$

(which is useful, since  $\llbracket B|A \rrbracket$  is directly obtainable from experts).

The following theorem states the *soundness* condition for the MP rule.

The Modus Ponens rule

$$\frac{\begin{array}{cc} A & m \\ A \rightarrow B & n \end{array}}{B \quad p} \quad (\text{we assume } m \text{ and } n \text{ are both non-zero})$$

is *sound* (and thus  $\llbracket B \rrbracket \neq 0$ ) if one of these four equivalent conditions hold:

1.  $m + n > 1$
2.  $\llbracket B|A \rrbracket > 0$
3.  $\llbracket A \wedge B \rrbracket > 0$
4. Either  $\llbracket A \rrbracket + \llbracket B \rrbracket > 1$  (and thus  $\llbracket B \rrbracket > 1 - m$ ) or both  $A$  and  $B$  are compatible ( $\alpha_{AB} > 0$ ) and not binary-valued.

In both sound and unsound cases we have the following easily computable bounds for the value  $\llbracket B \rrbracket$  of the MP conclusion (Sales [1992,96]):

$$(8) \quad \llbracket A \rrbracket + \llbracket A \rightarrow B \rrbracket - 1 \leq \llbracket B \rrbracket \leq \llbracket A \rightarrow B \rrbracket$$

or equivalently, in shorter notation:

$$m + n - 1 \leq p \leq n .$$



(Such bounds have been discovered again and again by quite diverse authors; see e.g. Genesereth & Nilsson [1987]). The lower bound—which equals  $\llbracket A \wedge B \rrbracket$ —is reached when  $\alpha_{AB} = 1$  and  $\llbracket A \rrbracket \geq \llbracket B \rrbracket$ , while the upper bound is reached when  $\alpha_{AB} = 0$  and  $\llbracket A \rrbracket + \llbracket B \rrbracket \geq 1$ . Naturally we usually know neither  $\llbracket B \rrbracket$  nor  $\alpha_{AB}$  beforehand, so we don't know whether the actual value  $\llbracket B \rrbracket$  reaches either bound or not, nor which is it; we can merely locate  $\llbracket B \rrbracket$  inside the  $[m + n - 1, n]$  interval.

But this interval can be narrowed. Since a very reasonable constraint a conditional  $A \rightarrow B$  may be expected to fulfill is that  $A$  and  $B$  be (assumedly) non-independent—and not binary—and *positively* correlated (i.e.:  $\llbracket A \wedge B \rrbracket > \llbracket A \rrbracket \cdot \llbracket B \rrbracket$ ), so we have:

$$(9) \quad \llbracket A \rrbracket + \llbracket A \rightarrow B \rrbracket - 1 \leq \llbracket B \rrbracket < \llbracket B|A \rrbracket.$$

Here, if  $A$  and  $B$  are fully or strongly compatible,  $\llbracket B \rrbracket$  will be nearer the lower bound. Thus, we can only increase our  $\llbracket B \rrbracket$  if we are assured that  $A$  and  $B$  are independent (in the sense that  $\llbracket A \wedge B \rrbracket$  equals  $\llbracket A \rrbracket \cdot \llbracket B \rrbracket$ , see above): we then obtain the highest value  $\llbracket B \rrbracket = \llbracket B|A \rrbracket$ . On the other hand, the more we confide instead in a strong logical relation between  $A$  and  $B$ , the more we should lean towards the low value given by

$$(10) \quad \llbracket B \rrbracket = \llbracket A \wedge B \rrbracket = m + n - 1.$$

Under very reasonable elementary hypotheses (like this one:  $\llbracket A \rightarrow B \rrbracket > \llbracket A \rightarrow \neg B \rrbracket$  or, equivalently,  $\sigma_{A(B)} > 1/2$ , see Sales [1996]), we easily get these bounds for  $\llbracket B \rrbracket$ :  $\llbracket A \rrbracket / 2 < \llbracket B \rrbracket \leq \llbracket A \rrbracket$ .

Now, if what we want is not an interval, however narrowed, but a precise value for  $\llbracket B \rrbracket$  we should favor the lower value, the one given by (10) above. There are lots of reasons (some mentioned in Sales [1996]) for this choice of value in the absence of more relevant information.

But if we want not merely a pair of bounds—or a favored lower bound—for the conclusion  $B$  of an MP but the *exact* value  $\llbracket B \rrbracket$ , the obvious candidate formula for this follows easily: suppose we are given not only  $\llbracket B|A \rrbracket$  but also  $\llbracket A|B \rrbracket$  (that we note by  $\sigma$  and  $\nu$ ) and we assume them estimated by experts. We then formulate MP as

$$\frac{A(m) \quad A \rightarrow B(\sigma, \nu)}{B(p)}$$

which is exactly (6) except that the conditional has prompted evaluation of relative truths of  $A$  and  $B$  in both directions. The value is computable at once from the above definition of  $\llbracket A|B \rrbracket$ :

$$\llbracket B \rrbracket = \frac{\llbracket B|A \rrbracket \cdot \llbracket A \rrbracket}{\llbracket A|B \rrbracket} \quad \text{or} \quad p = \frac{\sigma \cdot m}{\nu}.$$

Note that the above logical formula coincides formally with Bayes's theorem —whence the adjective (“Bayesian”) for any calculus that uses it— except that it deals not with hard-to-compute probabilities of events but with beliefs (or assertion strengths) of sentences. Note also that this value is the one that some *approximate reasoning* systems (e.g. Prospector) unqualifiedly assign to  $\llbracket B \rrbracket$  supposedly on purely probabilistic grounds —and falsely assuming that  $\llbracket B|A \rrbracket$  is the same as  $\llbracket A \rightarrow B \rrbracket$ —; see, for instance, the Genesereth & Nilsson [1987] text, where the logical equation above is said to be Bayes's formula.

If we wanted the MP presented in the more traditional *logical* way (5), first we would directly estimate the truth value  $\llbracket A \rightarrow B \rrbracket$  of the conditional, or compute it from  $\sigma$  through (7) —or both, and use each estimate as a cross-check on the other—, so we would now have, along with the expert guess of  $\nu$ :

$$\frac{\begin{array}{cc} A & m \\ A \rightarrow B & n(\nu) \end{array}}{B \quad p}$$

(where  $n = 1 - m \cdot (1 - \sigma)$ ), and so

$$(11) \llbracket B \rrbracket = \frac{\llbracket A \wedge B \rrbracket}{\llbracket A|B \rrbracket} = \frac{m + n - 1}{\nu}$$

that naturally fits the (9) bounds (when  $\nu$  runs along from 1 to  $\llbracket A \rrbracket$ ).

## 12 THREE COROLLARIES AND ONE EXTENSION

A few remarks can be made on some apparent advantages of the “logic-as-truth-valuation” approach that, following the advice of Popper *et alii*, we have advocated and described above. First, we mention three well-known fields that have been traditionally perceived as separate but that now automatically become special cases of a single formulation. Then, in subsection *b*, we hint at an obvious extension of the approach.

### 12.1 The special cases

1. *Classical* (two-valued) *logic* is the special case of our general logic in which *every* sentence is *binary* (i.e.  $\forall A \in \mathcal{L} \llbracket A \rrbracket \in \{0, 1\}$ ) or, equivalently, in which *every* sentence has *zero* imprecision (i.e.  $\forall A \in \mathcal{L} f(A) = 0$ ).
2. For *three-valued logics* first note that classical examples, especially Kleene's system of strong connectives [1938] and Łukasiewicz's [1920]

$L_3$ , give the following tables for the values of connectives (where  $U$  stands for “undetermined”):

$\wedge$	0	$U$	1		$\vee$	0	$U$	1		$\rightarrow$	0	$U$	1
0	0	0	0		0	0	$U$	1		0	1	1	1
$U$	0	$X$	$U$		$U$	$U$	$Y$	1		$U$	$U$	$Z$	1
1	0	$U$	1		1	1	1	1		1	0	$U$	1

with  $X = Y = U$  in both Kleene’s and Łukasiewicz’s tables, and  $Z = U$  in Kleene’s (but  $Z = 1$  in Łukasiewicz’s).

Now, if we abbreviate the “ $\llbracket A \rrbracket \in (0, 1)$ ” of our general logic by “ $\llbracket A \rrbracket = U$ ” (as in Kleene or Łukasiewicz) the values given in the above tables coincide exactly with those that would have been computed by our formulas, *except* that  $X$ ,  $Y$  and  $Z$  would remain undetermined until we knew  $\alpha_{AB}$ . In general, our values would match Kleene’s, but in certain cases they would yield differing results:

- (a) If  $\llbracket A \rrbracket + \llbracket B \rrbracket \leq 1$  and  $A$  and  $B$  are *incompatible* (typically because  $\mathbf{A} \cap \mathbf{B} = \emptyset$ ) then  $X = 0$ .
- (b) If  $\llbracket A \rrbracket + \llbracket B \rrbracket \geq 1$  and  $A$  and  $B$  are *incompatible* (typically because  $\mathbf{A} \cup \mathbf{B} = \Theta$ ) then  $Y = 1$ .
- (c) If  $\llbracket A \rrbracket \leq \llbracket B \rrbracket$  and  $A$  and  $B$  are *compatible* (typically because  $\mathbf{A} \subset \mathbf{B}$ ) then  $Z = 1$ .

Note that in the particular case in which  $B$  is  $\neg A$  we have *always*  $\llbracket A \wedge \neg A \rrbracket = 0$  and  $\llbracket A \vee \neg A \rrbracket = 1$  for any valuation, so that, for instance, the three classical Aristotelian principles ( $\vdash A \rightarrow A$ ,  $\vdash \neg(A \wedge \neg A)$  and  $\vdash A \vee \neg A$ ), which do not hold in these logics (except that the first one does in  $L_3$ ), do now hold in ours. These three results are perfectly classical and in full agreement with what is to be expected from a (Boolean) logic.

- (d) *Łukasiewicz & Tarski’s* [1930]  $L_\infty$  logic, as ours, generalizes classical (two-valued) logic in the sense that it allows the members of the sentential lattice to take values in  $[0,1]$  other than 0 or 1. Both systems of logic include classical two-valued logic as a special case. Nevertheless, while  $L_\infty$  renounces *Booleanity*, we renounce *functionality* (though not quite, since each connective is actually truth-functional in *three* arguments:  $\llbracket A \rrbracket$ ,  $\llbracket B \rrbracket$  and a third parameter such as, e.g.,  $\alpha_{AB}$ ). In fact, if the  $\{0, 1\}$  truth set is extended to  $[0,1]$  those two properties of classical logic cannot be both maintained, and one must be sacrificed; we find much easier to justify logically, and more convenient, the sacrifice of truth-functionality.

Our general logic admits  $L_\infty$  as a special case since, indeed,  $L_\infty$  behaves exactly as ours would do if *no* sentence in  $\mathcal{L}$  could be recognized as a *negation* of some other and, then, it would be assigned systematically the maximum compatibility ( $\alpha = 1$ ) connective formulas. Naturally that would give an error in the values of composite sentences involving non-fully-compatible subsentences, but it would also restore the lost truth-functionality of two-valued logic.  $L_\infty$  amounts—from our perspective—to viewing *all* sentences as having *always* maximum mutual compatibility. That means that  $L_\infty$  conceives *all* sentences as *nested* (i.e. for every  $A$  and  $A'$ , either  $\mathbf{A} \subset \mathbf{A}'$  or  $\mathbf{A}' \subset \mathbf{A}$ ) (Sales [1994]). Such a picture is strongly reminiscent of Shafer's [1976] description of conditions present in what he calls *consonant* valuations, and which entails the fiction of a *total*, linear order  $\vdash$  in  $\mathcal{L}$ —and  $\subset$  in  $\mathcal{P}(\Theta)$ —(that means a coherent, negationless universe) ... which is probably the best assumption we can make when information on sentences is lacking and negations are not involved—or cannot be identified as such. (Such an option is as legitimate as that of assuming, in the absence of information on sentences, that these are independent.) In  $L_\infty$  obvious negations may be a problem, but it can be solved by applying error-correcting *supervaluations* (van Fraassen [1968])—that our logic supplies automatically. And note that, in particular, the error incurred in by  $L_\infty$  when failing to distinguish between a sentence and its negation—thus not being able to recognize a contradiction—is just the quantity we called *imprecision* (or curiously, for the historical record, what Black called *vagueness* and defined formally as we did with imprecision, see Sales [1982b]).

### 12.2 Ignorance as subadditivity

If we now suppose that  $\mathcal{L}$  is still in a Boolean algebra but the  $v$  valuation we impose on  $\mathcal{L}$  is *subadditive*, i.e.:

$$\{[A \wedge B]\} + \{[A \vee B]\} \geq \{[A]\} + \{[B]\} \text{ (Subadditivity)}$$

(where we note explicitly by the  $\{[\ ]\}$  brackets that  $v$  is subadditive), then  $v$  can be characterized as a *lower probability* (see Good [1962]) or as a *belief*  $Bel(A)$  (in Shafer's [1976] sense). If the inequality sign were inverted, the valuation would become an *upper probability* or (Shafer's) *plausibility*  $Pl(A)$ . The defective value—the part of the value  $v(A)$  attributed neither to  $A$  nor to  $\neg A$ —can be expressed as:

$$\begin{aligned} \partial(A) &= 1 - (\{[A]\} + \{[\neg A]\}) \\ &= (1 - \{[\neg A]\}) - \{[A]\} \\ &= Pl(A) - Bel(A) \end{aligned}$$

In the finite  $\mathcal{L}$  case we know (Shafer [1976]) that a subadditive valuation like  $v : \mathcal{L} \rightarrow [0, 1] : A \mapsto \{[A]\}$  —or, better, the measure  $\mu : \mathcal{P}(\Theta) \rightarrow [0, 1] : \mathbf{A} \mapsto \mu(\mathbf{A})$  induced on  $\mathcal{P}(\Theta)$  by that valuation— defines a function  $m : \mathcal{P}(\Theta) \rightarrow [0, 1]$  (called “*basic assignment*” by Shafer) that satisfies:

1.  $m(\emptyset) = 0$
2.  $\sum_{\mathbf{A} \subset \Theta} m(\mathbf{A}) = 1$

so that the  $\mu(\mathbf{A})$  ( $= \{[A]\}$ ) values are computed from this measure through

$$\{[A]\} = \sum_{\mathbf{B} \subset \mathbf{A}} m(\mathbf{B})$$

and, conversely, the  $m(\mathbf{A})$  values can be obtained from  $\{[A]\}$  ( $= \mu(\mathbf{A})$ ) through

$$m(\mathbf{A}) = \sum_{\mathbf{B} \subset \mathbf{A}} (-1)^{|\mathbf{A}-\mathbf{B}|} \mu(\mathbf{B}) \text{ for any } \mathbf{A} \subset \Theta.$$

$Bel(A)$  and  $Pl(A)$  happen to coincide with the traditional concept (in Measure Theory) of *inner measure* ( $P_*$ ) and *outer measure* ( $P^*$ ), so that the following chain of equivalences has a transparent meaning (notice Shafer calls  $Bel(\neg A)$  “degree of doubt of  $A$ ”):

$$\begin{aligned} Pl(A) &= P^*(\mathbf{A}) = \sum_{\mathbf{B} \cap \mathbf{A} \neq \emptyset} m(\mathbf{B}) \\ &= \sum_{\mathbf{B} \subset \Theta} m(\mathbf{B}) - \sum_{\mathbf{B} \subset \mathbf{A}^c} m(\mathbf{B}) \\ &= 1 - P_*(\mathbf{A}^c) = 1 - Bel(\neg A) \end{aligned}$$

We know, also, that an *additive* valuation  $\mu$  is just a basic assignment  $m$  such that  $m(\mathbf{A}) = 0$  for all  $\mathbf{A} \subset \Theta$  *except* for the singletons  $\{\theta\}$  of  $\Theta$ . This is what Shafer calls ‘*Bayesian belief*’.

Subadditive belief derives from “non-rational” valuations of evidence by a reasoner (in the Ramsey/de Finetti sense). It can model and explain situations like this classic result: Confronted with the question “Should the Government allow public speeches against democracy?”, one user assented 25% of the time. Substituting the word “prohibit” for “allow” elicited a 54% of assenting responses. Since both words are antonyms (the contrary of prohibiting is allowing), it is clear that this user had an unattributed gap left between those two complementary concepts, thus:

$$\{[A]\} + \{[\neg A]\} = .25 + .54 \leq 1$$

which reveals that sentence  $A$  (=speeches allowed) was being valued *sub-additively*, and also that “allow” and “fail to prohibit” are here analyzable, in Shafer’s terms, as  $Bel(A)$  and  $Pl(A)$ , respectively, with values .25 (for belief) and .46 (for plausibility).

*Ignorance* (or, rather *total ignorance*) is the particular instance of sub-additive valuation in which all non-*true* sentences get the zero value, i.e.

$$\{[A]\} = 1 \quad \text{iff } A = \top, \quad \{[A]\} = 0 \quad \text{otherwise}$$

(this is what Shafer calls ‘*vacuous* belief function’). It is just the particular instance of valuation in which all non-*true* sentences get the zero value: indeed, we have, for a given  $A$ ,  $\{[A]\} = \{[\neg A]\} = 0$ . In a strict parallel with *total ignorance*, subadditive valuations can also adequately formalize *total certainty* (meaning that  $\{[A]\} = 1$  while at the same time  $\{[B]\} = 0$  for all  $B \vdash A$ ).

Subadditivity enables us to analyze other interesting situations related to Logic. For instance, suppose we have a subadditive valuation assigning  $A$  the value  $\mu(\mathbf{A}) = \{[A]\} = Bel(A)$ , and that a set  $\mathbf{A}'$  (not necessarily a subset of  $\mathbf{A}$ ) can be found in  $\mathcal{P}(\Theta)$  such that there is an *additive* valuation which assigns  $\mathbf{A}'$  precisely the same value. We denote  $\mathbf{A}'$  by “ $\square\mathbf{A}$ ” (where  $\square$  is a set-operator) and  $A' = \delta(\mathbf{A}')$  by “ $\square A$ ”. So we have:

$$Bel(A) = \{[A]\} = \llbracket \square A \rrbracket$$

The “ $\square$ ” is here a linguistic operator that acts on a sentence  $A$  and transforms it into another whose value is the belief (= subadditive truth-value) one can assign non-additively to  $A$ , so it seems proper to interpret “ $\square$ ” syntactically as “it is believed that”, and “ $\square_\alpha$ ” or “ $[\alpha]$ ” as “ $\alpha$  (= the name of a subject or agent) believes that”. Further, we would have

$$\begin{aligned} Pl(A) &= 1 - Bel(\neg A) = 1 - \{[\neg A]\} \\ &= 1 - \llbracket \square \neg A \rrbracket = \llbracket \neg \square \neg A \rrbracket = \llbracket \diamond A \rrbracket \end{aligned}$$

where we have defined a new operator “ $\diamond$ ” as an abbreviation for “ $\neg \square \neg$ ” to be interpreted as syntactically as “it is *plausible* that” or “ $\alpha$  admits as credible” (so that “ $\diamond_\alpha A$ ” —or  $\langle \alpha \rangle A$ — would read “ $\alpha$  finds that  $A$  can be believed”), because  $\alpha$  just does not believe the contrary.

Dubois & Prade [1987] speak rather of “necessity” (or “degree of knowledge”) and write  $Nec(A) =_{df} \{[A]\}$  (=  $\llbracket \square A \rrbracket$ ), and “possibility” (or “degree of admissibility”)  $Poss(A) =_{df} 1 - \{[\neg A]\}$  (=  $\llbracket \diamond A \rrbracket$ ); naturally,  $Nec(A) \leq \llbracket A \rrbracket \leq Poss(A)$  for any  $A$  (and also:

$$Nec(A) + Poss(\neg A) = Poss(A) + Nec(\neg A) = 1).$$

If necessity (or knowledge) of  $A$  and  $\neg A$  are totally incompatible (in the sense that  $Nec(\neg A) = 0$  whenever  $Nec(A) > 0$ ) then  $Nec(A \vee B) = \max(Nec(A), Nec(B))$  and  $Poss(A \wedge B) = \min(Poss(A), Poss(B))$ .

It is interesting to notice that a subadditive valuation may be superimposed on a sentential lattice without breaking its Boolean character, so that  $A \wedge \neg A = \perp$  (bivalence) and  $A \vee \neg A = \top$  (excluded middle) still hold,

while at the same time  $\{[A]\} + \{[\neg A]\} \leq 1$ . This slightly paradoxical fact may explain that many often-encountered situations —where mere subadditivity ( $\{[A]\} + \{[\neg A]\} \leq 1$ ) was probably the case— have been analyzed historically as invalidating the law of the *excluded middle*, because it was felt that there was a “third possibility” between  $A$  and  $\neg A$  making for the unattributed value  $1 - \{[A]\} - \{[\neg A]\}$ , covered neither by  $A$  nor  $\neg A$ . If our analysis is correct, such situations are analyzable in terms of incomplete valuations, but this does not imply the breaking of bivalence of any Boolean algebra (except, naturally, for intuitionistic logic, where the algebra is explicitly non-Boolean).

Subadditivity valuations on a Boolean algebra allow also analysis not only of the concept of *ignorance* and *certainty* (as we sketched above) but of the paradoxes of Quantum Logic as well. These arise, according to the ‘Quantum Logic’ proponents (e.g. Reichenbach in 1944), in explaining why the distributivity fails in this logic, following the standard interpretation of certain experimental results where:

$$p(a) \cdot p(b|a) + p(\bar{a}) \cdot p(b|\bar{a}) < p(b)$$

a relationship we write in this way:

$$(12) \{[A]\} \cdot \{[B|A]\} + \{[\neg A]\} \cdot \{[B|\neg A]\} < \{[B]\}.$$

What is odd is that this inequality is normally interpreted by quantum logicians (e.g. Watanabe) as meaning:

$$(13) (A \wedge B) \vee (\neg A \wedge B) \neq B$$

which obviously signals the breaking of distributivity. However, the even-handed transcription of the value-version (12) into the algebraic one (13) is clearly abusive: (13) is much stronger than (12), since it is equivalent to requiring that (12) hold for *all* conceivable valuations. Actually, in our notation, the reading of the experimental results translates immediately into (12), not (13). From there we conclude that:

$$\{[A \wedge B]\} + \{[\neg A \wedge B]\} < \{[B]\}$$

which is in clear violation of additivity, but not of distributivity. So we need not consider that quantum phenomena occur in a non-Boolean algebra (orthomodular lattices are the preferred alternatives) because a *sub-additively*-valued Boolean lattice surely would do for most quantum-logic applications.

Finally we mention that subadditive valuations can as well satisfactorily model how a scientific explanation frame (i.e. the appropriate lattice of theories that cover any observed or predictable true fact) is dynamically replaced by another once the first can no longer account for observed facts: as the valuation of explanations turns subadditive —reflecting that they

no longer cover all predictable facts— one is naturally forced to replace the original sentential structure of theories by a new one (still a Boolean lattice) provided with a new —now again additive— valuation on it that restores the balance; the augmented lattice generators are the new vocabulary, and the elementary components of the new structure are the required new explanatory elements (atomic theories) for the presently observed facts. Such a simple valuation-revision and lattice-replacement mechanism may serve to illustrate the basic dynamics of theory change in scientific explanation (see Sales [1982b]).

## APPENDIX

### A ON FUZZY LOGIC

#### A.1 The “Fuzzy Logic” tradition

In a joint reflection with Richard Bellman in 1964 Lotfi Zadeh, then a process control engineer, considered the nonsense of painstakingly computing numerical predictions in control theory contexts where the situation is complex enough to render them meaningless. So he proposed instead to rely confidently on broad —and inherently *vague*— linguistic descriptions like ‘high’ (for a temperature) or ‘open’ (for a valve) rather than on misleadingly precise values. He then went to suggest (in Zadeh [1965]) a non-standard extensional interpretation of Predicate Calculus. Though first advanced by Karl Menger (in a 1951 note to the French Académie entitled *Ensembles flous et fonctions aléatoires*), the idea was nevertheless original and simple: an atomic predicate sentence like ‘the temperature is high’ is assigned truth values in  $[0,1]$  reflecting the applicability of the sentence to circumstances (the actual temperatures); those values then define a “set”, a “fuzzy” set, by considering them to be the values of a generalized *characteristic* or *set membership* function, in a way that is strictly parallel to the standard procedure for defining predicate extensions (e.g. of ‘prime number’) by equating the  $\{0,1\}$  truth values with the characteristic function of the set (i.e. ‘the prime numbers’) so that if for instance  $\llbracket \text{Prime}(3) \rrbracket = 1$  then  $\chi_{\mathbf{Primes}}(3) = 1$ , i.e.  $3 \in \mathbf{Primes}$  (and so now, accordingly, if e.g.  $\llbracket \text{High}(170) \rrbracket = 0.8$  then  $\chi_{\mathbf{High\_temps}}(170) = 0.8$  or  $170 \in_{0.8} \mathbf{High\_temps}$ ).

Some snags soon arose to question the utter simplicity of the scheme, doubts such as: should set inclusion (defined as  $\mathbf{A} \subset \mathbf{B}$  iff  $\forall x \chi_{\mathbf{A}}(x) \leq \chi_{\mathbf{B}}(x)$ ) fail if a single point  $x$  does not satisfy the relation? or, more fundamentally: what are the appropriate formulas for the connectives? Zadeh had first proposed the usual Lukasiewicz connective formulas (the *min* and the *max* for the  $\wedge$  and  $\vee$ ) but then (in 1970, again with R. Bellman) considered that these were “non-interactive” and to be preferred only in the absence of more relevant information, so he offered further formulas he called “soft”



or “interactive” (the product and sum-minus-product). Bellman and Gierz [1973] showed that under certain pre-established conditions (notably, *truth-functionality*) the only possible connectives were the *min* and the *max*. For a (standard) logician this is an unfortunate result, since the *min* formula when applied to a (0,1)-valued sentence  $A$  and its negation  $\neg A$  yields always a non-zero value ( $\llbracket A \wedge \neg A \rrbracket = \min(\llbracket A \rrbracket, 1 - \llbracket A \rrbracket) \neq 0$ ), so explicitly *negating* the classical law of non-contradiction (never questioned before by any Logic) and thus placing Fuzzy Logic outside the standard logic mainstream, for which  $\neg(A \wedge \neg A)$  is always guaranteed theoremhood (and so, necessarily,  $\llbracket A \wedge \neg A \rrbracket = 0$  for any valuation —provided 1 is the only designated value, as we assume). Another unfortunate by-product, this one algebraic in character, is that the neatness and simplicity of the Boolean algebra structure —preservable only by sacrificing truth-functionality— are irrecoverably lost.

Note that (a) Boolean structure had to be sacrificed in Fuzzy Logic just by technical reasons (the *min* formula), not by a deliberate or methodological bias, and that (b) the choice of connectives was historically motivated, in fuzzy-set theory, by pragmatic reasons (prediction accuracy in applications) rather than by logical method: thus, many formulas were tried and discussed (see e.g. Rodder [1975] and Zimmerman [1977]). (Interestingly, while theoreticians stuck to Łukasiewicz’s *min*, practitioners —e.g. Mamdani [1977]— preferred the product, perhaps recognizing that in complex or poorly-known systems the best policy is to suppose sentences *independent* —in our sense, see above—; whence the product.) Anyway, after a brief flurry of discussion in the 1970s about the right connectives, fuzzy-set theory proceeded from then onwards *non-compatibly* by emphasizing that its basic theme is linguistic *vagueness*, not *logic* or *uncertainty*, and that the [0,1]-values are grounded on a (possibly non-additive) valuation called “possibility”, for which a complete subtheory has been elaborated since.

The initially intuitive “fuzzy logic” approach has since become an independent growth industry. From a logical point of view, some foundational points are arguable: (1) the apparently undisputable *truth-functionality* requirement, already present in Łukasiewicz, imposes a radical departure from (ordinary) logic: the sentences *cannot* form any longer a Boolean structure, and traditional logic principles are gone forever; (2) the theory’s justification for using (0,1)-values (values that reflect *imprecision*, since  $\llbracket A \rrbracket \in (0, 1)$  implies  $f(A) > 0$ ) is strictly linguistic: the cause of imprecision is attributed solely to the *vagueness* of language and explicitly excludes any non-linguistic component or dimension such as *uncertainty* (considered non-intersectingly to be the domain of Probability theory) or simply *approximation*.

The emphasis fuzzy theorists place on *vagueness* and its “orthogonality” with respect to other causes of unattributed truth value is understandable considering the basic tenets of the theory but questionable from a non-partisan stand.

First: in all cases we finally obtain a number, and this has a transparent function in reasoning: keeping track of our confidence in what we say. In the vagueness case the value we assign is clearly the *degree of applicability* of the sentence to the circumstances at hand (and this is seen by executing the (T) schema); in the uncertainty case it is our “belief” or its “probability” (in some more or less standard sense) what emerges from the (T) evaluation procedure. In either case (*vagueness* or *uncertainty*), what we try to capture and measure is the *degree of approximation* (to truth, or to full reliability) that we can confidently assign the sentence, and what we have in both cases is *imprecision* (difficulties in ascertaining the  $\{0,1\}$  truth value and also, consequently, doubts about committing ourselves to it along a whole inference process). Plausibly, it is easier to assign values consistently (in the  $[0,1]$  interval) to the sentence, regardless of where or why imprecision arose in the first place, because what we are mainly interested in is the *degree of confidence* we attach to this piece of information we are manipulating through the (hopefully truth-preserving) inferences.

Second: the two dimensions of imprecision, *linguistic* and *epistemic* (for vagueness and uncertainty, resp.), are not so separable as claimed, either conceptually or practically. (We do not consider here occasional claims – often made by Zadeh – that vagueness is *in things* –even in truth–, i.e. it is not linguistic, but ontological). As seen through application of the Tarski (T) schema, when the agent is incapable of making up his/her own mind as to the truth of the (unquoted) sentence, the *cause* for the under-attribution and the origin of the  $\varepsilon$  residual value need not be considered, only the resulting *confidence* matters. The non-attribution of “normal” (i.e.  $\{0,1\}$ ) truth values may originate in language (i.e. the sentence is *vague*) or in imperfect verification conditions (i.e. the sentence, even being linguistically precise, is nevertheless *uncertain* due to identification or measurement difficulties or other causes), but spotting the source is mostly an academic exercise: consider, for example, the sentence ‘this is probable’, which is imprecise in either or both senses, or the historically motivating illustration by Bellman/Zadeh (the ‘high temperature’ process-control case), where a discrimination of origins, either linguistic (the expression is vague) or epistemic (we don’t know what is the precise case, or we have difficulties in identifying it) is indifferent or pointless.

Moreover, not only such boundary examples cast doubts on the claimed vagueness/uncertainty orthogonality; even the *linguistic* character of fuzziness (or of “possibility”) is arguable. First, because a vague sentence, that to the utterer may mean just that the expression lacks straightforward *applicability* to actual fact, to the hearer –if not involved in the described situation– it may be *epistemic* information about what to expect (for instance, hearing a temperature is “high” may set the unknowing hearer into a –correspondingly imprecise– alert state; (s)he then even may usefully turn

the “membership” or “possibility” number into an *a priori*-probability or belief estimate). Second, because the number we assign to approximate membership in a class, which reflects the *applicability* of the sentence to the observed situation (a semantic quantity shared by the speakers of the language), is *constructed* and continually adjusted by the language speakers on the base of their experience of past cases, in a process which is the same as that of constructing all other  $\llbracket A \rrbracket$  assignments, however called (“truths”, “degrees”, “applicabilities”, “probabilities”, “beliefs”, “approximation” or whatever). The process, described above, consists in setting a universe  $\Theta$  (here the application instances of the sentences in the language) and then considering cases  $\theta \in \rho(A)$  (here the  $\theta$ s are application instances – *utterances*– of  $A$ ) from past experience; the weighed result is  $\llbracket A \rrbracket = \mu(\rho(A))$  (here the *applicability* of the –vague– sentence  $A$ , i.e. its “fuzziness”), and the relationship between  $A$  and all other sentences can also be user-evaluated through the compatibilities  $\alpha$  or the sufficiency/causality degrees  $\sigma$  introduced in the text (in a way that otherwise amounts to Popper’s ‘binary probability’ evaluation process).

This process of *constructing* values may be considered a further instance of our general approach to rationality based on the universe  $\Theta$  of *cases*, and the resulting number  $\llbracket A \rrbracket$  may be used in general reasoning as all other “truth values” are. We thus assure compatibility inside a shared formalism from which Logic, Probability and (e.g.) fuzzy reasoning or belief theories can be derived directly without costly or unnatural translations. To do this we merely need that the different interpretations (including the “fuzzy” one) obey the same laws and have the same formal components: (1) a common sentential language equipped with a *Boolean* structure (easy to justify and convenient for preserving the commonly accepted laws of logic), (2) *coherence* in attributing the values (regardless of the meaning we give them) to assure *additivity*, and (3) a predisposition to sacrifice truth-functionality when required. If the notion of fuzziness, however justified, could be conceived in this way, then the original 1964 Bellman/Zadeh proposal could be subsumed and solved in a very natural way, and we probably could do without a separate, costly and incompatible additional formalism. (In other words, we probably wouldn’t need “fuzzy logic”.)

## A.2 Gaines’s ‘Standard Uncertainty Logic’

In an interesting effort, born inside the “fuzzy” tradition, to construct a common formalization by discriminating between algebraic *structures* and truth *valuations* on them —somewhat paralleling and anticipating the aim of our present development— Gaines [1978] set out to define what he called ‘Standard Uncertainty Logic’, that covered and formalized two known sub-cases. This logic postulates (a) a proposition lattice with an algebraic structure that is initially assumed *Boolean* but —by technical reasons—

finally admitted to be merely distributive, and (b) a finitely *additive* valuation of the lattice on the  $[0,1]$  interval. Gaines then distinguishes two special cases of his logic: (1) what he calls ‘probability logic’ and defines to be the particular case of his logic where the law of the excluded middle holds, and (2) what he terms ‘fuzzy logic’, defined (in one among several alternative characterizations) to be his logic when that law does not hold. Apparently, Gaines’s encompassing logic should correspond to our general logic above (that covers classical logic as a special case, just as Gaines’s logic becomes his ‘probability logic’ when “all propositions are binary”), but closer examination reveals that Gaines’s confusingly called ‘probability logic’ turns out to be actually coextensional with *classical* logic, while his ‘fuzzy logic’ is, simply, *Lukasiewicz’s*  $L_\infty$ . This fact is spelled out by the property he mentions of propositional equivalence (here in our notation):

$$\llbracket A \leftrightarrow B \rrbracket = \min(1 - \llbracket A \rrbracket + \llbracket B \rrbracket, 1 - \llbracket B \rrbracket + \llbracket A \rrbracket)$$

in which the right hand is clearly  $1 - |\llbracket A \rrbracket - \llbracket B \rrbracket|$ . This is an expression that is deduced from Gaines’s postulates *only* if, necessarily, either  $A \vdash B$  or  $B \vdash A$  (note this either/or condition—exactly corresponding to our  $\alpha_{AB} = 1$  full-compatibility situation—is explicitly mentioned by Gaines as a characteristic property of his ‘fuzzy logic’). So Gaines makes the implicit assumption that the base lattice is *linearly* ordered (perhaps induced to it by the  $\leq$  symbol used for the –partial– propositional order in the lattice). A confirmation for this comes from the fact that classical logic principles—that hold, by definition, in his ‘probability logic’—do not hold in this one. No wonder, then, the base lattice cannot be but merely distributive.

*Departament de Llenguatges i Sistemes Informàtics, Universitat Politècnica de Catalunya. Spain.*

## BIBLIOGRAPHY

- [Adams, 1968] E. W. Adams. Probability and the logic of conditionals, in: J. Hintikka & P. Suppes, eds. *Aspects of inductive logic*, North Holland, 1968.
- [Aronson *et al.*, 1980] A. R. Aronson, B. E. Jacobs, and J. Minker. A note on fuzzy deduction, *Journal of the Assoc. for Computing Machinery*, **27**, 599–603, 1980.
- [Bacchus, 1990] F. Bacchus. *Representing and reasoning with probabilistic knowledge*, MIT Press, 1990.
- [Bar-Hillel and Carnap, 1952] Y. Bar-Hillel and R. Carnap. An outline of a theory of semantic information, TR 247 Research Lab. Electronics, MIT, 1952; reproduced in: Y. Bar-Hillel, *Logic and information*, Addison Wesley (1964)
- [Bellman and Gierz, 1973] R. Bellman and M. Gierz. On the analytical formalism of the theory of fuzzy sets, *Info. Sci.*, **5**, 149–156, 1973.
- [Bellman and Zadeh, 1970] R. Bellman and L. Zadeh. Decision-making in a fuzzy environment, *Management Science*, **17**, 141–162, 1970.
- [Bellman and Zadeh, 1976] R. Bellman and L. Zadeh. Local and fuzzy logics, in: J. M. Dunn & G. Epstein, eds., *Modern Uses of Multiple-Valued Logic*, Reidel, 1976.
- [Black, 1937] M. Black. Vagueness, *Phil. Sci.*, **4**, 427–455, 1937.

- [Bolc and Borowik, 1992] L. Bolc and P. Borowik. *Many-valued logics*, Springer, 1992.
- [Boole, 1854] G. Boole. *An Investigation of the Laws of Thought*, Dover, N.Y. (1958), 1954.
- [Carnap, 1950] R. Carnap. *Logical Foundations of Probability*, U. Chicago Press, 1950.
- [Carnap, 1962] R. Carnap. The aim of inductive logic, in: E. Nagel, P. Suppes & A. Tarski, eds., *Logic, Methodology and Philosophy of Science*, Stanford, 1962.
- [Cohen, 1990] J. Cohen. What I have learned (so far), *Am. Psychologist*, 1307–1308, 1990.
- [Cox, 1961] R. T. Cox. *The algebra of probable inferences*, Johns Hopkins U. Press, 1961.
- [de Finetti, 1931] B. de Finetti. Sul significato soggettivo della probabilità, *Fundamenta Mathematicae*, **17**, 298–329, 1931.
- [de Finetti, 1937] B. de Finetti. La prévision, ses lois logiques, ses sources subjectives, *Annales de l'Institut Henri Poincaré*, **7**, 1–68, English translation in: H. E. Kyburg Jr. & H.E. Smokler, eds., *Studies in subjective probability*, John Wiley (1964), 1931.
- [de Finetti, 1970] B. de Finetti. *Teoria delle probabilità*, Einaudi, English translation: *Theory of probability*, John Wiley (1974), 1070.
- [de Luca and Termini, 1972] A. de Luca and S. Termini. A definition of a non-probabilistic entropy in the setting of fuzzy sets theory, *Info. and Control*, **20**, 301, 1972.
- [Dubois and Prade, 1987] D. Dubois and H. Prade. *Théorie des possibilités*, Masson (2nd edition), 1987.
- [Dubois et al., 1993] D. Dubois, J. Lang and H. Prade. Possibilistic logic, in: S. Abramsky, D. Gabbay & T.S. Maibaum, eds., *Handbook of Logic in Artificial Intelligence* (Vol. 3), Oxford Univ. Press, 1993.
- [Duda et al., 1976] R. Duda, P. Hart and N. Nilsson. Subjective Bayesian methods for rule-based information systems, in: B.W. Webber & N. Nilsson, eds., *Readings in Artificial Intelligence*, Morgan Kaufmann, San Mateo (1981), 1976.
- [Fagin et al., 1990] R. Fagin, J. Y. Halpern and N. Megiddo. A logic for reasoning about probabilities, *Information & Computation*, *bf* 87, 78–128, 1990.
- [Fenstad, 1967] J. E. Fenstad. Representations of probabilities defined on first order languages, in: J.N. Crossley, ed., *Sets, Models and Recursion Theory*, North Holland, 1967.
- [Fenstad, 1968] J. E. Fenstad. The structure of logical probabilities, *Synthese*, **18**, 1–23, 1968.
- [Fenstad, 1980] J. E. Fenstad. The structure of probabilities defined on first order languages, in: R.C. Jeffrey, ed., *Studies in inductive logic and probability II*, U. of California Press, 1980.
- [Fenstad, 1981] J. E. Fenstad. Logic and probability, in: E. Agazzi, ed., *Modern Logic. A survey*, Reidel, 1981.
- [Field, 1977] H. H. Field. Logic, meaning and conceptual role, *J. of Phil. Logic*, **74**, 379–409, 1977.
- [Friedman and Halpern, 1995] N. Friedman and J. Y. Halpern. Plausibility measures; a user's guide (draft available on WWW at <http://robotics.stanford.edu>), 1995.
- [Gaifman, 1964] H. Gaifman. Concerning measures in first order calculi, *Israel J. of Mathematics*, **2**, 1–18, 1964.
- [Gaines, 1978] B. R. Gaines. Fuzzy and probability uncertainty logics, *Info. and Control*, **38**, 154–169, 1978.
- [Garbolino et al., 1991] P. Garbolino, H. E. Kyburg, et al.. Probability and logic, *J. of Applied Non-Classical Logics*, **1**, 105–197, 1991.
- [Gärdenfors, 1988] P. Gärdenfors. *Knowledge in flux*, MIT Press, 1988.
- [Genesereth and Nilsson, 1987] M. Genesereth and N. Nilsson. *Logical Foundations of Artificial Intelligence*, Morgan Kaufmann, 1987.
- [Gerla, 1994] G. Gerla. Inferences in probability logic, *Artificial Intelligence*, **70**, 33–52, 1994.
- [Good, 1950] I. J. Good. *Probability and the weighting of evidence*, Griffin, London, 1950.

- [Good, 1962] I. J. Good. Subjective probability as the measure of a non-measurable set, in: E. Nagel, P. Suppes & A. Tarski, eds., *Logic, Methodology and Philosophy of Science*, Stanford, 1962.
- [Gottwald, 1989] S. Gottwald. *Mehrwertige Logik*, Akademie-Verlag, 1989.
- [Grosz, 1986] B. Grosz. Evidential confirmation as transformed probability, in: Kanal, L.N., & Lemmer, J.F., eds. : 1986, *Uncertainty in Artificial Intelligence*, North Holland, 1986.
- [Haack, 1974] S. Haack. *Deviant logic*, Cambridge, 1974.
- [Hailperin, 1984] T. Hailperin. Probability logic, *Notre Dame J. of Formal Logic*, **25**, 198–212, 1984.
- [Hájek, 1996] P. Hájek. Fuzzy sets and arithmetical hierarchy, *Fuzzy Sets and Systems* (to appear)
- [Halpern, 1990] J. Y. Halpern. An analysis of first order logics of probability, *Artificial Intelligence*, **46**, 311–350, 1990.
- [Hamacher, 1976] H. Hamacher. On logical connectives of fuzzy statements and their affiliated truth function, *Proc. 3rd Eur. Meet. Cyber. & Systems Res.*, Vienna, 1976.
- [Harper, 1975] W. L. Harper. Rational belief change, Popper functions, and counterfactuals, *Synthese*, **30**, 221–262, 1975.
- [Heckermann, 1986] D. Heckermann. Probabilistic interpretations for MYCIN's certainty factors, in: Kanal, L.N., & Lemmer, J.F., eds. : 1986, *Uncertainty in Artificial Intelligence*, North Holland, 1986.
- [Hintikka and Pietarinen, 1968] J. Hintikka and J. Pietarinen. Probabilistic inference and the concept of total evidence, in: J. Hintikka & P. Suppes, eds. *Aspects of inductive logic*, North Holland, 1968.
- [Horn and Tarski, 1948] A. Horn and A. Tarski. Measures in Boolean algebras, *Trans. Am. Math. Soc.*, **64**, 467–497, 1948.
- [Kanal and Lemmer, 1986] L. N. Kanal and J. F. Lemmer, eds. *Uncertainty in Artificial Intelligence*, North Holland, 1986.
- [Keynes, 1921] J. M. Keynes. *A treatise on probability*, Macmillan, NY, 1921.
- [Kleene, 1938] S. C. Kleene. On notation for ordinal numbers, *J. of Symbolic Logic* **3**, 150–155, 1938.
- [Kolmogorov, 1933] A. N. Kolmogorov. *Grundbegriffe der Wahrscheinlichkeitsrechnung*, Springer, English translation: *Foundations of the theory of probability*, Chelsea, NY (1956), 1933.
- [Koopman, 1940] B. O. Koopman. The bases of probability, *Bulletin Am. Math. Soc.*, **46**, 763–774, 1940.
- [Koppelberg *et al.*, 1989] S. Koppelberg, J. D. Monk and R. Bonnet, eds. *Handbook of Boolean algebras* (3 vols.), North Holland, 1989.
- [Kyburg, 1993] H. E. Kyburg Jr. Uncertainty logics, in: S. Abramsky, D. Gabbay & T.S. Maibaum, eds., *Handbook of Logic in Artificial Intelligence* (Vol. 3), Oxford Univ. Press, 1993.
- [Jeffrey, 1965] R. Jeffrey. *The logic of decision*, North Holland, 1965.
- [Jeffrey, 1995] R. Jeffrey. From logical probability to probability logic, *10th Int. Congress of Logic, Methodology & Phil. of Sc.*, Aug. 19–25, 1995, Florence, 1995.
- [Jeffreys, 1939] H. Jeffreys. *Theory of probability*, Oxford, 1939.
- [Johnson-Laird, 1975] P. N. Johnson-Laird. Models of deduction, in: R.J. Falmagne, ed., *Reasoning: Representation and process in children and adults*, L. Erlbaum, 1975.
- [Johnson-Laird, 1983] P. N. Johnson-Laird. *Mental models*, Cambridge Univ. Press, 1983.
- [de Laplace, 1774] P. S. de Laplace. Mémoire sur la probabilité des causes par les événements, *Mémoires de l'Académie des Sciences de Paris*, Tome **VI**, 621, 1774.
- [de Laplace, 1820] P. S. de Laplace. *Théorie analytique des probabilités* (3rd ed.), Courcier, 1820.
- [Leblanc, 1960] H. Leblanc. On a recent allotment of probabilities to open and closed sentences, *Notre Dame J. of Formal Logic*, **1**, 171–175, 1960.
- [Leblanc, 2001] H. Leblanc. Alternatives to standard first-order semantics, in: D. Gabbay & F. Guenther, eds., *Handbook of Philosophical Logic*, Second Edition, Volume 2, pp. 53–132 Kluwer, Dordrecht, 2001.

- [Lee, 1972] R. C. T. Lee. Fuzzy logic and the resolution principle, *Journal of the Assoc. for Computing Machinery*, **19**, 109–110, 1972.
- [Lewis, 1976] D. Lewis. Probabilities of conditionals and conditional probabilities, *Philosophical Review*, **85**, 297–315, 1976.
- [Łoś, 1962] J. Łoś. Remarks on foundations of probability, *Proc. Int. Congress of Mathematicians*, Stockholm, 1962.
- [Lukasiewicz, 1920] J. Lukasiewicz. O logice trójwartościowej, *Ruch Filozoficzny* (Lwów), **5**, 169–171, 1920; see L. Borkowski, ed., *J. Lukasiewicz: Selected works*, North Holland (1970).
- [Lukasiewicz, 1938] J. Lukasiewicz. Die Logik und das Grundlagenproblem, in: F. Gonseth, ed., *Les entretiens de Zurich, 1938*, Leemann (1941) (pages 82–100, discussion pages 100–108).
- [Lukasiewicz and Tarski, 1930] J. Lukasiewicz and A. Tarski. Untersuchungen über den Aussagenkalkül, *Comptes Rendus Soc. Sci. & Lettres Varsovie*, Classe **III**, 51–77, 1930; translated by J.H. Woodger in: Alfred Tarski, *Logic, Semantics, Metamathematics*, Oxford (1956).
- [Mamdani, 1977] E. H. Mamdani. Application of fuzzy logic to approximate reasoning using linguistic synthesis, *IEEE Trans. on Computers*, **26**, 1182–1191, 1977.
- [MacColl, 1906] H. MacColl. *Symbolic Logic and its Applications*, London, 1906.
- [Miller, 1978] D. W. Miller. On distance from the truth as a true distance, in: J. Hintikka, I. Niiniluoto & E. Saarinen, eds., *Essays in Mathematical and Philosophical Logic*, Reidel, 1978.
- [Miller, 1981] D. W. Miller. A Geometry of Logic, in: H. Skala, S. Termini & E. Trillas, eds., *Aspects of Vagueness*, Reidel, 1981.
- [Nilsson, 1986] N. Nilsson. Probabilistic logic, *Artificial Intelligence*, **28**, 71–87, 1986.
- [Nilsson, 1993] N. Nilsson. Probabilistic logic revisited, *Artificial Intelligence*, **59**, 39–42, 1993.
- [Paass, 1988] G. Paass. Probabilistic logic, in: P. Smets, E.H. Mamdani, D. Dubois & H. Prade, eds., *Non-standard logics for automated reasoning*, Academic Press, 1988.
- [Pavelka, 1979] J. Pavelka. On fuzzy logic I, II, III, *Zeitschrift für mathematische Logik und Grundlagen der Mathematik*, **25**, 45–52, 119–134, 447–464, 1979.
- [Pearl, 1988] J. Pearl. *Probabilistic reasoning in intelligent systems*, Morgan Kaufmann, San Mateo, 1988.
- [Peirce, 1902] C. S. Peirce. *Minute Logic*, unpublished, 1902; (see ref. also in Rescher [1969]), in: C.S. Peirce, *Collected Papers*, Harvard (1933)
- [Popper, 1959] K. R. Popper. *The Logic of Scientific Discovery*, Hutchinson, 1959; (includes Appendix II, originally appeared in *Mind*, 1938, and Appendix IV, originally appeared in *Br. J. Phil. Sc.*, 1955)
- [Popper, 1962] K. R. Popper. Some comments on truth and the growth of knowledge, in: E. Nagel, P. Suppes & A. Tarski, eds., *Logic, Methodology and Philosophy of Science*, Stanford, 1962.
- [Popper, 1972] K. R. Popper. Two faces of commonsense, in: K. R. Popper, *Objective Knowledge*, Oxford, 1972.
- [Popper and Miller, 1987] K. R. Popper and D. W. Miller. Why probabilistic support is not inductive, *Phil. Trans. R. Soc. Lond.*, A **321**, 569–591, 1987.
- [Ramsey, 1926] F. P. Ramsey. Truth and probability, 1926; in: Ramsey, F. P.: 1931, *The Foundations of Mathematics*, Harcourt Brace, NY.
- [Reichenbach, 1935a] H. Reichenbach. Wahrscheinlichkeitslogik, *Erkenntnis*, **5**, 37–43, 1935.
- [Reichenbach, 1935] H. Reichenbach. *Wahrscheinlichkeitslehre*, 1935; English translation: *The Theory of Probability*, U. of California Press (1949)
- [Rescher, 1969] N. Rescher. *Many-Valued Logic*, McGraw-Hill, 1969.
- [Rodder, 1975] W. Rodder. On ‘And’ and ‘Or’ connectives in Fuzzy Set Theory, T. Report, *I. für Wirtschaftswiß.*, Aachen, 1975.
- [Sales, 1982a] T. Sales. Una lògica multivalent booleana, *Actes 1er Congrès Català de Lògica Matemàtica* (Barcelona, January 1982), 113–116, 1982.
- [Sales, 1982b] T. Sales. *Contribució a l’anàlisi lògica de la imprecisió*, Universitat Politècnica de Catalunya (Ph. D. dissertation), 1982.

- [Sales, 1992] T. Sales. *Propositional logic as Boolean many-valued logic*, TR LSI-92-20-R, Universitat Politècnica de Catalunya, 1992.
- [Sales, 1994] T. Sales. Between logic and probability, *Mathware & Soft Computing*, **1**, 99–138, 1994.
- [Sales, ] T. Sales. Logic of assertions, *Theoria*, **25**, 1996.
- [Savage, 1954] L. J. Savage. *Foundations of Statistics*, Dover, NY (1972), 1954.
- [Scott, 1973] D. Scott. Does many-valued logic have any use?, in: S. Körner, ed., *Philosophy of Logic*, Blackwell (1976), 1973.
- [Scott and Kraus, 1968] D. Scott and P. Kraus. Assigning probabilities to logical formulas, in: J. Hintikka & P. Suppes, eds., *Aspects of Inductive Logic*, North Holland, 1968.
- [Shafer, 1976] G. Shafer. *A Mathematical Theory of Evidence*, Princeton, 1976.
- [Shafer, 1996] G. Shafer. A unified semantics for logic and probability, *Artificial Intelligence and Mathematics*, Fort Lauderdale, Fla., Jan. 3-5 1996.
- [Shortliffe, 1976] E. H. Shortliffe. *MYCIN*, American Elsevier, 1976.
- [Smith, 1961] C. A. B. Smith. Consistency in statistical inference and decision, *J. of the Royal Stat. Soc.*, Ser. B **23**, 218–258, 1961.
- [Stalnaker, 1970] R. Stalnaker., R.: 1970, Probability and conditionals, *Philosophy of Science*, **37**, 64–80, 1970.
- [Suppes, 1968] P. Suppes. Probabilistic inference and the concept of total evidence, in: J. Hintikka & P. Suppes, eds., *Aspects of Inductive Logic*, North Holland, 1968.
- [Suppes, 1979] P. Suppes. *Logique du probable*, Flammarion, Paris, 1979.
- [Tarski, 1935a] A. Tarski. Wahrscheinlichkeitslehre und mehrwertige Logik, *Erkenntnis*, **5**, 174–175, 1935.
- [Tarski, 1935b] A. Tarski. The concept of truth in formalized languages, in: J.H. Woodger in: Alfred Tarski, *Logic, Semantics, Metamathematics*, Oxford (1956), 1935.
- [Trillas et al., 1982] E. Trillas, C. Alsina, and Ll. Valverde. Do we need max, min and 1-j in Fuzzy Set Theory?, in: R. Yager, ed., *Recent Developments of Fuzzy Set and Possibility Theory*, Pergamon, 1982.
- [Urquhart, 1986] A. Urquhart. Many-valued logic, in: D. Gabbay & F. Guentner, eds., *Handbook of Philosophical Logic* (Vol. 3), Reidel, 1986.
- [van Fraassen, 1968] B. C. van Fraassen. Presuppositions, implication and self-reference, *J. Phil.*, **65**, 1968.
- [van Fraassen, 1981] B. C. van Fraassen. Probabilistic semantics objectified I, II, *J. of Phil. Log.*, **10**, 371–394, 495–510, 1981.
- [Watanabe, 1969] S. Watanabe. Modified concepts of logic, probability and information based on generalized continuous characteristic function, *Info. and Control*, **15**, 1–21, 1969.
- [Wittgenstein, 1922] L. Wittgenstein. *Tractatus Logico-Philosophicus*, London, 1922.
- [Woodruff, 1995] P. W. Woodruff. Conditionals and generalized conditional probability, *10th Int. Congress of Logic, Methodology & Phil. of Sc.*, Aug. 19-25, 1995, Florence.
- [Yager, 1979] R. Yager. A note on fuzziness in a standard uncertainty logic, *IEEE Trans. Systems, Man & Cyb.*, **9**, 387–388, 1979.
- [Zadeh, 1965] L. A. Zadeh. Fuzzy sets, *Information and Control*, **8**, 338–353, 1965.
- [Zimmermann, 1977] H. J. Zimmermann. Results of empirical work on fuzzy sets, *Int. Conf. on Applied Gen. Systems Research*, Binghamton, NY, 1977.



# INDEX

- action and change in rewriting logic, 73
- Admissibility of cut
  - for intuitionistic implication, 209
- artificial intelligence, 331
  
- Bolzano, B., 327
- Boolean distance, 345
  
- CCS, 59
- Classical logic **C**, 205, 217
- classical probability, 324
- classical truth, 170
- complete, 297
- completeness, 186
- concurrent object-oriented programming, 56
- consequence relation, 287
- consistent, 297
- Contraction-free sequent calculi for intuitionistic logic, 230
  
- Dummett's argument, 171, 175
  
- entailment systems, 7
  
- Formula, 200
  - complexity, 200
  - substitution, 201
- fuzziness, 346
- fuzzy logic, 330, 358
  
- geometry of logic, 345
  
- Horn logic, 29
  
- ignorance as subadditivity, 354
- institution, 8
  
- internal negation, 290
- intuitionistic logics, 302
- Intuitionistic logic
  - axiomatization, 204
  - Kripke semantics, 205
  
- Lambek calculus, 259
- linear logic, 31, 303
- logic of rationality, 336
- Logic programming, 200, 207, 255
  - in linear logic, 275
- logical probability, 325
- Lukasiewicz, J., 352
  
- MacColl, H., 328
- Maude language, 18
- MaudeLog language, 18
- meaning for the (truth) value, 334
- meaning of the logical constants, 181
- meaning theories, 165, 177
- Modal logic
  - Kripke semantics, 231
- Modal logic **K**, 244
- Modal logic **K5**, 245
- Modal logic **K4**, 244
- Modal logic **K45**, 245
- Modal logic **KT**, 244
- Modal logic **S5**, 245
- Modal logic **S4**, 244, 264
- models of rewriting logic, 21
- multi-valued logic, 328
- multiplicative linear logic, 289
  
- Natural deduction
  - labelled, 257
  
- paraconsistent 3-valued logic PAC, 315

- paradox of inference, 179
- Peirce's axiom, 205
- Popper, K., 335
- probabilistic logic, 332
- probabilistic semantics, 335
- probability logic, 325
- proof calculi, 9
- proof theory for general assertions, 347
- propositional linear logic, 289
  
- quantifiers, 38
- Query
  - labelled, 233, 262
  
- $R_m$ , 306
- Realization, 268
- reflection, 13, 50
- relative truth, 343
- relevance logic, 289, 305
- Relevant logic **E**, 259
  - contractionless version **E-W**, 259
- relevant logic **T** (known as Ticket Entailment), 259
- Relevant logic **T** (Ticket Entailment)
  - contractionless version **T-W**, 259
- Resolution, 256
- Restart
  - bounded, 216
  - for classical logic, 217
  - for modal logics, 234
- rewriting logic, 4
- rewriting logic as a logical framework, 26
- $RMI$ , 308
  
- semantic framework, 54
- Sequent calculi
  - labelled, 257
- sequent calculi, 170
- sequent systems, 43
  
- smantics, 295
- standard uncertainty logic, 361
- structural operational semantics, 66
- subjective probability, 325
- Substructural logic **FL**, 259
- Substructural logics
  - axiomatization, 257, 269
  
- Tarski, A., 352
- Tarskian consequence relation, 295
- theories of meaning, 165
- theory morphism, 7
- three valued logics, 313
- three-valued logics, 352
- tonk, 179
- "truth" valuations, 340
- truth theories, 165
- type theory of Martin-Löf, 189
  
- universal algebra, 14